Notas de

Modelización matemática

Nota editorial.

Estos apuntes han sido redactados con inteligencia artificial a partir de múltiples materiales, bajo la dirección y revisión del autor. Pueden contener errores; ante cualquier discrepancia, manda el material oficial de la asignatura.

El temario sigue el proyecto docente de la asignatura *Modelización Matemática* del Grado en Matemáticas de la Universidad de Sevilla, y los exámenes resueltos proceden de sus convocatorias. Aun así, el contenido es el de un curso estándar de la materia y puede ser útil a estudiantes de otras universidades. Lo específico de cómo se imparte y se examina en Sevilla está reunido en el apéndice «Qué cae en el examen», que se puede saltar entero.

Versión de septiembre de 2026.

Índice general

Sesión cero de MATLAB **1**
Instalar MATLAB y decidir dónde se trabaja . . . 1
Tres formas de ejecutar una orden . . . 2
Variables y aritmética . . . 3
Vectores y matrices . . . 4
Funciones anónimas . . . 7
Valores lógicos, comparaciones y find . . . 8
M-funciones, ámbito de las variables y control de flujo . . . 9
Gráficos . . . 11
Leer datos de un fichero . . . 12
Resolver ecuaciones diferenciales: `ode45` . . . 13
Lo mismo en Python . . . 45

1. Algunos modelos discretos **47**
1.1. El modelo de Leslie . . . 48
1.1.1. Introducción . . . 48
1.1.2. Deducción del modelo . . . 51
1.1.3. Comportamiento a largo plazo de la población . . . 54
1.2. Medida de la relevancia de una página web: PageRank . . . 79
1.2.1. El modelo . . . 79
1.2.2. Cálculo del valor de PageRank mediante el método de la potencia 89
1.3. Prácticas de MATLAB del tema 1 . . . 95

2. Modelos con ecuaciones y sistemas diferenciales **119**
2.1. Modelos regidos por EDOs . . . 121
2.1.1. El modelo exponencial . . . 125
2.1.2. El modelo logístico . . . 130
2.1.3. El modelo de Gompertz . . . 136
2.1.4. Modelos de pesca . . . 138

2.2. Modelos regidos por SDOs . . . 144
2.2.1. Un segundo modelo de epidemias . . . 144
2.2.2. Epidemias con mortalidad: las dos leyes de incidencia . . . 148
2.2.3. Modelos vectoriales de poblaciones . . . 152
2.2.4. Modelo de Volterra (presa-depredador simplificado) . . . 153
2.2.5. Modelo de competición (simplificado) . . . 160
2.2.6. Modelo general de competición . . . 166
2.2.7. Modelo de simbiosis . . . 171
2.3. Prácticas de MATLAB del tema 2 . . . 175

3. Modelos obtenidos a partir del Cálculo de Variaciones 204
3.1. Principio de Lagrange . . . 205
3.2. Algunos problemas geométricos . . . 220
3.3. Aplicaciones en Mecánica (I): Principio de mínima acción . . . 229
3.3.1. Movimiento de una partícula en un campo conservativo . . . 231
3.3.2. Movimiento del péndulo . . . 233
3.3.3. Movimiento de una varilla . . . 240
3.4. Aplicaciones en Mecánica (II): Problemas de energía mínima . . . 242
3.4.1. Cuerda elástica: elasticidad lineal . . . 242
3.5. Prácticas de MATLAB del tema 3 . . . 247

4. Problemas de control óptimo para EDOs 280
4.1. Problemas básicos (I): EDO lineal sin restricciones con observación en $(0, T)$ 281
4.2. Problemas básicos (II): EDO lineal sin restricciones con observación final . . 295
4.3. Sistema de EDOs lineales sin restricciones . . . 300
4.4. EDO lineal de segundo orden . . . 307
4.5. EDO no lineal sin restricciones . . . 310
4.6. EDO lineal con restricciones de desigualdad . . . 316
4.7. Hamiltoniano . . . 321
4.8. Nuevos problemas más complicados . . . 324
4.8.1. Sistema de EDOs no lineales con restricciones . . . 325
4.8.2. EDO con restricciones de otro tipo . . . 326
4.8.3. Control óptimo bilineal . . . 327
4.9. El problema inverso: identificar parámetros por mínimos cuadrados . . . 330
4.10. Prácticas de MATLAB del tema 4 . . . 336

A. Complementos sobre ecuaciones y sistemas diferenciales ordinarios 357
A.1. Sistemas no autónomos . . . 358
A.1.1. Sistemas de primer orden en forma normal . . . 358
A.1.2. Ecuación de orden superior en forma normal . . . 359
A.1.3. Sistemas lineales . . . 359
A.1.4. Comparación de SDOs . . . 360
A.2. Sistemas autónomos . . . 361
A.2.1. Sistemas de primer orden en forma normal . . . 362
A.2.2. Sistema lineal homogéneo . . . 362
A.2.3. EDOs lineales de orden n con coeficientes constantes . . . 362
A.3. Estabilidad de SDOs autónomos . . . 363

A.3.1. Estabilidad de SDOs lineales de coeficientes constantes 364
A.3.2. Estabilidad de algunos SDOs autónomos no lineales 365
A.4. El lema de Gronwall 370

B. Qué cae en el examen **372**
B.1. Cómo se califica en 2026-27 372
B.2. El archivo de exámenes 372
B.3. La parte práctica, en el aula de informática 373
B.4. Qué hacer con ello 374

C. Exámenes resueltos **375**
C.1. Instrucciones oficiales 375
C.2. Simulacro de examen 376
C.3. Primera convocatoria de 2026: un modelo SIR con mortalidad 381
C.4. Examen final del 2 de junio de 2025 (primera convocatoria de 2024-25) . . . 392
C.5. Primera prueba de seguimiento del 5 de abril de 2024 412
C.6. Segunda convocatoria del 13 de julio de 2023 422
C.7. Primer parcial del 29 de marzo de 2023 430
C.8. Primer parcial del 28 de marzo de 2022 437
C.9. Ejercicios sueltos de las convocatorias de 2016 y 2017 443
C.9.1. Modelos de poblaciones 443
C.9.2. Cálculo de variaciones: extremos libres 449
C.9.3. Control óptimo: tres sistemas de optimalidad 453
C.9.4. Mecánica lagrangiana: tres sistemas 458
C.9.5. Prácticas: tres poblaciones y un problema inverso 464

D. Añadidos y omisiones **473**
Lo que el libro añade y no está en los apuntes 474
Lo del profesor que no está en el libro 475
Erratas detectadas 476

E. Cómo usar este libro **478**
Para quién es y desde dónde parte 478
La estructura 478
El código visual 479
Los ejercicios 479
Qué hacer si algo no cuadra 479

Sesión cero de MATLAB

EL CAMINO QUE NO SE TOMÓ: QUÉ ES ESTO Y POR QUÉ ESTÁ AQUÍ

Esta sesión introduce las herramientas de MATLAB que se usarán en las prácticas. Las construcciones más específicas se explican cuando aparecen. Se conserva sin número para mantener la correspondencia entre capítulos y temas del curso.
El material sale de dos sitios: de la carpeta de preparación de MATLAB del curso —su repaso comentado, sus cuatro ficheros de ejemplo y el guión de ecuaciones diferenciales, que es el que sostiene la última sección— y del capítulo *Introducción a MATLAB* del libro previo del autor, del que se conservan las explicaciones y el orden de las ideas.

LA PREGUNTA

Un biólogo tiene un censo de una población de tortugas repartido en cuatro clases de edad y sabe qué fracción de cada clase sobrevive al año siguiente. Quiere saber cuántas tortugas habrá dentro de cien años, y después repetir la cuenta con cincuenta políticas de protección distintas. ¿Cómo puede programar el cálculo para repetirlo cambiando solo los datos?

Instalar MATLAB y decidir dónde se trabaja

MATLAB es un programa de pago de la empresa MathWorks. Se descarga desde su página web con una cuenta asociada a la licencia de la universidad, que es la que se usa en las aulas de informática de la asignatura. Las prácticas ofrecen también versiones en Python, que es gratuito. Para preparar la prueba del curso se necesita practicar con MATLAB, el programa indicado en sus instrucciones.

Los paneles principales para trabajar son:

- **Command Window**, la ventana de órdenes. Tiene un símbolo » y espera que se escriba algo. Lo que se escribe se ejecuta al pulsar Intro.
- **Workspace**, el espacio de trabajo. Lista las variables que existen en este momento, con su valor.

- **Current Folder**, la carpeta actual. Es el punto de partida de las rutas relativas. MATLAB también busca funciones en su ruta de búsqueda; para los ejemplos basta seleccionar la carpeta que contiene el código y sus datos.
- **Editor**, donde se escriben los ficheros de código.

Conviene guardar cada ejercicio en una carpeta con su enunciado, su código y sus datos, y seleccionarla como carpeta actual. Los nombres descriptivos, como `solucion_leslie.m` o `biseccion.m`, ayudan a localizar y reutilizar el trabajo.

Dos reglas sobre nombres que conviene respetar desde el primer día (véase también el apéndice B):

1. Los nombres de scripts y funciones empiezan por una letra y continúan con letras, dígitos o guiones bajos; después se añade la extensión `.m`. Evitar espacios y acentos en las carpetas facilita compartir el código, aunque no sea una prohibición general del sistema de archivos. Las variables distinguen mayúsculas y minúsculas.
2. Las instrucciones de entrega del curso piden una carpeta con los apellidos sin espacios ni acentos, una línea de comentario con el nombre en cada programa y la carpeta comprimida al terminar. Son reglas de entrega; se recogen en el apéndice citado.

Tres formas de ejecutar una orden

Hay tres maneras de decirle a MATLAB que haga algo, y conviene distinguirlas desde el principio porque las tres se usan a lo largo del libro.

1. **En modo interactivo**, escribiendo una orden tras otra en la ventana de órdenes. Sirve para probar cosas y consultar variables. Las órdenes quedan en el historial; para conservar un procedimiento reproducible conviene guardarlo en un script.
2. **En un fichero** `.m`, que es un fichero de texto con una lista de órdenes. Si empieza directamente por órdenes es un *script*: al ejecutarlo se hace todo lo que contiene, en orden, como si se hubiera tecleado. Si empieza por la palabra `function` es una *M-función*, y de ellas se habla más abajo. Todo lo que se entrega en la asignatura son ficheros `.m`.
3. **En un fichero** `.mlx`, un *Live Script* que mezcla código, texto con formato y resultados en un mismo documento. Es cómodo para un repaso comentado, porque el texto y la salida van juntos. Aquí se usa `.m`: es texto plano y se lee en cualquier editor; la compatibilidad depende de las funciones utilizadas.

Para ejecutar un script basta escribir su nombre, sin la extensión, en la ventana de órdenes, o pulsar el botón *Run* del editor. En el editor, `Ctrl+Intro` (en macOS, `Cmd+Intro`) ejecuta la *sección* actual, que es el trozo de fichero delimitado por líneas que empiezan por `%%`; si no hay ninguna, ejecuta el fichero entero.

Tres órdenes de limpieza que se usan a cada rato: `clc` borra la pantalla de la ventana de órdenes, `clear` borra todas las variables del espacio de trabajo, y `close all` cierra todas las ventanas de gráficos. Cuando un programa no responde, `Ctrl+C` en la ventana de órdenes lo interrumpe.

Y la orden que hay que conocer antes que ninguna otra: `help nombre` escribe la ayuda de la orden `nombre`, y `helpwin nombre` la abre en una ventana aparte. Merece la pena acostumbrarse a ella: quien sabe leer `help polyfit` no necesita recordar el orden de los argumentos de ninguna orden.

Variables y aritmética

Una variable de MATLAB es una zona de memoria con un nombre. No es una variable matemática: por eso tiene sentido escribir `a = a + 2`, que en matemáticas sería falso y aquí significa «toma lo que hay en `a`, súmale 2 y guárdalo otra vez en `a`».

```
a = 1; % el punto y coma evita que MATLAB escriba el resultado
a = a + 2;
a % sin punto y coma: MATLAB muestra el valor
```

SALIDA ESPERADA

```
a =
     3
```

El punto y coma al final de una línea es la diferencia entre un programa legible y una pantalla llena de números. Se pone siempre, salvo cuando se quiere ver el resultado.

Las variables se distinguen por mayúsculas y minúsculas: `Inter` e `inter` son dos variables distintas, y también lo son `A0` (con un cero) y `AO` (con la letra o). Este último par avisa de un error que cuesta media hora encontrar.

```
Inter = 0.1;
inter = 0.2;
A0 = 0.3;
AO = 0.4;
whos % lista las variables con su tama\~no y su tipo
```

SALIDA ESPERADA

```
  Name Size Bytes Class Attributes
  A0 1x1 8 double
  AO 1x1 8 double
  Inter 1x1 8 double
  a 1x1 8 double
  inter 1x1 8 double
```

Los operadores aritméticos son los esperables: `+`, `-`, `*`, `/` y `^` para la potencia. Las funciones elementales tienen los nombres habituales: `sqrt`, `exp`, `log` (logaritmo neperiano), `sin`, `cos`, `abs`. La constante π es `pi`. Los valores `double` tienen unas quince o dieciséis cifras significativas de precisión. `format short` muestra cuatro decimales y `format long`, quince; ambos cambian la presentación, no la precisión del cálculo.

```
x = sqrt(2)
format long
x
format short
```

SALIDA ESPERADA

```
x =
    1.4142
x =
   1.414213562373095
```

Vectores y matrices

En los cálculos de esta sección se trabaja con matrices: un escalar es una matriz 1×1 y un vector tiene una sola fila o columna. Esta organización da nombre al programa (*MATrix LABoratory*); MATLAB admite además otros tipos de datos.

Un vector se escribe entre corchetes, separando las componentes con comas o con espacios. El punto y coma dentro de los corchetes cambia de fila. La comilla simple traspone y conjuga: para datos reales es la traspuesta habitual. Para trasponer sin conjugar se usa `.'`.

```
v = [1, 2, 3] % vector fila
w = [1; 2; 3] % vector columna
A = [1, 2; 3, 4] % matriz 2 x 2, fila a fila
v' % traspuesta: convierte la fila en columna
```

SALIDA ESPERADA

```
v =
     1 2 3
w =
     1
     2
     3
A =
     1 2
     3 4
ans =
     1
     2
     3
```

Cuando una expresión no se asigna a ninguna variable, MATLAB guarda el resultado en una variable llamada `ans`.

Dos maneras de construir vectores largos sin escribirlos. El operador dos puntos, `a:h:b`, genera los números desde `a` hasta `b` de `h` en `h`; si se omite `h`, el paso es 1. Y `linspace(a,b,n)` genera `n` puntos igualmente espaciados entre `a` y `b`; si se omite `n`, son 100. La primera se usa cuando importa el paso; la segunda, cuando importa el número de puntos, que es lo habitual para dibujar.

```
1:5
0:0.25:1
linspace(0, 1, 5)
```

SALIDA ESPERADA

```
ans =
     1 2 3 4 5
ans =
         0 0.2500 0.5000 0.7500 1.0000
ans =
         0 0.2500 0.5000 0.7500 1.0000
```

Para acceder a una componente se usan paréntesis, y **los índices empiezan en 1**, no en 0. La palabra `end` dentro de un índice significa «el último». Los dos puntos solos significan «todos».

```
x = [5, 0, -1, 3, 7];
x(2) % segunda componente
```

```
x(end) % la ultima
x(2:4) % de la segunda a la cuarta
A = [1, 2, 3; 4, 5, 6];
A(2, 3) % fila 2, columna 3
A(1, :) % toda la fila 1
A(:, 2) % toda la columna 2
```

SALIDA ESPERADA

```
ans =
     0
ans =
     7
ans =
     0 -1 3
ans =
     6
ans =
     1 2 3
ans =
     2
     5
```

Tres funciones para preguntar por el tamaño: `size(A)` devuelve el número de filas y de columnas; `length(v)` devuelve la dimensión mayor; `numel(A)` devuelve el número total de elementos. Para un vector, `length` y `numel` coinciden, y solo difieren en matrices. Y cuatro funciones para construir matrices especiales: `zeros(n,m)` y `ones(n,m)` llenan de ceros o de unos, `eye(n)` es la identidad, y `diag(v)` pone el vector `v` en la diagonal de una matriz de ceros. Esta última admite un segundo argumento: `diag(v,-1)` pone `v` en la subdiagonal, que es exactamente lo que necesita la matriz de Leslie del capítulo 1.

```
size(A)
numel(A)
diag([0.25, 0.5, 0.1], -1)
```

SALIDA ESPERADA

```
ans =
     2 3
ans =
     6
ans =
         0 0 0 0
    0.2500 0 0 0
         0 0.5000 0 0
         0 0 0.1000 0
```

La confusión más habitual: operar elemento a elemento

El asterisco entre dos matrices es el **producto de matrices**, con la regla de filas por columnas. Si lo que se quiere es multiplicar componente a componente, hay que poner un punto delante: `.*`, `./` y `.^`. Es la confusión más frecuente al empezar, y el síntoma es un error de dimensiones donde uno no lo esperaba. La suma y la resta no necesitan punto, porque solo tienen un significado.

```
v = [1, 2, 3];
v .^ 2 % cada componente al cuadrado
```

```
v * v' % producto de matrices: fila por columna, un numero
v .* v % producto componente a componente
```

SALIDA ESPERADA

```
ans =
     1 4 9
ans =
    14
ans =
     1 4 9
```

Escribir `v^2` da error, porque elevar al cuadrado una matriz que no es cuadrada no tiene sentido. En cambio, con una matriz cuadrada `L`, `L^2` es `L*L`, y `L^k` calcula la potencia k-ésima: es la orden con la que en el capítulo 1 se obtiene $L^k v(0)$.

Funciones que resumen un vector

`sum`, `max`, `min`, `mean` (media) y `std` (desviación típica) hacen lo que dicen. Dos detalles que importan. Primero, `max` y `min` pueden devolver dos cosas: el valor y la posición en que se alcanza. Para recoger las dos se escriben dos nombres entre corchetes a la izquierda del igual. Segundo, aplicadas a una matriz con varias filas, actúan **por columnas** y devuelven un resultado por columna; para un vector fila, reducen toda la fila a un valor. PageRank usa esta regla para sumar las columnas de su matriz. Por defecto, `std` usa el divisor $n-1$ si $n > 1$.

```
x = [5, 0, -1, 3, 7];
sum(x)
[m, pos] = max(x)
mean(x)
norm(x) % norma euclidea, raiz de la suma de cuadrados
norm(x, 1) % norma uno, suma de valores absolutos
```

SALIDA ESPERADA

```
ans =
    14
m =
     7
pos =
     5
ans =
    2.8000
ans =
    9.1652
ans =
    16
```

Funciones anónimas

Una función sencilla, de las que devuelven el valor de una expresión, se puede definir con una sola línea, en mitad de un programa o en la ventana de órdenes. Se llama función anónima y su sintaxis es

```
nombre = @(argumentos) expresion
```

Los ejemplos que siguen son los del fichero `Practica3.m` del curso. El primero define $f(x) = \sqrt{x^2 + 10}$.

```
ffun = @(x) sqrt(x.^2 + 10);
ffun(20)
log(ffun(5)) + 1
ffun([1, 2, 3])
```

SALIDA ESPERADA

```
ans =
   20.2485
ans =
    2.7777
ans =
    3.3166 3.7417 4.3589
```

Obsérvese el punto en `x.^2`. Gracias a él, la función admite un vector como argumento y devuelve el vector de imágenes. Sin el punto, `ffun([1,2,3])` daría error. Esta es la razón de que casi todas las funciones anónimas del libro lleven puntos: se escriben pensando en que se van a evaluar sobre una rejilla de puntos para dibujar.

En la expresión se pueden usar variables definidas antes. Su valor queda **congelado** en el momento de definir la función: si después se cambia `a`, la función sigue usando el valor antiguo.

```
a = 10;
ffun = @(x) sqrt(x.^2 + a);
a = 1000; % no afecta a ffun
ffun(0)
```

SALIDA ESPERADA

```
ans =
    3.1623
```

Un polinomio se maneja con dos órdenes que reaparecen en la primera práctica del capítulo 1. Un polinomio se representa por el vector de sus coeficientes **de mayor a menor grado**, y `polyval(coef, x)` lo evalúa en `x`, que puede ser un número o un vector. Así, $p(x) = x^3 + 3x^2 - x + 4$ es:

```
coef = [1, 3, -1, 4];
pfun = @(x) polyval(coef, x);
pfun(6)
```

SALIDA ESPERADA

```
ans =
   322
```

Las funciones anónimas pueden tener varias variables y pueden devolver vectores. Dos ejemplos: $h(x, y) = \max\{\|x\|_2^2, y^2\}$ y $g(x) = (x^2 - 1,\ 1 - x,\ x + 2)$:

```
hfun = @(x, y) max(norm(x)^2, y^2);
hfun(1, 2)
gfun = @(x) [x^2 - 1; 1 - x; x + 2];
gfun(-1)
```

SALIDA ESPERADA

```
ans =
     4
ans =
     0
     2
     1
```

La función `gfun` devuelve un vector columna, escrito con puntos y coma entre las componentes. En el capítulo 2, el segundo miembro de un sistema diferencial $y' = f(t, y)$ se escribirá exactamente así: una función anónima de dos argumentos que devuelve una columna.

Valores lógicos, comparaciones y `find`

Un valor lógico es `true` (verdadero) o `false` (falso). Los producen las comparaciones ==, <, >, <=, >= y ~= (distinto), y se combinan con `&` (y) y `|` (o). El doble igual es la comparación; el igual simple es la asignación, y confundirlos es un error clásico. Los ejemplos son los del fichero `Practica4.m` del curso.

```
3 == 7
a = 7;
a <= 3
x = [5, 0, -1, 3, 7];
x < 4
true & false
n = 6; m = 2;
n > 2 & m < 3
```

SALIDA ESPERADA

```
ans =
  logical
   0
ans =
  logical
   0
ans =
  1x5 logical array
   0 1 1 1 0
ans =
  logical
   0
ans =
  logical
   1
```

MATLAB muestra los valores lógicos como 1 y 0, pero su tipo es `logical`. Algunas operaciones los convierten automáticamente en números; otras funciones, como `sin`, requieren una conversión previa:

```
v = 3;
z = v < 5
```

```
z + 1
sin(z)  % da error
sin(z + 0)  % sumar 0 lo convierte en numero
```

SALIDA ESPERADA

```
z =
  logical
   1
ans =
    2
Undefined function 'sin' for input arguments of type 'logical'.
ans =
   0.8415
```

Esa aritmética con lógicos tiene un uso que conviene aprender: definir una función a trozos en una sola línea. Para $f(x) = 2$ si $x < 0$ y $f(x) = \cos x + 1$ si $x \geq 0$, se multiplica cada trozo por la condición que lo activa; la condición vale 1 donde se cumple y 0 donde no.

```
fun = @(x) 2*(x < 0) + (cos(x) + 1).*(x >= 0);
fun(-1)
fun(pi)
```

SALIDA ESPERADA

```
ans =
    2
ans =
    0
```

Por último, `find` devuelve las **posiciones** de las componentes que cumplen una condición, y un vector lógico usado como índice devuelve las **componentes** mismas. Las dos cosas se usan en la práctica de las cotizaciones del capítulo 1, para saber qué días el precio superó un umbral.

```
x = [5, 0, -1, 3, 7];
find(x < 4)  % posiciones
ind = x < 4;
x(ind)  % valores
```

SALIDA ESPERADA

```
ans =
    2 3 4
ans =
    0 -1 3
```

M-funciones, ámbito de las variables y control de flujo

Además de las funciones que trae MATLAB, se pueden crear otras nuevas, llamadas **M-funciones**, que se guardan cada una en un fichero **con el mismo nombre que la función** y extensión `.m`. Para que MATLAB la encuentre, el fichero tiene que estar en la carpeta actual o en su ruta de búsqueda. Una M-función tiene argumentos de entrada, por los que se le pasan valores, y argumentos de salida, por los que devuelve resultados. El ejemplo más corto, guardado en un fichero `suma.m`:

```
function [s] = suma(x, y)
% Esta funcion suma dos numeros
```

```
s = x + y;
```

Aquí `x` e `y` son los argumentos de entrada y `s` el de salida. La línea de comentario que va justo debajo de la cabecera es lo que muestra `help suma`. Desde la ventana de órdenes:

```
help suma
suma(3, 7)
suma(3/5, sqrt(23))
```

SALIDA ESPERADA

```
  Esta funcion suma dos numeros
ans =
    10
ans =
    5.3958
```

Si dentro de la función no se asigna ningún valor al argumento de salida, la función no devuelve nada. Conviene comprobarlo cambiando la última línea por `w = x + y;`: entonces `suma(3,7)` termina con un error que dice que la variable de salida `s` no está asignada.

El ámbito de las variables

Las variables creadas en la ventana de órdenes viven en el espacio de trabajo principal. Un script usa el espacio desde el que se ejecuta; una función tiene su propio espacio de variables locales. Es lo que muestra el fichero `mifun.m` del curso:

```
function mifun
%
% mifun sirve para mostrar cual es el
% ambito de las variables de una M-funcion
%
b = 5;
B = 2*b;
```

Tras ejecutar `mifun` en la ventana de órdenes, las variables `b` y `B` **no están** en el espacio de trabajo: `whos` no las lista. En cambio, si las mismas dos líneas se guardan como script `program1.m` y se ejecuta, las dos variables aparecen. Los argumentos permiten comunicar una función con el exterior sin mezclar sus variables locales con las del programa que la llama.

Conviene evitar nombres que **oculten funciones existentes**. Si se escribe `sin = 7`, entonces `sin(pi)` intenta indexar esa variable con π y da error; `clear sin` elimina la variable. Un fichero propio llamado `sin.m` puede ocultar la función de MATLAB. Del mismo modo, llamar `abs` a unas abscisas oculta la función valor absoluto.

Condicionales y bucles

El fichero `Maximo.m` del curso enseña el condicional `if–end`:

```
function [m] = Maximo(a, b)
%-------------------------------------------------------------
% m = Maximo(a,b) devuelve el maximo entre dos numeros
%-------------------------------------------------------------
m = a;
if b > a
  m = b;
end
```

La estructura general es `if condicion`, órdenes, y opcionalmente `elseif` otra condición u `else`, cerrando siempre con `end`. Los bucles tienen dos formas. El bucle `for` repite un bloque para cada valor de un vector, y es el que se usa para iterar $v(k+1) = Lv(k)$ veinte veces; el bucle `while` repite mientras se cumpla una condición, y es el que se usa cuando no se sabe de antemano cuántas iteraciones harán falta, como en el método de la potencia con un test de parada. La orden `break` sale del bucle y `return` sale de la función.

```
s = 0;
for k = 1:4
  s = s + k^2;
end
s
k = 0;
while 2^k < 100
  k = k + 1;
end
k
```

SALIDA ESPERADA

```
s =
    30
k =
     7
```

Escribir resultados

`disp(x)` escribe el valor de `x` sin el nombre. `num2str(x)` convierte un número en texto, para poder pegarlo a una frase entre corchetes. Y `fprintf` escribe con formato: el texto entre comillas lleva marcas que se sustituyen por los argumentos que siguen; `%i` es un entero, `%f` un real con seis decimales, `%9.4f` un real en un campo de al menos nueve caracteres con cuatro decimales, y `\n` es un salto de línea. `fprintf` reaparece para imprimir tablas en la práctica de Leslie, y conviene entender su formato antes de llegar allí.

```
N = 20;
disp(['N = ', num2str(N)])
fprintf('Autovalor dominante = %12.8f, es el numero %4i \n', 0.85142496, 1)
fprintf(' %3i %9.4f %9.4f \n', 0, 0.3571, 0.2500)
```

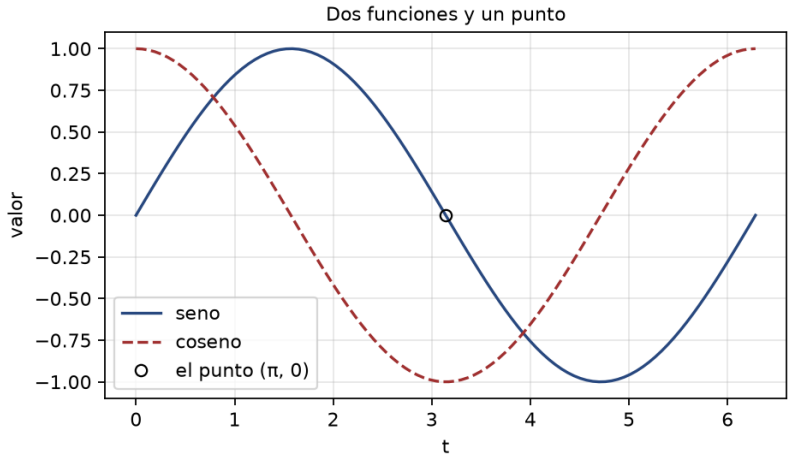
SALIDA ESPERADA

```
N = 20
Autovalor dominante = 0.85142496, es el numero 1
   0 0.3571 0.2500
```

Gráficos

Estas son las órdenes básicas para representar las soluciones. `plot(x,y)` dibuja los puntos de abscisas `x` y ordenadas `y` unidos por segmentos; si `x` es una rejilla fina, se ve una curva. `hold on` mantiene la figura para superponer más curvas y `hold off` la libera; sin `hold on`, el segundo `plot` borra el primero. `legend` pone las etiquetas **en el mismo orden en que se llamó a** `plot`. `xlabel`, `ylabel` y `title` rotulan, y `grid on` pone la rejilla. `figure` abre una ventana nueva, para no pisar la anterior.

```
t = linspace(0, 2*pi);
figure
plot(t, sin(t))
hold on
plot(t, cos(t), 'r--') % rojo y a trazos
plot(pi, 0, 'ko') % un punto negro con circulo
hold off
legend('seno', 'coseno', 'el punto (pi, 0)')
xlabel('t')
ylabel('valor')
title('Dos funciones y un punto')
grid on
```



Resultado del listado anterior: el seno (continuo), el coseno (a trazos) y el punto $(\pi, 0)$ marcado con un círculo.

Las figuras de esta sesión están dibujadas con el equivalente en Python de cada listado: `ode45` se sustituye por el Runge-Kutta 4(5) de Dormand-Prince de SciPy, que es el mismo método con las mismas tolerancias. En MATLAB cambian los colores y el aspecto de la ventana; las curvas son las mismas.

El tercer argumento de `plot` es una cadena con el color (`r` rojo, `g` verde, `b` azul, `k` negro, `m` magenta), el trazo (`-` continuo, `--` a trazos, `:` punteado) y el marcador (`o` círculo, `*` asterisco, `.` punto, `+` cruz). Un par de opciones que aparecen mucho en los ficheros del curso: `'LineWidth', 1.5` engorda la línea y `'MarkerSize', 15` agranda el marcador. La orden `shg` trae la ventana de gráficos al frente.

Leer datos de un fichero

La práctica de las cotizaciones del capítulo 1 parte de un fichero de texto, `Finanzas.dat`, con un número por línea. Un fichero así se lee con `load`, que devuelve directamente un vector columna. En la llamada siguiente, el fichero se busca por su nombre; guardarlo junto al script facilita encontrarlo.

```
A = load('Finanzas.dat');
size(A)
A(1:3)'
```

SALIDA ESPERADA

```
ans =
    88 1
ans =
       27046 27040 27020
```

Resolver ecuaciones diferenciales: `ode45`

Muchas prácticas del libro requieren aproximar soluciones de ecuaciones diferenciales. Esta sección introduce el integrador y las comprobaciones que permiten interpretar su salida.

La orden es `ode45`, y resuelve problemas de valor inicial

$$y' = f(t,y), \quad t \in [t_0, t_1], \qquad y(t_0) = y_0,$$

en su forma más simple así:

```
f = @(t,y) sin(t)*y + cos(t); % segundo miembro, siempre @(t,y)
ode45(f, [0, 3], 1) % resolver en [0,3] con y(0) = 1
```

Tres argumentos y en este orden: la función $f(t,y)$ como función anónima de *dos* variables (aunque no dependa de t, la t va delante y no se puede quitar), el intervalo $[t_0, t_1]$ como vector de dos componentes, y el dato inicial y_0. Llamada así, sin pedirle nada a la izquierda, `ode45` **dibuja** la solución y no devuelve ningún número.

LA IDEA: QUÉ ES LO QUE DEVUELVE DE VERDAD

Esta es la idea que hay que tener clara antes de seguir, porque de ella salen las tres formas de usar la orden. `ode45` no calcula «la solución»: calcula sus valores en un **conjunto finito de instantes**. En una integración hacia adelante que alcanza su extremo final, estos cumplen $t_0 = \tau_0 < \cdots < \tau_m = t_1$ y no suelen estar igualmente espaciados. El método adapta el paso según una estimación del error local. Los números 4 y 5 identifican los órdenes de las dos fórmulas de Runge-Kutta que se comparan para estimarlo.

De ahí dos consecuencias prácticas. La primera: si se quiere el valor en un instante concreto, casi seguro que ese instante *no* está en la malla, y hay que interpolar. La segunda: la malla depende de las tolerancias: dos ejecuciones pueden devolver vectores de distinta longitud. Por eso, la media de los valores de una malla no uniforme no representa sin más una media respecto del tiempo.

Las tres formas de llamar a `ode45`

Primera: dibujar. `ode45(f, [t0, t1], y0)` sin salida. Sirve para mirar y no para calcular. Para superponer varias soluciones se usa `hold on`, igual que con `plot`.

Segunda: los valores. `[t, y] = ode45(f, [t0, t1], y0)` no dibuja nada y devuelve dos vectores columna, los instantes de la malla y los valores en ellos. El último valor se lee con `y(end)`. Solo aproxima $y(t_1)$ si la integración ha llegado a t_1: se comprueba mirando `t(end)`. Si el dato inicial es un vector de varias componentes, `y` es una matriz que guarda **una componente por columna** y un instante por fila.

Tercera: la solución como objeto. `sol = ode45(f, [t0, t1], y0)` devuelve una *estructura*, es decir, una colección de datos con un solo nombre y con campos a los que se llega con un punto: `sol.x` contiene los instantes y cada columna de `sol.y` contiene el estado en uno de ellos. Lo valioso de esta forma es que la estructura se puede evaluar dentro del intervalo integrado con

```
deval(sol, tk)
```

que devuelve el valor en `tk` usando el mismo polinomio interpolador que el método usa por dentro. Es la forma que más se repite en el libro, porque casi siempre hace falta el valor en instantes que no son los de la malla.

Hay una cuarta variante que no es una forma nueva sino un cambio en un argumento: si en lugar del intervalo $[t_0, t_1]$ se pasa un **vector de instantes**, la salida se da exactamente en esos instantes y no en la malla interna.

```
f = @(t,y) y - y^3;
[t, y] = ode45(f, 1:0.1:2, 0.5); % salida en 1, 1.1, 1.2, ..., 2
```

El método sigue usando pasos adaptativos e interpola en los instantes pedidos; estos pueden influir en la elección del paso inicial, por lo que la malla interna no tiene por qué coincidir. Con solo dos componentes, el vector fija los extremos del intervalo. Para evaluar después en dos instantes concretos puede usarse `deval`.

LA IDEA: POR QUÉ HAY QUE APRETAR LAS TOLERANCIAS ANTES DE PUBLICAR UNA CIFRA

Por defecto, `ode45` usa tolerancias relativa 10^{-3} y absoluta 10^{-6}. Controlan una estimación del error local, no el error total de la solución. Para pedir más precisión se usa `opciones = odeset('RelTol', 1e-8, 'AbsTol', 1e-10)` y se pasa `opciones` como cuarto argumento. Antes de dar varias cifras conviene contrastarlas con una solución exacta o repetir el cálculo con tolerancias menores. El ejercicio 4 muestra cómo hacerlo.

Varias soluciones de la misma ecuación

Las dos primeras prácticas de EDO son la misma tarea con dos ecuaciones distintas: dibujar juntas las soluciones que salen de varios datos iniciales. Merece la pena hacerlas seguidas porque el dibujo de la primera es tranquilizador y el de la segunda es una advertencia.

Ejercicio 0.1. *Tres soluciones de $y' = y - y^3$*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 1 de la parte de EDO.

Dibujar en la misma ventana gráfica las soluciones del problema de valor inicial

$$y' = y - y^3, \qquad t \in [1, 5],$$

correspondientes a los datos iniciales $y(1) = 0.5$, $y(1) = 1$ e $y(1) = 2$.

Resolución. La idea, y por qué se integra con paso variable. Lo que se pide no es «dibujar una ecuación»: es comprobar con el ordenador algo que ya se sabe sin él. Por eso el ejercicio va en dos tiempos —primero qué tiene que salir, después el programa— y el programa solo sirve para confirmarlo o para descubrir que uno leyó mal el signo.

Se usa `ode45`, un método de Runge-Kutta encajado de órdenes 4 y 5. Compara dos aproximaciones que comparten evaluaciones de la función para estimar el error local y ajustar el paso. Así puede dedicar más pasos a los tramos que lo requieren y menos a los que son más suaves, sin imponer una malla uniforme.

Qué va a salir, antes de programar. La ecuación es autónoma y escalar, así que basta mirar el signo de $f(y) = y - y^3 = y(1-y)(1+y)$. Sus ceros son -1, 0 y 1, que son los equilibrios; entre 0 y 1 es positiva y por encima de 1 es negativa. Los tres datos del enunciado son positivos, luego: la que sale de 0.5 sube hacia 1, la que sale de 1 se queda quieta, y la que sale de 2 baja hacia 1. Las tres curvas se apilan sin cortarse —la unicidad lo prohíbe— y las dos de fuera se pegan a la de en medio. Este es el teorema de las EDO autónomas escalares del capítulo 2, y aquí se ve dibujado antes de demostrarlo.

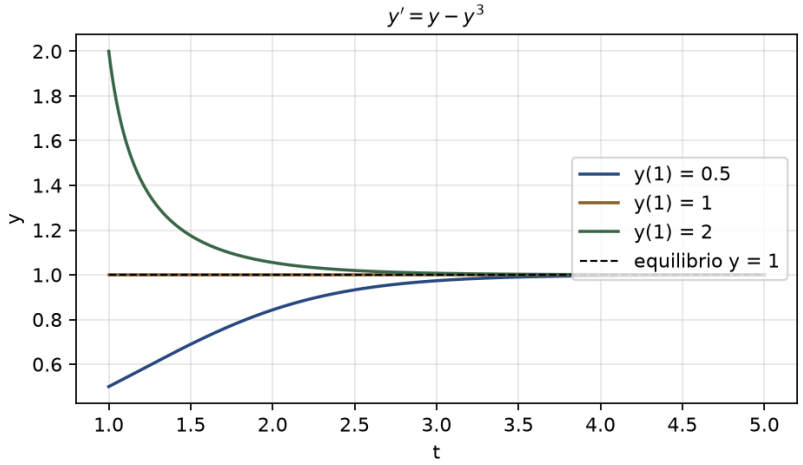
El programa. El bucle recorre los tres datos iniciales; `hold on` permite superponer sus curvas. La definición de f expresa el modelo y la llamada a `ode45` lo integra; las demás líneas fijan datos y tolerancias o presentan los resultados.

```
%% Ejercicio 1 - tres soluciones de y' = y - y^3 en [1,5]
f = @(t,y) y - y^3;
y0s = [0.5, 1, 2];
opc = odeset('RelTol', 1e-8, 'AbsTol', 1e-10);

figure, hold on
for k = 1:length(y0s)
  [t, y] = ode45(f, [1, 5], y0s(k), opc);
  plot(t, y, 'LineWidth', 1.4)
  fprintf('y(1) = %4.1f -> y(5) = %.6f \n', y0s(k), y(end))
end
plot([1, 5], [1, 1], 'k--') % el equilibrio y = 1
hold off
xlabel('t'), ylabel('y'), title('y'' = y - y^3'), grid on
legend('y(1) = 0.5', 'y(1) = 1', 'y(1) = 2', 'equilibrio y = 1', 'Location', 'east')
```

SALIDA ESPERADA

```
y(1) = 0.5 -> y(5) = 0.999497
y(1) = 1.0 -> y(5) = 1.000000
y(1) = 2.0 -> y(5) = 1.000126
```



Las tres soluciones de $y' = y - y^3$ en $[1, 5]$. Las dos de fuera se pegan a la de en medio, que es el equilibrio $y = 1$.

Las dos soluciones de fuera están ya a menos de una milésima del equilibrio en $t = 5$, y por lados opuestos. La central se mantiene exactamente en 1: el segundo miembro se anula allí y los incrementos calculados son cero. `1.000000` es su formato de salida.

Otra vía (la solución exacta). La ecuación es de Bernoulli, y el cambio $u = y^{-2}$ la vuelve lineal: $u' = -2y^{-3}(y - y^3) = -2u + 2$, de donde $u = 1 + Ce^{-2t}$ y

$$y(t) = \frac{1}{\sqrt{1 + (y_0^{-2} - 1)e^{2-2t}}}.$$

Con $y_0 = 0.5$ da $y(5) = 0.99949719$ y con $y_0 = 2$ da $y(5) = 1.00012582$, que son los dos números de la salida. Y la fórmula dice algo que la figura esconde: para $y_0 = 2$ el radicando es $1 - \frac{3}{4}e^{2-2t}$, que se anula en $t = 1 - \frac{1}{2}\log\frac{4}{3} = 0.856159$. Es decir, esa solución **explota hacia atrás** justo antes del instante inicial, y el intervalo $[1, 5]$ del enunciado es lo que hay a la derecha de la explosión. Hacia adelante no pasa nada, porque al bajar hacia el equilibrio se queda encerrada.

Dos comprobaciones adicionales. Si se pide integrar hacia atrás hasta $t = 0.5$ con $y(1) = 2$, se intenta atravesar la singularidad $t = 0.856159\ldots$ calculada arriba. El integrador reduce el paso y termina antes del extremo solicitado. La fórmula exacta permite identificar aquí la causa del fallo; el aviso por sí solo no la demuestra.

Si se parte de $y(1) = 10^{-12}$, la solución permanece próxima a $10^{-12}e^{t-1}$ y en $t = 15$ vale aproximadamente $1.20 \cdot 10^{-6}$. Una tolerancia absoluta de 10^{-6} puede ocultar un error relativo grande durante la fase inicial. Debe elegirse por debajo de la escala de la solución y comprobar que reducirla no cambia las cifras que se quieren conservar.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
f = lambda t, y: y - y ** 3
for y0 in (0.5, 1.0, 2.0):
    s = solve_ivp(f, (1, 5), [y0], dense_output=True, rtol=1e-8, atol=1e-10)
```

```
    tt = np.linspace(1, 5, 401)
    plt.plot(tt, s.sol(tt)[0], lw=1.4, label="y(1) = %.1f" % y0)
    print("y(1) = %4.1f -> y(5) = %.6f" % (y0, s.sol(5)[0]))
plt.plot([1, 5], [1, 1], "k--"); plt.legend(); plt.grid(True); plt.show()
```

Ejercicio 0.2. *Tres soluciones de $y' = -y + y^2$, y una que se escapa*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 2 de la parte de EDO.

Dibujar en la misma ventana gráfica las soluciones de

$$y' = -y + y^2, \qquad t \in [0, 1],$$

correspondientes a $y(0) = 0.1$, $y(0) = 1$ e $y(0) = 1.5$.

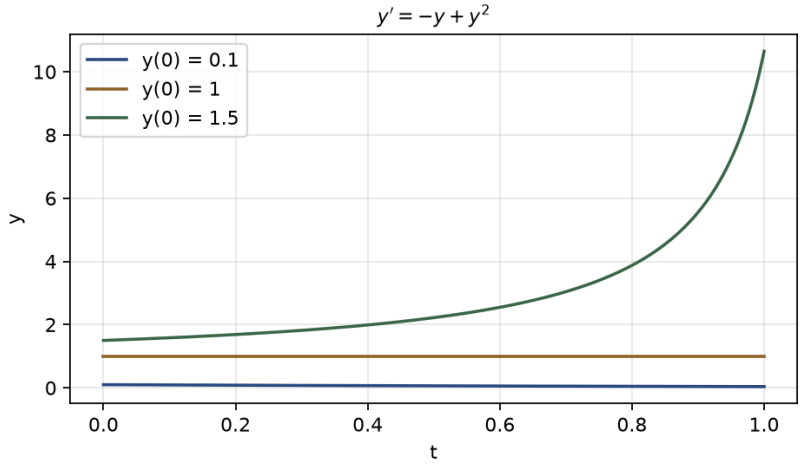
Resolución. La idea. El programa es el del ejercicio anterior con otro segundo miembro, y esa es justamente la gracia: **el método no cambia, y sin embargo el resultado sí**. Lo que se aprende aquí no es una orden nueva de MATLAB, sino que un integrador no protege de un modelo que no tiene solución en el intervalo pedido. El núcleo del programa vuelve a ser el par `f = @(t,y) -y + y^2` y la llamada a `ode45`; todo lo demás es dibujo. Y la pregunta que hay que hacerse antes de ejecutarlo es si el intervalo $[0, 1]$ cabe dentro del intervalo maximal de las tres soluciones, porque `ode45` no la contesta: la sufre.
Qué va a salir. Ahora $f(y) = y(y-1)$, con equilibrios 0 y 1, negativa entre ellos y positiva por encima de 1. El signo se lee como en el ejercicio anterior —sube donde $f > 0$, baja donde $f < 0$, se queda quieta en los ceros— y hay una diferencia, que es el punto del ejercicio: por encima de 1 la solución crece *y no tiene ningún equilibrio por delante que la frene*. En el ejercicio anterior sí lo había, el $y = 1$ que frenaba a la que salía de 2. Nada garantiza aquí, entonces, que la solución llegue viva hasta $t = 1$.

```
%% Ejercicio 2 - tres soluciones de y' = -y + y^2 en [0,1]
f = @(t,y) -y + y^2;
y0s = [0.1, 1, 1.5];
opc = odeset('RelTol', 1e-8, 'AbsTol', 1e-10);

figure, hold on
for k = 1:length(y0s)
  [t, y] = ode45(f, [0, 1], y0s(k), opc);
  plot(t, y, 'LineWidth', 1.4)
  fprintf('y(0) = %4.1f -> y(1) = %.6f \n', y0s(k), y(end))
end
hold off
xlabel('t'), ylabel('y'), title('y'' = -y + y^2'), grid on
legend('y(0) = 0.1', 'y(0) = 1', 'y(0) = 1.5', 'Location', 'northwest')
```

SALIDA ESPERADA

```
y(0) = 0.1 -> y(1) = 0.039270
y(0) = 1.0 -> y(1) = 1.000000
y(0) = 1.5 -> y(1) = 10.648940
```



Las tres soluciones de $y' = -y + y^2$ en $[0, 1]$. La que sale de 1.5 llega a 10.65 y aplasta a las otras dos contra el eje horizontal; la que sale de 1 es el equilibrio.

La tercera pasa de 1.5 a 10.65 en una unidad de tiempo, y aplasta a las otras dos contra el eje horizontal en el dibujo conjunto: conviene rehacerlo con `semilogy` o mirar las dos primeras aparte para verlas.
Dónde está la trampa. Esta ecuación también es de Bernoulli, y la solución exacta es

$$y(t) = \frac{y_0}{y_0 + (1 - y_0)e^t},$$

cuyo denominador se anula cuando $e^t = y_0/(y_0 - 1)$. Con $y_0 = 1.5$ eso ocurre en $t = \log 3 = 1.098612$. La solución **explota en tiempo finito**, y el intervalo $[0, 1]$ del enunciado termina 0.0986 unidades antes.
Esa fórmula es además *el control* del programa, y conviene usarla como tal: en $t = 1$ da 0.03927030 para $y_0 = 0.1$ y 10.64894033 para $y_0 = 1.5$, que son los dos números de la salida hasta la última cifra impresa. Un programa de integración numérica no se da por bueno porque dibuje una curva plausible; se da por bueno cuando reproduce un valor que se conoce por otra vía. Aquí esa vía existe porque la ecuación es de Bernoulli, y en la mayoría de las prácticas que vienen después no existirá: entonces habrá que buscar el control en una cantidad conservada, en un caso límite o en el destino que la teoría predice. Si se pide $[0, 1.2]$, `ode45` no da un error limpio: acorta el paso y avisa de que no puede cumplir las tolerancias. Devuelve solo el tramo alcanzado; aquí la fórmula exacta confirma que la causa es una explosión en tiempo finito.

LA IDEA: EL AVISO QUE HAY QUE SABER LEER

El aviso `Warning: Failure at t=... Unable to meet integration tolerances` indica que la integración ha fallado, pero no determina la causa: puede haber una singularidad, un problema de rigidez o una dificultad en la función programada. Se revisan el último instante alcanzado y el modelo.
La ausencia de un equilibrio por delante no basta para concluir explosión en tiempo finito: $y' = y$ crece sin cota y existe para todo tiempo. En el ejemplo presente, la singularidad se ha demostrado con la solución exacta. Del mismo modo, un método que no converge en el problema de Bratu no prueba por sí solo que no haya solución.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
f = lambda t, y: -y + y ** 2
for y0 in (0.1, 1.0, 1.5):
    s = solve_ivp(f, (0, 1), [y0], dense_output=True, rtol=1e-8, atol=1e-10)
    tt = np.linspace(0, 1, 201)
    plt.semilogy(tt, s.sol(tt)[0], lw=1.4, label="y(0) = %.1f" % y0)
    print("y(0) = %4.1f -> y(1) = %.6f" % (y0, s.sol(1)[0]))
print("explosion para y0 = 1.5 en t = log 3 =", np.log(3))
plt.legend(); plt.grid(True); plt.show()
```

Ejercicio 0.3. *Quedarse con una sola de las soluciones calculadas*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 3 de la parte de EDO.

Obtener solo la gráfica de la solución correspondiente al segundo dato inicial del ejercicio anterior.

Resolución. El ejercicio es corto y está puesto para fijar una idea de manejo de datos que después se usa en todas las prácticas de sistemas: **cómo se guarda lo que devuelve `ode45` y cómo se saca de ahí una sola curva**.
La idea, y el control que hay que ver antes de dibujar. El segundo dato inicial del ejercicio anterior es $y(0) = 1$, y 1 es un *equilibrio* de $y' = -y + y^2$. Así que la respuesta se conoce sin ordenador: la gráfica pedida es la recta horizontal $y \equiv 1$. Eso convierte el ejercicio en el más barato de los controles —si la curva que sale no es plana, el error está en la extracción de la columna, no en el integrador— y explica por qué la salida del programa no imprime valores sino tamaños: aquí lo que hay que comprobar es la forma del array, no la física. En una máquina el segundo miembro devuelve exactamente 0 en $y = 1$, de modo que esa componente sale constante **bit a bit**, sin ningún error acumulado.
De las dos versiones que siguen, la primera no tiene núcleo que señalar —es una llamada a `ode45` ya conocida— y en la segunda **el núcleo es el índice**, `y(:,2)`: toda la resolución cabe en entender qué guarda esa matriz y en qué orden. El resto —dibujo y rótulos— es lo de siempre.
Hay dos maneras, y las dos convienen.

La directa: volver a integrar solo con ese dato. Cuesta una línea y no hay nada que extraer.

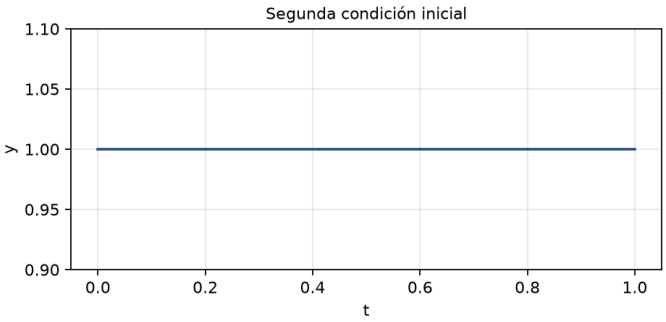
```
f = @(t,y) -y + y^2;
[t, y] = ode45(f, [0, 1], 1);
plot(t, y, 'LineWidth', 1.4), xlabel('t'), ylabel('y'), title('y(0) = 1'), grid on
```

La que enseña algo: integrar los tres datos **a la vez**. Como la ecuación es la misma, se puede pedir a `ode45` que resuelva un sistema de tres ecuaciones desacopladas, una por dato inicial; entonces `y` es una matriz de tres columnas y la segunda solución es `y(:,2)`. El segundo miembro tiene que devolver un **vector columna**, y por eso se escribe con `.^` y no con `^`: opera componente a componente.

```
F = @(t,y) -y + y.^2; % vectorial: tres ecuaciones desacopladas
[t, y] = ode45(F, [0, 1], [0.1; 1; 1.5]);
fprintf('%d instantes y %d columnas \n', size(y,1), size(y,2))
plot(t, y(:,2), 'LineWidth', 1.4) % la segunda columna: el dato y(0) = 1
xlabel('t'), ylabel('y'), title('Segunda condicion inicial'), grid on
```

SALIDA ESPERADA

```
45 instantes y 3 columnas
```



La segunda columna de la integración conjunta: el dato $y(0) = 1$ es un equilibrio, así que la gráfica es la recta $y = 1$.

El 3 es un dato del problema; el número de instantes **no lo es**, porque lo elige el integrador y cambia con las tolerancias y con la versión, así que la salida de arriba puede tener otro primer número. La lectura que hay que llevarse es la de siempre en MATLAB: **una columna por incógnita, una fila por instante**, y por tanto `y(:,k)` es la k-ésima componente a lo largo del tiempo y `y(end,:)` es el estado final entero.

EL CAMINO QUE NO SE TOMÓ: INTEGRAR LAS TRES A LA VEZ, Y POR QUÉ NO SALE LO MISMO

Integrar las tres a la vez no da lo mismo que integrarlas por separado, aunque las ecuaciones estén desacopladas: el control del paso es *común* a todas las componentes, así que la tercera solución, que se dispara, obliga a acortar el paso también para las otras dos. Se obtienen mallas distintas, y pueden cambiar el coste y los errores de cada componente. La integración conjunta no garantiza más precisión ni más rapidez. Para ecuaciones desacopladas puede ser útil compararla con la integración por separado.
Y hay una rotura que no es de precisión sino de todo o nada: **si una componente explota, se paran todas**. Basta cambiar el intervalo a $[0, 1.2]$ para verlo: la tercera se va a infinito en $t = \log 3$, el integrador se detiene ahí, y las dos primeras —que en $t = 1.2$ valdrían 0.032382 y 1, y nada les impedía llegar— se quedan sin su último tramo. En un sistema acoplado eso es inevitable; en tres ecuaciones desacopladas metidas en el mismo vector, es un precio que se paga por comodidad de escritura.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
F = lambda t, y: -y + y ** 2 # NumPy ya opera componente a componente
s = solve_ivp(F, (0, 1), [0.1, 1.0, 1.5], dense_output=True)
print(s.y.shape) # (3 componentes, n instantes): traspuesto de MATLAB
tt = np.linspace(0, 1, 201)
plt.plot(tt, s.sol(tt)[1], lw=1.4); plt.grid(True); plt.show()
```

Comparar con la solución exacta

Ejercicio 0.4. *La numérica contra la exacta*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 4 de la parte de EDO.

Comparar gráficamente la solución de

$$y' = y - y^3, \qquad y(1) = 0.5,$$

con la solución exacta $y = \dfrac{e^t}{\sqrt{3e^2 + e^{2t}}}$.

Resolución. La idea, y es la del libro entero. Este ejercicio no calcula nada nuevo: calcula lo mismo que el primero y lo pone al lado de la verdad. Es el prototipo del *control independiente*, que es lo que separa un programa que funciona de un programa del que uno se fía. La decisión de método aquí no es qué integrador usar —sigue siendo `ode45`— sino **con qué se compara y sobre qué malla**, y de eso van los dos apartados de abajo. El núcleo del programa son tres líneas: la que integra, la que evalúa la exacta y la resta; las diez restantes son dos `subplot` y sus rótulos.
De dónde sale esa fórmula. Es la del ejercicio 1 de esta sección con $y_0 = 0.5$: allí salió $y = (1 + 3e^{2-2t})^{-1/2}$, y multiplicando arriba y abajo por e^t queda $e^t / \sqrt{e^{2t} + 3e^2}$, que es la del enunciado. Conviene comprobarlo, porque comparar contra una fórmula equivocada es la manera más eficaz de perder una tarde.

Cómo se compara, y dónde está la única dificultad. La solución numérica vive en una malla que elige el integrador; la exacta se puede evaluar donde uno quiera. Así que hay dos formas de ponerlas frente a frente, y son las dos que hay que saber:

1. **Evaluar la exacta en la malla de la numérica.** Se toma `[t,y]`, se calcula `yex = exacta(t)` en esos mismos `t`, y se restan. Es lo natural y no hace falta interpolar nada.
2. **Evaluar la numérica en una malla fina propia.** Se toma `sol` y se usa `deval(sol, tt)` con `tt = linspace(1,5,401)`. Es lo que hay que hacer cuando la malla del integrador es demasiado basta para dibujar una curva suave, que aquí pasa: al integrador le bastan unas decenas de puntos en cuatro unidades de tiempo.

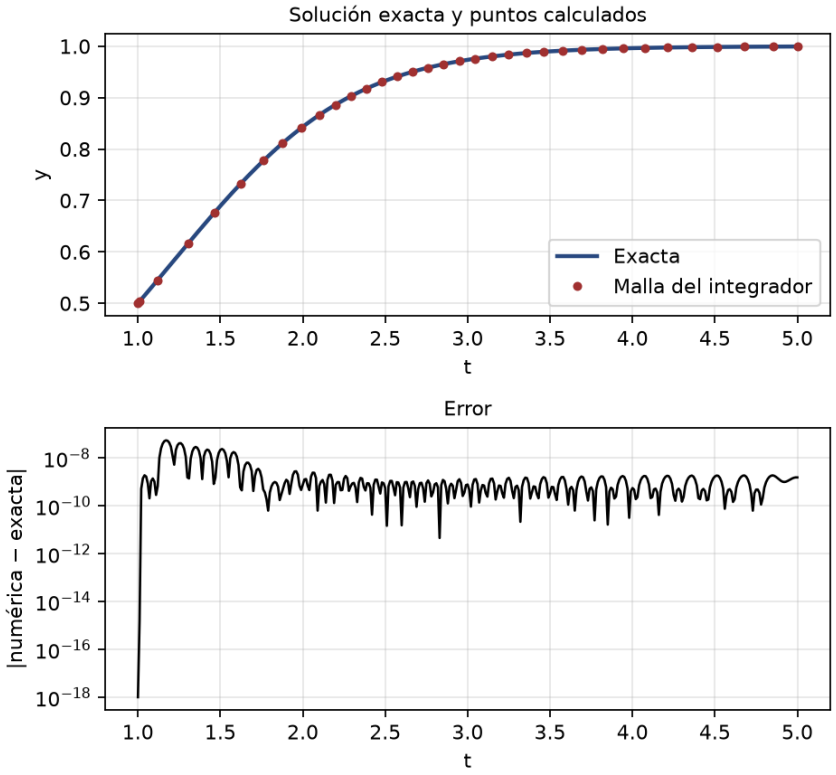
```
%% Ejercicio 4 - comparacion con la solucion exacta
f = @(t,y) y - y^3;
exacta = @(t) exp(t) ./ sqrt(3*exp(2) + exp(2*t));
opc = odeset('RelTol', 1e-8, 'AbsTol', 1e-10);

sol = ode45(f, [1, 5], 0.5, opc); % tercera forma: la solucion como objeto
tt = linspace(1, 5, 401);
yn = deval(sol, tt); % numerica en la malla fina
ye = exacta(tt);

figure
subplot(2,1,1)
plot(tt, ye, 'b', 'LineWidth', 2), hold on
plot(sol.x, sol.y, 'r.', 'MarkerSize', 12) % los puntos que calculo ode45
hold off, legend('Exacta', 'Malla de ode45', 'Location', 'southeast')
xlabel('t'), ylabel('y'), title('Solucion exacta y puntos calculados'), grid on

subplot(2,1,2)
semilogy(tt, abs(yn - ye), 'k', 'LineWidth', 1.2)
xlabel('t'), ylabel('|numerica - exacta|'), title('Error'), grid on

fprintf('Puntos de la malla: %d \n', length(sol.x)) % unas pocas decenas; varia
fprintf('Error maximo: %.1e \n', max(abs(yn - ye)))
fprintf('Error en t = 5: %.1e \n', abs(deval(sol,5) - exacta(5)))
```



Arriba, la solución exacta y los puntos que calcula el integrador (35 en este cálculo; ese número varía con el integrador). Abajo, el error en escala logarítmica, con un máximo de unos $5 \cdot 10^{-8}$.

Como referencia, la versión en Python de abajo produce un error máximo de unos $5.5 \cdot 10^{-8}$ en esa malla. La salida de MATLAB puede diferir porque su elección de pasos y su interpolación no son idénticas.

Qué mide el cálculo. `max(abs(yn-ye))` mide el mayor error en los 401 puntos de comparación, no el supremo exacto en todo el intervalo. Tanto ese valor como el número de pasos dependen del método y de las tolerancias; reducir estas permite comprobar si las cifras se estabilizan.

Cerca de $y = 1$, las soluciones exactas vecinas se aproximan porque $f'(1) = -2 < 0$. Esa estabilidad amortigua perturbaciones previas, pero no garantiza que cualquier cálculo numérico sea estable ni elimine los nuevos errores de cada paso. En Volterra, las órbitas cerradas no tienen esa atracción y conviene comprobar también una cantidad conservada.

CONEXIÓN: POR QUÉ LA MALLA ES TAN CORTA

Unas decenas de puntos para cubrir $[1, 5]$ parecen pocos, y bastan. La razón es que `ode45` no reconstruye la solución uniendo esos puntos con segmentos: el método guarda, en cada paso, un polinomio de grado cuatro que aproxima la solución *dentro* de ese paso, y es el que evalúa `deval`. La malla marca dónde se rehacen las cuentas, no la resolución de la curva. Por eso conviene dibujar con `deval` sobre una malla fina y no con `plot(sol.x, sol.y)`, que sí une con segmentos.

Cuando no se conoce una solución exacta. Puede repetirse el cálculo con tolerancias menores y comparar ambas soluciones sobre la misma malla. Si el error de la segunda es mucho menor, la diferencia aproxima el error de la primera. Aquí se puede comprobar esa idea frente a la fórmula exacta.
No hay una separación universal de tolerancias que garantice ese resultado. Se busca estabilización al refinarlas y, si hace falta, se contrasta con otro método, un residuo o un invariante. Dos cálculos que coinciden no descartan un error compartido en el modelo o en su programación.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
f = lambda t, y: y - y ** 3
exacta = lambda t: np.exp(t) / np.sqrt(3 * np.e ** 2 + np.exp(2 * t))
s = solve_ivp(f, (1, 5), [0.5], dense_output=True, rtol=1e-8, atol=1e-10) # dense_output
    = deval
tt = np.linspace(1, 5, 401)
print("error maximo: %.1e" % np.max(np.abs(s.sol(tt)[0] - exacta(tt))))
plt.semilogy(tt, np.abs(s.sol(tt)[0] - exacta(tt)), "k"); plt.grid(True); plt.show()
```

Conviene quedarse con dos hábitos de estos cuatro ejercicios. Uno es de cálculo: ningún número que dependa del número de puntos de la malla es un dato del problema, y las tolerancias mandan sobre las cifras que se publican. El otro es previo al ordenador: antes de programar una EDO escalar autónoma se mira el signo del segundo miembro, que dice si la solución sube o baja, hacia qué equilibrio, y si hay alguno delante que impida la explosión. Los dos ahorran más tiempo que cualquier truco de programación.

Preguntas que no son la gráfica: `interp1` y `fzero`

Los tres ejercicios que siguen usan modelos de población para pedir *números*, además de las gráficas: cuántos habitantes habrá o cuándo se llegará a cierta cifra. La primera pregunta se contesta evaluando la solución en un punto y la segunda resolviendo una ecuación. Ambas usan la aproximación numérica y, entre sus puntos, una interpolación.

Para evaluar en un punto ya hay dos herramientas, `deval` sobre la estructura y `interp1` sobre los vectores `[t, y]`. La segunda se escribe así:

```
[t, y] = ode45(f, [0, 10], y0);
yfun = @(s) interp1(t, y, s); % la solucion numerica, como funcion de s
yfun(3.7)
```

`interp1(t, y, s)` interpola linealmente entre los dos puntos de la malla que rodean a `s`. Es menos preciso que `deval`, que usa el polinomio de grado cuatro del propio método, pero se escribe con lo que ya hay y no necesita guardar la estructura.

Para la segunda pregunta —*en qué instante* vale la solución un valor dado— hay que resolver $y(t) = c$, que es buscar un cero de $g(t) = y(t) - c$. Eso lo hace `fzero`:

```
g = @(s) yfun(s) - 50;
ts = fzero(g, 5) % 5 es la aproximacion inicial
```

`fzero(g, x0)` busca un cero de `g` partiendo de `x0`. Admite también un *intervalo* en lugar de un punto, `fzero(g, [a, b])`, y entonces exige que `g(a)` y `g(b)` tengan signos distintos (o

un extremo sea raíz). Si g es continua, el cambio de signo garantiza que existe un cero dentro; `fzero` combina bisección e interpolación para aproximarlo. El intervalo debe quedar dentro del dominio donde se ha calculado la solución.

LA IDEA: POR QUÉ INVERTIR UNA FUNCIÓN ES RESOLVER UNA ECUACIÓN

Preguntar «¿en qué instante hay 50 millones?» es pedir $t = y^{-1}(50)$, y una solución numérica no viene con su inversa. Lo que sí se puede hacer siempre es *probar*: dado un t, calcular $y(t)$ y ver si pasa de 50 o no. Un buscador de ceros es exactamente eso automatizado, y por eso la pregunta inversa cuesta un algoritmo y la directa no cuesta nada.
De ahí sale la precaución que hay que tener siempre: **la respuesta puede no ser única**. Si la solución sube y luego baja —como la $I(t)$ del modelo SIR del capítulo 2— hay dos instantes con el mismo valor. `fzero` busca un cero y no garantiza encontrar el más cercano al punto inicial ni enumerar los demás. Mirar antes la gráfica no es un lujo: es lo que dice qué intervalo hay que pasarle. Cuando la solución es monótona —y en estos tres ejercicios lo es, porque el segundo miembro no cambia de signo— la respuesta es única y no hay nada que temer.

Ejercicio 0.5. *La ley de Malthus, y las dos preguntas que se le hacen a un modelo*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 5 de la parte de EDO.

La población de un país sigue la ley de Malthus $y' = 0.05\,y$, donde $y = y(t)$ es el número de habitantes en millones e t el tiempo en años. Sabiendo que inicialmente era de 39.5 millones: dibujar la solución aproximada en $[0, 10]$ y compararla con la exacta; decir cuántos habitantes habría a los diez años; y decir en qué instante se alcanzarían los 50 millones.

Resolución. La idea: son dos preguntas de distinta naturaleza y cuestan cosas distintas. «¿Cuántos habitantes a los diez años?» es *evaluar*: se pide el valor de una función en un punto y cuesta una interpolación. «¿En qué instante se llega a 50?» es *invertir*: se pide el punto donde la función vale algo, y una solución numérica no viene con su inversa. Esa segunda pregunta cuesta un algoritmo, y por eso aparece `fzero`. Toda la elección de método del ejercicio está ahí: `interp1` o `deval` para la primera, un buscador de ceros para la segunda, y `fzero` con **intervalo** y no con punto de partida, porque con intervalo se exige cambio de signo y se mantiene la búsqueda dentro del tramo integrado. Aquí la solución es continua y estrictamente creciente, y cruza 50 en $[0, 10]$: hay un único cero. Más adelante, en el modelo SIR, habrá que distinguir dos posibles instantes con el mismo número de infectados.
El núcleo del programa son también dos líneas —la llamada a `ode45` y la llamada a `fzero`—, y la tercera línea importante, `yfun = @(s) interp1(t, y, s)`, es la que las une: convierte una tabla de valores en algo que se puede evaluar donde haga falta.
La exacta, que aquí se sabe. La ecuación es lineal y homogénea, así que $y(t) = 39.5\,e^{0.05t}$. Conviene tenerla porque convierte el ejercicio en una comprobación: todo lo que devuelva el programa se puede contrastar a mano.

```
%% Ejercicio 5 - ley de Malthus
lambda = 0.05; y0 = 39.5;
```

```
f = @(t,y) lambda*y;
exacta = @(t) y0*exp(lambda*t);
opc = odeset('RelTol', 1e-8, 'AbsTol', 1e-10);

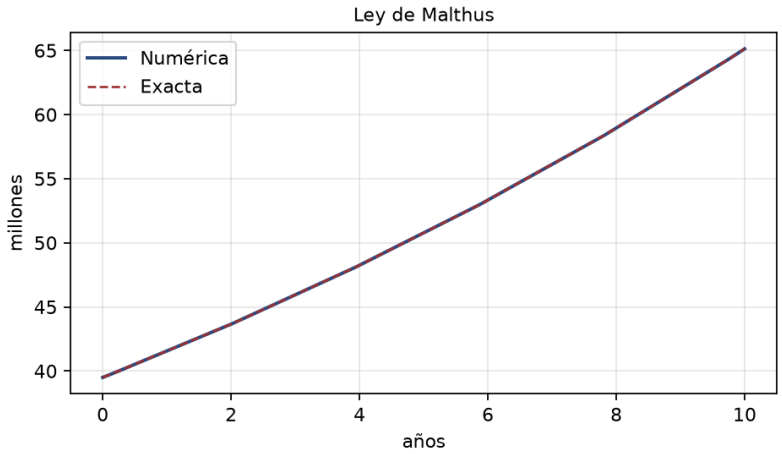
[t, y] = ode45(f, [0, 10], y0, opc);
yfun = @(s) interp1(t, y, s);  % la numerica, como funcion

figure
plot(t, y, 'b', 'LineWidth', 1.6), hold on
plot(t, exacta(t), 'r--', 'LineWidth', 1.2), hold off
legend('Numerica', 'Exacta', 'Location', 'northwest')
xlabel('anos'), ylabel('millones'), title('Ley de Malthus'), grid on

fprintf('A los 10 anos: %.6f millones (exacta %.6f) \n', y(end), exacta(10))
ts = fzero(@(s) yfun(s) - 50, [0, 10]);  % yfun es continua y cruza 50
fprintf('Se llega a 50 millones en t = %.6f anos \n', ts)
fprintf(' valor exacto: %.6f \n', log(50/y0)/lambda)
```

Valores de referencia de la solución exacta:

```
A los 10 anos: 65.124490 millones
Se llega a 50 millones en t = 4.714447 anos
```



La solución numérica y la exacta, $y(t) = 39.5\,e^{0.05t}$: a esta escala no se distinguen.

El instante obtenido con `interp1` depende de la malla: sus seis decimales impresos no tienen por qué coincidir con esta referencia.

Qué hay detrás de esos dos números. El primero es $39.5\,e^{0.5}$. El segundo sale de despejar: $50 = 39.5\,e^{0.05t}$ da $t = 20\log(50/39.5) = 4.714447$. Y hay una manera de tenerlo en la cabeza sin calculadora: en el modelo exponencial el **tiempo de duplicación** es $\log 2/\lambda = 13.86$ años, siempre el mismo, y pasar de 39.5 a 50 es multiplicar por 1.2658, algo menos de un tercio de una duplicación; $13.86/3 = 4.6$, que es la respuesta con un error del dos por ciento.

Los dos caminos para el segundo número, y en qué se diferencian. Aquí se ha usado `interp1` dentro de `fzero`, que es lo que sugiere el enunciado. Con `deval` sobre la estructura reduce el error de interpolación frente a las rectas de `interp1`:

```
sol = ode45(f, [0, 10], y0, opc);
```

```
ts2 = fzero(@(s) deval(sol,s) - 50, [0, 10]);
```

La diferencia entre las dos respuestas es del orden del error de la interpolación lineal sobre la malla, y para dar seis decimales conviene la segunda. El primer camino, en cambio, es el que se puede escribir cuando ya se tienen los vectores `[t, y]` de un cálculo anterior y no se guardó la estructura.

EL CAMINO QUE NO SE TOMÓ: EL MODELO SE CREE LO QUE LE PIDAS

El programa contesta 65.12 millones a los diez años y lo haría igual de contento con $t = 200$, donde predice 870 045 millones de habitantes. Esa cifra ilustra el problema de extrapolar esa tasa. La ley de Malthus no tiene ningún mecanismo que frene el crecimiento: es la que se corrige en el capítulo 2 poniéndole un techo, y así aparece la logística. Merece la pena ejecutar el programa con $[0, 200]$ una vez, ver el número, y recordarlo cada vez que un modelo se extrapole lejos de donde se ajustó.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_ivp
from scipy.optimize import brentq
lam, y0 = 0.05, 39.5
s = solve_ivp(lambda t, y: lam * y, (0, 10), [y0], dense_output=True, rtol=1e-8, atol=1e
    -10)
print("a los 10 anos: %.6f" % s.sol(10)[0])
print("t con y = 50 : %.6f" % brentq(lambda x: s.sol(x)[0] - 50, 0, 10)) # brentq =
    fzero con intervalo
```

Ejercicio 0.6. *Bacterias con techo: la logística*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 6 de la parte de EDO.

Una población de bacterias sigue la ley logística $y' = 2y - 0.04\,y^2$, donde $y = y(t)$ es el número de bacterias en miles y t el tiempo en días. Sabiendo que inicialmente había 40 000 bacterias, ¿cuántas habrá pasados 5 días? ¿En qué instante se alcanzarán las 48 000?

Resolución. La idea. Es el ejercicio anterior con el modelo corregido, y las dos preguntas son las mismas —evaluar e invertir— con las mismas herramientas. Lo único que cambia de método es que aquí se guarda la estructura `sol` y se usa `deval` en vez de `interp1`: la solución tiene una curvatura fuerte en el primer día y la interpolación lineal entre puntos de la malla perdería ahí precisión, mientras que `deval` evalúa el polinomio de grado cuatro que el propio método ha construido. El núcleo son las dos líneas de siempre, `ode45` y `fzero`; lo demás dibuja la curva y la recta del techo.
Y hay una decisión previa al programa que vale más que las dos: **comprobar que la pregunta tiene respuesta**. Invertir solo funciona si el valor buscado está dentro del recorrido de la solución en el intervalo, y eso se sabe leyendo el modelo, no ejecutándolo.
Leer el modelo antes de tocarlo. Escrito como $y' = \lambda y - \kappa y^2$ con $\lambda = 2$ y $\kappa = 0.04$, la capacidad de soporte es $\lambda/\kappa = 50$, es decir 50 000 bacterias: ese es el techo del que habla el capítulo 2. El dato inicial son 40 —ojo con las unidades: el enunciado da bacterias y la ecuación está en

miles— que está por debajo del techo, luego la solución crece y tiende a 50 sin llegar nunca. Con eso ya se sabe que la respuesta a la primera pregunta será algo menor que 50 000, y que la segunda pregunta tiene sentido porque $48 < 50$.

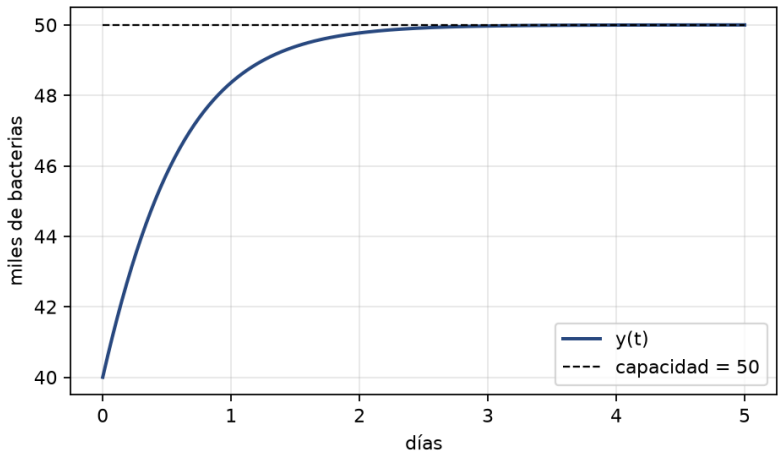
```
%% Ejercicio 6 - logistica: bacterias
f = @(t,y) 2*y - 0.04*y^2;
y0 = 40; % 40 000 bacterias, en miles
opc = odeset('RelTol', 1e-11, 'AbsTol', 1e-13);

sol = ode45(f, [0, 5], y0, opc);
tt = linspace(0, 5, 401);
plot(tt, deval(sol, tt), 'LineWidth', 1.6), hold on
plot([0, 5], [50, 50], 'k--'), hold off % la capacidad de soporte
xlabel('dias'), ylabel('miles de bacterias'), grid on
legend('y(t)', 'capacidad = 50', 'Location', 'southeast')

fprintf('A los 5 dias: %.6f miles \n', deval(sol, 5))
td = fzero(@(s) deval(sol,s) - 48, [0, 5]);
fprintf('Se llega a 48 000 en t = %.6f dias (%.1f horas) \n', td, 24*td)
```

SALIDA ESPERADA

```
A los 5 dias: 49.999433 miles
Se llega a 48 000 en t = 0.895880 dias (21.5 horas)
```



La logística $y' = 2y - 0.04y^2$ con $y(0) = 40$ (en miles): sube hacia la capacidad 50 sin llegar a ella.

Las dos respuestas, comprobadas a mano. La logística se integra por Bernoulli o por fracciones simples y da

$$y(t) = \frac{K}{1 + \left(\frac{K - y_0}{y_0}\right) e^{-\lambda t}} = \frac{50}{1 + \frac{1}{4} e^{-2t}},$$

de donde $y(5) = 50/(1 + \frac{1}{4}e^{-10}) = 49.999433$, y $y(t) = 48$ equivale a $\frac{1}{4}e^{-2t} = \frac{2}{48}$, es decir $e^{-2t} = \frac{1}{6}$ y $t = \frac{1}{2}\log 6 = 0.895880$. Los dos valores de referencia, redondeados a seis decimales.

Y la lectura, que es lo que el ejercicio quiere enseñar. En veintiuna horas y media la población

pasa de 40 000 a 48 000; en los cuatro días siguientes solo sube de 48 000 a 49 999. Ese frenazo es toda la diferencia entre la logística y la exponencial del ejercicio anterior, y se ve en la ecuación sin resolverla: la velocidad $f(y) = y(2 - 0.04y)$ es una parábola que vale $40 \cdot 0.4 = 16$ en $y = 40$ y solo $48 \cdot 0.08 = 3.84$ en $y = 48$, cuatro veces menos con solo ocho mil bacterias más. La velocidad disminuye de forma continua al acercarse al techo.

Comprobar que el valor se alcanza. La solución exacta satisface $y(t) < 50$ para todo tiempo finito, así que pedir el instante en que vale 50 no tiene respuesta. En cambio, 49.9999 sí se alcanza: ocurre en $t \approx 5.868$, fuera de $[0, 5]$. En este segundo caso hay que ampliar tanto la integración como el intervalo de búsqueda.

Para esta solución creciente, comparar los valores de los extremos permite decidir si un nivel está dentro del recorrido. En una función general, tener el mismo signo en ambos extremos no excluye raíces interiores. Cerca del techo, además, el redondeo puede hacer que la solución numérica parezca alcanzarlo: la fórmula exacta distingue ese efecto del modelo.

CONEXIÓN: EL MISMO MODELO, TRES VECES EN ESTE LIBRO

Esta ecuación reaparece en el capítulo 2 como el modelo logístico deducido desde las hipótesis, en el modelo SIS —donde la incógnita es el número de infectados y el «techo» es la población total menos γ/β—, y en la sección de pesca, donde se le resta la captura y aparecen umbrales. Merece la pena reconocerla aquí, en su forma más desnuda, para no tener que volver a deducir nada allí.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_ivp
from scipy.optimize import brentq
s = solve_ivp(lambda t, y: 2 * y - 0.04 * y ** 2, (0, 5), [40.], dense_output=True, rtol
    =1e-11, atol=1e-13)
print("a los 5 dias: %.6f" % s.sol(5)[0])
print("t con y = 48: %.6f" % brentq(lambda x: s.sol(x)[0] - 48, 0, 5))
```

Ejercicio 0.7. *Una epidemia que empieza con dos contagiados*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 7 de la parte de EDO.

En un modelo muy simplificado de epidemias, si $y(t)$ es el número de infectados en el instante t dentro de una población de 100 000 personas,

$$y' = 0.65 \cdot 10^{-6}\, y\,(100\,000 - y), \qquad t \geq 0.$$

Suponiendo que inicialmente hay dos personas contagiadas, ¿cuántos contagiados habrá pasados 50 días? ¿En cuánto tiempo se alcanzará el número de 500 infectados?

Resolución. La idea, y la decisión de método está en el eje. La ecuación es la misma del ejercicio anterior y el integrador también, así que aquí no hay nada nuevo que elegir en el cálculo. Lo que sí hay que elegir, y es lo que este ejercicio enseña, son **el intervalo y la escala del dibujo**, que no vienen dados: el enunciado pregunta por 50 días, pero la segunda pregunta no tiene respuesta dentro de $[0, 50]$ —a los cincuenta días hay 52 contagiados,

no 500— y hay que ampliar el intervalo. Aquí se elige $[0, 200]$ para mostrar también la aproximación al techo. La escala logarítmica permite distinguir el crecimiento inicial, que queda comprimido cerca del eje en una gráfica lineal de toda la población. El núcleo del programa vuelve a ser `ode45` más `fzero`; lo que se ha pensado antes de escribirlo son los dos números del intervalo y la palabra `semilogy`.

Es otra vez la logística. Con $\beta = 0.65 \cdot 10^{-6}$ y $N = 10^5$, la ecuación es $y' = \beta N y - \beta y^2$: crecimiento $\lambda = \beta N = 0.065$ por día y capacidad $\lambda/\beta = N$, la población entera. Lo que cambia respecto del ejercicio anterior es la *escala del dato inicial*: se arranca con 2 de 100 000, tan lejos del techo que durante mucho tiempo el término cuadrático es despreciable y la epidemia crece como una exponencial pura de tasa 0.065.

Para localizar 500 infectados se necesita un intervalo mayor que $[0, 50]$.

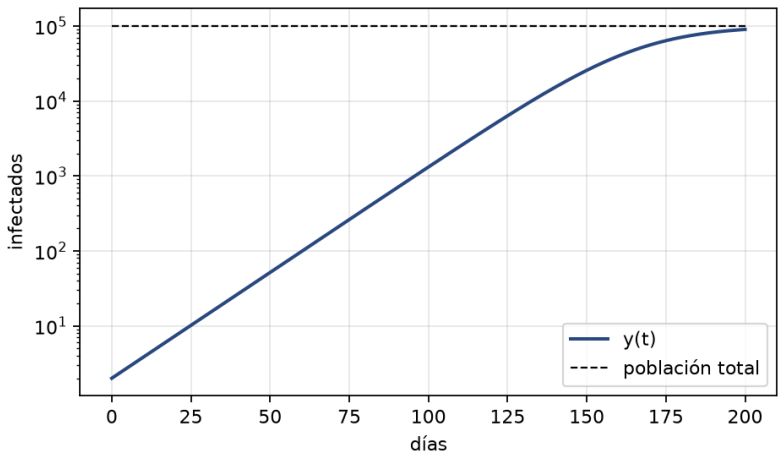
```
%% Ejercicio 7 - epidemia logistica
beta = 0.65e-6; N = 1e5; y0 = 2;
f = @(t,y) beta*y*(N - y);
opc = odeset('RelTol', 1e-11, 'AbsTol', 1e-13);

sol = ode45(f, [0, 200], y0, opc); % 200 dias: hay que ver la meseta
tt = linspace(0, 200, 801);
semilogy(tt, deval(sol, tt), 'LineWidth', 1.5), hold on
semilogy([0, 200], [N, N], 'k--'), hold off
xlabel('dias'), ylabel('infectados'), grid on
legend('y(t)', 'poblacion total', 'Location', 'southeast')

fprintf('A los 50 dias: %.6f infectados \n', deval(sol, 50))
tq = fzero(@(s) deval(sol,s) - 500, [0, 200]);
fprintf('Se llega a 500 infectados en t = %.6f dias \n', tq)
```

SALIDA ESPERADA

```
A los 50 dias: 51.555118 infectados
Se llega a 500 infectados en t = 85.022361 dias
```



La epidemia logística en escala logarítmica: durante mucho tiempo es casi una recta (crecimiento exponencial de tasa 0.065) y solo cerca del techo $N = 10^5$ se curva.

Por qué esos números no son una sorpresa. A los cincuenta días hay 52 infectados de cien mil:

representan todavía una fracción pequeña del total. Mientras $y \ll N$ la ecuación es casi $y' = 0.065\,y$, cuya solución es $2\,e^{0.065t}$, y en $t = 50$ eso vale $2e^{3.25} = 51.58$: la cuenta de la exponencial acierta con dos centésimas. El tiempo de duplicación es $\log 2/0.065 = 10.66$ días, y pasar de 2 a 500 son $\log_2 250 \approx 7.97$ duplicaciones, es decir 84.9 días. La respuesta exacta es 85.02; la diferencia de una décima es lo que ya empieza a frenar el término cuadrático.
El control exacto, que aquí también existe. La logística se integra, y con $y_0 = 2$ y $N = 10^5$ da

$$y(t) = \frac{N}{1 + \left(\frac{N}{2} - 1\right)e^{-0.065t}},$$

de donde $y(50) = 51.555118$ y, despejando $y(t) = 500$, $t = \frac{1}{0.065}\log\left(\frac{N/2-1}{N/500-1}\right) = 85.022361$. Son los dos valores de referencia redondeados a seis decimales. Conviene distinguir los dos controles que se han usado, porque sirven para cosas distintas: la exponencial pura es un control *de orden de magnitud* que se hace de cabeza y que avisa de un error grosero; la fórmula logística es un control *de cifras* y detecta un error pequeño. En una práctica sin solución cerrada solo se tendrá el primero, y por eso conviene practicarlo aquí, donde se puede comprobar cuánto se desvía.

LA IDEA: DÓNDE ESTÁ LA TRAMPA DE LAS EPIDEMIAS, Y POR QUÉ ES UN PROBLEMA DE ESCALA

No hay un cambio repentino de la ecuación: durante la fase inicial, el incremento es pequeño en valor absoluto porque hay pocos infectados, aunque su tasa relativa sea aproximadamente constante. Al crecer la población infectada, el factor $N - y$ reduce esa tasa; la gráfica logarítmica permite ver la salida del régimen aproximadamente exponencial.
El número reproductivo básico $\mathcal{R}_0$ de los modelos del capítulo 2 es una cantidad adimensional que determina un umbral, no una tasa por unidad de tiempo. En el modelo presente, la tasa relativa inicial es aproximadamente 0.065 por día.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_ivp
from scipy.optimize import brentq
b, N = 0.65e-6, 1e5
s = solve_ivp(lambda t, y: b * y * (N - y), (0, 200), [2.], dense_output=True, rtol=1e
    -11, atol=1e-13)
print("a los 50 dias: %.6f" % s.sol(50)[0])
print("t con y = 500: %.6f" % brentq(lambda x: s.sol(x)[0] - 500, 0, 200))
```

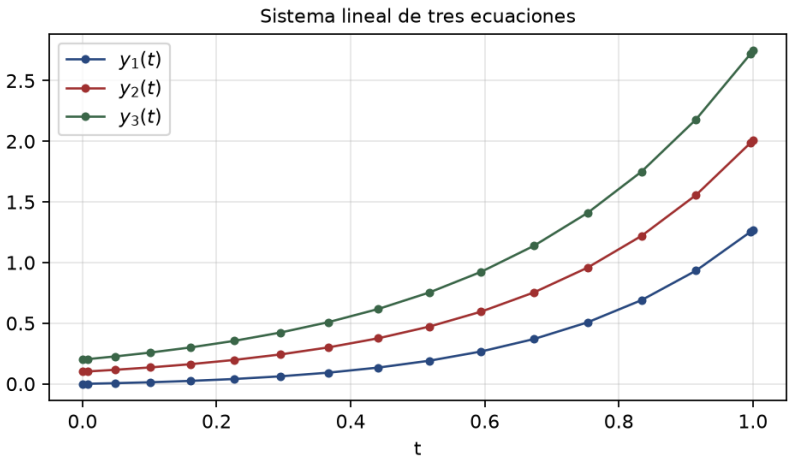
Sistemas: el segundo miembro devuelve un vector columna

Todo lo anterior vale para sistemas sin cambiar una letra de la llamada. Lo único que cambia es lo que hay dentro de la función y lo que hay dentro del dato inicial: donde había un número, hay un vector columna. Así se resuelve

$$x' = 3x - y + z, \qquad y' = -x + 5y - z, \qquad z' = x - y + 3z, \qquad t \in [0, 1],$$

con $x(0) = 0$, $y(0) = 0.1$, $z(0) = 0.2$:

```
f = @(t,y) [3*y(1) - y(2) + y(3);
            -y(1) + 5*y(2) - y(3);
            y(1) - y(2) + 3*y(3)];
y0 = [0; 0.1; 0.2];
ode45(f, [0, 1], y0)
legend('y_1(t)', 'y_2(t)', 'y_3(t)')
```



Las tres componentes del sistema lineal en $[0, 1]$; los puntos son los instantes que calcula el integrador.

Dos detalles que son la fuente de casi todos los errores al empezar. El primero: las incógnitas se numeran, `y(1)`, `y(2)`, `y(3)`, y no se llaman x, y y z aunque el enunciado sí. El segundo: el resultado tiene que ser una **columna**, con puntos y comas entre las filas; si se escriben comas, MATLAB construye una fila y `ode45` protesta diciendo que las dimensiones no coinciden.

Cuando el sistema es largo conviene sacarlo a una M-función en un fichero aparte, que se pasa con la arroba:

```
function dy = fsistem(t, y)
dy = zeros(3,1); % inicializacion: reserva la columna
dy(1) = 3*y(1) - y(2) + y(3);
dy(2) = -y(1) + 5*y(2) - y(3);
dy(3) = y(1) - y(2) + 3*y(3);
end
```

y entonces la llamada es `ode45(@fsistem, [0, 1], y0)`. La línea `zeros(3,1)` no es decorativa: es lo que garantiza que `dy` sea columna aunque después se rellene por partes.

Ejercicio 0.8. *Quedarse con una componente de un sistema*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 8 de la parte de EDO.

Dibujar únicamente la gráfica de $y_2(t)$ del sistema anterior.

Resolución. La idea. Es el ejercicio 3 otra vez, ahora con un sistema de verdad en lugar de tres ecuaciones desacopladas: se piden los valores con `[t, y]` y se toma la columna que interesa. Lo que cambia es que ahora las tres columnas están *acopladas* y no tiene sentido

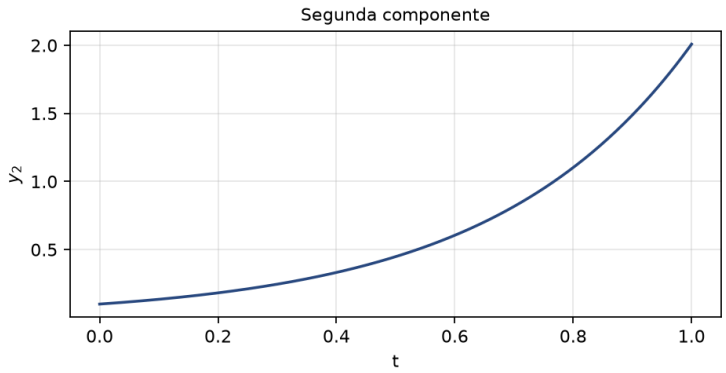
integrarlas por separado, de modo que la única decisión que queda es cómo se escribe el segundo miembro. Se escribe `@(t,y) A*y` en vez de las tres líneas de antes, y **esa es la única línea con contenido del programa**: es el núcleo, y el resto extrae una columna y la dibuja. Escribirlo así tiene una ventaja que no es de comodidad, y es la que da el control de este ejercicio.

```
%% Ejercicio 8 - solo la segunda componente
A = [3 -1 1; -1 5 -1; 1 -1 3];
f = @(t,y) A*y;  % el sistema es lineal: basta el producto
y0 = [0; 0.1; 0.2];
opc = odeset('RelTol', 1e-8, 'AbsTol', 1e-10);

[t, y] = ode45(f, [0, 1], y0, opc);
plot(t, y(:,2), 'LineWidth', 1.5)
xlabel('t'), ylabel('y_2'), title('Segunda componente'), grid on
fprintf('y(1) = (%.6f, %.6f, %.6f) \n', y(end,1), y(end,2), y(end,3))
```

SALIDA ESPERADA

```
y(1) = (1.269648, 2.008554, 2.747459)
```



La componente y_2 del mismo sistema en $[0, 1]$.

El control: $y(t) = e^{tA}y_0$. Escribir Ay hace evidente que el sistema es lineal de coeficientes constantes, y por tanto que la solución es la exponencial de la matriz. Eso da una comprobación exacta sin integrar nada, con `expm(A)*y0`: devuelve los mismos seis decimales que la salida de arriba.

Y una lectura de la matriz que además corrige la primera impresión. A es simétrica, así que admite una base ortogonal de autovectores reales; salen $\lambda = 2$ con $v_2 = (1, 0, -1)$, $\lambda = 3$ con $v_3 = (1, 1, 1)$ y $\lambda = 6$ con $v_6 = (1, -2, 1)$, los tres positivos: cada modo propio crece exponencialmente. El comportamiento de cada componente depende también del dato inicial. Lo tentador es decir que a la larga la solución se alinea con el autovector del autovalor mayor, como en el método de la potencia del capítulo 1. **Aquí no ocurre, y merece la pena ver por qué.** El dato inicial es $y_0 = (0,\ 0.1,\ 0.2)$, y

$$\langle y_0, v_6 \rangle = 0 \cdot 1 + 0.1 \cdot (-2) + 0.2 \cdot 1 = 0 :$$

y_0 es ortogonal al autovector dominante, de modo que ese modo **no está presente** y no

puede crecer, porque un sistema lineal no crea componentes que no había. Descomponiendo,

$$y(t) = 0.1\,e^{3t}\,(1,1,1) \;-\; 0.1\,e^{2t}\,(1,0,-1),$$

y en $t = 1$ eso da $(1.269648,\ 2.008554,\ 2.747459)$, los tres números de la salida. En particular $y_2(1) = 0.1\,e^3 \approx 2.008554$: lo que manda es e^3, no e^6. Y la dirección límite es $(1,1,1)/\sqrt{3}$, no v_6.

Es el caso degenerado del teorema de la potencia, y conviene tenerlo presente porque en el capítulo 1 volverá con otro traje: **el autovalor dominante manda sobre el dato *genérico***, y un dato ortogonal al autovector dominante no es genérico. En el método de la potencia los errores de redondeo pueden reintroducir una componente ausente, aunque no hay que contar con ello como garantía del método. En este sistema esa perturbación puede crecer más deprisa que la solución que se quiere calcular.

Qué cambia a tiempos largos. En $t = 10$ debe evaluarse `expm(10*A)*y0`, no `expm(A)*y0`. La fórmula exacta anterior da aproximadamente $(1.06860, 1.06865, 1.06870)\cdot 10^{12}$. Una evaluación en doble precisión puede perder cifras por cancelación entre términos grandes.

La sensibilidad se entiende comparando los modos: una perturbación inicial de tamaño ε en la dirección de v_6 crece como εe^{6t}, mientras el modo presente dominante crece como $0.1e^{3t}$. Su tamaño relativo es, por tanto, del orden de $10\varepsilon e^{3t}$. Esto explica la pérdida de precisión, pero no fija un tiempo universal a partir del cual falle cualquier cálculo: también influyen el algoritmo y los redondeos que introduce. La fórmula modal ofrece aquí una comprobación independiente.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_ivp
from scipy.linalg import expm
A = np.array([[3, -1, 1], [-1, 5, -1], [1, -1, 3.]])
y0 = np.array([0, 0.1, 0.2])
s = solve_ivp(lambda t, y: A @ y, (0, 1), y0, dense_output=True, rtol=1e-8, atol=1e-10)
print(s.sol(1)) # (1.269648, 2.008554, 2.747459)
print(expm(A) @ y0) # lo mismo, sin integrar
```

El plano de fases

Con dos incógnitas hay dos dibujos distintos y conviene no confundirlos. La **trayectoria** pone el tiempo en el eje horizontal y las dos componentes encima, `plot(t, y)`. La **órbita** tira el tiempo y dibuja una componente contra la otra, `plot(y(:,1), y(:,2))`: es la sombra del movimiento en el plano de las incógnitas.

LA IDEA: QUÉ SE GANA Y QUÉ SE PIERDE AL TIRAR EL TIEMPO

Se pierde *cuándo*, y se gana *qué forma*. En la órbita se ven de un vistazo cosas que la trayectoria esconde: en un sistema autónomo con unicidad, una órbita no estacionaria cerrada corresponde a un movimiento periódico; si espirala hacia un punto, hay un equilibrio atractivo; si se va al infinito, la solución no está acotada. Nada de eso exige resolver la ecuación, y por eso el plano de fases es la herramienta central de los

capítulos 2 y del apéndice de complementos.
Lo que no se puede leer en la órbita es la velocidad. Dos sistemas con las mismas órbitas pueden recorrerlas en tiempos muy distintos, y en Volterra ocurre justamente eso: la órbita pasa mucho tiempo en la parte de población baja y cruza deprisa el pico, de modo que un punto de la órbita no representa el mismo intervalo de tiempo que otro.

Ejercicio 0.9. *Un sistema presa-depredador y su órbita cerrada*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 9 de la parte de EDO.

Obtener las gráficas de la solución del sistema

$$y_1' = y_1 - 0.01\, y_1 y_2, \qquad y_2' = -y_2 + 0.02\, y_1 y_2, \qquad t \in [0, 15],$$

con $y_1(0) = y_2(0) = 20$, y de la órbita asociada.

Resolución. La idea. El sistema tiene una integral primera y las órbitas no estacionarias del cuadrante positivo son cerradas. Un método de Runge-Kutta no conserva necesariamente esa cantidad: los errores pueden desplazar la curva calculada hacia otras órbitas o alterar la fase. Aquí se usan tolerancias pequeñas y se contrasta la integral primera para distinguir los efectos numéricos de la dinámica del modelo.
El núcleo son dos líneas —el segundo miembro y la llamada a `ode45`— más una tercera que es de otra naturaleza: `Y = deval(sol, tt)`, que reconstruye la solución sobre dos mil puntos. Esa línea no calcula nada nuevo; decide cómo se dibuja lo ya calculado, y sin ella el ejercicio parece dar un resultado falso, como explica el aparte del final.
Qué sistema es. Es un Volterra: la primera especie crece sola y la segunda se muere sola, y el término $y_1 y_2$ transfiere de una a otra. Sus equilibrios son el origen y el punto interior $(c/d,\, a/b) = (1/0.02,\, 1/0.01) = (50, 100)$, y el dato inicial $(20, 20)$ está lejos de él, de modo que la órbita será grande. Se espera un ciclo cerrado, que es lo que el capítulo 2 demuestra a partir de la integral primera.

```
%% Ejercicio 9 - Volterra: trayectorias y orbita
f = @(t,y) [y(1) - 0.01*y(1)*y(2);
            -y(2) + 0.02*y(1)*y(2)];
y0 = [20; 20];
opc = odeset('RelTol', 1e-10, 'AbsTol', 1e-12); % orbita cerrada: conviene apretar

sol = ode45(f, [0, 15], y0, opc);
tt = linspace(0, 15, 2001);
Y = deval(sol, tt);

subplot(1,2,1)
plot(tt, Y(1,:), 'b', tt, Y(2,:), 'r', 'LineWidth', 1.3)
xlabel('t'), legend('y_1 (presa)', 'y_2 (depredador)'), title('Trayectorias'), grid on

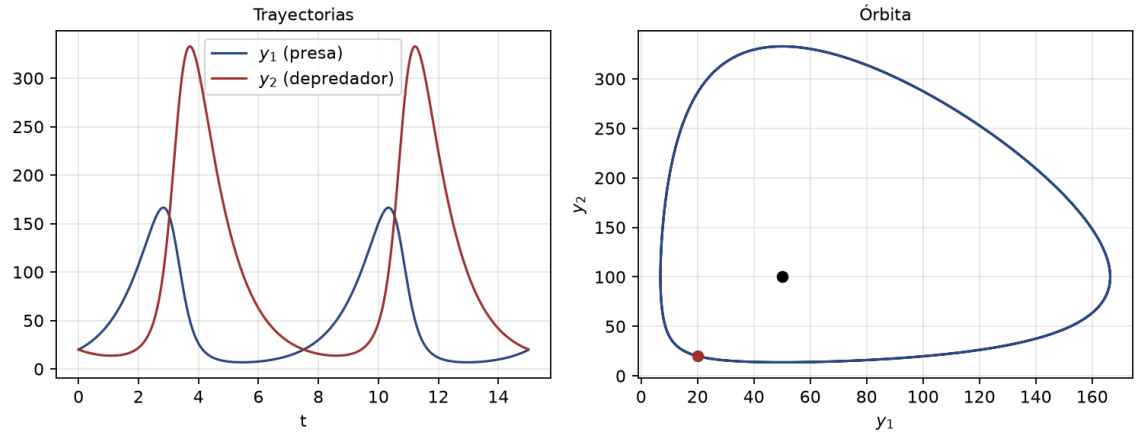
subplot(1,2,2)
plot(Y(1,:), Y(2,:), 'LineWidth', 1.3), hold on
plot(50, 100, 'k.', 'MarkerSize', 18) % el equilibrio interior
plot(y0(1), y0(2), 'r.', 'MarkerSize', 18), hold off
```

```
xlabel('y_1'), ylabel('y_2'), title('Orbita'), grid on

fprintf('y1 entre %.6f y %.6f \n', min(Y(1,:)), max(Y(1,:)))
fprintf('y2 entre %.6f y %.6f \n', min(Y(2,:)), max(Y(2,:)))
```

SALIDA ESPERADA

```
y1 entre 6.842435 y 166.407043
y2 entre 13.684862 y 332.813927
```



Volterra con $y_1(0) = y_2(0) = 20$. A la izquierda, las dos poblaciones en el tiempo; a la derecha, la órbita en el plano (y_1, y_2), cerrada alrededor del equilibrio $(50, 100)$ (punto negro). El punto rojo es el dato inicial.

Los extremos de la órbita. La salida anterior usa los puntos de la malla. Los extremos de las soluciones son aproximadamente $y_1 \in [6.842430, 166.407160]$ y $y_2 \in [13.684860, 332.814319]$. En la órbita exacta, los extremos de y_2 son **el doble** de los de y_1; los extremos muestreados no tienen por qué guardar exactamente esa relación. La razón se ve adimensionalizando: con $u = 0.02\, y_1$ y $v = 0.01\, y_2$ el sistema se vuelve

$$u' = u(1 - v), \qquad v' = v(u - 1),$$

que es el Volterra simétrico, con integral primera $H(u, v) = (u - \log u) + (v - \log v)$. Los extremos de u están donde $u' = 0$ con $u \neq 0$, es decir en $v = 1$, y allí $u - \log u = H - 1$; los extremos de v están en $u = 1$, y allí $v - \log v = H - 1$. **Es la misma ecuación**, luego u y v tienen exactamente el mismo par de extremos. Al deshacer el cambio, y_1 se recupera dividiendo por 0.02 y y_2 por 0.01, y de ahí el factor 2.
El periodo. Se estima buscando el siguiente retorno a $y_1 = 20$ con el mismo sentido de cruce que al inicio. Se obtiene $P \approx 7.509962$: las quince unidades de tiempo del enunciado cubren casi dos vueltas, sin completarlas. El periodo para oscilaciones pequeñas es $2\pi/\sqrt{ac} = 2\pi \approx 6.283185$; esta órbita tarda alrededor de un 19.5 % más.
Máximos de la malla y de la solución. `max(Y(1,:))` devuelve el mayor valor muestreado, que aproxima el máximo de la solución. Para refinarlo se localizan los ceros de y_1' y se comparan sus valores con los de los extremos del intervalo. La ecuación $u - \log u = H - 1$ proporciona aquí una comprobación independiente de los extremos de la órbita completa. Como $15 > P$, ambos se alcanzan en el intervalo considerado.

El control independiente, que es la integral primera. Con $u = 0.02\,y_1$ y $v = 0.01\,y_2$, la cantidad

$$H(u,v) = (u - \log u) + (v - \log v)$$

es constante a lo largo de cualquier solución —se comprueba derivando y sustituyendo—, y por tanto es la vara de medir del programa: si H se mueve, el error numérico ya es visible. Con el dato $(20, 20)$ vale $H_0 \approx 3.125729$. La variación de H sobre los valores calculados mide una parte del error, aunque no detecta por sí sola un desfase a lo largo de la misma órbita. **Esta es la forma de control que hay que buscar en toda práctica de sistemas conservativos**, porque no exige conocer la solución: una cantidad que la teoría dice constante, calculada sobre la salida del programa, y mirada su variación.

Comprobar la deriva. Al ampliar el intervalo conviene volver a medir la variación de H y repetir con tolerancias menores. Una curva que parece espiralar puede ser un artefacto numérico: la teoría demuestra que las órbitas exactas son cerradas. El sentido y la magnitud de la deriva dependen del método y de sus pasos; no hay que interpretar el dibujo aislado como un cambio cualitativo del modelo.

> **El camino que no se tomó: dibujar la órbita con la malla del integrador**
>
> Aquí se ha dibujado con `deval` sobre dos mil puntos, y no con `plot(y(:,1), y(:,2))` sobre la malla. Con la malla la órbita sale poligonal y, lo que es peor, **parece que no cierra**: los segmentos rectos cortan por dentro en las zonas de curvatura fuerte y dan la impresión de una espiral que se mete hacia dentro. Es un artefacto del dibujo y no del cálculo, pero ha hecho dudar a mucha gente de si Volterra es de verdad periódico.

Versión en Python (NumPy/Matplotlib)

```python
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
f = lambda t, y: [y[0] - 0.01 * y[0] * y[1], -y[1] + 0.02 * y[0] * y[1]]
s = solve_ivp(f, (0, 15), [20., 20.], dense_output=True, rtol=1e-10, atol=1e-12)
Y = s.sol(np.linspace(0, 15, 2001))
print(Y[0].min(), Y[0].max(), Y[1].min(), Y[1].max())
plt.plot(Y[0], Y[1]); plt.plot(50, 100, "k."); plt.grid(True); plt.show()
```

Ecuaciones de orden superior: el cambio de variables de siempre

`ode45` solo resuelve sistemas de primer orden, así que una ecuación de orden n hay que convertirla antes. El cambio es el que ya se conoce: se llama y_1 a la función, y_2 a su derivada, y_3 a la segunda, y cada nueva variable es la derivada de la anterior; la ecuación original se usa para la última. Para $y'' = g(t, y, y')$ queda

$$y_1' = y_2, \qquad y_2' = g(t, y_1, y_2),$$

con dato inicial $(y(0), y'(0))$. Es la traducción que se aplica en las prácticas de vigas y vibraciones del capítulo 3, y conviene automatizarla en la cabeza: la primera línea del segundo miembro es siempre `y(2)`.

Ejercicio 0.10. *Un péndulo forzado*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 10 de la parte de EDO.

Calcular y dibujar la solución de

$$y'' + 7\sin y + 0.1\cos t = 0, \qquad t \in [0,5], \qquad y(0) = 0,\ y'(0) = 1.$$

Resolución. La idea. `ode45` solo entiende sistemas de primer orden, así que el único trabajo de modelización de este ejercicio es la *traducción*: bautizar $y_2 = y'$ y escribir la ecuación como dos ecuaciones. No hay elección de método —sigue siendo el mismo integrador— y se fijan tolerancias pequeñas para comparar las cifras del estado final. Lo que sí hay que decidir es **qué se dibuja**: aquí se dibujan las dos cosas, la trayectoria y la órbita, porque dicen cosas distintas y en un sistema mecánico las dos son útiles.
El núcleo del programa son las dos filas del corchete del segundo miembro. La primera, `y(2)`, no es física: es la definición de la variable nueva, y en toda ecuación de orden superior será siempre esa. La segunda, `-7*sin(y(1)) - 0.1*cos(t)`, es la ecuación original despejada, y es la única línea del programa donde puede esconderse un error de modelo.
Qué ecuación es. Es un péndulo: $y'' = -7\sin y$ es la ecuación del péndulo con $g/\ell = 7$, y el término $-0.1\cos t$ es una fuerza externa pequeña que lo empuja. Sin ella, la energía $E = \frac{1}{2}y'^2 + 7(1-\cos y)$ se conservaría; con ella, oscila un poco.
El sistema. Con $y_1 = y$ e $y_2 = y'$:

$$y_1' = y_2, \qquad y_2' = -7\sin y_1 - 0.1\cos t,$$

Aquí el segundo miembro **depende de** t, como el ejemplo inicial de la sección. El forzamiento necesita ese argumento de la firma `@(t,y)`.

```
%% Ejercicio 10 - pendulo forzado, reducido a un sistema de primer orden
f = @(t,y) [y(2);
            -7*sin(y(1)) - 0.1*cos(t)];
opc = odeset('RelTol', 1e-11, 'AbsTol', 1e-13);

sol = ode45(f, [0, 5], [0; 1], opc);
tt = linspace(0, 5, 1001);
Y = deval(sol, tt);

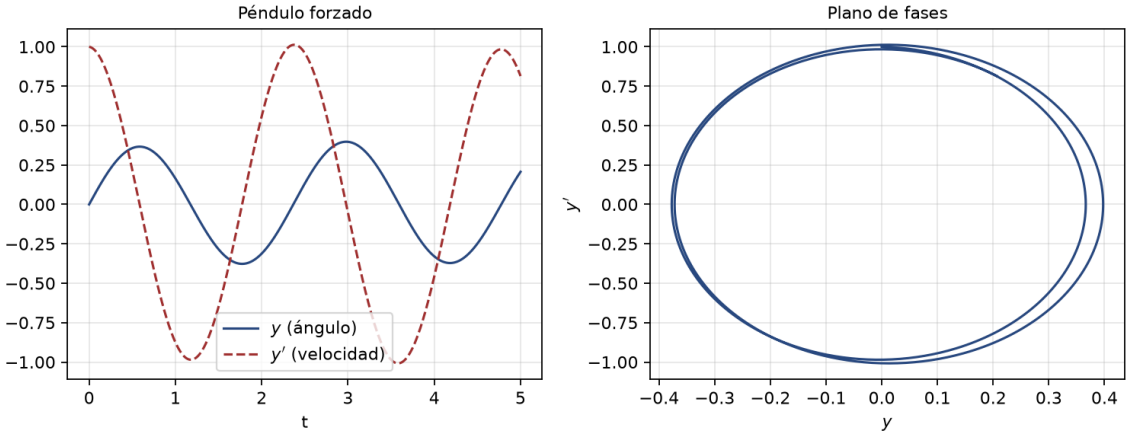
subplot(1,2,1)
plot(tt, Y(1,:), 'b', tt, Y(2,:), 'r--', 'LineWidth', 1.3)
xlabel('t'), legend('y (angulo)', 'y'' (velocidad)'), title('Pendulo forzado'), grid on

subplot(1,2,2)
plot(Y(1,:), Y(2,:), 'LineWidth', 1.3)
xlabel('y'), ylabel('y'''), title('Plano de fases'), grid on

fprintf('y(5) = %.8f, y''(5) = %.8f \n', Y(1,end), Y(2,end))
fprintf('maximo del angulo en la malla: %.8f rad \n', max(abs(Y(1,:))))
```

SALIDA ESPERADA

```
y(5) = 0.20688403, y'(5) = 0.81409003
maximo del angulo en la malla: 0.39755991 rad
```



El péndulo forzado $y'' = -7\sin y - 0.1\cos t$. A la izquierda, el ángulo y la velocidad; a la derecha, el plano de fases, cuya curva no llega a cerrarse.

Lectura. El ángulo no pasa de 0.398 radianes, unos 23 grados: el péndulo oscila cerca de abajo y no llega ni de lejos a dar la vuelta. Con oscilaciones tan pequeñas, $\sin y \approx y$ y la ecuación es casi la del muelle $y'' + 7y = 0$, cuyo periodo es $2\pi/\sqrt{7} = 2.375$; en la figura se cuentan poco más de dos oscilaciones en cinco unidades de tiempo, que es justo eso. Y la órbita del plano de fases no llega a cerrarse, porque el forzamiento le va inyectando y quitando energía.

El máximo de $|y|$ en la solución se alcanza cerca de $t = 2.979209$, donde $y' = 0$. Refinando ese cero se obtiene aproximadamente 0.39756072. La malla de 1001 puntos del programa da 0.39755991: la diferencia procede principalmente del muestreo. Imprimir ocho decimales no convierte el máximo de una malla en el máximo exacto.

El control, y aquí es exacto. Basta quitar el término forzante y comparar con lo que dice la conservación de la energía. Sin forzamiento, $E = \frac{1}{2}y'^2 + 7(1 - \cos y)$ es constante y vale $E_0 = \frac{1}{2}$ con el dato del enunciado; el punto de retorno es donde $y' = 0$, es decir $7(1 - \cos y) = \frac{1}{2}$, y da

$$y_{\text{máx}} = \arccos\left(1 - \tfrac{1}{14}\right) = 0.38025121.$$

La fórmula permite comprobar la amplitud del problema sin forzamiento. En el problema forzado, $E' = -0.1y'\cos t$: la fuerza puede aportar o retirar energía. El máximo angular calculado es aproximadamente un 4.6 % mayor que el del péndulo libre; es una comparación de amplitudes, no el porcentaje de energía aportada.

Cerca de la separatriz. En el péndulo sin forzamiento, la energía crítica es $E = 14$, que para $y(0) = 0$ corresponde a $|y'(0)| = \sqrt{28}$. Un pequeño error cerca de ese umbral puede cambiar la clasificación entre oscilación y rotación. Conviene comparar la deriva numérica de energía con la distancia al umbral.

Las soluciones dependen continuamente del dato inicial en todo intervalo finito fijo. Lo que cambia al cruzar la separatriz es su régimen a largo plazo; no debe confundirse con una discontinuidad de esa dependencia. Si la incertidumbre del dato permite caer a ambos lados, no puede decidirse el régimen solo por una trayectoria numérica.

CONEXIÓN: LA ENERGÍA COMO CRITERIO, Y EL CAPÍTULO 3

El péndulo sin forzamiento se deduce en el capítulo 3 del principio de acción estacionaria y cuya estabilidad se estudia con la función de Liapunov $V = \frac{1}{2}y'^2 + 7(1 - \cos y)$, que es la energía. Aquí sirve además de comprobación numérica: sin el término forzante, el valor calculado de V debe permanecer próximo a una constante. Su deriva puede indicar error de discretización o de programación; se contrasta al reducir tolerancias. La conservación proporciona una comprobación independiente de la forma de la curva.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_ivp
f = lambda t, y: [y[1], -7 * np.sin(y[0]) - 0.1 * np.cos(t)]
s = solve_ivp(f, (0, 5), [0., 1.], dense_output=True, rtol=1e-11, atol=1e-13)
print("y(5) = %.8f, y'(5) = %.8f" % tuple(s.sol(5)))
print("amplitud: %.8f" % np.max(np.abs(s.sol(np.linspace(0, 5, 1001))[0])))
```

Pasar parámetros a la ecuación

La firma de la función que resuelve `ode45` es obligatoriamente `@(t,y)`: dos argumentos y ni uno más. Pero casi todos los modelos del libro tienen parámetros —r y K en la logística, β y γ en el SIR, a, b, c, d en Volterra— y hace falta poder cambiarlos sin editar el fichero cada vez. Hay dos maneras y las dos aparecen en las prácticas.

La primera: envolver en una función anónima. Se escribe la M-función con todos los parámetros que haga falta y después se construye, en el programa principal, una función anónima de dos argumentos que fija los demás:

```
function dy = odeparam(t, y, a, b, c, d)
dy = zeros(2,1);
dy(1) = a*y(1) - b*y(1)*y(2);
dy(2) = -c*y(2) + d*y(1)*y(2);
end
```

```
a = 0.5; b = 0.2; c = 0.5; d = 0.1;
odefun = @(t,y) odeparam(t, y, a, b, c, d); % ahora si es de dos argumentos
[t, y] = ode45(odefun, [0, 25], [4; 4]);
```

La segunda: agrupar los parámetros en una estructura. Se define una variable con campos y se pasa entera como un solo argumento:

```
params.a = 0.5; params.b = 0.2; params.c = 0.5; params.d = 0.1;
odefun = @(t,y) odeparamS(t, y, params);
[t, y] = ode45(odefun, [0, 25], [4; 4]);
```

```
function dy = odeparamS(t, y, params)
dy = zeros(2,1);
dy(1) = params.a*y(1) - params.b*y(1)*y(2);
dy(2) = -params.c*y(2) + params.d*y(1)*y(2);
end
```

LA IDEA: CUÁL DE LAS DOS CONVIENE, Y POR QUÉ LA ENVOLTURA FUNCIONA

La envoltura anónima funciona porque una función anónima **congela** el valor de las variables que usa en el instante en que se define. Al escribir `@(t,y) odeparam(t, y, a, b, c, d)`, los cuatro números quedan guardados dentro del objeto; si después se cambia `a`, la función sigue usando el valor viejo. Eso, que en otros contextos es una trampa, es justo lo que hace falta aquí: la función que recibe `ode45` tiene dos argumentos y lleva los parámetros dentro.
Con dos o tres parámetros, la primera manera es más corta y se lee mejor. La estructura gana cuando son muchos, y sobre todo cuando el mismo juego de parámetros se pasa a varias funciones —el segundo miembro, la función de las condiciones de contorno, la que dibuja—, porque los campos identifican el significado de cada dato. Si una función anónima ya capturó la estructura, hay que volver a definirla para que use los valores modificados. En el capítulo 4, donde hay que llevar a, T, α, μ e y_d a la vez al estado y al adjunto, la estructura ahorra errores de orden de argumentos, que son los más difíciles de ver: si se cruzan dos parámetros del mismo tipo, el programa no falla, solo devuelve otro resultado.

Ejercicio 0.11. *Las órbitas de un Volterra con parámetros*

ORIGEN: hoja de preparación de MATLAB del curso, ejercicio 11 de la parte de EDO.

Escribir un script que contenga una M-función `odeparam` para dibujar las órbitas de la solución de

$$y_1' = a\,y_1 - b\,y_1y_2, \qquad y_2' = -c\,y_2 + d\,y_1y_2, \qquad t \in [0,25],$$

con $y_1(0) = y_2(0) = 4$ y $a = 0.5$, $b = 0.2$, $c = 0.5$, $d = 0.1$; una versión sin usar la estructura `params` y otra en que sí se use.

Resolución. La idea. Lo que se pide aquí no es resolver una ecuación nueva —es el Volterra del ejercicio anterior con otros números— sino **separar el modelo de sus parámetros**, que es la decisión de diseño que hace utilizable un programa. Un segundo miembro con los coeficientes escritos dentro sirve para un caso; uno que los reciba de fuera sirve para todos, y permite recorrer el espacio de parámetros sin editar el fichero, que es lo que se hace en las prácticas del capítulo 2 con β y γ y en las del 4 con μ. Las dos formas de conseguirlo —envoltura anónima y estructura— no son dos métodos numéricos: son dos maneras de escribir lo mismo, y por eso el ejercicio pide las dos.
El núcleo del programa no es ninguna de las dos llamadas a `ode45`, que hacen lo mismo, sino las dos líneas que construyen `fun1` y `fun2`: ahí es donde los parámetros entran en el modelo, y ahí es donde puede colarse un error que el ordenador no detecta.
El siguiente bloque se guarda como `solucion_volterra.m`. Contiene el programa y, al final, las dos funciones locales. Así pueden compararse ambas versiones en una misma ejecución.

```
%% Ejercicio 11 - las dos maneras de pasar parametros
a = 0.5; b = 0.2; c = 0.5; d = 0.1;
params.a = a; params.b = b; params.c = c; params.d = d;
opc = odeset('RelTol', 1e-10, 'AbsTol', 1e-12);

```

```
fun1 = @(t,y) odeparam(t, y, a, b, c, d); % sin estructura
fun2 = @(t,y) odeparamS(t, y, params); % con estructura

s1 = ode45(fun1, [0, 25], [4; 4], opc);
s2 = ode45(fun2, [0, 25], [4; 4], opc);
tt = linspace(0, 25, 2001);
Y1 = deval(s1, tt); Y2 = deval(s2, tt);

plot(Y1(1,:), Y1(2,:), 'b', 'LineWidth', 1.4), hold on
plot(Y2(1,:), Y2(2,:), 'r--', 'LineWidth', 1.4)
plot(c/d, a/b, 'k.', 'MarkerSize', 18) % equilibrio interior (5, 2.5)
plot(4, 4, 'r.', 'MarkerSize', 18), hold off
xlabel('y_1'), ylabel('y_2'), title('Orbita de Volterra'), grid on
legend('sin estructura', 'con estructura', 'equilibrio', 'dato inicial')

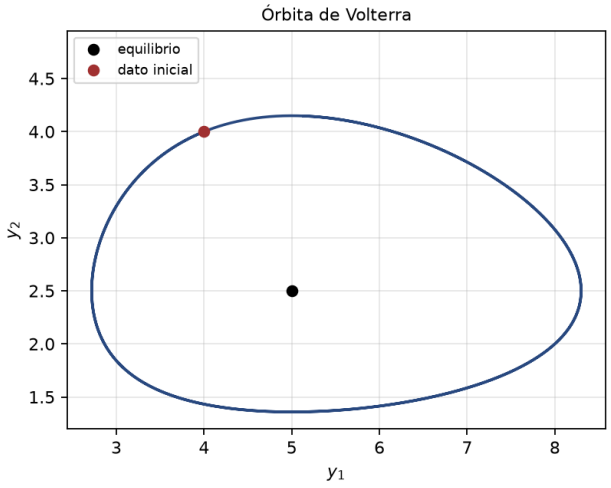
fprintf('Diferencia maxima entre las dos versiones: %.2e \n', max(max(abs(Y1 - Y2))))
fprintf('y1 entre %.6f y %.6f \n', min(Y1(1,:)), max(Y1(1,:)))
fprintf('y2 entre %.6f y %.6f \n', min(Y1(2,:)), max(Y1(2,:)))

function dy = odeparam(t, y, a, b, c, d)
dy = [a*y(1) - b*y(1)*y(2);
      -c*y(2) + d*y(1)*y(2)];
end

function dy = odeparamS(t, y, params)
dy = [params.a*y(1) - params.b*y(1)*y(2);
      -params.c*y(2) + params.d*y(1)*y(2)];
end
```

SALIDA ESPERADA

```
Diferencia maxima entre las dos versiones: 0.00e+00
y1 entre 2.718011 y 8.299465
y2 entre 1.359005 y 4.149729
```



La órbita del Volterra con $a = c = 0.5$, $b = 0.2$, $d = 0.1$, cerrada alrededor del equilibrio $(5, 2.5)$. Las dos versiones del programa dibujan la misma curva.

Las dos versiones coinciden bit a bit, y así debe ser: no son dos métodos, son dos maneras de escribir el mismo segundo miembro. Si dieran cifras distintas, habría que revisar las fórmulas, los parámetros y las opciones de ambas llamadas.
Y otra vez el factor dos. Los extremos de la órbita completa son aproximadamente $y_1 \in [2.718010, 8.299466]$ y $y_2 \in [1.359005, 4.149733]$. La salida anterior aproxima esos extremos mediante una malla. Los extremos exactos de y_1 son el doble de los de y_2, por una simetría análoga a la del ejercicio 9: aquí $a = c = 0.5$. Con $u = (d/c)\,y_1 = 0.2\,y_1$ y $v = (b/a)\,y_2 = 0.4\,y_2$ la integral primera se escribe $c(u - \log u) + a(v - \log v)$, y al ser $a = c$ los dos sumandos entran con el mismo peso: los extremos de u y los de v vuelven a resolver la misma ecuación. Al deshacer el cambio, el cociente de escalas es $0.4/0.2 = 2$. Si a y c fueran distintos, la simetría se rompería y no habría ningún factor limpio.
El periodo, estimado por el siguiente retorno a la sección $y_1 = 4$ con el mismo sentido de cruce, es $P \approx 12.889116$, frente a los $2\pi/\sqrt{ac} = 12.566371$ de las oscilaciones pequeñas: un 2.6 % más. Comparado con el 19 % del ejercicio 9, se ve que esta órbita es mucho más pequeña en relación con su equilibrio, y por eso la aproximación lineal acierta mucho mejor.
Los dos controles, y hacen falta los dos. Que las dos versiones coincidan comprueba que están *escritas* igual, no que estén *bien*: si los dos ficheros tienen el mismo error, las dos salidas lo tienen también. El control que sí es independiente vuelve a ser la integral primera, aquí $H = c(u - \log u) + a(v - \log v)$ con $u = (d/c)y_1$ y $v = (b/a)y_2$, que vale $H_0 \approx 1.0765700$ en el dato inicial. Se calcula su variación sobre la solución numérica y se contrasta al reducir las tolerancias.
Dónde se rompe, y es la rotura que no da ningún error. Basta cruzar `b` y `d` al montar la envoltura —escribir `odeparam(t, y, a, d, c, b)`— para que el programa funcione, dibuje una órbita cerrada perfectamente creíble e imprima seis decimales. Lo que devuelve es

$$y_1 \in [1.3590,\ 4.1497], \qquad y_2 \in [2.7180,\ 8.2995],$$

que son **los dos intervalos de arriba intercambiados**: han cambiado las escalas de población del sistema. Nada en la salida avisa. Lo que lo caza es la comprobación previa de siempre: el equilibrio interior tiene que estar en $(c/d,\ a/b) = (5,\ 2.5)$, y en la figura cruzada la órbita rodea el punto $(2.5,\ 5)$. Por eso el programa dibuja el equilibrio calculado a mano encima de la órbita: no es adorno, es el control.
Es también el argumento a favor de la estructura frente a la envoltura. Con `params.b` y `params.d` es más visible qué dato se usa, pero intercambiar dos campos válidos tampoco produce necesariamente un error de ejecución. La estructura facilita revisar el modelo; no sustituye esa revisión.

QUÉ DEJA. Un segundo miembro para `ode45` es siempre una función de dos argumentos que devuelve una columna, y esa restricción se sortea envolviendo en una función anónima, que congela los parámetros, o agrupándolos en una estructura. Una ecuación de orden n se convierte en un sistema llamando a las derivadas $y_2, \ldots, y_n$, y la última fila es la ecuación original despejada. Y con dos incógnitas hay dos dibujos que dicen cosas distintas: la trayectoria dice cuándo, la órbita dice qué forma.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_ivp
```

```
def odeparam(t, y, a, b, c, d):
    return [a * y[0] - b * y[0] * y[1], -c * y[1] + d * y[0] * y[1]]
a, b, c, d = 0.5, 0.2, 0.5, 0.1
s = solve_ivp(lambda t, y: odeparam(t, y, a, b, c, d), (0, 25), [4., 4.],
              dense_output=True, rtol=1e-10, atol=1e-12) # args=(a,b,c,d) tambien vale
Y = s.sol(np.linspace(0, 25, 2001))
print(Y[0].min(), Y[0].max(), Y[1].min(), Y[1].max())
```

Lo mismo en Python

Cada práctica del libro lleva debajo su versión en Python, como hacía el libro previo del autor, para poder comparar los dos entornos. Python es gratuito y se descarga de `python.org`; las bibliotecas que hacen las veces de MATLAB son **NumPy** (vectores y matrices), **Matplotlib** (gráficos) y **SciPy** (integración de ecuaciones diferenciales, problemas de contorno, optimización). Se instalan una sola vez desde una terminal:

```
python3 -m venv entorno
source entorno/bin/activate
python -m pip install numpy matplotlib scipy
```

La primera línea crea una carpeta `entorno` con un Python propio, la segunda lo activa en macOS o Linux, y la tercera instala las bibliotecas dentro. En Windows se usa `py -m venv entorno` para crearlo y, desde `cmd`, `entorno\Scripts\activate.bat` para activarlo. Con el entorno activo, un fichero `.py` se ejecuta con `python nombre.py`. La tabla que sigue recoge las equivalencias que se usan en el libro; conviene tenerla a mano al leer una versión en Python.

MATLAB	Python	Observación
`[1, 2, 3]`	`np.array([1, 2, 3])`	los índices empiezan en 0
`v(end)`	`v[-1]`	
`v(2:4)`	`v[1:4]`	el extremo derecho no entra
`A(i,j)`	`A[i-1, j-1]`	
`A * B`	`A @ B`	producto de matrices
`A .* B`	`A * B`	en NumPy `*` es elemento a elemento
`A^k`	`np.linalg.matrix_power(A, k)`	
`A.'`	`A.T`	matriz; sin conjugación
`A'`	`A.conj().T`	traspuesta conjugada
`linspace(a,b,n)`	`np.linspace(a, b, n)`	
`@(x) expr`	`lambda x: expr`	
`polyfit, polyval`	`np.polyfit, np.polyval`	mismo orden de coeficientes
`interp1(x,y,t)`	`np.interp(t, x, y)`	datos 1D; dentro del intervalo
`eig(A)`	`np.linalg.eigvals(A)`	solo autovalores
`norm(v,1)`	`np.linalg.norm(v, 1)`	
`find(cond)`	`np.where(cond)[0]`	
`load('f.dat')`	`np.loadtxt('f.dat')`	
`plot, hold on`	`plt.plot`	Matplotlib superpone por defecto
`legend, xlabel, title`	`plt.legend, plt.xlabel, plt.title`	
`ode45`	`scipy.integrate.solve_ivp`	sección anterior
`deval(sol,t)`	`sol.sol(t)`	con `dense_output=True`
`odeset('RelTol',r)`	`rtol=r`	argumento, no estructura
`fzero(g,[a,b])`	`brentq(g, a, b)`	cambio de signo
`bvp4c`	`scipy.integrate.solve_bvp`	capítulo 3

En NumPy, un vector `v` de una dimensión no cambia al escribir `v.T`; `v[:,None]` lo convierte en una columna bidimensional. `np.interp` mantiene por defecto el valor del extremo fuera del intervalo, mientras `interp1` lineal devuelve `NaN`. Para autovectores, `[V,D]=eig(A)` en

MATLAB corresponde a `valores,V=np.linalg.eig(A)`: el orden y la forma de las salidas son distintos. En la tabla, `brentq` se importa de `scipy.optimize`.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
import matplotlib.pyplot as plt

a = 1
a = a + 2
print(a) # 3
v = np.array([1, 2, 3])
print(v ** 2) # [1 4 9] (elemento a elemento)
print(v @ v) # 14 (producto escalar)
ffun = lambda x: np.sqrt(x ** 2 + 10)
print(ffun(np.array([1, 2, 3]))) # [3.3166 3.7417 4.3589]
x = np.array([5, 0, -1, 3, 7])
print(np.where(x < 4)[0] + 1) # [2 3 4] (posiciones, sumando 1)
t = np.linspace(0, 2 * np.pi, 100)
plt.plot(t, np.sin(t), label="seno")
plt.plot(t, np.cos(t), "r--", label="coseno")
plt.legend(); plt.xlabel("t"); plt.grid(True)
plt.show()
```

REPASO INTERCALADO: SESIÓN CERO

Con el libro cerrado:

1. ¿Qué diferencia hay entre `v*w` y `v.*w` si `v` y `w` son dos vectores fila de la misma longitud? ¿Cuál de las dos da error?
2. Escribe una función anónima que valga x^2 para $x < 1$ y $2x - 1$ para $x \geq 1$, sin usar `if`.
3. ¿Dónde busca MATLAB una M-función y qué diferencia hay entre sus variables locales y las del espacio de trabajo principal?
4. Se ejecuta `plot(x,y)` y después `plot(x,z)`. ¿Qué se ve en la figura, y qué línea faltaba?
5. ¿Qué devuelve `diag([1,2,3],-1)` y de qué tamaño es?

CAPÍTULO 1

Algunos modelos discretos

ANTES DE LEER

Intenta estas cuatro preguntas antes de leer. Sirven para activar lo que ya sabes y localizar lo que falta; contrasta tus respuestas al llegar a las secciones indicadas.

1. Una población de hembras se cuenta cada cinco años, repartida en cuatro clases de edad. ¿Qué dos números por clase bastan para predecir el censo siguiente, y por qué no hacen falta más?
2. Si multiplicas veinte veces un vector por la misma matriz, ¿de qué depende que el resultado se estire en una dirección fija, y de qué depende que oscile sin parar?
3. ¿Qué tiene que cumplir una matriz para que un teorema le garantice un autovalor real, positivo y mayor en módulo que todos los demás, sin calcular ninguno?
4. Un buscador tiene que ordenar millones de páginas sin leerlas. ¿Qué información, que no sea el contenido, dice cuál es más importante?

Las respuestas están, respectivamente, en las secciones 1.1.2, 1.1.3, 1.1.3.3 y 1.2.1.

Este capítulo presenta dos modelos discretos que usan sobre todo herramientas del Álgebra Lineal. El primero es un modelo de Dinámica de Poblaciones, el modelo de Leslie, que permite determinar bajo ciertas hipótesis cómo evoluciona el número de individuos de una población. El segundo es el método de cálculo del PageRank, desarrollado por los creadores de Google, que mide la relevancia de una página web a partir de un grafo dirigido. Desde el punto de vista matemático, los dos conducen al mismo sitio: al método de la potencia para calcular el autovalor dominante de una matriz, y al teorema que garantiza que ese autovalor existe y se comporta bien.

1.1 El modelo de Leslie

La pregunta

Una demógrafa cuenta una población de tortugas marinas y observa que casi ninguna cría llega a adulta. Se propone proteger los nidos para multiplicar los nacimientos. Con los datos que tiene, fracción de supervivencia y número medio de crías en cada edad, ¿puede saber *antes* de gastar el dinero si la población va a extinguirse o a crecer, y en qué edad hay que intervenir para que cambie el destino?

1.1.1 Introducción

El modelo se llama de Leslie en honor a su autor, Patrick Holt Leslie (1900–1974). El objetivo de la Dinámica de Poblaciones es describir cómo evolucionan los individuos de un colectivo de seres vivos tras reproducirse, fallecer, migrar o interactuar. Hay cuatro procesos que cambian el tamaño de una población: los nacimientos y las inmigraciones la aumentan; las defunciones y las emigraciones la disminuyen. **El modelo de Leslie se aplica a poblaciones que no se ven afectadas por migraciones.**

Tiene una segunda característica importante: **se contabilizan solo las hembras**, incluidas las clases que aún no se reproducen o ya han dejado de hacerlo. Para usarlo en poblaciones con machos y hembras se impone como hipótesis que el número de unos y otras coincide, o guarda una proporción fija, de modo que al final se obtiene una estimación del tamaño de la población completa.

Y una tercera: es razonable suponer que las tasas de mortalidad y de natalidad dependen de la edad. Hay poblaciones en que los individuos pasan por etapas vitales muy distintas, como los insectos que van de huevo a larva y de larva a adulto. Para tratar correctamente esas diferencias, el modelo hace dos cosas:

- divide la población en **clases de edad** dentro de las cuales la natalidad y la supervivencia se pueden considerar homogéneas;
- fija las clases con rangos de edades de amplitud constante τ y observa la población en incrementos de tiempo de esa misma longitud τ, para garantizar que, de un recuento al siguiente, todos los individuos de una clase que han sobrevivido están en la clase inmediatamente superior.

Estas tres decisiones (sin migraciones, solo hembras, clases equitemporales) son las que hacen el modelo, y merecen cada una su discusión. El libro previo del autor las discutió en forma de diálogo, y ese diálogo se conserva aquí porque las preguntas que plantea son exactamente las que hay que hacerse. Las intervenciones del autor van en cursiva y entre comillas, tal como se escribieron entonces.

> «*Queremos modelar una población. Asumimos que solo hay muertes y nacimientos. El modelo de Leslie descarta migraciones. ¿O eso es solo el caso que me han presentado? Las preguntas que también tenemos que responder son: ¿qué es lo que caracteriza al modelo de Leslie diferencialmente a otros? ¿Es el más simple o los hay aún más simples pero de utilidad? ¿O es este tan sencillo que deja de ser útil y es solo para construir otros? ¿O hay ocasiones en que este sí es útil?*

Eso de quitar las migraciones de en medio… entiendo que en cierto sentido lo que va a hacer es que el crecimiento y decrecimiento sea proporcional a la población que ya hay (mientras que si hay migraciones el hecho de que la población actual sea pequeña no dice necesariamente nada acerca de la cantidad de quienes vienen). ¿Es ese el motivo o es otro?»

Qué caracteriza al modelo y si es útil. El rasgo diferencial del modelo de Leslie es que introduce la **estructura interna** de la población. No es el más simple: el modelo discreto más simple es el de Malthus, $N_{k+1} = \lambda N_k$, en el que solo se describe el total y se supone un factor de crecimiento constante. No se distinguen las tasas por edad: el modelo no permite preguntar qué cambia al sustituir crías por adultas. Leslie da el salto de sustituir el escalar N_k por un vector, con una componente por clase de edad, y el escalar λ por una matriz. Y es útil de verdad: el ejemplo clásico es la conservación de las tortugas marinas de la pregunta que abre la sección. Un estudio clásico usó una matriz por *estadios*, del tipo de Lefkovitch que se presenta abajo, y encontró una mayor respuesta del crecimiento a mejorar la supervivencia de juveniles y subadultas que a proteger los huevos. Eso apoyaba intervenir también sobre la captura accidental en las redes.[1] La herramienta que permite ese diagnóstico es la que se construye en este capítulo: el autovalor dominante de la matriz y su sensibilidad a cada coeficiente.

EL CAMINO QUE NO SE TOMÓ: EL MODELO DE MALTHUS, UN SOLO NÚMERO

La alternativa. Describir la población por su tamaño total N_k y suponer que crece o decrece en proporción fija: $N_{k+1} = \lambda N_k$, luego $N_k = \lambda^k N_0$.
Su ventaja. Un solo parámetro, una sola cuenta, y captura lo esencial del crecimiento geométrico que Malthus describió en 1798. Sigue siendo el primer modelo que se ajusta a cualquier serie de censos, y en el capítulo 2 reaparece en versión continua como modelo exponencial.
Qué lo hunde. No distingue edades, y por eso no puede contestar a la pregunta de la demógrafa: en qué clase intervenir. Dos poblaciones con el mismo total pero una formada por crías y otra por adultas fértiles tienen futuros completamente distintos, y Malthus les asigna el mismo. Leslie es lo mínimo que hace falta para que la composición por edades importe.

La ausencia de migraciones. La intuición de la cita es correcta: al excluir la migración, la variación de la población se debe solo a procesos internos, nacimientos y muertes. Para que además sea **lineal** se supone que cada hembra de una clase aporta, en promedio, la misma fecundidad y supervivencia, independientemente de la densidad. Con tasas constantes resulta $v(k+1) = Lv(k)$: dados L y $v(0)$, el modelo fija el futuro, y eso permite aplicar la maquinaria del Álgebra Lineal, en concreto el teorema de Perron-Frobenius que se verá en la sección 1.1.3.3. Con migración el sistema sería $v(k+1) = Lv(k) + m(k)$, si la migración se representa mediante una entrada externa $m(k)$. Entonces hay que estudiar también esa entrada: el autovalor de L ya no basta por sí solo para decidir el comportamiento a largo plazo. El objetivo del modelo no es predecir el censo exacto del año que viene, sino responder a una pregunta más honda: con las tasas vitales actuales, y si la población estuviera aislada, ¿tiende a extinguirse o a crecer? Para medir esa «salud demográfica intrínseca» hay que aislarla de la migración.

[1] Crouse, Crowder y Caswell (1987), *A Stage-Based Population Model for Loggerhead Sea Turtles and Implications for Conservation*, *Ecology* 68, 1412–1423, `https://doi.org/10.2307/1939225`. Es un modelo por estadios, no una matriz de Leslie pura; comparte con ella el análisis mediante matrices no negativas.

> *«Por otro lado, se considera que todos los sujetos son hembras. A primera vista, con decir que "las hembras son la mitad de la población" (o una proporción constante) uno puede decir directamente la población total. Si el axioma es "la proporción de hembras y población total es constante" lo mismo me da tomar todo hembras que tomar a toda la población con una tasa de fertilidad multiplicada por la proporción de hembras, ¿no? Entonces… ¿hay otro motivo por el cual hacemos esa suposición de todo hembras?»*

Solo hembras. Si la fracción de hembras es una constante $p > 0$, igual en todas las clases, los vectores de hembras $F(k)$ y de población total $N(k)$ cumplen $F(k) = pN(k)$. Por tanto, $F(k+1) = LF(k)$ equivale a $N(k+1) = LN(k)$: *la misma matriz* cuando la conversión es uniforme. La fecundidad por hembra se mide en hijas; al contar toda la población se convierten a la vez progenitoras y descendencia.

El modelo de un solo sexo supone que la disponibilidad de machos no limita los nacimientos. Bajo esa hipótesis, seguir las hembras simplifica el cálculo. Si faltan parejas o las proporciones por edad cambian, puede ser necesario modelar ambos sexos; prescindir de uno no garantiza conservar toda la capacidad predictiva.

> *«Por otro lado, se considera que el número de nacimientos depende de la cantidad de hembras en una cierta edad. Entonces dividimos a las hembras en clases de edad, y por simplicidad entiendo que cogemos esas clases como equitemporales. Aquí se menciona el fenómeno del "cambio de fase" (huevo, larva, adulto), por cierto, de manera que para captar bien las circunstancias concretas del tipo de población manteniendo la equitemporalidad tendremos que reducir los intervalos temporales lo suficiente para que las suposiciones sean correctas (por ejemplo, la homogeneidad dentro de esa clase).»*

Las clases equitemporales. Aquí la cita señala la mayor limitación práctica del modelo. La matriz de Leslie solo tiene entradas no nulas en la primera fila y en la subdiagonal, y eso exige que el paso de tiempo sea exactamente la amplitud de la clase: en un paso, los supervivientes de la clase 1 pasan forzosamente a la clase 2. Si un insecto es huevo tres días, larva diez y adulto dos, un paso de un día y quince clases permite respetar exactamente esas tres duraciones. Es una descripción más detallada de la que se necesita si solo interesan las tres fases. La solución que adoptaron los ecólogos es la que la cita casi adivina, y se cuenta en la caja siguiente.

EL CAMINO QUE NO SE TOMÓ: CLASES POR ESTADIOS EN VEZ DE POR EDADES

La alternativa. Permitir que un individuo *permanezca* en su clase de un recuento al siguiente, añadiendo entradas P_i en la diagonal principal de la matriz. Es el modelo por estadios de Lefkovitch (1965), que se usa hoy en ecología tanto o más que el de Leslie.
Su ventaja. Las clases pueden ser huevo, larva y adulto, sin importar cuántos días dure cada fase; el paso de tiempo se desliga de la anchura de la clase y el modelo capta fenómenos como las plantas que no crecen si no hay luz.
Qué lo aparta aquí. No lo hunde nada: sigue vigente. Se deja fuera porque la asignatura estudia el caso de Leslie, en el que la estructura de la matriz permite escribir el autovector dominante con una fórmula explícita (sección 1.1.3.4) y demostrar la irreducibilidad con un solo ciclo. Con entradas en la diagonal, la teoría de Perron-Frobenius sigue aplicándose bajo sus hipótesis, pero esa fórmula concreta cambia.

1.1.2 Deducción del modelo

Consideremos una población formada por hembras y supongamos que la edad máxima que puede alcanzar una hembra es $E > 0$. Fijamos un entero positivo n, definimos $\tau = E/n$ y dividimos la población en n clases $C_1, \ldots, C_n$: la clase C_i está formada por los individuos de edad en el intervalo $[(i-1)\tau,\, i\tau)$ si $1 \leq i \leq n-1$, y por los de edad en $[(n-1)\tau,\, E]$ si $i = n$. Lo que interesa es cuántos individuos hay en cada clase en los tiempos $t_0 = 0$, $t_1 = \tau$, $t_2 = 2\tau$, y así sucesivamente.

Definición. Se denota $v(k)$ al vector de $\mathbb{R}^n$ cuya componente i-ésima, $v_i(k)$, es el número de hembras de la clase C_i en el instante $t_k = k\tau$. Se dice que $v(k)$ es el **vector de distribución por clases** (o por edades) en el instante t_k.

Es decir, $v_i(k)$ es el número de individuos del i-ésimo rango de edad en el k-ésimo instante de tiempo: el subíndice es la edad, el argumento es el tiempo.

Para modelar la natalidad y la supervivencia se suponen conocidos dos parámetros demográficos por clase:

1. la **fecundidad** $a_i \geq 0$, $1 \leq i \leq n$: promedio del número de hijas que tiene una hembra durante el tiempo τ que permanece en la clase C_i;
2. la **supervivencia** $b_i \in [0,1]$, $1 \leq i \leq n-1$: fracción de hembras de la clase C_i que se espera que sobreviva y pase a la clase C_{i+1} en un intervalo de tiempo de longitud τ.

Se admite que existe al menos un $a_i > 0$ y que todos los $b_j > 0$. Las dos hipótesis tienen su porqué. Si todos los a_i fueran nulos no habría nacimientos, y la población se extinguiría en tiempo finito: como máximo en n pasos, que es lo que tarda la clase más joven en llegar a la última y desaparecer. Si algún $b_r = 0$, las hembras de C_r no pasan a C_{r+1}. Se puede recortar a las r primeras clases cuando las posteriores están inicialmente vacías; si ya hay hembras en ellas, hay que conservar su supervivencia y sus posibles hijas durante el transitorio. El corte impide cruzar una edad, no fija una fecha de extinción de toda la población.

Con esto, fijado un instante t_k, se deben cumplir dos relaciones.

- Las hembras de la clase C_1 en el instante t_{k+1} son las nacidas entre t_k y t_{k+1}, y cada clase aporta las hijas de sus miembros:

$$v_1(k+1) = a_1 v_1(k) + a_2 v_2(k) + \cdots + a_n v_n(k).$$

- Las hembras de la clase C_i, con $2 \leq i \leq n$, en el instante t_{k+1} son las de la clase C_{i-1} en t_k que han sobrevivido:

$$v_i(k+1) = b_{i-1}\, v_{i-1}(k) \qquad \text{para todo } i = 2, 3, \ldots, n.$$

Las dos relaciones se expresan como una sola igualdad matricial.

Definición. El modelo de Leslie es la ecuación en diferencias

$$v(k+1) = L\, v(k) \qquad \text{para todo } k \geq 0,$$

donde $L \in \mathbb{R}^{n\times n}$, la **matriz de Leslie**, es

$$L = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ b_1 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & b_{n-1} & 0 \end{pmatrix}.$$

Fecundidades en la primera fila, supervivencias en la subdiagonal, y ceros en todo lo demás. Iterando, $v(1) = Lv(0)$, $v(2) = Lv(1) = L^2v(0)$, y en general

$$v(k) = L^k\, v(0) \qquad \text{para todo } k \geq 0.$$

Por tanto, a partir de la distribución inicial $v(0)$ y de la matriz L se puede determinar la distribución de las hembras en cualquier tiempo futuro t_k.

LA IDEA: LEER LA MATRIZ COMO UN AÑO DE VIDA

Multiplicar por L es *avanzar un período*. La fila 1 recoge: «cada clase envía sus hijas a la cuna». Cada fila $i \geq 2$ recoge: «la clase i del año que viene son los supervivientes de la clase $i-1$ de este año». Nada más ocurre. En el grafo que se dibujará en la sección 1.1.3.1, esto son dos tipos de flecha: envejecer, de i a $i+1$, y nacer, de i a 1. Todo lo que el modelo puede hacer está en esas flechas.

> *«En principio el hecho de que Leslie sea lineal es porque siempre vamos a poder encapsular un sistema autónomo aplicado sobre un vector (en lugar de una función) vía una matriz, ¿no? (esa matriz puede tener componentes no constantes, obviamente; en concreto aquí no sé si será constante, creo que a priori sí, si es que las tasas de fertilidad no cambian, que creo que ahí está el quid de la cuestión), ¿o todo este análisis es incorrecto por mi parte?»*

Este razonamiento del libro previo contiene un error que conviene dejar señalado, porque es instructivo. No es cierto que todo sistema autónomo sobre un vector se encapsule en una matriz. Tres casos distintos:

- si L dependiera del tiempo, $v(k+1) = L(k)v(k)$, el sistema sería lineal pero *no autónomo*;
- si L dependiera del propio estado, $v(k+1) = L(v(k))\, v(k)$, el sistema sería autónomo y, en general, *no lineal*: es lo que en ecología se llama densodependencia, y en el capítulo 2 el modelo logístico es exactamente eso en versión continua;
- el modelo de Leslie es lineal y autónomo porque se impone un supuesto biológico explícito: fecundidades y supervivencias *constantes*, que no cambian ni con el tiempo ni con el tamaño de la población.

Un sistema es autónomo cuando su ley de evolución no depende explícitamente del tiempo: $v(k+1) = f(v(k))$ es autónomo, $v(k+1) = f(v(k), k)$ no lo es. Añadir un término $b(v(k))$ conserva la autonomía y puede romper la linealidad (si b es lineal, se conserva). Una entrada externa $b(k)$ no constante hace que la ley dependa del tiempo; una entrada constante sigue dando un sistema autónomo, aunque afín. Y «cerrado» quiere decir sin flujos a través de la frontera del sistema: la población cambia solo por procesos internos. En el capítulo 2, el modelo SIR es cerrado porque $S + I + R$ se conserva.

CONEXIÓN: ECUACIONES DIFERENCIALES ORDINARIAS, EL ANÁLOGO CONTINUO

El modelo de Leslie es el hermano discreto del sistema lineal $y' = Ay$ con A constante que se estudió en Ecuaciones Diferenciales Ordinarias, y el apéndice A lo reenuncia con detalle. La correspondencia exacta es esta: la solución de $y' = Ay$ con $y(0) = y_0$ es $y(t) = e^{tA}y_0$, y la del modelo de Leslie es $v(k) = L^k v(0)$. En el caso continuo, la estabilidad del origen la deciden los signos de las *partes reales* de los autovalores de A; en el discreto la deciden los *módulos* de los autovalores de L: todo tiende a cero si y solo si todos tienen módulo menor que 1. La segunda cita de esta sección apuntaba, además, a que un sistema no lineal se estudia cerca de un equilibrio por su linealizado; eso es el primer método de Lyapunov, que el apéndice recoge

> y que el capítulo 2 usa a cada paso. Y en versión discreta funciona igual: si un modelo no lineal conserva la estructura de nacimientos y envejecimiento de Leslie, su linealización en el origen tiene esa misma forma, con las tasas evaluadas allí. La estructura por edades, por sí sola, no obliga a cualquier jacobiana a ser de Leslie.

Ejemplo. Este ejemplo acompañará todo el capítulo, y será el caso de prueba de la práctica 4. Para estudiar una especie cuya edad máxima es de 20 años se consideran períodos de 5 años, con lo que la población queda dividida en 4 clases. De la observación se deduce que solo una cuarta parte de los individuos del primer grupo (1–5 años) sobrevive hasta el siguiente período; que solo la mitad de los del segundo (6–10 años) sobreviven hasta los 11–15 años, y solo una décima parte de estos alcanza el último grupo (16–20 años). Se observa también que cada individuo del segundo grupo procrea en promedio un nuevo individuo, y los de los grupos 3 y 4 procrean 3 y 2, respectivamente.

Con la notación anterior, $E = 20$, $\tau = 5$ y $n = E/\tau = 4$; las clases son C_1 para $[0, 5)$, C_2 para $[5, 10)$, C_3 para $[10, 15)$ y C_4 para $[15, 20]$, y la población se observa en $t_0 = 0$, $t_1 = 5$, $t_2 = 10$, etc. La matriz de Leslie es

$$L = \begin{pmatrix} 0 & 1 & 3 & 2 \\ 0.25 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{pmatrix}.$$

Si la distribución inicial por grupos es de 100, 70, 70 y 40 hembras, la distribución dentro de 10 años es $v(2) = L^2 v(0)$. Hagamos la cuenta en dos pasos, para ver actuar cada fila. Primero,

$$v(1) = Lv(0) = \begin{pmatrix} 0 \cdot 100 + 1 \cdot 70 + 3 \cdot 70 + 2 \cdot 40 \\ 0.25 \cdot 100 \\ 0.5 \cdot 70 \\ 0.1 \cdot 70 \end{pmatrix} = \begin{pmatrix} 360 \\ 25 \\ 35 \\ 7 \end{pmatrix},$$

y después

$$v(2) = Lv(1) = \begin{pmatrix} 1 \cdot 25 + 3 \cdot 35 + 2 \cdot 7 \\ 0.25 \cdot 360 \\ 0.5 \cdot 25 \\ 0.1 \cdot 35 \end{pmatrix} = \begin{pmatrix} 144 \\ 90 \\ 12.5 \\ 3.5 \end{pmatrix},$$

Obsérvese lo que ha pasado en dos pasos: el total ha ido de 280 a 427 y de ahí a 250. Sube y baja. Que no se pueda decidir con dos pasos si la población crece o se extingue es justo lo que motiva la sección siguiente.

EXPLÍCATE. *En el cálculo de $v(1)$, la cuarta componente es $0.1 \cdot 70$, y no $0.1 \cdot 40$. Di por qué: ¿quiénes son las hembras de la clase 4 en t_1?*

1.1.3 Comportamiento a largo plazo de la población

LA PREGUNTA

Con la matriz del ejemplo, dos pasos no bastan para saber si la población crece o se extingue. ¿Existe alguna cantidad, calculable de una vez con los parámetros de la matriz, que decida el destino de la población sin necesidad de simular, y que además no dependa de con cuántas hembras se empiece?

Supongamos que estamos estudiando una población de hembras con el modelo de Leslie, de modo que $v(k) = L^k v(0)$ con L la matriz anterior. Interesa el **comportamiento asintótico**: cómo se comporta $v(k)$ cuando k crece hacia $+\infty$. Esa información dirá qué le depara a la población el futuro y, en particular, si está destinada a la extinción o si crecerá indefinidamente. Surgen de manera natural tres cuestiones:

1. ¿Es el comportamiento de los $v(k)$ para grandes valores de k independiente de $v(0)$?
2. ¿Existe el límite $v_\infty := \lim_{k\to+\infty} v(k)$ en $\mathbb{R}^n$?
3. En tal caso, ¿son sus componentes positivas? ¿Guarda alguna relación con los autovalores y autovectores de L?

> *«Cuando se pregunta por el comportamiento a largo plazo se preguntan estas tres cosas. ¿En principio se podría haber preguntado otras, no? En sistemas diferenciales nos preguntamos por estabilidad y por atractividad, y sus correspondientes casos uniformes o exponenciales. Entiendo que la estabilidad guarda relación con la idea de la primera pregunta: ¿depende el comportamiento de la población del punto inicial? La atractividad responde a la segunda. Y la tercera es si es viable como modelo o da cosas absurdas, y por supuesto su comportamiento en relación al espectro de la matriz.»*

La lectura es correcta en lo esencial, con dos matices. La primera pregunta es lo que en demografía se llama **ergodicidad**: la población «olvida» su estructura inicial de edades y se alinea con una dinámica dictada solo por la matriz. La segunda no es exactamente atractividad hacia un punto: en un sistema lineal, salvo que la población se extinga y vaya al origen, no hay ningún punto de equilibrio no trivial que atraiga. Lo que atrae es una **dirección**, una recta por el origen, y esa dirección es la del autovector dominante. Cuando el vector de población se alinea con ella, la proporción entre clases se congela aunque el total siga creciendo o decreciendo: es lo que se llama alcanzar la distribución estable de edades. Y la tercera no es de viabilidad, sino de clasificar el comportamiento del origen: en un sistema continuo $y' = Ay$ se mira el signo de la parte real de los autovalores; en uno discreto se mira si el módulo del autovalor dominante es mayor, menor o igual que 1.

Para responder hacen falta tres nociones del Álgebra Matricial. Se dan aquí primero con su definición exacta, y en la subsección 1.1.3.1 se desmonta lo que significan. En los resultados generales siguientes tomamos $n \geq 2$. El caso 1×1 se lee directamente: $[0]$ es irreducible por la definición de caminos entre vértices distintos, pero su radio espectral es cero. Si $A = [a]$ con $a > 0$, es primitiva y su único autovalor es a.

Definición. Una matriz cuadrada A de orden n es **irreducible** si, para cada par de índices i y j con $i \neq j$, existen índices $i_1, \ldots, i_m$ tales que

$$a_{i,i_1}\, a_{i_1,i_2} \cdots a_{i_m,j} \neq 0.$$

Definición. Sea A una matriz cuadrada de orden n. Se le asocia un **grafo dirigido** G_A de n vértices de modo que hay una arista desde el vértice j al vértice i si y solo si $a_{ij} \neq 0$. Entonces A es irreducible si y solo si G_A está **fuertemente conectado**, es decir, se puede llegar a cada vértice de G_A desde todos los demás.

Definición. Una matriz $A \geq 0$ se dice **primitiva** si existe un entero $m \geq 1$ tal que A^m tiene todas sus componentes positivas.

En lo que sigue se supone que en la matriz de Leslie es $a_n > 0$. La hipótesis se justifica así: si fuera $a_n = 0$, la clase C_n sería irrelevante para el comportamiento asintótico, porque las hembras que la componen no tienen capacidad reproductiva; luego C_n no afectaría a la evolución de C_1 y, por consiguiente, tampoco a la de las otras clases. Basta recortar la matriz a la última clase fértil. Y señalan un hecho que se demuestra en la subsección siguiente: $a_n > 0$ **si y solo si** L **es irreducible**.

> *«Y luego me sueltan un teorema que ya es aplicado a Leslie, donde se agrega a las condiciones que ya teníamos de Leslie que $a_n > 0$ (y justifican que esto sea así porque, según entiendo, puedes quitar todas las clases que ya no se reproducen, al igual que quitas a los hombres; no sé, me parece un poco fullero).»*

No es un truco: es aislar el motor reproductivo del sistema. Las hembras posreproductivas son un sumidero: consumen recursos y mueren, pero no devuelven individuos a la clase 1. Si se dejan en la matriz, esa parte del sistema no se comunica «hacia arriba» y la matriz se vuelve reducible, con lo que los teoremas que siguen no se aplican. Al truncar en la última edad fértil, todo estado alimenta al inicio del ciclo. El precio es que el modelo ya no cuenta a las abuelas; si se quiere su número, se recupera después aplicando las supervivencias a las clases fértiles, porque ellas no influyen en nadie.

Un primer resultado, que suele darse por bien conocido, se recuerda con su justificación porque es el punto de contacto con el Álgebra Lineal:

Proposición. $v(k) = L^k v(0)$ converge a cero, independientemente de $v(0)$, si y solo si el radio espectral de L es menor que 1, es decir, si y solo si todos los autovalores de L tienen módulo menor que 1. En ese caso la población va siempre hacia la extinción.

CONEXIÓN: ÁLGEBRA LINEAL, LA FORMA DE JORDAN Y LAS POTENCIAS DE UNA MATRIZ

Recuérdese el resultado que está detrás. Toda matriz $L \in \mathbb{R}^{n\times n}$ es semejante a su forma canónica de Jordan, $L = SJS^{-1}$, donde J es diagonal por bloques y cada bloque tiene un autovalor μ en la diagonal y unos en la superdiagonal. Entonces $L^k = SJ^kS^{-1}$, y la potencia de un bloque de Jordan de tamaño r tiene entradas de la forma $\binom{k}{j}\mu^{k-j}$ con $0 \leq j < r$. Si $|\mu| < 1$, cada una de esas entradas tiende a cero, porque un polinomio en k no puede con una exponencial que decrece; si $|\mu| \geq 1$, la entrada μ^k de la diagonal no tiende a cero. De ahí que $L^k \to 0$ si y solo si $\rho(L) < 1$, y esto es lo que dice la proposición. La misma descomposición es la que se usará para demostrar el teorema de convergencia de la sección 1.1.3.4.

Para obtener respuestas más completas hay que recurrir a los teoremas de Perron y de Frobenius. Antes de enunciarlos conviene entender de verdad qué es una matriz irreducible, porque es la hipótesis que lo sostiene todo.

1.1.3.1 Matrices irreducibles: lo que dice la definición y lo que modela

Esta subsección expande la definición anterior con el material del libro previo del autor, que la trabajó a fondo. La definición de arriba, con su cadena de subíndices, es opaca; lo que sigue la traduce.

LA IDEA: LA REGLA DE ORO DE LOS SUBÍNDICES

El número a_{ij}, fila i y columna j, significa «lo que sale *desde* el nodo j y llega *hacia* el nodo i». Es al revés de lo que uno leería, y la razón está en el producto Av: para calcular la componente i del resultado se usa la fila i entera,

$$(Av)_i = a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n,$$

o sea, «lo que sale de 1, de 2, …, de n entra en i». Con esta regla, el criterio de los subíndices deja de ser un jeroglífico: el producto $a_{i,i_1}a_{i_1,i_2}\cdots a_{i_m,j}$ es el producto de los pesos de las flechas del camino que sale de j, pasa por $i_m, \ldots, i_1$ y llega a i. Que sea distinto de cero significa que ninguna de esas flechas falta. La matriz es irreducible si todo nodo se alcanza desde todo nodo.

Hay una segunda forma de decir lo mismo, que es la que usan los textos de Álgebra Matricial y la que da nombre a la propiedad: A es **reducible** si existe una matriz de permutación P tal que

$$PAP^t = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix},$$

con X y Z bloques cuadrados de tamaño positivo y 0 un bloque rectangular de ceros. Multiplicar por P y por P^t no cambia el sistema dinámico: solo reordena las variables, es decir, les cambia el nombre. El bloque de ceros dice que las variables del grupo X no envían nada al grupo Z: Z evoluciona por su cuenta, y X recibe de Z a través de Y. Si no existe tal P, la matriz es irreducible.

Ejemplo. Un archipiélago de tres islas. El vector de estado $v = (v_1, v_2, v_3)^t$ es la población de cada isla; la matriz es el mapa de los puentes, con $a_{ij} = 1$ si hay puente de j a i. Si los puentes hacen un círculo $1 \to 2 \to 3 \to 1$,

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

y desde cualquier isla se llega a cualquier otra: es irreducible. Si en cambio las islas 1 y 2 están unidas en los dos sentidos, la 3 envía barcos a la 1, y de 1 y 2 no sale nada hacia la 3,

$$B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

quien desembarca en la isla 1 o en la 2 nunca llegará a la 3. La fila 3 es entera de ceros, y ese es ya el bloque de ceros: las islas 1 y 2 forman el bloque X y la 3 el bloque aislado Z. Con el criterio de los subíndices: de 3 a 2 hay camino, $a_{2,1}a_{1,3} = 1 \cdot 1 \neq 0$; de 1 a 3 no lo hay, porque toda cadena que termine en la fila 3 tiene un factor nulo.

Ejemplo. El mismo fenómeno en una matriz de Leslie, donde los números no son unos y ceros. Tres clases: crías, adultas con dos hijas de media, y abuelas sin hijas. Las crías sobreviven al 50 % y las adultas al 80 %:

$$L = \begin{pmatrix} 0 & 2.0 & 0 \\ 0.5 & 0 & 0 \\ 0 & 0.8 & 0 \end{pmatrix}.$$

La columna 3 es entera de ceros: las abuelas no producen crías, no rejuvenecen y después mueren. No hay ningún camino que salga de 3, luego L es **reducible**: las abuelas son un callejón sin salida. Para hacerla irreducible se borran las abuelas y queda la matriz 2×2 de crías y adultas, en la que $a_{1,2}a_{2,1} = 2.0 \cdot 0.5 \neq 0$ cierra el ciclo. A la topología del grafo le da igual si por un puente pasan diez personas o media: solo importa si el puente existe (> 0) o no ($= 0$).

Las distintas formas de decir «irreducible» son equivalentes, y merece la pena tener el teorema completo con su demostración, porque cada forma sirve para una cosa: la de bloques para el análisis teórico, la del grafo para la intuición, la de la potencia para programar, y la del producto de índices para demostrar a mano.

Teorema (caracterizaciones de la irreducibilidad)**.** Sea $A \geq 0$ una matriz de orden $n \geq 2$. Son equivalentes:

(I) no existe una matriz de permutación P tal que PAP^t sea triangular por bloques con un bloque de ceros fuera de la diagonal;
(II) el grafo dirigido G_A es fuertemente conexo;
(III) para todo par i, j existe un entero $m \geq 1$ tal que $(A^m)_{ij} > 0$;
(IV) para todo par $i \neq j$ existe una cadena $a_{i,k_1}a_{k_1,k_2} \cdots a_{k_{m-1},j} > 0$;
(V) la matriz $(I + A)^{n-1}$ tiene todas sus componentes estrictamente positivas.

LA IDEA: QUÉ SE VA A HACER, Y DÓNDE ESTÁ EL PESO

Cinco afirmaciones y una sola sustancia. La demostración no las prueba una a una: cierra un circuito, (I) $\Leftrightarrow$ (II) $\Leftrightarrow$ (IV) $\Leftrightarrow$ (III), y le cuelga la rama (II) $\Rightarrow$ (V) $\Rightarrow$ (III), de modo que cualquiera de las cinco lleva a cualquier otra siguiendo flechas.
Cada tramo traduce entre dos lenguajes distintos, y por eso cada uno es fácil por separado. El tramo (I) $\Leftrightarrow$ (II) traduce *álgebra* a *grafo*: un bloque de ceros fuera de la diagonal es exactamente «hay un grupo de islas del que no sale ningún barco hacia el otro grupo», y reordenar las variables es reetiquetar las islas. El tramo (II) $\Leftrightarrow$ (IV) es casi tautológico: un camino en el grafo es una cadena de entradas positivas, escrita con otra tipografía. Los dos últimos tramos ponen el reloj: (V) dice que, si el grafo va a conectarse, se conecta antes de $n - 1$ saltos.
El núcleo es (IV) $\Leftrightarrow$ (III), y cabe en una frase: *$(A^m)_{ij}$ es la suma de los pesos de todos los caminos de longitud m que van de j a i*. Multiplicar matrices es enumerar caminos. Todo lo demás es contabilidad alrededor de esa identidad. Y ahí mismo se ve por qué hace falta $A \geq 0$: en una suma de términos no negativos, «la suma es positiva» y «algún sumando es positivo» son lo mismo. Con signos, no.

Paso 1: del bloque de ceros a la incomunicación ((I) $\Leftrightarrow$ (II)). Supongamos que el grafo no es fuertemente conexo: existen dos nodos a y b tales que no hay camino de a a b. Sea Z el conjunto de todos los nodos alcanzables desde a (incluido a) y X el resto, donde vive b. No sale ninguna flecha de Z hacia X: si saliera, a la usaría para alcanzar un nodo de X, que por definición de Z no es alcanzable. Reordenamos las variables poniendo primero las de X y después las de Z. Una flecha de Z a X tendría origen en una columna de Z y destino en una fila de X, es decir, viviría en el bloque superior derecho; como no hay ninguna, ese bloque es de ceros y la matriz es reducible. Recíprocamente, si tras una permutación hay un bloque de ceros en las filas de un grupo X y las columnas de un grupo Z, entonces de Z no sale ninguna flecha hacia X, y no se puede viajar de un nodo de Z a uno de X: el grafo no es fuertemente conexo. (Con el convenio adoptado aquí la flecha va de la columna a la fila, así que el bloque nulo aparece arriba a la derecha o abajo a la izquierda según se ordene; en cualquier caso hay un bloque de ceros fuera de la diagonal.)

Paso 2: un camino es una cadena de factores positivos ((II) $\Leftrightarrow$ (IV)). Si el grafo es fuertemente conexo, para ir de j a i hay un camino $j \to k_{m-1} \to \cdots \to k_1 \to i$, y cada salto que existe tiene entrada positiva: $a_{k_{m-1},j} > 0, \ldots, a_{i,k_1} > 0$. El producto de positivos es positivo, que es (IV). Al revés, si el producto es positivo ningún factor es cero, todos los puentes existen y hay camino.

Paso 3: multiplicar matrices es contar caminos ((IV) $\Leftrightarrow$ (III)) — el núcleo de la prueba. Este es el paso que une la combinatoria con el álgebra, y descansa en la definición del producto de matrices:

$$(A^2)_{ij} = \sum_{k=1}^{n} a_{ik} a_{kj}, \qquad (A^3)_{ij} = \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} a_{i,k_1} a_{k_1,k_2} a_{k_2,j},$$

y en general $(A^m)_{ij}$ es la suma de los productos de *todos* los caminos de longitud m que van de j a i. Si existe un camino positivo de longitud m (IV), en esa suma hay un sumando positivo y ninguno negativo, porque $A \geq 0$; luego $(A^m)_{ij} > 0$, que es (III). Y si $(A^m)_{ij} > 0$ (III), como la suma es de términos no negativos, alguno de ellos es positivo: ese es el camino de (IV). **Aquí se usa $A \geq 0$ de forma esencial**: con entradas negativas, un camino de valor $+5$ podría cancelarse con otro de -5 y la equivalencia se vendría abajo.

Paso 4: el tope de $n-1$ saltos ((II) $\Rightarrow$ (V)). Si el grafo es fuertemente conexo, entre cualquier j y cualquier i hay un camino, y por tanto uno de longitud mínima. Un camino mínimo no repite vértices, porque si repitiera uno se podría recortar el bucle; como hay n vértices, tiene a lo sumo $n-1$ saltos. Ahora bien, como I y A conmutan, el binomio de Newton da

$$(I + A)^{n-1} = \sum_{m=0}^{n-1} \binom{n-1}{m} A^m,$$

con coeficientes positivos. La componente (i,j) de esta suma incluye el término $\binom{n-1}{m}(A^m)_{ij}$ para la longitud $m \leq n-1$ del camino mínimo, que es positivo por (III), y ningún término negativo. Luego $(I+A)^{n-1} > 0$.

Paso 5: cerrar el circuito ((V) $\Rightarrow$ (III)). Si $(I+A)^{n-1} > 0$, la componente (i,j) de la suma binomial anterior es positiva y todos sus sumandos son no negativos, así que

algún $(A^m)_{ij}$ es positivo, con $1 \leq m \leq n-1$ si $i \neq j$. Para $i = j$, el término I solo da el camino de longitud cero: elegimos otro vértice, viajamos hasta él y regresamos, usando los caminos ya obtenidos. Así hay también un retorno de longitud positiva. Esto completa (III). ■

Qué hace cada hipótesis. Solo hay dos, y las dos trabajan. La *no negatividad* $A \geq 0$ sostiene el paso 3 y, con él, todo el teorema: es lo que convierte «la suma de los caminos es positiva» en «hay un camino positivo». Bastaría una sola entrada negativa para romperlo: en $A = \left(\begin{smallmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right)$ los dos caminos de longitud dos que van de 1 a 1 pesan $+1$ y -1, se cancelan, y $(A^2)_{11} = 0$ aunque el grafo esté conectado. La *finitud del orden* n es la que da el paso 4: sin ella existiría el «para todo par hay un m» de (III), pero no habría ningún m común que sirviera de tope, y (V) — que es la caracterización computable, la que se programa — se quedaría sin enunciado.

Lo que hay que llevarse. Tres frases reconstruyen la prueba entera. (i) El número $(A^m)_{ij}$ es la suma de los pesos de todos los caminos de longitud m que van de j a i. (ii) Con $A \geq 0$ esa suma es positiva si y solo si hay al menos un camino, y un camino es exactamente una cadena de entradas no nulas; por eso grafo, cadena y potencia dicen lo mismo. (iii) Un camino sin rodeos entre n nodos da como mucho $n-1$ saltos, y por eso basta mirar la única matriz $(I+A)^{n-1}$ en lugar de todas las potencias.

LA IDEA: POR QUÉ $n-1$ Y POR QUÉ $I+A$

A es «dar exactamente un salto»; I es «quedarse quieto». Por eso $(I+A)^k$ suma los pesos de los caminos de longitud *menor o igual* que k, con los coeficientes binomiales que cuentan las maneras de intercalar esperas. Y $n-1$ es el máximo de saltos de un camino sin rodeos entre n nodos: quien necesita más ha pisado dos veces la misma isla. El criterio (V) es valioso porque pone un tope al «existe m» de (III): si en $n-1$ pasos el sistema no se ha conectado por completo, no lo hará jamás.

CONEXIÓN: MATEMÁTICA DISCRETA, GRAFOS DIRIGIDOS

La noción de grafo fuertemente conexo, y su distinción de la conexión débil (la del grafo no dirigido subyacente), es la de la asignatura de Matemática discreta de este mismo curso. Lo que añade el álgebra es el diccionario: matriz de adyacencia no negativa $\leftrightarrow$ grafo dirigido, $(A^m)_{ij}$ $\leftrightarrow$ número (o peso) de caminos de longitud m de j a i, irreducible $\leftrightarrow$ fuertemente conexo. Un grafo que no es fuertemente conexo se descompone en componentes fuertemente conexas, y la matriz, tras reordenar, en bloques diagonales irreducibles con ceros a un lado: es la forma normal de Frobenius. El libro previo del autor trabajó esta descomposición con un grafo de seis nodos sin ningún ciclo, cuyas componentes eran los seis nodos sueltos y cuya matriz, bien ordenada, era triangular con ceros en la diagonal.

Con el diccionario montado ya se puede demostrar, y no solo enunciar, la caracterización de las matrices de Leslie irreducibles.

Proposición. Sea L una matriz de Leslie de orden $n \geq 2$ con todos los $b_i > 0$. Entonces L es irreducible si y solo si $a_n > 0$.

La idea. El grafo de una matriz de Leslie es casi una escalera: las supervivencias $b_k > 0$

hacen que se pueda subir siempre un peldaño, de la clase k a la $k+1$, y las fertilidades $a_k > 0$ son los únicos toboganes que devuelven a la clase 1. Subir siempre se puede; lo que está en duda es *volver*. Y desde el último peldaño, la clase n, no hay escalón que subir: o hay tobogán ahí ($a_n > 0$) o la clase n es un sumidero del que no se sale. Toda la proposición es esa observación, mirada en los dos sentidos.

Paso 1: qué flechas tiene el grafo. En el grafo de L solo hay dos tipos de flecha, con el convenio de que $a_{ij} > 0$ es una flecha de j a i: las de *envejecer*, de k a $k+1$ para $k = 1, \dots, n-1$, que existen siempre porque $b_k > 0$; y las de *nacer*, de k a 1, que existen si y solo si $a_k > 0$.
Paso 2: si $a_n > 0$, un solo ciclo lo conecta todo (el núcleo). El ciclo $1 \to 2 \to \cdots \to n \to 1$ existe entero, y un ciclo que pasa por todos los vértices hace al grafo fuertemente conexo: para ir de j a i basta recorrer el ciclo desde j hasta i. Por la equivalencia (II) del teorema, L es irreducible.
Paso 3: si $a_n = 0$, la última clase es un sumidero. Del vértice n no sale ninguna flecha, ni de envejecer (no hay clase $n+1$) ni de nacer; ningún otro vértice es alcanzable desde n, y L es reducible. ■

Qué hace cada hipótesis. La hipótesis $b_i > 0$ para *todos* los i es la que garantiza que la escalera está completa; si una sola supervivencia se anulara, $b_k = 0$, la población quedaría cortada en dos tramos incomunicados y L sería reducible por mucho que a_n fuese positivo. Y la condición $a_n > 0$ no puede sustituirse por «hay alguna clase fértil»: con $a_n = 0$ y $a_{n-1} > 0$ el grafo tiene ciclos, pero ninguno pasa por la clase n, que sigue siendo un callejón sin salida. Es lo que ocurre en las abuelas del ejemplo anterior.

1.1.3.2 Matrices primitivas: irreducible no basta

La irreducibilidad garantiza que todo se comunica con todo. No garantiza que la población se *estabilice*, y el ejemplo que lo muestra es el más importante de esta sección.

Ejemplo. Una especie de insecto con un ciclo de vida rígido: huevo el año 1, larva el año 2, adulto el año 3, que pone huevos y muere. Con solo la última clase fértil,

$$L = \begin{pmatrix} 0 & 0 & a_3 \\ b_1 & 0 & 0 \\ 0 & b_2 & 0 \end{pmatrix}, \qquad v(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

La matriz es irreducible: el ciclo $1 \to 2 \to 3 \to 1$ está cerrado. Pero $v(1) = (0, b_1, 0)^t$, $v(2) = (0, 0, b_1 b_2)^t$, $v(3) = (a_3 b_1 b_2, 0, 0)^t$, y vuelta a empezar: la población recorre las clases con periodo 3 para siempre. El año 1 todo son huevos, el 2 todo larvas, el 3 todo adultos. Nunca hay a la vez huevos, larvas y adultos, y la proporción entre clases no converge a nada.

> *«Esto no lo veo del todo bien. O sea, he entendido que si solamente se reproduce una clase cada vez, entonces puede ser cíclico, pero estamos diciendo que las fertilidades son constantes. Entonces este análisis no es correcto; ese análisis sería correcto si las clases cambiaran su fertilidad en función del tiempo, ¿no?»*

Este era un malentendido del libro previo que conviene deshacer: **la oscilación no necesita tasas variables**. La matriz del ejemplo es constante, y aun así hace rotar el vector de estado.

El péndulo lo produce el espectro de una matriz fija: los autovalores de esa L son λ, $\lambda\omega$ y $\lambda\omega^2$ con $\omega = e^{2\pi i/3}$, tres números del *mismo módulo*. Ninguno domina a los otros: sus contribuciones no decaen *relativamente* entre sí. El tamaño total sí puede disminuir si $\lambda < 1$.

LA IDEA: IRREDUCIBLE FRENTE A PRIMITIVA

Irreducible es el mochilero paciente: «dime de dónde sales y a dónde vas y te encontraré una ruta; de la isla 1 a la 3 en dos saltos, de la 2 a la 1 en cinco». El número de saltos depende del par de islas. *Primitiva* es el teletransporte masivo: «existe un número m tal que, dando exactamente m saltos, todo el mundo puede llegar a cualquier parte a la vez». En la cigarra de tres años, en m saltos desde la clase 1 solo se llega a la clase $1 + m$ (mód 3): nunca a las tres a la vez. Toda primitiva es irreducible; la cigarra muestra que el recíproco es falso.

Proposición. Si $A \geq 0$ es primitiva, entonces es irreducible.

La idea, en una frase. Toda la proposición es el orden de dos cuantificadores: la primitividad pone el «existe m» *delante* del «para todo par», y la irreducibilidad lo pone *detrás*. Un billete válido para todo el mundo sirve, en particular, a cada uno.

Paso 1: escribir las dos definiciones con sus cuantificadores a la vista. Ser primitiva es

$$\exists\, m \geq 1 :\ \forall\, i, j, \quad (A^m)_{ij} > 0,$$

y ser irreducible, por la caracterización (III) del teorema anterior, es

$$\forall\, i, j,\ \exists\, m \geq 1 : \quad (A^m)_{ij} > 0.$$

Las dos frases tienen las mismas piezas y difieren solo en cuál de los dos cuantificadores manda sobre el otro.

Paso 2: el mismo m sirve para cada par (el núcleo). Sea m el exponente global que da la primitividad y sea (i, j) un par cualquiera. Ese mismo m, sin cambiarlo, cumple $(A^m)_{ij} > 0$, que es exactamente lo que (III) pide para ese par. Como el par era arbitrario, A es irreducible. ■

Y el recíproco es falso, que es lo que de verdad importa. Sacar el «existe» hacia fuera no es gratis: exige que todos los pares se conecten *a la vez*, en el mismo número de pasos. La cigarra de tres años del ejemplo anterior es irreducible y no primitiva, porque en m saltos desde la clase 1 solo se llega a la clase $1 + m \pmod 3$, y no hay ningún m que alcance las tres. La diferencia entre los dos órdenes es, literalmente, la diferencia entre una población que se estabiliza y una que oscila para siempre.

Lo que separa a una matriz irreducible de una primitiva es su **periodo**. Para un grafo fuertemente conexo, el máximo común divisor de las longitudes de todos sus ciclos se llama índice de imprimitividad h, y se cumple que una matriz irreducible es primitiva si y solo si $h = 1$. Este resultado se enuncia porque es la manera de entender la condición de primitividad de Leslie que viene después, y su demostración completa está en cualquier texto sobre matrices no negativas. En la matriz de Leslie se puede ver directamente: como las únicas flechas «hacia atrás» son las de nacer, todo ciclo simple, salvo el vértice por el que

se empiece a recorrerlo, es $1 \to 2 \to \cdots \to k \to 1$ para alguna clase fértil k, y su longitud es exactamente k. Luego

$$\{\text{longitudes de ciclos simples de } G_L\} = \{k : a_k > 0\}, \qquad h = \gcd\{k : a_k > 0\}.$$

Si hay dos clases consecutivas fértiles, $a_i > 0$ y $a_{i-1} > 0$, hay ciclos de longitudes i e $i-1$, que son coprimas, luego $h = 1$ y L es primitiva. En la cigarra solo hay un ciclo, de longitud 3, y $h = 3$.

LA IDEA: POR QUÉ MANDA EL MÁXIMO COMÚN DIVISOR

Si todos los ciclos midieran múltiplos de $h > 1$, una cohorte que sale de la clase 1 solo volvería a ella en tiempos múltiplos de h: la masa marcharía en h bloques sincronizados que nunca se mezclan. Dos longitudes coprimas rompen la sincronía. El argumento riguroso es el del problema de las monedas de Frobenius: si los tiempos de retorno a la clase 1 tienen máximo común divisor 1, todo entero suficientemente grande es un tiempo de retorno posible, y como en un número acotado de pasos se alcanza cualquier clase desde cualquier otra, se concluye que $L^m > 0$ para todo m grande. Eso es ser primitiva.

1.1.3.3 Los teoremas de Frobenius y de Perron

LA PREGUNTA

Una matriz 10×10 llena de números tiene un polinomio característico de grado 10, imposible a mano, y sus autovalores pueden ser complejos y girar sin descanso. ¿Puede haber una propiedad que se vea a simple vista, sin calcular nada, y que garantice que el autovalor mayor es real, positivo, y manda sobre todos los demás?

DE DÓNDE VIENE: PERRON, FROBENIUS Y EL SUELO DE LA REALIDAD

En el álgebra lineal pura el espacio es simétrico: los ejes van de $-\infty$ a $+\infty$ y a una matriz le da igual mandar un vector hacia arriba o hacia abajo. Pero la inmensa mayoría de las magnitudes de la naturaleza no tienen contraparte negativa: poblaciones, probabilidades, precios, flujos. Viven en el *ortante positivo*. Lo que conserva el ortante no negativo es tener entradas no negativas. Eso es compatible con autovalores negativos o complejos: la matriz de la cigarra los tiene, pero ninguna población inicialmente no negativa adquiere componentes negativas. Esos autovalores describen oscilaciones de la composición, no individuos negativos. Oskar Perron demostró en 1907, para matrices con todas las entradas positivas, que hay un autovalor real positivo que domina estrictamente a los demás, con un autovector de componentes positivas. Ferdinand Georg Frobenius extendió el resultado entre 1908 y 1912 a las matrices no negativas irreducibles, que es la forma en que se usa aquí. La teoría que nació de ahí, la de las matrices no negativas, es el motor oculto de casi todo lo que fluye en redes: cadenas de Markov, modelos económicos de Leontief, y el PageRank de la sección 1.2.

Teorema (Frobenius)**.** Sea $A \in \mathbb{R}^{n\times n}$, con $n \geq 2$, una matriz irreducible con componentes no negativas. Entonces:

1. existe un autovalor real simple $\lambda > 0$ y un autovector asociado $x \in \mathbb{R}^n$ tal que $\rho(A) = \lambda$ y x tiene todas sus componentes positivas;

2. además, si $v \in \mathbb{R}^n$ es un autovector de A y todas sus componentes son positivas, existe $\alpha > 0$ tal que $v = \alpha x$.

LA IDEA, EN VERSIÓN INTUITIVA: POR QUÉ EL AUTOVALOR GRANDE SE COME A LOS DEMÁS

Frobenius ya había dado un autovalor positivo con autovector positivo. Perron añade la palabra que hace útil todo el capítulo: ese autovalor es *estrictamente* dominante, y por eso al iterar la matriz todo lo demás se apaga y la población acaba con la forma del autovector.
La razón es fácil de ver, y es una desigualdad triangular con una pega. Si $\tilde{\lambda}$ es otro autovalor, con autovector w (que tendrá componentes de signos distintos, o complejas), al aplicar A se obtiene $|\tilde{\lambda}|\,|w_i| = \left|\sum_j a_{ij} w_j\right| \leq \sum_j a_{ij}|w_j|$: el módulo de una suma no supera la suma de módulos. Es decir, el vector de módulos $|w|$ crece al menos tanto como $|\tilde{\lambda}|$ lo exige, lo que fuerza $|\tilde{\lambda}| \leq \lambda$.
Lo interesante es cuándo puede darse la igualdad: solo si en esa desigualdad triangular no se pierde nada, es decir, si todos los w_j apuntan en la misma dirección del plano complejo. Y si todas las componentes de A son *estrictamente* positivas, todos los sumandos cuentan, así que la igualdad obliga a que w sea, salvo una fase, un vector positivo, o sea el propio autovector de Frobenius. En una palabra: **la positividad estricta es lo que impide que ningún otro autovalor empate**, porque impide que haya cancelaciones.
De ahí también se entiende por qué para matrices solo no negativas hace falta pedir *primitiva*. Si A tiene ceros, cada fila puede recibir solo componentes con la misma fase, aunque distintas filas reciban fases diferentes. Entonces la desigualdad triangular puede ser una igualdad sin que todo w tenga una fase común: aparecen autovalores del mismo módulo y se hacen posibles oscilaciones persistentes de la composición. Que alguna potencia de A sea estrictamente positiva dice algo más: hay un *mismo* número de pasos en el que todo puede influir sobre todo, y con eso se recupera la conclusión.

Teorema (Perron)**.** Sea $A \in \mathbb{R}^{n\times n}$ una matriz de orden n con todas sus componentes positivas. Entonces el autovalor λ que proporciona el teorema de Frobenius es **estrictamente dominante**: cualquier otro autovalor $\tilde{\lambda} \neq \lambda$ verifica $|\tilde{\lambda}| < \lambda$. Lo mismo ocurre si A tiene componentes no negativas y es primitiva.

Antes de dar una idea de las demostraciones, conviene leer los enunciados con cuidado, porque el libro previo del autor tropezó en un punto preciso.

> «*Entonces "el autovalor que proporciona el teorema anterior" (o sea, aquí entiendo que, aunque no lo han dicho explícitamente, estás añadiendo a las condiciones anteriores que A es irreducible, pues si no, no es aplicable el teorema anterior, obviamente, ¿no?) entonces ese λ es dominante; pero ya sabíamos que λ era simple y coincidía con el radio espectral, o sea que lo que me están diciendo en otras palabras es que $-\lambda$ no es autovalor, ¿no? Básicamente es eso, ¿no? ¿O lo estoy entendiendo mal?*»

Lo primero es correcto: una matriz estrictamente positiva o primitiva es irreducible, así que Frobenius se aplica y Perron lo refina. Lo segundo es la imprecisión que hay que corregir, y es grave: **es mucho más fuerte que descartar** $-\lambda$. El radio espectral traza en el plano complejo un círculo de radio λ. Frobenius dice que λ está en ese círculo y que es el único autovalor real

positivo de ese módulo; pero en el borde del círculo podría haber otros autovalores, reales negativos o complejos, como $-\lambda$, λi o las raíces $\lambda\omega$ de la cigarra. Cada uno de ellos aporta a L^k un término de módulo λ^k, que gira y no decae, y esas son las oscilaciones que no se apagan nunca. Perron **limpia el borde del círculo**: todos los demás autovalores están estrictamente dentro. Esa es la diferencia entre irreducible y primitiva, y es toda la diferencia entre «existe una proporción de equilibrio» y «la población converge a ella».

> *«Si el vector ya era positivo, entiendo que con que la matriz sea irreducible ya se tiene automáticamente el resultado (no hace falta que todas las componentes sean positivas ni que la matriz sea primitiva). Entonces, ¿el punto es cuando el vector inicial tiene algunas componentes positivas y otras cero?»*

Segundo cruce de cables que señaló el libro previo, y que se deshace con un contraejemplo mínimo. La primitividad es una propiedad de la *matriz*, no del vector inicial. Tomemos

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad x_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

A es irreducible, x_0 no tiene ningún cero, y la iteración oscila para siempre: $(2,1) \to (1,2) \to (2,1) \to \cdots$. La dirección nunca se alinea con el autovector de Perron $(1,1)$. La culpable es la matriz, de periodo 2 y autovalores ± 1, no los ceros de x_0. Lo que sí es cierto es lo contrario: si A es primitiva, **no hace falta** que x_0 sea estrictamente positivo. Basta $x_0 \geq 0$, $x_0 \neq 0$, porque a los m pasos con $A^m > 0$ cualquier cero se ha rellenado. Los tres niveles quedan así:

- **irreducible**: existe $\lambda > 0$ simple, igual al radio espectral, y un único autovector positivo salvo escala. Resultado estático, puramente algebraico;
- **primitiva**: además, la dinámica converge: para todo $x_0 \geq 0$ no nulo, la dirección de $A^k x_0$ tiende a la del autovector de Perron, olvidando el arranque;
- **irreducible pero no primitiva**: la dirección puede oscilar y ya no se garantiza convergencia para todo arranque. Por ejemplo, si x_0 es el autovector de Perron, $A^k x_0 = \lambda^k x_0$ y la dirección permanece fija. También puede converger un arranque que no sea ese autovector, si sus componentes en las otras direcciones espectrales de módulo λ son nulas. Lo que permite la oscilación persistente es que esas componentes estén presentes.

EXPLÍCATE. *En el contraejemplo 2×2, calcula los dos autovalores y comprueba que tienen el mismo módulo. Di qué conclusión del teorema de Frobenius se cumple y cuál del de Perron falla.*

La demostración de Frobenius que viene a continuación encuentra el autovalor sin tocar el polinomio característico: da a cada dirección una *nota*, el peor de los factores por los que crecen sus componentes, y se queda con la dirección de mejor nota. Conviene verlo funcionar primero en el caso más pequeño, donde todo se puede dibujar.

Ejemplo. Una población con dos clases, crías y adultas. Cada cría tiene en promedio 0.5 hijas por período, cada adulta 1.4, y el 60 % de las crías llega a adulta; las adultas no pasan de ahí. Su matriz de Leslie es

$$L = \begin{pmatrix} 0.5 & 1.4 \\ 0.6 & 0 \end{pmatrix},$$

que es irreducible porque $a_2 = 1.4 > 0$ y $b_1 = 0.6 > 0$. Una dirección no negativa de componentes que suman uno es $x = (t,\, 1-t)^t$, con $t \in (0,1)$ la fracción de crías. Al aplicar

L, cada clase queda multiplicada por su propio factor:

$$\frac{(Lx)_1}{x_1} = \frac{0.5t + 1.4(1-t)}{t} = \frac{1.4 - 0.9t}{t}, \qquad \frac{(Lx)_2}{x_2} = \frac{0.6t}{1-t}.$$

El primero es el factor de las crías y el segundo el de las adultas. Se leen así: cuantas más crías hay en la mezcla (t cerca de 1), menos adultas quedan para tenerlas, y el factor de las crías baja; las adultas del período siguiente, en cambio, salen de muchas crías, y su factor sube. Uno decrece con t y el otro crece.

La nota de la dirección es el menor de los dos, $r(x) = \min\{(1.4 - 0.9t)/t,\ 0.6t/(1-t)\}$. Con $t = 1/2$ los factores son 1.9 y 0.6, y la nota es 0.6: la mezcla crece al paso de las adultas. Con $t = 3/4$ son 0.967 y 1.8, y la nota es 0.967: ahora frenan las crías. Como un factor baja y el otro sube, el mínimo de los dos es más alto justo donde se cruzan (figura 1.1):

$$\frac{1.4 - 0.9t}{t} = \frac{0.6t}{1-t} \iff 0.3t^2 - 2.3t + 1.4 = 0 \iff t = \frac{2}{3} \ \text{(la otra raíz es } t = 7),$$

y ahí los dos factores valen 1.2. En la dirección ganadora, $x = (2/3,\ 1/3)^t$, las dos clases crecen *por el mismo factor*, y eso es literalmente ser autovector: $Lx = (0.5 \cdot \frac{2}{3} + 1.4 \cdot \frac{1}{3},\ 0.6 \cdot \frac{2}{3})^t = (0.8,\ 0.4)^t = 1.2\,x$. La nota ganadora es el autovalor dominante, y la dirección ganadora, su autovector positivo. El polinomio característico, $\lambda^2 - 0.5\lambda - 0.84 = (\lambda - 1.2)(\lambda + 0.7)$, lo confirma, pero no ha hecho falta.

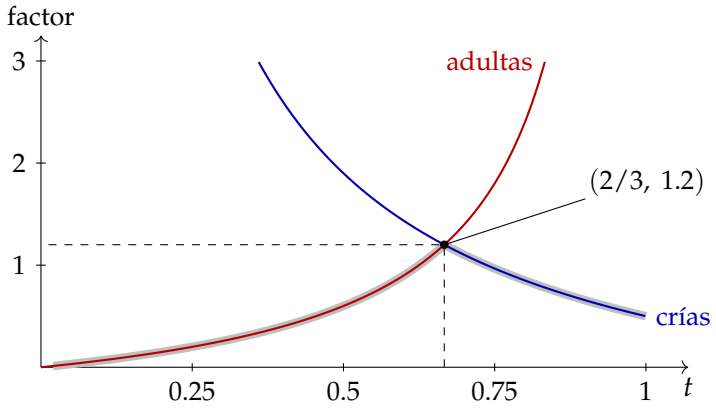


Figura 1.1: La nota de Collatz-Wielandt de la matriz L del ejemplo, en la dirección $x = (t, 1-t)^t$. En azul, el factor por el que se multiplican las crías; en rojo, el de las adultas; en gris, la nota $r(x)$, que es el menor de los dos. La nota es máxima donde los dos factores se igualan, en $t = 2/3$, y ahí vale 1.2: esa dirección es el autovector y ese valor el autovalor dominante.

Todo lo que la demostración general hace está ya en el dibujo. La nota tiene un máximo (paso 2). En el máximo no puede sobrar nada: si un factor superase al otro, esa clase tendría un excedente, y dar un poco más de peso a esa clase haría subir el factor de la otra, que es el peor de los dos (paso 3). Y la dirección ganadora no tiene ceros (paso 4). Lo que la dimensión dos esconde es por qué reforzar la clase con excedente siempre acaba ayudando a las demás: en general hay n factores y no dos curvas, y lo que garantiza que un excedente llegue a todas las clases es la irreducibilidad.

Idea de la demostración del teorema de Frobenius (existencia de λ y x). El argumento, debido a Helmut Wielandt (1950), es corto y hermoso, y explica de dónde sale el autovalor. Es un esbozo: se prueba que existen $\lambda > 0$ y $v > 0$ con $Av = \lambda v$, que es el corazón del teorema; la simplicidad y la unicidad del autovector positivo requieren algo más de trabajo y se dejan a la bibliografía.

La estrategia, simplificada al máximo. Nadie busca el autovalor resolviendo el polinomio característico. Se hace lo contrario: se define una *nota* $r(x)$ que puntúa a cada dirección x por lo mal que le va a su componente peor tratada, se toma la dirección v con la mejor nota, y se comprueba que en el ganador no puede sobrar nada. Si sobrara —si alguna componente creciera más de lo que dice su nota— ese excedente se podría repartir entre todas y subir la nota, contradiciendo que v era el ganador. Un vector en el que no sobra nada en ninguna componente es, exactamente, un autovector.

Paso 1: la función de Collatz-Wielandt, o la nota de cada dirección. Sea S el conjunto de los vectores no negativos y no nulos cuyas componentes suman 1: un símplex, que es compacto. Para $x \in S$ se define

$$r(x) = \min_{x_i > 0} \frac{(Ax)_i}{x_i},$$

el mayor factor c tal que $Ax \geq c\,x$. Es el cuello de botella de x: si el nodo 1 crece un 30 % y el 2 un 10 %, sus factores son 1.30 y 1.10, y $r(x) = 1.10$.

Paso 2: hay un ganador (compacidad). Escribamos $B = (I + A)^{n-1} > 0$. Si $Ax \geq r(x)x$, multiplicar por B, que es no negativa y conmuta con A, da $A(Bx) \geq r(x)Bx$. Por tanto, $r(Bx) \geq r(x)$: repartir por toda la red no empeora la nota. Como r no cambia al normalizar el vector, podemos buscar el supremo en

$$K = \{Bx / \|Bx\|_1 : x \in S\}.$$

Este conjunto es compacto (imagen continua de S) y todos sus vectores son estrictamente positivos. En K, r es continua: es el mínimo de un número finito de cocientes continuos con denominadores positivos. Weierstrass da un máximo, que es también el supremo sobre S. Sea $v \in K$ quien lo alcanza y $\lambda = r(v)$; entonces $Av \geq \lambda v$. Además $\lambda > 0$, porque evaluar r en e/n da un valor positivo: cada fila de una matriz irreducible de orden $n \geq 2$ tiene alguna entrada positiva.

Paso 3: en el ganador no sobra nada — el núcleo de la prueba. Supongamos $y = Av - \lambda v \geq 0$ con $y \neq 0$. Como A es irreducible, $(I + A)^{n-1} > 0$ por la caracterización (V), y por tanto $(I + A)^{n-1}y > 0$ en todas sus componentes. Como A conmuta con $(I + A)^{n-1}$, llamando $w = (I + A)^{n-1}v$ eso es $Aw > \lambda w$ componente a componente. Al ser desigualdades estrictas en un número finito de componentes, existe $\varepsilon > 0$ con $Aw \geq (\lambda + \varepsilon)w$, y normalizando w para que esté en S se obtiene $r(w) \geq \lambda + \varepsilon > \lambda$. Contradicción con que λ era el máximo. Luego $Av = \lambda v$.

Paso 4: el ganador es estrictamente positivo. Sabemos $v \geq 0$, $v \neq 0$ y $Av = \lambda v$, luego $(I + A)^{n-1}v = (1 + \lambda)^{n-1}v$. El lado izquierdo es estrictamente positivo, porque $(I + A)^{n-1} > 0$ y v tiene alguna componente positiva. Como $(1 + \lambda)^{n-1} > 0$, el propio v es estrictamente positivo. Que λ sea el radio espectral sale de la misma función r: para cualquier autovalor μ con autovector z, tomando módulos, $|\mu||z| \leq A|z|$, luego $|\mu| \leq r(|z|/\|z\|_1) \leq \lambda$. ■

Qué hace cada hipótesis. La *no negatividad* de A actúa dos veces y en el mismo sentido: hace que $Ax \geq 0$ cuando $x \geq 0$, con lo que la nota $r(x)$ tiene sentido, y hace que al tomar módulos en $\mu z = Az$ salga la desigualdad $|\mu||z| \leq A|z|$ del final —con entradas negativas, la desigualdad triangular daría $|Az| \leq |A||z|$, con $|A|$ tomado entrada a entrada, y no la comparación con $A|z|$ que necesitamos—. La *irreducibilidad* entra en los pasos 2, 3 y 4, y siempre por la misma puerta: la caracterización (V), $(I + A)^{n-1} > 0$. Es el mecanismo de reparto: convierte un excedente alojado en una sola componente en un excedente presente en *todas*, que es lo que permite mejorar la nota. Sin irreducibilidad ese reparto no llega a todas partes; la matriz $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ tiene dos autovectores no negativos independientes, $(1, 0)$ y $(0, 1)$, y la unicidad se pierde. Obsérvese, por último, que la *compacidad* del símplex S —cerrado y acotado en $\mathbb{R}^n$— es lo que, junto con la continuidad de r en K, garantiza que el supremo del paso 2 se alcanza. En dimensión infinita esa comodidad desaparece, y la versión del teorema para operadores positivos (Krein-Rutman) necesita pedir compacidad al operador a mano.

Lo que hay que llevarse. (i) Cada dirección del símplex recibe una nota, $r(x)$, el peor de los factores de crecimiento de sus componentes, y la compacidad garantiza una dirección ganadora. (ii) En la ganadora no sobra nada, porque la irreducibilidad reparte cualquier excedente entre todas las componentes y subiría la nota; una dirección sin excedentes cumple $Av = \lambda v$. (iii) La misma matriz de reparto, $(I + A)^{n-1} > 0$, rellena los ceros y hace $v > 0$. En la figura 1.1, las tres frases son el cruce de las dos curvas.

LA IDEA: LA FLOTA Y LOS EXCEDENTES

Los pasos 3 y 4 en versión física. Una flota (el vector x) navega en formación a la velocidad de su barco más lento, $r(x)$; el diseño de flota que navega más rápido en formación es v, a velocidad λ. Si en v algún nodo creciera *más* que λ, tendría un excedente; como la red no tiene callejones sin salida (irreducibilidad), en unos pasos el excedente se reparte por todos los nodos, y todos crecerían por encima de λ: una flota mejor que la mejor, absurdo. Y si un nodo de v estuviera vacío, los flujos de los demás lo llenarían en unos pasos, contra el hecho de que un autovector solo se reescala: absurdo otra vez. No hay excedentes y no hay nodos vacíos.

EL CAMINO QUE NO SE TOMÓ: EXIGIR POSITIVIDAD ESTRICTA, COMO HIZO PERRON

La alternativa. Enunciar el teorema solo para matrices con todas las entradas positivas, que es el teorema original de Perron de 1907.
Su ventaja. La demostración es más corta, no hace falta la noción de irreducibilidad, y se obtiene de golpe la dominancia estricta.
Qué la hunde para esta asignatura. Ni la matriz de Leslie ni la de transición de PageRank tienen todas las entradas positivas: la de Leslie está llena de ceros por construcción, y la de una red real también. Hacía falta la extensión de Frobenius a matrices no negativas irreducibles, y separar las dos conclusiones: existencia (irreducible) y dominancia (primitiva). Curiosamente, en PageRank se acaba volviendo al caso de Perron por un truco de modelado: la matriz de Google tiene todas las entradas positivas gracias al factor de amortiguación.

1.1.3.4 El teorema de Leslie: Perron-Frobenius aplicado

Como consecuencia de los dos teoremas anteriores se puede probar el resultado central del tema.

Teorema (de Leslie)**.** Supongamos que en la matriz de Leslie los $a_i \geq 0$ para $i = 1, \dots, n-1$, $a_n > 0$, y los $b_i \in (0,1)$ para $i = 1, \dots, n-1$. Entonces:

1. L tiene un único autovalor real positivo $\hat{\lambda}$, de multiplicidad 1, que coincide con su radio espectral: $\rho(L) = \hat{\lambda}$;
2. el vector $\hat{v}$ con
$$\hat{v}_1 = 1 \qquad \text{y} \qquad \hat{v}_i = \hat{\lambda}^{-(i-1)}\, b_1 \cdots b_{i-1} \quad \text{para } i = 2, \dots, n$$
es un autovector de L asociado a $\hat{\lambda}$ que tiene todas sus componentes positivas;
3. cualquier otro autovector w de L con componentes positivas está de nuevo asociado a $\hat{\lambda}$ y es de la forma $w = \alpha\hat{v}$ para algún $\alpha > 0$;
4. finalmente, si existe $i \geq 2$ tal que $a_i a_{i-1} > 0$, entonces L es primitiva y, por tanto, $\hat{\lambda}$ es estrictamente dominante.

El caso más pequeño, entero a mano. Antes de la demostración general conviene ver los cuatro apartados en una matriz en la que todo se calcula con la fórmula de la ecuación de segundo grado. Con dos clases,

$$L = \begin{pmatrix} a_1 & a_2 \\ b_1 & 0 \end{pmatrix}, \qquad \det(\lambda I - L) = \lambda^2 - a_1\lambda - a_2 b_1, \qquad \lambda_\pm = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2 b_1}}{2}.$$

Las dos raíces son reales, porque el discriminante es positivo. Se leen sin resolver nada:

- su **producto** es $-a_2 b_1 < 0$, así que una es positiva y la otra negativa: hay un único autovalor real positivo, $\hat{\lambda} = \lambda_+$ (apartado 1);
- su **suma** es $a_1 \geq 0$, así que la positiva es la mayor en módulo, $|\lambda_-| \leq \hat{\lambda}$, con igualdad exactamente cuando $a_1 = 0$. La dominancia estricta pide $a_1 a_2 > 0$, que con dos clases es ser fértiles las dos (apartado 4);
- la **segunda fila** de $Lv = \lambda v$ dice $b_1 v_1 = \lambda v_2$, y con $v_1 = 1$ da $v = (1,\, b_1/\lambda)^t$: la cascada del apartado 2 con un solo escalón.

Con los números de las crías y adultas de la figura 1.1, $a = (0.5,\, 1.4)$ y $b_1 = 0.6$, la ecuación es $\lambda^2 - 0.5\lambda - 0.84 = 0$, con raíces $\hat{\lambda} = 1.2$ y $\lambda_2 = -0.7$. El autovector de $\hat{\lambda}$ es $\hat{v} = (1,\, 0.5)^t$, que normalizado a suma uno es $(2/3,\, 1/3)^t$: a la larga, dos tercios de crías y un tercio de adultas, y un 20 % más de población en cada período. El autovector de -0.7 es $(1,\, -6/7)^t$, con signos mezclados: no describe ninguna población, y por eso el apartado 3 dice que todo autovector positivo es múltiplo de $\hat{v}$.

Y aquí se puede escribir la solución entera. Descomponiendo el dato inicial en los dos autovectores,

$$v(k) = c_1\,(1.2)^k \begin{pmatrix} 1 \\ 1/2 \end{pmatrix} + c_2\,(-0.7)^k \begin{pmatrix} 1 \\ -6/7 \end{pmatrix}.$$

Si se empieza con diez adultas y ninguna cría, $v(0) = (0, 10)^t$, los coeficientes son $c_1 = 140/19$ y $c_2 = -140/19$, y al dividir por 1.2^k,

$$1.2^{-k}\, v(k) = \frac{140}{19}\begin{pmatrix} 1 \\ 1/2 \end{pmatrix} - \frac{140}{19}\Big(-\frac{7}{12}\Big)^k \begin{pmatrix} 1 \\ -6/7 \end{pmatrix} \longrightarrow \frac{140}{19}\begin{pmatrix} 1 \\ 1/2 \end{pmatrix} \approx \begin{pmatrix} 7.37 \\ 3.68 \end{pmatrix}.$$

El primer término es la forma estable; el segundo cambia de signo en cada período y se encoge por 7/12 cada vez. En el censo eso se ve como una oscilación que se apaga:

k	$v(k)$	total	fracción de adultas
0	$(0,\ 10)$	10	1
1	$(14,\ 0)$	14	0
2	$(7,\ 8.4)$	15.4	0.545
3	$(15.26,\ 4.2)$	19.46	0.216
4	$(13.51,\ 9.16)$	22.67	0.404
5	$(19.57,\ 8.11)$	27.68	0.293

La fracción de adultas salta a un lado y a otro de $1/3$ con saltos cada vez menores, y el cociente entre totales sucesivos, 1.4, 1.1, 1.264, 1.165, 1.221, hace lo mismo alrededor de 1.2. Es exactamente lo que el teorema de convergencia de la subsección siguiente afirma para cualquier n.

Y si las crías no son fértiles. Con $a_1 = 0$ la ecuación es $\lambda^2 = 0.84$ y las raíces son $\pm\sqrt{0.84} \approx \pm 0.9165$: el mismo módulo. Desde $v(0) = (0, 10)^t$ el censo es $(14, 0)$, $(0, 8.4)$, $(11.76, 0)$, $(0, 7.056)\ldots$: la población es todo crías un período y todo adultas el siguiente, mengua (se multiplica por 0.84 cada dos períodos) y su composición no se estabiliza nunca. Es la cigarra de la subsección 1.1.3.2 con un ciclo de dos años: irreducible, porque $a_2 > 0$, y no primitiva, porque solo hay una clase fértil.

Qué se ha ganado. Con dos clases, el signo del producto de las raíces da el autovalor positivo único, el signo de la suma da la dominancia y la segunda fila da el autovector. Con n clases no hay fórmula para las raíces, y lo que hace la demostración general es llegar a las mismas tres conclusiones sin ella.

La demostración se da entera, porque cada pieza ya está preparada.

LA IDEA: DE DÓNDE SALE CADA APARTADO

Cuatro afirmaciones, y solo dos fuentes. Frobenius, cuya existencia se ha justificado y cuyas otras conclusiones se han enunciado, entrega lo *cualitativo*: hay un autovalor positivo simple, es el radio espectral y su autovector es positivo y único salvo escala. Lo que Frobenius *no* sabe es que L tenga forma de matriz de Leslie, y por eso no puede decir cuánto vale ese autovalor ni que sea el único real positivo. Eso lo pone la segunda fuente: la propia forma de L.

Y esa forma dice una sola cosa, repetida $n - 1$ veces: *la fila i de L tiene un único término*, b_{i-1}. Traducido, «la clase i del año que viene son los supervivientes de la clase $i - 1$». Si se exige que aplicar L solo reescale por λ, esa fila obliga a $v_i = (b_{i-1}/\lambda)v_{i-1}$, y la cascada queda determinada por v_1: hay un único candidato a autovector, no una familia. **Ese es el núcleo del teorema**: la estructura de escalera de L deja un solo grado de libertad.

Falta cerrar el circuito. La primera fila —la de los nacimientos, la única que no se ha usado— no aporta forma pero sí una ecuación escalar, $\varphi(\lambda) = 1$. Y como φ es una suma de potencias negativas de λ con coeficientes no negativos, cae monótonamente de $+\infty$ a 0: corta la altura 1 una vez y nada más que una. De ahí la unicidad. La imagen que conviene retener es la de una balanza: subir λ descuenta más el futuro, y hay

un único tipo de descuento con el que la descendencia esperada de una hembra vale exactamente uno, es decir, con el que la población se reemplaza a sí misma.

Paso 1: Frobenius pone lo cualitativo. Como $a_n > 0$ y todos los $b_i > 0$, la proposición de la subsección 1.1.3.1 dice que L es irreducible. Y $L \geq 0$. Luego se aplica el teorema de Frobenius: existe un autovalor real simple $\lambda > 0$ igual al radio espectral, con autovector de componentes positivas, y todo autovector positivo es múltiplo suyo. Eso son los apartados 1 y 3 salvo una cosa: el teorema de Frobenius no dice que λ sea el *único* autovalor real positivo. Eso lo va a dar la forma de L.

Paso 2: la escalera fuerza la forma del autovector, apartado 2 (el núcleo). Sea λ cualquier autovalor real positivo de L y v un autovector asociado. Escribimos $Lv = \lambda v$ y leemos la fila i-ésima con $i \geq 2$: en L esa fila tiene un solo término no nulo, el b_{i-1} de la columna $i-1$. Por tanto

$$(Lv)_i = b_{i-1} v_{i-1} = \lambda v_i \qquad \Longrightarrow \qquad v_i = \frac{b_{i-1}}{\lambda}\, v_{i-1}.$$

Es la traducción exacta de «aplicar L es avanzar un período» y «la clase i del año que viene son los supervivientes de la clase $i-1$»: pedir que v sea autovector obliga a que ese avance solo reescale por λ. Si fuera $v_1 = 0$, la cascada daría $v = 0$, que no es autovector; luego $v_1 \neq 0$ y se puede normalizar $v_1 = 1$. Iterando, cada escalón añade un factor b/λ:

$$v_i = \lambda^{-(i-1)}\, b_1 b_2 \cdots b_{i-1},$$

que es la fórmula del enunciado, y es positiva. La primera fila no interviene en la forma; lo que hace es **fijar el valor de** λ. Sustituyendo la cascada en $(Lv)_1 = \sum_{i=1}^{n} a_i v_i = \lambda v_1 = \lambda$ y dividiendo por λ,

$$\varphi(\lambda) := \sum_{i=1}^{n} a_i\, \lambda^{-i}\, b_1 \cdots b_{i-1} = 1,$$

con el convenio de que el producto vacío vale 1 para $i = 1$.

Paso 3: la fila de los nacimientos fija λ y lo hace único (apartado 1). La función φ es **estrictamente decreciente** en $\lambda > 0$: cada sumando con $a_i > 0$ es una constante positiva por λ^{-i}, que decrece estrictamente, y hay al menos uno porque $a_n > 0$. Cuando $\lambda \to 0^+$, $\varphi \to +\infty$, y cuando $\lambda \to +\infty$, $\varphi \to 0$. Por continuidad y monotonía, φ toma el valor 1 en **un único** punto $\hat{\lambda} > 0$. Como todo autovalor real positivo tiene que satisfacer $\varphi(\lambda) = 1$, solo hay uno, y por Frobenius es simple e igual al radio espectral. Esto cierra 1, 2 y 3.

Paso 4: dos ciclos coprimos dan la primitividad (apartado 4). Si $a_i > 0$ y $a_{i-1} > 0$ para algún $i \geq 2$, el grafo de L tiene ciclos de longitudes i e $i-1$, coprimas; por el criterio del índice de imprimitividad de la subsección 1.1.3.2, L es primitiva. Se puede también comprobar a mano que alguna potencia de L es estrictamente positiva, que es la definición directa, pero el argumento del máximo común divisor es el que explica por qué. Entonces el teorema de Perron, en su versión para matrices primitivas, da que $\hat{\lambda}$ es estrictamente dominante. ■

Qué hace cada hipótesis. Las tres trabajan en sitios distintos y se ve bien qué se rompe al quitarlas. La hipótesis $b_i \in (0, 1)$ sostiene el paso 2: si algún b_i fuera nulo la cascada se

cortaría y ya no se garantizaría un autovector estrictamente positivo; que además sean menores que 1 no hace falta para el álgebra, es la lectura demográfica (una probabilidad de supervivencia). La hipótesis $a_n > 0$ sostiene el paso 1 —da la irreducibilidad— y también el paso 3, porque garantiza que φ tiene al menos un sumando y no es la función cero, que valdría $0 \neq 1$ siempre. Y la hipótesis $a_i a_{i-1} > 0$ solo aparece en el paso 4: sin ella los apartados 1, 2 y 3 siguen siendo ciertos —la proporción de equilibrio *existe*— pero la dominancia estricta ya no queda garantizada por esta condición suficiente. Puede seguir habiendo primitividad: por ejemplo, fecundidades positivas solo en las clases 2 y 5 dan ciclos coprimos sin clases fértiles consecutivas. La cigarra de tres años cumple todo salvo esta última, y es exactamente la que oscila para siempre.

Lo que hay que llevarse. (i) Con $a_n > 0$ la matriz es irreducible, y Frobenius da lo cualitativo: un autovalor positivo simple que es el radio espectral, con autovector positivo único salvo escala. (ii) Las filas $2, \dots, n$ de L tienen un solo término, y al pedir $Lv = \lambda v$ fuerzan la cascada $v_i = (b_{i-1}/\lambda)v_{i-1}$. (iii) La primera fila da $\varphi(\lambda) = 1$, con φ estrictamente decreciente: una sola raíz positiva. (iv) Dos clases fértiles consecutivas dan ciclos de longitudes coprimas, luego primitividad, y Perron da la dominancia estricta.

DE DÓNDE VIENE: LA ECUACIÓN DE EULER-LOTKA

La ecuación $\varphi(\lambda) = 1$, es decir $\sum_i a_i b_1 \cdots b_{i-1} \lambda^{-i} = 1$, es la versión discreta de la ecuación de Euler-Lotka: Euler la escribió en 1760 para poblaciones humanas y Alfred Lotka la redescubrió en 1911 en su forma continua, $\int_0^\infty e^{-rx} \ell(x) m(x)\, dx = 1$, donde $\ell(x)$ es la supervivencia hasta la edad x y $m(x)$ la fecundidad a esa edad. El producto $b_1 \cdots b_{i-1}$ es la ℓ, el a_i es la m, y λ^{-i} es el e^{-rx} con $\lambda = e^{r\tau}$. La matriz la introdujeron, independientemente, Harro Bernardelli en 1941, E. G. Lewis en 1942 y Patrick Leslie en 1945; el nombre quedó para el tercero, que fue quien la analizó a fondo.

Cómo se lee la ecuación $\varphi(\lambda) = 1$, y qué se gana con leerla. Cada sumando de

$$\varphi(\lambda) = \sum_{i=1}^{n} a_i \, b_1 \cdots b_{i-1} \, \lambda^{-i}$$

tiene tres factores con significado propio. El producto $b_1 \cdots b_{i-1}$ es la probabilidad de que una hembra recién nacida llegue viva a la clase i; a_i es el número medio de hijas que tiene mientras está en ella; y λ^{-i} es un descuento: si la población se multiplica por λ en cada período, una hija que nace i períodos más tarde entra en una población λ^i veces mayor, y pesa en ella λ^{-i} veces lo que pesaría hoy. La ecuación $\varphi(\hat{\lambda}) = 1$ dice entonces que, al ritmo de crecimiento $\hat{\lambda}$, las hijas de una hembra, descontadas a ese mismo ritmo, la reemplazan exactamente.

Sin descuento, es decir en $\lambda = 1$, queda

$$\varphi(1) = \sum_{i=1}^{n} a_i \, b_1 \cdots b_{i-1},$$

el número medio de hijas que tiene una hembra a lo largo de toda su vida. Y como φ es estrictamente decreciente y $\varphi(\hat{\lambda}) = 1$, comparar $\varphi(1)$ con $1 = \varphi(\hat{\lambda})$ es comparar 1 con $\hat{\lambda}$ en sentido contrario:

$$\hat{\lambda} > 1 \iff \varphi(1) > 1, \qquad \hat{\lambda} = 1 \iff \varphi(1) = 1, \qquad \hat{\lambda} < 1 \iff \varphi(1) < 1.$$

Es la respuesta a la pregunta de la demógrafa que no exige resolver ningún polinomio: la población crece si y solo si cada hembra deja, en promedio, más de una hija. En el ejemplo, $\varphi(1) = 0 + 0.25 + 0.375 + 0.025 = 0.65$: cada hembra deja 0.65 hijas, la población no se reemplaza y $\hat{\lambda} < 1$. Si se duplica la supervivencia de las crías, $b_1 = 0.5$, los tres sumandos que llevan b_1 se duplican, $\varphi(1) = 1.3 > 1$, y la población pasa a crecer; el ejercicio del final de esta sección lo confirma acotando la raíz. Los dos números, $\varphi(1)$ y $\hat{\lambda}$, no miden lo mismo: $\varphi(1)$ cuenta por generación y $\hat{\lambda}$ por período, y aquí una generación dura más de un período, porque las madres tienen sus hijas a partir de la segunda clase. Lo que comparten es el lado del 1 en el que caen, y eso es lo que decide el destino.

CONEXIÓN: EL MISMO UMBRAL EN LAS EPIDEMIAS

En demografía, $\varphi(1)$ se llama *tasa neta de reproducción* y se escribe R_0. El nombre no es casual. En la sección 2.1.2 aparece el número básico de reproducción de una infección, $R_0 = N\beta/\gamma$: los contagios que causa en promedio un infeccioso en una población susceptible. Es la misma cuenta con otra clase de descendencia —los «hijos» de un infeccioso son sus contagios—, y el umbral es el mismo: la infección se vuelve endémica si $R_0 > 1$ y desaparece si $R_0 \leq 1$, como la población crece si cada hembra se reemplaza por más de una hija.

> *«Normalizamos la primera componente. Aplicamos la matriz de Leslie: para que v sea autovector, lo único que tiene que ocurrir es que L lo multiplique por ese λ. Pero multiplicar por Leslie es avanzar un tiempo, así que por fuerza v_2 tiene que ser b_1; y si volviéramos a multiplicar, v_3 tendría que ser $b_1 b_2$, y así hasta que v_i sería $b_1 \cdots b_{i-1}$. Pero claro: si hemos hecho i multiplicaciones y v es autovector, $\lambda^i v_i = b_1 \cdots b_{i-1}$; entonces se tiene el resultado.»*

Este razonamiento del libro previo es correcto en la idea y tiene un desajuste en el reparto de potencias: la clase i se alcanza con $i - 1$ multiplicaciones, no con i, y por eso el exponente es $\lambda^{-(i-1)}$. Es un error de una unidad, del tipo que se propaga sin avisar, y por eso se deja señalado.

Por qué la hipótesis $a_i a_{i-1} > 0$ es razonable. Si la fecundidad es positiva a lo largo de un intervalo de edades, una partición suficientemente fina puede situar ese intervalo en dos clases consecutivas. Es una motivación para la hipótesis, no una garantía biológica universal: la reproducción estacional o sincronizada puede exigir otro tratamiento. Matemáticamente, dos clases fértiles consecutivas son una condición suficiente; lo decisivo es el máximo común divisor de las edades fértiles.

1.1.3.5 El teorema de convergencia

Con L primitiva ya no solo *existe* la proporción estable: se *alcanza*, salga de donde salga la población. Dicho con precisión: en las condiciones del apartado 4, el método de la potencia iterada converge partiendo de cualquier $v(0) \neq 0$ con componentes $v_i(0) \geq 0$.

LA IDEA, EN VERSIÓN INTUITIVA: LA PIRÁMIDE DE EDADES SE OLVIDA DE CÓMO EMPEZÓ

Este es el teorema que hace de Leslie un modelo y no solo una matriz. Dice dos cosas, y conviene separarlas porque se confunden con facilidad.

La primera es sobre la **forma**: sea cual sea el reparto inicial por edades, con tal de que no sea idénticamente nulo y no sea absurdo, al cabo de bastantes años la pirámide de edades se parece a $\hat{v}$, el autovector de Frobenius. Cuántos individuos haya en total depende de cómo se empezó (ahí está el α), pero *la proporción entre las clases de edad no*: eso es lo que se llama distribución estable, y explica por qué poblaciones que arrancan de censos muy distintos acaban con la misma silueta.
La segunda es sobre el **tamaño**: el cociente entre el total de un año y el del anterior tiende a $\hat{\lambda}$. Es decir, pasado el transitorio la población crece (o decrece) de forma geométrica de razón $\hat{\lambda}$, y por eso $\hat{\lambda}$ se lee directamente como el factor de crecimiento por período τ de la especie: si $\hat{\lambda} > 1$ crece, si $\hat{\lambda} < 1$ se extingue, y si $\hat{\lambda} = 1$ se estabiliza.
Por qué es cierto ya se sabe del apartado anterior: se descompone el vector inicial en la base de autovectores, se aplica la matriz k veces, y cada componente queda multiplicada por su autovalor elevado a k. Como el de Frobenius es estrictamente el mayor, al dividir por $\hat{\lambda}^k$ todas las demás componentes tienden a cero. **El teorema de Perron es exactamente la hipótesis que permite hacer esto**, y la condición $a_i a_{i-1} > 0$ (dos clases fértiles consecutivas) es la que garantiza que la matriz de Leslie sea primitiva y no meramente irreducible: sin ella la población podría quedarse oscilando por cohortes en lugar de estabilizarse.

Teorema (comportamiento a largo plazo)**.** En las condiciones del teorema de Leslie, supongamos que para algún $i \geq 2$ es $a_i a_{i-1} > 0$. Sea $v(0) \neq 0$ con los $v_i(0) \geq 0$. Entonces:
1. existe $\alpha > 0$, que depende de $v(0)$, tal que

$$\lim_{k \to +\infty} \hat{\lambda}^{-k} v(k) = \alpha \, \hat{v};$$

2. además,

$$\lim_{k \to +\infty} \frac{\|v(k+1)\|_1}{\|v(k)\|_1} = \lim_{k \to +\infty} \frac{\sum_{j=1}^{n} v_j(k+1)}{\sum_{j=1}^{n} v_j(k)} = \hat{\lambda};$$

3. finalmente,

$$\lim_{k \to +\infty} \frac{v_i(k)}{\|v(k)\|_1} = \frac{\hat{v}_i}{\|\hat{v}\|_1} \qquad \text{para todo } i = 1, \ldots, n.$$

«Mi intuición: si L es primitiva, cualquier vector del primer octante —incluidos los de los ejes— se alinea, cuando $k \to \infty$, con el autovector; es decir, la población tiende a la proporción entre edades dada por ese autovector. Si $\lambda > 1$ la población se dispara, pero la proporción es constante. Así, el primer apartado es básicamente una normalización para ver que en el infinito el vector se alinea con el autovector; y el segundo dice que al dar un paso el módulo (en norma 1) crece por λ. No le veo más chiste que ese, pero me parece expresado de forma extraña, así que entiendo que me dice más cosas de las que estoy captando.»

La intuición es correcta y tiene nombre: es exactamente el **método de la potencia**. Iterar $v(k) = L^k v(0)$ es la iteración que en Cálculo Numérico halla el autovector dominante, y converge precisamente cuando hay un autovalor estrictamente dominante, que es lo que garantiza la primitividad. La población, sin saberlo, ejecuta el método de la potencia sobre L.

La trampa que hay que ver antes de empezar. La demostración parece un trámite: se descompone $v(0)$ en la base de autovectores, se aplica L^k, y como $\hat{\lambda}$ es el mayor, todo lo

demás se hace pequeño al dividir por $\hat{\lambda}^k$. Pero eso solo prueba que el límite es $c_1\hat{v}$ *para algún* $c_1 \geq 0$, y si c_1 fuese cero el límite sería el vector nulo y el teorema no diría nada. Uno tiende a pensar que eso no puede pasar porque $v(0) \geq 0$ y $\hat{v} > 0$ «apuntan hacia el mismo lado», pero esa intuición es falsa: los coeficientes de una base de autovectores no son proyecciones ortogonales, y nada impide a priori que la componente de $v(0)$ a lo largo de $\hat{v}$ se anule. **Probar** $c_1 > 0$ **es el verdadero contenido del apartado 1**, y lo que lo hace posible es un objeto nuevo: el autovector de Perron *por la izquierda*, $u > 0$, que actúa como el aparato de medida capaz de leer esa componente sin mirar a las demás.

Paso 1a: descomponer y dejar que el tiempo limpie. Escribimos $v(0)$ en la base de Jordan de L. Como $\hat{\lambda}$ es simple, su bloque es 1×1 y hay un autovector $\hat{v}$; el resto de la base son autovectores o autovectores generalizados w_j de los otros autovalores μ_j. Así

$$v(0) = c_1\hat{v} + \sum_{j\geq 2} c_j w_j.$$

Aplicar L^k y dividir por $\hat{\lambda}^k$ da

$$\hat{\lambda}^{-k} v(k) = c_1\hat{v} + \sum_{j\geq 2} c_j\, \hat{\lambda}^{-k} L^k w_j,$$

y por el cálculo de las potencias de un bloque de Jordan (recordado en la caja de conexión con Álgebra Lineal), cada $\hat{\lambda}^{-k}L^k w_j$ es combinación de términos $\binom{k}{s}(\mu_j/\hat{\lambda})^{k-s}$ con s acotado. Como L es primitiva, $|\mu_j| < \hat{\lambda}$ para todo $j \geq 2$, luego $|\mu_j/\hat{\lambda}| < 1$ y cada uno de esos términos tiende a cero. Solo sobrevive el término de Perron: $\hat{\lambda}^{-k}v(k) \to c_1\hat{v}$.

Paso 1b: el coeficiente de Perron es estrictamente positivo (el núcleo de la prueba). El coeficiente c_1 se calcula con el autovector de Perron *por la izquierda*: L^t es también no negativa e irreducible (mismo grafo con las flechas invertidas) y primitiva, y tiene por tanto un autovector $u > 0$ asociado a $\hat{\lambda}$, es decir, $u^t L = \hat{\lambda} u^t$. Multiplicando $v(0) = c_1\hat{v} + \sum c_j w_j$ por u^t a la izquierda, y usando que $u^t w_j = 0$ para los w_j de los otros autovalores (porque $u^t(L-\mu_j I)^s w = (\hat{\lambda}-\mu_j)^s u^t w$ y para w generalizado el lado izquierdo se anula para s grande), queda $u^t v(0) = c_1\, u^t\hat{v}$, luego

$$c_1 = \frac{u^t v(0)}{u^t \hat{v}}.$$

El denominador es positivo porque $u > 0$ y $\hat{v} > 0$. El numerador es positivo porque $u > 0$ y $v(0) \geq 0$ es no nulo: ningún vector no negativo y no nulo es ortogonal a un vector estrictamente positivo. Así $\alpha := c_1 > 0$.

Paso 2: contar cabezas mide la tasa (apartado 2). Como $\hat{\lambda}^{-k}v(k) \to \alpha\hat{v}$ y la norma es continua, $\hat{\lambda}^{-k}\|v(k)\|_1 \to \alpha\|\hat{v}\|_1 > 0$. Entonces

$$\frac{\|v(k+1)\|_1}{\|v(k)\|_1} = \hat{\lambda}\cdot\frac{\hat{\lambda}^{-(k+1)}\|v(k+1)\|_1}{\hat{\lambda}^{-k}\|v(k)\|_1} \longrightarrow \hat{\lambda}\cdot\frac{\alpha\|\hat{v}\|_1}{\alpha\|\hat{v}\|_1} = \hat{\lambda}.$$

Y como $v(k) \geq 0$, $\|v(k)\|_1 = \sum_j v_j(k)$ es la población total.

Paso 3: el α **se cancela y queda la forma (apartado 3).** Dividiendo numerador y denominador por $\hat{\lambda}^k$,

$$\frac{v_i(k)}{\|v(k)\|_1} = \frac{\hat{\lambda}^{-k}v_i(k)}{\hat{\lambda}^{-k}\|v(k)\|_1} \longrightarrow \frac{\alpha\hat{v}_i}{\alpha\|\hat{v}\|_1} = \frac{\hat{v}_i}{\|\hat{v}\|_1}.$$

■

Qué hace cada hipótesis. La *primitividad* es la que hace que los cocientes $|\mu_j|/\hat{\lambda}$ sean estrictamente menores que 1 en el paso 1a; si L fuera solo irreducible, algún $|\mu_j|$ igualaría a $\hat{\lambda}$, su término giraría sin decaer y el límite no existiría —la población de la cigarra—. La *no negatividad* de $v(0)$ y el hecho de que no sea nulo entran solo en el paso 1b, y entran de forma esencial: son lo que impide que $v(0)$ sea ortogonal a $u > 0$. Con un $v(0)$ de signos mixtos el numerador $u^t v(0)$ puede anularse, y entonces $\hat{\lambda}^{-k} v(k) \to 0$; el enunciado sería falso, aunque la matriz siguiera siendo primitiva. Y la *simplicidad* de $\hat{\lambda}$ es la que asegura que su bloque de Jordan es 1×1: si tuviera un bloque 2×2 aparecería un factor k que no se cancelaría al dividir por $\hat{\lambda}^k$, y el límite se iría a infinito en vez de estabilizarse.

Conexión con el Cálculo Numérico. Los pasos 1a y 3 son, literalmente, la demostración de convergencia del *método de la potencia*: iterar una matriz sobre un vector cuya componente en la dirección dominante no sea nula y normalizar en cada paso converge a esa dirección, con velocidad gobernada por el cociente $|\mu_2|/\hat{\lambda}$ entre el segundo autovalor y el primero. Lo que allí es un algoritmo aquí es un fenómeno: una población con reproducción regular ejecuta el método de la potencia sin saberlo, y su pirámide de edades es la salida.

Lo que hay que llevarse. (i) Se escribe $v(0)$ en una base de Jordan de L y se divide $L^k v(0)$ por $\hat{\lambda}^k$. (ii) La primitividad da $|\mu_j|/\hat{\lambda} < 1$ para los demás autovalores, y todo se apaga salvo $c_1\hat{v}$. (iii) El coeficiente $c_1 = u^t v(0)/u^t\hat{v}$ se lee con el autovector por la izquierda $u > 0$, y es positivo porque $v(0) \geq 0$ no es nulo. (iv) Los apartados 2 y 3 son cocientes en los que α se cancela. En el ejemplo de dos clases de la subsección anterior se ve todo a la vez: $u = (6,7)^t$, $c_1 = 70/9.5 = 140/19$, y el término que se apaga lleva el factor $(-7/12)^k$.

LA IDEA: EL DESTINO SE FACTORIZA EN TASA POR FORMA

El teorema separa dos cosas que parecían ir juntas. Quitando el crecimiento $\hat{\lambda}^k$, el vector no solo apunta hacia $\hat{v}$: converge a un múltiplo concreto $\alpha\hat{v}$, y α es lo único que recuerda de dónde salió. El apartado 2 hace $\hat{\lambda}$ *medible* sin calcular autovalores: el cociente entre el censo de un año y el del anterior tiende a $\hat{\lambda}$; contar cabezas es calcular la norma uno. Y el apartado 3 nombra un objeto propio de la demografía, la *distribución estable de edades*: el reparto porcentual al que tiende toda población, sea cual sea su arranque. A largo plazo, *cuánto* crece (un número) y *con qué forma* (un vector) se desacoplan, y los dos olvidan las condiciones iniciales.

De la igualdad del apartado 1 se deducen tres consecuencias:

- si $\hat{\lambda} < 1$, la población tiende a extinguirse; esto ya era conocido, porque $\hat{\lambda} = \rho(L)$;
- si $\hat{\lambda} > 1$, el tamaño de la población tiende a infinito: es la única posibilidad de que las $\hat{\lambda}^{-k} v_i(k)$ converjan a un número real positivo;
- si $\hat{\lambda} = 1$, el vector de distribución tiende a estabilizarse, acercándose a un vector de componentes finitas y positivas, un autovector asociado a $\hat{\lambda}$.

En los tres casos el comportamiento asintótico se produce **manteniendo las proporciones**: para valores grandes de k, las proporciones entre los grupos tienden a ser las que marca $\hat{v}$.

«Pero ¿esto no es válido para cualquier vector, también fuera del primer octante?»

Casi. La convergencia $\hat{\lambda}^{-k}v(k) \to c_1\hat{v}$ es álgebra lineal pura y vale para cualquier $v(0) \in \mathbb{R}^n$. Lo que cambia fuera del octante es *quién* es c_1: la fórmula $c_1 = u^t v(0)/u^t\hat{v}$ puede dar un número negativo (se alinea con $-\hat{v}$, una «población» de signos mezclados) o, en el hiperplano $u^t x = 0$, exactamente cero: entonces el término de Perron desaparece y $\hat{\lambda}^{-k}v(k) \to 0$, gobernado por los autovalores subdominantes. Ese hiperplano tiene medida nula pero existe. Y ahí está la gracia del primer octante: como $u > 0$, la ecuación $u^t x = 0$ obliga a signos mezclados, de modo que el octante no negativo solo toca ese hiperplano en el origen, que hemos excluido. Por eso el teorema se enuncia ahí: es lo que tiene sentido biológico y es la región donde $\alpha > 0$ está garantizado.

CONEXIÓN: ECUACIONES DIFERENCIALES ORDINARIAS, EL ANÁLOGO CONTINUO DEL TEOREMA

En el sistema lineal continuo $y' = Ay$ del apéndice A, cuando A tiene un autovalor real *simple* λ_1 con parte real estrictamente mayor que la de todos los demás, la solución $y(t) = e^{tA}y(0)$ satisface $e^{-\lambda_1 t}y(t) \to c_1 w_1$, por el mismo argumento de Jordan: los posibles factores polinómicos de los otros bloques quedan dominados por el decaimiento exponencial. La única diferencia es qué significa «dominar»: parte real mayor en el continuo, módulo mayor en el discreto. En el capítulo 2 este mecanismo reaparece al linealizar en un equilibrio.

Ejercicio 1.1. *La población del ejemplo: análisis completo del comportamiento a largo plazo*

ORIGEN: ejercicio 1.1 del curso; es también el núcleo del ejercicio 4 de la hoja de prácticas del tema 1.

Se considera la población del ejemplo, con matriz de Leslie

$$L = \begin{pmatrix} 0 & 1 & 3 & 2 \\ 0.25 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{pmatrix}, \qquad v(0) = \begin{pmatrix} 100 \\ 70 \\ 70 \\ 40 \end{pmatrix}.$$

Se pide:

1. obtener los autovalores y autovectores de L;
2. comprobar la existencia y unicidad de un autovalor positivo dominante $\hat{\lambda}$ de multiplicidad 1;
3. calcular un autovector asociado a $\hat{\lambda}$ con componentes positivas que sumen uno;
4. determinar los porcentajes de hembras en cada clase para tiempos grandes;
5. calcular los vectores $v(k)$ para $k = 1, \ldots, 20$ y el total de hembras en cada $t_k = 5k$, y comprobar que los valores se ajustan a lo que indica el apartado 2 del teorema de convergencia.

PISTAS, DE MENOR A MAYOR

1. *Antes de calcular nada, mira la matriz: ¿cuáles son a_n y las b_i? ¿Hay dos clases fértiles consecutivas? Eso decide qué teoremas se aplican.*
2. *El polinomio característico de una matriz de Leslie se puede desarrollar por la primera fila, o mejor: escribe la ecuación $\varphi(\lambda) = 1$ de la demostración del teorema y multiplica por λ^4.*
3. *El autovector no hace falta calcularlo resolviendo un sistema: la fórmula del apartado 2 del teorema lo da en cascada.*

4. *Para el apartado 5 no se piden veinte cuentas a mano; se pide la práctica 4 de la sección 1.3. Aquí basta hacer dos o tres pasos y leer la tabla.*

Resolución.
La idea, y ordena los cinco apartados en una sola frase. Multiplicar por L es avanzar cinco años. Iterar es avanzar varios períodos, y lo que decide qué pasa a la larga no es la matriz entera sino **su autovalor dominante y su autovector**: el autovector dice *qué forma* acaba teniendo la pirámide de edades, y el autovalor *cuánto crece o mengua* la población cada cinco años. Los cinco apartados son ese número y ese vector calculados con cuidado y después comprobado contra la iteración real.
La imagen que lo predice, y que conviene tener antes de calcular nada: si se descompone el vector inicial en la base de autovectores, aplicar L multiplica cada trozo por su autovalor. El trozo del autovalor grande se hace, con el tiempo, mucho mayor que los demás —porque $(|\lambda_2|/\hat{\lambda})^k \to 0$—, y al final **la población apunta en la dirección del autovector dominante haya salido de donde haya salido**. Por eso el apartado 4 pregunta por porcentajes y no por números: la forma se olvida del principio, el tamaño no.
Y una advertencia de método que este ejercicio enseña mejor que ninguno: **lo primero es mirar la matriz, no calcular**. Los dos teoremas del capítulo tienen hipótesis que se leen a simple vista —la última fecundidad no nula, dos clases fértiles consecutivas— y saber de antemano que el autovalor dominante existe, es simple y es estrictamente dominante cambia por completo lo que hay que demostrar después.
Paso 1: qué teoremas se aplican, antes de calcular nada. Es una matriz de Leslie con $n = 4$, $a = (0, 1, 3, 2)$ y $b = (0.25, 0.5, 0.1)$. Como $a_4 = 2 > 0$ y todos los $b_i > 0$, L es irreducible y se aplica el teorema de Leslie. Como $a_2 = 1 > 0$ y $a_3 = 3 > 0$ (y también $a_3 a_4 > 0$), hay clases fértiles consecutivas: L es primitiva y $\hat{\lambda}$ será estrictamente dominante. Antes de calcular nada ya sabemos que existe un único autovalor real positivo, que es simple, que es el mayor en módulo, y que la población convergerá en proporciones.
Paso 2 (el núcleo): el autovalor dominante. La ecuación $\varphi(\lambda) = 1$ del teorema es

$$\frac{a_1}{\lambda} + \frac{a_2 b_1}{\lambda^2} + \frac{a_3 b_1 b_2}{\lambda^3} + \frac{a_4 b_1 b_2 b_3}{\lambda^4} = 1,$$

es decir, con los números, $0 + \dfrac{0.25}{\lambda^2} + \dfrac{0.375}{\lambda^3} + \dfrac{0.025}{\lambda^4} = 1$. Multiplicando por λ^4:

$$p(\lambda) = \lambda^4 - 0.25\lambda^2 - 0.375\lambda - 0.025 = 0.$$

Este es, salvo el signo, el polinomio característico de L, y conviene comprobarlo desarrollando $\det(L - \lambda I)$ por la primera fila: se obtiene $\lambda^4 - a_1\lambda^3 - a_2 b_1 \lambda^2 - a_3 b_1 b_2 \lambda - a_4 b_1 b_2 b_3$, que con los datos es $p(\lambda)$. No tiene raíces racionales sencillas; se resuelve numéricamente (es lo que hace `eig` en la práctica) y salen

$$\hat{\lambda} \approx 0.851425, \qquad \lambda_{2,3} \approx -0.390784 \pm 0.517315\, i, \qquad \lambda_4 \approx -0.069856.$$

Se puede acotar $\hat{\lambda}$ a mano: $p(0.8) = 0.4096 - 0.16 - 0.3 - 0.025 = -0.0754 < 0$ y $p(0.9) = 0.6561 - 0.2025 - 0.3375 - 0.025 = 0.0911 > 0$, luego la raíz positiva está entre 0.8 y 0.9, y como p es creciente en esa zona es única ahí.
Apartado 2: dominancia. Los módulos son 0.851425, 0.648326 (los dos complejos), y 0.069856. El autovalor $\hat{\lambda}$ es real, positivo, simple, y estrictamente mayor en módulo que los demás:

es lo que predecían el teorema de Leslie y el de Perron. La unicidad como autovalor real positivo se ve también en que φ es decreciente, como en la demostración.
Apartado 3: autovector. Con la fórmula en cascada, $\hat{v}_1 = 1$, $\hat{v}_2 = b_1/\hat{\lambda} = 0.25/0.851425 \approx 0.293625$, $\hat{v}_3 = b_1 b_2/\hat{\lambda}^2 = 0.125/0.724924 \approx 0.172432$, $\hat{v}_4 = b_1 b_2 b_3/\hat{\lambda}^3 = 0.0125/0.617219 \approx 0.020252$. La suma es 1.486309, y dividiendo por ella el autovector con componentes que suman uno es

$$\frac{\hat{v}}{\|\hat{v}\|_1} \approx (0.6728,\ 0.1976,\ 0.1160,\ 0.0136)^t.$$

Comprobación con la primera fila: $1 \cdot 0.293625 + 3 \cdot 0.172432 + 2 \cdot 0.020252 = 0.851425 = \hat{\lambda} \cdot \hat{v}_1$.
Apartado 4: porcentajes a largo plazo. Por el apartado 3 del teorema de convergencia, son los del autovector normalizado: 67.3 % en la clase 1, 19.8 % en la 2, 11.6 % en la 3 y 1.4 % en la 4. Casi siete de cada diez hembras serán crías, y la clase de más edad será testimonial. Esto no depende de $v(0)$.
Apartado 5: la iteración. Ya se calcularon $v(1) = (360, 25, 35, 7)^t$ y $v(2) = (144, 90, 12.5, 3.5)^t$, con totales 280, 427, 250. La tabla completa la produce la práctica 4; los totales de los últimos pasos son 21.30, 18.06 y 15.38 para $k = 18, 19, 20$, cuyos cocientes sucesivos son 0.848 y 0.852, que oscilan ya alrededor de $\hat{\lambda} = 0.8514$. Las proporciones en $k = 20$ son $(0.6723, 0.1973, 0.1168, 0.0136)$, que se aproximan a las del autovector (el error máximo por componente es menor que 0.001). Y como $\hat{\lambda} < 1$, la población se extingue: pierde en torno a un 15 % cada cinco años.

Otra vía (qué cambiaría el destino). La pregunta de la demógrafa era dónde intervenir. El autovalor está determinado por $\varphi(\hat{\lambda}) = 1$, y φ crece con cada a_i y cada b_i; luego subir cualquiera de ellos sube $\hat{\lambda}$. Pero no todos por igual: el término en b_1 multiplica a toda la suma salvo el primero, mientras que b_3 solo aparece en el último sumando, que vale $0.025/\hat{\lambda}^4 \approx 0.048$. Duplicar la supervivencia de la clase 3 apenas mueve φ; duplicar b_1 la duplica casi entera. Con $b_1 = 0.5$ la ecuación queda $\lambda^4 - 0.5\lambda^2 - 0.75\lambda - 0.05 = 0$, y $p(1) = 1 - 0.5 - 0.75 - 0.05 < 0$ mientras $p(1.2) = 2.0736 - 0.72 - 0.9 - 0.05 > 0$: el autovalor pasa a estar entre 1 y 1.2, y la población crece. Esto es lo que descubrió el análisis de las tortugas: la sensibilidad de $\hat{\lambda}$ a cada coeficiente dice dónde rinde el esfuerzo.

Dónde se falla. En cinco sitios, y los cinco se ven en este ejercicio.

- *Al escribir L.* Las fecundidades van en la primera fila y las supervivencias en la subdiagonal, y hay $n - 1$ supervivencias, no n. Poner las b_i en la diagonal describe otra población, una en la que nadie pasa a la clase siguiente: la matriz queda triangular, sus autovalores son 0, 0.25, 0.5 y 0.1, y el dominante sale 0.5 en vez de 0.8514.
- *Al decidir con los primeros pasos.* El total pasa de 280 a 427; quien se pare ahí concluye que la población crece, y se extingue. El transitorio no dice nada del destino: lo dice $\hat{\lambda}$, o $\varphi(1) = 0.65$.
- *Al comparar autovalores sin tomar módulos.* Los complejos $-0.3908 \pm 0.5173\,i$ tienen parte real negativa, pero la dominancia se decide por el módulo, 0.648. Mirar la parte real es el criterio del sistema continuo $y' = Ay$, no el del discreto.
- *Al desplazar un exponente en la cascada.* Con $v_1 = 1$, la componente i lleva $\hat{\lambda}^{-(i-1)}$, no $\hat{\lambda}^{-i}$. Si se deja $v_1 = 1$ y se pone el exponente equivocado en las demás, sale un vector que no es autovector, y la comprobación con la primera fila lo detecta en una línea: $(Lv)_1$ da 1 donde

debería dar $\hat{\lambda} v_1 \approx 0.851$. (Si el exponente se desplaza en todas las componentes, incluida la primera, el vector es $\hat{v}/\hat{\lambda}$, que sí es autovector: el error de verdad es el mezclado.)
- *Al normalizar con la norma que no toca.* Los porcentajes piden componentes que sumen uno. La orden `eig` de MATLAB devuelve el autovector con norma euclídea uno, $(0.9465, 0.2779, 0.1632, 0.0192)^t$ o su opuesto, con todos los signos cambiados, y ese vector suma 1.407 (o -1.407): leerlo como porcentajes daría un 94.6 % de crías en vez del 67.3 %. Hay que dividir por la suma.

QUÉ DEJA. Ante una matriz de Leslie, lo primero no es calcular, sino leer: $a_n > 0$ da irreducibilidad, dos fértiles consecutivas dan primitividad, y con eso ya se sabe cómo va a acabar todo. La ecuación $\varphi(\lambda) = 1$ es el polinomio característico en forma útil, y la cascada $v_i = (b_{i-1}/\lambda)v_{i-1}$ da el autovector sin resolver ningún sistema.

1.2 Medida de la relevancia de una página web: PageRank

LA PREGUNTA

En 1998 había ya decenas de millones de páginas web y ninguna persona podía leerlas para decidir cuáles eran las buenas. Dos estudiantes de doctorado querían ordenarlas de más a menos importante *sin mirar su contenido*, usando solo qué página enlaza a cuál. ¿Puede una red de enlaces, por sí sola, decir qué página importa más, y cómo se evita que el resultado dependa de por dónde se empiece a contar?

PageRank asigna a cada página web un número de relevancia a partir de los enlaces de la red. Fue una pieza importante del buscador de Google, pero ese número no determina por sí solo el orden de los resultados. Lo desarrollaron los fundadores de Google, Larry Page (de cuyo apellido toma el nombre) y Sergey Brin, en la Universidad de Stanford durante su posgrado. Google explica que PageRank, muy evolucionado desde entonces, sigue formando parte de sus sistemas principales de clasificación junto con muchos otros factores.[2]

1.2.1 El modelo

PageRank interpreta un enlace de una página A a otra B como un **voto** de A para B: cuantos más votos tiene una página, más relevante se considera, y más aún cuando esos votos provienen de páginas que son a su vez muy votadas. Páginas y enlaces son, pues, nodos y aristas de un **grafo dirigido**.

Definición. Dada una red de n páginas, su **matriz de adyacencia** o de conectividad es la matriz cuadrada $A = \{a_{ij}\}$ de orden n, que solo contiene ceros y unos, dada por

$$a_{ij} = \begin{cases} 1 & \text{si hay un enlace } j \to i, \\ 0 & \text{en caso contrario.} \end{cases}$$

El grafo asociado G_A es el de la subsección 1.1.3.1: hay una arista de j a i si y solo si $a_{ij} \neq 0$.

[2] Google, *A guide to Google Search ranking systems*, apartado sobre análisis de enlaces y PageRank: `https://developers.google.com/search/docs/appearance/ranking-systems-guide`.

Es la misma regla de oro de los subíndices: la columna es el origen y la fila el destino. En el caso de Google, la fila i contiene los votos *recibidos* por la página i, y la columna j los *otorgados* por la página j. Para la red de cinco páginas de la figura 1.2, la matriz de conectividad es

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

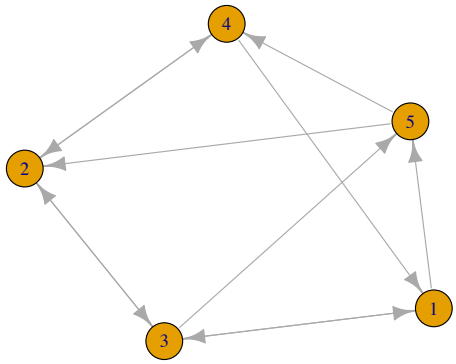


Figura 1.2: Una red con 5 páginas representada mediante un grafo dirigido con 5 nodos. Una arista de j a i significa que la página j contiene un enlace a la página i.

EXPLÍCATE. *Lee de la figura los enlaces que salen de la página 1 y comprueba que son la columna 1 de A. Después lee los que llegan a la página 2 y comprueba que son la fila 2.*

Los votos recibidos por una página son una primera medida de su relevancia. En el ejemplo, la página 2 tiene tres votos y podría considerarse la más importante. Pero esta medida no tiene en cuenta la *calidad* de los votos: tres enlaces desde páginas que nadie visita no valen lo que uno desde una página muy importante.

EL CAMINO QUE NO SE TOMÓ: CONTAR LOS ENLACES RECIBIDOS

La alternativa. Ordenar las páginas por el número de enlaces que reciben, es decir, por la suma de cada fila de A.
Su ventaja. Se calcula de un vistazo —es sumar cada fila, sin iterar nada— y ya captura la idea de que un enlace es un voto.
Qué la hunde. Dos cosas. La primera es la calidad: trata igual todos los votos, y por eso no distingue un enlace desde una página de referencia de uno desde una página vacía. La segunda es la manipulación: bastaría crear miles de páginas sin contenido que enlacen a la propia para subir en la lista. Hacer que el voto de una página valga lo que vale la página incorpora la calidad de los votos y convierte el problema en uno de autovectores. No lo hace inmune a la manipulación de enlaces.

Primer intento: la relevancia como autovector. Se busca un vector x de componentes no negativas que asigne a cada página su relevancia, medida teniendo en cuenta la relevancia de las páginas que enlazan con ella. En un primer intento se impone que cada x_i sea **proporcional**

a la suma de las relevancias de las páginas que enlazan con la página i. En el ejemplo, si $K > 0$ es la constante de proporcionalidad,

$$\begin{cases} x_1 = K(x_3 + x_4), \\ x_2 = K(x_3 + x_4 + x_5), \\ \quad \vdots \\ x_5 = K(x_1 + x_3), \end{cases} \qquad \text{esto es,} \qquad x = KAx.$$

De manera equivalente, $Ax = \lambda x$ con $\lambda = 1/K > 0$: x debe ser un autovector de A de componentes no negativas asociado a un autovalor positivo. Este es exactamente el tipo de objeto que el teorema de Frobenius garantiza cuando A es irreducible, y ya se ve por qué el capítulo junta Leslie y PageRank.

Esta estrategia da relevancia a la página i según el número de páginas que enlazan con ella y la relevancia que poseen. Pero se puede ir más allá y tener en cuenta si cada una de esas páginas tiene muchas o pocas posibilidades de enlazar con i: una página que reparte su voto entre cien enlaces debería pesar menos en cada uno que una que solo tiene dos. Con ese objetivo se analiza el problema desde una perspectiva probabilística.

La matriz de transición. Lo primero es construir a partir de A la **matriz de transición** P, que contiene las probabilidades de pasar de una página a otra. Basta dividir cada columna no nula de A por el número de componentes no nulas que posee, que es la suma de la columna, y dejar provisionalmente las columnas nulas como están: esas columnas necesitarán una corrección para conservar la probabilidad. En el ejemplo,

$$P = \begin{pmatrix} 0 & 0 & 1/3 & 1/2 & 0 \\ 0 & 0 & 1/3 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/3 & 0 & 0 \end{pmatrix}.$$

La columna 3, por ejemplo, tenía tres unos en A: la página 3 enlaza a 1, 2 y 5, y desde ella se salta a cada una con probabilidad $1/3$.

Supongamos de momento que no hay páginas sin enlaces, es decir, nodos sin salida (ninguna columna de A es toda de ceros), ni páginas sin votos (ninguna fila es toda de ceros). Imaginemos un internauta que navega por las páginas de manera aleatoria, en etapas sucesivas, y denotemos x^k al vector de $\mathbb{R}^n$ cuyas componentes $x_1^k, \ldots, x_n^k$ son las probabilidades de estar en las páginas $1, \ldots, n$ en la etapa k.

1. El internauta empieza en la página 1. Eso es $x^0 = e_1$, el primer vector de la base canónica: probabilidad 1 de estar en la página 1 y cero en el resto.
2. Cambia a otra página usando los enlaces de la página 1, que llevan a la 3 y a la 5. Acaba en la 3 con probabilidad $1/2$ o en la 5 con probabilidad $1/2$:

$$x^1 = (0,\, 0,\, 1/2,\, 0,\, 1/2)^t = Px^0.$$

3. Sigue navegando aleatoriamente. Desde la 3 va a 1, 2 o 5 con probabilidad $1/3$ cada una; desde la 5 va a 2 o 4 con probabilidad $1/2$. Sumando lo que llega a cada página,

$$x^2 = Px^1 = P^2x^0 = \left(\tfrac{1}{6},\, \tfrac{5}{12},\, 0,\, \tfrac{1}{4},\, \tfrac{1}{6}\right)^t.$$

Comprobemos la segunda componente: a la página 2 se llega desde la 3 con $\frac{1}{3} \cdot \frac{1}{2}$ y desde la 5 con $\frac{1}{2} \cdot \frac{1}{2}$, y $\frac{1}{6} + \frac{1}{4} = \frac{5}{12}$.

4. Y así sucesivamente: en la etapa $k \geq 1$,

$$x^k = Px^{k-1} = P^k x^0.$$

Este análisis permite interpretar las componentes del vector

$$x_P := \lim_{k \to \infty} P^k x^0$$

(si existe y es independiente de x^0) como las distintas relevancias de las páginas: las probabilidades de visita a largo plazo. En una cadena finita irreducible, la distribución estacionaria también da, casi seguramente, las fracciones de tiempo pasadas en cada página; esta afirmación sobre promedios temporales no exige que $P^k x^0$ converja. Y de nuevo surgen las tres cuestiones de la sección 1.1.3: ¿existe el límite?; en tal caso, ¿son sus componentes positivas y guardan relación con los autovalores y autovectores de P?; ¿es el límite independiente de x^0?

CONEXIÓN: CADENAS DE MARKOV

Lo que se acaba de construir es una cadena de Markov con espacio de estados las n páginas: condicionada a la historia previa, la probabilidad de estar en i en la etapa $k+1$ depende solo de dónde se estaba en la etapa k, y P es su matriz de transición escrita por columnas (en los textos de probabilidad suele escribirse traspuesta, por filas, y entonces se itera $x^t P$). El vector x_P es la distribución estacionaria de la cadena, y las tres preguntas son las clásicas de existencia, unicidad y convergencia a la distribución estacionaria. El teorema de Perron-Frobenius permite contestarlas: para una cadena finita, primitividad equivale a *irreducibilidad y aperiodicidad*. La irreducibilidad sola garantiza una distribución estacionaria única, pero puede haber oscilaciones en $P^k x^0$.

Teorema. Supongamos que $A \in \mathbb{R}^{n \times n}$ es la matriz de conectividad de un grafo dirigido con

$$\sum_{i=1}^{n} a_{ij} > 0 \qquad \text{para todo } j = 1, \ldots, n,$$

es decir, sin nodos sin salida. Sea P la correspondiente matriz de transición. Entonces:

1. P es **estocástica por columnas**: sus componentes son ≥ 0 y la suma de las componentes de cada columna vale siempre 1. Además, su radio espectral es $\rho(P) = 1$ y un autovalor principal es $\lambda_1 = 1$;
2. si $x \in \mathbb{R}^n$ es un vector estocástico (componentes ≥ 0 que suman 1), también lo es Px;
3. existe un autovector estocástico de P asociado a 1.

LA IDEA: LA CONSERVACIÓN DE LA MASA, Y EL PROBLEMA DE LOS CEROS

El teorema entero es la contabilidad de una única magnitud conservada. Si se reparte un kilo de probabilidad entre las n páginas y se deja que el navegante dé un paso, no se crea ni se destruye masa: cada página reenvía íntegramente lo que tenía, repartido entre sus enlaces. Eso es «las columnas suman 1», y de ahí salen solos los dos primeros apartados: el símplex de los vectores estocásticos es invariante, y ninguna dirección puede amplificarse, porque amplificar sería fabricar masa. Con masa que no crece, ningún autovalor puede tener módulo mayor que 1, y como el reparto total sí se conserva, el 1 se alcanza.

El apartado 3 es el que cuesta, y conviene ver por qué. Que 1 sea autovalor no garantiza que su autovector sea *estocástico*: podría tener componentes negativas, y un reparto de probabilidad negativa no significa nada. Perron daría la positividad de golpe, pero exige entradas estrictamente positivas y P está llena de ceros —la web real es casi toda ceros—. **El núcleo de la prueba es el rodeo que resuelve esto**: se ensucia la matriz sumando ε a todas las entradas, con lo que pasa a ser estrictamente positiva y Perron se aplica; se recoge el autovector positivo que Perron regala; y se deja que $\varepsilon \to 0$. La positividad no sobrevive al límite —puede degradarse a ≥ 0—, pero eso basta, y lo que sí sobrevive es la pertenencia al símplex, porque el símplex es **compacto**. La compacidad es el clavo del que cuelga todo el apartado.

Paso 1: la masa se conserva, luego nada se amplifica (apartado 1). Que P es no negativa y sus columnas suman 1 es la construcción: cada columna de A tiene $s_j > 0$ unos y se ha dividido por s_j. Sea e el vector de todas las componentes iguales a 1. Que las columnas de P sumen 1 se escribe $e^t P = e^t$, o traspuesto, $P^t e = e$: el vector e es autovector de P^t con autovalor 1. Como P y P^t tienen el mismo polinomio característico, 1 es autovalor de P. Para ver que ningún autovalor tiene módulo mayor que 1 se usa el teorema de Gershgorin, recordado en la caja siguiente, aplicado a P^t: sus discos están centrados en p_{jj} con radio $\sum_{i \neq j} p_{ij} = 1 - p_{jj}$, y el punto más alejado del origen de un disco así está a distancia $p_{jj} + (1 - p_{jj}) = 1$. Todo autovalor de P^t, y por tanto de P, tiene módulo ≤ 1. Como 1 es autovalor, $\rho(P) = 1$.
Paso 2: el símplex es invariante (apartado 2). Si x es estocástico, $Px \geq 0$ por ser producto de no negativos, y

$$\sum_{i=1}^{n} \Big(\sum_{j=1}^{n} p_{ij} x_j \Big) = \sum_{j=1}^{n} \Big(\sum_{i=1}^{n} p_{ij} \Big) x_j = \sum_{j=1}^{n} x_j = 1,$$

donde se ha intercambiado el orden de las sumas finitas y se ha usado que cada columna suma 1.
Paso 3: ensuciar la matriz para poder usar Perron, y limpiarla después (apartado 3, el núcleo de la prueba). Para cada $\varepsilon > 0$ sea $P_\varepsilon := P + \varepsilon U$, con U la matriz de orden n con todas las componentes iguales a 1. La matriz P_ε tiene todas sus componentes estrictamente positivas, luego se le aplica el teorema de Perron: existe un autovalor simple y estrictamente dominante $\lambda_\varepsilon = \rho(P_\varepsilon)$ y un único autovector estocástico asociado x_ε. Ahora dos observaciones:

- Los x_ε viven en el símplex $\{x \geq 0,\ \|x\|_1 = 1\}$, que es un conjunto compacto de $\mathbb{R}^n$. Por el teorema de Bolzano-Weierstrass, alguna subsucesión, que se sigue indicando con $\varepsilon \to 0$, converge a un x_P con $x_P \geq 0$ y $\|x_P\|_1 = 1$.
- Tomando la norma uno en $P_\varepsilon x_\varepsilon = \lambda_\varepsilon x_\varepsilon$, y usando que todo es no negativo, de modo que la norma uno es la suma de las componentes,

  $$\lambda_\varepsilon = \|P_\varepsilon x_\varepsilon\|_1 = \|P x_\varepsilon\|_1 + \varepsilon \|U x_\varepsilon\|_1 = 1 + n\varepsilon,$$

  porque Px_ε es estocástico por el apartado 2 y $Ux_\varepsilon = e \cdot \sum_j x_{\varepsilon,j} = e$, cuya norma uno es n. Luego $\lambda_\varepsilon \to 1$ cuando $\varepsilon \to 0$.

Pasando al límite en $P_\varepsilon x_\varepsilon = \lambda_\varepsilon x_\varepsilon$ a lo largo de la subsucesión, y usando que el producto por una matriz es continuo, queda $Px_P = x_P$: x_P es un autovector estocástico asociado al autovalor 1, que ya sabíamos que coincide con $\rho(P)$. ■

Qué hace cada hipótesis, y qué no se ha demostrado. La hipótesis $\sum_i a_{ij} > 0$ —ningún nodo sin salida— es la que permite dividir por s_j y construir P; sin ella habría columnas enteras de ceros, la masa desaparecería por el sumidero y ni siquiera el apartado 2 sería cierto. Conviene notar, además, lo que este teorema *no* dice: no dice que el autovector estocástico sea único, ni que la iteración $x^{k+1} = Px^k$ converja a él. Las dos cosas fallan de verdad en la web real: una red partida en dos islas sin enlaces entre ellas tiene dos distribuciones estacionarias distintas, y una red con un ciclo rígido oscila sin converger. Recuperar unicidad y convergencia queda garantizada si la matriz es primitiva. Es una condición suficiente, no necesaria, y una red real puede no cumplirla; una manera de asegurarla es *fabricarla*: eso es exactamente lo que hace el factor de amortiguación de la sección siguiente, que no es un parche heurístico sino la traducción a ingeniería del rodeo del paso 3 —sumar una matriz positiva pequeña— convertido esta vez en parte del modelo.

Lo que hay que llevarse. (i) «Las columnas suman uno» se escribe $e^t P = e^t$: el 1 es autovalor de P^t, luego de P, y Gershgorin impide que haya autovalores de módulo mayor. (ii) Por la misma razón el símplex es invariante: la masa ni se crea ni se destruye. (iii) Para tener un autovector *estocástico* se ensucia la matriz, $P + \varepsilon U > 0$, Perron da un autovector en el símplex, y la compacidad del símplex permite pasar al límite $\varepsilon \to 0$. El precio: se obtiene existencia, pero no unicidad ni convergencia.

CONEXIÓN: ÁLGEBRA LINEAL Y ANÁLISIS, DOS TEOREMAS QUE HACÍAN FALTA

Teorema de Gershgorin. Sea $B = \{b_{ij}\}$ una matriz compleja de orden n. Todo autovalor de B pertenece a la unión de los n discos $D_i = \{z \in \mathbb{C} : |z - b_{ii}| \leq \sum_{j \neq i} |b_{ij}|\}$. La demostración es corta: si $Bz = \mu z$ con $z \neq 0$, sea i el índice de la componente de z de mayor módulo; la fila i da $(\mu - b_{ii})z_i = \sum_{j \neq i} b_{ij} z_j$, y dividiendo por $|z_i|$ y acotando $|z_j|/|z_i| \leq 1$ sale $|\mu - b_{ii}| \leq \sum_{j \neq i} |b_{ij}|$. Aquí se aplica por filas a P^t, que es lo mismo que por columnas a P.
Teorema de Bolzano-Weierstrass en $\mathbb{R}^n$. Toda sucesión acotada de $\mathbb{R}^n$ tiene una subsucesión convergente; equivalentemente, un cerrado y acotado de $\mathbb{R}^n$ es compacto. Es el resultado de Topología y Análisis que se supone conocido de partida («compacidad en $\mathbb{R}^n$»), y es el que permite extraer el límite x_P de la familia x_ε. Reaparece en el capítulo 4 en versión débil, para espacios de funciones, donde la compacidad es muchísimo más delicada.

LA IDEA: POR QUÉ EL RODEO POR P_ε

La matriz P tiene ceros, quizá muchos, y puede no ser ni irreducible ni primitiva; el teorema de Perron no se le aplica directamente. Sumarle εU es añadir un enlace diminuto de toda página a toda página: la matriz se vuelve positiva y Perron regala un autovector estocástico. Después se deja $\varepsilon \to 0$, y lo único que hace falta para que sobreviva un autovector es la compacidad del símplex. Es la misma idea que, con una perturbación fija y una normalización que conserve las sumas de las columnas, reaparece en el factor de amortiguación de Google: $(P + \varepsilon U)/(1 + n\varepsilon)$ es $\alpha P + (1 - \alpha)U/n$ para $\alpha = 1/(1 + n\varepsilon)$. El precio del rodeo es que solo da *existencia*: el autovalor

> 1 es dominante pero no necesariamente estricto, y x_P no es necesariamente único.

Si P es además irreducible, el teorema de Frobenius dice que el autovector estocástico es único, y entonces, por definición, x_P **es el valor PageRank de la red** representada por A. Sus componentes se interpretan así: si un internauta se mueve aleatoriamente por la red, la fracción de tiempo que pasa en la página i converge casi seguramente a $x_{P,i}$. Si, además, la cadena es aperiódica, la probabilidad de estar en esa página en el instante k también converge a $x_{P,i}$.

Ejercicio 1.2. *El PageRank de la red de cinco páginas, calculado a mano*

ORIGEN: ejemplo del curso, donde el resultado se da en forma numérica; aquí se obtiene exacto.

Para la red de la figura 1.2, con la matriz de transición P de arriba:

1. comprobar que P es irreducible;
2. calcular el único autovector estocástico x_P de P asociado a 1, en forma exacta, y ordenar las páginas por relevancia;
3. explicar por qué las páginas 3, 4, 1 y 5, que reciben el mismo número de enlaces, no tienen la misma relevancia.

Resolución.

La idea. El PageRank no cuenta enlaces: **reparte reputación**. Cada página tiene una cantidad de prestigio, lo divide a partes iguales entre las páginas a las que enlaza y lo entrega; el prestigio de cada una es lo que recibe. Escrito en álgebra, eso es exactamente $Px = x$: el vector que se queda igual después de una ronda de reparto. De ahí que el problema sea de autovectores y no de contar, y de ahí también la respuesta del apartado 3, que es la moraleja del modelo: **un voto vale lo que vale quien lo da dividido entre cuántos votos reparte**, así que dos páginas con el mismo número de enlaces recibidos pueden acabar muy lejos una de otra.

El apartado 1 no es un trámite: es lo que garantiza que la pregunta tiene *una* sola respuesta. Que el grafo sea fuertemente conexo —que desde cualquier página se pueda llegar a cualquier otra siguiendo enlaces— es lo que hace irreducible a P, y sin eso podría haber dos repartos estables distintos, o un rincón de la red que se quedara con todo el prestigio. Comprobarlo es tan barato como exhibir *un* ciclo que pase por todos los vértices.

Y una comprobación que hay que hacer siempre y que aquí sale gratis: el sistema $Px = x$ tiene una ecuación **de sobra**, porque las columnas de P suman uno y por tanto las cinco ecuaciones son linealmente dependientes. Se resuelve con cuatro y la quinta sirve de control: si no cuadra, hay un error de cuenta.

Paso 1: irreducibilidad, que es lo que da la unicidad. El ciclo $1 \to 5 \to 4 \to 2 \to 3 \to 1$ pasa por todos los nodos: hay enlaces $1 \to 5$, $5 \to 4$, $4 \to 2$, $2 \to 3$ y $3 \to 1$, todos visibles en la figura y en las columnas correspondientes de A. Un ciclo dirigido que pasa por todos los vértices hace al grafo fuertemente conexo, y por la caracterización (II) de la subsección 1.1.3.1, P es irreducible. Por Frobenius, el autovector estocástico de 1 es único. Además, hay ciclos $1 \to 3 \to 1$ y $1 \to 5 \to 4 \to 1$, de longitudes 2 y 3; su máximo común divisor es 1. Por tanto, la matriz es primitiva y la iteración desde cualquier reparto inicial converge.

Paso 2 (el núcleo): resolver $Px = x$ con fracciones, usando la quinta ecuación como comprobación.

Las ecuaciones $Px = x$ son

$$\begin{aligned} x_1 &= \tfrac{1}{3}x_3 + \tfrac{1}{2}x_4, \quad x_2 = \tfrac{1}{3}x_3 + \tfrac{1}{2}x_4 + \tfrac{1}{2}x_5, \\ x_3 &= \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2, \quad x_4 = \tfrac{1}{2}x_2 + \tfrac{1}{2}x_5, \\ x_5 &= \tfrac{1}{2}x_1 + \tfrac{1}{3}x_3. \end{aligned}$$

Restando la primera de la segunda, $x_2 - x_1 = \frac{1}{2}x_5$. De la quinta, $x_5 = \frac{1}{2}x_1 + \frac{1}{3}x_3$, y de la tercera, $x_3 = \frac{1}{2}(x_1 + x_2)$. Sustituyendo esta en aquella, $x_5 = \frac{1}{2}x_1 + \frac{1}{6}(x_1 + x_2) = \frac{2}{3}x_1 + \frac{1}{6}x_2$, y entonces $x_2 - x_1 = \frac{1}{3}x_1 + \frac{1}{12}x_2$, es decir $\frac{11}{12}x_2 = \frac{4}{3}x_1$, luego $x_2 = \frac{16}{11}x_1$. De ahí, $x_3 = \frac{1}{2}\big(1 + \frac{16}{11}\big)x_1 = \frac{27}{22}x_1$, $x_5 = \frac{2}{3}x_1 + \frac{1}{6}\cdot\frac{16}{11}x_1 = \frac{44+16}{66}x_1 = \frac{10}{11}x_1$, y de la cuarta, $x_4 = \frac{1}{2}\big(\frac{16}{11} + \frac{10}{11}\big)x_1 = \frac{13}{11}x_1$. Comprobación con la primera ecuación, que no se ha usado: $\frac{1}{3}\cdot\frac{27}{22} + \frac{1}{2}\cdot\frac{13}{11} = \frac{9}{22} + \frac{13}{22} = 1$. Correcto. Tomando $x_1 = 22$ para quitar denominadores, $x = (22, 32, 27, 26, 20)$, cuya suma es 127. Por tanto

$$x_P = \frac{1}{127}\begin{pmatrix} 22 \\ 32 \\ 27 \\ 26 \\ 20 \end{pmatrix} \approx \begin{pmatrix} 0.1732283 \\ 0.2519685 \\ 0.2125984 \\ 0.2047244 \\ 0.1574803 \end{pmatrix},$$

La página 2 es la más relevante, seguida de la 3, la 4, la 1 y la 5.

Paso 3: por qué el mismo número de votos da relevancias distintas. Las páginas 3, 4, 1 y 5 reciben dos enlaces cada una. La 3 los recibe de la 1 y de la 2, que solo reparten su voto entre dos páginas cada una, y la 2 es la más relevante: recibe $\frac{1}{2}\cdot 22 + \frac{1}{2}\cdot 32 = 27$. La 5 recibe de la 1, que da la mitad, y de la 3, que reparte entre tres: $\frac{1}{2}\cdot 22 + \frac{1}{3}\cdot 27 = 20$. Mismo número de votos, pero votos de distinto peso y de distinta calidad. Eso es exactamente lo que el primer intento no medía y el modelo probabilístico sí.

Otra vía (por el método de la potencia). Iterando $x^{k+1} = Px^k$ desde $x^0 = e/5$, en 43 iteraciones el error en norma uno es menor que $5 \cdot 10^{-7}$ (la práctica 5 lo programa en un caso mayor). El segundo autovalor en módulo es ≈ -0.765, que explica el decaimiento asintótico como 0.765^k. En esta red sin nodos sin salida, la amortiguación multiplica por α los autovalores distintos de 1, acelerando ese decaimiento. Más adelante se demostrará una cota general del error proporcional a α^k.

Dónde se falla.

- *Al construir P.* Se divide cada columna de A por su suma, porque la columna j son los enlaces que salen de j. Dividir por las sumas de las filas da una matriz cuyas filas suman uno y cuyas columnas no: aquí las columnas sumarían 1, 1, 4/3, 5/6 y 5/6, y el internauta crearía y destruiría probabilidad. Comprobar que cada columna suma 1 cuesta cinco sumas y lo detecta.
- *Al tratar las cinco ecuaciones como independientes.* No lo son: $P - I$ es singular, y el sistema homogéneo tiene una recta entera de soluciones. Se resuelven cuatro y la quinta se usa de comprobación, o se sustituye una de ellas por $x_1 + \cdots + x_5 = 1$.
- *Al olvidar normalizar.* El vector $(22, 32, 27, 26, 20)$ es un autovector de 1, pero el PageRank es el estocástico, dividido por 127.
- *Al confundir votos con relevancia.* Las páginas 3 y 5 reciben dos enlaces cada una y quedan segunda y última. Contar la fila de A no ordena; hay que resolver.

QUÉ DEJA. El PageRank es el autovector estocástico de la matriz de transición, y se calcula a mano resolviendo $Px = x$ con una ecuación de sobra. Y la lección de modelado: dos páginas con los mismos votos pueden tener relevancias distintas, porque un voto vale lo que vale quien lo da dividido entre cuántos votos reparte.

Nodos sin salida. Supongamos ahora que en el grafo hay nodos sin salida. Las columnas correspondientes de A y de P solo contienen ceros, y P ya no es estocástica. Se define entonces la nueva matriz de transición

$$P' := P + w\,d^t, \qquad d = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}, \quad d_j := \begin{cases} 1 & \text{si la columna } j \text{ está llena de ceros,} \\ 0 & \text{si no,} \end{cases}$$

donde w es un vector estocástico. El producto $w\,d^t$ es una matriz de rango como mucho uno cuya columna j es $d_j w$: vale w en las columnas de los nodos sin salida y cero en las demás. Lo más habitual es elegir w uniforme, $w = \frac{1}{n}e$; con esa elección, P' se obtiene de P sustituyendo las columnas de los nodos sin salida por columnas con todas las componentes iguales a $1/n$. Equivale a añadir enlaces artificiales desde las páginas sin salida a todas las páginas, con igual probabilidad: el internauta que llega a una página sin enlaces salta a cualquiera.

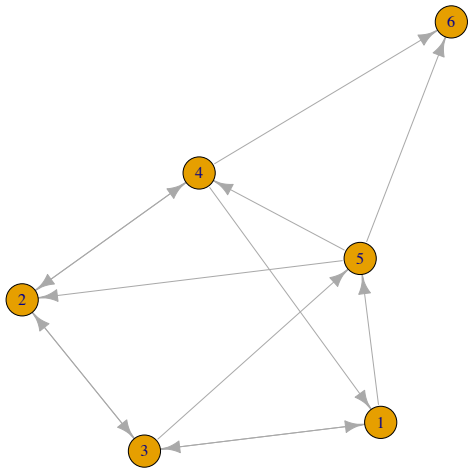


Figura 1.3: La red de la figura 1.2 con un nodo más, el 6, que no tiene salida: no contiene ningún enlace a otra página. Recibe enlaces de la 4 y de la 5.

Ejemplo. La red de la figura 1.3 contiene un nodo sin salida. Supongamos que se quiere privilegiar el acceso desde el nodo 6 a los nodos 4 y 5. Un posible w es $w = (1/6, 1/6, 1/6, 1/4, 1/4, 0)^t$, que es estocástico y da más peso a 4 y 5. Las matrices P y $P' = P + wd^t$ de esta red son

$$P = \begin{pmatrix} 0 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 \end{pmatrix}, \qquad P' = \begin{pmatrix} 0 & 0 & 1/3 & 1/3 & 0 & 1/6 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 1/6 \\ 1/2 & 1/2 & 0 & 0 & 0 & 1/6 \\ 0 & 1/2 & 0 & 0 & 1/3 & 1/4 \\ 1/2 & 0 & 1/3 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 \end{pmatrix}.$$

Obsérvese que las columnas 4 y 5 de P han cambiado respecto de la red de cinco nodos: ahora la página 4 enlaza a 1, 2 y 6, y la 5 a 2, 4 y 6, y cada una reparte entre tres. Solo $d_6 = 1$, así que wd^t es la matriz con w en la columna 6 y ceros en las demás. La matriz P' es de nuevo estocástica por columnas, y todo lo afirmado en el teorema para P vale para P'. Si además P' es irreducible, tiene sentido calcular su PageRank; aquí lo es, porque la columna 6 conecta el nodo 6 con todos los demás y el resto ya estaba conectado. Resolviendo $P'x = x$ como en el ejercicio anterior sale

$$x_P = \frac{1}{1600}\,(231,\ 319,\ 306,\ 294,\ 264,\ 186)^t \approx (0.1444,\ 0.1994,\ 0.1913,\ 0.1838,\ 0.1650,\ 0.1163)^t,$$

que se comprueba en una ecuación cualquiera: la sexta da $\frac{1}{3}\cdot 294 + \frac{1}{3}\cdot 264 = 98 + 88 = 186$. La página 2 sigue siendo la primera, y la 6, que no vota a nadie pero recibe votos, queda la última.

EL CAMINO QUE NO SE TOMÓ: QUITAR LOS NODOS SIN SALIDA

La alternativa. Eliminar de la red las páginas sin enlaces salientes antes de calcular, y asignarles después una relevancia a partir de las que las enlazan. Es lo que hicieron Page y Brin en sus primeras implementaciones, en las que los enlaces colgantes se retiraban durante el cálculo y se reincorporaban al final.
Su ventaja. La matriz que queda es estocástica sin retoques y más pequeña.
Qué la aparta. Quitar una página sin salida puede dejar sin salida a otras (las que solo enlazaban a ella), y hay que repetir el proceso; y en la web real las páginas sin enlaces salientes son muchísimas (ficheros PDF, imágenes, páginas aún no rastreadas). La corrección $P + wd^t$ las mantiene en el modelo con una regla clara, permite elegir a dónde saltar desde ellas, y no cambia el coste del cálculo, porque wd^t es de rango uno y se aplica a un vector con un producto escalar.

El factor de amortiguación. Demos un paso más en la deducción del modelo. Incluso si todas las páginas tuvieran enlaces, un internauta podría decidir libremente moverse a otra página sin usar los enlaces de la actual, por ejemplo escribiendo una dirección en el navegador. Este comportamiento se tiene en cuenta introduciendo el **vector de personalización** v, estocástico, y el **factor de amortiguación** $\alpha \in [0, 1]$. Se llega a la **matriz de Google**

$$\mathcal{P} = \alpha P' + (1-\alpha)\, v\, e^t = \alpha\big[P + w\, d^t\big] + (1-\alpha)\, v\, e^t, \qquad \alpha \in [0, 1].$$

El factor α indica que la probabilidad de que un internauta deje de seguir los enlaces y se mueva a otra página es $1 - \alpha$: con probabilidad α sigue un enlace de la página actual (columna de P') y con probabilidad $1 - \alpha$ salta a una página elegida según v. En sus trabajos originales Brin y Page usaron $\alpha = 0.85$, y en la mayoría de los artículos sobre PageRank α oscila entre 0.85 y 0.99.

Sobre la notación: es frecuente escribir la matriz de Google con la misma letra P que la de transición, distinguida tipográficamente. Este libro la escribe $\mathcal{P}$ para que no se confundan en una línea donde aparecen las dos, y advierte aquí que es la misma matriz.

Observación. Para cualquier $\alpha \in [0, 1]$, la matriz $\mathcal{P}$ sigue siendo estocástica por columnas. Sus componentes son no negativas, por ser suma de no negativos con coeficientes no negativos; y la columna j suma $\alpha \cdot 1 + (1-\alpha)\cdot 1 = 1$, porque las columnas de P' suman 1 y las de ve^t son todas iguales a v, que suma 1. Por tanto, una vez más, $\rho(\mathcal{P}) = 1$ y 1 es autovalor dominante. El valor de α es importante desde el punto de vista de la rapidez de convergencia del algoritmo, como se verá.

Hay diversas elecciones para v. Con $v = \frac{1}{n}e$, los internautas eligen al azar una página nueva cuando deciden no seguir los enlaces. Con $v = e_1 = (1, 0, \ldots, 0)^t$, cuando un internauta abandona la estructura de enlaces salta siempre a la página 1: es una forma de favorecer páginas concretas.

EL CAMINO QUE NO SE TOMÓ: SIN AMORTIGUACIÓN, $\alpha = 1$

La alternativa. Quedarse con P', es decir, tomar $\alpha = 1$: el internauta solo sigue enlaces.
Su ventaja. Es el modelo más fiel a la idea de «un enlace es un voto», sin ningún parámetro arbitrario que fijar.
Qué la hunde. La web real no es fuertemente conexa: hay grupos de páginas que solo se enlazan entre sí y que reciben enlaces del exterior sin devolverlos. En términos del teorema de Frobenius, P' puede ser reducible; si hay varias clases cerradas, el autovector estocástico no es único y el resultado puede depender del punto de partida; en términos del internauta, toda la probabilidad acaba atrapada en esos grupos y el resto de la web recibe relevancia cero. Además, aunque P' fuera irreducible, podría no ser primitiva, y entonces el método de la potencia oscilaría sin converger. Con $v_i > 0$ para todo i y $0 < \alpha < 1$, la mezcla con ve^t hace a $\mathcal{P}$ estrictamente positiva y arregla de golpe las dos cosas. Un parámetro arbitrario, α, a cambio de existencia, unicidad y convergencia: es un buen trato.

La escala de la web obliga a pensar en matrices de talla enorme, y adquiere gran importancia la elección de un buen método para calcular el autovector estocástico asociado a 1. El algoritmo original de Brin y Page se basa en el **método de la potencia**; después se propusieron otros basados en métodos iterados de resolución de sistemas lineales de gran talla.

EL CAMINO QUE NO SE TOMÓ: RESOLVER UN SISTEMA LINEAL EN VEZ DE ITERAR

La alternativa. Como $e^t x = 1$ para x estocástico, la ecuación $\mathcal{P}x = x$ equivale a $\alpha P'x + (1 - \alpha)v = x$, es decir, al sistema lineal $(I - \alpha P')\, x = (1 - \alpha)\, v$, cuya matriz es invertible para $\alpha < 1$ porque $\rho(\alpha P') = \alpha < 1$. Se puede resolver con cualquier método iterativo para sistemas grandes y dispersos.
Su ventaja. Los métodos de Krylov y los precondicionados convergen a menudo en menos iteraciones que la potencia, y la formulación como sistema permite reutilizar toda la maquinaria del álgebra lineal numérica.
Por qué se explica aquí la potencia. Porque cada iteración es un producto de una matriz dispersa por un vector, sin almacenar nada más, y porque su convergencia está garantizada por el teorema de Perron con una tasa explícita. Es el algoritmo original y el que pide la hoja de prácticas. La alternativa se usa, y se dice.

1.2.2 Cálculo del valor de PageRank mediante el método de la potencia

LA PREGUNTA

La matriz de Google de la web entera tiene miles de millones de filas y, con la amortiguación, ninguna entrada nula: no cabe en ningún ordenador. Sin embargo, Google recalcula el PageRank continuamente. ¿Cómo se multiplica un vector por una matriz que no se puede escribir?

Tomaremos $\alpha \in (0, 1)$ y, por simplicidad, w y v estándar:

$$w = v = \frac{1}{n}e = (1/n, \ldots, 1/n)^t.$$

En este caso la matriz de Google $\mathcal{P}$, a diferencia de P', tiene **todas sus componentes positivas**: cada entrada es al menos $(1-\alpha)/n > 0$. Se le aplica el teorema de Perron, en su forma original, y se deduce que 1 es un autovalor simple estrictamente dominante y que existe un único autovector estocástico asociado, que se denota x_∞ para distinguirlo de x_P, caracterizado por

$$x_\infty = \mathcal{P}x_\infty = \left[\alpha(P + v\,d^t) + (1-\alpha)\,v\,e^t\right]x_\infty, \qquad x_{\infty,i} \geq 0 \ \ \forall i, \qquad \sum_{i=1}^{n} x_{\infty,i} = 1.$$

Además, el método de la potencia converge, y el algoritmo que resulta es

$$\begin{cases} \text{elegir } x^0 \text{ estocástico, por ejemplo } x^0 = \frac{1}{n}e, \\ \text{calcular } x^{k+1} = \mathcal{P}x^k \text{ para } k \geq 0. \end{cases}$$

Sea cual sea $\alpha \in (0,1)$, $x^k \to x_\infty$ cuando $k \to +\infty$. Pero estos cálculos no pueden hacerse de forma directa: en la práctica $\mathcal{P}$ es una matriz **llena** y de talla enorme. Lo que sigue muestra cómo realizar las iteraciones sin construir $\mathcal{P}$, usando solo la matriz P, que es **hueca**: casi todas sus entradas son cero, y solo se almacenan las que no lo son.

LA IDEA: LA MATRIZ LLENA ES LA SUMA DE UNA HUECA Y DOS DE RANGO UNO

$\mathcal{P} = \alpha P + \alpha v d^t + (1-\alpha)ve^t$. El primer sumando es hueco. Los otros dos son matrices de rango como mucho uno, y multiplicar una matriz de rango uno vz^t por un vector x cuesta un producto escalar: $(vz^t)x = (z^t x)\,v$. Así que $\mathcal{P}x$ se calcula con un producto por la matriz hueca y dos productos escalares. Y todavía mejor: los dos productos escalares se pueden obtener de la norma uno de αPx, sin ni siquiera formar d. Esa es la observación que sigue.

Comenzamos con dos observaciones.

- Las matrices estocásticas por columnas conservan la norma uno de los vectores no negativos, por el apartado 2 del teorema. En consecuencia, todos los x^k obtenidos por el método de la potencia verifican $\|x^k\|_1 = 1$.
- La matriz $P' = P + wd^t$ es estocástica. Luego, si x es un vector estocástico, y con $w = v$,

$$1 = \|x\|_1 = \|(P + v\,d^t)x\|_1 = \|Px + (v\,d^t)x\|_1 = \sum_{i=1}^{n} \left|(Px)_i + ((v\,d^t)x)_i\right|$$
$$= \sum_{i=1}^{n} (Px)_i + \sum_{i=1}^{n} ((v\,d^t)x)_i = \|Px\|_1 + \|(v\,d^t)x\|_1,$$

donde se han quitado los valores absolutos porque todos los sumandos son no negativos. Pero $d^t x$ es un escalar no negativo y $\|v\|_1 = 1$, así que

$$\|(v\,d^t)x\|_1 = \|(d^t x)\,v\|_1 = (d^t x)\,\|v\|_1 = d^t x,$$

de donde, finalmente,

$$d^t x = 1 - \|Px\|_1.$$

Esta última igualdad es la clave: $d^t x$ es la probabilidad de estar en una página sin salida, y se obtiene sin conocer d, como lo que le falta a $\|Px\|_1$ para llegar a 1. Conocido x^k, y usando

que $e^t x^k = \sum_i x_i^k = \|x^k\|_1 = 1$,

$$\begin{aligned} x^{k+1} &= \big[\alpha(P + v\,d^t) + (1-\alpha)\,v\,e^t\big]x^k = \alpha(P + v\,d^t)x^k + (1-\alpha)\,v\,(e^t x^k) \\ &= \alpha P x^k + \alpha\,v\,(d^t x^k) + (1-\alpha)\,v = \alpha P x^k + \underbrace{\big[\alpha(d^t x^k) + (1-\alpha)\big]}_{\gamma_k}\,v = \alpha P x^k + \gamma_k\,v. \end{aligned}$$

Por último,

$$\gamma_k = \alpha(d^t x^k) + (1-\alpha) = \alpha\big[1 - \|Px^k\|_1\big] + (1-\alpha) = 1 - \alpha\|Px^k\|_1 = 1 - \|\alpha P x^k\|_1.$$

Leída en palabras, la fórmula $x^{k+1} = \alpha P x^k + \gamma_k v$ separa a los internautas en dos grupos. Los que siguen un enlace de verdad llevan la probabilidad $\alpha P x^k$. Todos los demás —los que deciden saltar y los que estaban en una página sin salida— suman γ_k y se reparten según v. Como la probabilidad total es uno, γ_k es lo que le falta a $\|\alpha P x^k\|_1$ para llegar a uno, y por eso no hace falta conocer d. La consecuencia es la reformulación del método de la potencia que sigue, y que es la que programa la práctica 5.

Algoritmo de la potencia para el cálculo del valor PageRank.

(a) Iniciar con $x^0 = \frac{1}{n}e$ (por ejemplo); tomar $k = 0$.

(b) Dados $k \geq 0$ y x^k, calcular

$$\begin{aligned} y^{k+1} &= \alpha P x^k \qquad (P \text{ es una matriz hueca}), \\ \gamma &= \|x^k\|_1 - \|y^{k+1}\|_1 = 1 - \|y^{k+1}\|_1, \\ x^{k+1} &= y^{k+1} + \gamma\,v, \\ \delta &= \|x^{k+1} - x^k\|_1. \end{aligned}$$

Si $\delta \leq \varepsilon$, parar; si no, cambiar k por $k+1$ y repetir (b).

EXPLÍCATE. *En el paso (b), γ se calcula como $1 - \|y^{k+1}\|_1$. Di qué dos cantidades del modelo se han sumado dentro de ese γ: la probabilidad de salir de una página sin enlaces y la de saltar por voluntad propia.*

Ejemplo. Una iteración a mano, en una red de tres páginas con un nodo sin salida. La 1 enlaza a la 2 y a la 3, la 2 enlaza a la 3, y la 3 no enlaza a ninguna. Entonces

$$P = \begin{pmatrix} 0 & 0 & 0 \\ 1/2 & 0 & 0 \\ 1/2 & 1 & 0 \end{pmatrix}, \qquad d = (0,0,1)^t,$$

con la tercera columna nula. Con $\alpha = 0.85$ y $v = x^0 = \frac{1}{3}e$:

- $Px^0 = (0,\, 1/6,\, 1/2)^t$, de norma uno $2/3$. Falta $1/3$, que es exactamente $x_3^0 = d^t x^0$: la probabilidad que estaba en la página 3 y que P no sabe adónde mandar. Es la identidad $d^t x = 1 - \|Px\|_1$ con números.
- $y^1 = \alpha P x^0 = (0,\, 17/120,\, 17/40)^t \approx (0,\, 0.1417,\, 0.4250)^t$, de norma uno $17/30 \approx 0.5667$.
- $\gamma = 1 - 17/30 = 13/30 \approx 0.4333$. Y se descompone en los dos sumandos de la pregunta anterior: $\alpha\,d^t x^0 = 0.85 \cdot \frac{1}{3} \approx 0.2833$ es la masa del nodo sin salida, que sigue un enlace artificial, y $1 - \alpha = 0.15$ es la de los internautas que saltan por voluntad propia; $0.2833 + 0.15 = 0.4333$.

▪ $x^1 = y^1 + \gamma v = (13/90,\ 103/360,\ 41/72)^t \approx (0.1444,\ 0.2861,\ 0.5694)^t$, que suma 1.

Formando la matriz de Google entera, $\mathcal{P}x^0$ da el mismo vector, pero exige escribir sus nueve entradas, ya sin ningún cero; el algoritmo solo ha usado las tres entradas no nulas de P y dos sumas. Iterando hasta el final, el PageRank de esta red es $(800,\ 1140,\ 2109)^t/4049 \approx (0.1976,\ 0.2816,\ 0.5209)^t$: la página 3, a la que enlazan las otras dos, se lleva más de la mitad, aunque no vote a ninguna.

Observación. Se pueden repetir los argumentos con otras elecciones de w y v. Por ejemplo, con $w = \frac{1}{n}e$ y $v = (v_1, \ldots, v_n)^t$ con los $v_i > 0$ arbitrarios de suma 1. La matriz $\mathcal{P}$ sigue siendo estrictamente positiva, porque $(1-\alpha)v_i > 0$, y el algoritmo solo cambia en que el término de rango uno de los nodos sin salida usa w y el de los saltos usa v. La identidad $d^t x^k = 1 - \|Px^k\|_1$ sigue valiendo, pues solo usa que w es estocástico. Tampoco ahora hace falta almacenar d:

$$x^{k+1} = \alpha P x^k + \alpha\big(1 - \|Px^k\|_1\big)w + (1-\alpha)v.$$

Proposición (velocidad de convergencia)**.** Da contenido preciso a la idea de que α controla la rapidez del algoritmo. Sea $\mathcal{P} = \alpha P' + (1-\alpha)ve^t$ con P' estocástica por columnas, v estocástico y $0 < \alpha < 1$, y sea x^0 estocástico. Existe un único punto fijo estocástico x_∞, incluso si v tiene ceros, y

$$\|x^k - x_\infty\|_1 \le \alpha^k \, \|x^0 - x_\infty\|_1 \qquad \text{para todo } k \ge 0.$$

La idea, en una frase. El error $x^k - x_\infty$ es la diferencia de dos repartos que suman lo mismo, así que *suma cero*: es un vector de excesos y defectos que se compensan. Y un vector de suma cero es invisible para el término de salto aleatorio, que solo sabe leer la masa total. De modo que sobre el error actúa nada más que $\alpha P'$, y P' no agranda la norma uno: el error queda multiplicado, como mucho, por α en cada paso. La convergencia geométrica no es un accidente numérico, es la vida media del recuerdo del punto de partida.

Paso 1: el error tiene suma cero, y eso apaga el término de rango uno (el núcleo). Como x^k y x_∞ son estocásticos, $e^t(x^k - x_\infty) = 1 - 1 = 0$, luego

$$x^{k+1} - x_\infty = \mathcal{P}x^k - \mathcal{P}x_\infty = \alpha P'(x^k - x_\infty) + (1-\alpha)\, v\, e^t(x^k - x_\infty) = \alpha P'(x^k - x_\infty).$$

El sumando $(1-\alpha)\, v\, e^t(x^k - x_\infty)$ se ha anulado entero: da igual qué vector de personalización v se elija, porque lo único que ese término mira del error es su suma, y su suma es cero.

Paso 2: una matriz estocástica no agranda la norma uno. Para $z \in \mathbb{R}^n$,

$$\|P'z\|_1 = \sum_i \Big|\sum_j p'_{ij} z_j\Big| \le \sum_i \sum_j p'_{ij}|z_j| = \sum_j \Big(\sum_i p'_{ij}\Big)|z_j| = \sum_j |z_j| = \|z\|_1.$$

donde la desigualdad es la triangular y las dos igualdades siguientes son el intercambio del orden de dos sumas finitas y la estocasticidad por columnas de P'.

Paso 3: iterar. Juntando los dos pasos, $\|x^{k+1} - x_\infty\|_1 \le \alpha\|x^k - x_\infty\|_1$, y por inducción sale la cota. Para justificar también la existencia y unicidad cuando v tiene ceros, los mismos dos pasos aplicados a dos repartos cualesquiera x, y dan $\|\mathcal{P}x - \mathcal{P}y\|_1 \le \alpha\|x - y\|_1$. Es una contracción del símplex cerrado en sí mismo; el teorema del punto fijo de Banach da el único x_∞. ■

Qué hace cada hipótesis. La condición $\alpha < 1$ es la única que hace falta para la contracción, y se ve dónde: es el factor que multiplica al P' del paso 1. Con $\alpha = 1$ el

argumento sigue siendo válido, pero la cota se convierte en $\|x^k - x_\infty\|_1 \leq \|x^0 - x_\infty\|_1$, que no dice nada —y en efecto, sin amortiguación la iteración puede oscilar sin converger—. Que x^0 sea estocástico entra solo en el paso 1, para que el error tenga suma cero; con un x^0 cualquiera la parte del error que no suma cero sobrevive y el argumento se cae. Obsérvese, por último, que la cota es de las que se pueden usar: es *a priori*, no depende de la red, y permite fijar de antemano cuántas iteraciones hacen falta para una precisión dada.

Observación (qué garantiza la parada)**.** En el algoritmo, $\delta = \|x^{k+1} - x^k\|_1$ mide el cambio entre iterados, no el error respecto del límite. Sumando la cola geométrica de cambios sucesivos se obtiene

$$\|x^{k+1} - x_\infty\|_1 \leq \frac{\alpha}{1-\alpha}\,\delta.$$

Por tanto, para garantizar error como mucho η en el vector devuelto, basta parar cuando $\delta \leq (1-\alpha)\eta/\alpha$. Esto traduce una tolerancia computable en una precisión del resultado.

Con $\alpha = 0.85$, el error se divide al menos por $0.85^{-1} \approx 1.18$ en cada paso y se reduce a la milésima parte en unas 43 iteraciones; con $\alpha = 0.99$ hacen falta unas 688 para garantizar esa reducción con la cota. Cuanto mayor es α, más conservadora resulta esta garantía de convergencia; la rapidez observada depende también de la red y del vector inicial. Es una de las razones para equilibrar el peso de los enlaces y el del salto aleatorio al elegir α.

DE DÓNDE VIENE: STANFORD, 1998

Sergey Brin y Lawrence Page describieron el algoritmo en 1998, en el artículo que presentaba el prototipo de su buscador, y en un informe técnico de Stanford con Motwani y Winograd que le dio nombre. La idea de medir importancia a partir de una red tenía antecedentes: Katz propuso en 1953 sumar la influencia de caminos de distintas longitudes con pesos decrecientes; no era simplemente calcular un autovector.[a] Pinski y Narin emplearon en 1976 un modelo de influencia para ordenar revistas por citas. Lo que añadieron Brin y Page fue el modelo del internauta aleatorio con amortiguación y saltos uniformes, que hace a la matriz estrictamente positiva y al problema resoluble a la escala de la web, y la decisión de programar el método de la potencia sin formar la matriz. Larry Page patentó el método a nombre de la Universidad de Stanford.

[a] L. Katz, *A New Status Index Derived from Sociometric Analysis*, Psychometrika 18 (1953), 39–43, `https://doi.org/10.1007/BF02289026`.

Ejercicio 1.3. *La matriz de Google de la red de cinco páginas*

ORIGEN: construido a partir del ejemplo anterior, para ver actuar el factor de amortiguación.

Para la red de la figura 1.2, con $\alpha = 0.85$ y $w = v = \frac{1}{5}e$:

1. escribir la matriz de Google $\mathcal{P}$ y comprobar que es estrictamente positiva y estocástica por columnas;
2. hacer a mano una iteración del algoritmo de la potencia desde $x^0 = \frac{1}{5}e$;
3. sabiendo que la iteración converge a $x_\infty \approx (0.1762,\ 0.2459,\ 0.2094,\ 0.2043,\ 0.1642)^t$, comparar con x_P del ejercicio del PageRank de cinco páginas y explicar en qué dirección ha

cambiado cada componente.

PISTAS, DE MENOR A MAYOR

1. *En esta red no hay nodos sin salida, luego* $d = 0$ *y* $\mathcal{P} = \alpha P + (1-\alpha)ve^t$.
2. *Cada entrada de* $\mathcal{P}$ *es* $0.85\,p_{ij} + 0.03$.
3. $\gamma = 1 - \|\alpha P x^0\|_1$, *y como* P *es estocástica y* x^0 *también,* $\|Px^0\|_1 = 1$.

Resolución.

La idea, y explica de un tirón los tres apartados. La matriz de Google no es una matriz nueva: es la de siempre **mezclada con la matriz de la ignorancia total**. Con probabilidad α el internauta sigue un enlace —eso es P— y con probabilidad $1-\alpha$ se olvida de todo y salta a una página al azar —eso es $\frac{1}{n}ee^t$, donde todas las entradas son iguales. La fórmula $\mathcal{P} = \alpha P + (1-\alpha)\frac{1}{n}ee^t$ es literalmente esa frase.

La mezcla explica las dos primeras respuestas y orienta la comparación de la tercera:

- *Por qué es estrictamente positiva.* Porque el segundo sumando lo es. Da igual cuántos ceros tenga P: al añadir 0.03 a cada entrada, ninguna se queda en cero. Y esa es toda la razón de existir del factor de amortiguación: **convierte una matriz con ceros, a la que Perron no se aplica, en una estrictamente positiva, a la que sí**. No es un truco numérico, es lo que hace que el problema tenga respuesta única.
- *Por qué sigue siendo estocástica.* Porque es una *media ponderada* de dos matrices estocásticas, con pesos α y $1-\alpha$ que suman uno; y una combinación convexa de matrices con columnas de suma 1 vuelve a tener columnas de suma 1.
- *Hacia dónde se mueve el resultado.* El salto uniforme introduce probabilidad en todas las páginas, incluidas las menos visitadas. En esta red veremos que todas las componentes se acercan a $1/5$, pero hay que comprobarlo: no es una regla general para cada componente, porque los pasos posteriores vuelven a distribuir esa masa por los enlaces.

Paso 1: la matriz. Como $(1-\alpha)/n = 0.15/5 = 0.03$, cada entrada de $\mathcal{P}$ es $0.85p_{ij} + 0.03$:

$$\mathcal{P} \approx \begin{pmatrix} 0.03 & 0.03 & 0.3133 & 0.455 & 0.03 \\ 0.03 & 0.03 & 0.3133 & 0.455 & 0.455 \\ 0.455 & 0.455 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.455 & 0.03 & 0.03 & 0.455 \\ 0.455 & 0.03 & 0.3133 & 0.03 & 0.03 \end{pmatrix},$$

donde 0.3133 aproxima $0.85/3 + 0.03$. Antes del redondeo, todas las entradas son al menos $0.03 > 0$. La columna 1 suma $0.03 + 0.03 + 0.455 + 0.03 + 0.455 = 1$, y lo mismo las demás, porque cada una es 0.85 por una columna de P, que suma 1, más cinco veces 0.03.

Paso 2 (el núcleo): una iteración a mano, sin construir $\mathcal{P}$. Conviene fijarse en que la cuenta que sigue **no usa la matriz recién escrita**: usa P y una constante. Ese es el motivo de que el algoritmo funcione con redes de mil millones de páginas, donde $\mathcal{P}$ no cabe en ninguna memoria pero P, que está llena de ceros, sí. $x^0 = (0.2, 0.2, 0.2, 0.2, 0.2)^t$. Primero $Px^0 = 0.2\,(Pe)$, y Pe es el vector de sumas de las filas de P: $(5/6,\ 4/3,\ 1,\ 1,\ 5/6)^t$. Luego $y^1 = \alpha Px^0 = 0.17\,(5/6, 4/3, 1, 1, 5/6)^t \approx (0.1417, 0.2267, 0.17, 0.17, 0.1417)^t$. Su norma uno es $0.17 \cdot 5 = 0.85$, así que $\gamma = 1 - 0.85 = 0.15$, y

$$x^1 = y^1 + 0.15 \cdot \tfrac{1}{5}e \approx (0.1717,\ 0.2567,\ 0.2000,\ 0.2000,\ 0.1717)^t,$$

que suma 1 antes del redondeo. La 2 ya aparece primera, pero aún hay empates: 3 y 4 comparten el segundo puesto, y 1 y 5 el último. El orden final exige seguir iterando.
Paso 3: comparar los resultados. Comparando con $x_P = (0.1732, 0.2520, 0.2126, 0.2047, 0.1575)^t$: las páginas con más relevancia (2, 3, 4) la pierden un poco y las de menos (1, 5) la ganan. Es el efecto del salto aleatorio: con probabilidad 0.15 el internauta olvida los enlaces y se reparte uniformemente; en esta red, el nuevo equilibrio acerca todas las componentes a $1/5 = 0.2$. El orden no cambia en este caso, pero no tiene por qué conservarse en general: α forma parte del modelo.

Qué hemos aprendido. La amortiguación hace dos cosas a la vez. Convierte una matriz con ceros en una estrictamente positiva, con lo que Perron garantiza un PageRank único y la convergencia desde cualquier reparto inicial. Y acelera esa convergencia: los autovalores de $\mathcal{P}$ distintos de 1 son los de P multiplicados por α, y aquí el segundo en módulo pasa de -0.765 a -0.650. El precio es que el resultado deja de depender solo de la red: depende también de α y de v.

Dónde se falla.

- *Al rellenar solo los ceros.* Poner 0.03 en las entradas nulas de P y dejar las demás como estaban no da $\mathcal{P}$: las columnas pasan a sumar 1.09 o 1.06. Hay que multiplicar toda P por α y sumar $(1-\alpha)/n$ a todas las entradas.
- *Al olvidar el factor α.* Sumar 0.03 a todas las entradas de P sin multiplicarla por 0.85 da columnas que suman 1.15.
- *Al esperar que γ cambie.* Sin nodos sin salida, $\|\alpha P x\|_1 = \alpha$ para todo x estocástico y γ vale siempre 0.15. Si un programa da otro valor, P está mal normalizada.
- *Al leer un orden en x^1.* Tras una iteración quedan empates que solo se deshacen iterando.

1.3 Prácticas de MATLAB del tema 1

La pregunta

Todo lo anterior se puede hacer a mano con cuatro clases de edad o con cinco páginas. Una población real tiene cien clases y la web tiene miles de millones de páginas; la matriz de transición de diez mil páginas, guardada entera, ocupa ya ochocientos megabytes de los que casi todas son ceros si cada página enlaza solo a unas pocas. ¿Qué hay que saber hacer con un ordenador para que el modelo sirva a esa escala, y qué parte de la teoría es la que dice que el cálculo va a terminar?

Esta sección resuelve la hoja de ejercicios del tema 1, que tiene cinco ejercicios, más los dos ejercicios complementarios del repaso de MATLAB (un 1.1 sobre la matriz de Vandermonde y un 3.1 sobre máximos, mínimos y medias). Los tres primeros son de manejo de datos: interpolación y ajuste, que no aparecen en la teoría pero sí en la hoja, y que son la forma de *obtener* los coeficientes de un modelo a partir de observaciones. Los dos últimos programan Leslie y PageRank. Cada ejercicio lleva el código MATLAB completo, la salida esperada, la figura, y la versión en Python. Todas las órdenes que se usan están en la sesión cero o se presentan aquí la primera vez que aparecen. Cada ejercicio va en su carpeta, con el fichero de datos si lo hay, y esa carpeta se selecciona como carpeta actual antes de ejecutar.

Ejercicio 1: interpolación de Lagrange y lineal a trozos

El libro previo del autor resolvió este ejercicio en diálogo, y aquí se integra tal cual: sus intervenciones en cursiva, y las respuestas, corregidas donde hacía falta.

Ejercicio 1.4. *Interpolación de Lagrange y lineal a trozos*

ORIGEN: hoja de ejercicios del tema 1, ejercicio 1.

Se consideran los puntos del plano de abscisas $x = (0, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15)$ y ordenadas $y = (10, 20, 30, -10, 10, 10, 10.5, 15, 50, 60, 85)$. Se pide:

a) calcular los coeficientes del polinomio de interpolación de Lagrange que pasa por estos puntos y, con ellos, construir una función anónima que lo evalúe en cualquier punto (órdenes `polyfit` y `polyval`);
b) construir otra función anónima que evalúe el polinomio de interpolación lineal a trozos que pasa por estos puntos en cualquier $x \in [0, 15]$ (orden `interp1`);
c) evaluar las dos funciones anteriores en $x = 1$;
d) dibujar juntas las gráficas de ambos interpolantes, con etiquetas y leyenda.

> *«Refresco cómo se saca el polinomio de Lagrange. Mi idea: $p(x) = \sum_i y_i L_i(x)$, con $L_i(x) = \prod_{j \neq i}(x - x_j) / \prod_{j \neq i}(x_i - x_j)$; así $p(x_i) = y_i$ y es de grado $n - 1$, luego debería valer. Había también un método de Newton; en ambos casos obtengo el mismo polinomio. Entiendo que "polinomio de interpolación de Lagrange" es básicamente la interpolación clásica de n nodos con grado $n - 1$. ¿Y la de Hermite, que tiene en cuenta las derivadas, daría $2n - 2$ grados de libertad? Tengo la idea de que esta es la interpolación más rudimentaria y que empezaba a fallar para soportes de muchos puntos ("overfitting"?). Para mí, `polyfit` toma mis dos vectores y devuelve el vector de coeficientes (¿de menor a mayor grado o al revés, y por qué?), y `polyval` toma ese vector y los puntos donde evaluar. Cuando me piden "construir una función anónima", ¿se refieren a algo como `f = @(x) polyval(a,x)`, no a reprogramar `polyval`, verdad?»*

Resolución.

La idea, y aquí la decisión de método es todo el ejercicio. No se pide calcular una curva: se pide calcular **dos**, la global y la local, y ponerlas al lado. Las dos pasan por los once puntos, las dos resuelven literalmente el problema planteado, y sin embargo en $x = 1$ una vale 15 y la otra -245.79. Ese contraste es el contenido del ejercicio, y la pregunta que hay que llevarse no es «¿cómo se hace?» sino **«¿qué le estoy pidiendo a los datos entre los nodos?»**: el polinomio global exige una única fórmula suave para todo el intervalo y paga ese lujo con oscilaciones; el interpolante a trozos no exige nada más allá de cada tramo y a cambio, en general, no es derivable en los nodos. Elegir uno u otro es una decisión de modelización, no de programación.

El núcleo del programa son dos líneas —`c = polyfit(x, y, numel(x)-1)` y `inlineal = @(t) interp1(x, y, t, 'linear')`—, una por método. Todo lo demás envuelve, evalúa y dibuja.

Las dos fórmulas dan el mismo polinomio. La forma de Lagrange y la de Newton no son dos polinomios distintos: por n puntos de abscisas distintas pasa un único polinomio de grado $\leq n - 1$, y Lagrange, Newton y Vandermonde (el del ejercicio 1.1) son tres maneras de

---

escribirlo. Los L_i cumplen $L_i(x_k) = \delta_{ik}$, de donde $p(x_i) = y_i$.
Hermite, con un ajuste. Interpola valores y derivadas: con n nodos, valor y primera derivada en cada uno son $2n$ condiciones, que determinan un polinomio de grado como mucho $2n-1$: hay $2n$ coeficientes, no $2n-2$.
Oscilaciones y sobreajuste son problemas distintos. Si los datos contienen ruido, interpolarlos exactamente puede producir sobreajuste: el polinomio reproduce también ese ruido. Incluso sin ruido pueden aparecer oscilaciones grandes. El fenómeno de Runge muestra que, para ciertas funciones, aumentar el grado con nodos equiespaciados puede empeorar la aproximación cerca de los extremos. No ocurre con toda función, ni podemos diagnosticar ruido en estos once datos, cuyas abscisas no son exactamente equiespaciadas. Si $f \in C^{n+1}$ en un intervalo que contiene los nodos y el punto x, el resto es

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^{n} (x - x_i)$$

para algún ξ en ese intervalo. Importan tanto las derivadas de f como la distribución de los nodos. Interpolar a trozos evita usar un polinomio global de grado alto; si se pueden elegir los nodos, los de Chebyshev ayudan a controlar las oscilaciones.
`polyfit` y `polyval`. `polyfit(x,y,d)` ajusta por mínimos cuadrados un polinomio de grado `d`; cuando `d` es el número de puntos menos uno, el ajuste pasa por todos y *es* el interpolante. Devuelve los coeficientes en orden decreciente de potencia, $[a_d, \ldots, a_0]$, porque así se escribe $p(x) = a_d x^d + \cdots + a_0$ y así los espera `polyval`. La función anónima es justo envolver `polyval`: `p = @(t) polyval(c,t)`.

> *«¿Qué hace exactamente `interp1`, a ver si me entero? Y ¿por qué se llama `interp1`? Si fuera solo con rectas lo entendería, pero para el `spline`, ¿por qué "1"?»*

`interp1`. Recibe los datos (x, y) y unos puntos de consulta t, y devuelve los valores interpolados en t. Con el método por defecto, `'linear'`, busca para cada t el tramo $x_i \le t \le x_{i+1}$ en que cae y evalúa la recta que une (x_i, y_i) con (x_{i+1}, y_{i+1}): es unir los puntos con segmentos. Las abscisas deben ser distintas; ordenarlas facilita leer los tramos. Con el método `'linear'` usado aquí, fuera de $[x_{\text{mín}}, x_{\text{máx}}]$ se devuelve `NaN`, salvo que se indique otra regla de extrapolación. El `1` es la dimensión de la variable independiente: se interpola $y = f(x)$ con una sola entrada; `interp2` es para superficies $z = f(x, y)$. El spline sigue siendo `interp1` porque sigue siendo función de una variable; el nombre es correcto. Otros métodos: `'nearest'`, `'pchip'` y `'spline'`. Y sobre el spline cúbico: con parábolas no se llega a C^2, porque la segunda derivada de una parábola es constante y pedir continuidad de curvatura las colapsaría en una sola; la cuenta de grados de libertad cierra justa con cúbicas, $4m$ coeficientes contra $4m-2$ condiciones, y los dos que sobran los fijan las condiciones de contorno.
El código. Es el del libro previo, con el añadido de la ampliación del ejercicio 1bis, que compara además con los ajustes de grado 2 y 3. El fichero se guarda como `solucion_interpolacion.m`.

```
%% Ejercicio 1 - Interpolacion (Lagrange y lineal a trozos)
% --- Datos ---
x = [0, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15];
y = [10, 20, 30, -10, 10, 10, 10.5, 15, 50, 60, 85];
```

```
5
6  % --- (a) Polinomio de Lagrange (polyfit / polyval) ---
7  c = polyfit(x, y, numel(x) - 1); % grado 10: pasa por los 11 puntos
8  plagrange = @(t) polyval(c, t);
9
10 % --- (b) Interpolante lineal a trozos (interp1) ---
11 inlineal = @(t) interp1(x, y, t, 'linear');
12
13 % --- (c) Evaluacion en x = 1 ---
14 disp('VALORES EN x = 1 (LAGRANGE Y LINEAL A TROZOS):')
15 disp([plagrange(1), inlineal(1)])
16
17 % --- (d) Grafica comparativa en [0, 15] ---
18 t = linspace(0, 15); % rejilla fina (100 puntos)
19 figure
20 plot(t, plagrange(t), 'b')
21 hold on
22 plot(t, inlineal(t), 'r')
23 plot(x, y, 'ko') % los datos, como circulos negros
24 hold off
25 legend('Lagrange (grado 10)', 'Lineal a trozos', 'Datos', 'Location', 'best')
26 xlabel('x'), ylabel('y')
27 title('Interpolacion: Lagrange frente a lineal a trozos')
28 grid on
```

SALIDA ESPERADA

```
VALORES EN x = 1 (LAGRANGE Y LINEAL A TROZOS):
 -245.7853 15.0000
```

MATLAB avisa además, al ejecutar `polyfit`, de que el polinomio está mal condicionado (*Polynomial is badly conditioned*): con grado 10 la matriz del ajuste tiene un número de condición enorme. El resultado es correcto, pero es delicado, y es justo lo que el ejercicio 1.1 va a poner de manifiesto.

El control, y hay uno para cada método. El valor 15 del interpolante lineal se comprueba a mano: $x = 1$ está entre $(0, 10)$ y $(2, 20)$, y la recta que los une pasa por el punto medio. El valor de Lagrange se puede comprobar con la fórmula de productos de la intervención inicial, que da $-245.785322\ldots$. Otro control es exigirle la propiedad que lo define: `max(abs(plagrange(x) - y))` tiene que ser cero, porque el polinomio pasa por los once puntos. Sale $5.2 \cdot 10^{-8}$, y ese número dice ya lo que va a decir el ejercicio siguiente: el error de redondeo se amplifica al calcular y evaluar los coeficientes. Un residuo absoluto no permite contar sin más cuántas cifras correctas quedan. Este control, junto con la evaluación independiente anterior, ayuda a separar «el polinomio está mal calculado» de «el polinomio está bien calculado y hace cosas raras entre los nodos», que es justo lo que ocurre aquí: -245.8 se aleja 260 unidades de los datos que tiene a un lado y a otro, y muestra grandes oscilaciones entre nodos. Prolongando la mirada a todo el intervalo, el polinomio baja hasta -348.21 en $x = 0.537991$ y sube hasta 108.79 en $x = 13.2348$, ambos extremos globales *dentro* del intervalo y no en sus bordes —en $x = 0$ y $x = 15$ vale exactamente 10 y 85, que son los datos—.

Dónde se rompe cada uno. El interpolante a trozos se rompe fuera del rango de los datos: `interp1` devuelve `NaN` para $t < 0$ o $t > 15$, y no es un fallo sino una negativa razonada, porque unir puntos con segmentos no dice nada de lo que hay más allá del último. El de

Lagrange no se niega a extrapolar, y ahí está el peligro: en $x = 16$ devuelve 2745.91, un número perfectamente formateado y sin ningún contenido. El otro sitio donde se rompe es el condicionamiento: la matriz del ajuste de grado diez tiene número de condición $1.7 \cdot 10^{14}$, a menos de dos ’ordenes de magnitud de la inversa de la precisión de máquina. Al aumentar el grado el problema puede agravarse, pero no hay un número universal de nodos a partir del cual falle. Centrar y escalar las abscisas mejora el cálculo; cambiar a interpolación local también evita las oscilaciones del polinomio global.

> *«Ya tengo el apartado (d) hecho de la forma estándar (dos `plot` con `hold on`), pero no sé cómo poner la leyenda. Quiero lo esencial, lo que siempre voy a necesitar, no mil florituras.»*

Lo esencial está en la sesión cero: `legend` etiqueta las curvas en el orden en que se llamó a `plot`. En el código de arriba, primer `plot` → «Lagrange», segundo → «Lineal a trozos», tercero → «Datos». El argumento `'Location','best'` deja que MATLAB elija la esquina que menos tapa. Un truco de escritura: como `plot(x,y)` une los datos con segmentos, la orden `plot(x, y, 'k')` *ya dibuja* el interpolante lineal a trozos, sin llamar a `interp1`.

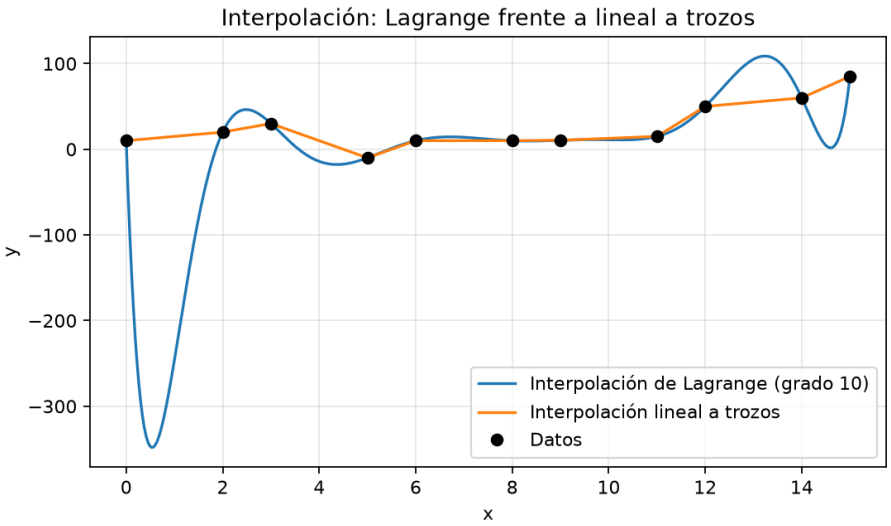


Figura 1.4: Los dos interpolantes sobre los once puntos. El lineal a trozos sigue los datos con calma; el polinomio de grado diez los clava todos pero se dispara entre nodos, hundiéndose hasta ≈ -348 cerca de $x \approx 0.5$ y volviendo a oscilar cerca de $x = 15$. Es el riesgo de oscilación al interpolar con grado alto.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
import matplotlib.pyplot as plt

x = np.array([0, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15], dtype=float)
y = np.array([10, 20, 30, -10, 10, 10, 10.5, 15, 50, 60, 85], dtype=float)

c = np.polyfit(x, y, x.size - 1) # (a) grado n-1: es el interpolante
plagrange = lambda t: np.polyval(c, t)
```

```
inlineal = lambda t: np.interp(t, x, y, left=np.nan, right=np.nan) # (b)
print(plagrange(1.0), inlineal(1.0)) # (c) -> -245.785... 15.0

t = np.linspace(0, 15, 100) # (d)
plt.plot(t, plagrange(t), label="Lagrange (grado 10)")
plt.plot(t, inlineal(t), label="Lineal a trozos")
plt.plot(x, y, "ko", label="Datos")
plt.legend(); plt.xlabel("x"); plt.ylabel("y"); plt.grid(True)
plt.title("Interpolacion: Lagrange frente a lineal a trozos")
plt.show()
```

Equivalencias: `np.polyfit` y `np.polyval` son `polyfit` y `polyval`, con el mismo orden de coeficientes; `np.interp(t, x, y)` es la interpolación lineal de `interp1(x,y,t)` dentro del intervalo. Fuera de él, NumPy devuelve por defecto el valor del extremo; por eso se han puesto `left=np.nan` y `right=np.nan` para igualar el comportamiento usado en MATLAB; `hold on` no hace falta porque Matplotlib superpone por defecto.

Ejercicio 1.1: la matriz de Vandermonde y los bucles

Ejercicio 1.5. *El sistema de Vandermonde*

ORIGEN: ejercicio 1.1 del repaso de MATLAB del curso. El libro previo lo dejó planteado sin resolver; se resuelve aquí.

Para recordar el manejo de bucles, se obtiene el polinomio de interpolación resolviendo el sistema de Vandermonde, aunque es bien sabido que no es un método numéricamente eficiente. Con las cinco primeras componentes de los vectores del ejercicio anterior, $x = (0, 2, 3, 5, 6)$ e $y = (10, 20, 30, -10, 10)$, construir la matriz de Vandermonde

1. usando un doble bucle `for`, o un bucle `for` y notación vectorial;
2. usando la orden `vander`;

resolver el sistema con cada matriz y comparar los coeficientes con los de `polyfit` usando la norma euclídea.

Resolución.
La idea, y el enunciado ya avisa de que el método es malo. Lo interesante de este ejercicio es que pide hacer a propósito lo que no se debe hacer, y merece la pena entender por qué se pide. Interpolar planteando y resolviendo el sistema de Vandermonde es la traducción literal de la definición —«que el polinomio pase por los puntos» son n ecuaciones lineales— y por eso es el camino que a uno se le ocurre primero. Es también el más caro ($O(n^3)$ frente a $O(n^2)$ de Newton) y el peor condicionado, porque las columnas de V son potencias de las mismas abscisas y se parecen cada vez más entre sí según sube el grado. El ejercicio sirve entonces para dos cosas a la vez: repasar bucles, que es lo que dice el enunciado, y **medir** el precio de un mal método en un caso donde la respuesta correcta se conoce por otra vía.
El núcleo son dos líneas y ninguna es un bucle: `V3 = vander(x)`, que construye la matriz, y `c3 = V3 \ y'`, que resuelve el sistema. Los dos bucles de arriba son el ejercicio de programación; la barra invertida es el ejercicio de cálculo. Y conviene saber qué hay detrás de esa barra, porque no es una división: MATLAB analiza la matriz, elige el algoritmo —aquí,

eliminación gaussiana con pivoteo parcial— y resuelve. Escribir `inv(V)*y` calcularía además una inversa que nadie ha pedido, con más operaciones y más error.
El sistema. Imponer que el polinomio pase por los cinco puntos da cinco ecuaciones lineales en los cinco coeficientes. Ordenando los coeficientes de mayor a menor grado, como `polyfit`, la matriz del sistema tiene en la fila i las potencias $x_i^4, x_i^3, x_i^2, x_i, 1$:

$$V = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 16 & 8 & 4 & 2 & 1 \\ 81 & 27 & 9 & 3 & 1 \\ 625 & 125 & 25 & 5 & 1 \\ 1296 & 216 & 36 & 6 & 1 \end{pmatrix}, \qquad Vc = y.$$

El determinante de V es $\prod_{i<j}(x_j - x_i) \neq 0$ porque las abscisas son distintas: por eso el interpolante existe y es único, que es el hecho que el ejercicio 1 daba por sabido.
Órdenes nuevas. Un bucle `for` anidado recorre filas y columnas. `zeros(n,n)` reserva la matriz antes de llenarla. La orden `vander(x)` construye exactamente esta matriz, con las potencias decrecientes de izquierda a derecha. Y la barra invertida, `V\y`, resuelve el sistema lineal $Vc = y$ por eliminación gaussiana con pivoteo; es la forma correcta de resolver un sistema en MATLAB, nunca `inv(V)*y`, que es más lento y menos preciso.

```
%% Ejercicio 1.1 - Matriz de Vandermonde y bucles
x = [0, 2, 3, 5, 6];
y = [10, 20, 30, -10, 10];
n = numel(x);

% --- 1. Matriz de Vandermonde con un doble bucle for ---
V1 = zeros(n, n);
for i = 1:n
  for j = 1:n
    V1(i, j) = x(i)^(n - j); % potencias decrecientes: n-1, ..., 0
  end
end

% --- 1 bis. Un solo bucle y notacion vectorial: toda la fila de golpe ---
V2 = zeros(n, n);
for i = 1:n
  V2(i, :) = x(i).^(n-1:-1:0);
end

% --- 2. Con la orden vander ---
V3 = vander(x);

% --- Resolver los sistemas y comparar con polyfit (norma euclidea) ---
c1 = V1 \ y'; % y' porque el sistema pide columna
c3 = V3 \ y';
cp = polyfit(x, y, n - 1)';
disp('Coeficientes por Vandermonde y por polyfit:')
disp([c1, c3, cp])
disp('Norma euclidea de las diferencias:')
disp([norm(c1 - cp), norm(c3 - cp), norm(V1 - V2), norm(V1 - V3)])
```

SALIDA ESPERADA

```
Coeficientes por Vandermonde y por polyfit:
```

```
    1.3611 1.3611 1.3611
  -15.9444 -15.9444 -15.9444
   55.5278 55.5278 55.5278
  -53.1667 -53.1667 -53.1667
   10.0000 10.0000 10.0000
Norma euclidea de las diferencias:
   1.0e-11 *
    0.3074 0.3074 0 0
```

Los tres métodos dan el mismo polinomio, como debía ser, y las diferencias son del orden de 10^{-12}: ruido de redondeo. Que no sean exactamente cero es la enfermedad de la matriz de Vandermonde: sus columnas, potencias de las mismas abscisas, son casi paralelas, y el sistema está mal condicionado. Con cinco puntos no importa; con once, `polyfit` ya avisaba. La línea `1.0e-11 *` es la forma que tiene MATLAB de decir que todos los números de la matriz hay que multiplicarlos por 10^{-11}.

El control, y hay dos, uno algebraico y otro numérico. El algebraico es el determinante: $\det V = \prod_{i<j}(x_j - x_i)$ debe salir 12960 con estas abscisas, y sale. El numérico es evaluar el polinomio obtenido en los propios nodos y exigir que devuelva los datos; con cinco puntos el residuo máximo es $1.4 \cdot 10^{-13}$. Los dos controles miden cosas distintas: el primero dice que la matriz que se ha construido es la que se quería, y el segundo, que el sistema se ha resuelto bien. Comparar con `polyfit`, que es lo que pide el enunciado, comprueba menos de lo que parece —dos métodos pueden coincidir y estar los dos mal— y sirve sobre todo para ver el tamaño del ruido.

Dónde se rompe, y se puede seguir el desmoronamiento paso a paso. El número de condición de V crece de forma brutal con el número de nodos:

$$n = 5:\ 9.9 \cdot 10^{3}, \qquad n = 8:\ 7.9 \cdot 10^{8}, \qquad n = 11:\ 1.7 \cdot 10^{14},$$

y el residuo en los nodos lo acompaña, $1.4 \cdot 10^{-13}$, $1.6 \cdot 10^{-10}$ y $2.8 \cdot 10^{-8}$. Con doble precisión, que da unas dieciséis cifras, un condicionamiento de 10^{14} puede consumir catorce en el peor caso al calcular los coeficientes. Es una estimación de sensibilidad, no una cuenta exacta de las cifras perdidas en cada resultado. El efecto depende de los nodos, de su escala y del algoritmo.

Y hay una segunda rotura que sorprende más, porque no depende del número de nodos sino de **dónde están**: multiplicar todas las abscisas por diez multiplica el condicionamiento por unos $4.6 \cdot 10^{8}$ —de $1.7 \cdot 10^{14}$ a $7.8 \cdot 10^{22}$ con los mismos once nodos—. La razón es que las columnas de V son $x^{10}, x^{9}, \ldots, 1$, y al escalar x las columnas dejan de tener tamaños comparables. Eso importa de verdad en el ejercicio siguiente, donde las abscisas son días del 1 al 88: ahí el ajuste de grado dos es inofensivo, pero uno de grado diez sobre esas mismas abscisas sería ruido. **El remedio, cuando haga falta, es centrar y escalar** —trabajar con $(x - \bar{x})/s$ en vez de con x—, que es exactamente lo que hace `polyfit` cuando se le piden tres salidas.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
x = np.array([0, 2, 3, 5, 6], dtype=float); y = np.array([10, 20, 30, -10, 10], dtype=
    float)
n = x.size
V1 = np.zeros((n, n))
```

```
for i in range(n):
    for j in range(n):
        V1[i, j] = x[i] ** (n - 1 - j)
V3 = np.vander(x) # potencias decrecientes, como vander
c1 = np.linalg.solve(V1, y) # equivale a V \ y
cp = np.polyfit(x, y, n - 1)
print(np.c_[c1, cp]); print(np.linalg.norm(c1 - cp), np.linalg.norm(V1 - V3))
```

Ejercicio 2: la recta de regresión

Ejercicio 1.6. *Recta de regresión*

ORIGEN: hoja de ejercicios del tema 1, ejercicio 2.

Con los mismos puntos del ejercicio 1, calcular y representar gráficamente la recta de regresión.

Resolución.
La idea, y es el cambio de pregunta. Los dos ejercicios anteriores exigían pasar por todos los puntos. Este renuncia a eso a propósito, y esa renuncia es la decisión de modelización: se supone que los datos traen ruido, o que el fenómeno de fondo es sencillo y las desviaciones no son información. En cuanto se admite eso, exigir que la curva pase por los puntos deja de ser una virtud y pasa a ser un defecto, porque obliga a la curva a seguir el ruido. **Interpolar y ajustar no son dos técnicas para el mismo problema: son dos problemas distintos**, y lo que decide cuál toca no es el ordenador sino lo que uno cree sobre los datos. La misma orden, `polyfit`, hace las dos cosas: con grado $n-1$ interpola y con grado menor ajusta. Esa unidad no es casual, y se ve al escribir el problema como un sistema lineal $Vc = y$ con V de n filas y $d+1$ columnas: si es cuadrado hay solución exacta, y si tiene más filas que columnas se busca la que minimiza el residuo.
El núcleo es una línea, `c1 = polyfit(x, y, 1)`. Las dos que la siguen —grados 2 y 3— no resuelven el ejercicio: están para poder comparar, y en esa comparación está lo que el ejercicio enseña.
Qué se calcula. Se busca la recta $y = mx + b$ que minimiza

$$S(m,b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2,$$

la suma de los cuadrados de los *residuos*, es decir, de las distancias verticales entre cada dato y la recta. S es un polinomio de grado dos en (m,b) y es convexo, porque es una suma de cuadrados de funciones afines; por eso su mínimo está exactamente donde se anulan las dos derivadas parciales:

$$\frac{\partial S}{\partial b} = -2\sum_i (y_i - mx_i - b) = 0, \qquad \frac{\partial S}{\partial m} = -2\sum_i x_i\,(y_i - mx_i - b) = 0.$$

Son las **ecuaciones normales**, y conviene leerlas antes de resolverlas. La primera dice que los residuos suman cero: la recta deja tanto error por encima como por debajo. La segunda,

que los residuos no tienen correlación con x: no queda en ellos ninguna tendencia lineal que la recta hubiera podido recoger. Como sistema lineal en (b, m) son

$$n\,b + \Big(\sum_i x_i\Big)m = \sum_i y_i, \qquad \Big(\sum_i x_i\Big)b + \Big(\sum_i x_i^2\Big)m = \sum_i x_i y_i.$$

Dividiendo la primera por n sale $b = \bar{y} - m\bar{x}$, y sustituyendo en la segunda se despeja m:

$$m = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}, \qquad b = \bar{y} - m\bar{x}.$$

El caso más pequeño, a mano. Con los tres puntos $(0,0)$, $(1,1)$ y $(2,3)$ las sumas son $n = 3$, $\sum x_i = 3$, $\sum x_i^2 = 5$, $\sum y_i = 4$ y $\sum x_i y_i = 7$, y las ecuaciones normales quedan $3b + 3m = 4$ y $3b + 5m = 7$. Restando, $2m = 3$: $m = 3/2$ y $b = -1/6$. Los residuos son $1/6$, $-1/3$ y $1/6$; suman cero, y $\sum x_i r_i = 0 - \frac{1}{3} + \frac{2}{6} = 0$, como dicen las dos ecuaciones. La suma de sus cuadrados es $1/6$, y ninguna otra recta deja menos.
Con los datos del ejercicio, $\bar{x} = 85/11 \approx 7.727$, $\bar{y} = 290.5/11 \approx 26.41$, y sale $m \approx 3.8267$, $b \approx -3.1612$. `polyfit(x,y,1)` hace exactamente esto.

La idea: por qué se llaman normales

Pedir que la recta pase por los once puntos es el sistema $Vc = y$, con $c = (m, b)^t$ y V la matriz de once filas cuyas dos columnas son el vector x de las abscisas y un vector de unos. Ese sistema no tiene solución. Al variar c, los vectores Vc recorren un plano de $\mathbb{R}^{11}$, el generado por esas dos columnas, y el vector de datos y está fuera de él. Mínimos cuadrados elige el punto del plano más cercano a y, que es su proyección ortogonal. En ese punto el residuo $r = y - Vc$ es perpendicular al plano, es decir, a sus dos columnas: $r \cdot (1, \ldots, 1) = 0$ y $r \cdot x = 0$, o en una línea $V^t(y - Vc) = 0$. Son las dos ecuaciones normales: aquí *normal* quiere decir perpendicular. Con un polinomio de grado d el plano pasa a tener $d + 1$ dimensiones, una por cada columna x^k, y la idea no cambia. Con grado $n - 1$ ese subespacio es todo $\mathbb{R}^n$, el residuo es cero y el ajuste es la interpolación del ejercicio 1.

Conexión: Inferencia Estadística, el modelo lineal

Esta recta es el estimador de mínimos cuadrados del modelo lineal simple $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ que cerró el curso de Inferencia Estadística en su tema de modelos lineales. Allí se probó que $\hat{\beta}_1 = S_{xy}/S_{xx}$ y $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$, que son las fórmulas de arriba, y que con errores normales independientes de media cero y varianza común son además los estimadores de máxima verosimilitud. Aquí no se hace inferencia, solo el ajuste; pero conviene saber que `polyfit` devuelve exactamente $\hat{\beta}$, y que el ajuste de grado d es el modelo lineal con regresores $x, x^2, \ldots, x^d$.

Y hacia delante. El criterio de minimizar la suma de los cuadrados de los residuos no se queda en los polinomios. En la sección 4.9 se usa para identificar los parámetros de un modelo de ecuaciones diferenciales a partir de observaciones: allí los residuos son las diferencias entre los datos y la solución numérica del modelo, que depende de los parámetros de forma no lineal. Ya no hay ecuaciones normales que resolver de una vez, y el mínimo se busca con un minimizador numérico; pero lo que se minimiza es la misma suma.

```
%% Ejercicio 2 - Recta de regresion (y parabola y cubica de ajuste)
x = [0, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15];
y = [10, 20, 30, -10, 10, 10, 10.5, 15, 50, 60, 85];

c1 = polyfit(x, y, 1); % recta: [pendiente, ordenada en el origen]
r1 = @(t) polyval(c1, t);
c2 = polyfit(x, y, 2); % parabola de ajuste
r2 = @(t) polyval(c2, t);
c3 = polyfit(x, y, 3); % cubica de ajuste
r3 = @(t) polyval(c3, t);
fprintf('Recta: y = %.4f x + (%.4f) \n', c1(1), c1(2))

m1 = min(x); m2 = max(x);
t = linspace(m1, m2);
figure
plot(x, y, 'r.', 'MarkerSize', 15)
hold on
plot([m1, m2], [r1(m1), r1(m2)], 'k', 'LineWidth', 2) % dos puntos bastan
plot(t, r2(t), 'b--', 'LineWidth', 1.3)
plot(t, r3(t), 'g:', 'LineWidth', 1.3)
hold off
legend('Datos', 'Recta de regresion', 'Parabola', 'Cubica', 'Location', 'northwest')
xlabel('x'), ylabel('y'), title('Ajuste por minimos cuadrados'), grid on
```

SALIDA ESPERADA

```
Recta: y = 3.8267 x + (-3.1612)
```

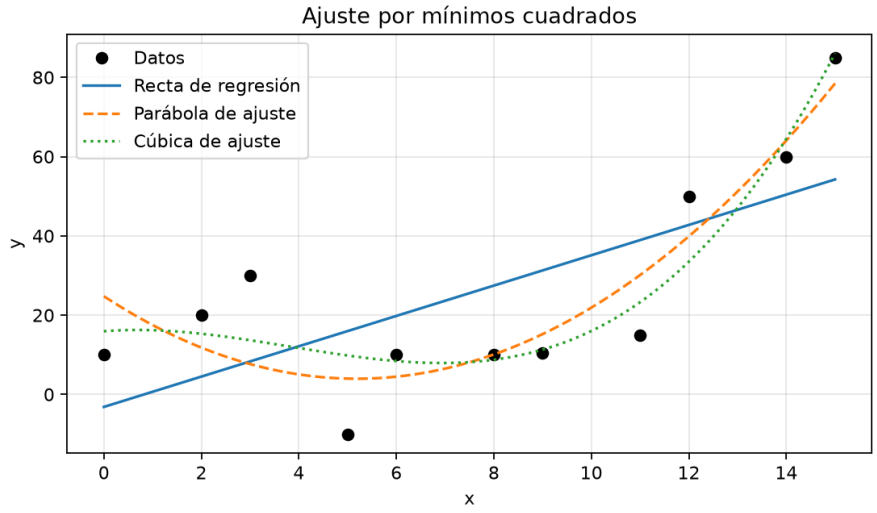


Figura 1.5: Los once datos con la recta de regresión, la parábola y la cúbica de ajuste. No se exige pasar por todos los puntos. Los grados bajos usados aquí suavizan los datos y evitan las grandes oscilaciones del interpolante de grado diez.

El control. Las fórmulas de arriba son el control, y por eso se han escrito: calculadas a mano dan $m = 3.826740$ y $b = -3.161172$, y `polyfit` devuelve 3.8267 y -3.1612. Hay un segundo control que no exige ninguna cuenta y conviene tener a mano: **la recta de regresión pasa siempre por el centro de gravedad de los datos**, es decir por $(\bar{x}, \bar{y}) = (7.7273, 26.4091)$,

porque la primera ecuación normal, la de los residuos de suma cero, dividida por n es exactamente $\bar{y} = m\bar{x} + b$. Comprobarlo cuesta una línea y detecta cualquier error en el término independiente.

Dónde se rompe, y aquí no se rompe el programa sino el criterio. La suma de cuadrados de los residuos baja al subir el grado: 4054.0 con la recta, 1422.6 con la parábola, 1083.4 con la cúbica. Y sigue bajando hasta llegar a cero con grado diez, que es el interpolante del ejercicio 1, el que se desploma hasta -348 entre dos nodos. **Por tanto el residuo no sirve para elegir el grado**: siempre premia al grado mayor. La elección se hace por fuera —sabiendo cuántos parámetros tolera el fenómeno, o midiendo el error sobre datos que no se han usado para ajustar—, y eso segundo es exactamente lo que hace el ejercicio siguiente reservando los ocho últimos días.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np, matplotlib.pyplot as plt
x = np.array([0, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15.]); y = np.array([10, 20, 30, -10, 10,
    10, 10.5, 15, 50, 60, 85.])
c1 = np.polyfit(x, y, 1); print("recta: y = %.4f x + (%.4f)" % (c1[0], c1[1]))
t = np.linspace(x.min(), x.max(), 100)
plt.plot(x, y, "r.", ms=12, label="Datos")
plt.plot(t, np.polyval(c1, t), "k", lw=2, label="Recta de regresion")
plt.plot(t, np.polyval(np.polyfit(x, y, 2), t), "b--", label="Parabola")
plt.plot(t, np.polyval(np.polyfit(x, y, 3), t), "g:", label="Cubica")
plt.legend(); plt.grid(True); plt.show()
```

Ejercicio 3: cotizaciones de una acción

Ejercicio 1.7. *Ajuste de una serie de cotizaciones*

ORIGEN: hoja de ejercicios del tema 1, ejercicio 3, y ejercicio 3.1 del repaso de MATLAB.

En el fichero `Finanzas.dat` está recogido el precio de una acción de la Bolsa española a lo largo de 88 días, tomado al cierre de cada día. Se quiere ajustar estos datos por una función que permita predecir el precio para un corto intervalo de tiempo más allá de la última cotización. Se pide:

a) leer los datos del fichero y representarlos mediante una poligonal;
b) calcular el polinomio de ajuste de grado 2 y representarlo en la misma gráfica, con etiquetas y leyendas.

Y, del repaso de MATLAB (ejercicios 3 ampliado y 3.1): calcular los ajustes de grado 1, 2 y 3 usando solo los primeros 80 días y el error cuadrático de cada uno en los días 81 a 88; obtener el precio máximo, el mínimo y el día en que se alcanza; qué días estuvo el precio por encima de 27800 y qué días entre 27800 y 28000; y el precio medio y su desviación típica.

Resolución.

La idea. Este es el primer ejercicio del libro con datos reales, y el cambio no es de técnica —se usa el mismo `polyfit`— sino de estatuto: aquí **no hay ninguna ley del mercado especificada detrás del ajuste**. En los modelos de población de la sesión cero la ecuación venía de unas hipótesis sobre el fenómeno y por eso extrapolar significaba algo; aquí el polinomio de

grado dos no dice nada sobre la bolsa, es una curva suave que pasa cerca de los datos. Toda la gracia del ejercicio está en llegar a ver esa diferencia con un número delante, y por eso el grueso de la resolución no es el ajuste sino la parte que reserva los ocho últimos días y mide el error fuera de la muestra.

La decisión de método que hay que saber justificar es **por qué grado dos**. No es que dos sea mágico: es el grado más bajo que puede reproducir un máximo, y la serie sube y luego baja. Grado uno no podría; grado ocho seguiría cada vaivén y no sería una tendencia.

El núcleo del programa son dos líneas —`A = load('Finanzas.dat')`, que trae los datos, y `c = polyfit(dd, A, 2)`, que los resume en tres números— y una tercera que es la que de verdad se examina, la que evalúa el ajuste en los días que no se usaron para construirlo. Todo lo demás es contabilidad de vectores: máximos, umbrales con `find` y estadísticos.

Lectura y poligonal. `A = load('Finanzas.dat')` devuelve una columna de 88 precios. Como los días no están en el fichero, se crean con `dd = (1:88)'`, columna para que tenga la misma forma que `A`. `plot(dd, A)` es la poligonal.

Ajuste y predicción. `polyfit(dd, A, 2)` da la parábola de mínimos cuadrados, $p(t) \approx -0.8957\,t^2 + 64.99\,t + 25959$. Evaluada en `1:95` se extiende una semana más allá de los datos; la predicción para el día 89 es ≈ 24648 y para el 95, ≈ 24049: la parábola, que ya baja, sigue bajando. Es un modelo de ajuste, no de mercado, y el ejercicio del repaso muestra su límite: ajustando con 80 días, el error en los ocho siguientes es de $2.2 \cdot 10^7$ para la recta, $4.6 \cdot 10^6$ para la parábola y $3.9 \cdot 10^6$ para la cúbica; mejora con el grado, pero errores de $\sqrt{3.9 \cdot 10^6/8} \approx 700$ unidades de cotización por día muestran una capacidad predictiva limitada en este tramo. El ejercicio no valida una estrategia de inversión.

Máximos, mínimos y umbrales. `[M, iM] = max(A)` da el máximo y su posición; `find(A > 27800)` da los días que cumplen la condición; `find(A > 27800 & A < 28000)` combina dos condiciones; `mean` y `std` dan media y desviación típica (`std` divide por $n-1$, como la cuasivarianza de Inferencia).

```
%% Ejercicio 3 - Cotizaciones: lectura, poligonal y ajuste de grado 2
A = load('Finanzas.dat'); % 88 precios, en columna
dd = (1:88)'; % dias, en columna

% --- (a) Poligonal ---
figure
h1 = plot(dd, A, 'b');
hold on

% --- (b) Parabola de ajuste, extendida a 95 dias ---
c = polyfit(dd, A, 2);
s = @(z) polyval(c, z);
d = 1:95;
h2 = plot(d, s(d), 'r', 'LineWidth', 1.5);
hold off
legend([h1, h2], 'Precio de cierre (poligonal)', 'Ajuste de grado 2', 'Location', 'best')
xlabel('Dias'), ylabel('Precio de la accion'), title('Parabola de ajuste'), grid on
fprintf('Coeficientes: %.6f %.6f %.6f \n', c)
fprintf('Prediccion dias 89 a 95: \n'), disp(round(s(89:95)))

% --- Repaso: ajustes con 80 dias y error cuadratico en los dias 81 a 88 ---
for g = 1:3
```

```
  cg = polyfit(dd(1:80), A(1:80), g);
  err = sum((polyval(cg, dd(81:88)) - A(81:88)).^2);
  fprintf('grado %i: error cuadratico dias 81-88 = %.4e \n', g, err)
end

% --- Repaso 3.1: maximo, minimo, umbrales, media y desviacion tipica ---
[M, iM] = max(A); [m, im] = min(A);
fprintf('Maximo %g en el dia %i; minimo %g en el dia %i \n', M, iM, m, im)
disp('Dias con precio > 27800:'), disp(find(A > 27800)')
disp('Dias con 27800 < precio < 28000:'), disp(find(A > 27800 & A < 28000)')
fprintf('Media %.4f, desviacion tipica %.4f \n', mean(A), std(A))
```

SALIDA ESPERADA

```
Coeficientes: -0.895746 64.986651 25959.374025
Prediccion dias 89 a 95:
      24648 24553 24455 24357 24256 24153 24049
grado 1: error cuadratico dias 81-88 = 2.2096e+07
grado 2: error cuadratico dias 81-88 = 4.6423e+06
grado 3: error cuadratico dias 81-88 = 3.9014e+06
Maximo 28720 en el dia 45; minimo 24060 en el dia 84
Dias con precio > 27800:
    44 45 46
Dias con 27800 < precio < 28000:
    46
Media 26499.5000, desviacion tipica 909.2879
```

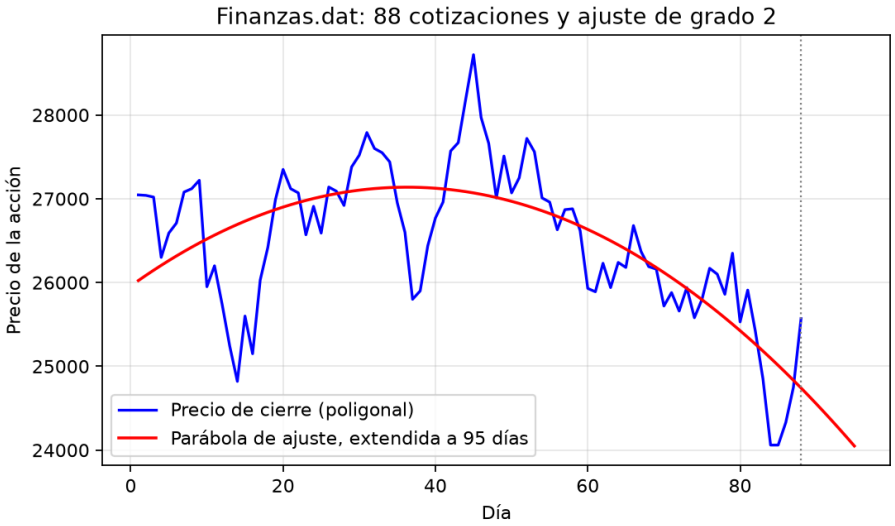


Figura 1.6: Las 88 cotizaciones como poligonal y la parábola de mínimos cuadrados, extendida hasta el día 95. La línea vertical marca el último dato.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np, matplotlib.pyplot as plt
A = np.loadtxt("Finanzas.dat"); dd = np.arange(1, 89)
c = np.polyfit(dd, A, 2); d = np.arange(1, 96)
plt.plot(dd, A, "b", label="Precio de cierre (poligonal)")
plt.plot(d, np.polyval(c, d), "r", lw=1.5, label="Ajuste de grado 2")
```

```
plt.legend(); plt.xlabel("Dias"); plt.ylabel("Precio de la accion"); plt.grid(True); plt.
    show()
for g in (1, 2, 3):
    cg = np.polyfit(dd[:80], A[:80], g)
    print(g, np.sum((np.polyval(cg, dd[80:]) - A[80:]) ** 2))
print(A.max(), dd[A.argmax()], A.min(), dd[A.argmin()])
print(dd[A > 27800], dd[(A > 27800) & (A < 28000)])
print(A.mean(), A.std(ddof=1)) # ddof=1: divide por n-1, como std de MATLAB
```

Los controles, y el segundo desmonta el ajuste. El primero es de lectura: la media 26499.50 tiene que caer entre el mínimo 24060 y el máximo 28720, y el máximo tiene que estar donde la poligonal tiene su pico. Además **ni el máximo ni el mínimo caen en el borde de la serie** —están en los días 45 y 84 de 88, y el último dato es 25570—, lo que permite comprobar que hay valores observados a ambos lados. Si el máximo estuviera en el último día, seguiría siendo el máximo de la muestra; lo que no podríamos afirmar es que la serie vaya a dejar de crecer después.

El segundo control es el que importa. El vértice de la parábola ajustada está en $t = -64.9867/(2 \cdot (-0.895746)) = 36.28$, mientras que el precio real hace su máximo en el día 45: **la curva y los datos ni siquiera coinciden en dónde está el pico**, y se diferencian en nueve días de ochenta y ocho. Ese desajuste no es un error del cálculo; es la señal de que la parábola solo recoge una tendencia gruesa. Tampoco se debe exigir que un modelo de tendencia reproduzca exactamente el día del mayor dato si hay fluctuaciones.

Dónde se rompe. En la extrapolación, y con una cifra: ajustando con ochenta días, el error cuadrático en los ocho siguientes es $3.9 \cdot 10^6$ para la cúbica, es decir $\sqrt{3.9 \cdot 10^6/8} \approx 698$ unidades de cotización de error típico por día sobre precios de 25 000, casi un tres por ciento diario. Y se rompe más lejos de forma absurda: el término cuadrático es negativo, así que la parábola baja sin límite y cruza el cero en $t = 210.33$; a partir del día 211 el modelo predice **precios negativos**. Ninguna de las dos cosas la detecta el programa, porque el programa no sabe que está modelando un precio. **Las restricciones del fenómeno —que un precio es positivo, que una población no supera su techo— las tiene que imponer quien modeliza, y un ajuste polinómico ordinario no las impone por sí solo.**

QUÉ DEJA. Leer un fichero, dibujar la poligonal y ajustar es el flujo de trabajo de cualquier modelado con datos. Y una lección de modelización: un polinomio que ajusta bien el pasado no tiene por qué predecir el futuro; el error fuera de la muestra es la medida honesta, y un ajuste no garantiza por sí solo las restricciones del fenómeno.

Ejercicio 4: el modelo de Leslie

Ejercicio 1.8. *Una M-función para el modelo de Leslie*

ORIGEN: hoja de ejercicios del tema 1, ejercicio 4. El código sigue el fichero `Leslie.m` del curso.

Escribir una M-función con datos de entrada la matriz de Leslie L y la inicialización $v(0)$, para:

1. obtener los autovalores y autovectores de L;
2. calcular el módulo de los autovalores y detectar la existencia de un autovalor positivo

dominante $\hat{\lambda}$;
3. proporcionar un autovector asociado a $\hat{\lambda}$ con componentes positivas que sumen uno (los porcentajes de hembras en cada clase para tiempos grandes);
4. calcular los vectores de distribución por clases $v(k)$ para $k \in \{1, \ldots, 20\}$.

Aplicarlo a la matriz y la inicialización del ejemplo, $a = (0, 1, 3, 2)$, $b = (0.25, 0.5, 0.1)$, $v(0) = (100, 70, 70, 40)^t$.

Resolución.

La idea, y explica la estructura entera de la función. El programa hace dos cosas que parecen independientes y no lo son: por un lado calcula el espectro de L, por otro itera $v(k+1) = Lv(k)$ veinte veces. La primera parte da la respuesta **teórica** —quién es $\hat{\lambda}$ y cómo se reparten las clases a la larga— y la segunda da la respuesta **experimental**. **Que las dos coincidan es el ejercicio**: la función está montada para que las últimas filas de la tabla se puedan comparar con la última línea, y esa comparación es a la vez la comprobación de que el programa funciona y la demostración numérica de la convergencia predicha por el teorema; no sustituye su demostración.

Hay una decisión de método, y es la que casi nadie ve: **no hace falta iterar para conocer el límite**. `eig` permite obtenerlo a partir del autovector de Perron. Aquí la matriz es 4×4 y ambos cálculos son baratos. Iterar permite ver el transitorio; para matrices grandes y dispersas, calcular solo el autovector necesario por iteración puede ser mucho más barato que obtener todo el espectro. No hay un factor fijo de ahorro entre ambos métodos.

El núcleo son tres líneas: `L = diag(b,-1)` con `L(1,:) = a`, que construye el modelo; `[V,D] = eig(L)`, que lo resuelve; y `v = L*v` dentro del bucle, que es el modelo aplicado. Las quince líneas de `fprintf` son formato, y conviene no confundirlas con contenido aunque ocupen la mitad del fichero.

Órdenes nuevas. `eig` calcula autovalores y autovectores; con dos salidas, `[V, D]`, las columnas de `V` son los autovectores y la diagonal de `D` los autovalores, en el mismo orden. Los autovalores complejos aparecen con su parte imaginaria, y un real puede aparecer como `0.8514 + 0.0000i` si algún otro es complejo: es real, solo que MATLAB imprime toda la columna en formato complejo. `abs` toma módulos y funciona igual con complejos. `fprintf` con un vector como argumento recicla el formato tantas veces como haga falta, y es así como se imprime cada fila de la tabla.

La lógica de la función. Se construye L; se piden autovalores y autovectores; se localiza el de mayor módulo con `max`, que devuelve el valor y la posición; se toma la columna correspondiente de `V`, se le quita el signo y se normaliza a suma uno; y se itera veinte veces imprimiendo el vector y sus proporciones. La comparación entre las últimas proporciones y el autovector es la comprobación del apartado 3 del teorema de convergencia. El código siguiente presupone la matriz primitiva del ejemplo y un vector inicial no negativo y no nulo. El máximo de los módulos por sí solo no comprueba esas hipótesis: si se cambian los datos, hay que verificar que el autovalor seleccionado sea real positivo, simple y estrictamente dominante antes de interpretar las proporciones como un límite. La tabla está preparada para cuatro clases.

```
function Leslie(a, b, v0)
% Leslie(a, b, v0): distribucion de la poblacion en el modelo de Leslie
% y analisis del comportamiento a largo plazo.
```

```
% a : vector de n componentes (fecundidades)
% b : vector de n-1 componentes (supervivencias)
% v0 : vector columna de n componentes (distribucion inicial)
% Prueba: a=[0,1,3,2]; b=[0.25,0.5,0.1]; v0=[100;70;70;40]; Leslie(a,b,v0)

% Construccion de la matriz de Leslie
L = diag(b, -1); % supervivencias en la subdiagonal
L(1, :) = a; % fecundidades en la primera fila

% 1. Autovalores y autovectores
[V, D] = eig(L); % V: autovectores por columnas; D: diagonal

% 2. Modulos y deteccion del dominante
lambda = abs(diag(D));
disp('Modulos de los autovalores de L:'), disp(lambda')
[lgorro, Ind] = max(lambda);
fprintf('Autovalor dominante = %12.8f, es el numero %4i \n', lgorro, Ind)

% 3. Autovector asociado, positivo y de suma uno
vv = abs(V(:, Ind)); % un autovector esta definido salvo signo
vv = vv / norm(vv, 1);

% 4. Distribucion en 20 periodos, con las proporciones frente al total
Niter = 20;
v = v0; S = sum(v); vS = v / S;
fprintf(' k v1(k) v2(k) v3(k) v4(k) v1/S v2/S v3/S v4/S \n')
fprintf(' %3i %9.3f %9.3f %9.3f %9.3f %7.4f %7.4f %7.4f %7.4f \n', 0, v, vS)
for k = 1:Niter
  v = L * v;
  S = sum(v); vS = v / S;
  fprintf(' %3i %9.3f %9.3f %9.3f %9.3f %7.4f %7.4f %7.4f %7.4f \n', k, v, vS)
end
fprintf(' Autovector de suma uno: %7.4f %7.4f %7.4f %7.4f \n', vv)
end
```

SALIDA ESPERADA

```
>> a=[0,1,3,2]; b=[0.25,0.5,0.1]; v0=[100;70;70;40]; Leslie(a,b,v0)
Modulos de los autovalores de L:
    0.8514 0.6483 0.6483 0.0699
Autovalor dominante = 0.85142496, es el numero 1
  k v1(k) v2(k) v3(k) v4(k) v1/S v2/S v3/S v4/S
   0 100.000 70.000 70.000 40.000 0.3571 0.2500 0.2500 0.1429
   1 360.000 25.000 35.000 7.000 0.8431 0.0585 0.0820 0.0164
   2 144.000 90.000 12.500 3.500 0.5760 0.3600 0.0500 0.0140
   3 134.500 36.000 45.000 1.250 0.6205 0.1661 0.2076 0.0058
   4 173.500 33.625 18.000 4.500 0.7556 0.1464 0.0784 0.0196
   5 96.625 43.375 16.812 1.800 0.6092 0.2735 0.1060 0.0113
  ...
  18 14.374 4.170 2.460 0.293 0.6749 0.1958 0.1155 0.0138
  19 12.135 3.593 2.085 0.246 0.6720 0.1990 0.1154 0.0136
  20 10.340 3.034 1.797 0.208 0.6723 0.1973 0.1168 0.0136
 Autovector de suma uno: 0.6728 0.1976 0.1160 0.0136
```

El orden de los autovalores que devuelve `eig` puede variar con la versión; lo que no varía es que el dominante es 0.8514 y que la población se extingue, perdiendo un 14.9 % cada

período asintóticamente; durante el transitorio la tasa varía. La figura 1.7 dibuja las cuatro columnas de proporciones.

El control, y tiene una forma cuantitativa que conviene usar. Las proporciones en $k = 20$ son $(0.6723, 0.1973, 0.1168, 0.0136)$ y el autovector de Perron normalizado es $(0.6728, 0.1976, 0.1160, 0.0136)$: con los valores sin redondear, el error máximo es $8.15 \cdot 10^{-4}$. El cociente espectral $0.6483/0.8514 \approx 0.7615$ explica la tasa asintótica, proporcional a 0.7615^k. Su potencia veinte es $4.3 \cdot 10^{-3}$, pero no es por sí sola una cota del error: falta una constante que depende del dato inicial y de los autovectores. Las oscilaciones proceden de los autovalores complejos. Un error menor de lo esperado no es señal de fallo: algunos modos pueden tener un peso pequeño o nulo.

Un segundo control, más barato, es `sum(L*v) / sum(v)` en las últimas iteraciones: ese cociente tiende a $\hat{\lambda}$ y en $k = 20$ vale 0.8533, a dos milésimas de 0.851425 —y su distancia también está gobernada por la misma tasa 0.7615^k—. Y un tercero, estructural: la suma de las componentes del autovector normalizado tiene que ser 1, y sus cuatro componentes positivas, que es lo que garantiza Perron-Frobenius y lo que hace que la respuesta se pueda leer como porcentajes.

Dónde se rompe. En dos sitios, y los dos aparecen en cuanto se cambian los datos. El primero: la línea `vv = abs(V(:,Ind))` toma valor absoluto para arreglar el signo, y eso funciona porque el autovector de Perron tiene todas sus componentes del mismo signo. Con una matriz que no fuese de Leslie —o con `Ind` apuntando por error a otro autovalor— el valor absoluto no arreglaría nada: convertiría un vector con signos mezclados en otro vector que puede dejar de ser autovector de L, y el programa seguiría imprimiendo cuatro números bonitos que suman uno. El segundo: si la matriz de Leslie **no es primitiva** —por ejemplo con una sola fecundidad no nula, $a = (0, 0, 0, 2)$— hay varios autovalores del mismo módulo máximo —los cuatro valen 0.397635— y las proporciones no convergen: se repiten con período cuatro, de modo que en $k = 20$ y en $k = 40$ se recupera exactamente el reparto inicial $(0.3571, 0.25, 0.25, 0.1429)$ y la tabla no se parece a la última línea por mucho que se alargue el bucle. El programa no avisa: hay que reconocerlo mirando la columna de módulos, y por eso el apartado 2 pide calcularlos y no solo quedarse con el mayor.

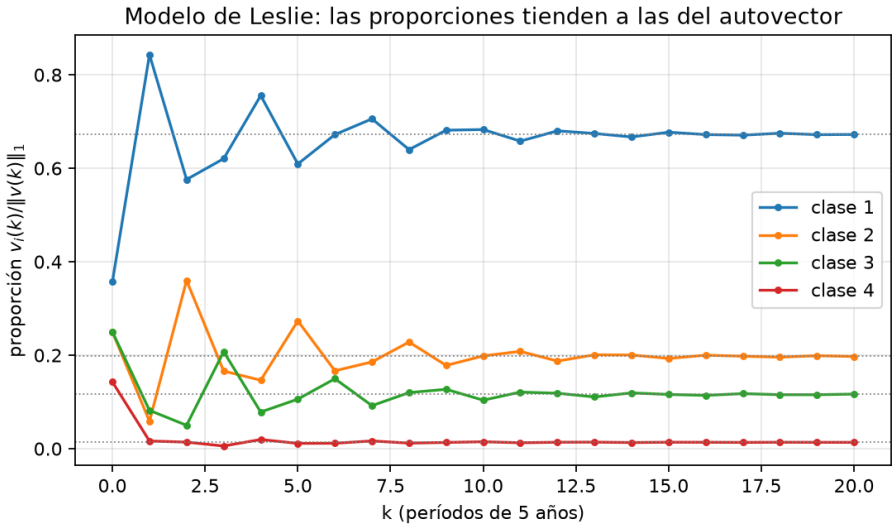


Figura 1.7: Proporción de cada clase de edad frente al total, $v_i(k)/\|v(k)\|_1$, a lo largo de veinte períodos. Las líneas de puntos son las componentes del autovector de Perron normalizado. Las oscilaciones de los primeros pasos las producen los autovalores complejos, de módulo 0.648, y se amortiguan como $(0.648/0.851)^k$.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
def leslie(a, b, v0, niter=20):
    L = np.diag(b, -1).astype(float); L[0, :] = a
    D, V = np.linalg.eig(L) # D autovalores, V autovectores por columnas
    ind = np.argmax(np.abs(D)); lam = D[ind].real
    vv = np.abs(V[:, ind].real); vv = vv / vv.sum()
    print("dominante %.8f, autovector de suma uno" % lam, vv)
    v = np.array(v0, dtype=float)
    for k in range(niter + 1):
        print("%3d" % k, np.round(v, 3), np.round(v / v.sum(), 4))
        v = L @ v
leslie([0, 1, 3, 2], [0.25, 0.5, 0.1], [100, 70, 70, 40])
```

Ejercicio 5: PageRank de una red aleatoria grande

Ejercicio 1.9. *PageRank de una red aleatoria con matrices huecas*

ORIGEN: hoja de ejercicios del tema 1, ejercicio 5. El código sigue el fichero `PageRank_2.m` del curso.

Escribir un programa, preferiblemente una M-función, para:

(a) construir de forma aleatoria una matriz de adyacencia asociada a una red con un número grande de nodos;

(b) calcular a partir de ella la matriz de probabilidades de transición, almacenarla en forma `sparse` y comparar la memoria usada en forma llena y en forma `sparse`;

(c) utilizar el algoritmo de la teoría para calcular el PageRank de la red.

Resolución.
La idea, y son dos decisiones de método, no una. El PageRank de la teoría es el autovector de Perron de la matriz de Google $\mathcal{P} = \alpha(P + vd^t) + (1-\alpha)v\,e^t$ (con e el vector de unos y $w = v$). Con diez mil páginas, $\mathcal{P}$ tiene 10^8 entradas y **todas distintas de cero**, ocupa ochocientos megabytes y calcular su espectro con `eig` es impensable. Las dos decisiones que hacen el cálculo posible son:
Primera: no diagonalizar, iterar. No hace falta todo el espectro, solo un autovector, y para eso basta el método de la potencia, que cuesta un producto matriz-vector por iteración. Es la misma decisión que en el ejercicio de Leslie se podía permitir el lujo de no tomar: allí la matriz era 4×4 y `eig` salía gratis.
Segunda, y es la fina: no formar $\mathcal{P}$. La matriz P sí es hueca —cada página enlaza a unas pocas páginas en promedio, no a diez mil—, y la parte que la densifica es el rango uno $(1-\alpha)v\,e^t$. Se evita escribiendo el producto por trozos: $\mathcal{P}x = \alpha Px + \big[\alpha d^t x + (1-\alpha)e^t x\big]v$. El corchete es un escalar, por lo que este sumando cuesta N operaciones en vez de N^2. En el código eso son las tres líneas de `x1 = alpha*(PS*x0)`, `gamma = norm(x0,1) - norm(x1,1)` y `x1 = x1 + gamma*v`. **Ese es el núcleo del programa**, y es donde vive la idea entera: la matriz que gobierna el proceso nunca llega a existir en la memoria del ordenador. Lo demás —sortear la red, dividir por las sumas de columnas, dibujar— es preparación y presentación.
Conviene ver por qué γ es lo que hay que sumar. Si $x_0 \geq 0$ tiene norma uno, entonces $\|\alpha P x_0\|_1$ vale α menos lo que se ha ido por los nodos sin salida, cuyas columnas son nulas; $\gamma = \|x_0\|_1 - \|\alpha P x_0\|_1$ es exactamente la masa que falta para volver a sumar uno, y repartirla según v es lo que hacen a la vez el salto aleatorio y el arreglo de los nodos colgantes. Por eso el programa funciona sin haber tratado por separado las columnas nulas.
Órdenes nuevas. `rand(m,1)` da m números al azar en $(0,1)$ y `randi(N,m,1)` da m enteros al azar entre 1 y N. `rng('shuffle')` cambia la semilla del generador en cada sesión, y `rng(1)` la fija para que el resultado sea reproducible; aquí se usa la segunda para poder comparar salidas. `sparse(i, j, s)` construye una matriz hueca con el valor `s(k)` en la posición `(i(k), j(k))`, sumando si una posición se repite; solo guarda las entradas no nulas. `min(PS, 1)` recorta a 1 las posiciones repetidas. `whos PS` dice cuánta memoria ocupa. `spy(PS)` dibuja el patrón de ceros y no ceros. `sum(PS)` suma por columnas. `histogram(x)` dibuja un histograma. Y el producto `PS*x0` con `PS` hueca solo hace las operaciones de las entradas no nulas: es lo que hace posible el cálculo.
La construcción de la red. Se sortean $N \cdot$`nme` pares (fila, columna) y se pone un 1 en cada uno con probabilidad 0.95 (el vector `H`); se anula la diagonal para que ninguna página se enlace a sí misma. Como algunas parejas se repiten, hay algo menos de $N \cdot$`nme` enlaces. Después, `z = sum(PS)` da las sumas de las columnas, se sustituyen por 1 las que valen 0 para poder dividir, y `PS * diag(1./z)` divide cada columna j por z_j: es la matriz de transición con las columnas nulas intactas. Multiplicar por la derecha por una diagonal escala las columnas; por la izquierda escalaría las filas.

```
function PageRank_red
% PageRank de una red aleatoria de N paginas, sin construir la matriz de Google.
N = 10000; % numero de paginas
nme = 7; % pares sorteados por pagina; algunos se descartan
alpha = 0.98; % factor de amortiguacion
```

```
tol = 1.e-15; % tolerancia del test de parada
rng(1); % semilla fija: resultados reproducibles

% (a) Matriz de adyacencia aleatoria, en forma hueca
Nnme = N * nme;
H = +(rand(Nnme, 1) < 0.95); % unos con probabilidad 0.95 (el + convierte a numero)
ind = randi(N, Nnme, 1); % filas (destinos)
jnd = randi(N, Nnme, 1); % columnas (origenes)
PS = min(sparse(ind, jnd, H), 1); % pares repetidos se recortan a 1
for k = 1:N
  PS(k, k) = 0; % sin enlaces de una pagina a si misma
end
whos PS % memoria en forma hueca
fprintf('Memoria en forma llena: %d bytes \n', N * N * 8)
figure(1), spy(PS(1:300, 1:300)), title('Patron de la esquina 300 x 300')

% (b) Matriz de transicion: dividir cada columna por su suma
z = sum(PS); % sumas de columnas (vector fila)
z0 = find(z == 0); % columnas nulas: nodos sin salida
z(z0) = 1; % se sustituye 0 por 1 para poder dividir
PS = PS * diag(sparse(1 ./ z)); % PS(:, j) / z(j); las columnas nulas siguen nulas

% (c) Metodo de la potencia sin formar la matriz de Google
v = 1 / N; % vector de personalizacion uniforme (escalar: v*ones)
x0 = v * ones(N, 1);
for k = 1:200
  x1 = alpha * (PS * x0);
  gamma = norm(x0, 1) - norm(x1, 1);
  x1 = x1 + gamma * v;
  err = norm(x1 - x0, 1);
  fprintf(' k = %4i, error = %20.16f \n', k, err)
  if err < tol
    fprintf(' El algoritmo converge en %3i iteraciones \n', k)
    [pmax, imax] = max(x1);
    fprintf(' Pagina mas relevante: %i, con PageRank %.4e \n', imax, pmax)
    figure(2), histogram(x1, 60), xlabel('PageRank'), ylabel('numero de paginas')
    return
  end
  x0 = x1;
end
warning('No se alcanzo la tolerancia en 200 iteraciones; revisar err.')
end
```

SALIDA ESPERADA

```
  Name Size Bytes Class Attributes
  PS 10000x10000 1144072 double sparse
Memoria en forma llena: 800000000 bytes
 k = 1, error = 0.3311710797585264
 k = 2, error = 0.1325946081282069
  ...
 k = 39, error = 0.0000000000000017
 k = 40, error = 0.0000000000000007
 El algoritmo converge en 40 iteraciones
 Pagina mas relevante: 9193, con PageRank 4.4218e-04
```

Qué de esa salida es un dato y qué no. La traza de arriba es la de la versión en Python con `default_rng(1)`, que es la que ha generado también las dos figuras. Con `rng(1)` de MATLAB el generador es otro, la red sorteada es otra, y la página ganadora y el número exacto de iteraciones serán distintos: **ninguno de esos dos números es un resultado del problema**. Sí lo son los órdenes de magnitud, que no dependen del sorteo: en esta realización, 66 504 enlaces de los 70 000 pares sorteados (el resto los quita la probabilidad 0.95, los pares repetidos y la diagonal), algo más de un megabyte en forma hueca frente a los 800 megabytes de la forma llena —una relación de 1 a 699—, y, en este ejemplo, parada en cuarenta iteraciones. La memoria hueca se puede predecir a mano: MATLAB guarda por cada entrada no nula un valor y un índice de fila, ocho bytes cada uno, más un puntero por columna, de modo que son $16 \cdot 66\,504 + 8 \cdot 10\,001 = 1\,144\,072$ bytes en esa estimación del almacenamiento de MATLAB. El bloque mezcla una traza numérica obtenida en Python con ese formato ilustrativo de memoria; no es una ejecución literal de `whos` en MATLAB.

El control de la convergencia. La proposición garantiza $\|x^k - x_\infty\|_1 \leq 2\alpha^k$, porque dos vectores estocásticos distan como mucho 2. Con $\alpha = 0.98$, esa cota garantiza error menor que 10^{-15} a partir de 1744 iteraciones en aritmética exacta. El test del programa, sin embargo, mide el cambio entre iterados: al parar con $\delta < 10^{-15}$ garantiza un error menor que $49 \cdot 10^{-15}$, salvo redondeo. La red concreta permite parar en cuarenta pasos porque mezcla mucho mejor que el peor caso. Asintóticamente intervienen los autovalores no estacionarios de P' (con los nodos sin salida ya corregidos), multiplicados por α. Cerca de la precisión de máquina, hay que vigilar también el estancamiento por redondeo.

El histograma muestra que casi todas las páginas tienen un PageRank cercano a $1/N = 10^{-4}$, con una cola hacia la derecha: las que reciben más enlaces. La ganadora tiene 4.42 veces la media, en esta realización de la red aleatoria; en la red real de Internet la cola es muchísimo más larga.

Y los dos controles que hay que ejecutar siempre en este algoritmo. El primero: `norm(x1, 1)` tiene que valer 1 en cada iteración, salvo pequeños errores de redondeo, porque el vector es una distribución de probabilidad; una desviación apreciable exige revisar el reparto de γ. El segundo: todas las componentes tienen que ser positivas, que es lo que garantiza Perron-Frobenius sobre $\mathcal{P}$. Los dos cuestan una línea, y son los que detectan el error más fácil de cometer aquí, que es dividir por las sumas de *filas* en vez de por las de columnas.

Dónde se rompe. Subiendo α hacia 1. Con $\alpha = 1$ desaparece la mezcla, $\mathcal{P}$ vuelve a ser P', que no tiene por qué ser primitiva, y el método puede no converger —es el mismo caso degenerado que la matriz de Leslie no primitiva del ejercicio anterior—. Con $\alpha = 0.999$ converge, pero la cota α^k pasa a exigir decenas de miles de iteraciones y el resultado se vuelve muy sensible a la estructura de enlaces: pequeños cambios en la red pueden mover mucho el orden. El valor 0.85 que se usa habitualmente no es una constante física, es un compromiso entre fidelidad a los enlaces reales y estabilidad del cálculo, y elegirlo es una decisión de modelización como cualquier otra de este libro.

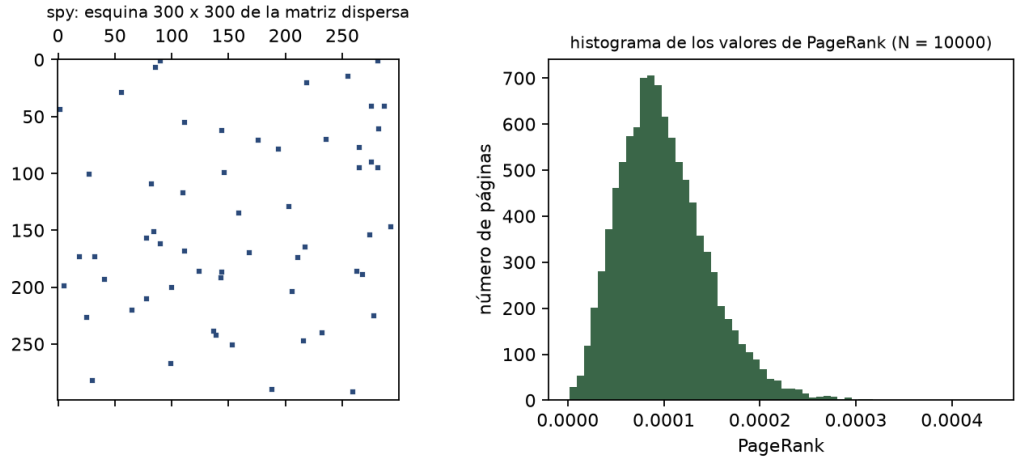


Figura 1.8: A la izquierda, el patrón de entradas no nulas de la esquina 300×300 de la matriz de transición hueca: cada punto es un enlace. A la derecha, el histograma de los diez mil valores de PageRank. Ambas figuras se han generado con la versión en Python; en MATLAB las producen `spy` y `histogram`.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy import sparse
N, nme, alpha, tol = 10000, 7, 0.98, 1e-15
rng = np.random.default_rng(1)
H = (rng.random(N * nme) < 0.95).astype(float)
ind = rng.integers(0, N, N * nme); jnd = rng.integers(0, N, N * nme)
PS = sparse.csc_matrix((H, (ind, jnd)), shape=(N, N)); PS.data[:] = np.minimum(PS.data,
    1)
PS.setdiag(0); PS.eliminate_zeros()
z = np.asarray(PS.sum(axis=0)).ravel(); z[z == 0] = 1
PS = PS @ sparse.diags(1.0 / z) # divide cada columna por su suma
v = 1.0 / N; x0 = np.full(N, v)
for k in range(1, 201):
    x1 = alpha * (PS @ x0); gamma = np.abs(x0).sum() - np.abs(x1).sum(); x1 += gamma * v
    err = np.abs(x1 - x0).sum()
    if err < tol: break
    x0 = x1
else:
    raise RuntimeError("No se alcanzo la tolerancia en 200 iteraciones")
print("test satisfecho en", k, "iteraciones; maximo", x1.max(), "en la pagina", x1.argmax
    () + 1)
```

QUÉ DEJA. Una matriz hueca solo guarda lo que no es cero, y multiplicarla por un vector cuesta lo que cuestan sus entradas no nulas. El algoritmo eficiente nunca forma la matriz de Google: multiplica por αP y reparte con γ lo que falta hasta uno. Es la diferencia entre formar una matriz enorme y realizar unas pocas decenas de productos dispersos.

REPASO INTERCALADO: CAPÍTULO 1, EN TRES RONDAS

Las tres rondas se hacen a libro cerrado, en sesiones distintas y separadas por días. Como es el primer capítulo, la mezcla es con las asignaturas ya cursadas en las que se apoya, no con capítulos anteriores.

Primera ronda, al terminar el capítulo.

1. Escribe la matriz de Leslie de una población de tres clases con $a = (0, 2, 1)$ y $b = (0.5, 0.4)$, y calcula $v(2)$ desde $v(0) = (10, 10, 10)^t$.
2. ¿Qué dos hipótesis sobre a y b hacen irreducible a L, y cuál la hace primitiva? Da un ejemplo de matriz de Leslie irreducible y no primitiva.
3. (Álgebra Lineal) Enuncia el teorema de Gershgorin y úsalo para acotar los autovalores de una matriz estocástica por columnas.
4. (Ecuaciones Diferenciales) Para $y' = Ay$ con A constante, ¿qué condición sobre los autovalores hace que toda solución tienda a cero? Compara con la condición discreta.
5. ¿Qué es $d^t x$ en el algoritmo de PageRank y cómo se calcula sin conocer d?

Segunda ronda, una semana después.

1. Para $n \geq 2$ y $b_i > 0$ para todo i, demuestra que $a_n > 0$ equivale a que la matriz de Leslie sea irreducible.
2. Deduce la ecuación $\sum_i a_i b_1 \cdots b_{i-1} \lambda^{-i} = 1$ y explica por qué tiene una sola raíz positiva.
3. (Inferencia Estadística) ¿Qué minimiza la recta de regresión y cómo se llaman los estimadores que `polyfit(x,y,1)` devuelve?
4. (Matemática discreta) Define grafo fuertemente conexo y componente fuertemente conexa. ¿Qué forma tiene la matriz de adyacencia de un grafo sin ciclos tras reordenar sus vértices?
5. ¿Por qué la matriz de Google es estrictamente positiva aunque la web tenga grupos de páginas aisladas, y qué teorema se aplica entonces?

Tercera ronda, antes de empezar el capítulo 3.

1. Sea P' estocástica por columnas y reducible. Construye un ejemplo 3×3 con dos autovectores estocásticos distintos asociados a 1, y explica qué significa para el internauta.
2. Demuestra que $\|x^k - x_\infty\|_1 \leq \alpha^k \|x^0 - x_\infty\|_1$ para el método de la potencia con la matriz de Google.
3. (Álgebra Lineal) Si $L = SJS^{-1}$ es la forma de Jordan, escribe L^k y di qué términos sobreviven al dividir por $\hat{\lambda}^k$ cuando $\hat{\lambda}$ es estrictamente dominante.
4. (Ecuaciones Diferenciales) El modelo logístico $y' = ry(1 - y/K)$ del capítulo 2 tiene un equilibrio en K. ¿Cuál es su linealización ahí y qué dice de la estabilidad? Compara con «$\hat{\lambda} < 1$» en Leslie.
5. Con $\alpha \in (0, 1)$ y v cualquiera estocástico con $v_i > 0$, ¿sigue siendo único el PageRank? ¿Y si algún $v_i = 0$?

CAPÍTULO 2

Modelos de procesos gobernados por ecuaciones y sistemas diferenciales ordinarios

ANTES DE LEER

Contesta estas cuatro preguntas antes de leer. No hace falta acertarlas: intenta razonar y vuelve a ellas después de estudiar.

1. Una población crece «en proporción a lo que hay». ¿Qué ecuación es esa, qué le pasa a su solución cuando el tiempo crece, y qué hipótesis biológica hay que añadir para que deje de crecer sin límite?
2. Si $y' = f(y)$ con f regular y $f(y_0) > 0$, ¿puede la solución que sale de y_0 cruzar un cero de f? ¿Puede oscilar? ¿A qué puede tender?
3. En una epidemia sin nacimientos ni muertes, ¿qué cantidad se conserva, y qué dice eso de cuántas ecuaciones hacen falta de verdad?
4. Zorros y conejos: si los zorros comen conejos y los conejos se reproducen solos, ¿se llega a un equilibrio, se extingue alguien, o pasa otra cosa? ¿Qué herramienta decide?

Las respuestas están, respectivamente, en las secciones 2.1.1 y 2.1.2, 2.1, 2.2.1 y 2.2.4.

En este tema se presentan y analizan varios modelos, con origen en distintas ramas de la Ciencia, que se formulan de manera eficiente con ecuaciones y sistemas diferenciales ordinarios (EDOs y SDOs). El proceso es siempre el mismo: se parte de un fenómeno que se quiere comprender y, si es posible, cuantificar; se eligen las variables «importantes»; se escriben las relaciones entre ellas y sus derivadas, basadas en leyes que tienen su origen en la observación; y el sistema que resulta, con frecuencia acompañado de condiciones adicionales, es el modelo que hay que analizar y, hasta donde se pueda, resolver. Los ejemplos se han elegido por su riqueza en Dinámica de Poblaciones, pero las mismas ecuaciones describen

procesos que en principio no tienen nada que ver: una masa radiactiva, una taza que se enfría, un depósito con sal, una noticia que se propaga.

El libro previo del autor abría con una reflexión general sobre qué es construir un modelo, antes de entrar en Leslie. Se conserva aquí porque es exactamente la discusión que este capítulo pone en práctica una y otra vez: cada modelo nace de decidir qué se desprecia.

> *«Entiendo que en la modelización de un fenómeno, parte fundamental de esa modelización es extraer las características relevantes para su estudio, y desechar las que no. Eso permite una suerte de aproximación iterada en la cual partimos de algunos supuestos fundamentales que desechan otros secundarios y después ajustar el modelo para contemplar esos supuestos adicionales. Entiendo que hay muchas ocasiones en que esta forma de operar es completamente impráctica porque añadir supuestos quiebra por completo la validez del modelo… De manera que debemos ser conscientes de: 1. El objetivo concreto que queremos solucionar. 2. El alcance concreto de nuestras simplificaciones, para no extrapolar aberrantemente.»*

Los dos puntos son dos problemas distintos. El primero, el objetivo, es un problema de *diseño* y mira hacia delante: fija el nivel de resolución que hace falta. Para predecir muchos movimientos orbitales sirve la mecánica de Newton; para explicar con precisión la precesión del perihelio de Mercurio se necesitan correcciones relativistas. El segundo, el alcance, es un problema de *validez* y mira hacia atrás: es saber dónde deja de valer lo que se ha construido. Este capítulo enseña las dos cosas con ejemplos: el modelo exponencial se diseña con una sola hipótesis y se ve dónde falla; el logístico añade una hipótesis y arregla ese fallo; el de pesca añade un agente externo; y los sistemas de dos especies muestran que, al pasar de una variable a dos, pueden aparecer órbitas periódicas no constantes, imposibles en una EDO autónoma escalar con unicidad. La dependencia de la extinción respecto del dato inicial ya puede aparecer en una variable, como veremos con el efecto Allee. En cada paso se dice qué se ha despreciado y qué se pierde por ello.

Conexión: Ecuaciones Diferenciales Ordinarias, qué se usa de allí

Este capítulo se apoya de forma continua en la asignatura de Ecuaciones Diferenciales Ordinarias. Todo lo que usa está reenunciado en el apéndice A, que es el apéndice de complementos del curso, y cada vez que se invoque un resultado se escribirá entero en el punto de uso. Para poder ir a mirar la demostración se cita además, capítulo y sección, el libro de EDO del autor (*Notas de Ecuaciones Diferenciales Ordinarias*, en adelante «libro de EDO»; su notación está traducida al principio del apéndice). El mapa de lo que se usa: la existencia y unicidad local de Picard (§2.3) y la solución maximal con su caracterización (§3.3 y §3.4); la comparación de soluciones (§3.5); los sistemas lineales, la exponencial de una matriz y las EDO lineales de orden n (§4.3, §4.6 y §4.9); las definiciones de estabilidad y la linealización (§6.1 y §6.4); las funciones de Liapunov (§7.2 y §7.3); y las órbitas con los teoremas de LaSalle y Poincaré-Bendixson (§8.2 a §8.5). Lo que se usa más veces es esto: que por cada dato inicial pasa una única solución maximal, que dos órbitas distintas no se cortan, y que una solución acotada en un compacto está definida para todo tiempo positivo. Y el teorema con el que abre el capítulo está también allí, como corolario del principio de invariancia (§8.4, «Corolario (caso unidimensional)»).

2.1 Modelos regidos por EDOs

LA PREGUNTA

Un biólogo mide la velocidad a la que crece un cultivo en función de cuántas células hay, y obtiene una curva $f(y)$ con un par de ceros. No sabe resolver la ecuación $y' = f(y)$ que resulta y no le hace falta: solo quiere saber si el cultivo se estabiliza, en qué valor, y si puede oscilar. ¿Se puede leer todo eso en el dibujo de f, sin resolver nada?

En una primera etapa se consideran modelos regidos por EDOs, es decir, descritos por el comportamiento de una función escalar. Nos restringimos a fenómenos modelados por una EDO **autónoma**, $y' = f(y)$, en la que el segundo miembro no depende del tiempo. Un resultado general acerca del comportamiento de las soluciones es el siguiente, y es la herramienta que resuelve de golpe casi todos los modelos escalares del capítulo. En este contexto, «punto crítico» significa **equilibrio**, es decir, un cero de f; no un cero de f'.

Teorema (comportamiento de las soluciones de una EDO autónoma escalar)**.** Sea φ la solución maximal del problema de Cauchy

$$\begin{cases} y' = f(y), \\ y(0) = y_0, \end{cases}$$

donde $f \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$ e $y_0 \in \mathbb{R}$, definida en el intervalo abierto $I(y_0)$.

- Si $f(y_0) = 0$, entonces y_0 es un punto de equilibrio, $I(y_0) = \mathbb{R}$ y $\varphi(t) \equiv y_0$.
- Si $f(y_0) > 0$, entonces φ es estrictamente creciente en $I(y_0)$. Además:
  a) si existe algún punto crítico de f mayor que y_0, entonces $I(y_0) \supset [0, +\infty)$ y existe $\lim_{t \to +\infty} \varphi(t) = \bar{y}$, donde $\bar{y}$ es el *menor* de los puntos críticos mayores que y_0;
  b) si no existe ningún punto crítico mayor que y_0, entonces $\lim_{t \to \sup I(y_0)} \varphi(t) = +\infty$, y $\sup I(y_0)$ puede ser finito o no.
- Un resultado análogo se tiene cuando $f(y_0) < 0$: φ es estrictamente decreciente en $I(y_0)$; si existe algún punto crítico menor que y_0, entonces $I(y_0) \supset [0, +\infty)$ y $\varphi(t) \to \bar{y}$, el *mayor* de los puntos críticos menores que y_0; y si no existe, $\varphi(t) \to -\infty$ cuando $t \to \sup I(y_0)$, que puede ser finito o no.

LA IDEA: LA RECTA COMO ESCENARIO

En una dimensión no hay sitio para dar vueltas. El estado es un punto de la recta real y $f(y)$ le dice hacia dónde moverse: a la derecha si $f(y) > 0$, a la izquierda si $f(y) < 0$, y quieto si $f(y) = 0$. Los ceros de f son paredes que no se pueden atravesar, porque por cada punto pasa una sola solución y la que pasa por un cero es la constante. Así que un punto que sale a la derecha sigue a la derecha para siempre, sin poder darse la vuelta (habría que pasar por un cero de f), hasta chocar con la primera pared que encuentre o hasta el infinito. Todo el teorema es esa imagen. Lo que hace falta demostrar es que «chocar con la pared» significa converger a ella y que el tiempo no se acaba antes.

Un caso concreto, leído en el dibujo. Antes de demostrar el teorema conviene verlo trabajar sobre una f concreta, elegida para que salgan todos sus casos:

$$y' = y^2 - 1 = (y-1)(y+1).$$

Los puntos críticos son -1 y 1. El signo de f se lee en su gráfica, una parábola que corta el eje en esos dos puntos (figura 2.1): positiva a la izquierda de -1, negativa entre -1 y 1, positiva a la derecha de 1. Debajo de la gráfica se dibuja la **línea de fases**, que es la recta de los estados con una flecha hacia la derecha donde $f > 0$ y hacia la izquierda donde $f < 0$. El teorema convierte esas flechas en afirmaciones:

- Si $y_0 < -1$: $f(y_0) > 0$, y el menor punto crítico mayor que y_0 es -1. La solución crece, está definida en $[0, +\infty)$ y tiende a -1 (caso a).
- Si $-1 < y_0 < 1$: $f(y_0) < 0$, y el mayor punto crítico menor que y_0 es -1. La solución decrece, está definida en $[0, +\infty)$ y tiende a -1.
- Si $y_0 > 1$: $f(y_0) > 0$ y no hay ningún punto crítico mayor que y_0. La solución crece y tiende a $+\infty$ cuando $t \to \sup I(y_0)$ (caso b).
- Si $y_0 = -1$ o $y_0 = 1$, la solución es constante.

El punto -1 atrae por los dos lados; el punto 1 repele por los dos lados y es la frontera entre acabar en -1 y escaparse. Todo esto se ha leído sin resolver la ecuación.

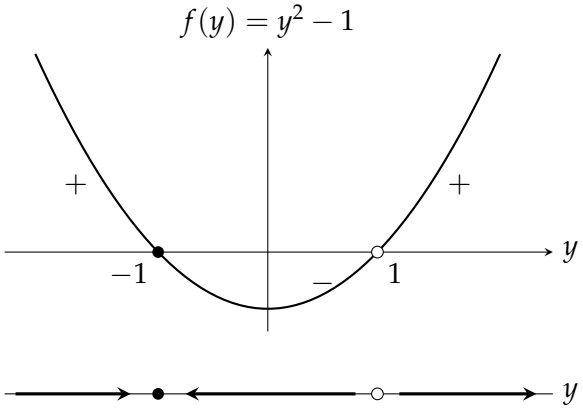


Figura 2.1: Arriba, la gráfica de $f(y) = y^2 - 1$ con su signo en cada tramo. Abajo, la línea de fases: cada flecha dice hacia dónde se mueve el estado. El punto lleno, -1, atrae por los dos lados; el hueco, 1, repele por los dos lados. A la derecha de 1 no hay ninguna pared, y la solución se va a $+\infty$.

La comprobación, resolviendo. Esta ecuación se resuelve separando variables, y conviene hacerlo una vez para ver que el dibujo no engaña. Con la descomposición

$$\frac{1}{y^2-1} = \frac{1}{2}\Big(\frac{1}{y-1} - \frac{1}{y+1}\Big)$$

se llega a $\frac{y-1}{y+1} = \frac{y_0-1}{y_0+1}e^{2t}$, y despejando, para $y_0 \neq \pm 1$,

$$y(t) = \frac{(y_0+1) + (y_0-1)e^{2t}}{(y_0+1) - (y_0-1)e^{2t}}.$$

Con $y_0 = 0$ sale $y(t) = \frac{1-e^{2t}}{1+e^{2t}} = -\tanh t$, que baja desde 0 hacia -1 sin llegar nunca: el caso con pared, ahora con fórmula. Con $y_0 = 2$ sale $y(t) = \frac{3+e^{2t}}{3-e^{2t}}$, cuyo denominador se anula en $t^* = \frac{1}{2}\log 3 \approx 0.549$: la solución llega a $+\infty$ en *tiempo finito*, y $\sup I(2) = \frac{1}{2}\log 3$.

Por qué el caso b) puede acabar en tiempo finito o no. Mientras $f > 0$ la solución es estrictamente creciente, y se puede usar la propia y como reloj: $dt = dy/f(y)$. El tiempo que tarda la solución en ir de y_0 al infinito es

$$\sup I(y_0) = \int_{y_0}^{+\infty} \frac{dy}{f(y)},$$

y todo depende de si esa integral converge. Para $f(y) = y^2 - 1$ el integrando se comporta como $1/y^2$ y la integral es finita: desde $y_0 = 2$ vale exactamente $\frac{1}{2}\log 3$, el mismo instante de antes. Para el modelo exponencial de la subsección siguiente, $f(y) = y$, el integrando es $1/y$ y la integral diverge: la solución $y_0 e^t$ también se va a infinito, pero tarda un tiempo infinito. Eso es todo lo que significa la frase «sup $I(y_0)$ puede ser finito o no» del enunciado.

Antes de la demostración, los tres resultados del apéndice que se van a usar, reenunciados. Los tres están también en el libro de EDO: la solución maximal en §3.3, su caracterización en §3.4, las órbitas en §8.2 y §8.3. Y el teorema mismo figura allí como «Corolario (caso unidimensional)» del libro de EDO, §8.4 «Principio de invariancia y Teorema de LaSalle» (página 174), con $f \in \mathrm{Lip}_{\mathrm{loc}}(a, b)$ en un intervalo (a, b) en lugar de $\mathbb{R}$ y los extremos b y a en el papel de $\pm\infty$; se deduce allí de LaSalle y de las órbitas. Aquí se da una demostración directa.

CONEXIÓN: LO QUE SE USA DE EDO EN LA DEMOSTRACIÓN

Existencia y unicidad de solución maximal (teorema de existencia y unicidad maximal del apéndice; libro de EDO, §3.3 «Prolongación de soluciones. Existencia y unicidad de solución maximal»). Si f es continua y localmente lipschitziana, por cada dato inicial pasa una única solución maximal, definida en un intervalo abierto $I(y_0)$. En un sistema autónomo, además, dos órbitas distintas no se cortan: si dos soluciones coinciden en un punto del espacio de fases, una es traslación en el tiempo de la otra (proposición sobre las órbitas de un sistema autónomo; §8.3 del libro de EDO).

Solución acotada en un compacto (proposición de prolongación en un compacto; §3.4 y §8.3 del libro de EDO). Si la semiórbita positiva $\gamma^+(y_0) = \{\varphi(t; y_0) : t \in I(y_0),\ t \geq 0\}$ está contenida en un compacto K del dominio, entonces $I(y_0) \supset [0, +\infty)$, y el conjunto $\Lambda^+(y_0)$ de los puntos límite positivos es compacto, invariante y no vacío.

Un límite es un equilibrio (observación sobre límites que son equilibrios). En las condiciones anteriores, si existe $\lim_{t\to+\infty} \varphi(t; y_0) = \bar{y}$, entonces $\bar{y}$ es un punto crítico. La razón es que $\{\bar{y}\}$ sería el conjunto límite, que es invariante: la órbita que sale de $\bar{y}$ tiene que quedarse en $\{\bar{y}\}$, luego es constante.

La demostración en una frase por paso. Se prueba que la solución *no puede darse la vuelta* (por unicidad, girar exigiría pisar un equilibrio); una función monótona tiene límite, así que hay destino; y ese destino, cuando es finito, no puede ser otra cosa que un equilibrio. Lo único delicado es descartar que el tiempo se acabe antes de llegar, y de eso se encarga la compacidad.

Paso 0: el caso trivial y la simetría. El caso $f(y_0) = 0$ lo resuelve la propia ecuación: la función constante $\varphi(t) \equiv y_0$ tiene derivada nula y $f(\varphi(t)) = f(y_0) = 0$, luego es solución en todo $\mathbb{R}$, y por unicidad es la solución maximal. Los casos $f(y_0) > 0$ y $f(y_0) < 0$ son simétricos, así que se supone $f(y_0) > 0$.
Paso 1: la unicidad prohíbe dar la vuelta — el núcleo de la prueba. Como órbitas distintas no se cortan y los equilibrios son órbitas formadas por un único punto, $\varphi(t)$ no puede ser un punto crítico para ningún t: si lo fuera, la órbita de φ y la del equilibrio se cortarían y serían la misma, y φ sería constante, contra $f(y_0) \neq 0$. Por tanto $\varphi'(t) = f(\varphi(t))$ no se anula en ningún t de $I(y_0)$. Como $\varphi'(0) = f(y_0) > 0$ y φ' es continua en un intervalo, si tomara algún valor negativo tendría que anularse en medio (teorema de Bolzano); luego $\varphi' > 0$ en todo $I(y_0)$ y φ es estrictamente creciente. En particular existe

$$\lim_{t \to \sup I(y_0)} \varphi(t) = \beta \in (-\infty, +\infty],$$

porque una función monótona siempre tiene límite en el extremo de su intervalo.
Paso 2: sin pared, el destino es el infinito (caso b). Entonces $\beta = +\infty$. En efecto, si β fuera finito, la semiórbita positiva estaría contenida en el compacto $[y_0, \beta]$, luego $I(y_0) \supset [0, +\infty)$ y, por la observación reenunciada arriba, β sería un punto crítico mayor que y_0, que no existe.
Paso 3: con pared, la trayectoria queda encerrada en un compacto y converge a la primera (caso a). En este caso la hipótesis da un punto crítico mayor que y_0; se llama $\tilde{y}$ a uno cualquiera de ellos, y de él se toma solo que es una pared por encima del punto de partida. La función φ no puede tomar nunca el valor $\tilde{y}$, por el argumento de las órbitas que no se cortan. Como es creciente y arranca en y_0,

$$y_0 \leq \varphi(t) < \tilde{y} \qquad \text{para todo } t \in I(y_0) \cap [0, +\infty).$$

Esto dice que $\gamma^+(y_0) \subset [y_0, \tilde{y}]$, un compacto; por la proposición del compacto, $I(y_0) \supset [0, +\infty)$. Además φ está acotada por $\tilde{y}$, luego β es finito, y por la observación es un punto crítico mayor que y_0. Es el menor de ellos: si hubiera otro $\hat{y}$ con $y_0 < \hat{y} < \beta$, la solución, que crece desde y_0 hasta β, tendría que pasar por $\hat{y}$, y eso es cortar la órbita de un equilibrio. ■

Qué hace cada hipótesis. La *lipschitzianidad local* de f es la hipótesis de fondo, y sostiene el paso 1 entero: sin unicidad las órbitas pueden cortarse, y con eso se cae todo lo demás. El ejemplo canónico es $y' = 3y^{2/3}$, continua pero no lipschitziana en el origen: por el dato inicial $y_0 = 0$ pasan la solución constante y también $\varphi(t) = t^3$, que se levanta del equilibrio y se marcha. Con eso una solución sí puede salir de una pared, y el teorema es falso. La *dimensión uno* es la otra hipótesis, y no está escrita en ninguna línea: vive en el hecho de que una función real monótona tiene límite. En el plano las órbitas tampoco se cortan, pero pueden rodearse, y el teorema deja de valer: una órbita acotada puede no converger a un punto y recorrer indefinidamente una órbita periódica. Esa es exactamente la frontera entre este teorema y el de Poincaré-Bendixson (libro de EDO, §8.5 «El caso de los sistemas planos. Teorema de Poincaré-Bendixson»), que se usa más adelante en este mismo capítulo con el modelo de Volterra.

Lo que hay que llevarse. Tres frases reconstruyen la prueba. (i) Una solución no constante

no puede pisar un equilibrio, porque las órbitas no se cortan; luego $f(\varphi(t))$ no cambia de signo y φ es estrictamente monótona. (ii) Una función monótona tiene límite; si es finito, la trayectoria vive en un compacto y la solución existe para todo $t \geq 0$. (iii) Un límite finito es un equilibrio, y tiene que ser la primera pared que la trayectoria encuentra en su sentido de marcha; si no hay pared, el límite es infinito, y cuándo se alcanza lo decide la integral de $1/f$. En la figura 2.1 se ven las tres: cada flecha es una solución monótona; las dos primeras acaban en la pared -1, y la tercera, sin pared, se va al infinito.

Observación. Hay una prueba directa, sin la proposición del compacto ni la observación del límite, de que si $f(y_0) > 0$ y β es finito entonces es un punto crítico. Se da aquí porque es corta e instructiva. Supongamos $\beta < +\infty$ y $f(\beta) \neq 0$; como φ crece hacia β y $f(\varphi(t)) > 0$, es $f(\beta) \geq 0$, luego $f(\beta) > 0$. Por continuidad existe $\delta > 0$ con $f(y) \geq f(\beta)/2 =: m > 0$ para $y \in [\beta - \delta, \beta]$. Desde el instante t_1 en que $\varphi(t_1) \geq \beta - \delta$, se tiene $\varphi'(t) \geq m$, luego $\varphi(t) \geq \varphi(t_1) + m(t - t_1)$: la solución va por encima de una recta de pendiente positiva. Si $\sup I(y_0) = +\infty$, la recta se va a infinito y φ no puede estar acotada por β. Y si $\sup I(y_0) = T < +\infty$, la solución permanece en el compacto $[y_0, \beta]$ hasta T, lo que contradice que una solución maximal con tiempo de vida finito tiene que salir de todo compacto (caracterización de las soluciones maximales, §3.4 del libro de EDO). Luego $f(\beta) = 0$.

EXPLÍCATE. *¿Dónde se ha usado que la ecuación es escalar? Señala el paso de la demostración que fallaría en $\mathbb{R}^2$, y piensa en un punto que gira alrededor de un equilibrio sin llegar a él.*

EL CAMINO QUE NO SE TOMÓ: MODELAR EN TIEMPO DISCRETO, COMO EN EL CAPÍTULO 1

La alternativa. Contar la población en instantes separados, $y_{k+1} = g(y_k)$, como hacía el modelo de Leslie. Es lo que de verdad hace un censo.
Su ventaja. No exige que la población sea grande ni que y sea derivable; el paso de tiempo es el de la observación; y en muchos organismos (insectos con una generación al año) la reproducción *es* discreta.
Qué la aparta aquí. El teorema que se acaba de demostrar no tiene análogo discreto: una iteración escalar $y_{k+1} = g(y_k)$ puede saltar por encima de un equilibrio, oscilar a su alrededor y hasta comportarse caóticamente (la logística discreta $y_{k+1} = r y_k(1 - y_k)$ es el ejemplo clásico). En tiempo continuo, la unicidad de solución prohíbe cruzar un equilibrio, y por eso el comportamiento cualitativo se lee en el signo de f sin resolver nada. Cuando la población es grande y los nacimientos están repartidos en el tiempo, el modelo continuo puede ser una aproximación útil y más sencilla de analizar, y es lo que hace la sección siguiente.

2.1.1 El modelo exponencial

LA PREGUNTA

Un arqueólogo encuentra un hueso y quiere saber de qué siglo es. No tiene ningún documento: solo el hueso, y el hecho de que el carbono 14 que contiene se va desintegrando. ¿Qué ley sigue esa desintegración, por qué la misma ley describe una población que crece o una taza de café que se enfría, y qué número, medible en el laboratorio, permite convertir «cuánto queda» en «cuánto tiempo ha pasado»?

Consideremos una población que evoluciona a lo largo del tiempo. El problema consiste en encontrar una función $y = y(t)$ que, en cada instante t, dé el número de individuos. Por

su propia definición esa función solo puede tomar valores discretos no negativos, luego si el número cambia, $t \mapsto y(t)$ no puede ser continua: una función continua con valores enteros en un intervalo es constante. Sin embargo, si el número de individuos es suficientemente grande, es razonable esperar que y se aproxime bien por una función continua e incluso diferenciable. Es la primera decisión de modelado, y conviene tenerla presente cada vez que un modelo dé, por ejemplo, 2.7 individuos.

Para escribir el modelo hay que hacer alguna suposición sobre el comportamiento de la población. La más simple es aceptar que **la velocidad de crecimiento en cada instante es proporcional al número de individuos en dicho instante**: con el doble de individuos esperamos el doble de nacimientos y el doble de muertes. Como $y(t)$ es el número de individuos, la velocidad de crecimiento es $y'(t)$, y la ecuación es

$$y' = \lambda y, \qquad \lambda \in \mathbb{R}.$$

Se dice que estamos ante el **modelo exponencial** o **modelo de Malthus**. El cociente y'/y entre la velocidad de crecimiento y la cantidad de individuos se llama **tasa de crecimiento**, y aquí se está suponiendo que es constante e igual a λ. Si $\lambda > 0$ hay más nacimientos que muertes; si $\lambda < 0$, al revés.

Las hipótesis, una a una. La ecuación $y' = \lambda y$ parece no suponer casi nada, y supone cuatro cosas. Conviene nombrarlas, porque cada modelo posterior del capítulo nace de quitar una de ellas.

- **La población es grande.** Es lo que permite tratar y como una función derivable. Se pierde todo lo que ocurre con pocos individuos: el azar de que los tres últimos sean del mismo sexo, o de que un año no nazca nadie. Con y pequeño el modelo estocástico suele ser más informativo; un valor como 2.7 puede tener sentido como esperanza, pero no como recuento exacto ni como garantía de supervivencia.
- **La tasa no depende de cuántos hay.** Un individuo se reproduce y muere al mismo ritmo medio en una población de cien que en una de un millón: el modelo desprecia la limitación de recursos en el intervalo estudiado. Es la hipótesis que se quita en el modelo logístico, y la primera que falla en la realidad.
- **La tasa no depende del tiempo.** No hay estaciones, ni años buenos y malos: la ecuación es autónoma. Quitarla lleva a $y' = \lambda(t)y$, que se resuelve igual de bien ($y = y_0 e^{\int_0^t \lambda(s)\,ds}$), aunque ya no permite aplicar directamente el teorema de las EDO autónomas escalares, como se advierte más adelante.
- **No se distinguen edades.** Se trabaja con tasas medias para toda la población, sin representar por separado crías y adultos. Esto no afirma que sean biológicamente iguales. Leslie distingue clases y permite estudiar cómo su composición modifica el crecimiento total. Además, en esta deducción se supone una población sin migraciones.

Con las cuatro, λ es una sola constante, nacimientos por individuo menos muertes por individuo, y el término λy dice «tantos más individuos, tantos más nacimientos y muertes, en la misma proporción». Nada más.

La ecuación es muy simple de resolver. Su solución general es $y(t) = Ce^{\lambda t}$, con C constante (de ahí el nombre). Para determinar C se impone una condición inicial; como el problema es autónomo, se puede fijar siempre en $t = 0$, y el problema de Cauchy

$$\begin{cases} y' = \lambda y, \\ y(0) = y_0 \end{cases}$$

tiene por solución maximal

$$y(t; y_0) = y_0 e^{\lambda t} \qquad \text{para todo } t \in \mathbb{R}.$$

Si $\lambda \neq 0$, el único equilibrio es $y = 0$. Si $\lambda = 0$, todas las soluciones son constantes y todo punto es un equilibrio. Para describir los conjuntos límite siguientes, suponemos $y_0 > 0$; con $y_0 = 0$, ambos conjuntos límite son $\{0\}$.

- Si $\lambda < 0$, la solución tiende a cero cuando $t \to +\infty$, y lo hace exponencialmente: el equilibrio nulo es **exponencialmente asintóticamente estable** (definición de estabilidad exponencial del apéndice; §6.3 del libro de EDO, página 143). Con la notación de los conjuntos límite, $\Lambda^-(y_0) = \emptyset$ y $\Lambda^+(y_0) = \{0\}$.
- Si $\lambda > 0$, el equilibrio nulo es **inestable**: $y(t; y_0) \to +\infty$ para todo $y_0 > 0$, por pequeña que sea la población inicial. Ahora $\Lambda^-(y_0) = \{0\}$ y $\Lambda^+(y_0) = \emptyset$.

CONEXIÓN: EL HERMANO DISCRETO DEL CAPÍTULO 1

El modelo de Malthus discreto, $N_{k+1} = \lambda_d N_k$, apareció en el capítulo 1 como la alternativa a Leslie que no distingue edades. Si se observa la solución continua en instantes $t_k = k\tau$, se obtiene $y(t_{k+1}) = e^{\lambda\tau} y(t_k)$: es el modelo discreto con $\lambda_d = e^{\lambda\tau}$. El umbral entre crecer y extinguirse es $\lambda = 0$ en el continuo y $\lambda_d = 1$ en el discreto, que es la misma frontera vista desde dos lados, y es la misma traducción «parte real frente a módulo» que se señaló al comparar $y' = Ay$ con $v(k+1) = Lv(k)$.

La ecuación $y' = \lambda y$ aparece en otros modelos clásicos de origen distinto. Aquí se recogen tres, y en los tres el trabajo de modelado consiste en identificar la variable y la ley de proporcionalidad.

Desintegración radiactiva. Los núcleos radiactivos se transforman espontáneamente emitiendo radiación. No siempre disminuye su número másico: el carbono 14, por ejemplo, se transforma en nitrógeno 14 por desintegración beta. Lo que disminuye aquí es la cantidad del *isótopo original que aún no se ha desintegrado*, no toda la materia de la muestra. El modelo supone una probabilidad de desintegración por unidad de tiempo constante para cada núcleo; para la cantidad media, o aproximadamente para una muestra grande, resulta **una velocidad de desintegración proporcional a la cantidad presente**. Llamando $y(t)$ a esa cantidad,

$$y' = -\lambda y,$$

con $\lambda > 0$ (la derivada tiene que ser negativa), y $y(t; y_0) = y_0 e^{-\lambda t}$ para todo $t \in \mathbb{R}$.

Se llama **semivida** o **periodo de semidesintegración**[1] de la sustancia al tiempo T que tarda una cantidad cualquiera en reducirse a la mitad. Si en t_0 hay $y_0 e^{-\lambda t_0}$, al cabo de T hay la mitad:

$$y_0 e^{-\lambda(t_0+T)} = \tfrac{1}{2} y_0 e^{-\lambda t_0} \quad \Longrightarrow \quad e^{-\lambda T} = \tfrac{1}{2} \quad \Longrightarrow \quad T = \frac{1}{\lambda} \log 2.$$

La semivida no depende ni de t_0 ni de la cantidad que haya en ese instante: solo de λ. Ese es el número que contesta a la pregunta del arqueólogo: se mide T en el laboratorio y con él se escribe la solución en términos con significado físico,

$$y(t; y_0) = y_0 e^{-(\log 2)\, t/T} = y_0\, 2^{-t/T} \qquad \text{para todo } t \in \mathbb{R}.$$

[1] También se encuentra con el nombre de «vida media». Conviene distinguirlo del tiempo medio de supervivencia de un núcleo, que para una ley exponencial es $1/\lambda$; la semivida es $T = (\log 2)/\lambda$.

Conocida y_0, la fórmula dice cuánta sustancia quedará en el futuro y cuánta había en el pasado (tiempos negativos): un hueso con la cuarta parte del carbono 14 que tiene un ser vivo lleva muerto dos semividas.

Ley de enfriamiento de Newton. Un cuerpo a temperatura $\theta(t)$ en un ambiente a temperatura fija τ se enfría o se calienta con una variación proporcional, en cada instante, a la diferencia entre la temperatura del medio y la del cuerpo:

$$\theta' = \lambda(\tau - \theta), \qquad \lambda > 0.$$

El signo es el correcto: si $\theta < \tau$ el cuerpo se calienta ($\theta' > 0$) y si $\theta > \tau$ se enfría. Con el cambio de variable $z = \tau - \theta$ se tiene $z' = -\lambda z$, otra vez el modelo exponencial; resolviendo y deshaciendo el cambio, $\theta(t) = \tau - Ce^{-\lambda t}$, y con $\theta(0) = \theta_0$,

$$\theta(t;\theta_0) = \tau - (\tau - \theta_0)e^{-\lambda t} \qquad \text{para todo } t \in \mathbb{R}.$$

Cuando $t \to +\infty$ todas las soluciones tienden al único equilibrio $\bar\theta = \tau$, que es de nuevo exponencialmente asintóticamente estable.

Modelo de mezclas. Un recipiente contiene un fluido que en el instante t ocupa un volumen $V(t)$; en él hay un soluto disuelto. Sea $x(t)$ la cantidad de soluto. Su variación es la diferencia entre lo que entra y lo que sale, $x'(t) = x_e(t) - x_s(t)$, y cada uno de esos términos es concentración por caudal: $x_e = c_e(t)v_e(t)$ y $x_s = c_s(t)v_s(t)$. La hipótesis simplificadora es que **el soluto está uniformemente repartido**, de modo que la concentración de salida es la del recipiente, $c_s(t) = x(t)/V(t)$. Queda

$$x'(t) = c_e(t)v_e(t) - \frac{v_s(t)}{V(t)}\,x(t), \qquad V'(t) = v_e(t) - v_s(t),$$

donde la segunda ecuación dice que el volumen cambia con la diferencia de caudales. Si $v_e \equiv v_s$ el volumen es constante, $V \equiv V_0$; si además los caudales y la concentración de entrada son constantes, $v_e = v_s = v > 0$, $V_0 > 0$ y c_e fijo,

$$x' = c_e v - \frac{v}{V_0}x = \frac{v}{V_0}\big(c_e V_0 - x\big).$$

El cambio $z = c_e V_0 - x$ da $z' = -\frac{v}{V_0}z$, y

$$x(t;x_0) = c_e V_0 - (c_e V_0 - x_0)e^{-vt/V_0} \qquad \text{para todo } t \in \mathbb{R}.$$

El único equilibrio es $\bar x = c_e V_0$, el caso en que la concentración del recipiente coincide con la de entrada, y todas las soluciones tienden a él.

LA IDEA: TRES MODELOS, UNA SOLA ECUACIÓN

Los tres son $z' = -\lambda z$ para una variable bien elegida: la cantidad que queda, la diferencia de temperaturas, el déficit de soluto. En los tres, z mide «lo que falta para el equilibrio», y la ley dice que lo que falta se reduce en proporción a lo que falta. Por eso todos convergen exponencialmente y todos tienen una constante de tiempo $1/\lambda$ (o una semivida $\log 2/\lambda$). Reconocer un modelo exponencial disfrazado es reconocer la

variable «lo que falta».

Ejercicio 2.1. *Datación por carbono 14 y una taza de café*

ORIGEN: construido a partir de los dos modelos del curso, para practicar la lectura de la semivida.

a) La semivida del carbono 14 es de 5730 años. Un hueso conserva el 23 % del carbono 14 de un organismo vivo. ¿Cuánto tiempo lleva muerto?
b) Una taza de café a 90 °C se deja en una habitación a 20 °C y a los cinco minutos está a 70 °C. ¿Cuándo estará a 40 °C?

Resolución.
La idea, que es la misma en los dos apartados. Los dos problemas parecen distintos —uno de arqueología y otro de una taza de café— y son *la misma ecuación*. En el primero decae la cantidad de carbono; en el segundo decae la **diferencia de temperatura** con la habitación, que es «lo que falta por enfriarse». En cuanto se mira esa diferencia y no la temperatura, el café obedece la misma ley que la cantidad de un isótopo radiactivo: se pierde siempre la misma *fracción* por unidad de tiempo, nunca la misma cantidad.
De ahí sale el método, y es uno solo y no dos: se escribe lo que decae como $(\text{valor inicial}) \cdot e^{-\lambda t}$; el dato intermedio sirve para despejar λ; y la pregunta se contesta tomando logaritmos. **Conocido el valor inicial, un dato posterior no degenerado basta para calibrar** λ, y esa es la razón de que estos problemas se puedan resolver con tan poca información. La comprobación que conviene hacer siempre es de orden de magnitud: contar a ojo cuántas semividas han pasado y ver que la respuesta cae donde debe.
Apartado a). La fracción que queda es $y(t)/y_0 = 2^{-t/T} = 0.23$. Tomando logaritmos, $-\frac{t}{T}\log 2 = \log 0.23$, luego

$$t = -T\,\frac{\log 0.23}{\log 2} = 5730 \cdot \frac{1.4697}{0.6931} \approx 12\,150 \text{ años.}$$

Como comprobación gruesa: dos semividas (11 460 años) dejan el 25 %, y el 23 % es un poco menos, luego un poco más de tiempo. Cuadra.
Apartado b). Con $\tau = 20$ y $\theta_0 = 90$, $\theta(t) = 20 + 70e^{-\lambda t}$. El dato $\theta(5) = 70$ da $70e^{-5\lambda} = 50$, es decir $e^{-5\lambda} = 5/7$ y $\lambda = \frac{1}{5}\log\frac{7}{5} \approx 0.0673$ por minuto. La ecuación $\theta(t) = 40$ es $70e^{-\lambda t} = 20$, luego $e^{-\lambda t} = 2/7$ y

$$t = \frac{\log(7/2)}{\lambda} = 5\,\frac{\log 3.5}{\log 1.4} \approx 5 \cdot \frac{1.2528}{0.3365} \approx 18.6 \text{ minutos.}$$

Y la misma comprobación gruesa que en (a): en cinco minutos la diferencia con la habitación ha pasado de 70 a 50, es decir, se ha reducido a poco menos de tres cuartos; para llegar a 20 hay que reducirla a 2/7. Tras tres intervalos de cinco minutos, la diferencia sería $70(5/7)^3 \approx 25.5$; tras cuatro, $70(5/7)^4 \approx 18.2$. El tiempo buscado debe estar entre 15 y 20 minutos, como el resultado 18.6.
Dónde está la trampa de este tipo de enunciado. En el café, en creer que lo que decae exponencialmente es θ. No lo es: θ tiende a 20, y una exponencial tiende a cero. Lo que decae es $\theta - 20$. Escribir la ley de Newton como $(\theta - \tau)' = -\lambda(\theta - \tau)$ y bautizar $z := \theta - \tau$ evita el

error de raíz, y es el mismo cambio de variable que convierte en una sola los tres modelos exponenciales disfrazados de este capítulo.

2.1.2 El modelo logístico

LA PREGUNTA

El modelo de Verhulst introduce un freno al crecimiento exponencial: los recursos son limitados. ¿Cuál es la corrección polinómica más sencilla que hace disminuir la tasa de crecimiento al aumentar la población, y qué nuevo número aparece con ella?

Cuando $\lambda > 0$, la solución $y_0 e^{\lambda t}$ crece hacia $+\infty$, y además exponencialmente. En la práctica, si la población es muy grande los recursos empiezan a escasear y el ritmo de crecimiento debe disminuir. Esto sugiere perfeccionar el modelo, y conviene hacerlo paso a paso, porque es el razonamiento de todo este capítulo: elegir la función f de $y' = f(y)$ por sus propiedades, no adivinarla.

1. Mientras la población sea pequeña respecto de los recursos disponibles, su crecimiento debería aproximarse por $y_0 e^{\lambda t}$.
2. Si el crecimiento en cada instante solo depende del número de individuos, el modelo es una EDO autónoma $y' = f(y)$ con f a determinar.
3. Una primera estrategia es aproximar f por su polinomio de Taylor de grado 1 en cero, $f(y) \approx f(0) + f'(0)y$. Pero $f(0) = 0$: sin individuos no hay ni nacimientos ni muertes. Luego esta estrategia devuelve el modelo exponencial con $\lambda = f'(0)$, y no aporta nada nuevo.
4. Aproximemos entonces f por su polinomio de Taylor de segundo grado, $f(y) \approx f(0) + f'(0)y + \frac{1}{2}f''(0)y^2$, con $f(0) = 0$ y $f'(0) > 0$.
5. Elegimos un término cuadrático negativo como primera forma de representar la competencia por recursos: suponemos $f''(0) < 0$. Este signo es una hipótesis del modelo, no una consecuencia del comportamiento de f lejos de cero. Poniendo $\lambda = f'(0)$ y $\kappa = -f''(0)/2$ se obtiene el **modelo logístico**:

$$y' = \lambda y - \kappa y^2, \qquad \lambda, \kappa > 0.$$

El desarrollo de Taylor motiva la elección local; usar la parábola también para poblaciones grandes es una hipótesis adicional que debe contrastarse con datos.

La tasa de crecimiento por individuo ya no es constante: vale $\lambda - \kappa y(t)$, y decrece con la población. A λ, la tasa que habría sin límite de recursos, se la llama **tasa vegetativa de crecimiento**.

Qué dice cada término y qué se pierde. El término λy es el de antes: nacimientos menos muertes, en proporción a cuántos hay. El término nuevo, $-\kappa y^2$, tiene una lectura concreta: el número de *parejas* de individuos que se pueden formar es $y(y-1)/2 \approx y^2/2$. Si además cada pareja tiene una tasa comparable de encuentros, el número de encuentros por unidad de tiempo es proporcional a y^2. Así, κy^2 representa el freno total debido a la competencia, con su intensidad recogida en κ. Esta es una posible interpretación del término cuadrático: un efecto de dos en dos. Con esa lectura se ve lo que se ha despreciado. Primero, que el freno actúa *en*

el acto: la falta de recursos de hoy reduce los nacimientos de hoy, cuando en muchas especies reduce los de la temporada siguiente; los modelos con retardo, $y' = \lambda y(1 - \kappa y(t-\tau)/\lambda)$, pueden producir oscilaciones, que el logístico autónomo escalar no produce. Segundo, que con pocos individuos no pasa nada malo: en el logístico la tasa es máxima cuando $y \to 0$, mientras que muchas especies necesitan una densidad mínima para encontrar pareja o defenderse, y por debajo de ella decrecen. Ese *efecto Allee fuerte* corresponde a una tasa per cápita negativa a densidades bajas; se estudia en la práctica 6 de este capítulo. No se caracteriza, en general, solo por el signo de $f''(0)$. Y tercero, sigue sin haber edades ni estaciones. El modelo se queda con una sola corrección, la de segundo orden, y a cambio se resuelve a mano y se lee de un vistazo.

EL CAMINO QUE NO SE TOMÓ: OTRAS FORMAS DE FRENAR EL CRECIMIENTO

La alternativa. Cualquier f con $f(0) = 0$, positiva para poblaciones pequeñas y negativa para poblaciones grandes, frena el crecimiento. Se podría haber tomado, por ejemplo, $f(y) = \lambda y \log(K/y)$, que es el modelo de Gompertz de la subsección 2.1.3, o un término cúbico, o una tasa que decae exponencialmente.
Su ventaja. Algunas de esas funciones ajustan mejor datos reales; Gompertz, en concreto, se utiliza para ajustar ciertos datos de crecimiento de tumores y peces, y por eso tiene subsección propia y reaparece en los modelos de pesca.
Por qué se eligió la parábola. Porque es el término *siguiente* en el desarrollo de Taylor de una f desconocida, y por tanto la corrección mínima que se puede hacer dentro de esta aproximación cuadrática. Su interpretación requiere las hipótesis anteriores; además, se resuelve exactamente. Cuando se sabe algo más de la especie, se sustituye la parábola por la f adecuada, y el teorema de la sección anterior sigue diciendo cómo se comportan las soluciones sin necesidad de resolver.

Resolución. La ecuación logística es fácil de resolver, como EDO de variables separadas o, más fácil aún, como **EDO de Bernoulli**. Dividiendo por y^2,

$$\frac{y'}{y^2} = \frac{\lambda}{y} - \kappa,$$

y con el cambio $z = 1/y$, que da $z' = -y'/y^2$, se obtiene $z' + \lambda z = \kappa$, una EDO lineal de primer orden. Se resuelve multiplicando por el factor integrante $e^{\lambda t}$:

$$\left(ze^{\lambda t}\right)' = \kappa e^{\lambda t} \quad \Longrightarrow \quad z = \frac{\kappa}{\lambda} + Ce^{-\lambda t}.$$

Volviendo a $y = 1/z$, la solución general es

$$y(t) = \frac{\lambda}{\kappa + C\lambda e^{-\lambda t}}.$$

Imponiendo $y(0) = y_0$ se determina C. Para datos $y_0 \geq 0$, la solución maximal es

$$y(t; y_0) = \frac{\lambda y_0}{\kappa y_0 + (\lambda - \kappa y_0)e^{-\lambda t}} \qquad \text{para todo } t \in I(y_0),$$

donde

$$I(y_0) = \begin{cases} \left(-\frac{1}{\lambda}\log\frac{\kappa y_0}{\kappa y_0 - \lambda},\, +\infty\right) & \text{si } y_0 > \lambda/\kappa, \\ \mathbb{R} & \text{si } 0 \leq y_0 \leq \lambda/\kappa. \end{cases}$$

EXPLÍCATE. *Comprueba de dónde sale el intervalo de definición: el denominador $\kappa y_0 + (\lambda - \kappa y_0)e^{-\lambda t}$ se anula en un t negativo exactamente cuando $\lambda - \kappa y_0 < 0$. Resuelve esa ecuación y verás la fórmula.*

Observación. La solución general no contiene la solución particular $y \equiv 0$, porque se dividió por y. Sin embargo esa función sí aparece en la fórmula de la solución maximal, que para $y_0 = 0$ da $y \equiv 0$.

De la fórmula se deduce

$$\lim_{t\to+\infty} y(t; y_0) = \begin{cases} \lambda/\kappa & \text{si } y_0 > 0, \\ 0 & \text{si } y_0 = 0, \end{cases}$$

y que 0 y λ/κ son los únicos puntos críticos. Este resultado se obtiene también, *sin resolver*, del teorema de las EDO autónomas escalares: $f(y) = \lambda y - \kappa y^2 = \kappa y(\lambda/\kappa - y)$ es positiva en $(0, \lambda/\kappa)$ y negativa a la derecha, luego si $0 < y_0 < \lambda/\kappa$ la solución es estrictamente creciente y tiende al menor punto crítico por encima, λ/κ; y si $y_0 > \lambda/\kappa$ es estrictamente decreciente y tiende al mayor punto crítico por debajo, que es el mismo. La figura 2.2 muestra varias trayectorias. Contrariamente al modelo exponencial, el logístico conduce a un estado de equilibrio cuando $t \to +\infty$: si $y_0 > 0$, sea cual sea, la población al cabo de un tiempo grande es aproximadamente λ/κ. (El enunciado se restringe a $y_0 \geq 0$, que es lo único que significa algo aquí. Un dato negativo tiene $f < 0$ y ningún punto crítico por debajo, así que la solución se va a $-\infty$, y además en tiempo finito.)

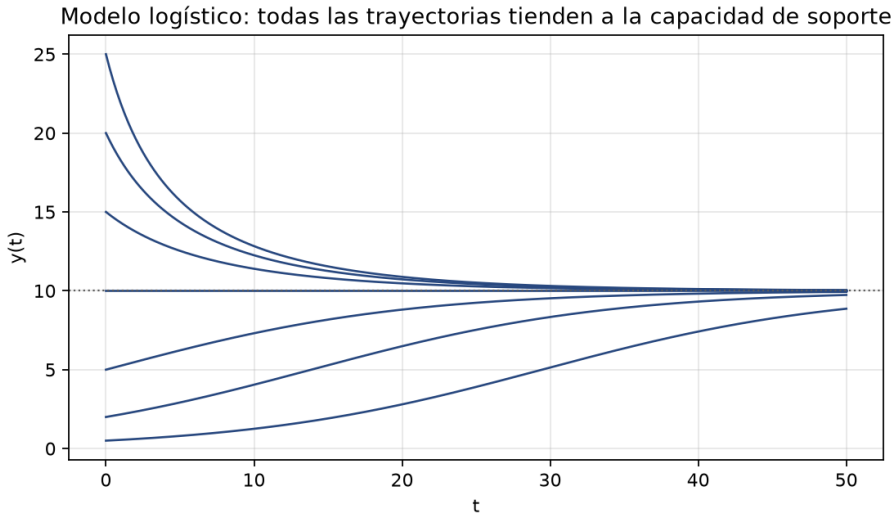


Figura 2.2: Trayectorias de la ecuación logística con $\lambda = 0.1$ y $\kappa = 0.01$. La capacidad de soporte es $\lambda/\kappa = 10$ (línea de puntos). Las que salen por encima bajan hacia ella, y las que salen entre 0 y 10 suben. Para $t \geq 0$, la forma de S solo la tienen las que arrancan entre 0 y $\lambda/(2\kappa) = 5$, que es donde está el punto de inflexión: la que sale justo de 5 ya empieza a frenar, y se ve que es cóncava desde el primer instante.

Observación. A la constante λ/κ se la llama **capacidad de soporte del medio**: es la población de equilibrio sostenible en este modelo, no un techo instantáneo: puede haber poblaciones iniciales mayores. Otra forma de obtener el modelo consiste en suponer que existe un umbral c_M a partir del cual hay más muertes que nacimientos, y que el modelo es una EDO autónoma $y' = f(y)$. Entonces f debe verificar $f(0) = f(c_M) = 0$, $f > 0$ en $(0, c_M)$ y $f < 0$ en $(c_M, +\infty)$,

y la función más simple con esas propiedades es $f(y) = ay(c_M - y)$ con $a > 0$: es la logística con $a = \kappa$ y $c_M = \lambda/\kappa$. El mismo modelo se usa si, en lugar de una población, se modela la transmisión de una noticia en un colectivo.

Observación. Los dos modelos dan valores parecidos mientras la corrección logística sea pequeña. La condición precisa se ve en la fórmula reescrita como

$$y(t; y_0) = \frac{\lambda y_0}{\lambda + \kappa y_0 (e^{\lambda t} - 1)}\, e^{\lambda t},$$

de donde $y(t) \approx y_0 e^{\lambda t}$ mientras $\kappa y_0 (e^{\lambda t} - 1) \ll \lambda$. Y una limitación de los dos modelos: la ley de crecimiento no depende explícitamente del tiempo, aunque $y'(t)$ sí cambie al cambiar la población. En muchas poblaciones la reproducción se concentra en ciertas épocas del año, lo que lleva a EDOs no autónomas $y' = f(t, y)$ con f periódica en t, cuyo análisis es mucho más complicado.

DE DÓNDE VIENE: VERHULST Y LA PALABRA «LOGÍSTICA»

Pierre-François Verhulst publicó la ecuación en 1838 y le puso el nombre de «logística» en 1845, sin explicar por qué. Ajustó con ella los censos de Francia, Bélgica y Essex, y predijo para Bélgica una población límite de unos 6.6 millones que la realidad superó de largo: la capacidad de soporte no es una constante de la naturaleza sino de la técnica de cada época. La ecuación cayó en el olvido y la redescubrieron Pearl y Reed en 1920 para la población de Estados Unidos. Hoy la curva en S está en todas partes: crecimiento de bacterias, adopción de tecnologías, y la misma función logística $1/(1 + e^{-x})$ es la que convierte en probabilidad la salida de un modelo de regresión logística y de las redes neuronales.

Un primer modelo de epidemias. Entre otros fenómenos que se modelan con EDOs logísticas destaca una epidemia en una población dividida en dos tipos de individuos: los **susceptibles** de contraer la enfermedad en el instante t, en número $S(t)$, y los que la **padecen**, en número $I(t)$. Los susceptibles que se contagian pasan a ser infecciosos y los infecciosos que se recuperan vuelven a ser susceptibles: el esquema es $S \to I \to S$ y el modelo se denota **SIS**. Las hipótesis son tres:

- la **tasa de contagio** es proporcional al número de susceptibles, con constante $\beta > 0$: cada infeccioso genera nuevos contagios a una tasa $\beta S(t)$, antes de restar las recuperaciones;
- la **velocidad de recuperación** es proporcional al número de infecciosos, con constante $\gamma > 0$: en I' aparece un término $-\gamma I(t)$;
- la población es **cerrada**: el número total N permanece constante, sin muertes ni nacimientos.

El modelo SIS se formula entonces como

$$\begin{cases} I' = \beta IS - \gamma I, \\ S' = -\beta IS + \gamma I. \end{cases}$$

Otra manera de entender los términos $\pm\beta IS$ es aceptar que la variación del número de susceptibles decrece en proporción al número de *encuentros* entre susceptibles e infecciosos, que es proporcional al producto IS.

Las hipótesis, una a una, y qué se pierde con cada una. Que los encuentros sean proporcionales a IS es la hipótesis de **mezcla homogénea**: cualquier infeccioso tiene la misma probabilidad de cruzarse con cualquier susceptible, como moléculas en un gas. Se pierde la

estructura de contactos (familias, colegios, la persona que ve a mil al día y la que ve a tres). Una red de contactos permite incorporar esas diferencias. Comparte con PageRank el uso de grafos, pero las ecuaciones de contagio son distintas y, en general, no lineales. Que la recuperación sea $-\gamma I$ es suponer que cada infeccioso se cura con la misma probabilidad por unidad de tiempo, **sin memoria** de cuánto lleva enfermo: la duración de la fase infecciosa queda distribuida exponencialmente, con media $1/\gamma$; una duración más concentrada o dependiente del tiempo ya transcurrido exige otro supuesto. Que el recuperado vuelva a ser susceptible es la hipótesis de **no inmunidad**; si pasa a una clase de recuperados inmunes durante el periodo estudiado, se obtiene el modelo SIR de la sección 2.2.1. Y que N sea constante es **población cerrada** durante la epidemia: sin nacimientos, muertes ni viajes relevantes en la escala temporal estudiada. Estas hipótesis explican los términos del sistema y orientan posibles refinamientos.

Al sumar las ecuaciones, $S' + I' \equiv 0$, luego $S + I \equiv N$ y $S \equiv N - I$: el sistema se reduce a una **EDO logística**,

$$I' = \beta I(N - I) - \gamma I = \beta I\big((N - \gamma/\beta) - I\big) =: f(I).$$

Los equilibrios son $I_1 = 0$ (con $S_1 = N$) e $I_2 = N - \gamma/\beta$ (con $S_2 = \gamma/\beta$). En Epidemiología, la cantidad

$$R_0 = \frac{N\beta}{\gamma}$$

se llama **número básico de reproducción** de la infección. Cuenta los contagios secundarios esperados que causa un infeccioso en una población completamente susceptible. En este modelo determina si, partiendo de $I_0 > 0$, la infección desaparece o se vuelve **endémica**, es decir, si los infecciosos tienden o no a cero.

LA IDEA: QUÉ CUENTA R_0

Un infeccioso contagia a razón de $\beta S \approx \beta N$ personas por unidad de tiempo al principio de la epidemia, cuando casi todos son susceptibles, y permanece infeccioso un tiempo medio $1/\gamma$. Luego $R_0 = \beta N/\gamma$ es el número de contagios que produce en promedio un solo infectado introducido en una población susceptible. Un valor menor que uno favorece el descenso inicial; uno mayor que uno, el crecimiento inicial. Es una interpretación media, no una garantía sobre cada cadena individual de contagios. El teorema de las EDO autónomas escalares va a convertir esa cuenta de servilleta en un teorema.

Con el teorema de las EDO autónomas escalares se analiza la solución maximal $I(t; I_0)$ del problema de Cauchy con $I(0) = I_0 \in (0, N)$. Todo depende del signo de $N - \gamma/\beta$, es decir, de si $R_0 \leq 1$ o $R_0 > 1$.

- Si $N - \gamma/\beta \leq 0$ (esto es, $R_0 \leq 1$), el único equilibrio relevante es $\bar{I}_1 = 0$ y $f(I) = \beta I((N - \gamma/\beta) - I) < 0$ para todo $I \in (0, N)$. Luego $I(t; I_0)$ está definida y es decreciente en $[0, +\infty)$ y tiende a 0; la correspondiente S crece hacia N. Es una **epidemia transitoria** que desaparece.
- Si $N - \gamma/\beta > 0$ ($R_0 > 1$), hay tres situaciones según I_0 sea menor, igual o mayor que $\bar{I}_2$: en la primera I crece, en la segunda es constante y en la tercera decrece. En todos los casos está definida en $[0, +\infty)$ y

$$\lim_{t \to +\infty} I(t; I_0) = \bar{I}_2 = N - \gamma/\beta.$$

La epidemia **persiste y se vuelve endémica**.

Ejercicio 2.2. *Análisis del modelo SIS por linealización*

ORIGEN: análisis teórico de la práctica 9 de la hoja del tema 2.

Para $f(I) = I(\beta N - \gamma - \beta I)$, calcular f' en los dos equilibrios y clasificar su estabilidad en función de R_0, incluido el caso $R_0 = 1$. Comparar con lo que da el teorema de las EDO autónomas escalares.

PISTAS, DE MENOR A MAYOR

1. *El criterio escalar de linealización es el caso $N = 1$ del teorema de estabilidad por linealización del apéndice (libro de EDO, §6.4 «Primer método de Lyapunov», página 146): $f'(\bar{y}) < 0$ da estabilidad asintótica exponencial y $f'(\bar{y}) > 0$ da inestabilidad.*
2. *En $R_0 = 1$ la derivada se anula y el criterio no decide: mira el signo de f a los dos lados.*

Resolución.

La idea. f es una parábola hacia abajo que corta al eje en 0 y en $\bar{I}_2 = N - \gamma/\beta$, y *toda* la respuesta está en ese dibujo. Una parábola hacia abajo sube al pasar por su primera raíz y baja al pasar por la segunda, así que las derivadas en los dos cortes tienen signos opuestos *si las raíces son distintas*: **uno de los dos equilibrios es estable y el otro inestable**. Lo único que decide R_0 es *cuál de los dos está a la derecha*, es decir, cuál de los dos ceros es el positivo. Ahí está la imagen que hay que llevarse: al pasar R_0 por 1 los dos ceros *se cruzan*, y al cruzarse se intercambian la estabilidad. El caso $R_0 = 1$ es justo el instante del cruce, cuando los dos son el mismo punto: la parábola se ha convertido en $-\beta I^2$, que toca el eje sin atravesarlo. Por eso ahí la linealización no puede decidir —no hay cambio de pendiente que mirar— y hay que mirar el signo de f, que es lo que el teorema escalar sí ve.

Paso 1 (el núcleo): las dos derivadas, que salen con signos opuestos. $f'(I) = \beta N - \gamma - 2\beta I$, luego $f'(0) = \beta N - \gamma$ y $f'(\bar{I}_2) = \beta N - \gamma - 2(\beta N - \gamma) = \gamma - \beta N$. Los dos tienen signos opuestos cuando $R_0 \neq 1$: en ese caso, exactamente uno de los equilibrios es exponencialmente asintóticamente estable y el otro inestable.

Paso 2: los tres casos, según de qué lado del cruce se esté.

- Si $\beta N - \gamma < 0$ ($R_0 < 1$): $f'(0) < 0$ y el equilibrio 0 es exponencialmente asintóticamente estable; $\bar{I}_2$ es inestable, pero además es negativo y no interesa.
- Si $\beta N - \gamma > 0$ ($R_0 > 1$): 0 es inestable y $\bar{I}_2 = N - \gamma/\beta > 0$ es exponencialmente asintóticamente estable.
- Si $\beta N = \gamma$ ($R_0 = 1$): la ecuación queda $I' = -\beta I^2$, cuyo único equilibrio es 0, con $f'(0) = 0$. La linealización no decide, pero $f(I) = -\beta I^2 < 0$ para $I \neq 0$: por el teorema de las EDO autónomas escalares toda solución con $I_0 > 0$ decrece hacia 0. El equilibrio es asintóticamente estable **dentro del dominio físico** $[0, N]$, aunque no exponencialmente: resolviendo, $I(t) = I_0/(1 + \beta I_0 t)$, que decae como $1/t$ para $I_0 > 0$. Si se admitieran datos negativos, se alejarían de 0: en toda la recta el equilibrio es semiestable, no estable.

Paso 3: comparación con el teorema escalar, que es lo que el enunciado pide de verdad. La linealización da información local (cerca del equilibrio) y de velocidad (exponencial); el teorema de la sección anterior da información global (desde cualquier $I_0 \in [0, N]$) pero no de velocidad. En una dimensión los dos coinciden en el veredicto salvo en el caso degenerado $f' = 0$, donde manda el signo de f.

Dónde se falla.

- *Al dar por bueno el equilibrio negativo.* Con $R_0 < 1$ el segundo cero, $\bar{I}_2 = N - \gamma/\beta$, es negativo: es una raíz de f, pero no un estado de la epidemia. Decir «hay dos equilibrios, uno estable y otro inestable» sin mirar el signo de $\bar{I}_2$ describe una población que no existe.
- *Al sacar una conclusión de $f'(\bar{y}) = 0$.* En $R_0 = 1$ la derivada se anula y la linealización no dice nada. Escribir «$f'(0) = 0$, luego es estable» o «luego es inestable» son dos errores distintos con la misma causa; lo que decide es el signo de f a cada lado.
- *Al olvidar el dominio.* En $R_0 = 1$ el origen atrae a los datos de $[0, N]$ y no a los negativos. La estabilidad se afirma dentro del dominio físico, y hay que decirlo.
- *Al comparar β con γ.* El umbral es $\beta N = \gamma$, no $\beta = \gamma$: el número de susceptibles entra en la cuenta. Con los datos de la práctica 9, $\beta = 0.001$ es ochenta veces menor que $\gamma = 0.08$, y aun así $R_0 = 200 \cdot 0.001/0.08 = 2.5$ y la infección se hace endémica.

QUÉ DEJA. En una EDO escalar, la estabilidad de un equilibrio se lee en $f'(\bar{y})$, y cuando $f'(\bar{y}) = 0$ en el signo de f a cada lado. El umbral $R_0 = 1$ es donde los dos equilibrios se cruzan e intercambian su estabilidad.

2.1.3 El modelo de Gompertz

LA PREGUNTA

El logístico frena el crecimiento con un término cuadrático y produce una curva en ese simétrica respecto de su punto de inflexión: si este ocurre en t_* y la capacidad es K, entonces $y(t_* + s) + y(t_* - s) = K$. Algunos datos de crecimiento presentan una curva asimétrica. ¿Cómo cambiar la tasa para representar esa asimetría, y qué se gana y qué se pierde al hacerlo?

Hay una tercera ley de crecimiento, y las tres se distinguen mirando **la tasa per cápita** y'/y, es decir, cuánto crece cada individuo por unidad de tiempo:

$$\underbrace{\frac{y'}{y} = \lambda}_{\text{exponencial}}, \qquad \underbrace{\frac{y'}{y} = \lambda - \kappa y}_{\text{logístico}}, \qquad \underbrace{\frac{y'}{y} = \lambda \log \frac{K}{y}}_{\text{Gompertz}}.$$

En el exponencial la tasa no baja nunca; en el logístico baja *linealmente* con la población; en el de Gompertz baja *logarítmicamente*. Escrito como EDO autónoma,

$$y' = \lambda y \log \frac{K}{y}, \qquad \lambda, K > 0,$$

y esta es la diferencia que lo cambia todo: el logaritmo se va a $+\infty$ cuando $y \to 0^+$, así que con muy pocos individuos la tasa per cápita es grande, aunque el crecimiento total $\lambda y \log(K/y)$ tienda a cero. Cerca de K, en cambio, $\log(K/y) \approx (K - y)/K$: la aproximación al equilibrio es exponencial, como en el logístico. La diferencia está en la forma de toda la trayectoria.

LA IDEA: DÓNDE ESTÁ EL PUNTO DE INFLEXIÓN, Y POR QUÉ ESO ES TODO

La curva de crecimiento tiene su punto de inflexión donde y' es máxima, es decir, donde la población produce más individuos por unidad de tiempo. En el logístico, $y' = \lambda y - \kappa y^2$ es una parábola con vértice en $y = \lambda/(2\kappa)$: **la mitad** de la capacidad de soporte, y por eso la curva es simétrica. En Gompertz, derivando $y\log(K/y)$ sale $\log(K/y) - 1$, que se anula en $y = K/e$: el punto de inflexión está en el $1/e \approx 36.8$ % de la capacidad, *antes* de la mitad. Toda la asimetría de la curva de Gompertz cabe en ese cambio de $1/2$ a $1/e$: sube antes, y luego le queda casi dos tercios del camino para frenar.

Dónde está definido, y por qué importa. El dominio es $(0, +\infty)$: el logaritmo lo exige, y por eso $y = 0$ **no es un punto crítico de este modelo**, al contrario de lo que pasaba en el exponencial y en el logístico. Se puede extender el segundo miembro a $y = 0$ por continuidad, con valor 0, porque $y\log(K/y) \to 0$; pero la extensión **no es localmente lipschitziana** ahí, ya que $\lambda(\log(K/y) - 1) \to +\infty$. En $y = 0$ no se puede aplicar Picard, y conviene tenerlo presente porque es el error clásico de este modelo.

Equilibrio y estabilidad. En $(0, +\infty)$ el único cero de $f(y) = \lambda y \log(K/y)$ es $y = K$, porque el factor y no se anula en el dominio. Y

$$f'(y) = \lambda\Big(\log\frac{K}{y} - 1\Big), \qquad f'(K) = -\lambda < 0,$$

luego K es exponencialmente asintóticamente estable. Con el teorema de las EDO autónomas escalares, aplicado en el intervalo abierto $(0, +\infty)$: si $0 < y_0 < K$ la solución crece hacia K, y si $y_0 > K$ decrece hacia K. La capacidad de soporte sigue siendo un atractor global del semieje positivo, como en el logístico; lo que ha cambiado es *cómo* se llega a ella.

Resolución: la propiedad que hace especial a Gompertz. Esta ecuación no lineal se convierte en **lineal** con el cambio $z = \log y$: medir la población en escala logarítmica. Es otra forma de simplificar una EDO mediante un cambio de variable, como $z = 1/y$ en el modelo logístico:

$$z' = \frac{y'}{y} = \lambda(\log K - z) \implies z' + \lambda z = \lambda \log K,$$

una EDO lineal de primer orden con coeficientes constantes. Con el factor integrante $e^{\lambda t}$ sale $z(t) = \log K + Ce^{-\lambda t}$, y de $z(0) = \log y_0$, $C = \log(y_0/K)$. Deshaciendo el cambio,

$$y(t) = K\exp\Big(\log\frac{y_0}{K}\,e^{-\lambda t}\Big) = K\Big(\frac{y_0}{K}\Big)^{e^{-\lambda t}}.$$

Es una **doble exponencial** —una exponencial en el exponente—, y esa es la firma de Gompertz. Como $e^{-\lambda t} \to 0$, el exponente tiende a 0 y $y(t) \to K$ para todo $y_0 > 0$. Si $y_0 \neq K$, hacia atrás el exponente diverge y la solución tiende a 0 cuando $y_0 < K$, o a $+\infty$ cuando $y_0 > K$, sin alcanzarlos en tiempo finito. Si $y_0 = K$, es constante. En los tres casos, la solución maximal está definida en todo $\mathbb{R}$, igual que la logística que sale por debajo de la capacidad.

EL CAMINO QUE NO SE TOMÓ: POR QUÉ NO SIEMPRE GOMPERTZ

La alternativa. Usar Gompertz en lugar del logístico en todos los modelos de población de este capítulo.
Su ventaja. Su asimetría puede ajustar mejor determinados datos de crecimiento; se resuelve exactamente, como el logístico. Elegir entre ambos requiere contrastarlos con los datos del problema.
Qué la aparta. La tasa logarítmica es una hipótesis diferente, no una mejora automática. También admite una interpretación: si $g = y'/y$, la ecuación implica $g' = -\lambda g$, de modo que la tasa per cápita decae exponencialmente hacia cero. Ambos modelos sirven para interpretar y ajustar, bajo supuestos distintos. Además, con parámetros positivos y población inicial positiva, ninguno de los dos modelos básicos describe una extinción: ambos convergen a su capacidad de soporte. Si se añade extracción constante a Gompertz, hay que vigilar especialmente el borde $y = 0$ de su dominio.

DE DÓNDE VIENE: UN ACTUARIO EN 1825

Benjamin Gompertz era matemático y actuario de la Alliance Assurance de Londres, y en 1825 estudiaba **tablas de mortalidad**. Su hipótesis de una fuerza de mortalidad ae^{bt}, con $a, b > 0$, conduce a una supervivencia $S(t) = \exp[-(a/b)(e^{bt} - 1)]$, normalizada por $S(0) = 1$. También es una doble exponencial, pero decreciente: no es la misma fórmula que la curva de crecimiento anterior. La relación entre ambas muestra cómo una familia de funciones puede aparecer en fenómenos distintos sin que sus interpretaciones sean intercambiables.

2.1.4 Modelos de pesca

LA PREGUNTA

Una cofradía de pescadores quiere fijar cuántas toneladas sacar cada año de un caladero. Si sacan pocas, pierden dinero; si sacan demasiadas, dentro de unos años no habrá peces. ¿Hay una cantidad máxima que se puede pescar cada año indefinidamente, y qué pasa exactamente el año en que se supera?

En los modelos anteriores la variación de la población depende solo del número de individuos. Se puede pensar, sin embargo, en agentes externos que influyan en la evolución: la pesca, o la caza, donde una cierta cantidad de individuos desaparece por la acción del hombre. El nuevo modelo completa los anteriores con un término que resta los individuos eliminados. Según se parta del exponencial o del logístico para la población $p = p(t)$ de peces:

$$p' = \lambda p - h(t, p) \qquad \text{o} \qquad p' = \lambda p - \kappa p^2 - h(t, p),$$

donde $h(t, p)$ es la cantidad de peces que se pesca por unidad de tiempo en el instante t en presencia de p peces. El segundo incorpora competencia por recursos. Si $\kappa p \ll \lambda$, los términos de crecimiento de ambos son próximos, aunque esa aproximación puede cambiar umbrales y predicciones a largo plazo. Se supone $\lambda > 0$ para estudiar una población capaz de crecer sin pesca a baja densidad y se denota $p(t; p_0)$ la solución maximal con $p(0) = p_0 \geq 0$, con intervalo $I(p_0)$.

Qué se está suponiendo. El término $-h(t, p)$ es lo único nuevo, y es un término de naturaleza distinta a los otros: no sale de la biología sino de una *decisión*. Los peces no eligen λ ni κ;

la cofradía sí elige h. Dos elecciones son las que se estudian: una cuota fija, h constante, que es lo que se legisla, y un esfuerzo fijo, $h = \mu p$, que es lo que hace una flota que sale siempre con los mismos barcos y pesca en proporción a lo que encuentra. Lo que se pierde en las dos: que pescar tiene un coste y un precio, que cambian con el tiempo y con p (eso es economía, y se recupera en el capítulo 4 como un problema de control con $\mu(t)$ como variable); que los peces tienen edades y las redes seleccionan tamaños (eso es Leslie, capítulo 1); y que la cuota fija h no puede seguir sacándose cuando ya no queda nada, que es el aviso que el modelo da por sí solo al producir $f(0) = -h \neq 0$.

Pesca constante. Se supone $h(t, p) \equiv h$, una constante positiva: cada año se saca la misma cantidad.

Primer caso, modelo exponencial: $p' = \lambda p - h$. Tiene un único punto crítico $\bar{p} = h/\lambda$. El modelo no es completamente correcto, porque $f(p) = \lambda p - h$ no se anula en $p = 0$: se le da validez solo mientras la solución sea positiva. El teorema de las EDO autónomas escalares dice que si $p_0 > h/\lambda$ la solución crece a $+\infty$ (y sería más adecuado usar el logístico), y que si $0 < p_0 < h/\lambda$ tiende a $-\infty$, lo que se interpreta como que **la población se extingue en tiempo finito**. La ecuación se resuelve con el factor integrante $e^{-\lambda t}$: toda solución es $p(t) = h/\lambda + Ce^{\lambda t}$, y con $p(0) = p_0$,

$$p(t; p_0) = \frac{h}{\lambda} + \Big(p_0 - \frac{h}{\lambda}\Big)e^{\lambda t} \qquad \text{para todo } t \in \mathbb{R}.$$

Si $0 < p_0 < h/\lambda$, la solución se anula en

$$T^* = \frac{1}{\lambda} \log \frac{h}{h - p_0\lambda},$$

y para $t > T^*$ es negativa: la población desaparece en ese momento. Con $\lambda = 0.1$, $h = 1$ y $p_0 = 5$, por ejemplo, $T^* = 10 \log 2 \approx 6.93$.

Segundo caso, modelo logístico: $p' = \lambda p - \kappa p^2 - h$. Ahora se aplica el teorema de las EDO autónomas escalares en vez de resolver, que es más laborioso. Los puntos críticos son las raíces de $\kappa p^2 - \lambda p + h = 0$,

$$p = \frac{\lambda \pm \sqrt{\lambda^2 - 4\kappa h}}{2\kappa},$$

y su existencia y número dependen del signo del discriminante.

1. Si $\lambda^2 - 4\kappa h < 0$ no hay equilibrios y $f(p) < 0$ siempre: sea cual sea p_0, la solución tiende a $-\infty$ y la población desaparece en tiempo finito.
2. Si $\lambda^2 - 4\kappa h = 0$ hay un único punto crítico $\bar{p} = \lambda/(2\kappa) > 0$, con $f(p) < 0$ para $p \neq \bar{p}$. Si $p_0 > \bar{p}$ la población decrece y converge a $\bar{p}$; si $0 < p_0 < \bar{p}$ tiende a $-\infty$ y desaparece en tiempo finito.
3. Si $\lambda^2 - 4\kappa h > 0$ hay dos puntos críticos positivos,

   $$0 < \bar{p}_1 = \frac{\lambda - \sqrt{\lambda^2 - 4\kappa h}}{2\kappa} < \bar{p}_2 = \frac{\lambda + \sqrt{\lambda^2 - 4\kappa h}}{2\kappa},$$

   con $f < 0$ en $[0, \bar{p}_1)$, $f > 0$ en $(\bar{p}_1, \bar{p}_2)$ y $f < 0$ para $p > \bar{p}_2$. Por tanto: si $0 < p_0 < \bar{p}_1$ la población desaparece en tiempo finito; si $\bar{p}_1 < p_0 < \bar{p}_2$ crece hacia $\bar{p}_2$; y si $p_0 > \bar{p}_2$ decrece hacia $\bar{p}_2$. El equilibrio $\bar{p}_1$ es inestable y $\bar{p}_2$ es asintóticamente estable; de hecho, aplicando el teorema de estabilidad por linealización (linealización; libro de EDO, §6.4 «Primer método de Lyapunov») con $f'(\bar{p}_2) < 0$, $\bar{p}_2$ es exponencialmente asintóticamente estable.

Si el dato inicial coincide con cualquiera de los equilibrios, la solución es constante. En los casos de extinción, el modelo solo se usa hasta alcanzar $p = 0$. La llegada ocurre en tiempo finito porque, cerca de cero, $f(p) \to -h < 0$: la velocidad de descenso no se anula al aproximarse a la frontera.

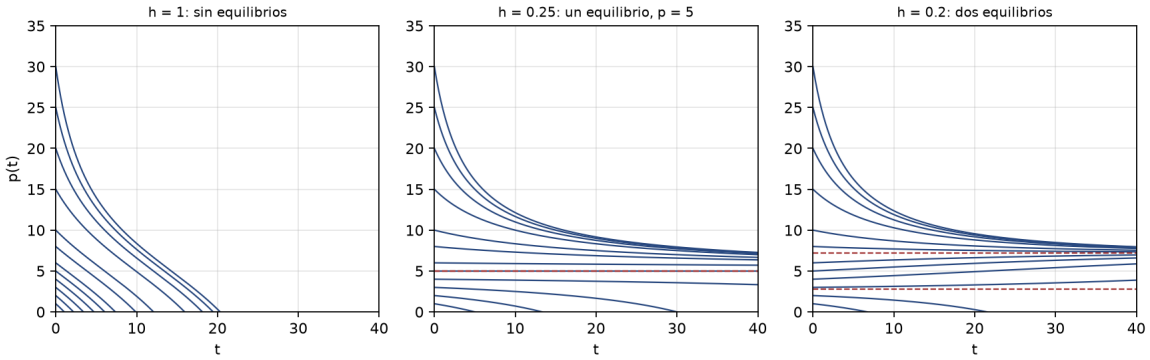


Figura 2.3: Pesca constante sobre el modelo logístico con $\lambda = 0.1$ y $\kappa = 0.01$, en los tres casos característicos. Con $h = 1$ no hay equilibrios y toda población se extingue en tiempo finito. Con $h = 0.25$ el discriminante se anula y hay un solo equilibrio $\bar{p} = 5$: las que salen por encima bajan hacia él, las que salen por debajo se extinguen. Con $h = 0.2$ hay dos, $\bar{p}_1 \approx 2.7639$ y $\bar{p}_2 \approx 7.2361$; el primero es la frontera entre extinción y supervivencia.

La idea: la cantidad máxima sostenible

La respuesta a la cofradía está en el discriminante. Una cuota $h > 0$ admite un equilibrio positivo si y solo si $h \leq \lambda^2/(4\kappa)$. Eso permite sostenerla para poblaciones iniciales adecuadas; no garantiza supervivencia desde cualquier p_0. Ese valor, $h_{\text{máx}} = \lambda^2/(4\kappa)$, es el *máximo rendimiento sostenible*: es el máximo de la parábola $\lambda p - \kappa p^2$, lo más que la población puede producir por unidad de tiempo, y se alcanza en $p = \lambda/(2\kappa)$, la mitad de la capacidad de soporte. Pescar justo $h_{\text{máx}}$ es vivir en el filo: el único equilibrio $\bar{p} = \lambda/(2\kappa)$ es semiestable, y cualquier perturbación hacia abajo (un año malo) lleva a la extinción en tiempo finito. El modelo aconseja dejar margen por debajo del máximo: con $0 < h < h_{\text{máx}}$ aparece un colchón, el intervalo $(\bar{p}_1, \bar{p}_2)$, dentro del cual la población se recupera sola.

El camino que no se tomó: pescar en proporción a lo que hay

La alternativa. En lugar de una cuota fija, un esfuerzo fijo: $h = \mu p$ con $\mu > 0$. Aquí se llama **pesca de arrastre**: la extracción por unidad de tiempo es proporcional a la población. Matemáticamente lo esencial es esa proporcionalidad, no el tipo concreto de arte de pesca.
Su ventaja. Los modelos quedan

$$p' = (\lambda - \mu)p \qquad \text{y} \qquad p' = (\lambda - \mu)p - \kappa p^2,$$

que son el exponencial y el logístico con λ sustituido por $\lambda - \mu$. No hay extinción en tiempo finito desde $p_0 > 0$. En el exponencial, la población crece sin límite si $\mu < \lambda$, permanece constante si $\mu = \lambda$ y tiende a cero si $\mu > \lambda$. En el logístico, tiende a $(\lambda - \mu)/\kappa$ si $\mu < \lambda$ y a cero si $\mu \geq \lambda$; en la igualdad, $p' = -\kappa p^2$ y el descenso es del orden de $1/t$. Cuando hay pocos peces se pescan pocos, pero eso no evita la extinción asintótica si el esfuerzo es demasiado alto.
Qué la aparta. No es que sea peor; es que responde a otra política. La cuota fija es lo que

se legisla (tantas toneladas al año) y la que produce la catástrofe del discriminante, que es la lección del modelo. Y hay una variante todavía más interesante, $\mu = \mu(t)$: elegir bien la estrategia de pesca en el tiempo es un problema de gran interés teórico y aplicado, y es exactamente el tipo de problema del capítulo 4, el control óptimo, donde $\mu(t)$ es el control y la ganancia acumulada es el coste.

Las dos políticas sobre Gompertz, y el rendimiento máximo sostenible. Todo lo anterior se ha hecho sobre el logístico, pero la ley de crecimiento es una elección del modelador, y conviene ver qué sobrevive al cambiarla. Sobre el modelo de Gompertz de la subsección 2.1.3 las dos políticas son

$$y' = \lambda y \log \frac{K}{y} - h \qquad \text{(cuota fija)}, \qquad y' = \lambda y \log \frac{K}{y} - qy \qquad \text{(esfuerzo fijo)},$$

y las dos se resuelven mirando una sola función,

$$g(y) := y \log \frac{K}{y}, \qquad g'(y) = \log \frac{K}{y} - 1, \qquad g''(y) = -\frac{1}{y} < 0.$$

La forma de g decide la cuota fija. Como $g'' < 0$, g es *estrictamente cóncava* en $(0, +\infty)$: sube desde 0, alcanza su máximo donde $g' = 0$, es decir en $y = K/e$, donde vale $g(K/e) = (K/e)\log e = K/e$, y baja cruzando el eje en $y = K$. Es una campana asimétrica. Los equilibrios de la cuota fija resuelven $g(y) = h/\lambda$, y basta cortar esa campana con una recta horizontal:

- si $h > \lambda K/e$, **ninguno**: la recta pasa por encima del máximo;
- si $h = \lambda K/e$, **uno**, tangencial, en $y = K/e$;
- si $0 < h < \lambda K/e$, **dos**, $0 < \bar{y}_1 < K/e < \bar{y}_2 < K$, sin fórmula cerrada mediante funciones elementales.

Como $f' = \lambda g'$ y g' es positiva a la izquierda de K/e y negativa a la derecha, $\bar{y}_1$ es inestable y $\bar{y}_2$ exponencialmente asintóticamente estable; en el caso tangencial $f \leq 0$ en todo el dominio, con igualdad solo en K/e, y el equilibrio es **semiestable**: atrae por la derecha y repele por la izquierda. Por debajo de $\bar{y}_1$ es $f < 0$ sin equilibrio debajo y con $f(y) \to -h < 0$ cuando $y \to 0^+$, luego la solución **alcanza el cero en tiempo finito**: extinción, y el modelo deja de valer ahí. Es la misma estructura de la cuota fija sobre el logístico —dos equilibrios, umbral y catástrofe al pasarse— con la campana λg en lugar de la parábola $\lambda p - \kappa p^2$.

En Gompertz, el esfuerzo fijo no tiene umbral de extinción. Con extracción qy el equilibrio es único, $\bar{y} = Ke^{-q/\lambda}$, y es estable para todo $q > 0$ (es el ejercicio de pesca de Gompertz): se pesque lo que se pesque, la población se asienta en un nivel positivo. Aquí interviene algo más que pescar menos cuando hay pocos peces: la tasa per cápita $\lambda \log(K/y)$ crece sin límite al acercarse y a cero y supera cualquier q finito. El logístico, cuya tasa per cápita está acotada por λ, sí tiene un umbral de extinción asintótica.

LA IDEA: MISMO RENDIMIENTO MÁXIMO, DISTINTA ESTABILIDAD

Compárense en el estado estacionario, que es lo que de verdad interesa a una cofradía: cuántas toneladas al año se pueden sacar *para siempre*.

- Con cuota fija, la mayor h que admite equilibrio es $h_{\text{máx}} = \lambda K/e$, y se sostiene con la población en $\bar{y} = K/e$.
- Con esfuerzo fijo, lo que se saca en el equilibrio es $q\bar{y} = qKe^{-q/\lambda}$; derivando respecto

de q, el máximo está en $q = \lambda$ y vale $\lambda K/e$, alcanzado también con $\bar{y} = Ke^{-1} = K/e$. **El rendimiento máximo sostenible es el mismo, $\lambda K/e$, y se consigue con la misma población, K/e.** La diferencia no está en cuánto se puede pescar, sino en qué pasa al equivocarse: con esfuerzo fijo, pasarse de q solo rebaja el equilibrio; con cuota fija, pasarse de h no rebaja nada, hace desaparecer el equilibrio y extingue la población en tiempo finito. Y aun sin pasarse, explotar cerca del máximo acerca $\bar{y}_1$ a $\bar{y}_2$ y reduce el margen: dentro del modelo, una perturbación que deje la población por debajo del umbral conduce a la extinción. Esa es la lección de modelado, y es la misma que daba el discriminante del logístico: *además del rendimiento importa la estabilidad de la política*.

CONEXIÓN: EL 50 % Y EL 36.8 %

Sobre el logístico, el rendimiento máximo sostenible es $\lambda^2/(4\kappa)$ y se alcanza con la población al 50 % de la capacidad de carga; sobre Gompertz es $\lambda K/e$ y se alcanza al $1/e \approx 36.8$ %. Los dos porcentajes salen *solo* de haber elegido una curva de crecimiento u otra, y dan predicciones distintas que habría que contrastar antes de usarlas para decidir una política pesquera. Es el ejemplo más limpio del capítulo de por qué la fase de modelado —elegir f— no es un trámite previo a las matemáticas: es donde se decide la respuesta.

Ejercicio 2.3. *El modelo de pesca con esfuerzo constante de Gompertz*

ORIGEN: análisis teórico de la práctica 7 de la hoja del tema 2.

Un modelo de pesca con esfuerzo constante es

$$y' = ry \log \frac{K}{y} - qy,$$

donde $y(t)$ es el peso total de la especie, r la constante de crecimiento, K la capacidad de soporte y qy la cantidad pescada por unidad de tiempo; se supone $r, K, q > 0$. Determinar el dominio de f, sus equilibrios, su estabilidad y el comportamiento de la solución de $y(0) = y_0$.

PISTAS, DE MENOR A MAYOR

1. *El logaritmo obliga a $y > 0$: el dominio de f es $(0, +\infty)$ y el cero no es un equilibrio de este modelo.*
2. *Un equilibrio es un cero de f; después f' en él.*
3. *Para el comportamiento, signo de $f(y_0)$ y teorema de las EDO autónomas escalares.*

Resolución.
La idea, y explica por qué este modelo es tan distinto del de cuota fija. Sáquese factor común y:

$$f(y) = y\left(r \log \frac{K}{y} - q\right).$$

Escrito así, el ejercicio entero se ve de un vistazo. El primer factor no se anula nunca en el dominio, luego **los equilibrios son los ceros del paréntesis**; y el paréntesis es estrictamente decreciente en y y recorre todo $\mathbb{R}$, así que se anula *exactamente una vez*, haya el q que haya. De ahí salen las tres respuestas de golpe: hay un único equilibrio, es positivo, y no existe ningún valor de q que lo haga desaparecer.

La imagen, en términos del modelo, es esta. Lo que se extrae es qy, *una fracción fija* de la población: cuando la población baja, la extracción baja con ella, automáticamente. Por eso la extracción disminuye cuando quedan pocos individuos; además, la tasa per cápita logarítmica supera a cualquier q finito cerca de cero. Ambas propiedades explican que no aparezca aquí ningún umbral de extinción. Compárese con la cuota fija $-h$ de la sección 2.1.4, donde lo que se extrae no depende de cuántos queden: allí el término constante puede superar a la producción biológica máxima, y entonces no hay equilibrio ninguno y la extinción es segura. *Un factor y de diferencia, y dos modelos con destinos opuestos.*

Paso 1: dominio y equilibrios. $f(y) = ry\log(K/y) - qy$ está definida en $(0, +\infty)$. Como $y > 0$, $f(y) = 0$ equivale a $r\log(K/y) = q$, es decir, $\log(K/y) = q/r$, luego $y = Ke^{-q/r}$: el único equilibrio es $\bar{y} = Ke^{-q/r}$.

Paso 2 (el núcleo): la estabilidad, y una sorpresa. $f'(y) = r\log(K/y) - r - q$ (derivando $y\log(K/y) = y\log K - y\log y$ sale $\log K - \log y - 1$). En el equilibrio, $\log(K/\bar{y}) = q/r$, luego $f'(\bar{y}) = q - r - q = -r < 0$: el equilibrio es exponencialmente asintóticamente estable. Merece la pena detenerse en el resultado: la tasa de convergencia es $-r$, y **no depende de** q. Pescar más rebaja el nivel en el que la población se asienta, pero no cambia la tasa exponencial local de retorno tras una perturbación pequeña. Es una propiedad de la ley de Gompertz y no del término de pesca, y se ve en que $-qy$ es lineal: al derivar aporta la constante $-q$, que se cancela contra el $+q$ que trae $\log(K/\bar{y})$ en el equilibrio.

Paso 3: comportamiento global, por el teorema escalar. Para $y_0 \in (0, \bar{y})$ es $y_0 < Ke^{-q/r}$, luego $\log(K/y_0) > q/r$ y $f(y_0) > 0$: la solución crece hacia $\bar{y}$. Para $y_0 > \bar{y}$, $f(y_0) < 0$ y decrece hacia $\bar{y}$. Si $y_0 = \bar{y}$, la solución es constante. Para todo $y_0 > 0$, $\lim_{t\to+\infty} y(t; y_0) = \bar{y}$; en este modelo la pesca con esfuerzo constante nunca extingue, solo rebaja el equilibrio de K a $Ke^{-q/r}$. Con los datos de la práctica, $r = 0.2$ y $K = 27$: para $q = 0.05$ el equilibrio es $27e^{-0.25} \approx 21.03$, y para $q = 0.3$ es $27e^{-1.5} \approx 6.02$.

Dónde se falla.

- *Al dar $y = 0$ por equilibrio*, por costumbre del logístico. Aquí el 0 está fuera del dominio: $\log(K/y)$ no está definido, y la extensión por continuidad no es lipschitziana, así que ni siquiera hay una solución única que salga de allí.
- *Al derivar $y\log(K/y)$.* Si se olvida el -1 de la regla del producto, $(y\log(K/y))' = \log(K/y) - 1$, sale $f'(\bar{y}) = q - q = 0$ y parece que la linealización no decide, cuando decide y da $-r$.
- *Al creer que, como no extingue, cualquier q da igual.* Con $q > r$ el equilibrio $Ke^{-q/r}$ queda por debajo de K/e, y lo que se saca cada año, $q\bar{y}$, es menos que el máximo sostenible rK/e. Con los datos, $q = 0.3$ da $0.3 \cdot 6.02 \approx 1.81$ al año frente a $0.2 \cdot 27/e \approx 1.99$: se pesca más fuerte para sacar menos. Es la caja «mismo rendimiento máximo, distinta estabilidad» vista desde el lado del error.

2.2 Modelos regidos por SDOs

LA PREGUNTA

Una epidemia puede crecer, alcanzar un máximo y después remitir sin que se haya contagiado toda la población. El SIS autónomo anterior no produce esa subida seguida de bajada: sus soluciones son monótonas. ¿Qué cambio permite representar una onda epidémica, y qué precio matemático se paga por introducirlo?

Normalmente los modelos de interés son más complejos que los estudiados hasta ahora, porque aparece más de una variable y existen interacciones entre ellas que producen efectos relevantes. Es más frecuente encontrarse con SDOs no escalares. En esta sección se estudian algunos sistemas de dos ecuaciones con orígenes variados, y se verá que aparece una dificultad nueva: en dimensión dos las órbitas pueden girar, y el teorema de las EDO autónomas escalares deja de valer. Cada modelo necesitará su propia herramienta: una cantidad conservada, un criterio de comparación, una función de Lyapunov o el teorema de Poincaré-Bendixson.

2.2.1 Un segundo modelo de epidemias

El modelo SIS se reducía a una EDO escalar porque la población total se descomponía en susceptibles e infecciosos. No contempla que los enfermos puedan curarse y hacerse *inmunes*. Con un sistema de tres EDOs se resuelve esa dificultad. Se supone que el número total de individuos es constante e igual a N y que la población se divide en tres categorías: los **susceptibles** $S(t)$, los **infecciosos** $I(t)$ y los **recuperados** $R(t)$, que ya han pasado la enfermedad y son inmunes. Los susceptibles que se contagian pasan a infecciosos y los infecciosos que se recuperan pasan a recuperados (y no vuelven a ser susceptibles): $S \to I \to R$.

Las hipótesis, una a una. Son las del SIS con una cambiada. Se mantienen la **mezcla homogénea** (contagios proporcionales a SI), la **recuperación sin memoria** (a ritmo γ por infeccioso, con duración media $1/\gamma$) y la **población cerrada** (N constante). Se cambia la no inmunidad por su contrario: la **inmunidad permanente**, el que se cura no vuelve a enfermar. Y se añade una que en el SIS también estaba implícita: **el contagiado es infeccioso desde el primer instante**, sin periodo de latencia. Latencia es el tiempo hasta ser infeccioso; incubación, el tiempo hasta presentar síntomas. El modelo SEIR añade un compartimento E para la primera. La inmunidad permanente excluye su pérdida posterior (que puede modelarse con SIRS). La vacunación es otra modificación: puede mover individuos de S a R sin pasar por I, y es compatible con suponer inmunidad permanente. El precio del cambio se paga en matemáticas, no en biología: con R la población total ya no determina el sistema y hacen falta de verdad dos ecuaciones.

En términos de EDOs: los susceptibles decrecen por la infección en la misma proporción en que crecen los infecciosos; los infecciosos disminuyen en la misma cantidad en que aumentan los recuperados; y los recuperados aumentan en proporción al número de infecciosos. Todo junto,

$$S' = -\beta SI, \qquad I' = \beta SI - \gamma I, \qquad R' = \gamma I,$$

con $\beta, \gamma > 0$. Añadiendo condiciones iniciales se obtiene el problema de Cauchy

$$\begin{cases} S' = -\beta SI, \\ I' = \beta SI - \gamma I, \\ R' = \gamma I, \\ S(0) = S_0,\ I(0) = I_0,\ R(0) = R_0, \end{cases} \qquad S_0, I_0, R_0 \geq 0, \quad S_0 + I_0 + R_0 = N.$$

Aquí $R_0 = R(0)$ denota los recuperados iniciales; no debe confundirse con el número básico de reproducción, que en el SIS se escribió con el mismo símbolo.

EL CAMINO QUE NO SE TOMÓ: SIS FRENTE A SIR, Y POR QUÉ NO SE ELIGE UNO PARA SIEMPRE

La alternativa. Quedarse con el modelo SIS, que es escalar y se resuelve del todo con el teorema de las EDO autónomas escalares.
Su ventaja. Es más simple, tiene un umbral limpio en el número básico de reproducción y representa el retorno a la susceptibilidad tras recuperarse.
Qué lo aparta aquí. No puede producir una onda que sube y baja: sus soluciones son monótonas. Para eso hace falta un tercer compartimento que retire individuos del juego, y el precio es que la conservación de N deja dos variables independientes y no proporciona una fórmula elemental para las soluciones generales. Sus equilibrios forman un continuo: la linealización no basta para clasificar todos los casos. Las propiedades cualitativas permitirán entenderlo sin despejar las soluciones. Se elige entre SIS, SIR y sus extensiones según las hipótesis pertinentes para cada problema.

Se denota $t \mapsto (S(t), I(t), R(t))$ la solución maximal y $I(S_0, I_0, R_0)$ su intervalo de definición. Sumando las EDOs, $(S + I + R)'(t) = 0$ para todo t, de donde $S + I + R \equiv N$: es coherente con la hipótesis de que no hay muertes ni nacimientos durante el intervalo de tiempo en que se estudia la enfermedad.

Por otra parte, existe una **infinidad de puntos de equilibrio**, todos los de la forma $(\bar{S}, 0, \bar{R})$ con $\bar{S}, \bar{R} \geq 0$ y $\bar{S} + \bar{R} = N$ para una población total fijada: sin infecciosos nada cambia. Al formar estos puntos críticos un continuo es imposible que sean atractivos: si se fija uno de ellos y se parte de otro tan próximo como se quiera pero distinto, la solución es constante e igual al segundo y no converge al primero. Y también queda claro que el método de perturbaciones de sistemas lineales, el teorema de estabilidad por linealización del apéndice (libro de EDO, §6.4 «Primer método de Lyapunov»), no va a servir para probar la estabilidad de estos puntos: en un equilibrio $(\bar{S}, 0, \bar{R})$ la matriz jacobiana tiene autovalores 0, 0 y $\beta\bar{S} - \gamma$. No permite concluir estabilidad asintótica; sí permite detectar inestabilidad cuando $\beta\bar{S} - \gamma > 0$. Si los demás autovalores son no positivos, los ceros impiden resolver la estabilidad con este criterio por sí solo.

No disponemos de una fórmula elemental general, pero sí se pueden obtener algunas órbitas particulares, y con ellas se ve que el intervalo de definición contiene a $[0, +\infty)$ y que las tres componentes son no negativas para $t \geq 0$, lo que muestra la vigencia del modelo.

Caso $I_0 = 0$. Entonces $(S_0, 0, R_0)$ es un punto crítico, la solución es constante y está definida en todo $\mathbb{R}$.

Caso $I_0 > 0$ y $S_0 = 0$ (luego $I_0 + R_0 = N$). Aquí $S(t) \equiv 0$: la función nula resuelve $S' = -\beta SI$ con dato cero, y la solución es única. Como $I' = -\gamma I$, se tiene $I(t) = I_0 e^{-\gamma t} > 0$. Y como

$R' = \gamma I$, R es estrictamente creciente y $R(t) = R_0 + \gamma \int_0^t I(s)\,ds$. Entonces $I(0, I_0, R_0) = \mathbb{R}$ y

$$\lim_{t\to+\infty} I(t) = 0, \qquad \lim_{t\to+\infty} R(t) = N.$$

Caso $S_0 > 0$ e $I_0 > 0$ (luego $I_0 + R_0 < N$). A partir del sistema se obtiene, resolviendo la ecuación lineal $S' = (-\beta I(t))S$ e integrando $R' = \gamma I$,

$$S(t) = S_0 \exp\Big(-\beta \int_0^t I(s)\,ds\Big), \qquad R(t) = R_0 + \gamma \int_0^t I(s)\,ds$$

en $I(S_0, I_0, R_0)$. Del mismo modo, la ecuación lineal $I' = (\beta S(t) - \gamma)I$ da

$$I(t) = I_0 \exp\Big(\int_0^t \big(\beta S(s) - \gamma\big)ds\Big).$$

Estas fórmulas justifican la positividad sin dividir previamente por una función que pudiera anularse. En $I(S_0, I_0, R_0) \cap [0, +\infty)$, S e I son positivas y R es no negativa (positiva para $t > 0$). Como su suma es N, las tres están acotadas entre 0 y N. Luego la semiórbita positiva vive en un compacto y $I(S_0, I_0, R_0) \supset [0, +\infty)$ por la proposición de prolongación en un compacto (caracterización de las soluciones maximales, §3.4 del libro de EDO). Además S es estrictamente decreciente (su derivada $-\beta SI$ es negativa), R es estrictamente creciente, y las dos están acotadas, luego existen los límites

$$\lim_{t\to+\infty} S(t) = \bar{S} \qquad \text{y} \qquad \lim_{t\to+\infty} R(t) = \bar{R}.$$

Por la conservación de la población total existe también $\lim_{t\to+\infty} I(t) = N - \bar{S} - \bar{R}$. Como todas las componentes convergen, el límite es un punto crítico (observación sobre límites que son equilibrios; los conjuntos límite son invariantes, §8.3 del libro de EDO), y los puntos críticos tienen $I = 0$; necesariamente

$$\lim_{t\to+\infty} I(t) = 0.$$

La enfermedad desaparece para tiempos grandes. La conclusión es que el conjunto $\Lambda^+(S_0, I_0, R_0)$ es siempre unitario y su único elemento es de la forma $(\bar{S}, 0, \bar{R})$ con $0 \le \bar{S}, \bar{R} \le N$ y $\bar{S} + \bar{R} = N$. La figura 2.4 muestra las órbitas en el plano (I, R).

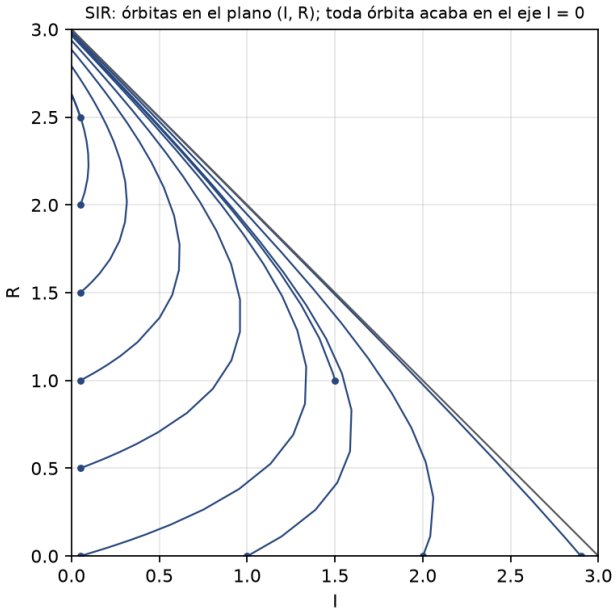


Figura 2.4: Órbitas del modelo SIR en el plano de fases (I, R) para $\beta = 1.5$, $\gamma = 1$ y $N = 3$. El triángulo $I + R \leq N$ es invariante, y toda órbita muere en el eje $I = 0$, en un punto que depende de dónde salió. Lo que decide la forma es el dato inicial: las siete órbitas que arrancan con $S_0 > \gamma/\beta = 2/3$ suben hasta un máximo de infecciosos y después bajan, mientras que las tres que arrancan con $S_0 < 2/3$ —las de $(I_0, R_0) = (0.05, 2.5)$, $(1.5, 1)$ y $(2.9, 0)$— no llegan a subir: se van hacia el eje desde el primer instante, porque ya no quedan bastantes susceptibles.

Lo que hay que llevarse. El argumento no usa ninguna fórmula para I, y se repite casi palabra por palabra en el SIR con mortalidad de la subsección siguiente y en casi todos los modelos de compartimentos. Son cuatro frases. (i) *Positividad*: S e I aparecen como factor en su propia ecuación, y una exponencial no se anula; R, que solo crece, no baja de $R_0 \geq 0$. (ii) *Acotación*: una cantidad conservada, aquí $S + I + R = N$, encierra la órbita en un compacto, y el compacto da existencia en $[0, +\infty)$. (iii) *Monotonía*: S solo baja y R solo sube, luego las dos tienen límite, y con ellas $I = N - S - R$. (iv) *El límite es un equilibrio*, y en este modelo los equilibrios son los estados sin infecciosos: $I \to 0$.

LA IDEA: POR QUÉ SE APAGA SIN CONTAGIAR A TODOS

Supongamos $S_0 > 0$ e $I_0 > 0$. De $I' = I(\beta S - \gamma)$ se lee que los infecciosos crecen mientras $S > \gamma/\beta$ y decrecen después. Si se arranca por debajo del umbral, $S_0 < \gamma/\beta$, la epidemia no llega a prender y I baja desde el primer instante. Y si se arranca por encima, como S solo baja, hay un instante en que cruza el umbral γ/β —lo cruza necesariamente, porque, si S permaneciera por encima, tendríamos $I' > 0$ siempre y por tanto $I(t) \geq I_0 > 0$, contradiciendo $I(t) \to 0$— y a partir de ahí la epidemia remite, aunque queden susceptibles: cada infectado ya no encuentra bastantes susceptibles para reemplazarse. Si $S_0 = \gamma/\beta$, $I'(0) = 0$, pero S baja inmediatamente y después I decrece. Una EDO autónoma escalar con unicidad no produce una solución no constante que primero suba y después baje. El número final de susceptibles $\bar{S}$ es

positivo: no todos se contagian. Para verlo y calcularlo, dividiendo S' entre R' se obtiene $dS/dR = -\beta S/\gamma$, luego $S = S_0 e^{-\beta(R-R_0)/\gamma}$ a lo largo de toda la órbita, y el punto final cumple $\bar{S} = S_0 e^{-\beta(N-\bar{S}-R_0)/\gamma}$, una ecuación trascendente en $\bar{S}$ que la práctica 10 comprueba numéricamente.

El tamaño final, con números. La ecuación del final no se despeja con una fórmula, pero se resuelve a mano con cuatro evaluaciones. Mídase la población en fracciones, $N = 1$, con $\beta = 2$ y $\gamma = 1$: al principio cada infeccioso contagia a razón de $\beta N = 2$ por unidad de tiempo durante un tiempo medio $1/\gamma = 1$, así que produce unos dos contagios. Se arranca con $S_0 = 0.99$, $I_0 = 0.01$, $R_0 = 0$. El umbral es $\gamma/\beta = 0.5$ y S_0 está muy por encima: hay brote. El punto final cumple

$$\bar{S} = 0.99\, e^{-2(1-\bar{S})},$$

y se busca por debajo del umbral, porque la epidemia solo remite cuando S ha bajado de γ/β y S ya no vuelve a subir. La función $g(S) = S - 0.99\, e^{-2(1-S)}$ vale -0.134 en $S = 0$ y $+0.136$ en $S = 0.5$, luego tiene un cero entre medias; y como $g(0.19) \approx -0.006$ y $g(0.2) \approx +0.0001$, ese cero está en $\bar{S} \approx 0.1998$. Una integración numérica del sistema da lo mismo, $\bar{S} = 0.19980$.

La lectura es la que anunciaba la caja: **uno de cada cinco no llega a contagiarse nunca**, aunque la epidemia haya alcanzado al 80 % de la población, $\bar{R} \approx 0.80$. La epidemia se apaga por falta de susceptibles *suficientes*, no por falta de susceptibles. Y la ecuación no contiene el tiempo: el tamaño final se sabe sin saber cuánto tarda.

DE DÓNDE VIENE: KERMACK Y MCKENDRICK, 1927

El modelo SIR lo propusieron William Kermack y Anderson McKendrick en 1927, en un artículo que ajustaba la epidemia de peste de Bombay de 1905-06 y que contenía ya el teorema del umbral: la epidemia solo prende si $S_0 > \gamma/\beta$. McKendrick era médico militar y había trabajado con Ronald Ross, que en 1911 había escrito las primeras ecuaciones diferenciales de la malaria. El modelo se ha refinado desde entonces con edades, redes de contacto y estocasticidad, pero la estructura de compartimentos y el umbral R_0 siguen siendo el lenguaje de la epidemiología.

EXPLÍCATE. *En el caso $S_0 > 0$, $I_0 > 0$, la demostración de que $I \to 0$ no usa ninguna fórmula para I: usa que S y R son monótonas acotadas y que un límite es un equilibrio. Di por qué ese argumento no se puede repetir para probar que I es monótona.*

2.2.2 Epidemias con mortalidad: las dos leyes de incidencia

LA PREGUNTA

El SIR anterior supone que la población total no cambia: los que se curan siguen ahí. En una epidemia que mata, no. Y hay una segunda hipótesis escondida en el término βSI: que cada persona se cruza con *más* gente cuanto más grande es la ciudad. ¿Cuál de las dos hipótesis se rompe antes, qué término hay que cambiar en cada caso, y qué se conserva cuando ya no se conserva la población?

Primera hipótesis: cómo se cuentan los contagios. En el modelo anterior el término de contagio es βSI, y se llama **incidencia de masa**. Detrás está la hipótesis de que cada individuo

tiene un número de encuentros *proporcional al tamaño de la población*: si la ciudad duplica sus habitantes, cada uno se cruza con el doble de gente. La alternativa es

$$\frac{\beta_f}{N} SI = \beta_f S \cdot \frac{I}{N}, \qquad N = S + I + R,$$

la **incidencia de frecuencia**: cada individuo tiene un número *fijo* de contactos por unidad de tiempo, y una fracción I/N de ellos son infecciosos. La forma más rápida de no confundirlas es mirar las unidades: β_f se mide en $1/\text{tiempo}$, mientras que β se medía en $1/(\text{individuo} \cdot \text{tiempo})$. Mientras N no cambie las dos son la misma con constantes distintas, $\beta = \beta_f/N$; en cuanto N varía, dejan de serlo.

LA IDEA: CUÁL DE LAS DOS ELEGIR

La de masa puede ser adecuada cuando aumentar la población en un espacio fijo aumenta proporcionalmente los contactos por persona. La de frecuencia representa un número de contactos por persona aproximadamente fijo, limitado por ejemplo por el tiempo disponible. La elección exige estudiar la estructura de contactos, no solo el nombre de la enfermedad. La regla práctica es preguntarse si al doblar la población se dobla el número de contactos por persona; si la respuesta es no, es incidencia de frecuencia. Y el aviso: en una población constante las dos ajustan igual de bien los datos, así que el ajuste no distingue entre ellas —distinguen las predicciones, en cuanto la población cambia—.

Segunda hipótesis: que nadie muere. Si solo una proporción $m \in (0,1)$ de los infecciosos llega a recuperarse y la proporción $1-m$ restante muere y sale del sistema, la tercera ecuación deja de ser $R' = \gamma I$ y pasa a ser $R' = m\gamma I$. El modelo completo es

$$\begin{cases} S' = -\dfrac{\beta_f}{N} SI, \\ I' = \dfrac{\beta_f}{N} SI - \gamma I, \\ R' = m\gamma I, \end{cases} \qquad N = S + I + R, \tag{2.1}$$

con $\beta_f, \gamma > 0$ y $m \in (0,1)$, y ahora N **no es constante**: sumando las tres ecuaciones, $N' = -(1-m)\gamma I \leq 0$ en el dominio físico. En lo que sigue suponemos $S_0, I_0 > 0$ y $R_0 \geq 0$; así $N_0 = S_0 + I_0 + R_0 > 0$.

Dónde está definido. El segundo miembro es una función racional cuyo único denominador es $N = S + I + R$, luego el campo es de clase C^∞ —y en particular localmente lipschitziano— en el abierto $\{N \neq 0\}$, que tiene **dos** componentes conexas: los semiespacios $\{N > 0\}$ y $\{N < 0\}$. Esos son los dos dominios maximales, y el que interesa es $\{N > 0\}$; dentro de él, los estados con sentido biológico tienen $S, I, R \geq 0$. Los puntos críticos son los del conjunto $\{I = 0\} \cap \{N > 0\}$: cualquier estado sin infecciosos es un equilibrio, y forman una porción del plano $I = 0$, no puntos aislados. Como en el SIR de la sección 2.2.1, la linealización no clasifica todos los casos, aunque sí detecta inestabilidad si $\beta_f \bar{S}/(\bar{S} + \bar{R}) - \gamma > 0$.

LA IDEA: POR QUÉ LA CONSERVACIÓN DE N NO SE PIERDE DEL TODO

Se ha roto la ley $N \equiv N_0$, que era la que daba el compacto y con él todo lo demás. Pero queda otra, y se encuentra buscando una combinación lineal de N y R cuya derivada se anule: como $N' = -(1-m)\gamma I$ y $R' = m\gamma I$, basta pesar la primera con m y la segunda con $1-m$,

$$\big(mN + (1-m)R\big)' = -m(1-m)\gamma I + (1-m)m\gamma I = 0.$$

Es decir, $mN + (1-m)R \equiv K := mN_0 + (1-m)R_0$ a lo largo de toda la solución. La lectura es que *los que mueren y los que se recuperan salen de los infecciosos en proporción fija*, así que pesando la población total con m y los recuperados con $1-m$ se vuelve a tener una cantidad conservada. Y de ella sale, en una línea, la cota que hacía falta: como $R \le S + I + R = N$,

$$K = mN + (1-m)R \le mN + (1-m)N = N,$$

luego $N(t) \ge K > 0$ mientras las componentes sean no negativas. Justificaremos enseguida esa propiedad para $t \ge 0$. El denominador nunca se acerca a cero, y la solución no puede escaparse por ahí: la conservación rota se ha sustituido por otra que hace exactamente el mismo trabajo.

Positividad, y la EDO lineal disfrazada. La observación que evita todo el trabajo es que S e I aparecen como *factores* en sus propias ecuaciones:

$$S' = -\Big(\frac{\beta_f I}{N}\Big)S =: a(t)S, \qquad I' = \Big(\frac{\beta_f S}{N} - \gamma\Big)I =: b(t)I.$$

Cada una es una EDO lineal escalar con coeficiente continuo, así que $S(t) = S_0 e^{\int_0^t a}$ e $I(t) = I_0 e^{\int_0^t b}$: si arrancan positivas, se quedan positivas en todo el intervalo maximal, porque una exponencial no se anula. Además, N no cambia de signo: es continua y no puede atravesar el plano excluido $N = 0$. Para $t \ge 0$, $R(t) = R_0 + m\gamma \int_0^t I(s)\,ds \ge 0$. Con eso, $S' < 0$ y $R' > 0$ estrictamente, y $N' < 0$: S y N decrecen, R crece.

Existencia global y comportamiento final. Con $K \le N \le N_0$, y $0 < S \le S_0$, $0 \le R \le K/(1-m)$, $0 < I \le N_0$, la órbita está encerrada en un compacto contenido en el dominio; luego la solución maximal está definida en todo $[0, +\infty)$. Y siendo S y N decrecientes y acotadas inferiormente, y R creciente y acotada superiormente, existen

$$S_\infty = \lim_{t\to\infty} S \ge 0, \qquad R_\infty = \lim_{t\to\infty} R > 0, \qquad N_\infty = \lim_{t\to\infty} N > 0,$$

y de $I = N - S - R$ sale el cuarto límite, $\lim I = \ell \ge 0$. Necesariamente $\ell = 0$: si fuera positivo, para tiempos grandes tendríamos $R' = m\gamma I \ge m\gamma\ell/2 > 0$, contradiciendo que R converge. La epidemia se apaga bajo estas hipótesis.

El umbral, y el pico. El signo de I' es el de $\beta_f S/N - \gamma$, es decir, el de

$$\frac{S}{N} - \frac{1}{\mathcal{R}_0}, \qquad \mathcal{R}_0 := \frac{\beta_f}{\gamma},$$

y aquí lo que decide no es S sino la *fracción* de susceptibles S/N. El número básico de reproducción $\mathcal{R}_0$ es adimensional, como también lo era $N\beta/\gamma$ en el modelo de masa. Aquí no depende de N porque se ha supuesto un número fijo de contactos por persona. En el caso límite $m = 1$ —nadie muere, $N \equiv N_0$— la función $g := S/N - 1/\mathcal{R}_0$ es estrictamente decreciente, porque S lo es y N es constante; luego cambia de signo a lo sumo una vez. Si $S_0/N_0 > 1/\mathcal{R}_0$, empieza positiva, I crece, y como $I \to 0$ tiene que acabar decreciendo: existe un **único** $T_{\text{máx}} \in (0, \infty)$ con I estrictamente creciente en $(0, T_{\text{máx}})$ y estrictamente decreciente después. Ese $T_{\text{máx}}$ es el pico de la epidemia, y el argumento entero es «el signo de una función estrictamente monótona cambia una vez y solo una».

Y con mortalidad, $m < 1$. El argumento sigue valiendo, con una cuenta más: ahora N también baja, y nada garantiza de antemano que la fracción S/N lo haga, porque un cociente de dos funciones decrecientes puede subir. Derivando el cociente y usando $S' = -\beta_f SI/N$ y $N' = -(1-m)\gamma I$,

$$\left(\frac{S}{N}\right)' = \frac{S'N - SN'}{N^2} = \frac{-\beta_f SI + (1-m)\gamma SI}{N^2} = -\frac{SI}{N^2}\left(\beta_f - (1-m)\gamma\right).$$

Se lee como una carrera entre dos flujos: los contagios sacan susceptibles de la fracción a ritmo β_f, y las muertes la hacen subir, porque encogen el total, a ritmo $(1-m)\gamma$. Gana el primero si $\beta_f > (1-m)\gamma$, es decir, si $\mathcal{R}_0 > 1 - m$. Y eso está garantizado en cuanto hay brote: si $S_0/N_0 > 1/\mathcal{R}_0$, como $S_0/N_0 < 1$ (hay infecciosos al principio), es $\mathcal{R}_0 > 1 > 1 - m$. Entonces S/N es estrictamente decreciente, $g = S/N - 1/\mathcal{R}_0$ también, y el resto del argumento no cambia: g empieza positiva, no puede seguir positiva siempre porque $I \to 0$, y cambia de signo una sola vez. Hay un único pico también con $m < 1$. Con $\beta_f = 2$, $\gamma = 1$, $m = 0.5$ y el dato $(0.99,\ 0.01,\ 0)$, una integración numérica da un solo máximo, $I \approx 0.195$ en $t \approx 4.50$.

La idea: por qué $\mathcal{R}_0$ y $R(0)$ no son lo mismo

Es la confusión más frecuente de todo el tema, y la culpa es de la notación. $R(t)$ es el número de recuperados en el instante t, y $R_0 = R(0)$ es un dato inicial, medido en individuos. $\mathcal{R}_0 = \beta_f/\gamma$ es un número sin unidades que dice cuántos contagios produce un infeccioso en una población totalmente susceptible: es la duración media de la infección, $1/\gamma$, por la velocidad de contagio, β_f. Cuando en un periódico se dice «el erre cero de esta variante es 3», se habla del segundo. En esta sección el primero se escribe R_0 y el segundo $\mathcal{R}_0$, con letra caligráfica; en el SIS se usó R_0 para el número básico de reproducción.

Conexión: qué se pierde, y por dónde crece el modelo

Los compartimentos siguen suponiendo mezcla homogénea, sin edades ni redes de contacto, y la salida de la fase infecciosa sigue teniendo tasa constante por individuo: su duración se modela mediante una exponencial. La ausencia de memoria es una simplificación que puede ser insuficiente para ciertos datos. Añadir un compartimento de expuestos todavía no infecciosos da el modelo SEIR, y separar los asintomáticos de los sintomáticos, el SEAIR de la sección 4.9, que es donde aparece el problema realmente interesante: no simular la epidemia con los parámetros dados, sino *leer los parámetros* en los datos observados.

2.2.3 Modelos vectoriales de poblaciones

LA PREGUNTA

Dos especies comparten un bosque. Sin saber nada de su biología, solo que cada una crece o decrece según cuántos individuos haya de cada una, ¿cuántas formas cualitativamente distintas de relacionarse pueden tener, y qué es lo mínimo que hay que medir para saber en cuál están?

Hasta ahora los modelos describen colectivos aislados. En la práctica es más común que varias poblaciones compartan el mismo hábitat, y aquí se estudia el caso de dos especies que conviven e interaccionan. Estudiaremos tres tipos de interacción recíproca según el signo del efecto de cada especie sobre la otra: **presa-depredador**, **competición** y **simbiosis** (aquí, cooperación o mutualismo). No se pretende incluir todas las relaciones ecológicas posibles: quedan fuera, por ejemplo, las que solo afectan a una de las dos especies.

Sean $u = u(t)$ y $v = v(t)$ las cantidades de individuos de las dos poblaciones. Como en el caso de una sola especie, el modelo es un SDO autónomo

$$\begin{cases} u' = g(u, v), \\ v' = h(u, v), \end{cases}$$

y la cuestión es cómo deben ser g y h. Como antes, se buscan modelos para poblaciones no demasiado grandes, de modo que g y h se aproximen por un desarrollo de Taylor de primer o, a lo más, segundo orden. Se empieza observando que se debe tener

$$g(0, v) \equiv 0 \qquad \text{y} \qquad h(u, 0) \equiv 0 :$$

si no hay individuos de una población y no hay inmigración, no aparece espontáneamente, haya o no individuos de la otra. En consecuencia, si g y h son suficientemente regulares, derivando esas identidades respecto de v y de u respectivamente,

$$g(0,0) = \partial_v g(0,0) = \partial^2_{vv} g(0,0) = 0, \qquad h(0,0) = \partial_u h(0,0) = \partial^2_{uu} h(0,0) = 0.$$

Con desarrollos de Taylor de *primer* orden, estas igualdades dejan $g(u, v) \approx \partial_u g(0,0)\, u$ y $h(u, v) \approx \partial_v h(0,0)\, v$: dos EDOs independientes, $u' = au$ y $v' = bv$, en las que **las poblaciones no interaccionan**. Para representar interacción en esta aproximación hay que ir al segundo orden, y entonces

$$\begin{cases} u' = \partial_u g(0,0)\, u + \frac{1}{2}\big(\partial^2_{uu} g(0,0)\, u^2 + 2\, \partial^2_{uv} g(0,0)\, uv\big), \\ v' = \partial_v h(0,0)\, v + \frac{1}{2}\big(\partial^2_{vv} h(0,0)\, v^2 + 2\, \partial^2_{uv} h(0,0)\, uv\big). \end{cases}$$

La interacción queda determinada por las **derivadas cruzadas** $\partial^2_{uv} g(0,0)$ y $\partial^2_{uv} h(0,0)$: para que la aproximación de segundo orden represente interacción en ambos sentidos, se suponen no nulas. En modelos de orden superior puede haber interacción aunque estas derivadas se anulen. Según sus signos:

- **Presa-depredador**: signos distintos, por ejemplo $\partial^2_{uv} g(0,0) > 0$ y $\partial^2_{uv} h(0,0) < 0$. Si aumenta v aumenta la tasa de crecimiento de u, y un aumento de u hace decrecer la de v: u es el **depredador** y v la **presa**. (Conviene fijarlo desde ahora, porque es al revés de lo que uno espera por orden alfabético: aquí, y en todo el capítulo, u come y v es comida.)

- **Competición**: $\partial^2_{uv} g(0,0) < 0$ y $\partial^2_{uv} h(0,0) < 0$. El aumento de cada especie frena a la otra.
- **Simbiosis**: las dos positivas. El crecimiento de cada una favorece a la otra.

Respecto de las otras derivadas, $\partial^2_{uu} g(0,0)$ y $\partial^2_{vv} h(0,0)$ juegan un papel análogo al de $-\kappa$ en el logístico (el coeficiente cuadrático es la mitad de cada derivada): representan autolimitación dentro de la especie, y se suponen ≤ 0. A continuación se hace el estudio completo de los modelos presa-depredador y competición en el caso simplificado en que estas derivadas son *nulas* —sin autolimitación, solo interacción—, y después el de competición general y el de simbiosis, donde sí aparecen.

LA IDEA: LA RESPUESTA A LA PREGUNTA

Lo mínimo que hay que medir son dos signos: cómo responde la tasa de crecimiento de cada especie a la abundancia de la otra. Con dos signos hay cuatro combinaciones, y una de ellas $(+,-)$ aparece dos veces intercambiando los papeles; quedan tres tipos. Es notable que la clasificación de estos tres tipos de interacción se lea en dos coeficientes. El desarrollo de Taylor justifica una aproximación local; su utilidad para otras poblaciones requiere contraste con datos. En la sección 2.2.6 se verá qué cambia al añadir los términos que faltan.

2.2.4 Modelo de Volterra (presa-depredador simplificado)

LA PREGUNTA

Durante la Primera Guerra Mundial disminuyó la pesca en el Adriático y aumentó la proporción de peces depredadores en las capturas. ¿Cómo podría explicar el modelo que reducir la pesca favorezca al depredador respecto de la presa, y qué forma tiene la relación entre los dos a lo largo de los años: se equilibran, o se persiguen sin fin?

Además de los signos de presa-depredador se supone aquí

$$\partial_u g(0,0) < 0, \qquad \partial_v h(0,0) > 0, \qquad \partial^2_{uu} g(0,0) = \partial^2_{vv} h(0,0) = 0.$$

La primera hipótesis se justifica así: si $v = 0$ los depredadores no tienen alimento y su número no hace otra cosa que disminuir, luego $g(u,0) < 0$, y como $g(0,0) = 0$, $\partial_u g(0,0) = \lim_{u\to 0^+} g(u,0)/u \leq 0$. Que sea estrictamente negativa es una hipótesis adicional: una tasa de mortalidad positiva incluso a baja densidad. La segunda dice que, en ausencia de depredadores, las presas se reproducen de manera natural. Y suponer $\partial^2_{vv} h(0,0) = 0$ equivale a despreciar su autolimitación. Puede ser una aproximación si las presas permanecen lejos de agotar los recursos; no queda garantizada solo por la presencia de depredadores. Con estas precisiones, el **modelo de Volterra** es

$$\begin{cases} u' = u(-a + bv), \\ v' = v(c - du), \end{cases} \qquad a, b, c, d > 0.$$

Qué dice cada término. Conviene leer las dos ecuaciones en palabras antes de tocarlas. En la de los depredadores, $-au$ es la mortalidad sin alimento: sin presas los depredadores decaen exponencialmente, con semivida $\log 2/a$; y $+buv$ es su reproducción, proporcional al

número de encuentros uv con las presas (otra vez la mezcla homogénea del SIS: se supone una tasa comparable de encuentros para cada pareja), donde b combina la frecuencia de esos encuentros y su efecto reproductivo. En la de las presas, $+cv$ es su reproducción sin límite en ausencia de depredadores, es decir, el modelo exponencial con ausencia de autolimitación; y $-duv$ es la pérdida total de presas, con la tasa de encuentros y la eficacia de captura recogidas en d. Cuatro constantes, cuatro frases. Lo que se ha despreciado se ve en lo que falta: no hay $-\kappa v^2$, luego las presas no se autolimitan (solo las limitan los depredadores); no hay saciedad, luego un depredador come el doble si hay el doble de presas, sin tope, cuando un lince come lo que come; y no hay refugios, edades ni estaciones. Las dos primeras omisiones son las que producen el resultado más llamativo y más frágil del modelo, las órbitas cerradas, y se discuten en cuanto aparezca.

Adimensionalización. El sistema tiene cuatro constantes. Con cambios adecuados de unidades de medida de u, v y del tiempo se reduce a una sola. Qué quiere decir eso se ve mejor primero en el caso más simple.

El caso más simple: la logística. En $y' = \lambda y - \kappa y^2$ hay dos constantes. Se cambia la unidad de población y la escala de tiempo a la vez: $\hat{y}(s) = p\,y(rs)$, con $p, r > 0$ a elegir. En palabras, $\hat{y}(s)$ es la población del instante rs contada en unidades de $1/p$ individuos, y cada unidad del tiempo nuevo s vale r unidades del tiempo antiguo. La regla de la cadena da

$$\frac{d\hat{y}}{ds}(s) = p\,r\,y'(rs) = p\,r\big(\lambda\,y(rs) - \kappa\,y(rs)^2\big),$$

y como $y(rs) = \hat{y}(s)/p$,

$$\frac{d\hat{y}}{ds} = r\Big(\lambda\hat{y} - \frac{\kappa}{p}\hat{y}^2\Big).$$

Hay dos coeficientes y dos libertades: $p = \kappa/\lambda$ iguala los dos coeficientes, y $r = 1/\lambda$ los hace valer 1. Queda $\hat{y}' = \hat{y}(1 - \hat{y})$, **sin ningún parámetro**: todas las logísticas son la misma curva, vista con otra unidad de población —la capacidad λ/κ— y con otro reloj —el tiempo $1/\lambda$—. La cuenta que hay que retener es *constantes menos unidades libres*: aquí, $2 - 2 = 0$.

El de Volterra, igual, con tres unidades. Se pone

$$\hat{u}(s) = p\,u(rs), \qquad \hat{v}(s) = q\,v(rs),$$

con $r, p, q > 0$ a elegir. Por la regla de la cadena, $\frac{d\hat{u}}{ds} = pr\,u'(rs) = r\hat{u}(-a + \frac{b}{q}\hat{v})$ y $\frac{d\hat{v}}{ds} = r\hat{v}(c - \frac{d}{p}\hat{u})$. La primera igualdad es la regla de la cadena; la segunda es el paso que se suele saltar. Se sustituye la ecuación, $u'(rs) = u(rs)\big(-a + b\,v(rs)\big)$, y se escriben $u(rs) = \hat{u}(s)/p$ y $v(rs) = \hat{v}(s)/q$:

$$pr\,u'(rs) = pr\,\frac{\hat{u}}{p}\Big(-a + b\,\frac{\hat{v}}{q}\Big) = r\,\hat{u}\Big(-a + \frac{b}{q}\,\hat{v}\Big).$$

La p se cancela porque cada término de la ecuación lleva un factor u; la q se queda, dividiendo a b. La segunda ecuación es igual con los papeles cambiados. Tomando $q = b/a$ y $p = d/c$ queda

$$\frac{d\hat{u}}{ds} = ar\,\hat{u}(-1 + \hat{v}), \qquad \frac{d\hat{v}}{ds} = cr\,\hat{v}(1 - \hat{u}),$$

y tomando $r = 1/a$, llamando de nuevo u, v a $\hat{u}$, $\hat{v}$ y λ a $cr = c/a$, se obtiene el **sistema adimensionalizado**

$$\begin{cases} u' = u(-1 + v), \\ v' = \lambda v(1 - u). \end{cases}$$

Cómo se eligen p, q y r. Cada unidad libre se gasta en hacer igual a 1 un coeficiente. La elección $q = b/a$ permite sacar factor común a en el primer paréntesis, $-a + a\hat{v} = a(-1 + \hat{v})$; $p = d/c$ hace lo mismo con c en el segundo; y $r = 1/a$ convierte en 1 el factor ar de delante. Con tres unidades se han fijado tres coeficientes, y de las cuatro constantes sobrevive una combinación, $\lambda = c/a$: $4 - 3 = 1$. Cuál sobrevive depende de la elección. Con $r = 1/c$ en lugar de $1/a$ queda $u' = \mu\, u(-1 + v)$, $v' = v(1 - u)$, con $\mu = a/c = 1/\lambda$: el mismo modelo, con el parámetro en la otra ecuación. Lo que no depende de la elección es cuántos quedan: uno. El mismo método, con cinco constantes, se aplica a la simbiosis de la subsección 2.2.7, y el apéndice C lo aplica a varios modelos más.

Con números. Sean $a = 0.5$ y $c = 1$ por mes, y $b = 0.01$ y $d = 0.02$ por individuo y mes. Entonces $q = b/a = 0.02$, $p = d/c = 0.02$, $r = 1/a = 2$ y $\lambda = c/a = 2$. Deshacer el cambio es leer las unidades: $\hat{v} = 1$ son $1/q = a/b = 50$ presas, $\hat{u} = 1$ son $1/p = c/d = 50$ depredadores, y $s = 1$ son $r = 2$ meses. Una solución del sistema adimensionalizado se traduce con

$$u(t) = \frac{1}{p}\,\hat{u}\Big(\frac{t}{r}\Big) = 50\,\hat{u}\Big(\frac{t}{2}\Big), \qquad v(t) = \frac{1}{q}\,\hat{v}\Big(\frac{t}{r}\Big) = 50\,\hat{v}\Big(\frac{t}{2}\Big)\,:$$

el equilibrio $(1, 1)$ es el de 50 depredadores y 50 presas, y un intervalo de tiempo de longitud Δs en el sistema nuevo dura $2\,\Delta s$ meses.

LA IDEA: QUÉ ES ADIMENSIONALIZAR

Es elegir la unidad de presas de modo que «una presa» sea lo que hace falta para que los depredadores ni crezcan ni decrezcan ($v = 1$ es a/b en unidades originales), la unidad de depredadores de modo que $u = 1$ sea el nivel que deja estables a las presas (c/d), y la unidad de tiempo como el tiempo medio de supervivencia de un depredador sin comida bajo la hipótesis exponencial ($1/a$). Después de eso solo sobrevive un número, $\lambda = c/a$, el cociente entre la tasa de reproducción de las presas y la de mortalidad de los depredadores. La forma del campo adimensional solo depende de λ; las trayectorias dependen además de los datos iniciales adimensionalizados. Para recuperar cantidades y tiempos físicos hay que deshacer el cambio de unidades. La técnica simplifica la comparación entre modelos sin perder esa información.

Dónde se falla al adimensionalizar.

- *Al olvidar la r de la regla de la cadena.* La derivada de $\hat{u}(s) = p\,u(rs)$ es $pr\,u'(rs)$, no $p\,u'(rs)$. Sin esa r desaparece la tercera libertad, y parece que solo se pueden quitar dos constantes.
- *Al sustituir al revés.* De $\hat{u} = p\,u$ sale $u = \hat{u}/p$, no $u = p\,\hat{u}$. Con el error aparecen bq donde debía estar b/q y un factor p^2 que no debería estar, y las elecciones de antes ya no dejan coeficientes iguales a 1.
- *Al no deshacer el cambio.* El equilibrio $(1, 1)$, los periodos y los máximos que salen del sistema adimensionalizado están en las unidades nuevas. En el ejemplo numérico, el periodo de las oscilaciones pequeñas es $2\pi/\sqrt{\lambda} \approx 4.44$ en la variable s, que son unos 8.9 meses; decir «4.44 meses» mezcla las dos escalas.
- *Al esperar que sobreviva siempre la misma constante.* Otra elección de r deja $\mu = 1/\lambda$ en la otra ecuación, y las dos respuestas son correctas. Cuando el enunciado fija la forma final del sistema, la elección de p, q y r queda obligada.

Soluciones semitriviales y positividad. Con $u \equiv 0$ o $v \equiv 0$ el sistema se reduce a $v' = \lambda v$ o a $u' = -u$, con soluciones

$$(u(t), v(t)) = (0, Ce^{\lambda t}) \qquad \text{y} \qquad (u(t), v(t)) = (Ce^{-t}, 0) \qquad \text{para todo } t \in \mathbb{R}.$$

Tomando $C = 0$ o $C = 1$ se obtienen las órbitas $\{(0,0)\}$, $\{(0,v) : v > 0\}$ y $\{(u,0) : u > 0\}$: el origen y los dos semiejes positivos. Como órbitas distintas no se pueden cortar, si se parte de (u_0, v_0) con $u_0 > 0$ y $v_0 > 0$, la órbita no puede salir del primer cuadrante:

$$u(t; u_0, v_0) > 0 \quad \text{y} \quad v(t; u_0, v_0) > 0 \qquad \text{para todo } t \in I(u_0, v_0).$$

Si las dos poblaciones tienen individuos en el instante inicial, ninguna se extingue en tiempo finito.

Equilibrios y linealización. Resolviendo $u(-1+v) = 0$, $\lambda v(1-u) = 0$ se obtienen dos puntos, $(0,0)$ y $(1,1)$, que en las variables originales son $(0,0)$ y $(c/d, a/b)$. Para estudiar su estabilidad se intenta aplicar el teorema de linealización (teorema de estabilidad por linealización; libro de EDO, §6.4 «Primer método de Lyapunov», donde está demostrado con más detalle del que cabe aquí, incluida la versión de inestabilidad para $f \in C^2$) a $F(u,v) = (u(-1+v), \lambda v(1-u))$, cuya matriz jacobiana es

$$DF(u,v) = \begin{pmatrix} -1+v & u \\ -\lambda v & \lambda(1-u) \end{pmatrix}.$$

En el origen, $DF(0,0) = \operatorname{diag}(-1, \lambda)$: un autovalor negativo y otro positivo, luego $(0,0)$ **es inestable** (es un punto de silla: se entra por el eje de los depredadores, que mueren, y se sale por el de las presas, que se reproducen). En $(1,1)$,

$$DF(1,1) = \begin{pmatrix} 0 & 1 \\ -\lambda & 0 \end{pmatrix},$$

con autovalores $\pm\sqrt{\lambda}\, i$, imaginarios puros. **El teorema de linealización no se aplica**: no dice nada cuando hay autovalores con parte real nula.

Una primera información sale de los signos de las componentes de F:

$$\begin{aligned} &u' < 0,\ v' > 0 \quad \text{si } (u,v) \in (0,1) \times (0,1), \\ &u' > 0,\ v' > 0 \quad \text{si } (u,v) \in (0,1) \times (1,\infty), \\ &u' > 0,\ v' < 0 \quad \text{si } (u,v) \in (1,\infty) \times (1,\infty), \\ &u' < 0,\ v' < 0 \quad \text{si } (u,v) \in (1,\infty) \times (0,1). \end{aligned}$$

Mirando la dirección del vector (u', v') se ve que las soluciones parecen girar alrededor de $(1,1)$ en el sentido de las agujas del reloj (con u en el eje horizontal): pocas presas y pocos depredadores, las presas crecen; muchas presas, los depredadores crecen; muchos depredadores, las presas caen; pocas presas, los depredadores caen. Pero esto no aclara si el equilibrio es estable: las órbitas podrían ser espirales que se acercan a $(1,1)$ o que se alejan al infinito.

Las órbitas, explícitamente. Sin una fórmula elemental para las soluciones como funciones del tiempo, podemos obtener una ecuación implícita de sus órbitas. La idea es buscar localmente u como función de v, donde $v' \neq 0$:

$$\frac{du}{dv} = \frac{u'}{v'} = \frac{u(-1+v)}{\lambda v(1-u)} \quad \Longrightarrow \quad \lambda\Big(1-\frac{1}{u}\Big)du = -\Big(1-\frac{1}{v}\Big)dv,$$

que es de variables separadas. Integrando, existe $C \in \mathbb{R}$ con $\lambda(u - \log u) = -(v - \log v) + C$. Es decir: a lo largo de cualquier solución con componentes positivas, la cantidad $\lambda(u(t) - \log u(t)) + (v(t) - \log v(t))$ es constante. La derivación directa por la regla de la cadena, que se hará al estudiar la estabilidad, comprueba la conservación también cuando $v' = 0$. Cambiando la definición de la constante conviene escribirlo así:

$$\lambda\big(u(t) - \log u(t) - 1\big) + \big(v(t) - \log v(t) - 1\big) = C.$$

Esto significa que la órbita está contenida en la curva

$$\Gamma(C) := \{(u,v) \in (0,+\infty)^2 : \lambda\psi(u) + \psi(v) = C\}, \qquad \psi(s) := s - \log s - 1 \quad (s > 0).$$

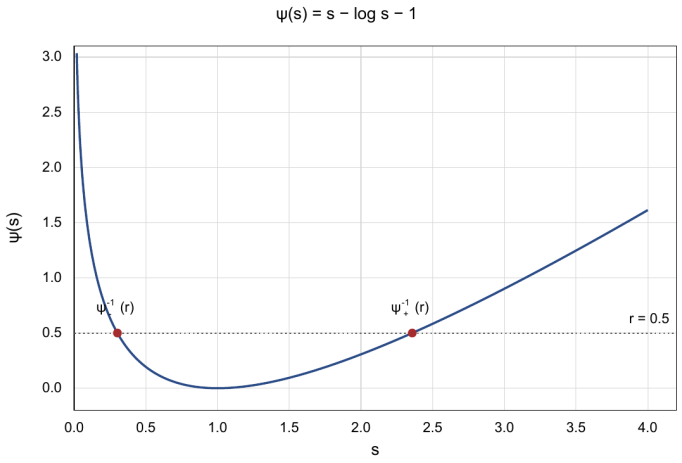


Figura 2.5: La función $\psi(s) = s - \log s - 1$: tiende a $+\infty$ en 0^+ y en $+\infty$, decrece en $(0,1)$, vale 0 en $s = 1$ y crece en $(1,+\infty)$. La ecuación $\psi(s) = r$ con $r > 0$ tiene exactamente dos raíces, $\psi_-^{-1}(r) < 1 < \psi_+^{-1}(r)$.

La figura 2.5 muestra ψ. Se tiene $\lim_{s\to 0^+} \psi(s) = +\infty$, $\lim_{s\to+\infty} \psi(s) = +\infty$, ψ decreciente en $(0,1)$, $\psi(1) = 0$ y creciente en $(1,+\infty)$; todo sale de $\psi'(s) = 1 - 1/s$. Por tanto $\psi \geq 0$, con igualdad solo en 1. Si $u(0) = v(0) = 1$, entonces $C = (\lambda+1)\psi(1) = 0$ y $\Gamma(0)$ es el equilibrio $\{(1,1)\}$. En cualquier otro caso con $u(0), v(0) > 0$ es $C = \lambda\psi(u(0)) + \psi(v(0)) > 0$, y las órbitas del primer cuadrante distintas del equilibrio giran alrededor de él en el sentido de las agujas del reloj y están contenidas en las curvas $\Gamma(C)$. **En particular no pueden describir espirales.**

Las curvas $\Gamma(C)$ con $C > 0$ y la recta $u = 1$ se cortan en dos puntos: los de la forma $(1,s)$ con $\psi(s) = C$, uno con $s = \psi_-^{-1}(C) \in (0,1)$ y otro con $s = \psi_+^{-1}(C) \in (1,\infty)$. Del mismo modo cortan a $v = 1$ en dos puntos. Son curvas cerradas y regulares: el gradiente de $\lambda\psi(u) + \psi(v)$ solo se anula en $(1,1)$, y $\psi''(s) = 1/s^2 > 0$ hace estrictamente convexas sus subregiones de nivel. Así, para $C > 0$, $\Gamma(C)$ es una curva simple que rodea a $(1,1)$. Además es compacta

y está separada de los ejes, porque ψ tiende a $+\infty$ tanto en cero como en infinito. La órbita permanece en ella y, por prolongación en un compacto, está definida para todo tiempo; la curva no contiene equilibrios. Aplicando el **teorema de Poincaré-Bendixson** (teorema de Poincaré–Bendixson del apéndice, §8.5 del libro de EDO: una semiórbita positiva contenida en un compacto sin puntos críticos de un sistema plano tiende en espiral a una órbita cíclica o es ella misma una órbita cíclica), y teniendo en cuenta que no puede haber espirales, se deduce que **las órbitas son periódicas**. La figura 2.6 las muestra. Interesa observar que cuando nos alejamos del equilibrio las órbitas se acercan mucho a los ejes, lo que en la práctica podría conducir a la aniquilación de alguna de las especies.

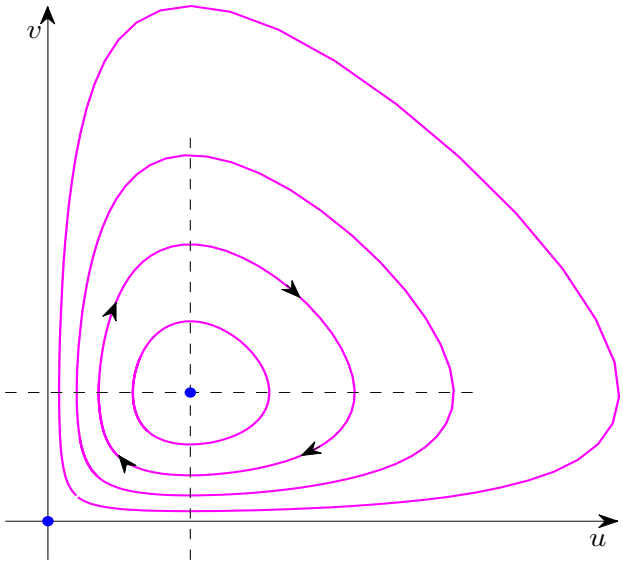


Figura 2.6: Órbitas del sistema de Volterra adimensionalizado con $\lambda = 1$: las curvas $\psi(u) + \psi(v) = C$ para varios valores de C. Todas son cerradas y rodean al equilibrio $(1,1)$. Las exteriores alcanzan valores muy pequeños de cada población en distintos tramos del ciclo. Las flechas indican el sentido horario del recorrido. El dibujo representa posiciones, no el tiempo empleado en recorrerlas.

Además, **el equilibrio** $(1,1)$ **es estable, pero no atractivo**. La estabilidad se deduce del teorema de Liapunov (teorema de Liapunov del apéndice; §7.3 del libro de EDO) sin más que considerar la función

$$V(u,v) := \lambda\psi(u) + \psi(v) \qquad \text{para } (u,v) \in \mathbb{R}_+ \times \mathbb{R}_+ :$$

es definida positiva alrededor de $(1,1)$, porque $\psi \geq 0$ con cero solo en 1, y su derivada respecto del sistema es

$$\begin{aligned} \dot{V} &= \lambda\psi'(u)\,u' + \psi'(v)\,v' = \lambda\Big(1-\frac{1}{u}\Big)u(-1+v) + \Big(1-\frac{1}{v}\Big)\lambda v(1-u) \\ &= \lambda(u-1)(v-1) + \lambda(v-1)(1-u) = 0. \end{aligned}$$

Luego V es una función de Liapunov y $(1,1)$ es estable. No es atractivo porque las órbitas cercanas son cerradas y no se acercan a él.

Conexión: el segundo método de Lyapunov y las integrales primeras

El teorema de Liapunov, reenunciado: si $\bar{y}$ es un equilibrio de $y' = f(y)$ y existe $V \in C^1$ definida positiva en una bola $\bar{B}(\bar{y};\rho)$ con $\dot{V}(y) = \nabla V(y) \cdot f(y) \leq 0$ en ella, entonces $\bar{y}$ es estable; si además $\dot{V}$ es definida negativa, es asintóticamente estable. Está en el libro de EDO, §7.3 «Condiciones suficientes de estabilidad» (página 155), con las definiciones de función de Liapunov y de $\dot{V}$ en §7.2 y la intuición de la energía que no aumenta. Aquí $\dot{V} = 0$: V es una *integral primera*, se conserva exactamente, y por eso las órbitas son sus curvas de nivel. Es la misma situación que la conservación de la energía en el péndulo sin rozamiento del capítulo 3, y la misma técnica: cuando la linealización da autovalores imaginarios puros, buscar una cantidad conservada es la salida.

La idea: la respuesta al Adriático

El promedio de una órbita periódica se puede calcular sin resolver nada. Integrando $u'/u = -1 + v$ a lo largo de un periodo T, el lado izquierdo da $\log u(T) - \log u(0) = 0$, luego $\int_0^T v\,dt = T$: el promedio de presas en un ciclo es exactamente 1 (en las variables originales, a/b), y del mismo modo el promedio de depredadores es 1 (o c/d). Los promedios son los valores del equilibrio y no dependen de la órbita. La pesca indiscriminada quita una fracción de las dos especies: en las variables originales cambia a por $a + \mu$ y c por $c - \mu$. Si $0 < \mu < c$, siguen existiendo ciclos y el promedio de presas pasa a $(a + \mu)/b$ (sube), mientras que el de depredadores pasa a $(c - \mu)/d$ (baja). Reducir la pesca produce el efecto contrario dentro del modelo: es la explicación propuesta para el aumento relativo de depredadores observado en las capturas. Esto es lo que Volterra explicó a su yerno, el biólogo D'Ancona, en 1926.

De dónde viene: Lotka, Volterra y las órbitas cerradas

Vito Volterra escribió el sistema en 1926 para explicar los datos de pesca del Adriático que le había enseñado Umberto D'Ancona; Alfred Lotka lo había obtenido en 1920 para una reacción química oscilante y lo publicó en su libro de 1925, y por eso se llama de Lotka-Volterra. Su familia de órbitas cerradas es frágil: existen perturbaciones arbitrariamente pequeñas del campo que cambian ese retrato de fases. Por ejemplo, añadir autolimitación puede hacer que las órbitas se acerquen a un equilibrio; otros cambios pueden producir ciclos límite. Eso es inestabilidad estructural, aunque no *toda* perturbación destruya los ciclos. La cuestión es qué conclusiones sobreviven al añadir al modelo efectos que se habían omitido. La lección se conserva: Volterra explica *que* presas y depredadores oscilan y por qué las fases van desfasadas, y eso sí se observa en la naturaleza, en los linces y liebres de la Compañía de la Bahía de Hudson.

El camino que no se tomó: añadir autolimitación a las presas

La alternativa. Tomar $\partial^2_{vv} h(0,0) < 0$, es decir, $v' = v(c - du - \kappa v)$: las presas siguen una logística cuando no hay depredadores.
Su ventaja. Es más realista (los conejos no crecen sin límite aunque no haya zorros) y cambia la estabilidad cuando existe equilibrio interior. Este requiere $bc > \kappa a$ y está en $u_* = (c - \kappa a/b)/d$, $v_* = a/b$. Con la función adaptada

$$\widetilde{V} = d\big(u - u_* - u_* \log(u/u_*)\big) + b\big(v - v_* - v_* \log(v/v_*)\big)$$

se obtiene $\dot{\widetilde{V}} = -b\kappa(v - v_*)^2 \leq 0$. El teorema de LaSalle (teorema de LaSalle; §8.4 del libro de EDO) da estabilidad asintótica: en la recta $v = v_*$, el único conjunto invariante es el equilibrio.

La aproximación puede ser en espiral o de tipo nodo; no siempre oscila.
Por qué se deja fuera. Porque el modelo simplificado es el que se puede tratar entero con lápiz: órbitas explícitas, periodicidad, promedios. Y porque el objetivo del tema es enseñar las herramientas (integral primera, Liapunov, Poincaré-Bendixson), no el modelo más fiel. La versión con autolimitación se estudia en la sección 2.2.6 para la competición, y para presa-depredador es el primer ejercicio de la tercera ronda del repaso, al final de este capítulo: pide justamente el equilibrio interior, la linealización y la estabilidad asintótica que se acaban de anunciar.

EXPLÍCATE. *Comprueba con la regla de la cadena que $\dot{V} = 0$, y después di en una frase por qué $\dot{V} = 0$ implica que las órbitas están en las curvas de nivel de V y por qué eso no basta por sí solo para saber que son periódicas: hay que usar además la geometría de esas curvas y la ausencia de equilibrios, como en el argumento de Poincaré-Bendixson.*

2.2.5 Modelo de competición (simplificado)

LA PREGUNTA

Dos especies de paramecios comparten alimento en un cultivo. En sus experimentos de 1934, Gause observó que una podía desplazar a la otra. El modelo simplificado plantea además una pregunta: ¿puede cambiar la ganadora al cambiar las poblaciones iniciales, y qué frontera separa «gana la primera» de «gana la segunda»?

Se considera ahora el modelo

$$\begin{cases} u' = u(a - bv), \\ v' = v(c - du), \end{cases} \qquad a, b, c, d > 0,$$

que corresponde al desarrollo de segundo orden con los signos de competición y $\partial^2_{uu} g(0,0) = \partial^2_{vv} h(0,0) = 0$. Leído término a término: cada especie, sola, crece exponencialmente (au, cv: recursos ilimitados dentro de la especie), y cada encuentro entre las dos (uv) le cuesta a las dos ($-buv$, $-duv$): compiten por lo mismo y el encuentro perjudica a ambas. Lo que se pierde es lo mismo que en Volterra y algo más grave: al no haber autolimitación, una especie sola crece sin tope, y eso es justo lo que hace que el modelo prediga, salvo los datos de una curva especial, que una gana y crece sin cota mientras la otra tiende a cero. Es un defecto del modelo, no de los paramecios de Gause, y se corrige en la sección 2.2.6. Los mismos cambios de variables que en presa-depredador conducen al sistema adimensionalizado

$$\begin{cases} u' = u(1 - v), \\ v' = \lambda v(1 - u). \end{cases}$$

De nuevo hay soluciones semitriviales: si $u_0 = 0$ entonces $u \equiv 0$ y $v(t) = v_0 e^{\lambda t}$; si $v_0 = 0$ entonces $v \equiv 0$ y $u(t) = u_0 e^t$. Las órbitas $\{(0,0)\}$, $\{(u,0) : u > 0\}$ y $\{(0,v) : v > 0\}$ garantizan, como antes, que si el dato inicial tiene componentes positivas entonces $u(t) > 0$ y $v(t) > 0$ allá donde estén definidas.

En principio no se puede deducir directamente que todas las soluciones estén definidas en todo $\mathbb{R}$: cada especie sola crece exponencialmente, y el sistema podría explotar. Gracias a la positividad se puede usar un **criterio de comparación** y probar que el intervalo de definición de toda solución que parte de datos positivos contiene a $[0, +\infty)$.

Teorema. Supongamos $u_0 > 0$ y $v_0 > 0$, y sea $t \mapsto (u(t), v(t))$ la solución maximal del sistema adimensionalizado con $u(0) = u_0$, $v(0) = v_0$, definida en $I(u_0, v_0)$. Entonces

$$0 < u(t) < u_0 e^t, \qquad 0 < v(t) < v_0 e^{\lambda t}, \qquad t \in I(u_0, v_0) \cap (0, +\infty).$$

En $t = 0$ se tiene igualdad en las cotas superiores. En consecuencia, $I(u_0, v_0) \supset [0, +\infty)$.

La idea. La pregunta de fondo es si la competición puede hacer explotar a alguien en tiempo finito. La respuesta se ve sin cuentas: la competición solo *resta*. Cada especie, si estuviera sola y sin freno, crecería exponencialmente; el término $-uv$ le quita población, nunca se la añade. De modo que el crecimiento exponencial libre es un **techo**, y una solución que vive por debajo de un techo finito no puede escaparse a infinito en tiempo finito. La demostración es esa frase, escrita con un factor integrante para poder integrar la desigualdad.

Paso 1: tirar el término de competición deja una desigualdad diferencial (el núcleo). Sabemos que $u(t) > 0$ y $v(t) > 0$ en $I(u_0, v_0)$. Por tanto

$$u'(t) = u(t)\big[1 - v(t)\big] < u(t) \qquad \text{y} \qquad v'(t) = \lambda v(t)\big[1 - u(t)\big] < \lambda v(t)$$

para todo $t \in I(u_0, v_0)$: se ha quitado un término $-uv$ negativo. Ese es todo el contenido del teorema; lo que sigue es la manera de convertir una desigualdad entre derivadas en una desigualdad entre funciones.

Paso 2: el factor integrante convierte la desigualdad en monotonía. De la primera desigualdad, $u'(t) - u(t) < 0$. Multiplicando por $e^{-t} > 0$, $e^{-t}u'(t) - e^{-t}u(t) < 0$, es decir,

$$\frac{d}{dt}\big(e^{-t}u(t)\big) < 0 \qquad \text{para todo } t \in I(u_0, v_0).$$

El truco es el de siempre en las ecuaciones lineales de primer orden: e^{-t} se elige precisamente para que la regla del producto reconstruya el binomio $u' - u$, y así una información sobre la derivada de u pasa a ser una información sobre el *crecimiento* de una función auxiliar, que sí se puede integrar.

Paso 3: integrar y cerrar con el criterio de prolongación. Fijado $t \in I(u_0, v_0) \cap (0, +\infty)$, integrando en $[0, t]$ se obtiene $e^{-t}u(t) - u_0 < 0$, que es la cota para u. La de v es análoga con $e^{-\lambda t}$. Finalmente, si el extremo derecho del intervalo maximal fuera un número finito M, las dos cotas mantendrían la trayectoria, para $0 \le t < M$, en un rectángulo compacto de $\mathbb{R}^2$. El campo es polinómico y está definido en todo ese espacio: la proposición de prolongación mediante cotas del apéndice permitiría prolongar la solución más allá de M, contradiciendo la maximalidad. Esta es la caracterización de §3.4 del libro de EDO. ■

Qué hace cada hipótesis. La positividad de u_0 y v_0 es lo único que se pide, y hace dos cosas distintas. Primero garantiza que u y v no cambian de signo, y sin eso el paso 1 se invierte: si v fuera negativa, el término $-uv$ pasaría a *sumar* y el techo exponencial dejaría de serlo. Y segundo, es la que da sentido al modelo, porque una población negativa no significa nada. Conviene notar también qué es lo que *no* sostiene la cota: sirve para $t \ge 0$ y no dice nada hacia atrás. Hacia el pasado el signo de la desigualdad se invierte y las soluciones sí pueden explotar en tiempo finito; por eso el enunciado habla de $[0, +\infty)$ y no de $\mathbb{R}$.

Lo que hay que llevarse. Es el método que sirve cada vez que hay que probar que una solución maximal existe en todo $[0, +\infty)$, y tiene tres pasos. (i) *Signo*: primero la positividad, porque es la que dice qué términos restan. (ii) *Techo*: se tiran los términos que restan, queda una desigualdad lineal como $u' < u$, y el factor integrante la convierte en $u(t) < u_0 e^t$. (iii) *Compacto*: en cualquier intervalo acotado $[0, M)$ las cotas encierran la solución en un rectángulo, y una solución maximal no puede terminar dentro de un compacto del dominio. La cota no tiene que ser uniforme en el tiempo —$u_0 e^t$ no lo es—: basta con que no explote en tiempo finito. Con el mismo esquema se prueba la existencia global en el SIR con mortalidad de la subsección 2.2.2, donde el techo lo da una cantidad conservada, y en el modelo general de competición de la subsección siguiente, donde lo dan dos logísticas.

CONEXIÓN: COMPARACIÓN DE SOLUCIONES

El argumento del factor integrante es la demostración del criterio de comparación escalar del apéndice (teorema de comparación escalar): si $f \leq g$ y las soluciones de $y' = f(t, y)$ y $y' = g(t, y)$ salen del mismo dato, la primera queda por debajo de la segunda para $t \geq t_0$. Aquí $g(t, u) = u$ y f es el segundo miembro real, que es menor. Está en el libro de EDO, §3.5 «Comparación de soluciones» (página 51, segundo teorema), y la caracterización de las soluciones prolongables, que es la otra pieza que se usa, en §3.4.

Equilibrios. Como en presa-depredador, los puntos críticos son $(0, 0)$ y $(1, 1)$. Su estabilidad se puede estudiar ahora con los sistemas lineales perturbados:

$$DF(u, v) = \begin{pmatrix} 1 - v & -u \\ -\lambda v & \lambda(1 - u) \end{pmatrix}.$$

$$DF(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \qquad DF(1, 1) = \begin{pmatrix} 0 & -1 \\ -\lambda & 0 \end{pmatrix}.$$

En el origen los dos autovalores son positivos: $(0, 0)$ **es inestable**, y de hecho repulsivo, las soluciones se alejan de él en todas las direcciones. En $(1, 1)$ los autovalores son $\pm\sqrt{\lambda}$, reales de signos opuestos: $(1, 1)$ **es inestable**, un punto de silla, repulsivo en la dirección del autovector del autovalor positivo y atractivo en la del negativo.

De los signos de u' y v' en el plano de fases:

$$\begin{aligned} &u' > 0,\ v' > 0 \quad \text{si } (u, v) \in (0, 1) \times (0, 1), \\ &u' < 0,\ v' > 0 \quad \text{si } (u, v) \in (0, 1) \times (1, \infty), \\ &u' < 0,\ v' < 0 \quad \text{si } (u, v) \in (1, \infty) \times (1, \infty), \\ &u' > 0,\ v' < 0 \quad \text{si } (u, v) \in (1, \infty) \times (0, 1). \end{aligned}$$

Una consecuencia es que **no hay órbitas periódicas no constantes**. Para verlo sin adivinar el recorrido, supongamos que u es periódica: en su máximo y en su mínimo, $u' = 0$ obliga a $v = 1$, y allí $u'' = -\lambda u(1 - u)$. En el máximo esto exige $u \leq 1$; en el mínimo, $u \geq 1$. Solo es posible $u \equiv 1$, que obliga a $v \equiv 1$. La clasificación de las curvas de nivel que sigue precisará los límites de todas las órbitas; la mera ausencia de ciclos, en un sistema plano general, no excluye conexiones entre equilibrios.

Las órbitas. Operando con las EDOs como antes,

$$\frac{du}{dv} = \frac{u(1 - v)}{\lambda v(1 - u)} \quad \Longrightarrow \quad \lambda\Big(1 - \frac{1}{u}\Big)du = \Big(1 - \frac{1}{v}\Big)dv,$$

y las órbitas están contenidas en las curvas

$$\Lambda(C) := \{(u,v) \in \mathbb{R}^2_+ : \lambda\psi(u) - \psi(v) = C\}.$$

La conservación vale también donde se anulaba el denominador anterior: basta derivar $\lambda\psi(u) - \psi(v)$ a lo largo de la solución y se obtiene cero. La diferencia con Volterra es un signo: ahora es una *resta*. Para esta función separable, las ramas que vamos a describir son abiertas, en vez de los óvalos de Volterra. Se estudian según el valor de C, usando que para cada $r > 0$ la ecuación $\psi(s) = r$ tiene dos raíces, $\psi_-^{-1}(r) \in (0,1)$ y $\psi_+^{-1}(r) \in (1,+\infty)$.

- **Caso** $C = 0$**.** Las curvas $\lambda\psi(u) = \psi(v)$ contienen al punto $(1,1)$ y

$$\Lambda(0) = \{(1,1)\} \cup \Lambda_{--}(0) \cup \Lambda_{-+}(0) \cup \Lambda_{+-}(0) \cup \Lambda_{++}(0),$$

  donde

$$\begin{aligned}
\Lambda_{--}(0) &:= \{(u,v) : 0 < u < 1,\ v = \psi_-^{-1}(\lambda\psi(u))\},\\
\Lambda_{-+}(0) &:= \{(u,v) : u > 1,\ v = \psi_-^{-1}(\lambda\psi(u))\},\\
\Lambda_{+-}(0) &:= \{(u,v) : 0 < u < 1,\ v = \psi_+^{-1}(\lambda\psi(u))\},\\
\Lambda_{++}(0) &:= \{(u,v) : u > 1,\ v = \psi_+^{-1}(\lambda\psi(u))\}.
\end{aligned}$$

  Así, $\Lambda(0)$ es la unión del equilibrio y cuatro curvas: $\Lambda_{--}(0)$ va de $(0,0)$ a $(1,1)$ y las otras tres conectan geométricamente $(1,1)$ con el infinito. Esto describe la forma de las curvas; el sentido del tiempo se determina después. Los dos subíndices se leen así: el primero dice qué rama de ψ^{-1} da v, es decir, si $v < 1$ (signo $-$) o $v > 1$ (signo $+$); el segundo dice de qué lado de 1 está u. Cada curva vive, por tanto, en uno de los cuatro cuadrantes que las rectas $u = 1$ y $v = 1$ dibujan alrededor de la silla: $\Lambda_{--}(0)$ en $u < 1,\ v < 1$; $\Lambda_{-+}(0)$ en $u > 1,\ v < 1$; $\Lambda_{+-}(0)$ en $u < 1,\ v > 1$; y $\Lambda_{++}(0)$ en $u > 1,\ v > 1$.
- **Caso** $C > 0$**.** Entonces $C + \psi(v) > 0$ para todo $v > 0$, y se puede despejar u:

$$u = \psi_+^{-1}\Big(\frac{C+\psi(v)}{\lambda}\Big) \qquad \text{o bien} \qquad u = \psi_-^{-1}\Big(\frac{C+\psi(v)}{\lambda}\Big),$$

  y $\Lambda(C) = \Lambda_-(C) \cup \Lambda_+(C)$ con $\Lambda_\pm(C)$ la gráfica de cada una de esas dos funciones de $v > 0$. Las dos curvas no tienen ningún punto en común.
- **Caso** $C < 0$**.** Entonces $\lambda\psi(u) - C > 0$ para todo $u > 0$, se despeja v, $v = \psi_\pm^{-1}(\lambda\psi(u) - C)$, y $\Lambda(C) = \Sigma_-(C) \cup \Sigma_+(C)$, dos curvas sin puntos comunes.

Las curvas con $C \neq 0$ no tienen autointersecciones ni contienen puntos críticos; en consecuencia cada una **coincide con la órbita completa** de cualquiera de sus puntos. Con las propiedades de ψ se hace el análisis cualitativo para datos $u_0, v_0 > 0$, que se visualiza en la figura 2.7.

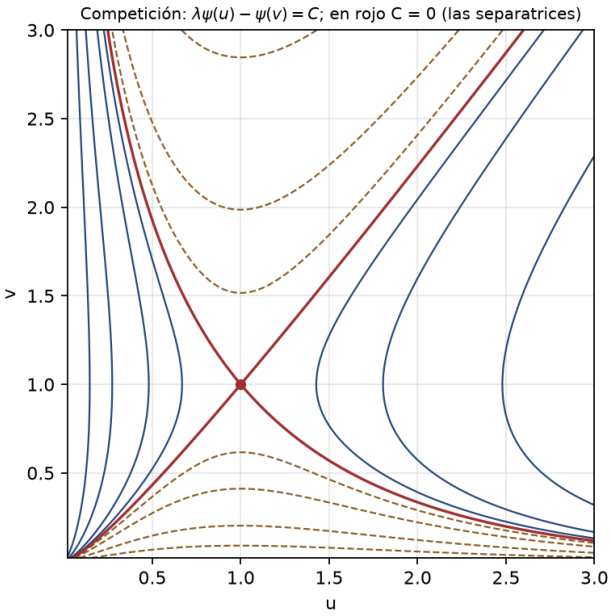


Figura 2.7: Curvas $\lambda\psi(u) - \psi(v) = C$ para $\lambda = 1.4$ y distintos valores de C. En rojo, $C = 0$: dos ramas llegan al punto de silla $(1, 1)$ y dos se alejan de él. En azul, $C > 0$; en ocre, $C < 0$. Cada rama sin el equilibrio es una órbita; el equilibrio es otra. Los colores distinguen niveles, no la especie ganadora.

Caso 1: $C = \lambda\psi(u_0) - \psi(v_0) > 0$**.** Entonces (u_0, v_0), y con él toda la órbita, está en $\Lambda_-(C)$ o en $\Lambda_+(C)$.

- Si está en $\Lambda_-(C)$, $u(t) = \psi_-^{-1}\big((C + \psi(v(t)))/\lambda\big)$ para todo t. Cuando $v \to 0^+$ o $v \to +\infty$, $\psi(v) \to +\infty$ y ψ_-^{-1} de algo grande tiende a 0: $u \to 0$. Con los signos de los cuadrantes, esto significa que para $t \to -\infty$ las dos tienden al equilibrio $(0, 0)$; después ambas crecen hasta un instante t_1 en que $v(t_1) = 1$ y $u(t_1) < 1$, y a partir de ahí v sigue creciendo hasta $+\infty$ mientras u decrece a 0. **Una población se extingue y la otra crece indefinidamente.**
- Si está en $\Lambda_+(C)$, $u = \psi_+^{-1}\big((C + \psi(v))/\lambda\big)$, y ahora cuando $v \to 0^+$ o $v \to +\infty$ es $u \to +\infty$. Al retroceder al extremo izquierdo del intervalo maximal, ambas crecen sin cota: no se puede sustituir ese extremo por $-\infty$, pues puede haber explosión hacia atrás. Cuando t crece ambas decrecen hasta un t_1 con $v(t_1) = 1$, $u(t_1) > 1$, y después u crece hacia $+\infty$ mientras v decrece a cero. Esto no implica $u(t) > v(t)$ en todo instante: son poblaciones adimensionalizadas con escalas distintas.

Caso 2: $C < 0$**.** El análisis es análogo intercambiando los papeles de u y v.

Caso 3: $C = 0$**.** Ahora hay cuatro curvas.

- $\Lambda_{--}(0)$ es estrictamente creciente, con $u \to 0^+$ cuando $v \to 0^+$ y $u \to 1$ cuando $v \to 1^-$. Cuando $t \to -\infty$ las dos poblaciones se acercan a $(0, 0)$; ambas crecen siempre y cuando $t \to +\infty$ se acercan a $(1, 1)$. **No hay extinción.**
- $\Lambda_{-+}(0)$ es estrictamente decreciente, con $v \to 1$ cuando $u \to 1^+$ y $v \to 0$ cuando $u \to +\infty$. Cuando $t \to -\infty$ se acercan a $(1, 1)$; cuando $t \to +\infty$, u crece a $+\infty$ y v decrece a 0: **la población v se extingue.**

- $\Lambda_{+-}(0)$ es estrictamente decreciente, con $v \to +\infty$ cuando $u \to 0^+$ y $v \to 1$ cuando $u \to 1^-$. Cuando $t \to -\infty$ se acercan a $(1,1)$; cuando t crece, u decrece a 0 (**se extingue**) y v crece a $+\infty$.
- $\Lambda_{++}(0)$ es una gráfica estrictamente creciente: $v \to 1$ cuando $u \to 1^+$ y $v \to +\infty$ cuando $u \to +\infty$. Recorrida hacia atrás, ambas poblaciones tienden a infinito al acercarse al extremo izquierdo del intervalo maximal. Recorrida hacia delante, ambas decrecen estrictamente y tienden a 1 cuando $t \to +\infty$. **Ninguna se extingue.**

  Las cuatro ramas, puestas en fila, son la cruz roja de la figura 2.7 recorrida con flechas:

rama	cuadrante	viene de	va a	quién se extingue
$\Lambda_{--}(0)$	$u<1,\ v<1$	$(0,0)$	$(1,1)$	nadie
$\Lambda_{++}(0)$	$u>1,\ v>1$	el infinito	$(1,1)$	nadie
$\Lambda_{-+}(0)$	$u>1,\ v<1$	$(1,1)$	$u\to\infty,\ v\to 0$	v
$\Lambda_{+-}(0)$	$u<1,\ v>1$	$(1,1)$	$u\to 0,\ v\to\infty$	u

Las dos primeras *llegan* a la silla: con ella forman su variedad estable. Las dos últimas *salen* de ella: son su variedad inestable, y cada una acaba en la victoria de una especie. Una integración numérica con $\lambda = 1.4$ desde un punto de cada rama lo confirma, con una precaución: la silla repele los errores de redondeo, así que una trayectoria calculada sobre $\Lambda_{--}(0)$ se acerca a $(1,1)$ durante un rato y después se escapa por una de las ramas inestables. Es el mismo fenómeno que obliga a integrar hacia atrás en la práctica 13.

LA IDEA: LA FRONTERA DE GAUSE

La frontera son las dos ramas rojas *crecientes*, junto con $(1,1)$: $\Lambda_{--}(0) \cup \{(1,1)\} \cup \Lambda_{++}(0)$, la variedad estable de la silla. Los datos sobre ella tienden a la coexistencia; a sus lados gana una u otra especie. Tiene área cero, pero eso no significa que sea imposible elegir un dato sobre ella. El signo de C no decide quién gana: un mismo signo aparece en ramas distintas.

El caso $\lambda = 1$ permite verlo de un vistazo. Restando las ecuaciones, $(u-v)' = u - v$, luego $u(t) - v(t) = e^t(u_0 - v_0)$: la frontera es la recta $u = v$. Los datos $(2,1)$ y $(1/2,1)$ tienen ambos $C > 0$, pero gana u en el primero y v en el segundo. Para otros λ, la frontera suele ser curva.

El modelo muestra así exclusión competitiva salvo sobre esa frontera; no es una ley universal de toda interacción entre especies. La comparación con Leslie debe hacerse con precisión: para una matriz de Leslie primitiva, lo que se independiza del dato positivo es el perfil relativo límite, no la magnitud total de la población.

Observación. El modelo contempla casos en que al menos una población tiende a $+\infty$, lo que no es realista por la limitación de recursos. Se debe a la excesiva simplificación del desarrollo de segundo orden: tendría sentido suponer los signos de competición junto con $\partial^2_{uu} g(0,0) < 0$ y/o $\partial^2_{vv} h(0,0) < 0$, es decir, añadir términos que conviertan en logísticas las EDOs cuando una de las poblaciones está ausente. Es lo que se hace en la sección siguiente.

Observación. Todo lo dicho para la competición simplificada se adapta a los modelos de simbiosis sin más que invertir el sentido del tiempo. En efecto, el cambio $t \to -t$ lleva el sistema $u' = u(-a+bv)$, $v' = v(-c+du)$ de la simbiosis simplificada al de competición $u' = u(a-bv)$, $v' = v(c-du)$. Las órbitas son las mismas recorridas al revés. En simbiosis simplificada puede haber crecimiento conjunto sin cota, pero también extinción conjunta: por

ejemplo, si ambas poblaciones adimensionales empiezan por debajo de 1, las dos decrecen hacia cero. La inversión del tiempo no convierte todas las trayectorias en crecimiento.

EXPLÍCATE. *Sin mirar las fórmulas: en el cuadrante $(1,\infty)\times(0,1)$, ¿hacia dónde se mueve una órbita, y qué le puede pasar a largo plazo? Compáralo con Volterra en el mismo cuadrante.*

2.2.6 Modelo general de competición

LA PREGUNTA

Si cada una de las dos especies, sola, se estabiliza en su propia capacidad de soporte, ¿puede la competencia hacer que las dos coexistan de forma estable, o sigue ganando siempre una? ¿De qué dos números depende?

Con el objetivo de completar la información se considera el modelo, ya adimensionalizado,

$$\begin{cases} u' = u(1-v-\alpha u), \\ v' = \lambda v(1-u-\beta v), \\ u(0)=u_0,\ v(0)=v_0, \end{cases}$$

con $\alpha,\beta,\lambda>0$: una generalización del anterior que puede ser más adecuada y realista. Los términos nuevos son $-\alpha u^2$ y $-\lambda\beta v^2$, asociados a las derivadas $\partial^2_{uu}g(0,0)$ y $\partial^2_{vv}h(0,0)$ que el modelo simplificado había puesto a cero: cada especie se estorba a sí misma, como en la logística, con $1/\alpha$ y $1/\beta$ como capacidades de soporte propias. Se recupera así lo que se había perdido, que una especie sola no crezca sin límite. En general ya no disponemos de la integral primera anterior: el estudio se apoya en el análisis cualitativo. Sí podemos calcular las soluciones semitriviales y los puntos críticos. Se obtienen los puntos

$$(0,0), \qquad (0,1/\beta), \qquad (1/\alpha,0),$$

y las órbitas semitriviales $\{(0,v):0<v<1/\beta\}$, $\{(0,v):v>1/\beta\}$, $\{(u,0):0<u<1/\alpha\}$ y $\{(u,0):u>1/\alpha\}$: cada especie sola sigue una logística con capacidad de soporte $1/\alpha$ o $1/\beta$. Cuando $\alpha\beta\neq 1$ y además $\alpha,\beta<1$ o $\alpha,\beta>1$ hay un punto crítico adicional en el interior del cuadrante,

$$\Big(\frac{\beta-1}{\alpha\beta-1},\ \frac{\alpha-1}{\alpha\beta-1}\Big),$$

que sale de resolver $1-v-\alpha u=0$, $1-u-\beta v=0$; las condiciones sobre α,β son las que hacen positivas las dos coordenadas. Si $\alpha\beta=1$, las rectas son paralelas; solo coinciden cuando $\alpha=\beta=1$. En ese caso hay una infinidad adicional de equilibrios: el segmento $\bar u+\bar v=1$, con $\bar u,\bar v\geq 0$. Para comprobar la excepción basta multiplicar $u+\beta v=1$ por α: su lado izquierdo pasa a ser $\alpha u+v$, pero el derecho es α, y debe valer 1.

La órbita de toda solución que parte de (u_0,v_0) con $u_0,v_0>0$ permanece en el primer cuadrante, por el mismo argumento de las órbitas semitriviales. Esto conduce a las desigualdades

$$u'\leq u(1-\alpha u), \qquad v'\leq \lambda v(1-\beta v),$$

y aplicando el criterio de comparación (teorema de comparación escalar; §3.5 del libro de EDO) con las logísticas de la derecha, cuyas soluciones tienden a $1/\alpha$ y $1/\beta$, la solución

maximal verifica $u(t) \leq \max\{u_0, 1/\alpha\}$ y $v(t) \leq \max\{v_0, 1/\beta\}$ para $t \geq 0$ mientras esté definida. Por prolongación en el rectángulo compacto que dan estas cotas, su intervalo de definición contiene a $[0, +\infty)$. Ahora **ninguna población tiende a infinito**.

Se supone por simplicidad $\lambda = 1$. Con argumentos adecuados se prueba el siguiente comportamiento cualitativo, que cambia en función de α y β y se resume en la figura 2.8.

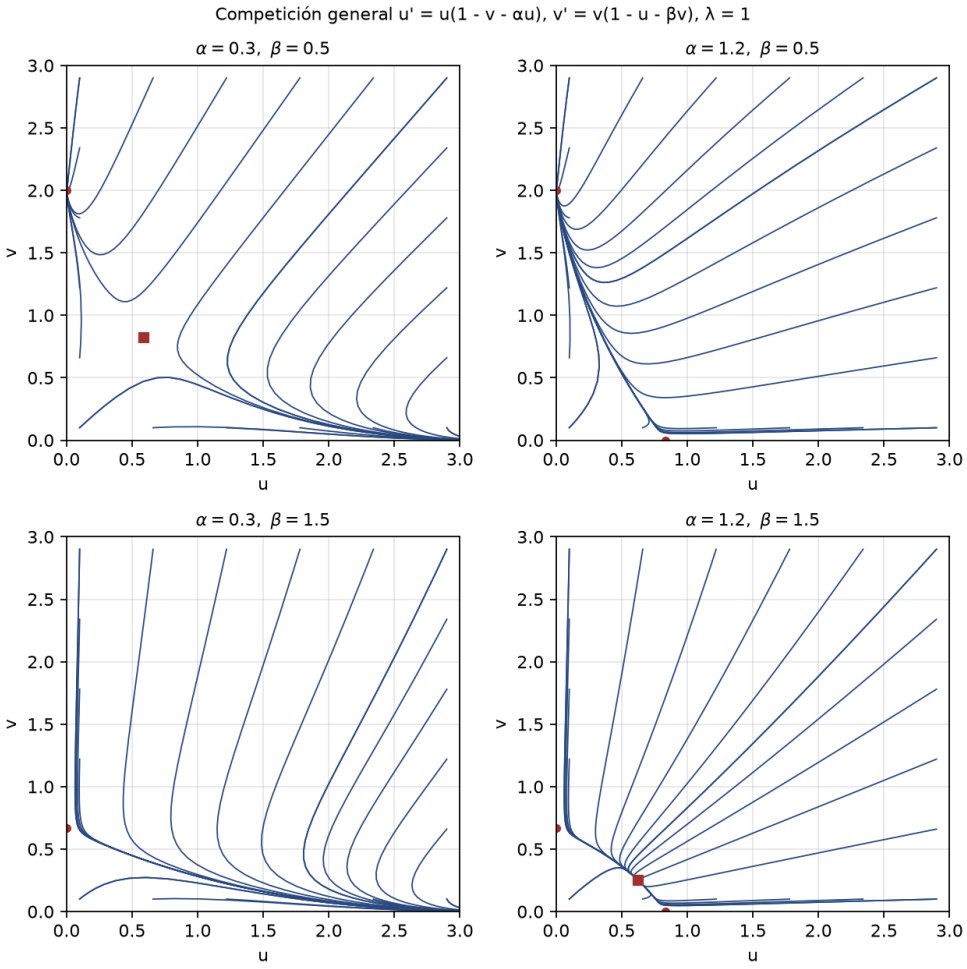


Figura 2.8: Trayectorias del modelo de competición general con $\lambda = 1$ y los cuatro pares (α, β) característicos. Arriba a la izquierda, $\alpha, \beta < 1$: el equilibrio interior (cuadrado) es una silla y casi toda órbita acaba en un eje. Arriba a la derecha y abajo a la izquierda, un α o β mayor que 1: gana siempre la misma especie. Abajo a la derecha, $\alpha, \beta > 1$: el equilibrio interior atrae a todo el cuadrante abierto y las dos coexisten. Los cuatro comportamientos se refieren a datos con las dos componentes positivas.

Caso $\alpha, \beta < 1$. Si $(1-\beta)v_0 = (1-\alpha)u_0$, la solución converge al equilibrio interior. Pero si $(1-\beta)v_0 > (1-\alpha)u_0$ entonces $(u(t), v(t)) \to (0, 1/\beta)$, y si $(1-\beta)v_0 < (1-\alpha)u_0$ entonces $(u(t), v(t)) \to (1/\alpha, 0)$. Aunque ahora ninguna población tiende a infinito, salvo en condiciones muy especiales una de las dos se acaba extinguiendo.

Caso $\alpha < 1$, $\beta \geq 1$, o $\alpha = 1$, $\beta > 1$ (los casos simétricos $\alpha \geq 1$, $\beta < 1$ y $\alpha > 1$, $\beta = 1$ son similares). Se tiene siempre $u(t) \to 1/\alpha$ y $v(t) \to 0$: la población v siempre acaba en

extinción.

Caso $\alpha = \beta = 1$. Hay una infinidad de puntos críticos $(\bar{u}, \bar{v})$ con $\bar{u} + \bar{v} = 1$, además de $(0, 0)$. Este caso es *degenerado* y se resuelve entero con una observación: con $\alpha = \beta = \lambda = 1$ las dos ecuaciones se escriben $u' = u(1 - u - v)$ y $v' = v(1 - u - v)$, es decir, comparten el mismo factor. Por tanto

$$\Big(\frac{u}{v}\Big)' = \frac{u'v - uv'}{v^2} = \frac{uv(1-u-v) - uv(1-u-v)}{v^2} = 0 :$$

el cociente u/v se conserva, cada semirrecta que sale del origen es invariante, y sobre ella la suma $s = u + v$ cumple la logística $s' = s(1 - s)$ y tiende a 1. La conclusión es que toda solución con $u_0, v_0 > 0$ converge al punto del segmento $\bar{u} + \bar{v} = 1$ que está en su propia semirrecta,

$$\lim_{t \to +\infty} (u(t), v(t)) = \Big(\frac{u_0}{u_0 + v_0}, \frac{v_0}{u_0 + v_0}\Big),$$

y **ninguna de las dos poblaciones se extingue**: la proporción inicial se conserva exactamente y solo cambia el tamaño total. La bisectriz es el caso $u_0 = v_0$ y da el límite $(1/2, 1/2)$, con las dos órbitas $\gamma = \{(u, u) : 0 < u < 1/2\}$, que viene de $(0, 0)$, y $\tilde{\gamma} = \{(u, u) : u > 1/2\}$, que viene del infinito. Este caso es la frontera entre la exclusión de $\alpha, \beta < 1$ y la coexistencia de $\alpha, \beta > 1$, y no se parece a ninguno de los dos: en vez de un punto de llegada hay un segmento entero de ellos, y el dato inicial elige cuál. Basta mover α o β una milésima para que el segmento se rompa: según la perturbación, aparece exclusión o coexistencia en un único equilibrio; es el ejemplo más limpio del capítulo de un modelo estructuralmente inestable.

Caso $\alpha, \beta > 1$. Se tiene siempre

$$\lim_{t \to +\infty} (u(t), v(t)) = \Big(\frac{\beta - 1}{\alpha\beta - 1}, \frac{\alpha - 1}{\alpha\beta - 1}\Big) :$$

ambas poblaciones acaban en equilibrio sin que ninguna resulte aniquilada, a pesar de que compitan.

LA IDEA: CUÁNDO SE PUEDE COEXISTIR

α mide cuánto se frena u a sí misma comparado con cuánto la frena v (en las variables adimensionales, el coeficiente de v en u' es 1). Si $\alpha > 1$ y $\beta > 1$, cada especie se estorba a sí misma *más* de lo que estorba a la otra: al crecer, se frena antes de poder eliminar a la competidora, y las dos caben. Si $\alpha, \beta < 1$, cada una estorba más a la otra que a sí misma, y gana la que empieza con ventaja. La respuesta a la pregunta son esos dos números, y la coexistencia estable es posible exactamente cuando la competencia dentro de cada especie es más fuerte que la competencia entre especies. Es la lectura que el modelo da a ocupar nichos suficientemente distintos; fuera de él, la coexistencia puede depender también de otras interacciones y de la variación del ambiente.

Ejercicio 2.4. *Linealización en los equilibrios del modelo general*

ORIGEN: construido sobre la sección 2.2.5 del curso, que da los resultados sin justificación; el ejercicio justifica los que se pueden justificar por linealización.

Con $\lambda = 1$, calcular la matriz jacobiana del modelo general en $(0,0)$, $(1/\alpha, 0)$, $(0, 1/\beta)$ y, cuando exista, en el equilibrio interior, y clasificar su estabilidad en los cuatro casos de la figura. Comprobar que el veredicto coincide con lo que muestra la figura.

PISTAS, DE MENOR A MAYOR

1. *$F(u,v) = (u - uv - \alpha u^2,\ v - uv - \beta v^2)$; deriva y evalúa.*
2. *En los equilibrios de los ejes la matriz es triangular: los autovalores están en la diagonal.*
3. *En el interior, usa que $1 - v - \alpha u = 0$ y $1 - u - \beta v = 0$ para simplificar; el determinante decide si es silla (< 0) o nodo, y la traza el signo.*

Resolución.

La idea, y el atajo que evita todos los autovalores. En dimensión dos no hace falta calcular autovalores para clasificar un equilibrio: basta con **la traza y el determinante**, porque el determinante es el producto de los dos autovalores y la traza su suma. Si $\det < 0$, los autovalores son reales de signos opuestos y el equilibrio es una silla; si $\det > 0$, los autovalores reales tienen el mismo signo; si son complejos conjugados, su parte real es la mitad de la traza. Por tanto, con $\det > 0$ y traza negativa hay estabilidad asintótica en ambos casos. Si la traza es cero, ese criterio no decide. Toda la clasificación de este ejercicio sale de ahí, y en tres de los cuatro puntos ni siquiera hace falta eso, porque la matriz sale triangular y los autovalores están en la diagonal.

El otro movimiento, y es el que ahorra el trabajo sucio, es **usar las ecuaciones del equilibrio para simplificar la jacobiana antes de mirarla**. En el punto interior se cumple $1 - v^* - \alpha u^* = 0$, así que $1 - v^* - 2\alpha u^*$, que es lo que aparece en la matriz, no es más que $-\alpha u^*$. La jacobiana pasa de tener entradas ilegibles a tener cuatro términos limpios, y el determinante sale factorizado de un golpe. Es una técnica general: en un equilibrio, todo lo que se anula por definición se puede restar de la jacobiana.

Y la lectura que hay que retener antes de hacer nada: **la condición de coexistencia $\alpha, \beta > 1$ se lee en el determinante positivo del equilibrio interior**. Hay que exigir también que ese equilibrio exista en el cuadrante: el producto $\alpha\beta > 1$, por sí solo, no lo garantiza. Con α y β grandes cada especie se estorba sobre todo a sí misma, y entonces pueden convivir; con α y β pequeños se estorban más entre ellas que a sí mismas, el punto interior se vuelve silla y gana una de las dos según de qué lado se empiece.

Paso 1: la jacobiana general. $DF(u,v) = \begin{pmatrix} 1 - v - 2\alpha u & -u \\ -v & 1 - u - 2\beta v \end{pmatrix}$.

Paso 2: los tres equilibrios de los ejes, donde la matriz es triangular. Origen. $DF(0,0) = I$: dos autovalores iguales a 1, inestable (fuente), en todos los casos.

$(1/\alpha, 0)$. $DF = \begin{pmatrix} -1 & -1/\alpha \\ 0 & 1 - 1/\alpha \end{pmatrix}$, autovalores -1 y $1 - 1/\alpha$. Si $\alpha > 1$, el segundo es positivo: silla, inestable. Si $\alpha < 1$, negativo: nodo asintóticamente estable. Coincide con la figura: con $\alpha < 1$, el estado con solo u atrae a los datos cercanos (casos primero y tercero). En el primero también existe otra cuenca de atracción: estabilidad local no significa que gane desde todo

dato.
$(0, 1/\beta)$. Simétrico: autovalores $1 - 1/\beta$ y -1. Estable si $\beta < 1$.
Paso 3 (el núcleo): el equilibrio interior, simplificado con sus propias ecuaciones. Interior, $u^* = \frac{\beta-1}{\alpha\beta-1}$, $v^* = \frac{\alpha-1}{\alpha\beta-1}$. Usando las ecuaciones del equilibrio, $1 - v^* - 2\alpha u^* = -\alpha u^*$ y $1 - u^* - 2\beta v^* = -\beta v^*$, luego

$$DF(u^*, v^*) = \begin{pmatrix} -\alpha u^* & -u^* \\ -v^* & -\beta v^* \end{pmatrix}.$$

$$\text{tr} = -(\alpha u^* + \beta v^*) < 0, \qquad \det = u^* v^* (\alpha\beta - 1).$$

Si $\alpha, \beta > 1$, det > 0 y tr < 0: los dos autovalores tienen parte real negativa y el equilibrio es exponencialmente asintóticamente estable: coexistencia. Si $\alpha, \beta < 1$, det < 0: silla, inestable, y su variedad estable es la semirrecta positiva $(1 - \beta)v = (1 - \alpha)u$. Para comprobar su invariancia, si $v = ku$, con $k = (1 - \alpha)/(1 - \beta) > 0$, las dos tasas per cápita coinciden: $1 - v - \alpha u = 1 - u - \beta v$. Así $v' = ku'$ y la trayectoria permanece en ella. Todo cuadra con la figura.
Dónde llega la linealización y dónde no. Lo que la linealización no da es la afirmación *global* de que toda órbita acaba en tal o cual punto, que necesita argumentos globales: en esencia, que no hay órbitas cíclicas ni conexiones que formen un circuito, y localizar los posibles conjuntos límite. La linealización por sí sola tampoco decide los casos frontera con autovalor cero.
Dónde se falla.

- *Al no usar las ecuaciones del equilibrio.* Sustituir directamente $u^* = \frac{\beta-1}{\alpha\beta-1}$ y $v^* = \frac{\alpha-1}{\alpha\beta-1}$ en $1 - v - 2\alpha u$ lleva a fracciones que casi nunca se simplifican bien. Restando antes $1 - v^* - \alpha u^* = 0$, la entrada es $-\alpha u^*$ sin más.
- *Al leer la coexistencia en $\alpha\beta > 1$ a secas.* Con $\alpha = 4$ y $\beta = 1/2$ es $\alpha\beta = 2 > 1$, pero el punto «interior» es $(u^*, v^*) = (-1/2, 3)$: no está en el cuadrante, y el caso es de los que gana siempre la misma especie, aquí v. Primero se comprueba que el equilibrio existe en el cuadrante; después se mira el determinante.
- *Al tomar por global la estabilidad local.* Con $\alpha, \beta < 1$ los dos equilibrios de los ejes son asintóticamente estables a la vez, y cada uno atrae solo a los datos de su lado de la variedad estable de la silla. «$(1/\alpha, 0)$ es estable, luego gana u» es falso para todos los datos del otro lado.

QUÉ DEJA. En un sistema plano, traza y determinante de la jacobiana clasifican el equilibrio sin calcular autovalores: det < 0 es silla; det > 0 con tr < 0 da estabilidad asintótica. En el equilibrio interior de este modelo, una vez comprobada su existencia, el cambio de signo de $\alpha\beta - 1$ separa la silla de la coexistencia estable.

2.2.7 Modelo de simbiosis

LA PREGUNTA

La clasificación de la subsección 2.2.3 dejaba tres tipos y solo se han estudiado dos. Falta el tercero: dos especies que se ayudan, cada una necesitando a la otra para sobrevivir. ¿Basta con que se ayuden para garantizar la supervivencia? ¿Puede haber un umbral por debajo del cual las dos desaparezcan?

Se llama **simbiosis obligada** a la relación en la que ninguna de las dos especies sobrevive sola. Asociaciones como los líquenes, las micorrizas o las termitas con sus microorganismos permiten pensar en la ayuda mutua; que la dependencia sea obligatoria para *ambas* especies es una hipótesis adicional que debe comprobarse en cada caso. El modelo, con las dos derivadas cruzadas positivas como manda la clasificación, y con un término de autolimitación en la segunda especie, es

$$\begin{cases} \bar{u}' = \bar{u}(-a + b\bar{v}), \\ \bar{v}' = \bar{v}(-c + d\bar{u} - e\bar{v}), \end{cases} \qquad a, b, c, d, e > 0.$$

Los signos dicen la historia entera: los términos $-a\bar{u}$ y $-c\bar{v}$ son la muerte de cada especie *en ausencia de la otra*, que es lo que significa «obligada»; $+b\bar{u}\bar{v}$ y $+d\bar{u}\bar{v}$ son la ayuda mutua; y $-e\bar{v}^2$ es el $-\kappa y^2$ del logístico, la competencia por los recursos dentro de la propia especie $\bar{v}$.

Paso previo: quitar constantes. Hay cinco constantes y solo dos son esenciales. Midiendo cada población y el tiempo en unidades propias —$\bar{u} = Au$, $\bar{v} = Bv$, $t = \tau s$— se pueden fijar tres constantes a 1, porque hay tres unidades libres. Es el método de la subsección 2.2.4 con una diferencia de escritura: allí la variable nueva era un múltiplo de la antigua, $\hat{u} = p\,u$; aquí se escribe la antigua como múltiplo de la nueva, $\bar{u} = Au$, así que A hace el papel de $1/p$ y τ el de r. Las cuentas salen de sustituir y aplicar la regla de la cadena: $u' = \tau u(-a + bBv)$ y $v' = \tau v(-c + dAu - eBv)$, y para que esto tenga la forma buscada basta tomar

$$\tau = \frac{1}{c}, \qquad A = \frac{c}{d}, \qquad B = \frac{a}{b},$$

con lo que queda

$$\begin{cases} u' = \lambda u(-1 + v), \\ v' = v(-1 + u - \beta v), \end{cases} \qquad \lambda = \frac{a}{c} > 0, \qquad \beta = \frac{ae}{bc} > 0. \tag{2.2}$$

Cinco constantes menos tres unidades libres dejan dos parámetros en este campo adimensional. Las trayectorias dependen además de los datos iniciales en las nuevas unidades. Obsérvese que β **no depende de** d: la intensidad con la que u ayuda a v se ha absorbido entera en la unidad con la que se mide u. Cambiar d conserva la forma del campo adimensional, pero cambia el dato adimensional $u_0 = d\bar{u}_0/c$ si se mantiene fija la población física inicial; por tanto, sí puede cambiar su destino.

Equilibrios y estabilidad. En el primer cuadrante cerrado hay exactamente dos. Si $u = 0$, la segunda ecuación da $v = 0$ o $v = -1/\beta < 0$; si $v = 0$, la primera da $u = 0$; y en el interior hay

que tener $v = 1$ y $u - \beta v = 1$, es decir, $u = 1 + \beta$. Luego los equilibrios son $(0, 0)$ y $(1 + \beta, 1)$. La matriz jacobiana es

$$DF(u, v) = \begin{pmatrix} \lambda(v-1) & \lambda u \\ v & u - 1 - 2\beta v \end{pmatrix},$$

y en los dos puntos vale

$$DF(0,0) = \begin{pmatrix} -\lambda & 0 \\ 0 & -1 \end{pmatrix}, \qquad DF(1+\beta, 1) = \begin{pmatrix} 0 & \lambda(1+\beta) \\ 1 & -\beta \end{pmatrix}.$$

El origen es diagonal con las dos entradas negativas: **exponencialmente asintóticamente estable**. Y para el otro no hace falta calcular autovalores, basta el determinante:

$$\det DF(1+\beta, 1) = 0 \cdot (-\beta) - \lambda(1+\beta) \cdot 1 = -\lambda(1+\beta) < 0,$$

y un determinante negativo en dimensión dos significa autovalores reales de signos opuestos: es una **silla**, inestable. Los dos hechos juntos ya dicen casi todo lo que va a pasar.

LA IDEA: LA SIMBIOSIS OBLIGADA NECESITA MASA CRÍTICA

El origen es aquí asintóticamente estable; en Volterra era una silla, y en la competición una fuente. La lectura dentro del modelo es esta: si las dos poblaciones son pequeñas, se encuentran poco, la ayuda mutua —que es un término de producto, uv— es de segundo orden y no compensa las muertes $-a\bar{u}$ y $-c\bar{v}$, que son de primer orden. Cada una muere antes de que la otra pueda ayudarla. Hace falta una **masa crítica** para arrancar, y quien la marca es la variedad estable de la silla: por encima de ella las dos poblaciones acaban creciendo sin cota, por debajo las dos se extinguen. No se afirma que ambas sean monótonas desde el instante inicial: primero pueden pasar por una región donde una crece y la otra decrece. Es la misma estructura de umbral que aparecía en el efecto Allee, pero con dos especies y en el plano.

Los ejes son órbitas, y de ahí la positividad. La restricción del sistema al eje $u = 0$ es la EDO escalar

$$y' = y(-1 - \beta y),$$

cuyos puntos críticos son $y = 0$ y $y = -1/\beta$; en el semieje $y \geq 0$ solo está el primero, y $f'(0) = -1 < 0$, luego es exponencialmente asintóticamente estable. Para $y_0 > 0$ es $f(y_0) < 0$ y la solución decrece hacia 0. Se puede resolver —es una Bernoulli, como la logística— y sale

$$y(t) = \frac{y_0 e^{-t}}{1 + \beta y_0 (1 - e^{-t})},$$

cuyo denominador se anula en $t^* = -\log\left(1 + \frac{1}{\beta y_0}\right) < 0$: la explosión ocurre en **tiempo negativo**, así que la solución maximal está definida en $(t^*, +\infty) \supset [0, +\infty)$ y la órbita es $\gamma(y_0) = (0, +\infty)$ para todo $y_0 > 0$ —la misma órbita para todos, como corresponde a una EDO escalar autónoma sin equilibrios intermedios—. Lo mismo, y más fácil, en el eje $v = 0$: allí $u' = -\lambda u$.

En consecuencia los dos semiejes son unión de órbitas, y como dos órbitas distintas de un sistema autónomo no se cortan, una solución que arranque con $u_0 > 0$ y $v_0 > 0$ no puede

tocarlos: $u(t) > 0$ y $v(t) > 0$ en todo su intervalo de definición. El primer cuadrante abierto es invariante, que es lo que hacía falta para que el modelo tenga sentido.

Otra vía (la positividad sin hablar de órbitas). Escribiendo la primera ecuación como $u' = \lambda(-1 + v(t))\, u$, con v ya conocida, es una EDO lineal escalar y $u(t) = u_0 \exp\big(\lambda \int_0^t (-1 + v)\big)$, que no se anula nunca. Igual con v. Es el mismo truco de la sección 2.2.2: cuando la incógnita aparece como factor en su propia ecuación, hay una exponencial escondida y el signo está decidido de antemano.

Las cuatro regiones, y por qué no hay ciclos. En el primer cuadrante abierto las curvas donde se anulan las derivadas son la recta horizontal $v = 1$ (donde $u' = 0$) y la recta $u = 1 + \beta v$ (donde $v' = 0$), que se cortan en la silla. Dividen el cuadrante en cuatro regiones, y en cada una el signo del campo es constante:

$$u' > 0 \iff v > 1, \qquad v' > 0 \iff u > 1 + \beta v.$$

El núcleo: dos de las cuatro regiones son positivamente invariantes. Sea $R_+ = \{v > 1,\ u > 1 + \beta v\}$, donde las dos crecen. Su frontera tiene dos trozos, y en los dos el campo apunta hacia dentro: sobre $\{v = 1,\ u > 1 + \beta\}$ es $u' = 0$ pero $v' = u - 1 - \beta > 0$, así que la trayectoria sube y entra; sobre $\{u = 1 + \beta v,\ v > 1\}$ es $v' = 0$ pero $u' = \lambda u(v - 1) > 0$, así que se va a la derecha y entra. Luego de R_+ no se sale. Con el mismo argumento y los signos cambiados, la región $R_- = \{v < 1,\ u < 1 + \beta v\}$, donde las dos decrecen, también es positivamente invariante.

De ahí sale la respuesta a la pregunta con la que abría la subsección: **no hay órbitas cíclicas en el primer cuadrante abierto**. Una órbita cerrada tendría que rodear al único equilibrio interior y visitar por tanto las cuatro regiones; pero en cuanto entrase en R_+ o en R_- ya no podría salir, y no podría cerrarse. Este modelo no tiene ciclos periódicos; eso no excluye cambios de crecimiento a decrecimiento durante un transitorio.

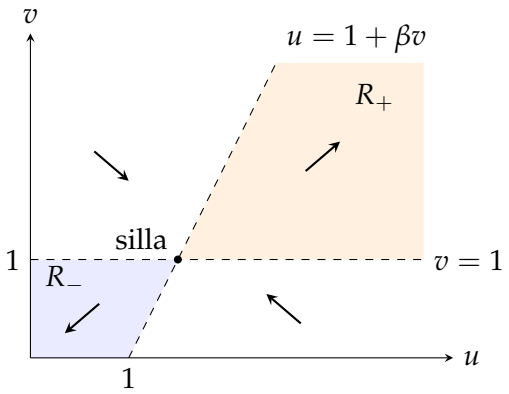


Figura 2.9: Signos del campo de simbiosis, dibujados con $\beta = 1/2$. Las flechas indican el sentido de cambio de las dos poblaciones, no órbitas calculadas. En las regiones sombreadas ambas cambian en el mismo sentido y el campo de la frontera impide salir: disminuyen en R_- y aumentan en R_+.

Otra vía (el criterio de Dulac). Con la función $B(u, v) = 1/(uv)$, positiva y C^1 en el primer

cuadrante abierto,

$$\partial_u\Big(\frac{\lambda(v-1)u}{uv}\Big)+\partial_v\Big(\frac{v(-1+u-\beta v)}{uv}\Big)=\partial_u\Big(\frac{\lambda(v-1)}{v}\Big)+\partial_v\Big(\frac{-1+u-\beta v}{u}\Big)=0-\frac{\beta}{u}<0,$$

de signo constante, luego por el criterio de Dulac no hay órbitas cerradas contenidas en el primer cuadrante abierto, que es simplemente conexo. La autolimitación hace que esta divergencia sea estrictamente negativa y permite usar Dulac. Sin ella, con $e=0$, la divergencia es cero y este criterio no decide: el argumento de las regiones todavía excluye ciclos. Tener una integral primera no implica tener órbitas periódicas.

LA IDEA: POR QUÉ LA SIMBIOSIS NO PUEDE OSCILAR, Y VOLTERRA SÍ

Merece la pena poner los tres modelos en fila, porque la diferencia es un solo signo. En Volterra la integral primera es una **suma**, $\lambda\psi(u)+\psi(v)$, con $\psi(s)=s-\log s-1$: sus niveles positivos son óvalos sin equilibrios, que las órbitas recorren periódicamente. En la competición simplificada se conserva la **diferencia** $\lambda\psi(u)-\psi(v)$, y sus ramas tienen la geometría de una silla. Al invertir el tiempo se obtiene la simbiosis simplificada con las mismas curvas.

El contraste útil es la realimentación: en Volterra, más presas favorecen al depredador, que después reduce las presas. En los modelos bilineales de competición y simbiosis estudiados aquí no aparece ese recorrido cíclico. No es una regla sobre cualquier modelo biológico: otros términos, retardos o especies pueden cambiar la conclusión.

Qué pasa entonces, y qué se pierde. El retrato es este: las órbitas que empiezan por debajo de la variedad estable de la silla caen al origen, las dos especies se extinguen; las que empiezan por encima acaban en R_+, donde ambas crecen sin cota y explotan en tiempo finito. Para justificar el último paso no se puede despreciar $-\beta v^2$: puede tener el mismo orden que uv. La comparación correcta usa $r=u/v$:

$$r'=r\big[(\lambda+\beta-r)v+1-\lambda\big].$$

Cuando v es grande, si $r<\beta+\lambda/2$, se tiene $r'/r\geq 1$ a partir de cierto valor de v, y r alcanza esa cota; en la frontera $r=\beta+\lambda/2$ también apunta hacia arriba. Así se alcanza y se conserva $r\geq\beta+\lambda/2$. Desde que además $v\geq 4/\lambda$,

$$v'=v\big[(r-\beta)v-1\big]\geq(\lambda/4)v^2,$$

y la comparación con una EDO de crecimiento cuadrático obliga a la explosión en tiempo finito.

Hay un caso exacto que la deja ver: si $\lambda=1$, la recta $u=(1+\beta)v$ es invariante. Sobre ella, $v'=v(v-1)$; con $v_0>1$, la explosión ocurre en $t_*=\log\big(v_0/(v_0-1)\big)$.

Añadir $-\lambda\alpha u^2$, $\alpha>0$, no garantiza coexistencia estable. Las rectas de equilibrio serían $v=1+\alpha u$ y $u=1+\beta v$: solo se cortan en el cuadrante cuando $\alpha\beta<1$, y allí el determinante de la jacobiana es $\lambda u_* v_*(\alpha\beta-1)<0$, otra silla. Con $\alpha\beta\geq 1$ no hay equilibrio interior. Para representar una convivencia estable hay que revisar también cómo crece la ayuda mutua, por ejemplo incorporando saturación, y volver a analizar el modelo.

2.3 Prácticas de MATLAB del tema 2

LA PREGUNTA

Algunos de estos modelos admiten una solución explícita; otros se estudian integrándolos numéricamente. Después hay que responder a preguntas concretas: en qué instante la población se dobla, cuándo baja de un umbral, cuánto vale el error frente a la solución exacta. ¿Qué hay que pedirle a un ordenador para aproximar una solución que no sabemos escribir con una fórmula, y cómo se comprueba la fiabilidad del resultado?

Esta sección resuelve los ejercicios 6 a 13 de la hoja del tema 2. Todos usan la misma orden, `ode45`, introducida en la sesión cero y recordada a continuación; las demás órdenes nuevas se presentan la primera vez que aparecen. Cada ejercicio va en su carpeta y se ejecuta con esa carpeta como carpeta actual. Los ficheros que acompañan a la hoja (`Allee.m`, `Pesca.m`, `SIS.m`, `SIR.m`, `PresaDepred.m`, `Difusion.m`, `Competicion.m`) se siguen fielmente; el que falta (`UmbralTecho.m`) se escribe aquí.

Un aviso sobre las cifras. La hoja da resultados con cinco cifras (la población del efecto Allee se dobla en $t = 0.95542$, por ejemplo). Esas cifras salen de `ode45` con sus tolerancias por defecto y de interpolar linealmente sobre la malla de tiempos que `ode45` devuelve, que es gruesa. Con tolerancias más finas, o con otro integrador, la tercera o cuarta cifra cambia: un cálculo más preciso da 0.95477. Son aproximaciones con distinta precisión, no cinco decimales igualmente garantizados. Se comprueba qué cifras se estabilizan al refinar el cálculo y, cuando es posible, se contrasta con una cuenta independiente. También se contrasta la forma de las curvas con la teoría: un método mal resuelto puede falsear incluso el comportamiento cualitativo.

La orden `ode45`

`ode45` integra numéricamente un problema de Cauchy $y' = f(t, y)$, $y(t_0) = y_0$, con un método de Runge-Kutta de orden 4 y 5 que adapta el paso de tiempo. Se llama

```
[t, y] = ode45(f, [t0, T], y0)
```

donde `f` es una función de *dos* argumentos, tiempo y estado, anónima o definida en un fichero, aunque el sistema sea autónomo y no use el primero; `[t0, T]` es el intervalo; e `y0` es el dato inicial (MATLAB admite un vector fila o columna). Devuelve una columna `t` con los instantes de salida elegidos automáticamente para esta llamada y una matriz `y` con **una fila por instante y una columna por componente**: `y(:,1)` es la primera componente en todos los tiempos, `y(end,:)` es el estado final. Para un sistema, el segundo miembro devuelve una columna, como las funciones anónimas vectoriales de la sesión cero: `f = @(t,y) [y(2); -y(1)]`. Dos detalles que se usan enseguida: si la EDO es escalar y se pasa un *vector* de datos iniciales, `ode45` resuelve todos a la vez y devuelve una columna por dato inicial, siempre que `f` esté escrita con operaciones elemento a elemento. Si se guardan solo `t` e `y`, se puede interpolar esa tabla con `interp1`; si se pide la estructura `sol = ode45(f, [t0,T], y0)`, `deval(sol, tk)` usa el interpolador del propio integrador, como se vio en la sesión cero. La tabla de salida no es toda la información que el método puede proporcionar.

```
f = @(t,y) -y;  % y' = -y
[t, y] = ode45(f, [0, 2], 1);  % y(0) = 1, solucion exp(-t)
size(t), size(y)
[t(end), y(end), exp(-2)]
```

SALIDA ESPERADA. Una salida orientativa es la siguiente. El número de filas depende de las opciones y de la versión; lo fijo es que `t` e `y` tengan el mismo número de filas.

```
ans =
    45 1
ans =
    45 1
ans =
    2.0000 0.1353 0.1353
```

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
from scipy.integrate import solve_ivp
import numpy as np
s = solve_ivp(lambda t, y: -y, [0, 2], [1.0], dense_output=True)
print(s.t.shape, s.y.shape) # s.y tiene una FILA por componente (al reves que MATLAB)
print(s.t[-1], s.y[0, -1], np.exp(-2))
print(s.sol(1.0)) # dense_output=True: solucion evaluable en cualquier t
```

En Python, `solve_ivp` hace lo mismo; dos diferencias que hay que tener en la cabeza: devuelve las componentes por *filas*, y con `dense_output=True` ofrece una función `s.sol(t)` que evalúa la aproximación en cualquier instante del intervalo integrado. Es el papel que desempeña `deval` en MATLAB; ambos usan el interpolador del integrador, que no es la interpolación lineal de una tabla.

Ejercicio 6: el efecto Allee

Ejercicio 2.5. *El efecto Allee*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 6, con el fichero `Allee.m` del curso.

Se llama efecto Allee a la caída de la densidad de una población que se reproduce sexualmente cuando el número de individuos desciende por debajo de un nivel crítico. La extensión de la ecuación logística que lo modela es

$$N' = rN(N-a)\left(1-\frac{N}{K}\right),$$

con $r>0$, $0<a<K$, K la capacidad de soporte, a el nivel crítico y r la constante del modelo. Escribir una M-función que resuelva numéricamente el modelo para $t\in[0,6]$ con $r=10^{-2}$, $a=100$, $K=500$ y datos iniciales $N(0)=50,100,150,200$; represente las soluciones en la misma figura, comprobando que si $N_0<a$ la población tiende a anularse y si $N_0>a$ tiende a K; y estime el tiempo que tarda una población de 200 individuos en doblarse.

Resolución.
La idea, y es un cambio de signo. La logística de siempre tiene el 0 como equilibrio *inestable*: por pocos individuos que queden, la población se recupera. El efecto Allee invierte precisamente eso, y lo hace con un solo factor: el $(N-a)$ que multiplica. Cerca de cero ese

factor vale $-a < 0$, así que $f \approx -raN$ y **la población decrece**. El origen pasa de repeler a atraer, y entre él y K aparece necesariamente un tercer equilibrio, $N = a$, que hace de **frontera**.

La dificultad de encontrar pareja es una posible interpretación: por debajo de a la reproducción no compensa las pérdidas. Este es un *efecto Allee fuerte*, con umbral de extinción; no todo efecto Allee exige que la tasa neta llegue a ser negativa.

Aquí los tres ceros son **simples**, y por eso f cambia de signo al cruzar cada uno: los dos extremos son estables y el del medio inestable. Saber que entre ceros consecutivos el signo es constante no basta para afirmar que alterna. Por ejemplo, en $y' = y^2$ el signo es positivo a ambos lados del equilibrio. La alternancia de este modelo se lee en sus tres factores lineales distintos.

Paso 1 (el núcleo): los tres equilibrios y la alternancia de estabilidad. Los equilibrios son $\bar{N}_1 = 0$, $\bar{N}_2 = a$ y $\bar{N}_3 = K$. Derivando el producto,

$$f'(N) = r(N-a)\Big(1 - \frac{N}{K}\Big) + rN\Big(1 - \frac{N}{K}\Big) - \frac{r}{K}N(N-a),$$

luego $f'(0) = -ra < 0$, $f'(a) = ra(1 - a/K) > 0$ y $f'(K) = -\frac{r}{K}K(K-a) = -r(K-a) < 0$: 0 y K son exponencialmente asintóticamente estables y a es inestable. Con el teorema de las EDO autónomas escalares: para $N_0 \in (0, a)$ es $f(N_0) < 0$ y la solución decrece hacia 0; para $N_0 \in (a, K)$, $f(N_0) > 0$ y crece hacia K; para $N_0 > K$, decrece hacia K. El nivel a es una frontera: por debajo, la tasa neta del modelo es negativa y la población tiende a cero. Con el dato $N_0 = a$ permanece exactamente en ese equilibrio inestable.

Órdenes nuevas. `axis([xmin, xmax, ymin, ymax])` fija los ejes antes de dibujar. `fzero(g, [a, b])` calcula una raíz de la función escalar `g` en el intervalo $[a, b]$, en cuyos extremos `g` debe cambiar de signo; `fzero(g, x0)` arranca de un punto y busca ella el intervalo. Y el truco para convertir la tabla de `ode45` en función: `N4 = @(t) interp1(times, N(:,4), t) - 400` interpola la cuarta columna en el instante `t` y le resta 400; su cero es el instante buscado.

```
function Allee
% Efecto Allee: N' = r*N*(N-a)*(1-N/K), cuatro datos iniciales a la vez.
a = 100; r = 0.01; K = 500;
N0 = [50, 100, 150, 200]; % vector de datos iniciales
Tfinal = 6;

fun = @(t,y) r*y.*(y-a).*(1-y/K); % con puntos: vale para el vector N0
[times, N] = ode45(fun, [0, Tfinal], N0); % una columna por dato inicial

close all
axis([0, Tfinal, -5, K+5])
hold on
h1 = plot(times, N(:,1), 'b', 'LineWidth', 1);
h2 = plot(times, N(:,2), 'k', 'LineWidth', 1);
h3 = plot(times, N(:,3), 'g', 'LineWidth', 1);
h4 = plot(times, N(:,4), 'r', 'LineWidth', 1);

% Tiempo en que la poblacion N(0)=200 se dobla: raiz de N(t) - 400
N4 = @(t) interp1(times, N(:,4), t) - 400;
tcrit = fzero(N4, [0, 2]);
plot([0, Tfinal], [400, 400], 'k')
```

```
h5 = plot(tcrit, 400, 'b.', 'MarkerSize', 20);
legend([h1,h2,h3,h4,h5], 'N(0)=50', 'N(0)=100', 'N(0)=150', 'N(0)=200', ...
       'Punto de corte', 'Location', 'Best')
title(['La poblacion se dobla en t = ', num2str(tcrit)], 'FontSize', 12)
hold off
shg
end
```

SALIDA ESPERADA

```
>> Allee
(figura con el titulo: La poblacion se dobla en t = 0.95542)
```

La curva de $N(0) = 100$ es constante: es el equilibrio inestable, y `ode45` lo respeta porque $f(100) = 0$ exactamente. La de 50 tiende a cero, y las de 150 y 200 a 500. El 0.95542 de la salida ilustrada incluye el error de la integración y el de `interp1`; no es el tiempo exacto redondeado a cinco decimales.

Hay una comprobación independiente que evita resolver $N(t)$. Entre 200 y 400 la población crece, así que separar variables da directamente el tiempo de paso:

$$t_* = \int_{200}^{400} \frac{dN}{0.01N(N-100)(1-N/500)} = 0.9547712524\ldots$$

La integral se evalúa por fracciones simples; la figura muestra 0.95477. El título del programa se construye pegando texto y número con corchetes y `num2str`.

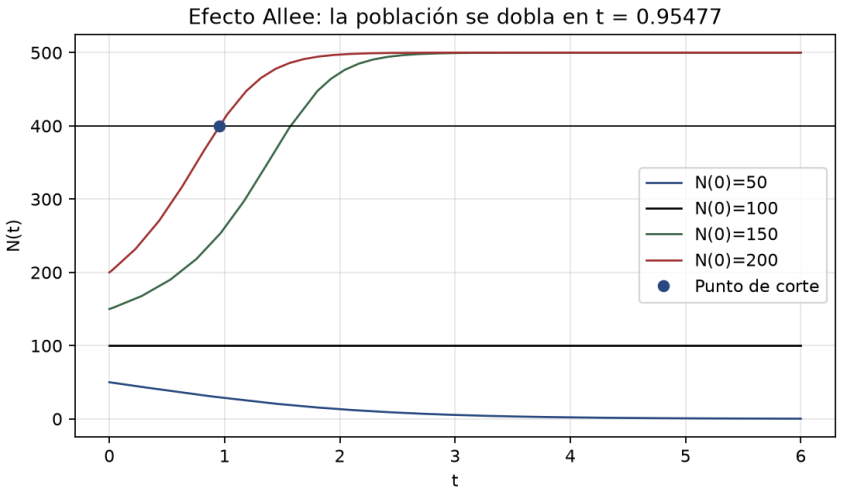


Figura 2.10: Las cuatro soluciones del modelo de Allee. La de $N(0) = 100$ es el equilibrio inestable; por debajo, extinción; por encima, la capacidad de soporte. El punto marca el instante en que $N(0) = 200$ llega a 400.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
from scipy.optimize import brentq
a, r, K, T = 100, 0.01, 500.0, 6
f = lambda t, N: r * N * (N - a) * (1 - N / K)
```

```
soluciones = {}
for N0 in [50, 100, 150, 200]:
    s = solve_ivp(f, [0, T], [N0], dense_output=True, rtol=1e-6, atol=1e-9)
    soluciones[N0] = s.sol
    plt.plot(s.t, s.y[0], label="N(0)=%d" % N0)
tcrit = brentq(lambda t: soluciones[200](t)[0] - 400, 0, 2) # equivale a fzero en [0,2]
plt.axhline(400, color="k"); plt.plot(tcrit, 400, "bo")
plt.title("La poblacion se dobla en t = %.5f" % tcrit); plt.legend(); plt.show()
```

Aquí se usa un bucle para integrar cada dato por separado. También se podría llamar a `solve_ivp(f, [0,T], [50,100,150,200])`: la función actúa componente a componente y representa cuatro EDO independientes, igual que en MATLAB. El bucle permite que cada trayectoria use su propia malla. `brentq(g,a,b)` y `fzero(g,[a,b])` buscan un cero acotado por un cambio de signo, aunque usan algoritmos distintos; `s.sol` proporciona la interpolación continua del integrador.

Dónde se falla. Casi siempre en la programación, y casi siempre en los mismos cinco sitios.

- *Al olvidar los puntos.* `ode45` trabaja internamente con el estado en columna, y aquí esa columna tiene cuatro componentes, una por dato inicial. Escrito sin punto, `y*(y-a)` es el producto matricial de dos columnas 4×1, que no está definido, y el programa se detiene con un error de dimensiones. Con un solo dato inicial el mismo código funciona, y por eso el fallo aparece justo al pasar el vector `N0`.
- *Al leer la matriz por filas.* `N(:,4)` es la trayectoria que sale de 200; `N(4,:)` es el cuarto instante de la malla para los cuatro datos a la vez. Las dos órdenes son válidas y no dan ningún error: la equivocada simplemente dibuja o interpola otra cosa.
- *Al dar a `fzero` un intervalo sin cambio de signo.* Con `[0, 2]` la función vale -200 en $t = 0$ y unos $+97$ en $t = 2$, y hay un cero dentro. Con `[0, 0.5]` vale unos -115 en el extremo derecho, porque la población aún no ha llegado a 400, y `fzero` se niega a buscar.
- *Al salirse de la malla.* `interp1` devuelve `NaN` fuera de $[0,6]$, así que un intervalo como `[0, 8]` tampoco sirve. En Python es peor, porque no avisa: `np.interp` devuelve el último valor de la tabla fuera de su rango.
- *Al buscar un instante que no existe.* La población que sale de 50 decrece y la que sale de 100 es constante: ninguna de las dos se dobla, y ningún `fzero` lo encontrará. Lo dice antes el signo de f, como en el paso 1; el programa solo debe buscar lo que la teoría asegura que existe.

QUÉ DEJA. `ode45` devuelve una tabla de valores en instantes que elige ella; para preguntar por un instante concreto se interpola, y para preguntar «cuándo vale tanto» se busca un cero con `fzero`. Y la teoría va antes: con el signo de f se sabe qué tiene que salir, y el programa solo lo confirma y le pone números.

Ejercicio 7: un modelo de pesca con esfuerzo constante

Ejercicio 2.6. *Pesca con esfuerzo constante sobre el modelo de Gompertz*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 7, con el fichero `Pesca.m` del curso. El análisis teórico es el ejercicio de pesca de Gompertz.

Para el modelo $y' = ry\log(K/y) - qy$ del ejercicio de pesca de Gompertz, escribir una M-función `[flag1, flag2] = Pesca(q, y0, T)` que, con $r = 0.2$ y $K = 27$: resuelva el modelo en $[0, T]$ y represente la solución; calcule, si existe, el instante en que el peso es la mitad del inicial, devolviendo `flag1 = true`; calcule, si existe, el primer instante en que $y(t) \leq 5$, devolviendo `flag2 = true`; y compare, para $y(0) = 10$, las soluciones con $q = 0.05$ y $q = 0.3$ con el comportamiento asintótico predicho.

Resolución.
La idea, y aquí lo que se programa es una decisión, no una cuenta. El ejercicio pide dos instantes que *pueden no existir*, y devolver dos banderas que digan si existen. Eso cambia la estructura del programa: antes de buscar un cero hay que decidir si hay algo que buscar, y esa decisión no la puede tomar el ordenador probando, porque probar en un intervalo finito no distingue «no ocurre» de «no ocurre todavía». La toma la teoría. Como la solución es monótona y tiende a $\bar{y} = Ke^{-q/r}$ sin cruzarlo (el análisis del ejercicio de pesca de Gompertz), un nivel distinto de y_0 se alcanza en tiempo finito si está *estrictamente* entre y_0 y $\bar{y}$. El nivel inicial se alcanza en $t = 0$; el equilibrio, si no se parte de él, solo en el límite. Los `if` traducen estas distinciones.
El núcleo es fijar el equilibrio, decidir si existe cada cruce e integrar la EDO. Las órdenes `figure`, `legend` y `text` solo presentan el resultado. Hay una precaución adicional: que un cruce exista no garantiza que ocurra antes de T. Las banderas indicarán existencia en todo el futuro; $[0, T]$ será la ventana de la gráfica.
Órdenes nuevas. Una función con dos argumentos de salida se escribe `function [flag1, flag2] = Pesca(q, y0, T)` y dentro se asigna a cada uno; los valores lógicos `true` y `false` son de la sesión cero. `text(x, y, 'mensaje')` escribe un texto dentro de la figura en las coordenadas (x, y) de los ejes. `pause` detiene el programa hasta que se pulsa una tecla, y aquí sirve para ver la primera figura antes de que aparezca la segunda. Los mensajes se construyen concatenando texto y `num2str`.
La lógica. La existencia de los dos instantes se decide con la teoría, no probando: t_1 existe si $y_0/2 > \bar{y}$; t_2 existe si $y_0 \leq 5$ (y vale 0) o si $\bar{y} < 5$. En los cruces que quedan dentro de la ventana se busca el cero de la solución interpolada menos el nivel. Para los demás se dispone de un control exacto: con $A = \log K - q/r$, la solución satisface

$$\log y(t) = A + (\log y_0 - A)e^{-rt}, \qquad t_\ell = -\frac{1}{r}\log\left(\frac{\log \ell - A}{\log y_0 - A}\right).$$

La fórmula de t_ℓ se usa solo para niveles estrictamente intermedios. Así, si el cruce queda fuera de $[0, T]$, el programa puede dar su instante sin extrapolar datos. Dentro de la ventana también se conserva como alternativa si la malla numérica no acota el cero. Esta es la corrección del código de partida que evita confundir «todavía no» con «nunca».

```
function [flag1, flag2] = Pesca(q, y0, T)
% Pesca(q, y0, T): modelo y' = r*y*log(K/y) - q*y en [0,T], con r=0.2, K=27.
```

```
% flag1 = true si existe t con y(t) = y0/2; flag2 = true si existe t con y(t) <= 5.
% Pruebas: Pesca(.4, 14, 20) Pesca(.3, 14, 20) Pesca(.05, 10, 20) Pesca(.3, 10, 20)
assert(q >= 0 && y0 > 0 && T > 0, 'Se requiere q >= 0, y0 > 0 y T > 0')
r = 0.2; K = 27; y02 = 0.5*y0;
A = log(K) - q/r;
tiempo = @(ell) -log((log(ell)-A)/(log(y0)-A))/r;
ybar = K*exp(-q/r); % equilibrio

% (1) Resolucion y representacion
f = @(t,y) r*y.*log(K/y) - q*y;
[times, y] = ode45(f, [0, T], y0);
close all
figure(1)
axis([0, T, 0, 1.2*max(y0, ybar)])
hold on
plot(times, y, 'LineWidth', 1.1)
plot([0, T], [ybar, ybar], 'r', 'LineWidth', 1.1)
plot(0, y0, 'r.', 'MarkerSize', 20)
title(['Evolucion de la masa de peces en [0,', num2str(T), ']; ', ...
      'Factor de pesca = ', num2str(q)], 'FontSize', 14)
legend(' Solucion', ['Equilibrio = ', num2str(ybar)], ...
      [' Dato inicial = ', num2str(y0)], 'Location', 'Best')
text(2, 2, 'Pulsa RETURN para continuar ...', 'FontSize', 14)
shg
pause

% (2) Instante en que el peso es la mitad del inicial (solo si y0/2 > ybar)
flag1 = false;
if (y02 > ybar)
  t1 = tiempo(y02); % referencia exacta, aunque t1 > T
  if t1 <= T && y(end) < y02 % cero acotado por los datos
    p1 = @(t) interp1(times, y, t) - y02;
    t1 = fzero(p1, [0, T]);
  end
  msg1 = ['La mitad del dato inicial se alcanza para t = ', num2str(t1)];
  flag1 = true;
else
  msg1 = 'No se alcanza la mitad del dato inicial';
end

% (3) Primer instante con y <= 5
flag2 = false;
if (y0 <= 5)
  msg2 = ' Se tiene y(t) <= 5 para t = 0'; t2 = 0; flag2 = true;
else
  if (ybar < 5)
    t2 = tiempo(5);
    if t2 <= T && y(end) < 5
      p2 = @(t) interp1(times, y, t) - 5;
      t2 = fzero(p2, [0, T]);
    end
    msg2 = ['Se tiene y(t) <= 5 para t = ', num2str(t2)]; flag2 = true;
  else
    msg2 = 'No existe t tal que y(t) <= 5';
  end
```

```
end

% Representacion de (2) y (3)
figure(2)
axis([0, T, -2, 1.2*max(y0, ybar)])
hold on
plot(times, y, 'LineWidth', 1.1)
plot([0, T], [ybar, ybar], 'r', 'LineWidth', 1.1)
plot([0, T], [y02, y02], 'k--', 'LineWidth', 1.1)
plot([0, T], [5, 5], 'k', 'LineWidth', 1.1)
plot(0, y0, 'r.', 'MarkerSize', 20)
if flag1 && t1 <= T, plot(t1, y02, 'k.', 'MarkerSize', 20), end
if flag2 && t2 <= T, plot(t2, min(y0,5), 'k.', 'MarkerSize', 20), end
legend('Solucion', 'Equilibrio', 'y = mitad dato inicial', 'y = 5', 'Location', 'Best')
text(1, 0, msg1, 'FontSize', 14)
text(1, -1, msg2, 'FontSize', 14)
hold off
shg
end
```

SALIDA ESPERADA

```
% Salida ilustrativa del codigo de partida; las cifras pueden variar.
>> [f1, f2] = Pesca(0.4, 14, 20)
(figura 1: equilibrio = 3.6541; figura 2 con los mensajes)
La mitad del dato inicial se alcanza para t = 3.6388
Se tiene y(t) <= 5 para t = 7.2795
f1 =
  logical
   1
f2 =
  logical
   1
```

Con $q = 0.4$ el equilibrio es $27e^{-2} \approx 3.6541$, por debajo de 7 y de 5: los dos instantes existen. Con $q = 0.3$ el equilibrio es 6.0245: se alcanza la mitad de 14 pero nunca se baja de 5, y `flag2` es falso. El apartado 4 compara, para $y_0 = 10$, $q = 0.05$ (equilibrio 21.03, la población *crece* hacia él) y $q = 0.3$ (equilibrio 6.02, decrece hacia él): en los dos casos la solución tiende al equilibrio, sin cruzarlo, como dice la teoría; ninguno de los dos instantes existe.

El control, y esta ecuación se resuelve, así que el control es exacto. El cambio $z = \log y$ convierte $y' = ry\log(K/y) - qy$ en la ecuación lineal $z' = r(\log K - z) - q$, de donde

$$z(t) = A + (z_0 - A)e^{-rt}, \qquad A = \log K - \frac{q}{r},$$

y el instante en que y vale ℓ se despeja: $t = -\frac{1}{r}\log\big((\log \ell - A)/(z_0 - A)\big)$. Con $q = 0.4$ e $y_0 = 14$ eso da $t_1 = 3.628701$ y $t_2 = 7.273519$. Obsérvese que $\bar{y} = e^A = Ke^{-q/r}$, que es el equilibrio: la fórmula recupera de paso el análisis cualitativo.

La comparación con la salida enseña algo que hay que llevarse. La salida ilustrada imprime 3.6388 y 7.2795, mientras que los valores de referencia son 3.6287 y 7.2735. Buscar con mucha precisión el cero de una aproximación no elimina el error de esa aproximación. Aquí se interpola linealmente sobre la malla de `ode45`: unir con segmentos una curva convexa puede desplazar el cruce. Para mejorar el resultado se puede guardar `sol = ode45(...)` y evaluar con `deval(sol,t)`, además de reducir `RelTol` y `AbsTol`. Las cifras fiables se comprueban al

refinar y, en este caso, comparando con la fórmula exacta.
La regla general: el error del instante reúne los de integración, interpolación y búsqueda del cero. Además, si la curva cruza el nivel con pendiente muy pequeña, un pequeño error vertical puede producir un error grande en el tiempo.
Dos bordes que conviene probar. Con `Pesca(.4,14,1)`, los dos instantes existen, pero quedan fuera de la figura: la versión corregida los obtiene con la fórmula exacta y no llama a `fzero` fuera de sus datos. En cambio, si un nivel coincide con $\bar{y}$ y se parte de otro valor, no hay cruce finito. Cerca de esa igualdad, el redondeo puede alterar una comparación estricta; conviene contrastar los parámetros y el equilibrio antes de interpretar un tiempo muy grande. **Estar en el recorrido futuro del modelo no es lo mismo que estar en el recorrido de la ventana de integración.**

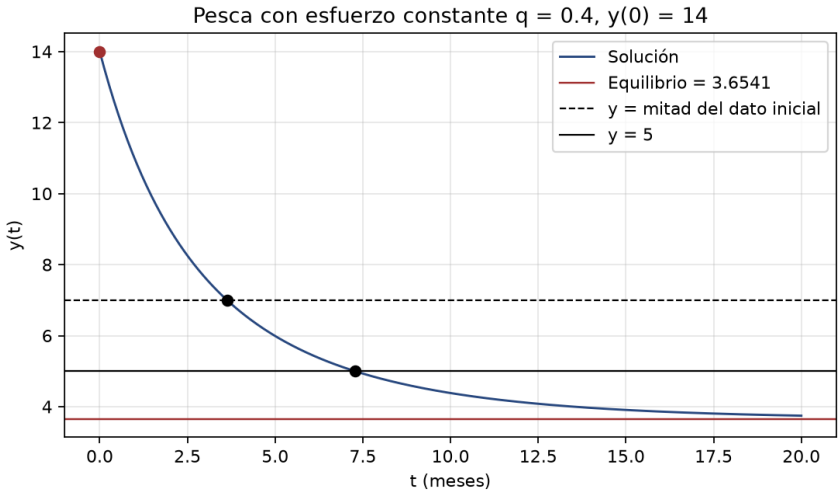


Figura 2.11: Pesca con $q = 0.4$ desde $y_0 = 14$: la masa de peces baja hacia el equilibrio 3.65 y cruza la mitad del dato inicial y el nivel 5 en los instantes marcados.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_ivp
from scipy.optimize import brentq
def pesca(q, y0, T):
    if q < 0 or y0 <= 0 or T <= 0:
        raise ValueError("Se requiere q >= 0, y0 > 0 y T > 0")
    r, K = 0.2, 27.0
    A = np.log(K) - q/r
    tiempo = lambda ell: -np.log((np.log(ell)-A)/(np.log(y0)-A))/r
    ybar = K * np.exp(-q / r)
    s = solve_ivp(lambda t, y: r * y * np.log(K / y) - q * y, [0, T], [y0], dense_output=
        True)
    if not s.success:
        raise RuntimeError(s.message)
    flag1 = flag2 = False
    if 0.5 * y0 > ybar:
        t1 = tiempo(0.5*y0)
        if t1 <= T and s.y[0, -1] < 0.5*y0:
```

```
            t1 = brentq(lambda t: s.sol(t)[0] - 0.5*y0, 0, T)
        flag1 = True
        print("mitad del dato inicial en t =", t1)
    if y0 <= 5:
        flag2 = True; print("y <= 5 en t = 0")
    elif ybar < 5:
        t2 = tiempo(5)
        if t2 <= T and s.y[0, -1] < 5:
            t2 = brentq(lambda t: s.sol(t)[0] - 5, 0, T)
        flag2 = True
        print("y <= 5 en t =", t2)
    return flag1, flag2
print(pesca(0.4, 14, 20)) # (True, True); tiempos aproximados
print(pesca(0.4, 14, 1)) # (True, True); cruces posteriores a T
```

Ejercicio 8: umbral y techo

Ejercicio 2.7. *Un modelo con umbral y techo*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 8. No hay fichero de partida; se escribe aquí.

Sea $N(t)$ el número de individuos de una población en el instante t, en años, con

$$N' = r(N - N_1)(N_2 - N), \qquad r > 0, \quad 0 \le N_1 < N_2,$$

donde N_1 es el umbral y N_2 el techo. Escribir una M-función `UmbralTecho(r, N1, N2, N0, T)` que resuelva el modelo en $[0, T]$, represente la solución y calcule, si existe, el instante en que la población se dobla. Obtener la solución con $r = 5 \cdot 10^{-8}$, $N_1 = 10^3$, $N_2 = 10^6$ y $N(0) = 2 \cdot 10^3$, y decir qué se puede afirmar de la población.

Resolución.
La idea. Este modelo tiene dos equilibrios, uno de ellos posiblemente cero, y la recta de datos iniciales se parte en tres regiones con destinos distintos: extinción por debajo del umbral, crecimiento hacia el techo entre los dos, y descenso hacia el techo por encima. **Lo que el programa tiene que hacer, antes que integrar, es situar el dato inicial en una de las tres**, y la condición que se escribe en el `if` —`(N0 > N1) && (2*N0 < N2)`— es la condición para doblarse en algún instante finito, cuando $N_0 > 0$. El núcleo del programa son esa condición y la llamada a `ode45`; el `fzero` y las siete líneas de `plot` vienen después.
Y hay una decisión de escala que se ve en la figura y merece atención: el umbral vale 10^3 y el techo 10^6, mil veces mayor. En un eje lineal el umbral es indistinguible del cero, de modo que la frontera entre extinguirse y llegar al millón queda pegada al eje y desaparece del dibujo. La figura de abajo es honesta —muestra lo que el programa dibuja— pero no muestra el umbral: para verlo hay que dibujar aparte la ventana $[0, 2 \cdot 10^3]$, o usar escala logarítmica, que es lo que se hizo en la epidemia de la sesión cero por la misma razón.
Análisis teórico. Los equilibrios son N_1 y N_2. Como $f'(N) = r(N_2 - N) - r(N - N_1)$, es $f'(N_1) = r(N_2 - N_1) > 0$, inestable, y $f'(N_2) = -r(N_2 - N_1) < 0$, exponencialmente asintóticamente estable. Si $N_0 \in (0, N_1)$, la solución decrece y alcanza cero en tiempo finito: mientras $0 \le N \le N_0$, se cumple $N' \le -r(N_1 - N_0)(N_2 - N_0) < 0$. En ese primer

cero, el instante de extinción, el modelo deja de valer. Su prolongación matemática toma valores negativos y explota hacia $-\infty$ también en tiempo finito; no existe hasta $t \to \infty$. Si $N_0 \in (N_1, N_2)$, $f(N_0) > 0$ y crece hacia N_2. Si $N_0 > N_2$, decrece hacia N_2. Con los datos, $N_0 = 2000$ está entre umbral y techo: la población crece hacia el millón, y como $2N_0 = 4000 < N_2$, se dobla.

Una discrepancia de la hoja. El enunciado dice $N(0) = 2 \cdot 10^3$, pero la figura de la solución oficial está hecha con $N(0) = 2 \cdot 10^5$ (el rótulo dice «Dato inicial $y(0) = 200000$» y «Doble del dato inicial $= 400000$»), y da un tiempo de doblado de 19.6761. Con $N(0) = 2000$ el doblado tarda 22.03 años y en $T = 200$ la población todavía no ha llegado al techo ($\approx 9.6 \cdot 10^5$). Se dan los dos.

El código admite $N_0 > 0$. Si se parte por debajo del umbral, limita la integración al instante de extinción, calculado separando variables:

$$t_{\text{ext}} = \frac{1}{r(N_2 - N_1)} \log \frac{N_1(N_2 - N_0)}{N_2(N_1 - N_0)} \qquad (0 < N_0 < N_1).$$

El tiempo de doblado se obtiene de la misma integral; si cae dentro de la ventana y los datos acotan el cruce, se calcula también numéricamente. Así se mantiene la distinción entre existencia del instante y visibilidad en la gráfica.

```
function UmbralTecho(r, N1, N2, N0, T)
% UmbralTecho(r, N1, N2, N0, T): N' = r*(N-N1)*(N2-N) en [0,T].
% N1 umbral (inestable), N2 techo (estable). Calcula cuando se dobla la poblacion.
% Pruebas: UmbralTecho(5e-8, 1e3, 1e6, 2e3, 200) UmbralTecho(5e-8, 1e3, 1e6, 2e5, 200)
assert(r > 0 && 0 <= N1 && N1 < N2 && N0 > 0 && T > 0)
a = r*(N2-N1);
tfin = T;
if N0 < N1
  textincion = log(N1*(N2-N0)/(N2*(N1-N0)))/a;
  tfin = min(T, textincion); % hasta el primer cero
end
f = @(t,N) r*(N-N1).*(N2-N);
[times, N] = ode45(f, [0, tfin], N0);

close all
hold on
h1 = plot(times, N, 'b', 'LineWidth', 1.2);
h2 = plot([0, T], [N1, N1], 'g--');
h3 = plot([0, T], [N2, N2], 'r--');
h4 = plot([0, T], [2*N0, 2*N0], 'k');
h5 = plot(0, N0, 'r.', 'MarkerSize', 20);

% La poblacion se dobla si 2*N0 queda entre N0 y el limite de la solucion
if (N0 > N1) && (2*N0 < N2)
  td = log((2*N0-N1)*(N2-N0)/((N2-2*N0)*(N0-N1)))/a;
  if td <= T && N(end) > 2*N0
    Nfun = @(t) interp1(times, N, t) - 2*N0;
    td = fzero(Nfun, [0, T]);
  end
  if td <= T, plot(td, 2*N0, 'k.', 'MarkerSize', 20), end
  title(['La poblacion se dobla en el tiempo t = ', num2str(td)])
else
  title('La poblacion no se dobla')
```

```
end
legend([h1,h2,h3,h4,h5], 'Solucion', ['Umbral : y = ', num2str(N1)], ...
      ['Techo : y = ', num2str(N2)], ['Doble del dato inicial = ', num2str(2*N0)], ...
      ['Dato inicial y(0) = ', num2str(N0)], 'Location', 'Best')
xlabel('t (anos)'), ylabel('N(t)'), grid on
hold off
shg
end
```

SALIDA ESPERADA

```
% Valores de referencia para contrastar los titulos numericos:
N0 = 200000: tiempo de doblado = 19.6865
N0 = 2000: tiempo de doblado = 22.0344
```

Estos valores proceden de la fórmula exacta, no de una garantía de precisión de la malla por defecto. Se pueden comprobar a mano porque la ecuación se integra: separando variables y descomponiendo en fracciones simples, $t = \frac{1}{r(N_2-N_1)} \log \frac{(N-N_1)(N_2-N_0)}{(N_2-N)(N_0-N_1)}$, que con $N = 2N_0$ da 19.686460 y 22.034401. La hoja muestra 19.6761 para la primera; la diferencia ilustra por qué hay que controlar el error numérico antes de fiarse de todas las cifras impresas. La operación `&&` es el «y» lógico entre escalares, que además no evalúa el segundo término si el primero es falso; `&` actúa elemento a elemento y evalúa ambos lados.

Un segundo control, y no cuesta integrar nada. En $T = 200$ la población vale 956 282 con $N_0 = 2 \cdot 10^3$ y 999 816 con $N_0 = 2 \cdot 10^5$: las dos por debajo del techo y subiendo, como debe ser, porque N_2 es un equilibrio y ninguna solución lo cruza. Comprobar que el resultado **no supera el techo** detecta ciertos fallos de implementación o de precisión: si el programa devolviera $1.1 \cdot 10^6$, habría que revisar la ecuación y la integración numérica. Pasar esta prueba, por sí solo, no demuestra que todo esté bien. Y la velocidad de la parte final también se puede predecir: cerca del techo la linealización da $N_2 - N \sim Ce^{-r(N_2-N_1)t} = Ce^{-0.04995\,t}$, con tiempo característico $1/0.04995 = 20.0$ años. Con $N_0 = 2 \cdot 10^5$ el hueco al techo pasa de 800 000 a 184 en doscientos años; con $N_0 = 2 \cdot 10^3$ solo baja a 43 718, y la diferencia no es un fallo de la estimación sino su límite: la exponencial vale cerca del equilibrio, y la solución que arranca de dos mil tarda casi un siglo en llegar a esa zona.

Dónde se rompe el modelo. Por debajo del umbral, la prolongación matemática pasa a valores negativos. Con $N_0 = 500$, la extinción ocurre en $t \approx 13.87$ años. Más adelante hay una singularidad en tiempo finito: no existe una solución matemática en $[0, 300]$ que el integrador pueda aproximar. Un valor negativo enorme o una integración interrumpida no son una predicción de población; indican que se ha sobrepasado el dominio de validez.

Por eso el código se detiene en el primer cero. En una ecuación cuya solución no se conoce, la misma condición puede implementarse como un *evento* del integrador. Es la misma precaución que con las cotizaciones negativas: **el dominio de validez del modelo lo pone quien modeliza, y hay que programarlo si se quiere que el ordenador lo respete**.

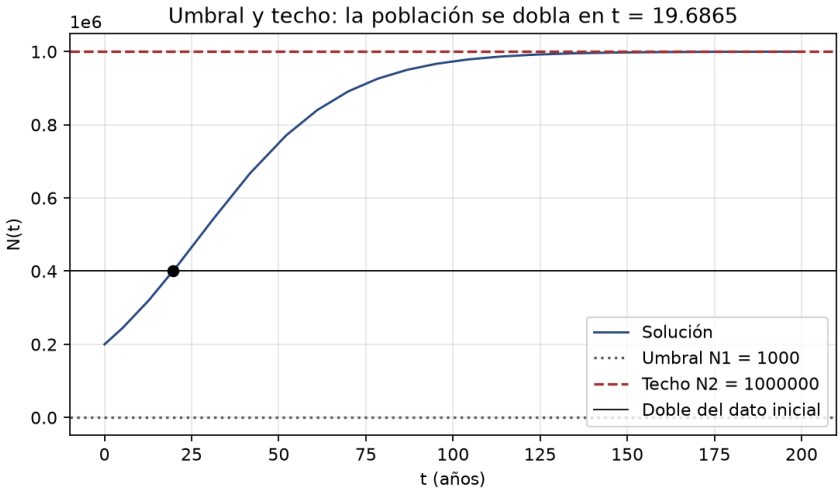


Figura 2.12: Umbral y techo con $N(0) = 2 \cdot 10^5$: la población crece con forma de S hacia el techo de un millón y se dobla a los 19.7 años. El umbral de 1000 es invisible a esta escala, y es una frontera: por debajo de él la población se extingue en tiempo finito.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_ivp
from scipy.optimize import brentq
def umbral_techo(r, N1, N2, N0, T):
    if not (r > 0 and 0 <= N1 < N2 and N0 > 0 and T > 0):
        raise ValueError("Parametros fuera del dominio del modelo")
    a = r*(N2-N1)
    tfin = T
    if N0 < N1:
        textincion = np.log(N1*(N2-N0)/(N2*(N1-N0)))/a
        tfin = min(T, textincion)
    s = solve_ivp(lambda t, N: r*(N-N1)*(N2-N), [0, tfin], [N0],
                  dense_output=True, rtol=1e-8, atol=1e-8)
    if not s.success:
        raise RuntimeError(s.message)
    if N0 > N1 and 2*N0 < N2:
        td = np.log((2*N0-N1)*(N2-N0)/((N2-2*N0)*(N0-N1)))/a
        if td <= T and s.y[0, -1] > 2*N0:
            td = brentq(lambda t: s.sol(t)[0] - 2*N0, 0, T)
        print("se dobla en t =", td)
    else:
        print("la poblacion no se dobla")
    return s
umbral_techo(5e-8, 1e3, 1e6, 2e5, 200) # 19.686
umbral_techo(5e-8, 1e3, 1e6, 2e3, 200) # 22.034
```

Ejercicio 9: el modelo SIS

Ejercicio 2.8. *El modelo SIS*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 9, con el fichero `SIS.m` del curso. El análisis teórico es el ejercicio sobre el SIS y su estabilidad.

Escribir una M-función `SIS(beta, gamma, N, I0, T)` que resuelva numéricamente el modelo SIS en $[0, T]$ y represente (I, S), y obtenerla con $\beta = 0.001$, $\gamma = 0.08$, $N = 200$, $I_0 = 30$ y $T = 90$.

Resolución.
La idea, y dice qué hay que mirar en la figura. El modelo SIS no tiene clase de recuperados: el que se cura vuelve a ser susceptible. Por eso $S + I$ se conserva y todo el sistema se reduce a *una* ecuación escalar para I, que es la que el ejercicio sobre el SIS y su estabilidad analiza. Programarlo con dos componentes es cómodo, pero conviene saber que la segunda es redundante: dibujar S es dibujar $N - I$ boca abajo, y que las dos curvas sumen N, salvo el error de redondeo, es una **primera comprobación** del programa.
La segunda comprobación es el destino, y se conoce de antemano: I tiende a $N - \gamma/\beta$ si $R_0 > 1$ e $I_0 > 0$, y a cero si $R_0 \leq 1$. Si $I_0 = 0$, no aparece infección en este modelo. **Un programa de este tipo no se da por bueno porque dibuje algo bonito, sino porque llega adonde la teoría dice que tiene que llegar.** Ese es el hábito que estas prácticas pretenden instalar.
Paso 1: las órdenes nuevas. `plot(times, y)` con `y` de varias columnas dibuja una curva por columna, con colores distintos y en el orden de las columnas, que es el orden que después sigue `legend`. En un título, `'\beta'` escribe la letra griega β: los rótulos de MATLAB entienden un poco de LaTeX.
Paso 2: qué tiene que salir, calculado antes de ejecutar. $R_0 = 200 \cdot 0.001/0.08 = 2.5 > 1$: la epidemia es endémica y los infectados tienden a $\bar{I}_2 = N - \gamma/\beta = 200 - 80 = 120$, con $S \to 80$. Como $I_0 = 30 < 120$, I crece. Con $\beta = 0.0001$ (segunda prueba de la hoja), $R_0 = 0.25$ e $I \to 0$.

```
function SIS(beta, gamma, N, I0, T)
% SIS(beta, gamma, N, I0, T): modelo SIS, I' = beta*S*I - gamma*I, S' = -I', S + I = N.
% Pruebas: SIS(0.001, 0.08, 200, 30, 90) SIS(0.0001, 0.08, 200, 30, 90)
S0 = N - I0;
y0 = [I0; S0]; % estado en columna: (I, S)
f = @(t,y) ((beta*y(2) - gamma).*y(1)).*[1; -1]; % I' y S' = -I'
[times, y] = ode45(f, [0, T], y0);

close all
figure(1)
axis([0, T, 0, 1.1*N])
hold on
plot(times, y, 'LineWidth', 1.3) % dos curvas: I y S
plot([0, T], [N, N], 'k--', 'LineWidth', 1.3)
legend('Infectados', 'Susceptibles', ['Poblacion total= ', num2str(N, '%5.0f')], ...
       'Location', 'Best')
R0 = N*beta/gamma;
xlabel(' Tiempo '), ylabel(' Numero de individuos ')
title(['Evolucion del modelo SIS. \beta=', num2str(beta), ', \gamma= ', num2str(gamma),
```

```
    ...
      ', R_0= ', num2str(R0)])
hold off
shg
end
```

SALIDA ESPERADA

```
>> SIS(0.001, 0.08, 200, 30, 90)
(figura: I sube de 30 hacia 120 y S baja de 170 hacia 80; titulo con R_0 = 2.5)
```

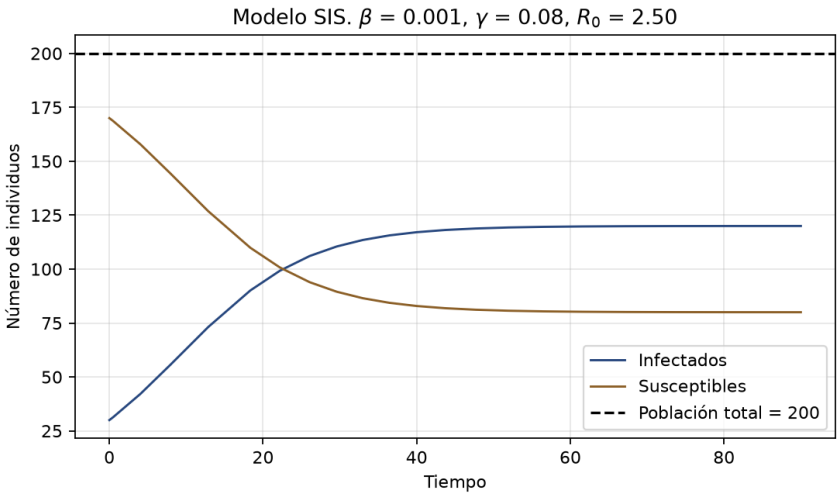


Figura 2.13: Modelo SIS con $R_0 = 2.5$: los infectados crecen con forma de S hacia el equilibrio endémico 120 y los susceptibles bajan hacia $\gamma/\beta = 80$. La suma se mantiene en 200.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
def sis(beta, gamma, N, I0, T):
    f = lambda t, y: [(beta * y[1] - gamma) * y[0], -(beta * y[1] - gamma) * y[0]]
    s = solve_ivp(f, [0, T], [I0, N - I0], rtol=1e-6)
    plt.plot(s.t, s.y[0], label="Infectados"); plt.plot(s.t, s.y[1], label="Susceptibles"
        )
    plt.axhline(N, ls="--", color="k", label="Poblacion total = %d" % N)
    plt.title(r"SIS. $\beta$=%g, $\gamma$=%g, $R_0$=%.2f" % (beta, gamma, N * beta /
        gamma))
    plt.legend(); plt.show()
sis(0.001, 0.08, 200, 30, 90)
```

Ejercicio 10: el modelo SIR

Ejercicio 2.9. *El modelo SIR*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 10, con el fichero `SIR.m` del curso. La teoría es la sección 2.2.1.

Escribir una M-función `SIR(beta, gamma, I0, R0, N, T)` que resuelva numéricamente el modelo SIR en $[0, T]$ y represente las trayectorias de I, R y S y la órbita (I, R). Obtenerla con $\beta = 0.001$, $\gamma = 0.007$, $I_0 = 10$, $R_0 = 0$, $N = 130$ y $T = 365$.

Resolución.
La idea, y es la que hace interesante este ejercicio. Las dos ejecuciones que se se comparan se diferencian en un solo parámetro, γ, y cambian tanto la altura del pico como la duración. Bajar γ aumenta el tiempo medio de recuperación del modelo: I sigue creciendo hasta que S desciende a un umbral γ/β menor, por lo que el pico resulta más alto. En este ejemplo también queda una cola de infecciosos más larga; cuando ya casi no hay susceptibles, domina $I' \approx -\gamma I$.
De ahí sale lo que hay que mirar en las dos figuras, y es lo que este ejercicio enseña sobre integrar numéricamente: **hay que elegir una ventana que permita ver el fenómeno que se quiere estudiar**. La EDO es la misma sea cual sea T, pero el gráfico puede ocultar una cola larga. Con $\gamma = 0.07$, en 365 días se ve prácticamente toda la evolución; con $\gamma = 0.007$, el dibujo termina cuando aún queda un número apreciable de infecciosos.
Paso 1: qué tiene que salir, calculado antes de ejecutar. $\gamma/\beta = 7$, y $S_0 = 120 \gg 7$: la epidemia prende con fuerza, I sube hasta un máximo y baja a cero, S cae casi a 0 (quedan $\bar{S} \approx 0$ susceptibles: casi todos se contagian) y $R \to N - \bar{S}$. Con $\gamma = 0.07$ (segunda prueba de la hoja), $\gamma/\beta = 70$, el pico es mucho menor y quedan unos 28 susceptibles sin contagiar: es el efecto que explica la sección 2.2.1. Con $\gamma = 0.07$ la infección es numéricamente despreciable en $t = 365$ ($I \approx 5 \cdot 10^{-5}$); con $\gamma = 0.007$, **no**: quedan todavía unos 11.7 infecciosos, porque el tiempo medio de recuperación es $1/\gamma \approx 143$ días y la cola tarda años en irse. Se trata de poblaciones continuas: un valor positivo muy pequeño no representa literalmente una fracción de persona. En la solución matemática, $I(t)$ no alcanza cero en tiempo finito. Como control adicional, la integral primera del SIR da los picos sin integrar en el tiempo:

$$I_{\text{máx}} = N - R_0 - k + k\log(k/S_0), \qquad k = \gamma/\beta, \quad S_0 > k.$$

Con los datos propuestos valen 103.11 y 22.27, respectivamente.

```
function SIR(beta, gamma, I0, R0, N, T)
% SIR(beta, gamma, I0, R0, N, T): modelo SIR, estado (I, R, S) con S + R + I = N.
% Pruebas: SIR(0.001, 0.07, 10, 0, 130, 365) SIR(0.001, 0.007, 10, 0, 130, 365)
S0 = N - I0 - R0;
y0 = [I0; R0; S0];
f = @(t,y) [ (beta*y(3) - gamma).*y(1); ... % I'
             gamma*y(1); ... % R'
            -(beta*y(1)).*y(3) ]; % S'
[times, y] = ode45(f, [0, T], y0);

close all
figure(1) % trayectorias I, R, S
axis([0, T, 0, 1.1*N])
```

```
hold on
plot(times, y, 'LineWidth', 1.3)
plot([0, T], [N, N], 'k--', 'LineWidth', 1.3)
legend('Infectados', 'Recuperados', 'Susceptibles', ...
      ['Poblacion total= ', num2str(N, '%5.0f')], 'Location', 'NorthEast')
xlabel(' Tiempo '), ylabel(' Numero de individuos ')
title(['Evolucion del modelo SIR. \beta=', num2str(beta), ', \gamma= ', num2str(gamma)])
hold off
shg

figure(2) % orbita (I, R) y region invariante
hold on
plot([0, N], [N, 0], 'k', 'LineWidth', 1.3) % recta I + R = N
plot(y(:,1), y(:,2), 'r', 'LineWidth', 1.3)
title('Region invariante'), xlabel('Infectados'), ylabel('Recuperados')
hold off
shg
end
```

SALIDA ESPERADA

```
>> SIR(0.001, 0.007, 10, 0, 130, 365)
(figura 1: I alcanza unos 103 hacia t = 45; al final R vale unos 118 y S casi 0)
(figura 2: sale de (10, 0) y termina cerca de (11.7, 118.3); tiende al eje I = 0)
```

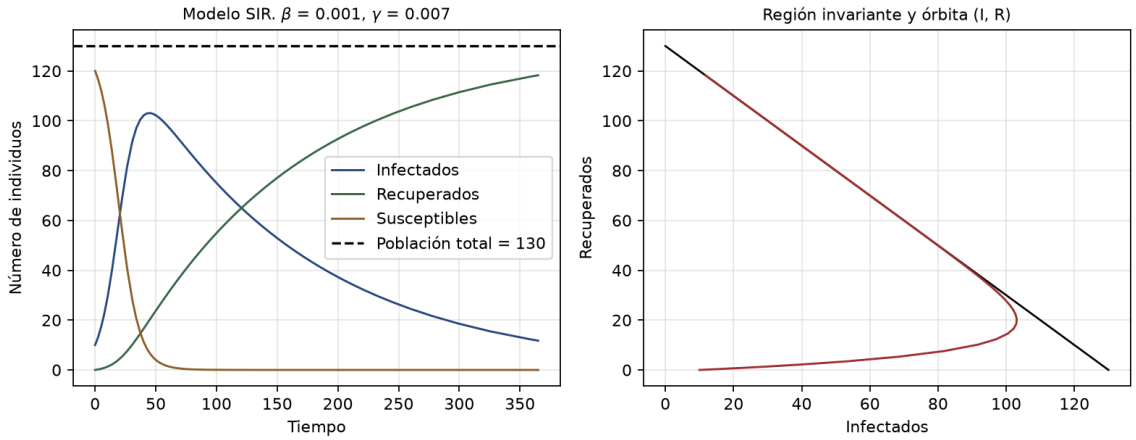


Figura 2.14: Modelo SIR con $\beta = 0.001$ y $\gamma = 0.007$. Izquierda: la onda epidémica; los susceptibles quedan reducidos a un valor muy pequeño. Derecha: la órbita en el plano (I, R) dentro del triángulo invariante. Su destino es el eje $I = 0$, pero en los 365 días dibujados se representa solo hasta unos 11.7 infecciosos, parte de la cola larga que produce un γ pequeño.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
def sir(beta, gamma, I0, R0, N, T):
    f = lambda t, y: [(beta * y[2] - gamma) * y[0], gamma * y[0], -beta * y[0] * y[2]]
    s = solve_ivp(f, [0, T], [I0, R0, N - I0 - R0], rtol=1e-6)
    fig, ax = plt.subplots(1, 2, figsize=(10, 4))
    for k, nombre in enumerate(["Infectados", "Recuperados", "Susceptibles"]):
```

```
        ax[0].plot(s.t, s.y[k], label=nombre)
    ax[0].axhline(N, ls="--", color="k"); ax[0].legend()
    ax[1].plot([0, N], [N, 0], "k"); ax[1].plot(s.y[0], s.y[1], "r")
    ax[1].set_xlabel("Infectados"); ax[1].set_ylabel("Recuperados"); plt.show()
sir(0.001, 0.007, 10, 0, 130, 365)
```

Ejercicio 11: presa-depredador y el insecticida

Ejercicio 2.10. *Presa-depredador con control biológico*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 11, con el fichero `PresaDepred.m` del curso.

Sea $N(t)$ la densidad de una especie de insectos y $P(t)$ la de su depredador, usado como agente de control biológico, con

$$P' = -P + 2PN, \qquad N' = 5N - 3PN.$$

Escribir `PresaDepred(P0, N0, T)` que resuelva el modelo, represente las trayectorias y la órbita con los equilibrios, y obtener la solución para $P_0 = 2$, $N_0 = 0.5$ y para $P_0 = 2$, $N_0 = 0.1$, que corresponde a haber aplicado antes un insecticida que redujo los insectos a la quinta parte. Comparar y estudiar las implicaciones a largo plazo de la decisión de aplicar el insecticida.

Resolución.
La idea, y este ejercicio no se hace para ver una figura sino para contestar una pregunta de decisión. Lo que se compara no son dos condiciones iniciales cualesquiera: son *la misma situación con y sin intervención humana*. Aplicar el insecticida es mover el dato inicial de $(2, 0.5)$ a $(2, 0.1)$, y la pregunta es qué le hace eso al sistema a medio plazo. La respuesta la da la estructura, no el dibujo: las órbitas son cerradas y rodean al equilibrio, pero para saber si la nueva órbita es mayor hay que comparar la integral primera, no solo la distancia al equilibrio. Aquí reducir N desde $1/2$ aumenta H, pues $\partial H/\partial N = 2 - 1/N < 0$ para $N < 1/2$. Eso coloca el dato en una curva de nivel exterior, cuyo máximo de insectos es mayor. El ordenador pone cifras a esta comparación.
De hecho las cifras se pueden obtener *sin integrar*, y conviene saberlo porque es la diferencia entre usar el modelo y ejecutarlo. La integral primera $H(P, N) = 3P - 5\log P + 2N - \log N$ es constante sobre cada órbita; los extremos de N están donde $N' = 0$ con $N \neq 0$, es decir en $P = c/d = 5/3$, y los de P donde $N = a/b = 1/2$. Sustituyendo y resolviendo la ecuación escalar resultante salen los cuatro números de cada órbita.
El núcleo del programa son dos líneas —el segundo miembro y `ode45`— y el resto son dos figuras. Y hay una línea que parece decorativa y no lo es: `plot(c/d, a/b, 'k.')`, que dibuja el equilibrio calculado a mano encima de la órbita. Una órbita completa con datos positivos, distintos del equilibrio, debe rodearlo. Si no lo hace, conviene comprobar el código, la precisión y si se ha integrado durante una vuelta completa.
Análisis teórico. Con $f_1(P, N) = -P + 2PN$ y $f_2(P, N) = 5N - 3PN$, los equilibrios son $(0, 0)$ y $(5/3, 1/2)$. La jacobiana es $\begin{pmatrix} 2N - 1 & 2P \\ -3N & 5 - 3P \end{pmatrix}$; en $(0, 0)$ vale $\mathrm{diag}(-1, 5)$, inestable (silla);

en $(5/3, 1/2)$ vale $\begin{pmatrix} 0 & 10/3 \\ -3/2 & 0 \end{pmatrix}$, con autovalores $\pm\sqrt{5}\,i$, y la linealización no decide. La sección 2.2.4 resuelve la duda: el equilibrio es estable, no atractivo, y las órbitas son cerradas.

El insecticida. Los extremos de la órbita se calculan sin integrar, con la integral primera $H(P,N) = 3P - 5\log P + 2N - \log N$ de la sección 2.2.4: N alcanza su máximo y su mínimo cuando $P = c/d = 5/3$, y P los suyos cuando $N = a/b = 1/2$. Con $P_0 = 2$, $N_0 = 0.5$ se parte de un punto de la recta $N = 1/2$ que pasa por el equilibrio, desplazado solo en P; por eso $P_0 = 2$ es exactamente uno de los dos extremos de P sobre la órbita. La órbita es pequeña: los insectos oscilan entre 0.318 y 0.741 —sí bajan de $1/2$, porque el equilibrio queda dentro—, y los depredadores entre 1.373 y 2. Con $N_0 = 0.1$ se ha empujado el sistema lejos del equilibrio, y como la órbita es cerrada y *rodea* al equilibrio, la población de insectos vuelve a subir y alcanza un máximo de 1.497, el doble del máximo sin insecticida —y tres veces la población de partida—. El insecticida hace que la plaga alcance un pico mayor más adelante, *dentro de este modelo y para estos datos*. La perturbación concreta aumenta H; otras perturbaciones pueden reducirlo o incluso dejarlo igual. Por tanto, no se puede concluir que toda intervención aumente la órbita.

Los máximos, comprobados uno a uno. El período de oscilaciones pequeñas, $2\pi/\sqrt{ac} \approx 2.81$, es una aproximación cerca del equilibrio, no el período exacto de cualquier órbita. Para los dos datos de este ejercicio se obtienen aproximadamente 2.835 y 3.067: en $T = 10$ caben tres vueltas completas. Con $N_0 = 0.1$, el primer máximo de insectos, 1.4973, ocurre en $t \approx 1.947$, y el primer máximo de depredadores, 2.8742, en $t \approx 2.428$; se repiten cada período.

Con $N_0 = 0.5$, el máximo de P es 2 y ya se alcanza en $t = 0$: el dato satisface $N_0 = 1/2$, y $N'(0) = (5 - 3P_0)N_0 < 0$, de modo que $P''(0) = 2P_0N'(0) < 0$. También se vuelve a alcanzar en las vueltas siguientes. Un máximo en el borde de una ventana puede ser un extremo de la órbita; lo que lo justifica es este análisis, no su posición en la tabla. Del mismo modo, el máximo de una tabla de valores puede quedar algo por debajo del máximo entre sus nodos.

El control. La integral primera. Con $(2, 0.5)$ vale $H_0 = 4.2274113$ y con $(2, 0.1)$, $H_0 = 5.0368492$; una integración de control en $[0, 10]$ con `solve_ivp`, `rtol=1e-12` y `atol=1e-13` conserva H con una deriva menor que 10^{-11}. Además, los cuatro extremos deben satisfacer $H(P,N) = H_0$ con $P = 5/3$ o $N = 1/2$, según la componente. Esta cuenta permite comprobarlos sin depender de la malla temporal.

Con tolerancias más holgadas puede aparecer una deriva visible hacia curvas interiores o exteriores. Su sentido depende del método, de los pasos y de la órbita: no hay que interpretar un estrechamiento o ensanchamiento numérico como un efecto biológico. Antes hay que comprobar que H se conserva y repetir con mayor precisión. El error acumulado de discretización, además del redondeo, puede simular una amortiguación que la ecuación no tiene.

Dónde se rompe el modelo, que es lo que más se le puede reprochar. Volterra no tiene ningún término de autolimitación, así que la conclusión «el insecticida duplica el pico» es válida dentro de un modelo en el que los insectos, en ausencia de depredador, crecerían exponencialmente para siempre. Al añadir $-\kappa N^2$, con $\kappa > 0$, en la ecuación de la presa, la coexistencia requiere $5 - \kappa/2 > 0$: entonces el equilibrio es $((5 - \kappa/2)/3, 1/2)$ y es atractivo, como se estudió en la sección 2.2.4. Puede ser foco o nodo, según κ. La comparación de los picos transitorios

exige estudiar ese nuevo sistema; no se deduce que el insecticida siga empeorándolos ni que los duplique. Esta distinción separa lo que demuestra el modelo de lo que habría que contrastar en el campo.

```
function PresaDepred(u0, v0, T)
% PresaDepred(u0, v0, T): u = P depredador, v = N presa.
% u' = (-a + b*v)*u, v' = (c - d*u)*v con a=1, b=2, c=5, d=3.
% Pruebas: PresaDepred(2, 0.5, 10) PresaDepred(2, 0.1, 10)
a = 1; b = 2; c = 5; d = 3;
f = @(t,y) [(-a + b*y(2)).*y(1); (c - d*y(1)).*y(2)];
[tout, yout] = ode45(f, [0, T], [u0; v0]);

close all
figure(1) % trayectorias
hold on
plot(tout, yout(:,1), tout, yout(:,2), 'LineWidth', 1.5)
plot(tout(1), yout(1,1), '*', tout(1), yout(1,2), '*')
legend('P depredador', 'N presa', 'P inicial', 'N inicial')
xlabel('t'), title(['Trayectorias; P_0 = ', num2str(u0), '; N_0 = ', num2str(v0)])

figure(2) % orbita y equilibrios
hold on
plot(yout(:,1), yout(:,2), 'LineWidth', 1.5)
plot(u0, v0, '*', 'MarkerSize', 10)
plot(0, 0, 'k.', c/d, a/b, 'k.', 'MarkerSize', 25)
legend('Orbita', 'Dato inicial', 'Equilibrios')
xlabel('P: depredadores'), ylabel('N: presas'), title('Plano de fases')
shg
end
```

SALIDA ESPERADA

```
>> PresaDepred(2, 0.5, 10)
(orbita cerrada pequena alrededor de (1.667, 0.5); N oscila entre 0.318 y 0.741)
>> PresaDepred(2, 0.1, 10)
(orbita cerrada grande; N alcanza 1.497 y P alcanza 2.874)
```

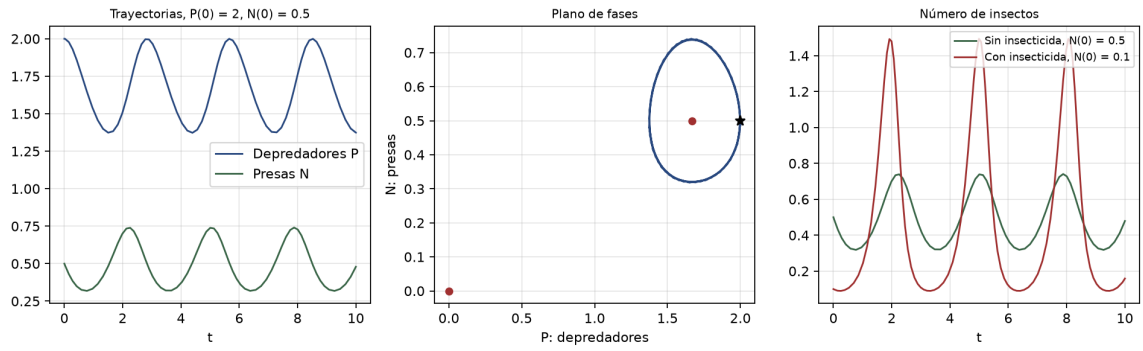


Figura 2.15: Izquierda y centro: trayectorias y órbita para $P_0 = 2$, $N_0 = 0.5$. Derecha: el número de insectos con y sin insecticida. Reducir los insectos a la quinta parte aproximadamente duplica el pico posterior en este modelo y para los datos indicados.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
from scipy.optimize import brentq
a, b, c, d = 1.0, 2.0, 5.0, 3.0
H = lambda P, N: 3*P - 5*np.log(P) + 2*N - np.log(N)
f = lambda t, y: [(-a + b * y[1]) * y[0], (c - d * y[0]) * y[1]]
for P0, N0 in [(2.0, 0.5), (2.0, 0.1)]:
    s = solve_ivp(f, [0, 10], [P0, N0], rtol=1e-8)
    plt.plot(s.t, s.y[1], label="N(0) = %g" % N0)
    H0 = H(P0, N0)
    # En estos dos casos H(5/3, .5) < H0 < H(5/3, 5).
    maxN = brentq(lambda N: H(5/3, N) - H0, .5, 5)
    print("max de la orbita con N(0)=%g: %.3f" % (N0, maxN))
plt.legend(); plt.title("Insectos con y sin insecticida"); plt.show()
```

Ejercicio 12: difusión a través de una membrana

Ejercicio 2.11. *Difusión y el error de la aproximación numérica*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 12, con el fichero `Difusion.m` del curso.

Dos recipientes Ω_1, Ω_2 de volúmenes $V_1, V_2 > 0$ están separados por una membrana que una sustancia puede atravesar. Si $y_1(t)$ e $y_2(t)$ son las cantidades de sustancia en cada uno, con datos iniciales no negativos,

$$y_1' = \frac{k}{V_2} y_2 - \frac{k}{V_1} y_1, \qquad y_2' = \frac{k}{V_1} y_1 - \frac{k}{V_2} y_2, \qquad y_1(0) = y_1^0,\ y_2(0) = y_2^0,$$

con $k > 0$ un coeficiente de transporte que depende de la permeabilidad. Escribir `Difusion(k, V1, V2, y01, y02, T)` que resuelva el sistema, dibuje las cantidades y las concentraciones, y calcule el error medio $E = \frac{1}{m}\sum_{i=1}^{m} \|y(t_i) - z^i\|_2$ entre la solución exacta y y la aproximación z^i en los instantes calculados. Aplicarlo a $k = 0.1$, $V_1 = 1$, $V_2 = 2$, $y_1^0 = 1$, $y_2^0 = 0$ en $[0, 40]$.

PISTAS, DE MENOR A MAYOR

1. *Suma las dos ecuaciones: $y_1 + y_2$ se conserva. Sustituye $y_2 = N - y_1$ y queda una EDO lineal en y_1 del tipo del modelo de mezclas.*
2. *Las concentraciones son y_i / V_i; la teoría dice que se igualan.*
3. *El error pide evaluar la solución exacta en los instantes `t` de `ode45`, que es una función vectorial, y la norma fila a fila.*

Resolución.

La idea, y reduce el sistema a una sola ecuación. La sustancia no se crea ni se destruye: solo cambia de recipiente. Esa frase, escrita en álgebra, es $(y_1 + y_2)' = 0$, y convierte un sistema de dos ecuaciones en **una** —porque conocida y_1 ya se conoce $y_2 = N - y_1$— y esa una es lineal. Todo lo que hay que saber del modelo sale de ahí.

La imagen que predice el resultado, y conviene tenerla antes de resolver: el flujo por la membrana va de donde hay más *concentración* a donde hay menos, no de donde hay más cantidad. Por tanto el estado final no es «la mitad en cada recipiente» sino **la misma**

concentración en los dos, y eso reparte la sustancia en proporción a los volúmenes. Con $V_1 = 1$ y $V_2 = 2$, el recipiente grande se queda con el doble, y la concentración común es $N/(V_1 + V_2) = 1/3$. La solución numérica debe aproximarse a ese límite.

La hoja llama a este proceso «ósmosis», pero la ecuación modela un intercambio difusivo de sustancia a volúmenes fijos. En la ósmosis se estudia específicamente el paso de disolvente a través de una membrana selectiva. Aquí k multiplica una diferencia de concentraciones para dar una cantidad por unidad de tiempo; por tanto, tiene unidades de volumen por tiempo.

Paso 1 (el núcleo): la cantidad conservada, que reduce el sistema a una ecuación. Sumando, $(y_1 + y_2)' = 0$, luego $y_1 + y_2 \equiv N := y_1^0 + y_2^0$. Con $y_2 = N - y_1$,

$$y_1' = a - by_1, \qquad a = \frac{kN}{V_2}, \qquad b = \frac{k}{V_1} + \frac{k}{V_2} = \frac{k(V_1 + V_2)}{V_1 V_2},$$

que es el modelo de mezclas: $y_1(t) = \frac{a}{b} + (y_1^0 - \frac{a}{b})e^{-bt}$ e $y_2(t) = N - y_1(t)$, definidas matemáticamente en todo $\mathbb{R}$. Con datos no negativos, las cantidades permanecen no negativas para $t \geq 0$. Las concentraciones límite son

$$c_1 = \lim_{t\to\infty} \frac{y_1(t)}{V_1} = \frac{a}{bV_1} = \frac{N}{V_1 + V_2}, \qquad c_2 = \lim_{t\to\infty} \frac{y_2(t)}{V_2} = \frac{N}{V_1 + V_2}:$$

las concentraciones se igualan, exponencialmente, a la concentración media. Con los datos, $a/b = 1/3$ y las dos concentraciones tienden a $1/3$.

Órdenes nuevas. `subplot(1,2,k)` divide la ventana en una fila y dos columnas y activa la k-ésima; después `plot` dibuja en ella. `set(gcf, 'Position', [x, y, ancho, alto])` fija el tamaño de la ventana; `gcf` es la figura actual y `gca` los ejes actuales. `sum(A, 2)` suma por filas (el 2 es la dimensión), frente a `sum(A)`, que suma por columnas; `sqrt` sobre un vector actúa componente a componente. Con eso, $\sum_i \|y(t_i) - z^i\|_2$ es `sum(sqrt(sum(Err1.*Err1, 2)))`, donde `Err1` es la matriz de diferencias, una fila por instante.

```
function Difusion(k, V1, V2, y10, y20, T)
% Difusion(k, V1, V2, y10, y20, T): difusion entre dos recipientes.
% Prueba: Difusion(0.1, 1, 2, 1, 0, 40)
N = y10 + y20; kv1 = k/V1; kv2 = k/V2;
a = kv2*N; b = kv1 + kv2; c = a/b;
F = @(t,y) [kv2*y(2) - kv1*y(1); kv1*y(1) - kv2*y(2)];
[t, y] = ode45(F, [0, T], [y10; y20]);

close all
figure
set(gcf, 'Position', [100, 100, 1100, 500])
subplot(1,2,1) % cantidades
axis([0, T, 0, max(1,1.1*N)]), hold on
plot(t, y(:,1), t, y(:,2), 'LineWidth', 1.3)
legend('Cantidad sustancia conjunto 1', 'Cantidad sustancia conjunto 2', 'Location', '
    Best')
title('Evolucion modelo de difusion: cantidades'), hold off
subplot(1,2,2) % concentraciones
cmax = max([y10/V1, y20/V2]);
axis([0, T, 0, max(1,1.1*cmax)]), hold on
plot(t, y(:,1)/V1, t, y(:,2)/V2, 'LineWidth', 1.3)
legend('Concentracion conjunto 1', 'Concentracion conjunto 2', 'Location', 'Best')
title('Evolucion modelo de difusion: concentraciones'), hold off
```

```
23
24 % Error medio frente a la solucion exacta, en los instantes de ode45
25 z = @(t) [c + (y10-c)*exp(-b*t), N - c - (y10-c)*exp(-b*t)]; % una fila por t
26 Err1 = z(t) - y; % matriz length(t) x 2
27 Err = sum(sqrt(sum(Err1.*Err1, 2)))/length(t); % media de las normas de las filas
28 subplot(1,2,1)
29 text(T/4, 0.8*N, ['Error = ', num2str(Err, '%15.10f')], 'FontSize', 15)
30 fprintf('El error es: %15.10f \n', Err)
31 shg
32 end
```

SALIDA ESPERADA

```
% Una salida ilustrativa; depende de la version y de la malla.
>> Difusion(0.1, 1, 2, 1, 0, 40)
El error es: 0.0000039964
```

En la salida ilustrada, el error medio absoluto es del orden de 10^{-6}. No es la misma magnitud que `RelTol`: esta interviene, junto con `AbsTol`, en el control de la estimación del error local, y no promete directamente una cota del error global medio.
La cuenta compara cada aproximación con la solución exacta en *el mismo instante*. Además, esta media da igual peso a todos los nodos, aunque estén espaciados de forma desigual: no es una media temporal. Para comparar dos métodos de manera homogénea conviene evaluar ambos en una malla común. Conocer la solución exacta permite comprobar también el error máximo y cómo disminuye al refinar las tolerancias.

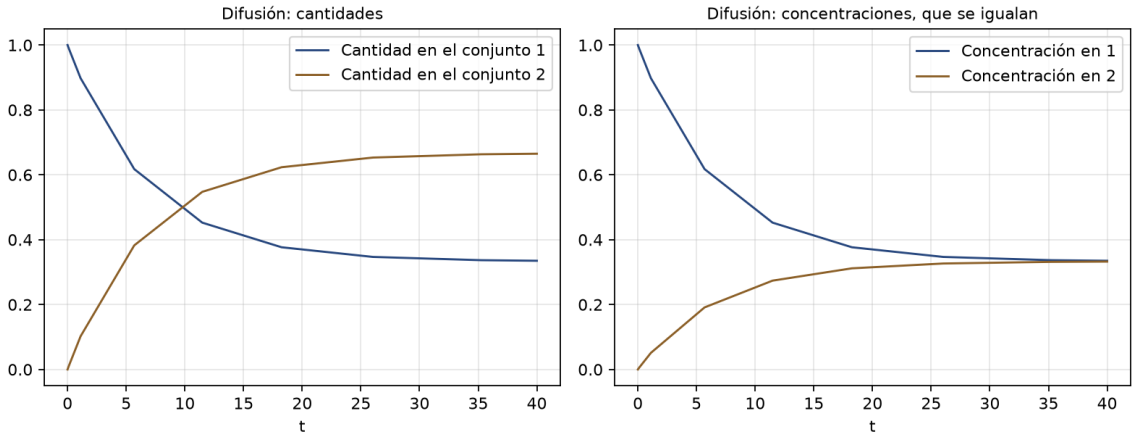


Figura 2.16: Difusión con $k = 0.1$, $V_1 = 1$, $V_2 = 2$. Las cantidades se reparten 1/3 y 2/3 (proporcionales a los volúmenes) y las concentraciones tienden a 1/3. Se unen los nodos con segmentos: el error E se mide en los nodos, no en todo el trazo entre ellos.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_ivp
def difusion(k, V1, V2, y10, y20, T):
    N = y10 + y20; kv1, kv2 = k / V1, k / V2; a = kv2 * N; b = kv1 + kv2; c = a / b
    F = lambda t, y: [kv2 * y[1] - kv1 * y[0], kv1 * y[0] - kv2 * y[1]]
    s = solve_ivp(F, [0, T], [y10, y20], rtol=1e-3, atol=1e-6) # tolerancias de ode45
    exacta = np.c_[c + (y10 - c) * np.exp(-b * s.t), N - c - (y10 - c) * np.exp(-b * s.t)
```

```
        ]
    err = np.mean(np.linalg.norm(exacta - s.y.T, axis=1)) # media de las normas por filas
    print("error medio = %.10f con %d instantes" % (err, s.t.size))
difusion(0.1, 1, 2, 1, 0, 40)
```

Ejercicio 13: competición

Ejercicio 2.12. *Especies en competición*

ORIGEN: hoja de ejercicios del tema 2, ejercicio 13, con los ficheros `Competicion.m` y `CompeticionBis.m` del curso. La teoría es la sección 2.2.5.

Dos especies con densidades $N(t)$ y $P(t)$ siguen

$$N' = aN - bNP, \qquad P' = cP - dNP, \qquad a, b, c, d > 0.$$

Escribir `Competicion(N0, P0)` que resuelva el modelo con $a = 0.1$, $b = 5$, $c = 0.2$, $d = 1$, $T = 30$, represente la solución y el diagrama de fases con los equilibrios; obtener la solución para $N_0 = 0.1$, $P_0 = 0.01$; y confirmar con varias ejecuciones que la configuración de las órbitas es la de la teoría.

Resolución.
La idea, y la decisión de método es integrar hacia atrás. El equilibrio interior es una silla. Los puntos de su variedad estable se acercan a él hacia delante, pero un pequeño error transversal termina alejando una trayectoria numérica. Para trazar esa variedad puede partirse muy cerca de la silla, en la dirección de su autovector estable, e integrar hacia atrás: el movimiento recorre la rama alejándose de la silla y contrae inicialmente el error transversal. Los papeles estable e inestable se intercambian al invertir el tiempo.
El código de abajo integra órbitas desde varios datos, en los dos sentidos. Eso muestra la geometría de las cuencas, aunque no calcula por sí solo la separatriz. Hay que recordar otra limitación: algunas soluciones explotan hacia atrás en tiempo finito, de modo que no siempre se puede llegar a $-T$.
El núcleo del programa son entonces *tres* líneas: el segundo miembro, la integración hacia delante y la integración hacia atrás. Y una cuarta que es de otra clase, `eq2 = [c/d; a/b]`, porque marca en la figura el punto respecto del cual hay que leerlo todo.
Que el ejercicio pida «confirmar con varias ejecuciones» no es un adorno: una órbita no distingue una silla de un nudo, y solo cuatro o cinco datos iniciales repartidos a los dos lados enseñan la partición del plano en cuencas. Es el equivalente numérico de dibujar un campo de direcciones.
Análisis teórico. Con $f_1 = aN - bNP$, $f_2 = cP - dNP$, la jacobiana es $\begin{pmatrix} a - bP & -bN \\ -dP & c - dN \end{pmatrix}$. En $(0,0)$ vale $\operatorname{diag}(a,c)$, los dos positivos: fuente, inestable. En $(c/d, a/b)$ vale $\begin{pmatrix} 0 & -bc/d \\ -da/b & 0 \end{pmatrix}$, con autovalores $\pm\sqrt{ac}$: silla, inestable. Es lo que se demostró en la sección 2.2.5: gana una u otra según el lado de la separatriz.
Órdenes nuevas. `ode45(f, [0, -T], y0)` integra *hacia atrás*, de 0 a $-T$; con ello se obtiene el

tramo de semiórbita negativa hasta donde exista la solución. La salida que no interese se descarta con la virgulilla: `[~, y2] = ode45(...)`.

Qué sale con $(0.1, 0.01)$. Como $bP_0 = 0.05 < a = 0.1$, N crece al principio; y como $dN_0 = 0.1 < c = 0.2$, P también. Se está en el cuadrante donde las dos crecen. Después, la separatriz decide: en $t = 30$, $N = 1.3 \cdot 10^{-5}$ y $P = 0.423$: gana P. Conviene subrayar que ese 0.423 **no es un máximo**: P sigue creciendo, y de hecho crece sin cota, porque una vez que N se ha desplomado la segunda ecuación es casi $P' = cP$. Es el valor en el borde de la ventana, y aquí el borde es lo que hay que mirar precisamente porque el interés está en el destino, no en un pico. El máximo verdadero de la figura es el de N, que sube hasta 0.1245 en $t = 8.22$ antes de caer: la especie perdedora crece durante ocho unidades de tiempo antes de extinguirse, y quien mirase solo los primeros instantes concluiría lo contrario de lo que pasa. Hacia atrás, las dos vienen del origen. El programa `CompeticionBis` dibuja seis órbitas a la vez y muestra las dos cuencas separadas por la silla.

El control, y es exacto. Este sistema también tiene integral primera, y de la clase que la caja de intuición de la sección 2.2.5 llamaba «diferencia»: dividiendo las dos ecuaciones y separando variables,

$$H(N, P) = a \log P - bP - c \log N + dN$$

es constante sobre cada órbita —derivando, $H' = (a - bP)P'/P - (c - dN)N'/N = (a - bP)(c - dN) - (c - dN)(a - bP) = 0$—. Con los datos del ejercicio y $(N_0, P_0) = (0.1,\ 0.01)$ vale exactamente

$$H_0 = 0.1 \log(0.01) - 5(0.01) - 0.2 \log(0.1) + 0.1 = 0.05,$$

porque $0.1 \log(0.01) = 0.2 \log(0.1)$. Una integración de control con `solve_ivp`, `rtol=1e-12` y `atol=1e-15`, mantiene la deriva de H por debajo de $6 \cdot 10^{-12}$ en $[0, 30]$. Comprobarla es especialmente útil cuando las poblaciones recorren varios órdenes de magnitud: un integrador puede controlar bien la especie abundante y perder la precisión relativa de la que casi ha desaparecido.

Sirve además para localizar la separatriz sin integrar, porque la separatriz está contenida en la curva de nivel que pasa por la silla, $H(0.2,\ 0.02) = 0.030685$. Conviene, eso sí, no pedirle a H más de lo que da: **el valor de H_0 por sí solo no decide quién gana**. La silla es también punto de silla de H —su hessiana ahí es $\operatorname{diag}(c/N^2,\ -a/P^2)$, con signos opuestos— y por eso la curva de nivel crítica es una cruz que parte el entorno en cuatro sectores, dos con $H > H_s$ y dos con $H < H_s$. Los datos $(0.1,\ 0.01)$ y $(0.3,\ 0.02)$ tienen casi el mismo H_0 —0.0500 y 0.0496— y ganan especies distintas. Lo que H da es la curva; el lado lo da la posición sobre ella.

Cerca de la separatriz. Si el error numérico es comparable a la distancia entre el dato y la frontera, la especie que parece ganar puede depender de la tolerancia. No hay una distancia universal, como 10^{-6}, por debajo de la cual el cálculo sea imposible: hay que comparar la precisión con la separación relevante y comprobar la estabilidad al refinar.

Aquí puede hacerse una comprobación geométrica independiente. Para $N_0 = 0.1$, la rama estable situada por debajo de la silla se encuentra resolviendo $H(0.1, P_s) = H_s$ con $0 < P_s < 0.02$: sale $P_s \approx 0.00714806$. El dato $P_0 = 0.01$ queda por encima y gana P. No basta usar $|H - H_s|$ como si fuera una distancia: cerca de la silla $\nabla H = 0$, y el nivel crítico contiene también la variedad inestable. Hay que identificar la rama que separa las cuencas.

La limitación biológica es que ninguna especie tiene techo. Añadir autolimitación puede cambiar quién gana e incluso permitir coexistencia estable, como muestra el modelo general de competición; no se limita a frenar el crecimiento de la ganadora.
Dos ajustes numéricos que responden a estos problemas. Para datos positivos usamos $z_1 = \log N$, $z_2 = \log P$:

$$z_1' = a - be^{z_2}, \qquad z_2' = c - de^{z_1}.$$

Así se trabaja con logaritmos cuando una población es extremadamente pequeña, y al dibujar se recupera $y = \exp(z)$ componente a componente. No es un modelo distinto.
Al integrar hacia atrás, un *evento* detiene el cálculo al salir del rectángulo que se representará. Esto evita intentar atravesar una explosión: para $(0.3, 0.02)$, por ejemplo, el extremo pasado está cerca de -10.78, antes de -30. La opción `Events` recibe una función con tres salidas: `value` se anula en el borde, `isterminal=1` pide parar y `direction=0` acepta ambos sentidos de cruce. La función auxiliar va al final del mismo fichero. El corte es una decisión de visualización, no una extinción ni un equilibrio.

```
function Competicion(N0, P0)
% Competicion(N0, P0): N' = aN - bNP, P' = cP - dNP, con a=.1, b=5, c=.2, d=1, T=30.
% Pruebas: Competicion(0.1, 0.01) Competicion(0.1, 0.002) Competicion(0.3, 0.02)
close all
a = 0.1; b = 5; c = 0.2; d = 1; T = 30;
assert(N0 > 0 && P0 > 0, 'Se requieren poblaciones iniciales positivas')
g = @(t,z) [a - b*exp(z(2)); c - d*exp(z(1))];
eq1 = [0; 0]; eq2 = [c/d; a/b];
disp(['Equilibrio no trivial: (', num2str(eq2(1)), ',', num2str(eq2(2)), ')'])

z0 = log([N0; P0]);
Nmax = max(N0, eq2(1)); Pmax = max(P0, eq2(2));
topes = [2*Nmax; 2*Pmax];
opt = odeset('RelTol', 1e-9, 'AbsTol', 1e-11);
[t1, z1] = ode45(g, [0, T], z0, opt);
optAtras = odeset(opt, 'Events', @(t,z) salirCaja(t,z,topes));
[t2, z2] = ode45(g, [0, -T], z0, optAtras);
y1 = exp(z1); y2 = exp(z2);
ymax = max([y1(:); y2(:)]);
H1 = a*z1(:,2)-b*y1(:,2)-c*z1(:,1)+d*y1(:,1);
fprintf('Deriva maxima de H: %g\n', max(abs(H1-H1(1))))

figure(1) % trayectorias en [-T, T]
axis([-T, T, -0.1, 1.1*ymax]), hold on
plot(t1, y1(:,1), 'b', t1, y1(:,2), 'r', 'LineWidth', 1.5)
plot([-T, T], [0, 0], 'k--')
plot([-T, T], [eq2(1), eq2(1)], 'b--', [-T, T], [eq2(2), eq2(2)], 'r--')
plot(t2, y2(:,1), 'b', t2, y2(:,2), 'r', 'LineWidth', 1.5)
legend('N(t)', 'P(t)', 'Equilibrio nulo', 'Equilibrio de N', 'Equilibrio de P', 'Location
    ', 'Best')
title('Modelo de especies en competicion'), xlabel('Tiempo'), ylabel('Poblaciones')
hold off

figure(2) % plano de fases
axis([0, 2*Nmax, 0, 2*Pmax]), hold on
plot(y1(:,1), y1(:,2), 'b', 'LineWidth', 1.5)
plot(y2(:,1), y2(:,2), 'g', 'LineWidth', 1.5)
```

```
plot(N0, P0, 'r.', 'MarkerSize', 15)
plot(eq1(1), eq1(2), 'k.', eq2(1), eq2(2), 'k.', 'MarkerSize', 30)
legend('Orbita para t > 0', 'Orbita para t < 0', 'Dato inicial', 'Equilibrios', 'Location
    ', 'Best')
title('Especies en competicion. Plano de fases'), xlabel('Poblacion 1 (N)'), ylabel('
    Poblacion 2 (P)')
hold off
end

function [value, isterminal, direction] = salirCaja(t,z,topes)
value = max(exp(z)./topes) - 1;
isterminal = 1;
direction = 0;
end
```

SALIDA ESPERADA

```
>> Competicion(0.1, 0.01)
Equilibrio no trivial: (0.2,0.02)
(tambien se imprime la deriva de H; debe disminuir al refinar)
(figura 1: N sube y luego cae a 0; P sube hasta 0.42 en t = 30)
(figura 2: la orbita sale del origen, pasa cerca de la silla y se va por el eje P)
```

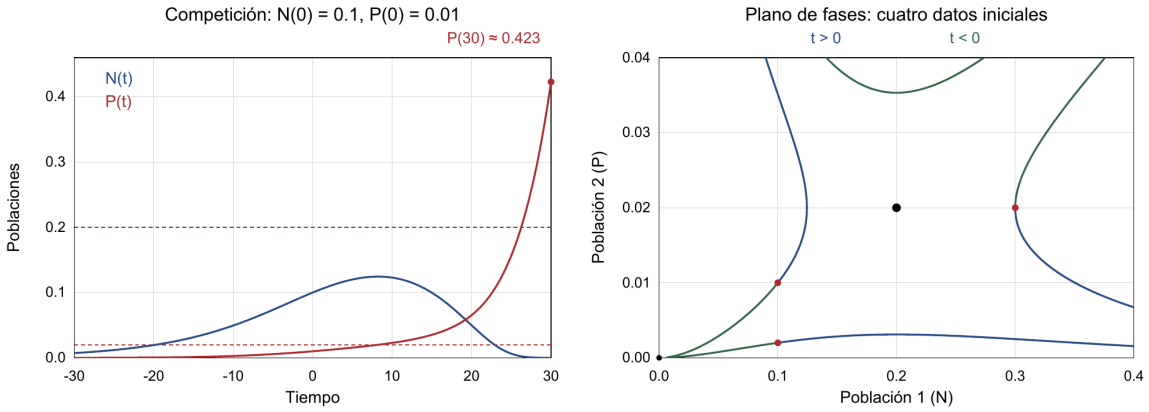


Figura 2.17: Competición con $a = 0.1$, $b = 5$, $c = 0.2$, $d = 1$. Izquierda: las trayectorias desde $(0.1, 0.01)$ en $[-30, 30]$, con $P(30)$ visible. Derecha: los cuatro datos de la versión Python, hacia delante en azul y hacia atrás en verde, mostrando sus tramos dentro del recuadro. La variedad estable de la silla en $(0.2, 0.02)$ separa las dos cuencas: en cada una una población tiende a cero y la otra crece sin cota.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
a, b, c, d, T = 0.1, 5.0, 0.2, 1.0, 30.0
g = lambda t, z: [a-b*np.exp(z[1]), c-d*np.exp(z[0])]
for N0, P0 in [(0.1, 0.002), (0.1, 0.01), (0.1, 0.05), (0.3, 0.02)]:
    z0 = np.log([N0, P0])
    topes = 2*np.maximum([N0, P0], [c/d, a/b])
    def salir_caja(t, z):
        return np.max(np.exp(z)/topes) - 1
```

```
    salir_caja.terminal = True
    salir_caja.direction = 0
    adelante = solve_ivp(g, [0, T], z0, rtol=1e-9, atol=1e-11)
    atras = solve_ivp(g, [0, -T], z0, rtol=1e-9, atol=1e-11, events=salir_caja)
    if not (adelante.success and atras.success):
        raise RuntimeError("Integracion incompleta; revisar el mensaje del solver")
    y1, y2 = np.exp(adelante.y), np.exp(atras.y)
    H1 = a*adelante.y[1]-b*y1[1]-c*adelante.y[0]+d*y1[0]
    print("deriva de H:", np.max(np.abs(H1-H1[0])))
    plt.plot(y1[0], y1[1], "b"); plt.plot(y2[0], y2[1], "g")
    plt.plot(N0, P0, "r.")
plt.plot([0, c/d], [0, a/b], "k.", ms=10)
plt.xlim(0, 0.4); plt.ylim(0, 0.04); plt.show()
```

QUÉ DEJA. Integrar hacia atrás permite explorar de dónde viene una órbita, dentro de su intervalo de existencia. En este modelo, la variedad estable de la silla separa las dos cuencas. La integral primera y el control del error ayudan a interpretar lo que el dibujo muestra cerca de esa frontera.

REPASO INTERCALADO: CAPÍTULO 2, EN TRES RONDAS

A libro cerrado, en sesiones distintas. Un tercio del capítulo 2 y dos tercios del capítulo 1 y de las asignaturas en que se apoya.

Primera ronda, al terminar el capítulo.

1. Enuncia el teorema de las EDO autónomas escalares y úsalo para describir todas las soluciones de $y' = y(1-y)(y-3)$ según y_0.
2. (Capítulo 1) Escribe la matriz de Leslie con $a = (0, 2, 1)$, $b = (0.5, 0.4)$. ¿Es irreducible? ¿Primitiva? Compara «$\hat{\lambda} < 1$» con «$\lambda < 0$» en el modelo exponencial.
3. (Capítulo 1) ¿Qué es γ en el algoritmo de PageRank y por qué evita construir la matriz de Google?
4. ¿Qué es R_0 en el modelo SIS y qué dos cosas distintas ocurren según sea menor o mayor que 1?
5. (EDO) Enuncia el teorema de estabilidad por linealización y di en qué equilibrios de este capítulo no se puede aplicar.
6. ¿Qué orden de MATLAB integra un problema de Cauchy, qué devuelve, y cómo se evalúa la solución en un instante que no está en la tabla?

Segunda ronda, una semana después.

1. Deduce el modelo logístico a partir del desarrollo de Taylor de f y explica el signo de $f''(0)$. Resuélvelo como ecuación de Bernoulli.
2. (Capítulo 1) Demuestra que $\varphi(\lambda) = \sum a_i b_1 \cdots b_{i-1} \lambda^{-i} = 1$ tiene una única raíz positiva. ¿Qué papel juega en el modelo exponencial discreto?
3. (EDO) Enuncia el criterio de comparación escalar y úsalo para probar que en la competición simplificada $u(t) < u_0 e^t$ para $t > 0$, con datos iniciales positivos.
4. Escribe la cantidad conservada del modelo de Volterra y demuestra que se conserva. ¿Por qué eso no basta para la periodicidad, y qué teorema la cierra?
5. (Capítulo 1) En el ejercicio de las cotizaciones, ¿por qué el error fuera de la muestra es la medida honesta de un ajuste? Relaciónalo con lo que dice el modelo logístico sobre extrapolar el exponencial.

6. Con pesca constante h sobre la logística, ¿cuál es el máximo h sostenible y por qué pescar exactamente ese valor es peligroso?

Tercera ronda, antes de empezar el capítulo 4.

1. Añade al modelo de Volterra adimensionalizado una autolimitación de las presas, $v' = \lambda v(1 - u - \kappa v)$, con $0 < \kappa < 1$. Calcula el equilibrio interior, linealiza y demuestra que es asintóticamente estable. ¿Por qué hay que adaptar $V = \lambda\psi(u) + \psi(v)$ al nuevo equilibrio para usarlo como función de Liapunov?
2. (Capítulo 1) Un internauta se mueve por la red de cinco páginas del capítulo 1. Escribe la cadena de Markov como un sistema $x^{k+1} = Px^k$ y compárala con $y' = Ay$: ¿qué autovalor decide la convergencia en cada caso?
3. (EDO) Enuncia el teorema de LaSalle y úsalo en el modelo SIR reducido a (I, R), sobre el triángulo invariante, con $V = S = N - I - R$. Comprueba primero que $\dot{V} = -\beta SI \leq 0$ y di por qué $V = I + R$ *no* sirve para esto, aunque sea la elección que uno hace primero. Después identifica el mayor invariante contenido en $\{\dot{V} = 0\}$ y explica qué le falta al argumento para concluir que $I \to 0$ desde un dato con $S_0 > 0$.
4. Clasifica los cuatro casos del modelo general de competición con $\lambda = 1$. Comprueba primero cuándo el equilibrio no axial tiene ambas componentes positivas; usa después la traza y el determinante para decidir su estabilidad y analiza los equilibrios axiales.
5. (Capítulo 3, anticipo) El péndulo sin rozamiento conserva la energía. Escribe el sistema de primer orden, encuentra su cantidad conservada y compárala con la de Volterra: ¿en qué se parecen las órbitas y en qué no?

CAPÍTULO 3

Modelos obtenidos a partir del Cálculo de Variaciones

ANTES DE LEER

Cuatro preguntas antes de leer, a libro cerrado.

1. Para minimizar una función de una variable se iguala la derivada a cero. Si lo que se quiere minimizar es una integral que depende de una *función* entera, ¿qué hace el papel de la derivada, y qué tipo de ecuación sale al igualarla a cero?
2. ¿Por qué el camino más corto entre dos puntos es la recta? No la respuesta de siempre: una demostración que valga también para el camino más rápido, o el más barato.
3. Un péndulo sin rozamiento, ¿tiene alguna posición de equilibrio estable que el criterio de linealización no sepa decidir? ¿Con qué se decide?
4. Una cuerda colgada de dos puntos toma una forma. ¿Qué es lo que la naturaleza está minimizando al elegir esa forma, y por qué al añadir muelles en los extremos cambian las condiciones de contorno y no la ecuación?

Las respuestas están en las secciones 3.1, 3.2, 3.3.2 y 3.4.

Este tema cambia la clase de problema. En el capítulo 2 el modelo era una ecuación diferencial y la pregunta era cómo se comportan sus soluciones. Aquí el modelo es un **problema variacional**: se parte de una cantidad que depende de toda una trayectoria o de toda una forma. Primero se estudia cómo minimizarla (un tiempo, una longitud o una energía). La ecuación diferencial aparece como una *consecuencia*: una condición necesaria del mínimo. En Mecánica se aplicará la misma cuenta a la acción, cuya primera variación se anula aunque la trayectoria no siempre sea un mínimo: es una acción *estacionaria*. Esa es la idea del Cálculo de Variaciones, y es una idea de modelado tan potente como la del capítulo anterior, porque muchas leyes de la Física se enuncian con más naturalidad como un principio de mínimo que como una ecuación. El tema se desarrolla en cuatro secciones: el

principio general (3.1), tres problemas geométricos clásicos (3.2), la mecánica de partículas y del péndulo vista como mínima acción (3.3), y las configuraciones de equilibrio de una cuerda vistas como energía mínima (3.4). Al final, las prácticas de MATLAB del tema, que introducen el resolvedor de problemas de contorno `bvp4c`.

CONEXIÓN: QUÉ SE USA DE OTRAS ASIGNATURAS

De Cálculo Diferencial, el teorema de Fermat (en un mínimo interior la derivada se anula) y la integración por partes: son las dos únicas herramientas de la sección 3.1, y se reenuncian donde se usan. De EDO, la teoría de problemas de contorno para una EDO de segundo orden (capítulo 5 del libro de EDO del autor, «Problemas de contorno para un SDO lineal», reenunciado en el apéndice A): allí se ve (§5.1 y §5.2) que un problema de contorno, a diferencia del de Cauchy, puede no tener solución o tener infinitas, y el teorema de alternativa de §5.4 dice cuándo; aquí aparecen los dos casos. De Geometría local de curvas y superficies, la noción de geodésica, cuyos tramos suficientemente cortos minimizan la longitud entre sus extremos sobre una superficie, que reaparece en la sección 3.2. Y de Mecánica, las energías cinética y potencial y la segunda ley de Newton, que aquí se *deducen* en vez de postularse.

3.1 Principio de Lagrange

LA PREGUNTA

En junio de 1696 Johann Bernoulli publicó un desafío: dados dos puntos a distinta altura, no en la misma vertical, encontrar la curva por la que una bola que se deja caer sin rozamiento llega antes del uno al otro. Se cuenta que Newton lo resolvió en una noche; la solución no era ni la recta ni el arco de circunferencia. Pero la pregunta que abrió medio siglo de trabajo fue otra: ¿hay un método general, uno solo, para encontrar la función que hace mínima una cantidad que depende de toda la función?

Se considera un abierto no vacío $\mathcal{O} \subset \mathbb{R}^{2N+1}$ y una función $f : \mathcal{O} \to \mathbb{R}$. Los puntos de $\mathcal{O}$ se denotan (t, x, x'), con $t \in \mathbb{R}$, $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ y $x' = (x'_1, \dots, x'_N) \in \mathbb{R}^N$; aquí x' es el nombre de la tercera variable, no todavía una derivada. Se escribe

$$\nabla_x f := \Big(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_N}\Big), \qquad \nabla_{x'} f := \Big(\frac{\partial f}{\partial x'_1}, \dots, \frac{\partial f}{\partial x'_N}\Big).$$

Se supone $f \in C^0(\mathcal{O})$ y que existen $\nabla_x f, \nabla_{x'} f \in C^0(\mathcal{O}; \mathbb{R}^N)$. Dados un intervalo cerrado $[a, b] \subset \mathbb{R}$, con $a < b$, y dos puntos $x_a, x_b \in \mathbb{R}^N$, interesa el **problema de mínimos**

$$\begin{cases} \text{Minimizar } \displaystyle\int_a^b f\big(t, x(t), x'(t)\big)\, dt \\ \text{Sujeto a: } x \in C^1([a,b])^N,\ \big(t, x(t), x'(t)\big) \in \mathcal{O}\ \forall t \in [a,b],\ x(a) = x_a,\ x(b) = x_b. \end{cases} \tag{P}$$

Es uno de los problemas más elementales del Cálculo de Variaciones. Lo que se minimiza es un **funcional**: una aplicación J que a cada función x admisible le asigna el número $J(x) = \int_a^b f(t, x, x')\, dt$. Las incógnitas no son números sino funciones, y el conjunto donde se busca el mínimo, las funciones C^1 que pasan por x_a en a y por x_b en b, tiene dimensión infinita.

LA IDEA: UN PROBLEMA DE MÍNIMOS EN DIMENSIÓN INFINITA

En Cálculo se minimiza una función $F : \mathbb{R}^n \to \mathbb{R}$ y la condición necesaria es $\nabla F = 0$: n ecuaciones. Aquí hay «infinitas variables», el valor de x en cada t, y la condición necesaria va a ser «una ecuación por cada t», es decir, una ecuación diferencial. Toda la sección es traducir «la derivada se anula en el mínimo» a este contexto. La herramienta es siempre la misma y conviene recordarla paso a paso: se perturba el mínimo x en una dirección h, se mira la función de una variable $\sigma \mapsto J(x + \sigma h)$, que tiene un mínimo en $\sigma = 0$, se deriva y se iguala a cero. El resto es integrar por partes y un lema que dice que si una integral se anula contra todas las h, el integrando es cero.

El problema puede no tener solución. Con $N = 1$, $[a, b] = [-1, 1]$, $x_a = -1$, $x_b = 1$ y

$$J(x) = \int_{-1}^{1} t^2 x'(t)^2 \, dt,$$

el problema (P) no tiene solución. En efecto, las funciones $x_n(t) = \arctan(nt)/\arctan(n)$ cumplen $x_n(\pm 1) = \pm 1$ y $J(x_n) \to 0$ cuando $n \to +\infty$, mientras que $J(x) > 0$ para toda $x \in C^1([-1, 1])$ con $x(-1) = -1$ y $x(1) = 1$. La comprobación merece hacerse. Para el segundo punto: si $J(x) = 0$ entonces $t^2 x'(t)^2 = 0$ para todo t, luego $x' \equiv 0$ salvo quizá en $t = 0$, y por continuidad de x' es $x' \equiv 0$: x constante, incompatible con $x(-1) = -1$, $x(1) = 1$. Para el primero, $x_n'(t) = \frac{n}{(1+n^2t^2)\arctan n}$ y

$$\begin{aligned} J(x_n) &= \frac{n^2}{\arctan^2 n} \int_{-1}^{1} \frac{t^2}{(1+n^2t^2)^2} \, dt = \frac{n^2}{\arctan^2 n} \cdot \frac{1}{n^3} \int_{-n}^{n} \frac{s^2}{(1+s^2)^2} \, ds \\ &\leq \frac{1}{n \arctan^2 n} \int_{-\infty}^{\infty} \frac{s^2}{(1+s^2)^2} \, ds = \frac{\pi/2}{n \arctan^2 n} \to 0, \end{aligned}$$

con el cambio $s = nt$. El ínfimo es 0 y no se alcanza: las x_n tienden a la función escalón, que no es C^1, y el mínimo «querría» ser esa función que no está en el conjunto admisible.

Para poder hablar de las soluciones de (P) hay que imponer hipótesis adicionales a f. Aquí se pasa por encima de esa dificultad: **se supone que existe al menos una solución** $x = x(t)$ de (P) y se trata de caracterizarla y, cuando sea posible, calcularla. El objetivo principal es obtener las **condiciones necesarias de optimalidad** que debe cumplir cualquier solución; cuando esas condiciones sean también suficientes, se podrá identificar la solución de forma completa. Es exactamente el esquema del capítulo 4, donde además se demostrará la existencia.

EL CAMINO QUE NO SE TOMÓ: DEMOSTRAR PRIMERO QUE EL MÍNIMO EXISTE

La alternativa. Antes de caracterizar el mínimo, probar que existe, como se hace en el capítulo 4 con el método directo: tomar una sucesión minimizante y extraer una subsucesión convergente.
Su ventaja. Distingue dos preguntas: si hay un mínimo y qué ecuación satisface. Una condición necesaria puede incluso ayudar a demostrar que no existe, pero no garantiza por sí sola que alguno de sus candidatos minimice.
Por qué no se hace aquí. Porque en $C^1([a, b])$ con la norma natural no hay compacidad automática: una sucesión minimizante puede no converger a una función C^1, como las x_n del ejemplo. Una vía frecuente es trabajar en espacios de Sobolev, junto con estimaciones que den compacidad débil y una propiedad de semicontinuidad que permita pasar al límite. Ampliar el espacio no basta: el peso t^2 del ejemplo permite pendientes enormes cerca de cero, y ese problema tampoco alcanza el mínimo en H^1 (funciones con primera derivada débil en

L^2). El capítulo 4 desarrollará el método directo en $L^2(0,T)$ para el control. Aquí se obtienen condiciones necesarias en C^1 y se justifica la existencia cuando las propiedades del ejemplo lo permiten.

Hace falta el siguiente lema, que es la pieza que convierte «la integral se anula contra toda h» en «el integrando es cero».

LA IDEA, EN VERSIÓN INTUITIVA: SI NO SE ANULARA, SE LA PILLA CON UNA JOROBA

Este lema es el que convierte una condición integral en una condición puntual, y es el motor de todo el cálculo de variaciones: sin él, de «la primera variación es nula» no se saldría nunca a la ecuación de Euler-Lagrange.
Lo que dice es de sentido común llevado al límite. Se sabe que g tiene integral nula *contra todas* las funciones de prueba h que se anulan en los extremos. Si g no fuera idénticamente cero, habría un punto t_0 donde no se anula; por continuidad conserva su signo en todo un intervalito alrededor. Y entonces basta con ser malintencionado al elegir h: se toma una *joroba*, una función positiva concentrada justo en ese intervalito y nula fuera de él. Contra esa h, el integrando gh tiene signo constante y no es idénticamente nulo, luego la integral no puede ser cero. Contradicción.
La imagen que conviene guardar es esa: **la familia de funciones de prueba es tan grande que puede señalar con el dedo cualquier trocito del intervalo**. Si g escondiera algo en algún sitio, habría una h que lo saca a la luz. Por eso, en la práctica, el lema se usa siempre igual: se llega a una expresión de la forma «$\int(\text{algo})\,h = 0$ para toda h admisible» y se concluye que ese algo es cero en todo punto, que es la ecuación diferencial buscada.

Teorema (lema fundamental del Cálculo de Variaciones)**.** Sea $g \in C^0([a,b])$ y supongamos que

$$\int_a^b g(t)h(t)\,dt = 0 \qquad \text{para toda } h \in C^1([a,b]) \text{ con } h(a) = h(b) = 0.$$

Entonces $g \equiv 0$.

Paso 1: si g falla, falla en una franja entera (la continuidad como lupa). Por reducción al absurdo. Si $g \not\equiv 0$, existe $t_0 \in [a,b]$ con $g(t_0) \neq 0$. Por continuidad existe un intervalo $[c,d] \subset [a,b]$ con $c < d$ y $g \neq 0$ en (c,d); cambiando g por $-g$ si hace falta, se puede suponer $g > 0$ en (c,d).
Este paso es el que hace falsa la versión del lema sin continuidad: una g que valiera 1 en un único punto y 0 en el resto tendría integral nula contra cualquier h, y no sería idénticamente nula. Lo que la continuidad garantiza es que un valor no nulo *ocupa sitio*, y ese sitio es lo que después se puede señalar.
Paso 2: construir la joroba que señala esa franja — el núcleo de la prueba. Se define

$$h(t) = \begin{cases} (t-c)^2(d-t)^2 & \text{si } t \in [c,d], \\ 0 & \text{si } t \in [a,b] \setminus [c,d]. \end{cases}$$

La elección tiene dos exigencias enfrentadas, y el cuadrado en cada factor es lo que las concilia: h tiene que *anularse fuera de* (c,d), para no recoger información de donde

no se sabe nada de g, y a la vez tiene que ser *admisible*, es decir, C^1 y nula en los extremos, para poder aplicarle la hipótesis. Con factores simples $(t-c)(d-t)$ la primera exigencia se cumple pero la segunda no: la derivada saltaría en c y en d.
Esta h es de clase C^1 en $[a,b]$: dentro de $[c,d]$ es un polinomio, fuera es cero, y en los puntos de empalme c y d tanto h como h' valen cero por los dos lados (el cuadrado en cada factor es lo que hace que la derivada también se anule). Además $h(a) = h(b) = 0$ porque $a \le c < d \le b$ y fuera de $[c,d]$ es nula.
Paso 3: la integral filtrada es positiva, y eso contradice la hipótesis. Como h y g son estrictamente positivas en (c,d) y h vale cero fuera,

$$\int_a^b g(t)h(t)\,dt = \int_c^d g(t)h(t)\,dt > 0,$$

porque el integrando es continuo, positivo en (c,d) y no idénticamente nulo. Pero la hipótesis dice que esa integral es cero. Contradicción. ■

Qué hace cada hipótesis. Solo hay dos, y cada una bloquea un contraejemplo distinto. La *continuidad de g* sostiene el paso 1: sin ella el enunciado es falso tal como está escrito; en la versión con g integrable, la conclusión es $g = 0$ salvo en un conjunto de medida nula. Y que la igualdad valga *para toda h* admisible sostiene el paso 2: si el conjunto de funciones de prueba fuese pequeño, la conclusión se debilitaría hasta desaparecer. Con una sola h fija, «$\int gh = 0$» dice únicamente que g es ortogonal a esa h, que es casi nada; con todas las h que se anulan en los extremos, g resulta ortogonal a un conjunto tan grande que no le queda sitio donde esconderse. La riqueza de la familia de prueba *es* el teorema.

EXPLÍCATE. *¿Por qué se toma $(t-c)^2(d-t)^2$ y no $(t-c)(d-t)$? Calcula la derivada de la segunda en $t = c$ por la derecha y verás que no empalma con el cero de la izquierda.*

Teorema (condición necesaria de Euler-Lagrange)**.** Si $x = x(t)$ es una solución (local) del problema (P) tal que la función $t \mapsto \nabla_{x'} f(t, x(t), x'(t))$ pertenece a $C^1([a,b])^N$, entonces x es solución del problema de contorno

$$\begin{cases} -\dfrac{d}{dt}\nabla_{x'} f(t,x,x') + \nabla_x f(t,x,x') = 0 \quad \text{en } (a,b), \\ x(a) = x_a, \quad x(b) = x_b. \end{cases} \tag{EL}$$

Ejemplo (el teorema en el caso más simple, antes de demostrarlo)**.** Con $N = 1$ y $f(t,x,x') = \frac{1}{2}x'^2 + gx$ (una constante g), las dos derivadas parciales son $\partial_{x'} f = x'$ y $\partial_x f = g$. La ecuación (EL) dice $-\frac{d}{dt}(x') + g = 0$, es decir, $x'' = g$: la única función que puede minimizar $\int (\frac{1}{2}x'^2 + gx)\,dt$ entre las que valen x_a en a y x_b en b es una parábola $x = \frac{g}{2}t^2 + c_1 t + c_2$, con c_1, c_2 fijados por los extremos. Es exactamente la cuerda colgada de la sección 3.4. Lo que la demostración va a probar es que esa deducción («deriva f respecto de x', deriva el resultado respecto de t, resta la derivada respecto de x, iguala a cero») es correcta en general.

La idea: la demostración en una frase por paso

Se toma el mínimo x y se le suma un pequeño múltiplo de una función h que vale cero en los extremos, para no estropear las condiciones de contorno (paso 1). Al variar el múltiplo σ, el valor del funcional es una función de un solo número, con mínimo en $\sigma = 0$, así que su derivada en 0 es cero (paso 2). Esa derivada tiene un término con h y otro con h'; integrando por partes el de h', todo queda multiplicado por h (paso 3). Y si una función continua multiplicada por *cualquier* h integra cero, la función es cero (paso 4). Los cuatro pasos son el método de todo el capítulo y del siguiente.

Dónde está el peso. No en el paso 2, que es el teorema de Fermat de toda la vida, sino en el **paso 3**. Ahí está la única dificultad real, y conviene nombrarla: la derivada $F'(0)$ sale con *dos* términos, uno multiplicado por h y otro por h', y mientras queden los dos no se puede aplicar directamente el lema fundamental: h y h' no son dos funciones de prueba independientes, sino una función y su derivada. La integración por partes es lo que rompe esa atadura: traslada la derivada de h al coeficiente y deja a h sola, libre y arbitraria. Y el precio que cobra —un término de frontera— es exactamente el que las condiciones de contorno $h(a) = h(b) = 0$ estaban preparadas para pagar. Por eso las perturbaciones se eligieron nulas en los extremos desde el paso 1: preservan los datos del problema y, además, hacen desaparecer la contribución del borde al integrar por partes.

Y una advertencia sobre qué se ha demostrado. Es tentador leer Euler-Lagrange como «la ecuación de los mínimos», y no lo es. Todo el argumento parte de suponer que el mínimo *existe* y es C^1, y lo único que hace es extraerle una consecuencia. Es una condición *necesaria*: sus soluciones son candidatos, igual que en $\mathbb{R}^n$ los puntos con gradiente nulo incluyen máximos y puntos de silla. Puede ocurrir que la ecuación tenga solución y el problema no tenga mínimo. Si se plantea la minimización en un espacio más amplio que C^1, también puede existir un mínimo no derivable al que esta versión clásica del teorema no se aplique.

Paso 1: las perturbaciones admisibles. Sea $h \in C^1([a,b])^N$ con $h(a) = h(b) = 0$. El conjunto $\{(t, x(t), x'(t)) : t \in [a,b]\}$ es la imagen continua de un compacto, luego es compacto, y está contenido en el abierto $\mathcal{O}$; por tanto su distancia a la frontera $\partial\mathcal{O}$ es positiva. Como h y h' están acotadas en $[a,b]$, existe $\varepsilon_0 > 0$ tal que para todo σ con $|\sigma| < \varepsilon_0$ la función $x + \sigma h$ es admisible: es de clase C^1, $(t, (x+\sigma h)(t), (x'+\sigma h')(t)) \in \mathcal{O}$ para todo t (está a distancia menor que la de la frontera), y $(x+\sigma h)(a) = x_a$, $(x+\sigma h)(b) = x_b$ porque h se anula en los extremos. Por definición de mínimo,

$$\int_a^b f(t,x,x')\,dt \le \int_a^b f(t, x+\sigma h, x'+\sigma h')\,dt \qquad \text{para todo } |\sigma| < \varepsilon_0.$$

Paso 2: reducción a una variable. La función $F : (-\varepsilon_0, \varepsilon_0) \to \mathbb{R}$,

$$F(\sigma) = \int_a^b f(t, x+\sigma h, x'+\sigma h')\,dt,$$

alcanza un mínimo en $\sigma = 0$. Las hipótesis de derivabilidad sobre f (gradientes continuos) permiten derivar bajo el signo integral, porque el integrando y su derivada

respecto de σ son continuos en el compacto $[a,b]\times[-\varepsilon_0/2,\varepsilon_0/2]$, y por la regla de la cadena

$$F'(\sigma)=\int_a^b\Big(\nabla_x f(t,x+\sigma h,x'+\sigma h')\cdot h+\nabla_{x'} f(t,x+\sigma h,x'+\sigma h')\cdot h'\Big)\,dt.$$

Como F alcanza un mínimo en un punto interior, por el teorema de Fermat $F'(0)=0$:

$$\begin{aligned}&\int_a^b\Big(\nabla_x f(t,x,x')\cdot h+\nabla_{x'} f(t,x,x')\cdot h'\Big)\,dt=0\\&\text{para toda } h\in C^1([a,b])^N \text{ con } h(a)=h(b)=0.\end{aligned}\tag{3.5}$$

Leída en palabras: la *variación primera* del funcional en la dirección h, que es la suma de «cuánto cambia f si se mueve x» por h más «cuánto cambia f si se mueve x'» por h', integrada en todo el intervalo, es cero para cualquier dirección h admisible. Es el análogo exacto de «$\nabla F\cdot h=0$ para todo h» en $\mathbb{R}^n$.

Paso 3: integración por partes, para dejar a h sola — el núcleo de la prueba. Suponiendo que $t\mapsto\nabla_{x'} f(t,x(t),x'(t))$ es de clase C^1, se integra por partes el segundo sumando:

$$\int_a^b\nabla_{x'} f\cdot h'\,dt=\Big[\nabla_{x'} f\cdot h\Big]_a^b-\int_a^b\frac{d}{dt}\nabla_{x'} f\cdot h\,dt=-\int_a^b\frac{d}{dt}\nabla_{x'} f\cdot h\,dt,$$

donde el término de frontera desaparece porque $h(a)=h(b)=0$. Queda

$$\begin{aligned}&\int_a^b\Big(\nabla_x f(t,x,x')-\frac{d}{dt}\nabla_{x'} f(t,x,x')\Big)\cdot h\,dt=0\\&\text{para toda } h\in C^1([a,b])^N \text{ con } h(a)=h(b)=0.\end{aligned}\tag{3.6}$$

Ahora h aparece sola, sin derivar: la integración por partes ha pasado la derivada de h a $\nabla_{x'} f$, y el término de frontera ha desaparecido por las condiciones $h(a)=h(b)=0$. Lo que multiplica a h dentro de la integral es, en cada t, un vector de N componentes, y el paso siguiente prueba que es el vector cero.

Paso 4: el lema. Tomando $h=(0,\dots,\eta,\dots,0)$ con $\eta\in C^1([a,b])$, $\eta(a)=\eta(b)=0$, en la componente i, el lema fundamental aplicado a la función continua $g_i=\partial_{x_i} f-\frac{d}{dt}\partial_{x_i'} f$ da $g_i\equiv 0$ para cada i. Son las N EDOs de (EL); las condiciones de contorno son las del problema. ■

Qué hace cada hipótesis. Se puede seguir en qué pasos intervienen y qué se rompe al quitarla.

La *continuidad de los gradientes de f en x y x'*, es lo que autoriza a derivar bajo el signo integral en el paso 2; sin ella F podría no ser derivable y el teorema de Fermat no tendría dónde agarrarse.

Que $\mathcal{O}$ sea *abierto* y la curva viva dentro es lo que da el $\varepsilon_0>0$ del paso 1, y ahí se usa la compacidad de la curva: es lo que garantiza que la distancia a la frontera es positiva y no un ínfimo nulo. Si el mínimo tocara la frontera del dominio, no habría margen para perturbar en las dos direcciones, F tendría el mínimo en un extremo del intervalo y $F'(0)=0$ sería falso. Es la misma diferencia que en $\mathbb{R}^n$ entre un mínimo interior y uno en el borde.

Que $\nabla_{x'} f$ compuesta con la solución sea *de clase* C^1 es la hipótesis del paso 3, y es la más incómoda de las tres porque es una hipótesis sobre la solución, no sobre los datos. Se puede evitar: la observación de Du Bois-Reymond que sigue integra por partes al revés, sobre el otro sumando, y obtiene la ecuación en forma integral sin pedirle esa regularidad a nadie.
Y las condiciones $h(a) = h(b) = 0$ son las que matan el término de frontera del paso 3. Cuando el problema no fija un extremo —la observación final del capítulo 4— ese término *no* desaparece, y lo que aporta no es un estorbo: es precisamente la condición de contorno natural que sustituye al dato que falta.

Lo que hay que llevarse. Cuatro frases reconstruyen la prueba, y son las mismas en todo este capítulo y en el siguiente. (i) *Perturbar sin salirse*: $x + \sigma h$, con h nula en los extremos, sigue siendo admisible, y $\sigma \mapsto J(x + \sigma h)$ es una función de una sola variable con mínimo en $\sigma = 0$. (ii) *Fermat*: su derivada en 0 se anula, y esa derivada es la variación primera (3.5), con un término en h y otro en h'. (iii) *Partes*: se pasa la derivada de h al coeficiente, y el término de borde no cuenta porque $h(a) = h(b) = 0$. (iv) *Lema*: lo que multiplica a h se anula en cada punto, y eso es la ecuación. Cuando un extremo no está fijado solo cambia la frase (iii): el borde sí cuenta, y lo que aporta es la condición de contorno que falta.

CONEXIÓN: EL TEOREMA DE FERMAT Y LA REGLA DE LA CADENA

Teorema de Fermat. Si $F : (-\varepsilon_0, \varepsilon_0) \to \mathbb{R}$ es derivable en 0 y tiene en 0 un extremo relativo, entonces $F'(0) = 0$. Demostración: si fuera $F'(0) > 0$, el cociente incremental sería positivo cerca de 0 y $F(\sigma) > F(0)$ para $\sigma > 0$ pequeño, $F(\sigma) < F(0)$ para $\sigma < 0$; no sería extremo. Lo mismo con < 0. Es el punto de partida de las condiciones de primer orden de este capítulo y del siguiente. **Derivación bajo la integral.** Si $\Phi(t, \sigma)$ y $\partial_\sigma \Phi$ son continuas en $[a, b] \times [-\delta, \delta]$, entonces $\frac{d}{d\sigma} \int_a^b \Phi \, dt = \int_a^b \partial_\sigma \Phi \, dt$; es el teorema de Leibniz de Cálculo de varias variables. Aquí $\partial_\sigma f(t, x + \sigma h, x' + \sigma h') = \nabla_x f \cdot h + \nabla_{x'} f \cdot h'$ por la regla de la cadena.

DE DÓNDE VIENE: EULER, LAGRANGE Y LA CARTA DE 1755

Euler obtuvo la ecuación en 1744 en su *Methodus inveniendi*, discretizando la curva en una poligonal, escribiendo la integral como una suma, minimizando respecto de cada vértice y pasando al límite: un método geométrico y laborioso. Comparte con los métodos numéricos la idea de discretizar, pero `bvp4c` resuelve ecuaciones de contorno por colocación, no minimiza esa poligonal. En 1755 Lagrange, con diecinueve años, le escribió desde Turín con el método de las variaciones que se acaba de ver, el δ que perturba la función entera, y Euler adoptó el nombre «cálculo de variaciones» y el método. La ecuación lleva los dos nombres. El lema de Du Bois-Reymond que se presenta después permite precisar la regularidad necesaria para esta deducción.

La condición necesaria de mínimo es un problema de contorno que, en los casos regulares, toma la forma de un sistema de N ecuaciones de segundo orden. Si la dependencia de f en x' es degenerada, ese orden puede reducirse. Si ese problema tiene a lo más una solución, queda asegurada la unicidad de solución de (P), aunque no necesariamente la existencia. Las condiciones $x(a) = x_a$ y $x(b) = x_b$ fijan los valores de la incógnita en la frontera de su dominio: se llaman **condiciones de Dirichlet**. Otras restricciones en el problema de mínimos, o la ausencia de ellas, conducen a otras condiciones de contorno.

Extremo libre: la condición de Neumann. Se parte del problema

$$\begin{cases} \text{Minimizar } \displaystyle\int_a^b f(t,x,x')\,dt \\ \text{Sujeto a: } x \in C^1([a,b])^N,\ (t,x(t),x'(t)) \in \mathcal{O}\ \forall t,\ x(a) = x_a, \end{cases}$$

donde no se impone ninguna restricción en el extremo b. Si x es solución, el mismo razonamiento lleva a la igualdad (3.5), con la diferencia de que ahora solo hay que imponer $h(a) = 0$ (para que $x + \sigma h$ siga valiendo x_a en a), y h puede tomar cualquier valor en b. Integrando por partes, el término de frontera ya no desaparece:

$$\nabla_{x'} f(b, x(b), x'(b)) \cdot h(b) + \int_a^b \left(\nabla_x f - \frac{d}{dt}\nabla_{x'} f\right) \cdot h\,dt = 0$$
$$\text{para toda } h \in C^1([a,b])^N \text{ con } h(a) = 0.$$

Se procede en dos tiempos. Primero, tomando h con $h(b) = 0$ además, desaparece el término de frontera y el lema da la EDO de (EL). Segundo, sabiendo ya que la integral es cero para toda h, queda $\nabla_{x'} f(b, x(b), x'(b)) \cdot h(b) = 0$ para todo valor $h(b) \in \mathbb{R}^N$, luego

$$\nabla_{x'} f(b, x(b), x'(b)) = 0.$$

La consecuencia es que x debe ser solución del problema de contorno

$$\begin{cases} -\dfrac{d}{dt}\nabla_{x'} f(t,x,x') + \nabla_x f(t,x,x') = 0 \quad \text{en } (a,b), \\ x(a) = x_a, \quad \nabla_{x'} f(b, x(b), x'(b)) = 0. \end{cases}$$

La igualdad en b es una **condición natural**, llamada aquí de Neumann; fija $\nabla_{x'} f = 0$, que no siempre equivale a $x' = 0$. Si no se impone ninguna restricción ni en a ni en b, la cuenta es la misma con h arbitraria y aparecen *dos* términos de frontera en vez de uno: la EDO de (EL) sale igual, y quedan las dos condiciones naturales $\nabla_{x'} f(a, x(a), x'(a)) = 0$ y $\nabla_{x'} f(b, x(b), x'(b)) = 0$, una por cada extremo que se ha dejado suelto. El detalle, con el signo del término en a y un ejemplo resuelto, está en el ejercicio de los dos extremos libres.

LA IDEA: LO QUE SE DEJA LIBRE SE PAGA CON UNA CONDICIÓN NATURAL

Cada condición de contorno que el problema *no* impone, la impone el mínimo por su cuenta: si el extremo puede moverse, en el mínimo la «fuerza» $\nabla_{x'} f$ en ese extremo tiene que ser cero, porque si no, moviéndolo se mejoraría. Por eso a las condiciones de Neumann se las llama *naturales*: no están en el enunciado, salen de la minimización. En la cuerda de la sección 3.4, $\nabla_{x'} f = \mu u'$ es la tensión vertical en el extremo, y «extremo libre» significa «tensión cero», que es lo que uno esperaría. El recuento da, para un sistema regular de N ecuaciones de segundo orden, $2N$ condiciones entre las impuestas y las naturales. Tener ese número de condiciones no garantiza por sí solo existencia ni unicidad de un problema de contorno.

Observación (lema de Du Bois-Reymond)**.** No hace falta suponer de antemano que la función

$$t \longmapsto \nabla_{x'} f(t, x(t), x'(t))$$

sea de clase C^1. Si se tiene (3.5), se puede probar esa regularidad y que se cumple (3.6). Es consecuencia de un resultado con interés propio, el **lema de Du Bois-Reymond**: si $u \in C^0([a,b])$ y

$$\int_a^b u(t)h'(t)\,dt = 0 \qquad \text{para toda } h \in C^1([a,b]) \text{ con } h(a) = h(b) = 0,$$

entonces u es constante. Un comentario análogo vale para el problema con extremo libre.

La idea. El lema fundamental prueba «g es cero» eligiendo una h que señale el sitio donde g falla. Aquí no se puede: lo que aparece en la integral es h', y h' no es libre —tiene que ser la derivada de algo que se anula en los dos extremos, es decir, una función de *media nula*—. Así que u solo puede ser ortogonal a las funciones de media nula, y el complemento ortogonal de ese conjunto son las constantes. Eso es todo el enunciado.

Y hay una manera de verlo que además *construye* la prueba: si u ha de ser constante, esa constante solo puede ser su media c. Así que se toma c, se mide la desviación $u - c$, y se elige como función de prueba precisamente la primitiva de esa desviación. Con esa elección la hipótesis dice, literalmente, que la desviación es ortogonal a sí misma; y lo único ortogonal a sí mismo es cero. **Toda la prueba cabe en esa frase**: se le hace la pregunta al enunciado usando como pregunta la propia respuesta que se busca.

Del lema de Du Bois-Reymond. Paso 1: elegir como función de prueba la primitiva de la desviación (el núcleo). Sea $c = \frac{1}{b-a}\int_a^b u$ la media de u y $h(t) = \int_a^t (u(s) - c)\,ds$. Esta h es C^1 por el teorema fundamental del cálculo, con $h' = u - c$; cumple $h(a) = 0$ por construcción y $h(b) = \int_a^b u - c(b-a) = 0$ justo porque c es la media, ni un valor mayor ni menor. Es admisible, y esa es la única razón por la que se ha elegido la media y no otra constante.

Paso 2: la hipótesis se convierte en una norma nula. Con esa h, $0 = \int_a^b u\,h' = \int_a^b u(u-c)$. Como también $\int_a^b c(u-c) = c\int_a^b (u-c) = 0$ —de nuevo porque c es la media—, restando queda

$$\int_a^b \big(u(t) - c\big)^2 dt = 0.$$

Una función continua y no negativa con integral nula es idénticamente nula: si valiera algo positivo en un punto, por continuidad valdría algo positivo en un intervalo y la integral sería positiva. Luego $u \equiv c$. ■

Qué hace cada hipótesis, y para qué sirve el lema. La continuidad de u se usa dos veces: para que h sea C^1 en el paso 1 y para el argumento final del paso 2. Y la condición $h(a) = h(b) = 0$ es lo que restringe las h' admisibles a las de media nula, que es justo lo que hace que la conclusión sea «u es constante» y no «u es cero»: sin ella se podría tomar $h(t) = t$, con $h' \equiv 1$, y saldría además $\int u = 0$, forzando $c = 0$. La utilidad requiere un paso más. Escribamos

$$A(t) = \nabla_x f(t, x(t), x'(t)), \qquad B(t) = \nabla_{x'} f(t, x(t), x'(t)),$$
$$Q(t) = \int_a^t A(s)\,ds.$$

Como $Q' = A$ y h se anula en los extremos, (3.5) da

$$0 = \int_a^b (A \cdot h + B \cdot h')\, dt = \int_a^b (B - Q) \cdot h'\, dt.$$

Ahora sí se aplica el lema, componente a componente, a $B - Q$: es constante. Por tanto, $B = Q +$ constante es C^1 y $B' = Q' = A$, que es Euler-Lagrange. La derivabilidad de B sale como conclusión; no se ha supuesto para integrar por partes.

LA IDEA: BELTRAMI ES LA CONSERVACIÓN DE LA ENERGÍA, Y POR QUÉ

Antes de la cuenta conviene saber qué se está haciendo, porque la cuenta la esconde.
La afirmación. Si el integrando no mira el reloj —si f depende de dónde se está y de cómo se va, pero no de *cuándo*—, entonces a lo largo de cualquier solución de Euler-Lagrange hay una cantidad conservada. Para una incógnita escalar permite reducir el orden de la ecuación; para un sistema aporta una integral primera, que no resuelve por sí sola todas las componentes.
Por qué tiene que ser verdad. La imagen es la de un paisaje que no cambia con el tiempo. Si el terreno es el mismo hoy que mañana, un recorrido óptimo empezado hoy y el mismo recorrido empezado mañana cuestan lo mismo: el problema es *invariante frente a traslaciones en el tiempo*. Esta invariancia corresponde a las leyes locales: desplazar una trayectoria en el tiempo no preserva, en general, los extremos fijados del problema variacional. Por eso no podemos usar sin más $h = x'$ como una perturbación admisible en la prueba anterior. La simetría orienta la búsqueda de una cantidad conservada; la comprobación rigurosa será derivar $f - \nabla_{x'} f \cdot x'$ y usar Euler-Lagrange. El factor x' conecta esa cuenta con el cambio de posición que produce una pequeña traslación temporal.
Qué es la cantidad conservada. En Mecánica, con $f = \mathcal{L} = E_c - E_p$, la lagrangiana, el vector $\nabla_{x'}\mathcal{L}$ es el momento $p = mx'$ y la expresión $\nabla_{x'}\mathcal{L} \cdot x' - \mathcal{L}$ es $mx'^2 - (\frac{1}{2}mx'^2 - E_p) = \frac{1}{2}mx'^2 + E_p$: la energía total. La identidad dice, con el signo cambiado, que la energía se conserva. Y dice también *cuándo* se conserva: la ausencia de dependencia explícita del tiempo en la lagrangiana garantiza esta conservación. Es una condición suficiente, no una equivalencia para toda trayectoria. Un péndulo cuya cuerda alguien va acortando no conserva la energía, y la razón es exactamente que su $\mathcal{L}$ mira el reloj.
Y la metáfora predice el enunciado. Si en vez de invariancia temporal el problema tuviera invariancia frente a traslaciones *espaciales* —si f no dependiera de una componente x_i— el mismo argumento con esa dirección daría que $\partial_{x_i'} f$ es constante: la conservación del momento en esa dirección. Se ve directamente en la ecuación de Euler-Lagrange, cuya componente i-ésima queda $\frac{d}{dt}\partial_{x_i'} f = \partial_{x_i} f = 0$. Una simetría, una constante del movimiento: es el teorema de Noether, y estos dos casos son sus dos ejemplos más simples.

Observación (identidad de Beltrami)**.** Con frecuencia f no depende explícitamente de t: $f = f(x, x')$. En tal caso, si x es dos veces continuamente diferenciable, multiplicando

escalarmente el SDO de (EL) por x',

$$0 = \Big(\nabla_x f(x,x') - \frac{d}{dt}\nabla_{x'} f(x,x')\Big)\cdot x' = \nabla_x f\cdot x' + \nabla_{x'} f\cdot x'' - \frac{d}{dt}\big(\nabla_{x'} f\cdot x'\big)$$
$$= \frac{d}{dt}\Big(f(x,x') - \nabla_{x'} f(x,x')\cdot x'\Big),$$

donde en el segundo paso se ha sumado y restado $\nabla_{x'} f\cdot x''$ (regla del producto para $\frac{d}{dt}(\nabla_{x'} f\cdot x')$) y en el tercero se ha reconocido $\frac{d}{dt}f(x,x') = \nabla_x f\cdot x' + \nabla_{x'} f\cdot x''$. Por tanto existe $C\in\mathbb{R}$ tal que x satisface la EDO de *primer* orden

$$f(x,x') - \nabla_{x'} f(x,x')\cdot x' = C \qquad \text{en } [a,b],$$

conocida como **identidad de Beltrami**. Es una integral primera de la ecuación de Euler-Lagrange, y en Mecánica es la conservación de la energía.

Dónde está el peso, y qué hace la hipótesis. Los tres pasos de la cadena no pesan igual. El primero es multiplicar por x', y esa es la única decisión del argumento: es la traducción algebraica de perturbar la solución en el tiempo. El segundo, sumar y restar $\nabla_{x'} f\cdot x''$, es la **maniobra clave**, y lo que hace es reconocer que dos sumandos sueltos son en realidad la derivada de un producto; a partir de ahí ya no hay elecciones, solo se lee. El tercero es la regla de la cadena, que aquí tiene *dos* sumandos y no tres, y ahí está toda la hipótesis: si f dependiera explícitamente de t aparecería un $\partial_t f$ suelto que no encaja en ninguna derivada total, y en su lugar quedaría

$$\frac{d}{dt}\Big(f - \nabla_{x'} f\cdot x'\Big) = \partial_t f,$$

que describe la variación de esa cantidad cuando existe $\partial_t f$. En Mecánica, con $f=\mathcal{L}$ y energía $E = \nabla_{x'}\mathcal{L}\cdot x' - \mathcal{L}$, el signo se invierte: $E' = -\partial_t\mathcal{L}$. Para un integrando general, esta identidad no tiene por qué expresar una energía física. La hipótesis $x\in C^2$ permite la cadena de derivadas escrita arriba; con menor regularidad se necesita una justificación adicional.

Observación (problemas de máximos)**.** Análogamente se consideran problemas de máximos: basta cambiar el signo de f, y la conclusión es la misma que en el teorema de Euler–Lagrange. La condición de Euler-Lagrange no distingue mínimos de máximos, igual que $F'=0$ en Cálculo.

El teorema de Euler–Lagrange da una condición necesaria. Como ocurre con funciones de una variable, con hipótesis adicionales de **convexidad** la condición resulta suficiente.

Teorema (condición suficiente de mínimo por convexidad)**.** Si $\mathcal{O}$ es convexo y para todo $t\in[a,b]$ la función $(x,x')\mapsto f(t,x,x')$ es convexa, entonces toda solución del problema de contorno (EL) es solución (global) de (P).

La idea, en una frase. Euler-Lagrange dice que la variación primera es cero, y eso es información *local*: en $\mathbb{R}^n$ sería «el gradiente se anula», que no distingue un mínimo de una silla. La convexidad es lo que globaliza esa información, y lo hace por una sola propiedad: una función convexa *queda siempre por encima de su plano tangente*. Con eso, un punto donde el gradiente se anula no está solo por debajo de sus vecinos: está por debajo de todo el mundo, porque el plano tangente es horizontal y la función entera vive encima. La demostración escribe esa desigualdad punto a punto, la integra, y reconoce que lo que queda es justo la ecuación de Euler-Lagrange, que vale cero.

Sea x una solución de (EL) y sea $\bar{x}$ cualquier función admisible: $\bar{x} \in C^1([a,b])^N$, $(t, \bar{x}(t), \bar{x}'(t)) \in \mathcal{O}$, $\bar{x}(a) = x_a$, $\bar{x}(b) = x_b$. Hay que probar $J(x) \leq J(\bar{x})$.

Paso 1: la función convexa queda por encima de su plano tangente — el núcleo de la prueba. Por la convexidad de f en (x, x'), para todo t y todo $\lambda \in (0,1)$,

$$(1-\lambda) f(t,x,x') + \lambda f(t,\bar{x},\bar{x}') \geq f\big(t, (1-\lambda)x + \lambda\bar{x}, (1-\lambda)x' + \lambda\bar{x}'\big),$$

donde todo se evalúa en t y el punto intermedio está en $\mathcal{O}$ por ser convexo. Reordenando,

$$f(t,\bar{x},\bar{x}') - f(t,x,x') \geq \frac{f\big(t, x+\lambda(\bar{x}-x), x'+\lambda(\bar{x}'-x')\big) - f(t,x,x')}{\lambda}.$$

Tomando límite cuando $\lambda \to 0^+$, el cociente tiende a la derivada direccional, que por la regularidad de f es

$$f(t,\bar{x},\bar{x}') - f(t,x,x') \geq \nabla_x f(t,x,x')\cdot(\bar{x}-x) + \nabla_{x'} f(t,x,x')\cdot(\bar{x}'-x'), \quad \text{para todo } t \in [a,b].$$

Es la desigualdad «una función convexa queda por encima de su plano tangente».

Paso 2: integrar, y reconocer que lo que sobra es (EL), que vale cero. Integrando entre a y b,

$$J(\bar{x}) - J(x) \geq \int_a^b \Big(\nabla_x f\cdot(\bar{x}-x) + \nabla_{x'} f\cdot(\bar{x}'-x')\Big)\,dt.$$

Integrando por partes el segundo término, y usando que $\bar{x}(a) - x(a) = \bar{x}(b) - x(b) = 0$ (los términos de frontera se anulan) y que x es solución de (EL),

$$\int_a^b \Big(\nabla_x f\cdot(\bar{x}-x) + \nabla_{x'} f\cdot(\bar{x}'-x')\Big)\,dt = \int_a^b \Big(\nabla_x f - \frac{d}{dt}\nabla_{x'} f\Big)\cdot(\bar{x}-x)\,dt = 0.$$

Luego $J(x) \leq J(\bar{x})$ para toda $\bar{x}$ admisible, y x es solución de (P). ■

LA IDEA: DÓNDE SE USA CADA HIPÓTESIS

La convexidad en (x, x') *conjuntamente*, no en cada variable por separado, es lo que da la desigualdad del plano tangente; sin ella, (EL) puede dar un máximo, un punto de silla, o un mínimo local que no es global. La convexidad de $\mathcal{O}$ garantiza que el segmento entre x y $\bar{x}$ es admisible. Y la regularidad de f permite pasar del cociente a la derivada. En el primero de los problemas geométricos siguientes, $f = \sqrt{1+u'^2}$ es convexa en u' y no depende de u; en el de la superficie de revolución, $f = u\sqrt{1+u'^2}$ no es convexa en (u, u'). Por eso el primer mínimo se justifica con este criterio y el de la superficie de revolución requiere otro análisis.

Lo que hay que llevarse. Dos frases. (i) Una función convexa queda por encima de su plano tangente, y eso, escrito en cada t e integrado, dice que $J(\bar{x}) - J(x)$ es al menos la variación primera en la dirección $\bar{x} - x$. (ii) Esa variación primera, integrada por partes con los extremos comunes, es la ecuación de Euler-Lagrange multiplicada por $\bar{x} - x$, y vale cero. No hace falta ningún σ pequeño: la dirección $\bar{x} - x$ puede ser tan grande como se quiera, y por eso el

mínimo que sale es global.

Cómo se comprueba la convexidad. El teorema pide que $(x, x') \mapsto f(t, x, x')$ sea convexa para cada t fijo, y la prueba práctica es la de Cálculo: si f es de clase C^2 en (x, x') sobre un convexo, es convexa si y solo si su hessiana en esas variables es semidefinida positiva en cada punto. La variable t no interviene: hace de parámetro. Con $N = 1$ la hessiana es la matriz

$$\begin{pmatrix} f_{xx} & f_{xx'} \\ f_{xx'} & f_{x'x'} \end{pmatrix},$$

y ser semidefinida positiva son tres desigualdades: $f_{xx} \geq 0$, $f_{x'x'} \geq 0$ y $f_{xx} f_{x'x'} - f_{xx'}^2 \geq 0$. Tres casos a mano:

- $f = \frac{1}{2}x'^2 + \frac{1}{2}x^2$: la hessiana es la identidad. Es convexa y el criterio se aplica; es la del ejercicio siguiente.
- $f = x'^4 + x^2 - x$, el mismo tipo que el funcional cuártico resuelto en el apéndice C: la hessiana es diagonal, con $f_{xx} = 2$ y $f_{x'x'} = 12x'^2$, los dos ≥ 0, y el término cruzado es nulo. Es convexa. El término lineal $-x$ no estorba nunca, porque sus segundas derivadas son cero; y que $f_{x'x'}$ se anule en $x' = 0$ tampoco, porque la hessiana sigue siendo semidefinida.
- $f = u\sqrt{1 + u'^2}$, la de la superficie de revolución: $f_{uu} = 0$, pero el término cruzado $f_{uu'} = u'/\sqrt{1 + u'^2}$ no se anula, y el determinante vale $-u'^2/(1 + u'^2) < 0$ en cuanto $u' \neq 0$. No es convexa, como se dijo en la caja anterior.

Un error frecuente es mirar solo la diagonal. En el tercer caso f es convexa en u' con u fija, y lineal (luego convexa) en u con u' fija, y aun así no es convexa en el par: el término cruzado es el que lo impide.

Qué pasa cuando falta: una candidata que no es mínimo. La intuición que precede a la demostración de Euler-Lagrange advertía que la ecuación puede tener solución sin que el problema tenga mínimo. El ejemplo más pequeño es

$$J(x) = \int_0^T \big(x'^2 - x^2\big)\,dt, \qquad x(0) = x(T) = 0,$$

cuya hessiana es $\operatorname{diag}(-2, 2)$: no es convexa en x. Euler-Lagrange da $-2x'' - 2x = 0$, es decir, $x'' + x = 0$, y la función nula es solución, con $J(0) = 0$. ¿Es el mínimo? Depende de la longitud del intervalo.

- *Si $T > \pi$, no.* Con $x = c\sin(\pi t/T)$, que se anula en los dos extremos, sale $J = \frac{c^2 T}{2}\big(\frac{\pi^2}{T^2} - 1\big) < 0$, y al crecer c el funcional baja hasta $-\infty$: no hay mínimo, y la solución de Euler-Lagrange es solo un punto estacionario. Con $T = 4$ se ve con un polinomio: $x = t(4 - t)$ da $\int_0^4 (4 - 2t)^2 dt - \int_0^4 t^2(4 - t)^2 dt = \frac{64}{3} - \frac{512}{15} = -\frac{64}{5}$.
- *Si $T < \pi$, sí, y es el único.* Se toma $s(t) = \sin\big(t + \frac{\pi - T}{2}\big)$, que es positiva en todo $[0, T]$ precisamente porque $T < \pi$, y cumple $s'' = -s$. Toda x admisible se escribe $x = s\,w$, con $w = x/s$ de clase C^1 y $w(0) = w(T) = 0$, y derivando (y usando $s'' = -s$) se comprueba la identidad

  $$x'^2 - x^2 = s^2 w'^2 + \big(s\,s'\,w^2\big)'.$$

  Al integrar, el último término es una derivada exacta que se anula en los dos extremos, así que $J(x) = \int_0^T s^2 w'^2\,dt \geq 0 = J(0)$, con igualdad solo si w es constante, es decir, nula.

Hay que sacar dos lecturas distintas. La convexidad es *suficiente*, no necesaria: con $T < \pi$ el cero es el mínimo aunque f no sea convexa, y para verlo hace falta otro argumento (aquí, completar un cuadrado con una s que no se anula). Y sin ella, la condición de Euler-Lagrange puede señalar un candidato que no minimiza nada. Es la acción del muelle de la sección 3.3.1 con $m = k = 2$, es decir, con $\omega = 1$: allí reaparece el mismo umbral, $\tau > \pi/\omega$, visto desde la Mecánica.

Ejercicio 3.1. *Los dos extremos libres, y el recuento de condiciones*

ORIGEN: ejercicio propuesto en el curso tras la condición de Neumann.

Obtener el problema de contorno que debe verificar una solución del problema de mínimos en el que no se impone ninguna restricción ni en a ni en b. Comprobar después que, para $f = \frac{1}{2}x'^2 + \frac{1}{2}x^2$ con $N = 1$, el problema resultante tiene solución única, y calcularla.

PISTAS, DE MENOR A MAYOR

1. *Repite el paso 2 de la demostración con h arbitraria, sin ninguna condición en los extremos; salen dos términos de frontera.*
2. *Primero h con $h(a) = h(b) = 0$ para tener la EDO; después h con $h(a) = 1$, $h(b) = 0$, y al revés.*
3. *Con la f dada, (EL) es $x'' = x$ y las condiciones naturales son $x'(a) = x'(b) = 0$.*

Resolución.
La idea, y es una regla de contabilidad. Un problema de contorno regular de segundo orden lleva dos condiciones por incógnita, aunque eso no asegura que tenga solución única. Aquí el problema variacional las reparte de la siguiente manera, y conviene decirlo como un principio antes de hacer la cuenta: **cada extremo aporta exactamente una condición por incógnita; o la pone el enunciado, o la pone el propio problema.** Nunca las dos, nunca ninguna.
Por qué es así se ve en la integración por partes. El término de frontera que aparece es $\nabla_{x'} f \cdot h$ evaluado en los dos extremos. Si el enunciado fija $x(a)$, entonces $h(a) = 0$ —no se puede perturbar lo que está clavado— y el término de frontera muere solo: la condición en a ya la puso el enunciado. Si no lo fija, $h(a)$ es libre, y para que ese término se anule para *toda* h no queda más remedio que $\nabla_{x'} f(a, \dots) = 0$: la condición la pone el problema, y se llama *natural* justamente por eso. Aquí no hay ninguna impuesta, luego las $2N$ son naturales.
Paso 1 (el núcleo): perturbar sin restricciones e integrar por partes. Con $h \in C^1([a,b])^N$ arbitraria, $x + \sigma h$ es admisible para σ pequeño, y $F'(0) = 0$ da (3.5) para toda h. Integrando por partes,

$$\nabla_{x'} f(b, x(b), x'(b)) \cdot h(b) - \nabla_{x'} f(a, x(a), x'(a)) \cdot h(a) + \int_a^b \left(\nabla_x f - \frac{d}{dt} \nabla_{x'} f \right) \cdot h \, dt = 0.$$

Paso 2: usar la arbitrariedad de h dos veces, primero dentro y luego en el borde. Con $h(a) = h(b) = 0$, el lema da la EDO. Sabido eso, la integral es siempre cero y queda $\nabla_{x'} f(b, \dots) \cdot h(b) = \nabla_{x'} f(a, \dots) \cdot h(a)$ para toda h; con $h(b) = 0$ y $h(a)$ arbitrario sale $\nabla_{x'} f(a, x(a), x'(a)) = 0$, y con los papeles cambiados $\nabla_{x'} f(b, x(b), x'(b)) = 0$. El problema de contorno es la EDO con **condiciones naturales en ambos extremos**. Para un sistema regular de segundo orden, el recuento es N ecuaciones y $2N$ condiciones, todas naturales.
Paso 3: el ejemplo concreto, que además comprueba la unicidad por otra vía. Para $f = \frac{1}{2}x'^2 + \frac{1}{2}x^2$:

$\partial_{x'} f = x'$, $\partial_x f = x$, y (EL) es $-x'' + x = 0$, es decir, $x'' = x$, con $x'(a) = x'(b) = 0$. La solución general es $x = c_1 e^t + c_2 e^{-t}$ (teorema de las EDO lineales de coeficientes constantes del apéndice, raíces ± 1), y $x' = c_1 e^t - c_2 e^{-t}$. Las condiciones $x'(a) = x'(b) = 0$ dan $c_1 e^a = c_2 e^{-a}$ y $c_1 e^b = c_2 e^{-b}$; dividiendo, $e^{b-a} = e^{a-b}$, imposible si $a < b$ salvo que $c_1 = c_2 = 0$. La única solución es $x \equiv 0$, y como f es convexa (suma de cuadrados), es el mínimo global: el funcional $\int \frac{1}{2}(x'^2 + x^2)$ sin ninguna restricción se minimiza, como era de esperar, con la función nula.

Dónde se falla.

- *En el signo del término de la izquierda.* La integración por partes deja $\nabla_{x'} f \cdot h$ evaluado en b *menos* evaluado en a. Aquí no importa, porque cada término se iguala a cero por separado; pero en cuanto el funcional lleva un término propio en el borde (un muelle, un coste final), ese signo decide la condición. En la cuerda con muelles de la sección 3.4 sale $ku(0) - \mu u'(0) = 0$ en la izquierda y $ku(L) + \mu u'(L) = 0$ en la derecha.
- *Al confundir $\nabla_{x'} f = 0$ con $x' = 0$.* Coinciden en este ejemplo porque $\partial_{x'} f = x'$. Con $f = \frac{1}{2}(x' - 1)^2 + \frac{1}{2}x^2$ la condición natural es $x'(a) = x'(b) = 1$, y quien escriba $x' = 0$ resuelve otro problema.
- *Al dar por hecha la unicidad.* Si se quita el término $\frac{1}{2}x^2$, con $f = \frac{1}{2}x'^2$, sale $x'' = 0$ con $x'(a) = x'(b) = 0$, y toda constante es solución: todas dan $J = 0$ y todas son mínimos. El recuento de condiciones es el mismo; lo que se ha perdido es que el problema homogéneo solo tenga la solución nula. Es la viga libre del ejercicio de la viga elástica, donde sumar una constante tampoco cambia nada.

QUÉ DEJA. Cada extremo libre aporta una condición natural $\nabla_{x'} f = 0$; en un sistema regular de segundo orden, las condiciones impuestas y naturales suman $2N$. Y el ejemplo enseña la unicidad por otra vía: cuando la EDO es lineal y el problema de contorno homogéneo solo tiene la solución nula, el mínimo es único.

Qué se ha ganado en esta sección. La pregunta de Bernoulli pedía un método único para encontrar la función que minimiza una cantidad que depende de toda la función, y la respuesta tiene dos mitades. La necesaria: todo mínimo de clase C^1 cumple la ecuación de Euler-Lagrange (el lema de Du Bois-Reymond evita pedirle más regularidad), con las condiciones de contorno del enunciado y, donde el enunciado calla, las naturales. La suficiente: si el integrando es convexo en (x, x') (y $\mathcal{O}$ es convexo), toda solución de ese problema de contorno es el mínimo global. Queda abierto lo que el método no hace por sí solo, que es resolver la ecuación. La sección siguiente lo hace en los problemas geométricos clásicos, y en dos de ellos, la braquistócrona y la superficie mínima, la herramienta es la identidad de Beltrami, que baja un orden cuando f no depende de la variable independiente.

3.2 Algunos problemas geométricos

LA PREGUNTA

Un ingeniero de caminos tiene que trazar una vía entre dos pueblos sobre un terreno con colinas, y un arquitecto tiene que decidir la forma de una pompa de jabón sostenida por dos anillos. Los dos quieren lo mínimo: longitud, área. ¿Se pueden resolver los dos problemas con la misma ecuación, y qué forma sale en cada caso?

En esta sección se aplican los resultados anteriores a varios problemas de origen geométrico. Se deducen sus problemas de contorno y se resuelven los casos accesibles a mano; en cada caso se distingue obtener una candidata de justificar que es un mínimo.

Ejemplo (distancia mínima entre dos puntos del plano)**.** Dados $A = (x_1, y_1)$ y $B = (x_2, y_2)$ en $\mathbb{R}^2$, es bien sabido que el camino de menor longitud que los une es el segmento AB. Es fácil de probar con técnicas elementales, pero veamos una prueba basada en el teorema de Euler–Lagrange y el criterio de convexidad, en el caso particular en que solo se consideran caminos dados por la gráfica de funciones $y = u(x)$ o $x = v(y)$. Suponemos $A \neq B$; si la recta que los une es vertical se intercambian los ejes. Ordenando los extremos queda $x_1 < x_2$, y el problema es

$$\begin{cases} \text{Minimizar } \displaystyle\int_{x_1}^{x_2} \sqrt{1 + u'(x)^2}\, dx \\ \text{Sujeto a: } u \in C^1([x_1, x_2]),\ u(x_1) = y_1,\ u(x_2) = y_2. \end{cases}$$

Aquí $f(x, u, u') = \sqrt{1 + u'^2}$, con $\partial_u f = 0$ y $\partial_{u'} f = u' / \sqrt{1 + u'^2}$. Si u es solución, la condición de Euler-Lagrange es

$$\frac{d}{dx}\Big(\frac{u'(x)}{\sqrt{1 + u'(x)^2}}\Big) = 0 \text{ en } (x_1, x_2), \qquad u(x_1) = y_1,\ u(x_2) = y_2.$$

En particular, para alguna constante C, $u' / \sqrt{1 + u'^2} = C$ en $[x_1, x_2]$. La función $s \mapsto s/\sqrt{1 + s^2}$ es estrictamente creciente (su derivada es $(1 + s^2)^{-3/2} > 0$), luego inyectiva, y $u'(x) = m$ constante para todo x. Por tanto la única solución del problema de contorno es la recta

$$u(x) = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \qquad \text{para } x \in [x_1, x_2].$$

Gracias a que f es convexa (es convexa en u', porque $\sqrt{1 + s^2}$ tiene segunda derivada $(1 + s^2)^{-3/2} > 0$, y no depende de u), la condición de Euler-Lagrange no solo es necesaria sino suficiente, y u es la única solución. La gráfica de u es el segmento AB. Un razonamiento análogo se aplica a caminos $x = v(y)$.

LA IDEA: POR QUÉ SALE UNA RECTA SIN SABER QUE SALE UNA RECTA

La ecuación dice que la *pendiente* es constante, porque $\partial_{u'} f$ es una función inyectiva de la pendiente y su derivada es cero. El mecanismo es general: cuando f no depende de u, la cantidad $\partial_{u'} f$ se conserva a lo largo de la solución. En mecánica esa cantidad es el momento, y su conservación es lo que en un problema sin fuerzas da movimiento rectilíneo uniforme: es el mismo teorema.

Ejercicio 3.2. *Geodésicas sobre una superficie dada como gráfica*

ORIGEN: ejemplo 3.8 del curso, que deja como ejercicio el problema de contorno.

Dada la superficie $S = \{(x,y,z) : z = F(x,y)\}$, con $F \in C^2(\mathbb{R}^2)$, y dos puntos $A = (a,\alpha,F(a,\alpha))$ y $B = (b,\beta,F(b,\beta))$ sobre ella, con $a < b$, buscamos una curva de longitud mínima que los una. Toda minimizante regular es una **geodésica**, aunque una geodésica no tiene por qué minimizar entre extremos lejanos. Restringimos aquí la búsqueda a curvas que se puedan escribir como $y = u(x)$, $z = F(x,u(x))$, con $x \in [a,b]$, $u(a) = \alpha$, $u(b) = \beta$; el problema es

$$\text{Minimizar } \int_a^b \sqrt{1 + u'(x)^2 + \big(F_x(x,u(x)) + F_y(x,u(x))u'(x)\big)^2}\,dx$$

sobre las $u \in C^1([a,b])$ con $u(a) = \alpha$, $u(b) = \beta$. Determinar el problema de contorno que deben verificar las posibles soluciones, y particularizarlo al plano inclinado $F(x,y) = cy$.

PISTAS, DE MENOR A MAYOR

1. *El integrando es $f(x,u,u') = \sqrt{1+u'^2+(F_x+F_yu')^2}$ con F_x, F_y evaluadas en (x,u); llama Q a la raíz y deriva con la regla de la cadena.*
2. *$\partial_{u'}f = (u' + (F_x + F_yu')F_y)/Q$; $\partial_u f = (F_x + F_yu')(F_{xy} + F_{yy}u')/Q$.*
3. *En el plano $F = cy$, $F_x = 0$, $F_y = c$, y las derivadas segundas son cero.*

Resolución.

La idea. Detrás de la maraña de derivadas parciales hay una sola frase: *la longitud de una curva sobre la superficie es la integral del módulo de su vector tangente, y la ecuación de Euler-Lagrange de esa integral es la ecuación de las geodésicas*. Todo lo que parece complicado viene de que la curva vive en el espacio y se está describiendo con una sola función, u; el vector tangente tiene tres componentes, la tercera es $F_x + F_yu'$ por la regla de la cadena, y la raíz cuadrada es su módulo. Ni más ni menos.

La imagen que predice el resultado, y que conviene tener antes de derivar nada: **una goma elástica tensada sobre la superficie busca una curva de longitud mínima**. Sobre un plano inclinado la goma se queda recta —una rotación lleva ese plano a uno horizontal sin cambiar las longitudes—, y por eso el caso particular del enunciado tiene que dar una recta. Si no diera una recta, habría un error en la cuenta.

Y hay un atajo estructural que hay que aprender a ver, porque es el que resuelve el caso particular sin resolver ninguna ecuación diferencial: cuando F es un plano, el integrando **no depende de** u, solo de u'. Entonces $\partial_u f = 0$, la ecuación se reduce a $\frac{d}{dx}\partial_{u'}f = 0$, y lo que dice es que $\partial_{u'}f$ se conserva. Es la misma conservación que en mecánica da el momento cuando el lagrangiano no ve una coordenada.

Paso 1: de dónde sale el integrando. La curva es $\gamma(x) = (x, u(x), F(x,u(x)))$, luego $\gamma' = (1, u', F_x + F_yu')$ y su longitud es $\int_a^b |\gamma'|\,dx$. Sea $Q = \sqrt{1+u'^2+(F_x+F_yu')^2}$. Entonces

$$\partial_{u'}f = \frac{u' + (F_x + F_yu')F_y}{Q}, \qquad \partial_u f = \frac{(F_x + F_yu')(F_{xy} + F_{yy}u')}{Q},$$

Paso 2: la ecuación general, que no se resuelve a mano y no pasa nada. El problema de contorno de

Euler-Lagrange es

$$-\frac{d}{dx}\Big[\frac{u' + (F_x + F_y u')F_y}{Q}\Big] + \frac{(F_x + F_y u')(F_{xy} + F_{yy} u')}{Q} = 0, \qquad u(a) = \alpha,\ u(b) = \beta,$$

con todas las derivadas de F evaluadas en $(x, u(x))$. Es una EDO de segundo orden no lineal, que en general no se resuelve a mano. `bvp4c` permite buscar aproximaciones numéricas; la existencia de solución y la convergencia del cálculo requieren su propia comprobación.
Paso 3 (el núcleo): el caso particular, resuelto por la conservación y no por la EDO. En el plano inclinado $F = cy$: $F_x = 0$, $F_y = c$, derivadas segundas nulas, luego $\partial_u f = 0$ y $\partial_{u'} f = \frac{u'(1+c^2)}{\sqrt{1+(1+c^2)u'^2}}$, que es función inyectiva de u'. La ecuación dice que $\partial_{u'} f$ es constante, luego u' es constante: la geodésica es una recta en el plano (x, y), y por tanto una recta en el plano inclinado. Es lo que debía salir, porque el plano inclinado es isométrico al plano.
Dónde se falla.

- *Al derivar respecto de u.* F_x y F_y se evalúan en $(x, u(x))$, así que también dependen de u, y $\partial_u f$ lleva las derivadas segundas F_{xy} y F_{yy}. Quien las trate como constantes obtiene $\partial_u f = 0$ en cualquier superficie, y con ello la conservación de $\partial_{u'} f$, que solo vale cuando el integrando no depende de u. En el plano inclinado el error no se ve, porque allí esas derivadas segundas son de verdad nulas; por eso el caso particular no basta para comprobar la ecuación general.
- *Al olvidar qué curvas se han dejado fuera.* Escribir la curva como $y = u(x)$ exige que avance siempre en x. Si los dos puntos tienen la misma abscisa, o si el camino corto tiene que retroceder en x para rodear una colina, la formulación no lo contiene; entonces se parametriza por y o por la longitud de arco.
- *Al llamar «el camino más corto» a cualquier geodésica.* Sobre una esfera de radio R, dos puntos no antipodales están unidos por dos arcos del mismo círculo máximo, y los dos son geodésicas; el largo mide más de πR y no minimiza. La ecuación de Euler-Lagrange es una condición necesaria y no distingue entre ellos, igual que $\nabla F = 0$ en $\mathbb{R}^n$ no distingue un mínimo de un punto de silla.

CONEXIÓN: LAS GEODÉSICAS DE GEOMETRÍA LOCAL

Una geodésica tiene curvatura geodésica nula: su aceleración, al parametrizarla por longitud de arco, es normal a la superficie. La ecuación obtenida expresa esa misma condición usando x como parámetro. Al pasar a longitud de arco toma la forma habitual de la ecuación de geodésicas con símbolos de Christoffel; no se debe suponer que x es ya un parámetro afín.
El cálculo anterior prueba la necesidad. Para una superficie general, la convexidad del integrando no da la suficiencia. La geometría aporta otro argumento, el lema de Gauss, que permite demostrar que los tramos suficientemente cortos de una geodésica minimizan la longitud entre sus extremos. Esto no asegura que una geodésica entera sea el camino más corto.

Ejemplo (la braquistócrona)**.** La **braquistócrona** (del griego *brachistos*, el más corto, y *cronos*, tiempo), o curva de descenso más rápido, es la curva que une dos puntos y es recorrida en tiempo mínimo por un móvil que se desliza sin rozamiento desde el más alto, del que parte con velocidad cero, hasta el más bajo, bajo la acción de la gravedad. Es el problema de la pregunta que abre la sección 3.1, y en su estudio tiene su origen el Cálculo de Variaciones.

Se toma un sistema de coordenadas en el plano vertical que contiene a los dos puntos, con origen en A y con el eje de ordenadas vertical *apuntando hacia abajo*, de modo que $B = (\alpha, \beta)$ con $\alpha, \beta > 0$. El problema es

$$\begin{cases} \text{Minimizar } \dfrac{1}{\sqrt{2g}} \displaystyle\int_0^\alpha \sqrt{\frac{1+u'(x)^2}{u(x)}}\, dx \\ \text{Sujeto a: } u \in C([0,\alpha]) \cap C^1((0,\alpha]), \\ u(0) = 0,\ u(\alpha) = \beta,\ u(x) > 0 \text{ si } x > 0,\ T(u) < \infty. \end{cases}$$

con g la aceleración de la gravedad y la integral entendida como impropia en $x = 0$. Se permite una tangente vertical inicial: exigir C^1 en todo el intervalo cerrado excluiría la propia cicloide que buscamos. La explicación: la longitud de un trozo de curva es $ds = \sqrt{1+u'^2}\, dx$, y el tiempo en recorrerlo es ds/v, donde $v(x)$ es la velocidad del móvil en $(x, u(x))$; luego

$$T(u) = \int_0^{T(u)} dt = \int_0^\alpha \frac{\sqrt{1+u'(x)^2}}{v(x)}\, dx.$$

Como el eje vertical apunta hacia abajo, la energía potencial es $-mgu$: disminuye cuando u crece. La conservación de la energía da

$$-mgu(x) + \tfrac{1}{2}mv(x)^2 = 0,$$

pues el móvil parte del origen en reposo. Por tanto $v(x) = \sqrt{2gu(x)}$, y se obtiene el funcional anterior.

Las hipótesis, una a una. El funcional lleva dentro cinco decisiones. Que el móvil sea una **masa puntual**: si una esfera maciza pequeña rueda sin deslizar y sin pérdida de energía, parte de ella se emplea en girar y $v^2 = \frac{10}{7}gu$ en lugar de $2gu$. En ese modelo de rodadura ideal cambia la constante que multiplica el tiempo, pero no la curva óptima del centro de masas. La fricción estática que permite rodar no disipa energía. Que en el modelo de deslizamiento **no haya rozamiento**: la conservación de la energía permite escribir v en función de la altura sola. Con disipación, en general v depende de todo el camino recorrido y ya no basta este funcional de u y u'. Que **parta del reposo**: si saliera con velocidad v_0, la constante de la energía sería $\frac{1}{2}mv_0^2$ y el integrando $\sqrt{(1+u'^2)/(u+v_0^2/2g)}$, que es el mismo problema con el origen desplazado hacia arriba. Que la curva sea una **gráfica** $y = u(x)$: se excluyen las curvas que retroceden en x, lo que no quita nada porque la cicloide que sale no retrocede (con α/β grande baja por debajo de B y vuelve a subir, pero siempre avanzando en x). Y que todo ocurra en un **plano vertical**: salirse del plano alarga el camino sin ganar velocidad. Son las hipótesis de este modelo ideal; para una pista real habría que valorar además el rozamiento y las restricciones sobre su forma.

Es un ejemplo singular, en condiciones que escapan del marco de los teoremas de Euler–Lagrange y de convexidad, porque f no es continua en $u = 0$, que es justo el valor en $x = 0$. En cualquier subintervalo cerrado de $(0, \alpha)$ sí se pueden hacer variaciones de soporte interior: allí se obtiene Euler–Lagrange y, para una candidata C^2 en el interior, la identidad de Beltrami. Como $f = \sqrt{(1+u'^2)/u}$ no depende de x, Beltrami dice que $f - u'\,\partial_{u'} f$ es constante. Se calcula por partes, sin prisa. Escribiendo $f = (1+u'^2)^{1/2}u^{-1/2}$ (el factor $1/\sqrt{2g}$ se deja aparte, porque multiplica a todo), la derivada respecto de u' es

$$\partial_{u'} f = \frac{1}{2}(1+u'^2)^{-1/2} \cdot 2u' \cdot u^{-1/2} = \frac{u'}{\sqrt{u(1+u'^2)}},$$

luego $u'\,\partial_{u'} f = \dfrac{u'^2}{\sqrt{u(1+u'^2)}}$. Para restar f y esto se escribe f con el mismo denominador: multiplicando numerador y denominador de $\sqrt{(1+u'^2)/u}$ por $\sqrt{1+u'^2}$ queda $f = \dfrac{1+u'^2}{\sqrt{u(1+u'^2)}}$. Entonces

$$f - u'\,\partial_{u'} f = \frac{(1+u'^2) - u'^2}{\sqrt{u(1+u'^2)}} = \frac{1}{\sqrt{u(1+u'^2)}},$$

y Beltrami, con el $1/\sqrt{2g}$ absorbido en la constante, es

$$\frac{1}{\sqrt{u(x)\big(1+u'(x)^2\big)}} = C > 0 \quad \text{si } 0 < x \le \alpha, \qquad u(0) = 0,\ u(\alpha) = \beta.$$

Es decir, $u(1+u'^2) = 1/C^2 =: 2r$ es constante a lo largo de la curva.

LA IDEA: QUÉ DICE $u(1+u'^2) = 2r$

Es una relación entre la profundidad y la pendiente en cada punto: $1 + u'^2 = 2r/u$, luego *cuanto más abajo está la curva, menor es el valor absoluto de su pendiente*. Al acercarse al punto de salida, $u \to 0$ y $u' \to +\infty$: la braquistócrona arranca **vertical**, para ganar velocidad cuanto antes (es lo que muestra la curva verde de la figura). Y en el punto más bajo, $u = 2r$, la pendiente es cero. Todo el comportamiento cualitativo está en esa ecuación de primer orden, antes de resolverla; la cicloide que obtendremos realiza ese comportamiento. Con números: para llegar de $(0,0)$ a $(1, 0.5)$ sale $r \approx 0.26$, así que la curva baja hasta $2r \approx 0.52$, un poco por debajo del punto de llegada, y vuelve a subir.

La candidata es una **cicloide** que une A con B:

$$\begin{cases} x = r(\theta - \sin\theta), \\ y = r(1 - \cos\theta), \end{cases} \qquad \theta \in [0, \theta_0],$$

donde θ_0 es solución de

$$\frac{\alpha}{\beta} = \frac{\theta_0 - \sin\theta_0}{1 - \cos\theta_0} \qquad \text{y} \qquad r = \frac{\beta}{1 - \cos\theta_0}.$$

La idea. La ecuación $u(1+u'^2) = 2r$ liga la altura con la pendiente, y en esa forma no se integra cómodamente. Pero dice algo geométrico: cuanto más ha bajado la curva, menos inclinada está. Eso es lo que hace una rueda que gira: el punto marcado del borde cae verticalmente al empezar y sale horizontalmente en el fondo. El truco de la demostración es *dejar de usar la altura como variable* y usar en su lugar el ángulo girado por esa rueda. Con el cambio adecuado, la relación entre altura y pendiente se convierte en una identidad trigonométrica, y lo que quedaba por integrar se integra solo.

Cómo se obtiene la cicloide a partir de $u(1+u'^2) = 2r$. Integramos la rama descendente y comprobamos después la continuación.
Paso 1: despejar la pendiente. $u' = \sqrt{(2r-u)/u}$ (se toma la raíz positiva en el tramo en que la curva baja).
Paso 2: el cambio de variable que convierte la raíz en un cociente de senos (el núcleo).

El cambio $u = r(1-\cos\theta) = 2r\sin^2(\theta/2)$ da $2r-u = 2r\cos^2(\theta/2)$, luego $u' = \cot(\theta/2)$, y $du = r\sin\theta\,d\theta$. El cambio está elegido para eso: la identidad $\sin^2+\cos^2=1$ es exactamente la forma que tiene la ecuación, y por eso la raíz cuadrada desaparece sin dejar rastro.

Paso 3: integrar el desplazamiento horizontal. Entonces

$$\begin{aligned} dx = \frac{du}{u'} &= \frac{r\sin\theta\,d\theta}{\cot(\theta/2)} = r\cdot 2\sin(\theta/2)\cos(\theta/2)\cdot\frac{\sin(\theta/2)}{\cos(\theta/2)}\,d\theta \\ &= 2r\sin^2(\theta/2)\,d\theta = r(1-\cos\theta)\,d\theta, \end{aligned}$$

e integrando con $x=0$ en $\theta=0$, $x = r(\theta-\sin\theta)$. Es la curva que describe un punto de una circunferencia de radio r que rueda sin deslizar por debajo del eje x. Las dos condiciones sobre θ_0 y r son imponer que la cicloide pase por (α,β): dividiendo $x(\theta_0)=\alpha$ entre $y(\theta_0)=\beta$ desaparece r y queda la ecuación en θ_0, que tiene solución única en $(0,2\pi)$ porque el cociente $(\theta-\sin\theta)/(1-\cos\theta)$ crece de 0 a $+\infty$.

Y este último detalle es más que un cierre técnico: dice que **por cada par de puntos pasa exactamente una cicloide de este tipo**, de modo que la condición de contorno no elige entre varias candidatas sino que determina una sola. Es lo que convierte la ecuación diferencial en un problema con respuesta. ■

Qué hace cada hipótesis, y qué se gana con el cambio de variable. La única hipótesis estructural que permite esta reducción es que f no depende de x: es la que permite aplicar Beltrami y quedarse con una ecuación de *primer* orden, $u(1+u'^2)=2r$, en lugar de la de segundo orden que da Euler-Lagrange. Sin ella no habría nada que integrar aquí. Dentro de la demostración, en cambio, hay una elección que conviene no confundir con una hipótesis: la raíz positiva del paso 1. Elegirla equivale a mirar solo el tramo en que la curva baja, y podría parecer que la solución obtenida vale solo ahí; pero no es así, y la razón es el cambio de variable. Escrita en θ, la pendiente es $u' = \cot(\theta/2)$, que es positiva para $\theta<\pi$ y **negativa** para $\theta>\pi$: la parametrización cruza el fondo $u=2r$ y sigue por el tramo ascendente sin que haya que hacer nada. Eso es lo que permite que θ_0 pueda pasar de π, como ocurre en la figura 3.1 ($\theta_0\approx 3.51$): la curva más rápida baja por debajo del punto de llegada y vuelve a subir para alcanzarlo. Al sustituirla se comprueba, también en el fondo, la ecuación completa de Euler–Lagrange,

$$2uu'' + 1 + u'^2 = 0 \qquad (0<x<\alpha).$$

Esta comprobación importa: la ecuación de Beltrami sola también admite tramos constantes $u=2r$, que no satisfacen Euler–Lagrange.

En la salida, $u'(x)\to+\infty$: la gráfica no es derivable allí. La parametrización por θ es suave, pero su derivada se anula en $\theta=0$, así que tampoco es regular en ese extremo. La longitud de arco sí proporciona una parametrización regular desde la salida:

$$s = 4r\big(1-\cos(\theta/2)\big), \qquad \frac{d(x,u)}{ds} = \big(\sin(\theta/2),\cos(\theta/2)\big) \longrightarrow (0,1).$$

La tangente inicial es vertical. El cambio de variable permite describirla sin exigir una pendiente finita de la gráfica.

CONEXIÓN: POR QUÉ LA CANDIDATA ES REALMENTE EL MÍNIMO

La condición necesaria no bastaría. Aquí se puede cerrar el argumento con el criterio de convexidad anterior, tras cambiar de variable: pongamos $z = \sqrt{u}$. Entonces

$$\sqrt{\frac{1+u'^2}{u}} = G(z,z'), \qquad G(z,w) = \sqrt{z^{-2}+4w^2} = \big\|(z^{-1},2w)\big\|.$$

Para $z > 0$, esta función es estrictamente convexa en (z,w). Se ve sin una hessiana: $1/z$ es estrictamente convexa, la norma crece con su primera componente positiva y la desigualdad triangular controla las combinaciones convexas. Si los valores de z coinciden pero los de w son distintos, la desigualdad triangular es estricta.
Sea z_* la cicloide en esta variable. Cumple Euler–Lagrange en el interior. La prueba del plano tangente, integrada entre ε y α, da para cualquier competidora z

$$\int_\varepsilon^\alpha \big(G(z,z') - G(z_*,z_*')\big)\,dx \geq \big[\partial_w G(z_*,z_*')(z-z_*)\big]_\varepsilon^\alpha.$$

El término en α es cero. El de ε tiende a cero porque $z(0) = z_*(0) = 0$ y $|\partial_w G| \leq 2$. Pasando a la integral impropia se obtiene $T(u) \geq T(u_*)$. La convexidad estricta asegura la unicidad. Así se justifican tanto el mínimo como el tratamiento del extremo singular, dentro de la familia de gráficas admisibles.

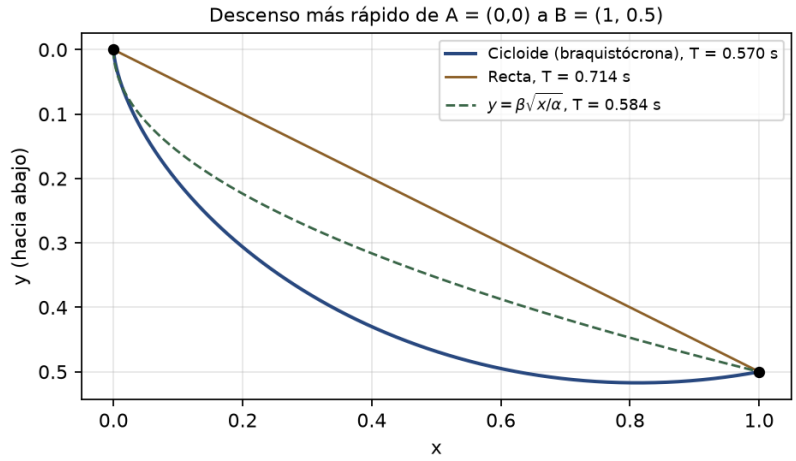


Figura 3.1: De $A = (0,0)$ a $B = (1,0.5)$, con el eje y hacia abajo. La cicloide, con $\theta_0 \approx 3.508$ y $r \approx 0.2586$, tarda $\sqrt{r/g}\,\theta_0 = 0.570$ s; la recta, $\sqrt{2L/(g\sin\vartheta)} = 0.714$ s, un 25 % más. La curva verde, que arranca vertical como la cicloide, tarda 0.584 s: solo un 2.4 % más. Esta comparación muestra la ventaja de ganar velocidad pronto; compartir la tangente inicial no basta para alcanzar el mínimo. Con $\alpha/\beta = 2$, $\theta_0 > \pi$: la cicloide baja por debajo de B y vuelve a subir para llegar a él. (Los tres tiempos de la leyenda son los mismos que los de este pie: se calculan con la fórmula exacta $\sqrt{r/g}\,\theta_0$ para la cicloide y con cuadratura adaptada para las otras dos, porque el integrando $\sqrt{1+y'^2}/\sqrt{y}$ se hace infinito en el origen y una suma sobre los puntos dibujados los subestimaría.)

EL CAMINO QUE NO SE TOMÓ: LA CIRCUNFERENCIA DE GALILEO

La alternativa. Galileo, en 1638, afirmó que el arco de circunferencia era la curva de descenso más rápido, tras probar que era más rápido que la recta.
Su ventaja. Es cierto que vence a la recta, y en la época no había método para comparar con *todas* las curvas: solo se podían comparar unas pocas familias.
Qué la hundió. El método de Bernoulli y de Newton en 1697, y después el de Euler y Lagrange: en vez de proponer curvas, se deduce la ecuación que debe cumplir la mejor. La cicloide mejora a los arcos de circunferencia, y la condición de Euler–Lagrange permite comprobar que estos no son la curva óptima. Es el ejemplo fundacional de por qué hace falta una condición necesaria: para no depender de la intuición sobre qué curvas probar.

DE DÓNDE VIENE: LA SOLUCIÓN DE BERNOULLI POR LA ÓPTICA

Johann Bernoulli resolvió su propio problema con una analogía: por el principio de Fermat, la luz sigue el camino de tiempo mínimo, y en un medio cuyo índice de refracción varía con la altura la ley de Snell dice que $\sin\theta / v$ es constante a lo largo del rayo, con θ el ángulo con la vertical. Con $v = \sqrt{2gy}$, $\sin\theta = 1/\sqrt{1+u'^2}$, eso es exactamente $u(1+u'^2) = \text{const}$, la ecuación de arriba. Según una anécdota atribuida a la sobrina de Newton, este recibió el problema a las cuatro de la tarde y lo mandó resuelto a la mañana siguiente, sin firmar; Bernoulli reconoció «al león por sus garras». Y Jakob Bernoulli, el hermano, dio la solución más cercana al método general, que es la que Euler desarrolló.

Ejercicio 3.3. *Superficie mínima de revolución*

ORIGEN: ejemplo 3.10 del curso, que deja como ejercicio el problema de contorno.

El problema de la superficie mínima de un sólido de revolución sobre el eje OX generado por una curva $y = u(x)$ que une $A = (a, \alpha)$ y $B = (b, \beta)$, con $a < b$ y $\alpha, \beta > 0$, es

$$\text{Minimizar } 2\pi \int_a^b u(x)\sqrt{1+u'(x)^2}\,dx \quad \text{sobre } u \in C^1([a,b]),\ u(a)=\alpha,\ u(b)=\beta,\ u>0.$$

Hallar el problema de contorno que debe verificar toda posible solución, y resolverlo con la identidad de Beltrami.

PISTAS, DE MENOR A MAYOR

1. *$f(u,u') = u\sqrt{1+u'^2}$ no depende de x: Beltrami.*
2. *$f - u'\partial_{u'} f$ se simplifica mucho; llama C a la constante y despeja u'.*
3. *La EDO $u' = \sqrt{u^2/C^2 - 1}$ se integra con el cambio $u = C\cosh s$.*

Resolución.
La idea, y cabe en dos frases. El integrando $u\sqrt{1+u'^2}$ **no depende de** x. Siempre que eso pasa hay una cantidad conservada —la identidad de Beltrami— y una ecuación de segundo orden se derrumba en una de primer orden, que sí se integra. Ese es el único movimiento del ejercicio; el resto es álgebra.
La razón de fondo merece decirse porque reaparece en todo el capítulo y en el siguiente: que f no dependa de x significa que el problema es *el mismo en todas partes del eje*, y a esa invariancia le corresponde una ley de conservación. Es exactamente el mismo fenómeno que en mecánica hace que la energía se conserve cuando el lagrangiano no depende del tiempo,

y la fórmula $f - u'\partial_{u'} f$ corresponde allí a la energía con el signo cambiado. En el capítulo 4 volverá con otra cara: cuando el hamiltoniano no depende del tiempo, se conserva.
Y la imagen que orienta el resultado: una película de jabón entre dos aros iguales y próximos forma un cuello más estrecho en el centro. La cuenta determinará su perfil, una catenaria, la misma familia de curvas que adopta una cadena colgada. Los problemas son físicamente distintos: el área de revolución integra $2\pi u\, ds$, mientras que la cadena minimiza energía gravitatoria con longitud fija. El multiplicador de esa restricción desplaza la altura por una constante, y explica por qué reaparece la familia de catenarias.
Paso 1: de dónde sale el funcional, y qué se supone. Una película de jabón tiene una energía proporcional a su área (la tensión superficial σ es energía por unidad de área), y en reposo elige la forma de energía mínima: minimizar el área *es* el modelo físico, y σ sale fuera del mínimo. El área de la superficie de revolución es $\int 2\pi u\, ds = 2\pi \int_a^b u\sqrt{1+u'^2}\, dx$: cada trozo de curva de longitud ds barre una banda de radio u. Las hipótesis: que la superficie sea **de revolución** (los dos anillos son coaxiales y se restringe la búsqueda a películas que respeten esa simetría; las superficies mínimas generales requieren una EDP), que la película **no pesa** (sin gravedad, que la deformaría hacia abajo) y que **no hay diferencia de presión** entre los dos lados (una pompa cerrada, con aire dentro, minimiza área a volumen fijo y da una esfera). Con eso el problema es de una sola función de una variable.
Paso 2: Euler-Lagrange, que da la ecuación pero no la resuelve. $\partial_u f = \sqrt{1+u'^2}$, $\partial_{u'} f = uu'/\sqrt{1+u'^2}$, y el problema de contorno es

$$-\frac{d}{dx}\Big(\frac{u\,u'}{\sqrt{1+u'^2}}\Big) + \sqrt{1+u'^2} = 0, \qquad u(a) = \alpha,\ u(b) = \beta.$$

Paso 3 (el núcleo): Beltrami baja el orden y la ecuación se integra. Como f no depende de x,

$$f - u'\partial_{u'} f = u\sqrt{1+u'^2} - \frac{uu'^2}{\sqrt{1+u'^2}} = \frac{u(1+u'^2) - uu'^2}{\sqrt{1+u'^2}} = \frac{u}{\sqrt{1+u'^2}} = C,$$

con $C > 0$ porque $u > 0$. Luego $1 + u'^2 = u^2/C^2$. Para integrar sin dividir por u' en el cuello, escribimos $u' = \sinh s$, de donde $u = C\cosh s$. La ecuación completa de Euler–Lagrange se simplifica a $uu'' = 1 + u'^2$, y por tanto

$$s' = \frac{u''}{\sqrt{1+u'^2}} = \frac{\sqrt{1+u'^2}}{u} = \frac{1}{C}.$$

Así, $s = (x - x_0)/C$ y

$$u(x) = C\cosh\frac{x - x_0}{C},$$

una **catenaria**. *Paso 4: qué se ha demostrado y qué no.* Las dos constantes C y x_0 se fijan con $u(a) = \alpha$ y $u(b) = \beta$, y es un sistema que *no siempre tiene solución*: si los dos anillos están demasiado separados para su radio, no hay catenaria que pase por los dos puntos. En ese caso el problema planteado, que exige una gráfica positiva, no tiene mínimo. Los dos discos que cierran los anillos pertenecen a una familia ampliada de superficies desconectadas; no son una solución admisible de este enunciado.
Además $f = u\sqrt{1+u'^2}$ no es convexa en (u, u'): para $u' \neq 0$, su hessiana tiene determinante $-u'^2/(1+u'^2) < 0$. El criterio de convexidad no se aplica. Cuando hay dos catenarias, la

de cuello más ancho es un mínimo local y la otra es inestable; eso todavía no prueba que la primera sea el mínimo global. Por ejemplo, para $a = -0.6$, $b = 0.6$ y $\alpha = \beta = 1$, ambas catenarias tienen área mayor que 2π. Una sucesión de perfiles con un cuello cada vez más fino aproxima los dos discos y su área tiende a 2π, así que ninguna de esas dos candidatas es un mínimo global. La existencia de una solución de Euler–Lagrange y la existencia del mínimo son preguntas distintas.

QUÉ DEJA. Beltrami convierte un problema de segundo orden en uno de primer orden integrable cuando f no depende de x; es la herramienta de todos los problemas geométricos clásicos. Y este ejemplo enseña que el problema de contorno puede no tener solución, o tener dos, y que la condición necesaria no distingue cuál es el mínimo.

3.3 Aplicaciones en Mecánica (I): Principio de mínima acción

LA PREGUNTA

Una piedra lanzada al aire describe una parábola. Newton lo explica con una fuerza que actúa instante a instante. Pero hay otra forma de contarlo, con antecedentes en la mínima acción de Maupertuis y formulada aquí como principio de Hamilton: entre los caminos que unen las mismas posiciones en los mismos instantes, la trayectoria hace *estacionaria* una cierta integral. ¿Cómo puede una piedra «saber» de antemano cuál es ese camino, y por qué las dos descripciones dicen exactamente lo mismo?

Supongamos un sistema de partículas cuyo movimiento viene descrito por las variables x_i, $1 \leq i \leq m$. El **Principio de Mínima Acción**, en su formulación de Hamilton, dice que la trayectoria entre dos posiciones fijadas a y b en instantes $t_0 < t_1$ hace estacionaria la acción

$$S(x) = \int_{t_0}^{t_1} \mathcal{L}\big(t, x(t), x'(t)\big)\, dt.$$

Es decir, el problema consiste en

$$\begin{cases} \dfrac{d}{d\sigma} S(x + \sigma h)\Big|_{\sigma=0} = 0 \quad \text{para toda } h \in C^1([t_0, t_1])^m \text{ con } h(t_0) = h(t_1) = 0, \\ x \in C^1([t_0, t_1])^m, \quad (t, x(t), x'(t)) \in \mathcal{O}\ \forall t, \quad x(t_0) = a,\ x(t_1) = b. \end{cases}$$

La función $\mathcal{L} : \mathcal{O} \subset \mathbb{R}^{2m+1} \to \mathbb{R}$ es la **lagrangiana**; reservamos el nombre «acción» para su integral S. Buscar un mínimo de S conduce a esta condición necesaria. El principio físico impone la condición estacionaria: no afirma que toda trayectoria sea un mínimo. El muelle mostrará la diferencia.

Se considera el caso especialmente importante en que las fuerzas son **conservativas**: solo dependen de la posición y su resultante es el gradiente de una función $\phi = \phi(x)$ cambiado de signo, $F = -\nabla_x \phi(x)$. En este caso las leyes de la Física dicen que $\mathcal{L}$ es independiente de t, $\mathcal{L} = \mathcal{L}(x, x')$, y que se escribe como la diferencia entre la energía cinética E_c y la energía potencial E_p del sistema:

$$\mathcal{L} = E_c - E_p.$$

Qué se está suponiendo al escribir $\mathcal{L} = E_c - E_p$. Cuatro cosas, y conviene nombrarlas porque cada aplicación de esta sección las hereda. Que las fuerzas sean **conservativas**: el trabajo que hacen entre dos posiciones no depende del camino, y por eso existe ϕ. El rozamiento no lo es (su trabajo depende de la longitud del camino), y no queda incluido en la lagrangiana autónoma $E_c - E_p$ de este modelo. Las fuerzas disipativas pueden añadirse como fuerzas generalizadas en las ecuaciones de Lagrange, como se hará en el péndulo con $-c\varphi'$. Esto no significa que sea imposible construir cualquier otra formulación variacional para un sistema disipativo. Que las **masas sean constantes**: un cohete que quema combustible no cumple $\frac{d}{dt}(mx') = mx''$. Que las **velocidades sean pequeñas** frente a la de la luz, para que $E_c = \frac{m}{2}|x'|^2$; la lagrangiana relativista es otra. Y que las **ligaduras sean ideales y fijas**: para las curvas o superficies inmóviles de estos ejemplos, la fuerza de reacción no hace trabajo en los desplazamientos permitidos. Por eso puede eliminarse al escribir $\mathcal{L}$ en coordenadas adaptadas a la ligadura. Lo que se gana a cambio es lo que dice la caja anterior: se puede escribir $\mathcal{L}$ en cualquier coordenada, y las conservaciones salen solas.

El teorema de Euler–Lagrange dice que, si $t \mapsto \nabla_{x'}\mathcal{L}(x(t), x'(t))$ es de clase C^1, necesariamente

$$-\frac{d}{dt}\nabla_{x'}\mathcal{L}(x, x') + \nabla_x\mathcal{L}(x, x') = 0 \qquad \text{en } [t_0, t_1],$$

y por la identidad de Beltrami existe $C \in \mathbb{R}$ tal que

$$\mathcal{L}(x, x') - \nabla_{x'}\mathcal{L}(x, x') \cdot x' = C \qquad \text{en } [t_0, t_1].$$

Esto, junto con informaciones adicionales sobre x (los valores en t_0 y t_1, o la posición y velocidad iniciales), permite con frecuencia determinar x.

EL CAMINO QUE NO SE TOMÓ: LAS LEYES DE NEWTON DIRECTAMENTE

La alternativa. Escribir $F = ma$ para cada partícula, en coordenadas cartesianas, y resolver.
Su ventaja. En coordenadas cartesianas ofrece una vía directa para obtener las aceleraciones a partir de las fuerzas.
Qué la aparta. Tres cosas. Primera, las ligaduras: en el péndulo la partícula está obligada a moverse en una circunferencia, y con Newton hay que introducir la tensión de la varilla como incógnita auxiliar y eliminarla después; con la acción se escribe $\mathcal{L}$ en la única coordenada libre, el ángulo, y la tensión no aparece. Segunda, las coordenadas: la acción es un escalar y se puede escribir en las coordenadas que convengan (ángulos, centro de masas), mientras que $F = ma$ es una ecuación vectorial que hay que transformar. Tercera, las leyes de conservación salen solas: si $\mathcal{L}$ no depende de t se conserva la energía (Beltrami), si no depende de una coordenada se conserva su momento. Por eso la Mecánica teórica se escribe con lagrangianos. Newton sigue siendo lo que se usa para una partícula en cartesianas, y el primer ejemplo enseña que las dos formulaciones coinciden.

3.3.1 Movimiento de una partícula en un campo conservativo

Se buscan las ecuaciones del movimiento de una partícula en un campo conservativo de potencial $\phi \in C^1(D)$, $D \subset \mathbb{R}^3$ abierto: la partícula «vive» en D y sobre ella actúa la fuerza $F(x) = -\nabla_x \phi(x)$. Siendo m la masa y $x = x(t)$ la posición, la **energía cinética** en el instante t es $E_c(t) = \frac{m}{2}|x'(t)|^2$ y la **energía potencial** es $E_p(t) = \phi(x(t))$. La lagrangiana es

$$\mathcal{L}(x, x') = E_c - E_p = \frac{m}{2}|x'|^2 - \phi(x),$$

con $\nabla_{x'}\mathcal{L} = mx'$ y $\nabla_x \mathcal{L} = -\nabla_x \phi$, y la ecuación de Euler-Lagrange queda

$$-\frac{d}{dt}(mx') - \nabla_x \phi(x) = 0, \qquad \text{es decir,} \qquad mx'' = -\nabla_x \phi(x).$$

El Principio de Mínima Acción se reduce a la Segunda Ley de Newton: masa por aceleración igual a fuerza. Y la identidad de Beltrami es

$$\frac{m}{2}|x'|^2 - \phi(x) - mx' \cdot x' = C, \qquad \text{es decir,} \qquad \frac{m}{2}|x'|^2 + \phi(x) = E :$$

la suma de la energía cinética y la potencial permanece constante, el **Principio de Conservación de la Energía**. En el caso general, la ecuación de Euler-Lagrange se interpreta de modo análogo: se llama a $\nabla_{x'}\mathcal{L}$ **impulso generalizado** y a $\nabla_x \mathcal{L}$ **fuerza generalizada**, y la ecuación dice que la derivada del impulso es la fuerza. Por analogía, Beltrami se interpreta como un principio generalizado de conservación de la energía.

LA IDEA: POR QUÉ NO HAY QUE CONOCER EL DESTINO

La piedra no anticipa su destino. La condición de acción estacionaria equivale aquí a $mx'' = F$, una ley local. Fijar dos extremos sirve para comparar trayectorias en el argumento variacional; para calcular un movimiento se pueden dar, en cambio, la posición y la velocidad iniciales e integrar la misma ecuación. La formulación variacional facilita las coordenadas y las leyes de conservación. Que una trayectoria estacionaria sea además un mínimo requiere otra comprobación, como veremos con el muelle.

Caso particular: movimiento en el campo gravitatorio terrestre cerca de la superficie. La partícula está sometida a la gravedad, con aceleración constante $g \approx 9.8$ m/s^2. Con $x = (x_1, x_2, x_3)$, siendo x_3 la altura, la fuerza es $F = (0, 0, -mg) = -\nabla\phi(x)$ para el potencial $\phi(x) = mgx_3$. La hipótesis es que g es constante, es decir, que la altura es pequeña frente al radio de la Tierra: el potencial exacto es $-GMm/r$, y mgx_3 es su desarrollo de primer orden alrededor de la superficie. Se desprecian también la resistencia del aire (que no es conservativa) y la rotación terrestre (la fuerza de Coriolis, que desvía un proyectil de largo alcance). El sistema $mx'' = -\nabla\phi$ se reduce a

$$x_1'' = 0, \qquad x_2'' = 0, \qquad x_3'' = -g.$$

Conocidas la posición x_0 y la velocidad v_0 iniciales, integrando dos veces,

$$x(t) = x_0 + v_0 t - \tfrac{1}{2} g t^2 e_3 \qquad \text{para } t \geq 0, \quad e_3 = (0, 0, 1).$$

Y conocidas las posiciones x_a y x_b en dos instantes $t_a < t_b$, que es la formulación del principio de mínima acción, la trayectoria es

$$x(t) = \frac{t_b - t}{t_b - t_a}\, x_a + \frac{t - t_a}{t_b - t_a}\, x_b - \tfrac{1}{2} g (t - t_a)(t - t_b)\, e_3 \qquad \text{para } t \in [t_a, t_b].$$

Se comprueba: en $t = t_a$ vale x_a, en $t = t_b$ vale x_b, la segunda derivada de los dos primeros términos (lineales en t) es cero y la del tercero es $-g e_3$. Aquí el problema de contorno tiene siempre solución única: la fórmula da una solución y la diferencia de otras dos sería afín y nula en dos instantes distintos, por lo que sería cero.

Caso particular: movimiento de un muelle. Una partícula de masa m sujeta al extremo de un resorte vertical, sin gravedad ni resistencia del medio. Sea $x(t)$ la desviación respecto de la posición de equilibrio $x = 0$. Se acepta que la fuerza elástica es $F(x) = -kx$ con $k > 0$, proporcional a la desviación y opuesta a ella: es la **ley de Hooke**, base de la Elasticidad Lineal. Es la misma decisión que la logística del capítulo 2: una fuerza $F(x)$ desconocida con $F(0) = 0$ (en el equilibrio no hay fuerza) se sustituye por su primer término de Taylor, $F'(0)x$, suponiendo aquí $F'(0) = -k < 0$ para tener una fuerza restauradora lineal. Si además F es impar, el siguiente término posible es $-k_3 x^3$: un muelle que se endurece o se ablanda al estirarlo mucho, cuyo período depende de la amplitud (el oscilador de Duffing); y se desprecian la masa del propio muelle y cualquier amortiguamiento, que no es conservativo. El potencial es $\phi(x) = \frac{k}{2}|x|^2$ y la ecuación

$$x'' = -\frac{k}{m} x,$$

con soluciones (teorema de las EDO lineales de coeficientes constantes del apéndice, raíces $\pm i\sqrt{k/m}$)

$$x(t) = c_1 \cos\Big(\sqrt{\tfrac{k}{m}}\, t\Big) + c_2 \sin\Big(\sqrt{\tfrac{k}{m}}\, t\Big).$$

La partícula sigue un **movimiento armónico simple**, periódico con período $T = 2\pi\sqrt{m/k}$. Conocidas x_0 y v_0 iniciales,

$$x(t) = x_0 \cos\Big(\sqrt{\tfrac{k}{m}}\, t\Big) + v_0 \sqrt{\tfrac{m}{k}} \sin\Big(\sqrt{\tfrac{k}{m}}\, t\Big) \qquad \text{para todo } t \in \mathbb{R}.$$

Por otra parte, conocidas las posiciones x_a y x_b en dos instantes $t_a < t_b$, **no siempre existe solución** del problema de contorno. Sea $\omega = \sqrt{k/m}$ y supongamos $t_b - t_a = \pi/\omega$, medio período. Toda solución cumple $x(t_a + \pi/\omega) = -x(t_a)$, porque el seno y el coseno cambian de signo al avanzar π. Luego si $x_b \neq -x_a$ no hay ninguna solución, y si $x_b = -x_a$ hay infinitas ($x(t) = x_a \cos(\omega(t - t_a)) + B \sin(\omega(t - t_a))$ sirve para cualquier B). Así, si $x_b \neq -x_a$, no hay trayectoria estacionaria con esos datos, ni por tanto mínimo diferenciable de la acción correspondiente.

También puede haber una trayectoria estacionaria sin mínimo. Sea $\tau = t_b - t_a$ y $h(t) = \sin(\pi(t - t_a)/\tau)$, que se anula en ambos extremos. Para cualquier trayectoria estacionaria x, al desarrollar la acción cuadrática y anular el término lineal se obtiene

$$S(x + \varepsilon h) - S(x) = \frac{m\varepsilon^2 \tau}{4} \left[\Big(\frac{\pi}{\tau}\Big)^2 - \omega^2 \right].$$

Si $\tau > \pi/\omega$, el valor es negativo para todo $\varepsilon \neq 0$: hay perturbaciones arbitrariamente pequeñas que reducen la acción. Esta es la razón para decir *estacionaria* en el principio físico, aunque se conserve el nombre histórico de «mínima acción».

CONEXIÓN: PROBLEMAS DE CONTORNO DE EDO

Es el fenómeno del capítulo 5 del libro de EDO del autor, «Problemas de contorno para un SDO lineal»: el problema de contorno $x'' + \omega^2 x = 0$, $x(0) = x_a$, $x(\tau) = x_b$ tiene solución única si y solo si $\omega\tau \notin \pi\mathbb{Z}$, que es cuando el problema homogéneo asociado solo tiene la solución nula. Es el «Teorema (de alternativa)» de §5.4 «Problema de contorno lineal» (página 89): si el homogéneo solo tiene la solución nula, hay solución única para todo dato; si tiene soluciones no nulas, o no hay solución o hay infinitas. Y el ejemplo de §5.5 (página 90) es exactamente este, $y'' + y = 0$ con $y(0) = y(\pi) = 0$, que tiene las infinitas soluciones $B \operatorname{sen} x$. En §5.2 se ve como intersección de dos variedades de soluciones que puede ser vacía, un punto o una recta. El mismo fenómeno gobierna `bvp4c` en las prácticas: cuando el problema homogéneo tiene soluciones no nulas, el resolvedor puede detectar una singularidad o producir resultados que dependen del arranque; una salida numérica por sí sola no resuelve la falta de unicidad o de compatibilidad. El problema de Bratu de la práctica 20 es el caso no lineal: dos soluciones, una o ninguna según λ.

EXPLÍCATE. *En el muelle, ¿cuál es la identidad de Beltrami y qué dice? Escríbela y comprueba con la solución explícita que $\frac{m}{2}x'^2 + \frac{k}{2}x^2$ es constante.*

3.3.2 Movimiento del péndulo

LA PREGUNTA

Según el relato tradicional, Galileo cronometró con su pulso una lámpara de la catedral de Pisa. Su estudio de las oscilaciones le llevó a defender que el período de un péndulo no depende de la amplitud. Los relojeros construyeron sobre eso los relojes de péndulo, y descubrieron que no es exacto: un péndulo que oscila mucho atrasa. ¿Qué ecuación rige el péndulo de verdad, cuándo vale la de Galileo, y qué pasa si se le da tanta energía que da la vuelta completa?

Un cuerpo de masa m está en el extremo de una varilla delgada de masa despreciable y longitud ℓ, que solo puede girar alrededor del otro extremo sin salir del plano.

Las hipótesis, una a una. Esa frase contiene cinco idealizaciones, y cada una tiene un precio conocido.

- **Varilla rígida y sin masa.** Rígida es lo que reduce el problema a un solo grado de libertad, el ángulo; sin masa es lo que permite poner toda la energía en la partícula del extremo. Si la varilla pesa, hay que sumar su energía cinética de rotación, y eso es la sección 3.3.3. Y si en vez de varilla es un *hilo*, hay una hipótesis escondida: el hilo solo tira, no empuja, y en las vueltas completas de más abajo la tensión tiene que seguir siendo positiva en el punto más alto; con un hilo, un péndulo con energía justa para pasar por arriba se descuelga antes de llegar. Las vueltas completas de la varilla no exigen esa condición adicional. Con un hilo también son posibles si la rapidez en lo alto satisface $v^2 \geq g\ell$, para que la tensión no resulte negativa.
- **Masa puntual.** Un cuerpo extenso en el extremo añade su propio momento de inercia. Si está unido rígidamente a la varilla, la ecuación conserva su forma, con una longitud efectiva determinada por el momento de inercia y la posición del centro de masas.
- **Movimiento plano.** El péndulo real puede salirse del plano y describir un cono: es el

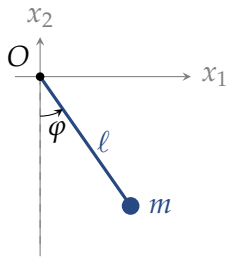
péndulo esférico, con dos ángulos y una cantidad conservada más, el momento angular vertical. El péndulo de Foucault requiere además tener en cuenta la rotación terrestre: el giro de su plano de oscilación no se explica solo pasando al péndulo esférico.

- **Sin rozamiento ni aire.** La energía se conserva y las oscilaciones se repiten sin perder amplitud. Con rozamiento $-c\varphi'$ los equilibrios bajos atraen a los datos próximos, como se verá con LaSalle. Un forzamiento periódico permite comportamientos más variados, incluido el caos para ciertos parámetros; no todo péndulo forzado es caótico. El fichero `quasipar.m` del tema 2 ofrece un ejemplo de forzamiento.
- **Punto de suspensión fijo.** Si el punto se mueve (un péndulo colgado de un carro, o el péndulo de Kapitza con el soporte vibrando), cambia la ecuación. En el segundo caso, una vibración vertical suficientemente rápida y adecuada puede estabilizar el equilibrio invertido.

En lugar de coordenadas cartesianas es más simple usar el ángulo φ que forma la varilla con la vertical. Con $\varphi = \varphi(t)$, la posición es

$$x_1 = \ell \sin\varphi(t), \qquad x_2 = -\ell\cos\varphi(t),$$

(el origen en el punto de suspensión, $\varphi = 0$ colgando hacia abajo).



Derivando, $x_1' = \ell\cos\varphi\,\varphi'$, $x_2' = \ell\sin\varphi\,\varphi'$. La energía cinética es

$$E_c = \frac{1}{2}m|x'|^2 = \frac{1}{2}m\ell^2(\varphi')^2,$$

porque $\cos^2 + \sin^2 = 1$, y la energía potencial gravitatoria es masa por g por altura, $E_p = mgx_2 = -mg\ell\cos\varphi$. La lagrangiana, en la variable φ, es

$$\mathcal{L}(\varphi,\varphi') = \frac{1}{2}m\ell^2|\varphi'|^2 + mg\ell\cos\varphi.$$

Leída término a término: $\ell\varphi'$ es la velocidad de un punto que recorre una circunferencia de radio ℓ con velocidad angular φ', luego el primer sumando es $\frac{1}{2}mv^2$ sin más; y $-\ell\cos\varphi$ es la altura del cuerpo medida desde el punto de suspensión, que vale $-\ell$ colgando y $+\ell$ invertido, luego el segundo es $-mg\cdot$altura. La ecuación de Euler-Lagrange, con $\partial_{\varphi'}\mathcal{L} = m\ell^2\varphi'$ y $\partial_\varphi\mathcal{L} = -mg\ell\sin\varphi$, es $\frac{d}{dt}(m\ell^2\varphi') = -mg\ell\sin\varphi$, esto es,

$$\varphi'' + \frac{g}{\ell}\sin\varphi = 0.$$

Es la **ecuación no lineal del péndulo** en ausencia de rozamiento. Obsérvese que ni la masa ni la tensión de la varilla aparecen: la primera se simplifica y la segunda no hace trabajo. Usando Beltrami (o multiplicando por φ' e integrando) se obtiene la ecuación de conservación de la energía dividida por $m\ell^2$,

$$\frac{1}{2}|\varphi'|^2 - \frac{g}{\ell}\cos\varphi = E,$$

una EDO de primer orden que resulta de gran ayuda para estudiar el comportamiento. Al variar E en $\mathbb{R}$ se consiguen todas las órbitas.

El sistema y sus equilibrios. Llamando $\psi = \varphi'$, la ecuación es el sistema de primer orden

$$\begin{cases} \varphi' = \psi, \\ \psi' = -\dfrac{g}{\ell}\sin\varphi. \end{cases}$$

Los puntos críticos son $(k\pi, 0)$ con $k \in \mathbb{Z}$: el péndulo quieto colgando (k par) o quieto invertido (k impar). La matriz jacobiana en ellos es

$$D(k\pi,0) = \begin{pmatrix} 0 & 1 \\ -\frac{g}{\ell}\cos(k\pi) & 0 \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 & 1 \\ -g/\ell & 0 \end{pmatrix} & k \text{ par}, \\ \begin{pmatrix} 0 & 1 \\ g/\ell & 0 \end{pmatrix} & k \text{ impar}, \end{cases}$$

con autovalores

$$\lambda = \begin{cases} \pm\sqrt{g/\ell}\, i & k \text{ par}, \\ \pm\sqrt{g/\ell} & k \text{ impar}. \end{cases}$$

Se separan los equilibrios en dos conjuntos, $\Lambda_1 = \{(2k\pi, 0)\}$ y $\Lambda_2 = \{((2k+1)\pi, 0)\}$. El teorema de linealización (teorema de estabilidad por linealización) dice que los de Λ_2 son **inestables** (un autovalor real positivo: sillas), y sobre los de Λ_1 no dice nada, porque los autovalores son imaginarios puros. Es la misma situación que en Volterra, y se resuelve igual: con la energía.

LA IDEA: EL POZO DE POTENCIAL, ANTES DE LAS CUENTAS

Todo el análisis que sigue se lee en un dibujo. La energía $E = \frac{1}{2}\psi^2 + U(\varphi)$ con $U(\varphi) := -\frac{g}{\ell}\cos\varphi$ es como la de una bola que rueda por un perfil de montaña rusa de forma U: valles en $\varphi = 0, \pm 2\pi, \ldots$ (el péndulo colgando, $U = -g/\ell$) y cumbres en $\varphi = \pm\pi, \pm 3\pi, \ldots$ (invertido, $U = g/\ell$). Se dibuja U y, encima, la recta horizontal a altura E_0. La bola solo puede estar donde la recta queda por encima del perfil, porque $\frac{1}{2}\psi^2 = E_0 - U(\varphi)$ tiene que ser ≥ 0; y su velocidad en cada punto es la raíz del doble de la distancia entre la recta y el perfil. Con eso se ven los cuatro casos: si E_0 es el fondo del valle, la bola está parada; si la recta corta al perfil dentro de un valle, la bola va y viene entre los dos cortes (oscila); si la recta pasa justo por las cumbres, la bola llega a la cumbre con velocidad cero, y como la cumbre es un equilibrio, no llega nunca; y si la recta pasa por encima de las cumbres, la bola no se para en ningún sitio y sigue siempre en el mismo sentido (da vueltas). Con $\ell = 1$ y $g = 9.8$: $E_0 = 0$ es la recta a media altura, y corta al perfil en $\cos\varphi = 0$, es decir, el péndulo oscila hasta la horizontal, $\varphi_{\text{máx}} = \pi/2$; $E_0 = 9.8$ es la separatriz.

Las órbitas por niveles de energía. De la conservación de la energía, cada órbita está contenida en el conjunto

$$\Lambda(E_0) := \left\{(\varphi,\psi) \in \mathbb{R}^2 : \tfrac{1}{2}|\psi|^2 - \tfrac{g}{\ell}\cos\varphi = E_0\right\}, \qquad E_0 := \tfrac{1}{2}|\psi_0|^2 - \tfrac{g}{\ell}\cos\varphi_0.$$

Como $0 \leq \frac{1}{2}\psi^2 = E_0 + \frac{g}{\ell}\cos\varphi$, se tiene $E_0 \geq -\frac{g}{\ell}\cos\varphi$ y $\psi = \pm\sqrt{2(E_0 + \frac{g}{\ell}\cos\varphi)}$; en particular $E_0 \geq -g/\ell$. En palabras: la velocidad angular al cuadrado es el doble de lo

que la energía total supera a la potencial, luego el péndulo solo puede visitar los ángulos donde la potencial no supera a E_0, y en cada uno de ellos su velocidad está determinada salvo el signo. Como la potencial mínima es $-g/\ell$ (colgando) y la máxima es g/ℓ (invertido), los valores $E_0 = -g/\ell$ y $E_0 = g/\ell$ son las dos fronteras entre comportamientos distintos. Se analiza la órbita de (φ_0, ψ_0) según el valor de E_0.

- $E_0 = -g/\ell$**.** Entonces $\cos\varphi = 1$ y $\psi = 0$: $\varphi(t) \equiv 2k\pi$, $\psi \equiv 0$. El punto inicial es un equilibrio de Λ_1: el péndulo está en reposo colgando y permanece así.
- $-g/\ell < E_0 < g/\ell$**.** Se deduce $\cos\varphi \geq -E_0\ell/g$, luego φ se mueve en el intervalo $[-\varphi_{\text{máx}}, \varphi_{\text{máx}}]$ con $\varphi_{\text{máx}} = \arccos(-E_0\ell/g)$ (o ese intervalo más un múltiplo de 2π), y $\psi = 0$ justo en los extremos $\varphi = \pm\varphi_{\text{máx}}$. El conjunto $\Lambda(E_0)$ está formado por una infinidad de **curvas cerradas** que no contienen puntos críticos: hay un punto de Λ_1 en el interior de cada una, y los de Λ_2 no están sobre ellas, porque en ellos la energía vale $-\frac{g}{\ell}\cos((2k+1)\pi) = g/\ell > E_0$. El movimiento es **periódico**, entre los ángulos $\pm\arccos(-E_0\ell/g)$: el péndulo oscila.
- $E_0 = g/\ell$**.** El nivel de energía contiene puntos sobre todos los valores de φ, pero una órbita individual no recorre todos esos ángulos. La órbita está en

$$\Lambda(g/\ell) = \left\{(\varphi,\psi) : \tfrac{1}{2}\psi^2 - \tfrac{g}{\ell}(\cos\varphi + 1) = 0\right\},$$

  que contiene a Λ_2 (en $((2k+1)\pi, 0)$ la expresión vale 0) y es una unión de órbitas. O bien el punto inicial es un equilibrio de Λ_2 y el péndulo permanece invertido, o bien $\varphi_0 \neq (2k+1)\pi$, lo que implica $\psi \neq 0$ a lo largo de la órbita. Tomemos k tal que $(2k-1)\pi < \varphi_0 < (2k+1)\pi$. Si $\psi_0 > 0$, el ángulo aumenta y la semiórbita positiva converge a $((2k+1)\pi, 0)$ cuando $t \to +\infty$; si $\psi_0 < 0$, converge a $((2k-1)\pi, 0)$. La trayectoria permanece entre esas dos posiciones invertidas. Son las **separatrices**: el péndulo tiene justo la energía para llegar arriba, y tarda un tiempo infinito en llegar.
- $E_0 > g/\ell$**.** De nuevo φ toma todos los valores, $\Lambda(E_0)$ no contiene ningún equilibrio y $\psi = \pm\sqrt{2(E_0 + \frac{g}{\ell}\cos\varphi)} \neq 0$ para todo φ. Si $\psi_0 > 0$, entonces $\psi > 0$ en toda la órbita y el péndulo **da vueltas indefinidamente** en sentido contrario a las agujas del reloj; si $\psi_0 < 0$, en el sentido de las agujas.

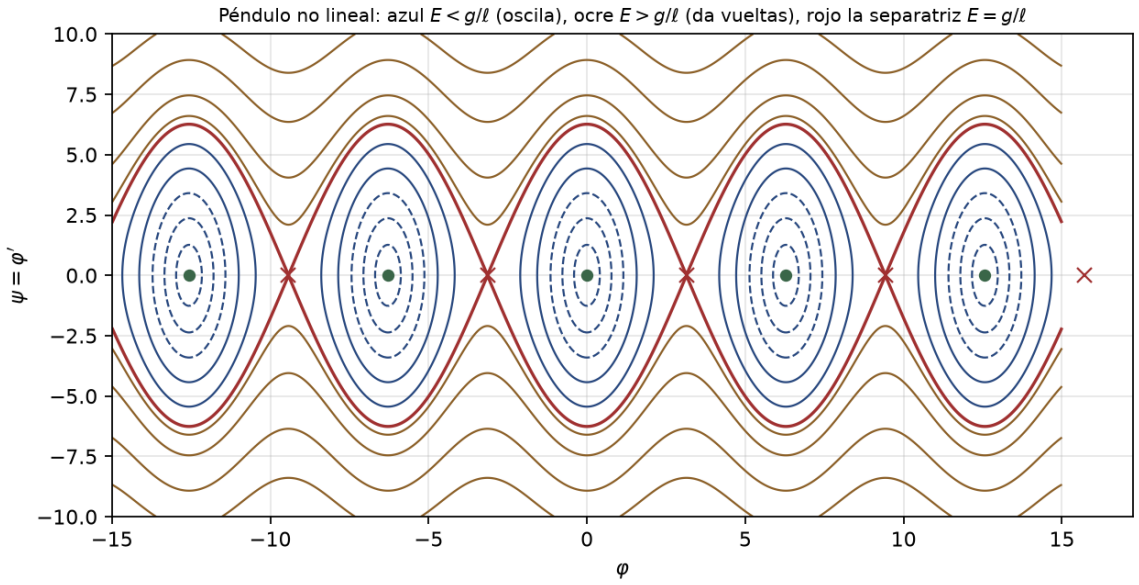


Figura 3.2: Órbitas del péndulo con $\ell = 1$ m y $g = 9.8$. Las curvas cerradas (azul) son oscilaciones, con un equilibrio estable (verde) en su interior; las onduladas (ocre) son vueltas completas; la separatriz (rojo, $E = g/\ell$) une los equilibrios inestables (aspas). Las abscisas que difieren en 2π son el mismo ángulo, así que también las onduladas se cierran al enrollar el dibujo en un cilindro. La excepción es la roja: sobre ella el movimiento no es periódico, porque tarda un tiempo infinito en llegar al equilibrio invertido.

En la figura 3.2 están dibujadas varias órbitas para $\ell = 1$ m y $g = 9.8$ m/s^2. Las diferencias de 2π en abscisas corresponden al mismo ángulo, y por eso las onduladas, que en el plano parecen abiertas, son en realidad periódicas: cada vuelta completa devuelve el péndulo al mismo estado físico, y el período es el tiempo que tarda φ en crecer 2π. Conviene no extender la frase a *todas* las órbitas, porque las separatrices rojas del dibujo no lo son: el análisis de más arriba da sobre ellas una solución que se acerca al equilibrio invertido y tarda un tiempo infinito en llegar, así que no vuelve nunca a su punto de partida. Enrollar el cilindro cierra la curva, no la órbita. Las de forma elíptica corresponden a $E < 9.8 = g/\ell$: la velocidad ψ pasa de positiva a negativa y el péndulo va a un lado y a otro. Para $E > 9.8$ la velocidad no cambia de signo y el péndulo está siempre girando en el mismo sentido. Los equilibrios estables son $\varphi = 2k\pi$, $\psi = 0$; los inestables, $\varphi = (2k+1)\pi$, $\psi = 0$.

La idea, y por qué no hay que buscar nada. El método de Liapunov exige inventar una función que decrezca a lo largo de las órbitas, y ese invento es su gran dificultad: no hay ningún procedimiento general para construirla. Aquí, sin embargo, no hace falta inventar nada, y esa es toda la gracia del argumento. La función ya existe y tiene nombre físico: es la *energía*. Un péndulo sin rozamiento no gana ni pierde energía, luego $\dot{V} = 0$; y un péndulo con rozamiento solo la pierde, luego $\dot{V} \leq 0$. La estabilidad sale de ahí sin resolver la ecuación: si la energía no puede crecer, el péndulo no puede alejarse del fondo más de lo que su energía inicial le permite. El pozo de potencial es una trampa de la que no se sale sin que alguien empuje.

Por qué los equilibrios de Λ_1 son estables. La figura lo hace creíble, pero no lo demuestra; el argumento es el de Volterra.
Paso 1: la energía es definida positiva. La función $V(\varphi,\psi)=\frac{1}{2}\psi^2+\frac{g}{\ell}(1-\cos\varphi)$ es definida positiva alrededor de $(0,0)$ (vale 0 allí, y $1-\cos\varphi>0$ para $0<|\varphi|<2\pi$), y su derivada respecto del sistema es

$$\dot V=\psi\psi'+\frac{g}{\ell}\sin\varphi\,\varphi'=\psi\Big(-\frac{g}{\ell}\sin\varphi\Big)+\frac{g}{\ell}\sin\varphi\,\psi=0.$$

Paso 2: la cancelación exacta es la conservación de la energía (el núcleo). Los dos sumandos son iguales y de signo contrario: son las tasas de cambio de la energía cinética y de la potencial, divididas por $m\ell^2$. Lo que aumenta una lo pierde la otra. Es la identidad de Beltrami de la sección anterior, escrita en el lenguaje de los sistemas dinámicos. Con $\dot V=0$, V es una función de Liapunov y por el teorema de Liapunov el origen es estable.
Paso 3: estable pero no atractivo, y qué cambia con rozamiento. No es atractivo, porque las órbitas vecinas son cerradas: la energía se conserva, así que una órbita que empieza con $V(\varphi_0,\psi_0)>0$ conserva ese valor y no puede tender al equilibrio, donde $V=0$. Con rozamiento, $\psi'=-\frac{g}{\ell}\sin\varphi-c\psi$, con $c>0$, sale $\dot V=-c\psi^2\le 0$. Para aplicar LaSalle, tomamos la componente que contiene al origen de $\{V\le\delta\}$, con $0<\delta<2g/\ell$: es compacta e invariante hacia adelante. Dentro de ella, $\dot V=0$ implica $\psi=0$, y una trayectoria permanece en ese eje solo si también $\sin\varphi=0$; el único punto posible en esa componente es el origen. LaSalle (corolario de estabilidad asintótica por LaSalle) da por tanto estabilidad asintótica local del equilibrio bajo. Obsérvese que $\dot V$ no es definida negativa —se anula en todo el eje $\psi=0$, no solo en el origen—, y por eso el teorema de Liapunov a secas no basta y hace falta LaSalle. ■

Dónde está esto demostrado con más profundidad. Este mismo cálculo, con la misma función de Liapunov y en su versión general, está en el libro de EDO del autor: la libro de EDO, §7.4 «Ejemplos» trata la *ecuación de Newton* $x''=-kf(x)$ con $xf(x)>0$, que es el péndulo con $f(x)=\sin x$ en un entorno de 0 contenido en $(-\pi,\pi)$, toma $V=\frac{1}{2}|y_2|^2+kF(y_1)$ con $F(y_1)=\int_0^{y_1}f$ —cinética más potencial— y obtiene $\dot V=0$ y estabilidad uniforme; y a continuación la *ecuación de Newton con fricción* $x''+\beta\psi(x')=-kf(x)$, que da $\dot V=-\beta y_2\psi(y_2)\le 0$. Los enunciados de Liapunov y de LaSalle están en libro de EDO, §7.3 «Condiciones suficientes de estabilidad» y libro de EDO, §8.4 «Principio de invariancia y Teorema de LaSalle». La dificultad del método está en construir una función adecuada: en problemas mecánicos, la energía proporciona a menudo una candidata natural, como en este péndulo.

Observación (ecuación linealizada del péndulo)**.** Para ángulos pequeños se usa $\sin\varphi\approx\varphi$ y se obtiene la **ecuación linealizada**

$$\varphi''+\frac{g}{\ell}\varphi=0,$$

con soluciones $\varphi(t)=C_1\cos(kt)+C_2\sin(kt)$, $k=\sqrt{g/\ell}$: el péndulo de Galileo, con período $2\pi\sqrt{\ell/g}$ independiente de la amplitud. Su energía es $E=\frac{1}{2}|\varphi'|^2+\frac{g}{2\ell}|\varphi|^2$ y las órbitas son elipses (figura 3.3). Solo hay similitud con el péndulo no lineal cuando E es pequeño.

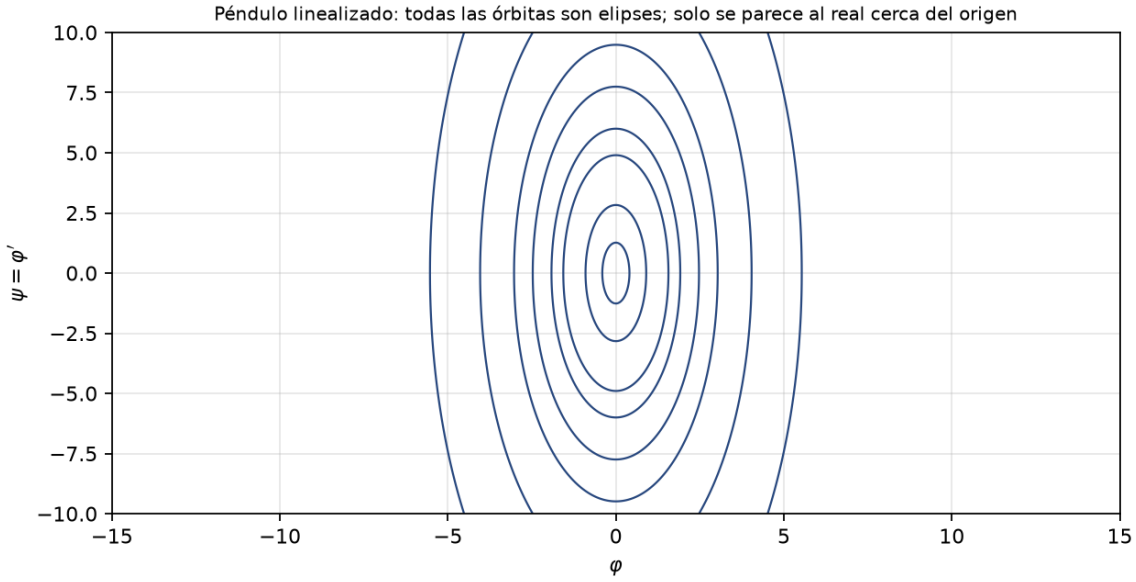


Figura 3.3: Órbitas del péndulo linealizado con $\ell = 1$: elipses concéntricas para todo nivel de energía. Comparado con la figura anterior, coincide cerca del origen y no ve ni las separatrices ni las vueltas completas.

La idea: cuánto atrasa el reloj

La respuesta al relojero se calcula con la energía. En la órbita de amplitud $0 < \varphi_0 < \pi$, $\psi = \sqrt{\frac{2g}{\ell}(\cos\varphi - \cos\varphi_0)}$, y el período es cuatro veces el tiempo de subir de 0 a φ_0: $T = 4\int_0^{\varphi_0} d\varphi/\psi$. Con el cambio $\sin(\varphi/2) = \sin(\varphi_0/2)\sin\theta$ sale $T = 4\sqrt{\ell/g}\int_0^{\pi/2} \frac{d\theta}{\sqrt{1-\sin^2(\varphi_0/2)\sin^2\theta}}$, una integral elíptica que vale $\pi/2$ en $\varphi_0 = 0$ (Galileo) y crece con la amplitud. Con $\ell = 1$: para $\varphi_0 = 0.1$ rad, $T = 2.0083$ s frente a 2.0071; para 1 rad, 2.140 s (un 6.6 % más); para 2 rad, 2.667 s (un 33 % más); y para 3 rad, 5.161 s, más de dos veces y media el período pequeño. El crecimiento se dispara al acercarse a π, porque el movimiento se hace cada vez más lento cerca de la posición invertida. Un reloj de péndulo se regula con amplitud pequeña y constante. Huygens publicó en 1673 su estudio del péndulo cicloidal en el *Horologium oscillatorium*; su período no depende de la amplitud: es la misma cicloide de la braquistócrona, que también es la *tautócrona*.

Explícate. *Sin mirar: en el caso $E_0 = g/\ell$ con $\psi_0 > 0$, ¿por qué el péndulo tarda un tiempo infinito en llegar arriba? Piensa en que el punto de llegada es un equilibrio y en la unicidad de solución.*

3.3.3 Movimiento de una varilla

Se considera una barra de anchura despreciable cuyo movimiento se realiza en el plano; con esa simplificación bastan tres variables (para un sólido rígido general hacen falta seis). La barra tiene longitud 2ℓ y densidad constante ρ, se encuentra en caída libre sometida solo a la gravedad, y su eje está en un plano perpendicular a la superficie terrestre.

Las hipótesis, una a una, porque cada una fija un término de lo que sigue. **Anchura despreciable**: la barra es un segmento, y sus puntos se numeran con un solo parámetro r; una barra gruesa necesitaría integrar en volumen. **Rígida**: la distancia r de cada punto al centro no cambia con el tiempo, y por eso la posición de la barra entera queda determinada por dónde está el centro y hacia dónde apunta: tres números, p_1, p_2 y φ. Una barra flexible tendría infinitos grados de libertad y su modelo es la viga de la práctica 19. **Densidad constante**: el centro de masas está en el centro geométrico, y eso es lo que hace que un término se anule más abajo. **Movimiento plano**: en el espacio un sólido rígido necesita tres coordenadas de posición y tres de orientación; las rotaciones espaciales no conmutan en general, y las ecuaciones de la rotación (las de Euler) son no lineales; en el plano el ángulo es uno solo y las rotaciones conmutan entre sí. **Solo la gravedad**: sin aire, que frenaría y haría girar de otra manera una barra que cae de plano que una que cae de canto. Con estas cinco cosas, el resultado va a ser exacto y sorprendentemente simple.

Parametrización. Sea $p = (p_1(t), p_2(t))$ el centro de gravedad y sea $\varphi(t)$ el ángulo de la barra con la horizontal, de modo que un extremo es $q = (p_1 + \ell\cos\varphi,\ p_2 + \ell\sin\varphi)$. Cualquier otro punto de la barra es $x^r = p + y^r$ con $y^r = (r\cos\varphi, r\sin\varphi)$, $r \in (-\ell, \ell)$. Su velocidad es $(x^r)' = p' + (y^r)'$ con $(y^r)' = (-r\sin\varphi\,\varphi',\ r\cos\varphi\,\varphi')$.

Energía cinética. Es la suma de las de las partículas, es decir, la integral

$$E_c(t) = \frac{1}{2}\int_{-\ell}^{\ell} \rho\,|p' + (y^r)'|^2\,dr = \rho\ell|p'|^2 + \rho\,p'\cdot\int_{-\ell}^{\ell}(y^r)'\,dr + \frac{\rho}{2}\int_{-\ell}^{\ell}|(y^r)'|^2\,dr,$$

desarrollando el cuadrado: $|p' + (y^r)'|^2 = |p'|^2 + 2p'\cdot(y^r)' + |(y^r)'|^2$, y al integrar en r el primer sumando da $|p'|^2$ por la longitud 2ℓ, de ahí el $\rho\ell|p'|^2$. Los tres sumandos se leen así: el primero es la energía cinética que tendría toda la masa, $2\ell\rho$, concentrada en el centro y moviéndose con él; el tercero es la energía de girar alrededor del centro; y el segundo mide cuánto «se ayudan» traslación y rotación, y va a valer cero. El término cruzado se anula: $\int_{-\ell}^{\ell}(y_i^r)'\,dr = \frac{d}{dt}\int_{-\ell}^{\ell} y_i^r\,dr = 0$ porque $\int_{-\ell}^{\ell} y_i^r\,dr = \cos\varphi\int_{-\ell}^{\ell} r\,dr = 0$ (y lo mismo con $\sin\varphi$): es la definición de centro de gravedad, la que hace que la descomposición $x^r = p + y^r$ sea la buena. Y $|(y^r)'|^2 = r^2(\sin^2\varphi + \cos^2\varphi)|\varphi'|^2 = r^2|\varphi'|^2$, luego

$$E_c(t) = \rho\ell|p'(t)|^2 + \frac{\rho}{2}|\varphi'(t)|^2\int_{-\ell}^{\ell} r^2\,dr = \rho\ell|p'(t)|^2 + \frac{1}{3}\rho\ell^3|\varphi'(t)|^2.$$

Comprobación con números: una barra de 2 m ($\ell = 1$) y 3 kg/m que se traslada a 2 m/s sin girar tiene $E_c = 3\cdot 1\cdot 4 = 12$ J, que es $\frac{1}{2}\cdot 6\cdot 4$, la de una partícula de 6 kg; si además gira a $\varphi' = 1$ rad/s, se añaden $\frac{1}{3}\cdot 3\cdot 1 = 1$ J, que es $\frac{1}{2}I\varphi'^2$ con $I = 2\ \mathrm{kg\,m^2}$. **Un factor que conviene comprobar.** A veces se ve escrito el segundo término como $\frac{2}{3}\rho\ell^3|\varphi'|^2$, que es el doble de lo que sale. La integral $\int_{-\ell}^{\ell} r^2\,dr$ vale $2\ell^3/3$, y multiplicada por $\rho/2$ da $\rho\ell^3/3$. El valor $\frac{2}{3}\rho\ell^3$ es el momento de inercia I de la barra respecto de su centro (masa $2\ell\rho$ por $(2\ell)^2/12$), y la energía

cinética de rotación es $\frac{1}{2}I|\varphi'|^2 = \frac{1}{3}\rho\ell^3|\varphi'|^2$: falta el $\frac{1}{2}$. En este caso de giro libre el desliz no cambia la ecuación $\varphi'' = 0$, porque su coeficiente no nulo se cancela. Sí cambiaría la relación entre aceleración y momento de fuerzas si hubiera un par aplicado.

Energía potencial. Es la suma de las alturas por ρg:

$$E_p(t) = \int_{-\ell}^{\ell} \rho g\, x_2^r\, dr = \int_{-\ell}^{\ell} \rho g\big(p_2(t) + y_2^r(t)\big)dr = 2\ell\rho g\, p_2(t),$$

otra vez porque $\int y_2^r\, dr = 0$. Es la energía potencial de una partícula de masa $m = 2\ell\rho$ a la altura del centro de masas.

La lagrangiana y las ecuaciones. Con lo anterior,

$$\mathcal{L}(p, \varphi, p', \varphi') = \rho\ell|p'|^2 + \frac{1}{3}\rho\ell^3|\varphi'|^2 - 2\ell\rho g\, p_2.$$

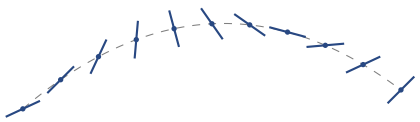
Las ecuaciones de Euler-Lagrange en p: $\nabla_{p'}\mathcal{L} = 2\rho\ell p'$, $\nabla_p\mathcal{L} = (0, -2\ell\rho g)$, luego $2\rho\ell p'' = (0, -2\ell\rho g)$, es decir,

$$p_1'' = 0, \qquad p_2'' = -g.$$

En φ: $\partial_{\varphi'}\mathcal{L} = \frac{2}{3}\rho\ell^3\varphi'$, $\partial_\varphi\mathcal{L} = 0$, luego

$$\varphi'' = 0.$$

El sistema para p dice que **el centro de masas se mueve como una partícula sometida a la gravedad**, y la ecuación para φ que **la barra gira a velocidad constante alrededor del centro de masas**. Es el resultado que se ve en cualquier bastón lanzado al aire.



Instantes igualmente espaciados; el tiempo avanza hacia la derecha.

LA IDEA: POR QUÉ SE DESACOPLAN

La lagrangiana se separa en un término de (p, p') y otro de (φ, φ'): se anula el término cruzado de la energía cinética y, además, la energía potencial del campo uniforme depende solo de la altura del centro de masas. Y eso se debe a haber elegido el centro de gravedad como origen de la descomposición: con cualquier otro punto de la barra, $\int y^r\, dr \neq 0$ y aparecería un acoplamiento. La coordenada φ es *cíclica* (no aparece en $\mathcal{L}$), luego su momento $\partial_{\varphi'}\mathcal{L}$ se conserva: es el momento angular respecto del centro de masas, y de ahí φ' constante. La traslación obedece a la fuerza total; la rotación, al momento de las fuerzas respecto del centro de masas, que aquí es cero.

CONEXIÓN: MECÁNICA, EL SÓLIDO RÍGIDO

La descomposición «traslación del centro de masas más rotación alrededor de él», con la energía cinética $\frac{1}{2}m|p'|^2 + \frac{1}{2}I|\varphi'|^2$, es el teorema de König de la asignatura de Mecánica, y el $I = \frac{2}{3}\rho\ell^3 = \frac{1}{12}m(2\ell)^2$ es el momento de inercia de una barra. El apartado lo obtiene sumando las energías de sus puntos y aplicando el principio de Hamilton, bajo las hipótesis de rigidez y campo uniforme ya formuladas.

Qué se ha ganado en esta sección: un método en cuatro pasos. El péndulo y la varilla se han resuelto con la misma secuencia, y es la que sirve para cualquier sistema mecánico con fuerzas conservativas y ligaduras ideales. Lo que el principio aporta frente a $F = ma$ está en los pasos 1 y 4: libertad para elegir coordenadas y leyes de conservación que salen solas.

1. *Coordenadas.* Se eligen las mínimas que describen la posición y que ya incorporan las ligaduras: el ángulo φ en el péndulo; el centro p y el ángulo φ en la varilla. Las reacciones de las ligaduras ideales no aparecen en las ecuaciones de esas coordenadas.
2. *Energía cinética.* Se escribe la posición cartesiana en función de las coordenadas, se deriva y se toma $\frac{m}{2}|x'|^2$, o se integra en la masa, como en la varilla. Es el paso donde se cometen casi todos los errores: en el péndulo, $x' = (\ell \cos\varphi\, \varphi', \ell \sin\varphi\, \varphi')$ da $\ell^2\varphi'^2$ solo después de usar $\cos^2\varphi + \sin^2\varphi = 1$.
3. *Energía potencial*, sumando todas las contribuciones (gravedad, muelles) en las mismas coordenadas.
4. *Euler-Lagrange* para $\mathcal{L} = E_c - E_p$, una ecuación por coordenada. Antes de derivar conviene mirar si alguna coordenada no aparece en $\mathcal{L}$: es *cíclica*, y su momento $\partial\mathcal{L}/\partial q_i'$ se conserva, como el de φ en la varilla. Y si $\mathcal{L}$ no depende de t, Beltrami da la energía, como en el péndulo.

En el apéndice C (subsección C.9.4) hay tres sistemas más montados con estos cuatro pasos: un péndulo sujeto además a un muelle, una partícula entre dos muelles y una partícula sobre un cono, en coordenadas polares.

3.4 Aplicaciones en Mecánica (II): Problemas de energía mínima

LA PREGUNTA

Un cable colgado entre dos postes adopta un perfil que depende de su peso, su tensión y las cargas que sostiene. ¿Qué está minimizando al quedarse quieto, y cómo cambia la respuesta si en vez de clavar los extremos se sujetan con muelles?

Se consideran sistemas físicos **en equilibrio**, en estado estacionario e independiente del tiempo. En los modelos que siguen, el equilibrio se obtiene minimizando la **energía potencial**. En general, ser un punto estacionario de la energía no basta para ser un mínimo; aquí la convexidad permitirá comprobarlo. Como se está ante estados estacionarios, toda la energía es potencial: no hay energía cinética.

3.4.1 Cuerda elástica (con pequeños desplazamientos). Elasticidad lineal

Una cuerda de longitud de referencia L ocupa, sin fuerzas verticales, el segmento horizontal $[0, L] \times \{0\} \subset \mathbb{R}^2$. Puede estar ya sometida a tensión horizontal. Al someterla a una fuerza estacionaria vertical, la posición de equilibrio cambia. Se supone que las partículas de la cuerda solo experimentan **desplazamientos verticales**: la que estaba en $(x, 0)$ pasa a $(x, u(x))$. La función u es la **función de desplazamiento** y da el perfil de la cuerda. Suponiendo que u y u' no toman valores grandes (**pequeños desplazamientos**), se acepta que la energía potencial elástica que adquiere la cuerda al apartarla de su posición de equilibrio es

$$E_{pe} = \frac{\mu}{2} \int_0^L |u'(x)|^2 \, dx,$$

con $\mu > 0$ un coeficiente efectivo de rigidez transversal, que en la interpretación de cuerda tensa siguiente es la tensión inicial. Los ejemplos que siguen tienen una energía potencial total que es E_{pe} más otra cantidad determinada por las fuerzas exteriores.

De dónde sale $\frac{\mu}{2}\int u'^2$, y qué se supone. Una cuerda tensa con tensión μ (una fuerza, la que tira de cada extremo) se opone a que la alarguen: alargarla una longitud δ cuesta un trabajo $\mu\delta$, si la tensión no cambia al alargar. Al deformarse en u, el trozo que ocupaba $[x, x+dx]$ pasa a medir $\sqrt{1+u'(x)^2}\,dx$, luego el alargamiento total es $\int_0^L(\sqrt{1+u'^2}-1)\,dx$, y con la pendiente pequeña, $\sqrt{1+u'^2}-1 \approx \frac{1}{2}u'^2$: la energía es $\frac{\mu}{2}\int u'^2$. Las hipótesis están todas en esa cuenta. **Solo desplazamientos verticales**: cada partícula sube o baja sin moverse de su abscisa; en una cuerda real también se desplaza algo en horizontal, y aquí se desprecia ese movimiento dentro de la aproximación. **Pendiente pequeña**: es lo que permite quedarse con $\frac{1}{2}u'^2$, y lo que hace lineal todo lo demás; con pendientes grandes hay que conservar la dependencia no lineal. Eso no produce por sí solo la catenaria: el cable inextensible de la caja siguiente tiene además una restricción de longitud y otro modo de contar su peso. **Tensión constante**, independiente de cuánto se ha estirado la cuerda; una goma que se endurece al estirarse necesitaría una ley que relacione tensión y deformación, y su correspondiente energía. Y **estado estacionario**: la cuerda no vibra; si vibrara habría energía cinética $\frac{\rho}{2}\int u_t^2$ y la ecuación de Euler-Lagrange sería la de ondas, $\rho u_{tt} = \mu u_{xx}$ en ausencia de fuerza vertical exterior, en derivadas parciales. El modelo lineal da una parábola bajo carga uniforme por unidad de abscisa. Para pendientes pequeñas aproxima perfiles de cables pesados; permite resolver el problema a mano y sumar soluciones de la ecuación lineal.

EL CAMINO QUE NO SE TOMÓ: LA CATENARIA EXACTA

La alternativa. Modelar un cable flexible e inextensible, con densidad constante por unidad de longitud de arco. Su longitud no cambia; se minimiza la energía gravitatoria $\int \rho g u\,ds$ con $\int \sqrt{1+u'^2}\,dx$ fija y extremos dados. La solución, cuando la longitud permite unirlos con un cable colgante, es $u = C\cosh((x-x_0)/C) + D$, con $C > 0$: una catenaria. La constante D fija la altura; en la superficie de revolución anterior, u era un radio y no había esa traslación libre.
Su ventaja. Describe el cable ideal bajo su propio peso sin exigir pendientes pequeñas; la tensión puede variar a lo largo de él.
Por qué se elige la versión lineal. Porque $\frac{\mu}{2}\int u'^2$ es el primer término del desarrollo de $\mu\int(\sqrt{1+u'^2}-1)$, la energía elástica de una cuerda que se estira, y con él la ecuación de Euler-Lagrange es *lineal*: $-\mu u'' = F$. Eso es la **Elasticidad Lineal**: sustituir la energía exacta por su parte cuadrática, a cambio de una ecuación que se resuelve y que se puede superponer. Para una cuerda de tender, que apenas se curva, las dos dan lo mismo (una parábola y una catenaria de poca flecha son muy próximas cuando las pendientes son pequeñas); para una cadena que cuelga mucho, no. Es la misma decisión que Hooke con el muelle: proporcionalidad como aproximación de primer orden.

Ejemplo (cuerda sujeta en los extremos bajo la acción de la gravedad)**.** La cuerda está en el campo gravitatorio y sus extremos han sido fijados. La energía potencial total se obtiene integrando las densidades de energía elástica y gravitatoria. En esta subsección se absorbe la densidad lineal en g: este símbolo representa el peso por unidad de longitud de referencia, no solo la aceleración gravitatoria.

$$E = \int_0^L \Big(\frac{\mu}{2}|u'|^2 + gu\Big)dx,$$

y la función u que da la posición de la cuerda es solución del problema

$$\text{Minimizar } \int_0^L \Big(\frac{\mu}{2}|u'|^2 + gu\Big)dx \quad \text{sobre } u \in C^1([0,L]),\ u(0) = u(L) = 0.$$

Con $f(u,u') = \frac{\mu}{2}u'^2 + gu$, $\partial_{u'}f = \mu u'$ y $\partial_u f = g$, Euler-Lagrange conduce al problema de contorno

$$\begin{cases} -\mu u'' = -g & \text{en } (0,L), \\ u(0) = u(L) = 0. \end{cases}$$

Gracias a la convexidad de $(u,u') \mapsto \frac{\mu}{2}|u'|^2 + gu$ (un cuadrado más un término lineal), el criterio de convexidad dice que u es solución del problema de mínimos si y solo si lo es del de contorno. Integrando dos veces $u'' = g/\mu$, $u = \frac{g}{2\mu}x^2 + c_1 x + c_2$, y con $u(0) = u(L) = 0$ el perfil de la cuerda es el arco de parábola

$$u(x) = -\frac{g}{2\mu}x(L-x) \qquad \text{para } x \in [0,L].$$

Cuanto mayor es μ, más rígida es la cuerda y menor su deformación: la flecha en el centro es $gL^2/(8\mu)$. La propia solución permite comprobar la aproximación: máx $|u'| = gL/(2\mu)$ debe ser pequeño. En el caso más general de una carga exterior vertical continua $F = F(x)$, por unidad de longitud de referencia, el problema es análogo y se llega a

$$\begin{cases} -\mu u'' = F(x) & \text{en } (0,L), \\ u(0) = u(L) = 0. \end{cases}$$

Ejemplo (cuerda no homogénea)**.** Si la cuerda no es homogénea, tomamos $\mu \in C^1([0,L])$ con $\mu(x) > 0$, y la energía elástica es $E_{pe} = \frac{1}{2}\int_0^L \mu(x)|u'(x)|^2\,dx$. Con una distribución de esfuerzos verticales $F = F(x)$, $\partial_{u'}f = \mu(x)u'$, y el problema de contorno es

$$\begin{cases} -\big(\mu(x)u'\big)' = F(x) & \text{en } (0,L), \\ u(0) = u(L) = 0. \end{cases}$$

Es la forma autoadjunta $(p(x)y')' + q(x)y = b(x)$ del libro de EDO, §5.5 «EDOs lineales de segundo orden con condiciones de contorno separadas» (página 90), con $p = \mu$, $q = 0$ y $b = -F$, y el operador elíptico autoadjunto L de §5.7; es el que resuelve `bvp4c` en la práctica 17.

Ejemplo (cuerda sujeta con muelles elásticos)**.** En los ejemplos anteriores la posición de los extremos es fija. Ahora la cuerda está acoplada en los extremos a sendos muelles que actúan en la dirección vertical, con posición de equilibrio a altura $y = 0$. Con la cuerda en el campo gravitatorio, la energía potencial es

$$E = \int_0^L \Big(\frac{\mu}{2}|u'|^2 + gu\Big)dx + \frac{k}{2}|u(0)|^2 + \frac{k}{2}|u(L)|^2,$$

donde $k > 0$ es la constante de recuperación de los muelles (la misma para los dos, por simplicidad) y μ es constante. Los dos últimos términos son la energía elástica de cada muelle, $\frac{k}{2}$ por el cuadrado de su alargamiento, que es el desplazamiento del extremo. El problema es

$$\text{Minimizar } \int_0^L \Big(\frac{\mu}{2}|u'|^2 + gu\Big)dx + \frac{k}{2}|u(0)|^2 + \frac{k}{2}|u(L)|^2 \quad \text{sobre } u \in C^1([0,L]).$$

Contrariamente a los ejemplos anteriores, no se conoce el valor de u en ninguno de los extremos, y los dos últimos sumandos no son integrales. Para resolverlo es mejor usar directamente las ideas de la demostración del teorema de Euler–Lagrange.

LA IDEA: LO QUE VA A SALIR, Y POR QUÉ

El procedimiento es el de siempre, perturbar con $u+\sigma h$, derivar en $\sigma=0$ e integrar por partes, con dos novedades. Primera: como el problema no fija u en los extremos, h puede valer lo que quiera en 0 y en L, y por eso la integración por partes deja dos términos de frontera vivos. Segunda: los dos sumandos de los muelles se derivan como un cuadrado cualquiera y producen $ku(0)h(0)$ y $ku(L)h(L)$, que son también términos de frontera. Al final habrá, en cada extremo, un coeficiente multiplicando a $h(0)$ (o a $h(L)$) que tiene que ser cero, y ese coeficiente será «fuerza del muelle, ku, menos fuerza con que tira la cuerda, $\mu u'$» (con el signo que le toque a cada extremo). O sea, que la cuerda se para donde cada muelle equilibra a la cuerda: eso es lo que uno esperaría, y es lo que sale.

Sea u solución y $h\in C^1([0,L])$ arbitraria, sin ninguna condición en los extremos. Para todo $\sigma\in\mathbb{R}$, $u+\sigma h$ es admisible y por tanto

$$\int_0^L\Big(\frac{\mu}{2}|u'+\sigma h'|^2+g(u+\sigma h)\Big)dx+\frac{k}{2}|u(0)+\sigma h(0)|^2+\frac{k}{2}|u(L)+\sigma h(L)|^2$$
$$\geq\int_0^L\Big(\frac{\mu}{2}|u'|^2+gu\Big)dx+\frac{k}{2}|u(0)|^2+\frac{k}{2}|u(L)|^2.$$

La derivada respecto de σ en $\sigma=0$ es cero (teorema de Fermat), y derivando término a término (el cuadrado $\frac{\mu}{2}|u'+\sigma h'|^2$ da $\mu(u'+\sigma h')h'$, y $\frac{k}{2}|u(0)+\sigma h(0)|^2$ da $k(u(0)+\sigma h(0))h(0)$) se obtiene

$$\int_0^L\big(\mu u'h'+gh\big)dx+ku(0)h(0)+ku(L)h(L)=0\qquad\text{para toda } h\in C^1([0,L]).$$

Es la variación primera de la energía total, igual a cero en cualquier dirección h: la integral es la variación de la energía de la cuerda (elástica más gravitatoria) y los dos sumandos sueltos son las variaciones de las energías de los muelles. Suponiendo u dos veces diferenciable, una integración por partes del primer término, $\int_0^L\mu u'h'\,dx=\mu u'(L)h(L)-\mu u'(0)h(0)-\int_0^L\mu u''h\,dx$, lleva a

$$\int_0^L(-\mu u''+g)h\,dx+\big(ku(0)-\mu u'(0)\big)h(0)+\big(ku(L)+\mu u'(L)\big)h(L)=0,$$
$$\text{para toda } h\in C^1([0,L]).$$

En particular, para las h con $h(0)=h(L)=0$ queda $\int_0^L(-\mu u''+g)h\,dx=0$, y como en el teorema de Euler–Lagrange, el lema fundamental da la EDO $\mu u''=g$ en $(0,L)$. Sustituyendo la EDO en la identidad anterior,

$$\big(ku(0)-\mu u'(0)\big)h(0)+\big(ku(L)+\mu u'(L)\big)h(L)=0\qquad\text{para toda } h\in C^1([0,L]),$$

y tomando h con $h(0)=1$, $h(L)=0$, y después $h(0)=0$, $h(L)=1$, se obtiene que u es solución de

$$\begin{cases}\mu u''=g & \text{en } (0,L),\\ ku(0)-\mu u'(0)=0,\quad ku(L)+\mu u'(L)=0.\end{cases}$$

Las condiciones de contorno son **de tipo Robin** (o de tipo Fourier): una combinación del valor y de la derivada en cada extremo. La solución es un arco de parábola,

$$u(x) = -\frac{g}{2\mu}x(L-x) - \frac{gL}{2k}.$$

Se comprueba: $u'' = g/\mu$; $u(0) = u(L) = -gL/(2k)$; $u'(0) = -gL/(2\mu)$ y $u'(L) = gL/(2\mu)$, luego $ku(0) - \mu u'(0) = -gL/2 + gL/2 = 0$ y $ku(L) + \mu u'(L) = -gL/2 + gL/2 = 0$. La solución es la del ejemplo de la cuerda bajo gravedad *desplazada hacia abajo* en $gL/(2k)$: cada muelle sostiene la mitad del peso gL y se estira lo que dicta Hooke. Como era de esperar, si $k \to +\infty$ el muelle se hace tan potente que los extremos no se desplazan y se recupera la solución con extremos fijos. La suficiencia y la unicidad se comprueban juntas: para cualquier $h \in C^1([0, L])$, la primera variación se anula y el desarrollo exacto deja

$$E(u+h) - E(u) = \frac{\mu}{2}\int_0^L h'^2\,dx + \frac{k}{2}\left(h(0)^2 + h(L)^2\right) \geq 0.$$

La igualdad obliga a $h' = 0$ y $h(0) = h(L) = 0$, luego a $h = 0$: el mínimo es único.

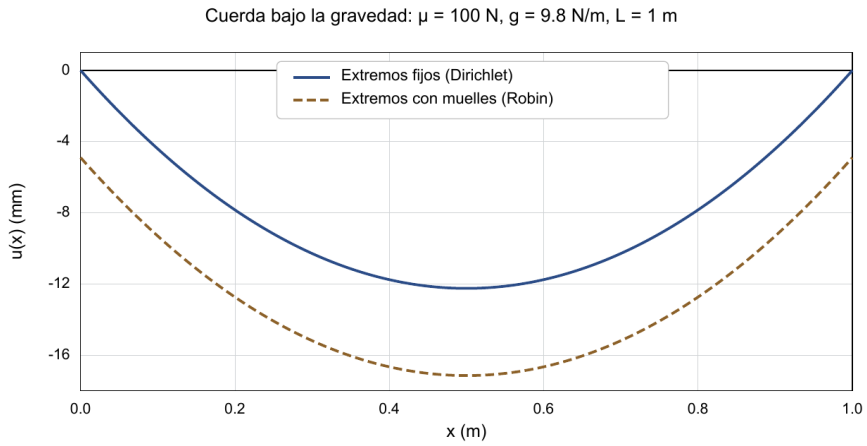


Figura 3.4: Cuerda con tensión $\mu = 100$ N, peso lineal $g = 9.8$ N/m y $L = 1$ m. Con extremos fijos, la flecha es 12.25 mm; con muelles de constante $k = 1000$ N/m, el perfil baja 4.9 mm. En ambos casos máx $|u'| = 0.049$: las pendientes son pequeñas, como exige el modelo.

LA IDEA: LOS TRES TIPOS DE CONDICIÓN DE CONTORNO, EN UNA CUERDA

Dirichlet ($u(0) = 0$): el extremo está clavado; se impone la posición y la cuerda «devuelve» una fuerza, $\mu u'(0)$, que es la reacción del clavo. Neumann ($u'(0) = 0$): el extremo está libre, nadie tira de él, y la cuerda llega horizontal; la posición la elige la cuerda. Robin ($ku(0) - \mu u'(0) = 0$): el extremo está sujeto a un muelle, y la fuerza que hace la cuerda, $\mu u'(0)$, es la que hace el muelle, $ku(0)$: equilibrio de fuerzas en el extremo. Los tres son el mismo esquema, y Robin contiene a los otros dos como límites: $k \to \infty$ es Dirichlet, $k = 0$ es Neumann. Esto es un límite de las condiciones; no garantiza un límite finito de las soluciones. Bajo el peso $gL > 0$, si ambos muelles desaparecen no hay equilibrio: $\mu u'' = g$ no puede satisfacer $u'(0) = u'(L) = 0$, y el desplazamiento $-gL/(2k)$ diverge cuando $k \to 0^+$. En la práctica 16 hay un muelle en un extremo y un clavo en el otro, y las dos condiciones salen exactamente así.

EXPLÍCATE. *En el problema con muelles no hay ninguna condición de contorno en el enunciado y salen dos. ¿De dónde sale cada una, y qué habría pasado con el recuento de condiciones si la EDO fuera de cuarto orden, como la viga de la práctica 19?*

CONEXIÓN: DEL EQUILIBRIO DE LA CUERDA A SUS ONDAS

La energía elástica que acaba de usarse es también la parte potencial del modelo vibratorio. Para una cuerda homogénea con densidad lineal $\rho > 0$, tensión constante $\mu > 0$ y pequeños desplazamientos transversales, la acción sin fuerza exterior es

$$\mathcal{A}[u] = \int_{t_0}^{t_1} \int_0^L \left(\frac{\rho}{2} u_t^2 - \frac{\mu}{2} u_x^2 \right) dx\, dt.$$

Ahora la incógnita depende de posición y tiempo. Si los extremos de la cuerda están fijos, las variaciones h se anulan en $x = 0, L$; para obtener la ecuación de movimiento se fijan también las configuraciones en t_0, t_1. Derivar $\mathcal{A}[u + \varepsilon h]$ e integrar por partes en ambas variables da

$$0 = \int_{t_0}^{t_1} \int_0^L (\rho u_t h_t - \mu u_x h_x)\, dx\, dt = \int_{t_0}^{t_1} \int_0^L (-\rho u_{tt} + \mu u_{xx}) h\, dx\, dt.$$

El lema fundamental, para soluciones suficientemente regulares, produce $\rho u_{tt} = \mu u_{xx}$. Es la ecuación de ondas: la misma energía espacial sirve para estudiar el equilibrio y, al añadir la energía cinética a la acción, el movimiento. La acción es estacionaria; no se está afirmando que toda solución la minimice.

Los modos normales recuperan además el problema de autovalores conocido. Al buscar $u(x,t) = q(t)\psi(x)$ se obtiene

$$-\psi'' = \lambda \psi, \quad \psi(0) = \psi(L) = 0, \qquad q'' + \frac{\mu}{\rho} \lambda q = 0.$$

Los modos no nulos son $\psi_n(x) = \operatorname{sin}(n\pi x/L)$, con $\lambda_n = (n\pi/L)^2$ y frecuencia angular $\omega_n = \sqrt{\mu/\rho}\, n\pi/L$. Su ortogonalidad permite separar las amplitudes, igual que los ejes principales separan una forma cuadrática en mecánica. Si se necesita una base ortonormal en $L^2(0,L)$, cada seno se multiplica por $\sqrt{2/L}$.

3.5 Prácticas de MATLAB del tema 3

LA PREGUNTA

En varios modelos de este tema se conoce el desplazamiento en los dos extremos, pero no la pendiente inicial que necesitaría `ode45`. ¿Cómo se resuelve numéricamente ese problema de contorno, y qué pasa si tiene dos soluciones o ninguna? Compararemos el método de tiro, que busca el dato inicial que falta, con la resolución simultánea en todo el intervalo.

Esta sección resuelve los ejercicios 14 a 23 de la hoja del tema 3. El 14 y el 18 son problemas de Cauchy y se resuelven con `ode45`; los demás son problemas de contorno y usan la orden nueva del tema, `bvp4c`, que se presenta en el ejercicio 15. Los ficheros que acompañan a la hoja (`DosParticulas.m`, `VigaElastica.m`, `VibForzadas.m`, `VigaOrden4.m`, `Bratu.m`) se siguen; los que faltan (`VigaElasticaF.m`, `VigaElasticaNH.m`, `Autofun.m`) se escriben aquí.

Ejercicio 14: dos partículas que se atraen

Ejercicio 3.4. *Dos partículas bajo la gravedad terrestre y su atracción mutua*

ORIGEN: hoja de ejercicios del tema 3, ejercicio 14, con el fichero `DosParticulas.m` del curso.

Dos partículas A y B, de masas m_A y m_B, están sometidas a la atracción gravitatoria terrestre y a la atracción mutua: por la ley de la gravitación universal, la fuerza entre ambas es proporcional al producto de las masas e inversamente proporcional al cuadrado de la distancia, y apunta a la otra partícula. Deducir las ecuaciones del movimiento por el principio de mínima acción y escribir una M-función `DosParticulas(mA, mB, g, G, y0, yp0, T)` que las resuelva en $[0, T]$ con $x(0) = (0, 0, 1000)$, $x'(0) = 0$, $y(0) = y_0$, $y'(0) = y_0'$, y represente las órbitas en el espacio y las trayectorias. Ejecutarla con `DosParticulas(1, 1, 9.8, 3e4, [100;0;1000], [0;1;0], 40)`.

Resolución.
La idea: reunir las fuerzas en una energía potencial. El principio de acción estacionaria permite obtener las seis ecuaciones del movimiento a partir de las energías cinética y potencial. La ventaja aquí es organizar los signos de la interacción: **las fuerzas se recuperan como menos el gradiente del potencial.**
Las dos interacciones son conservativas, de modo que podemos usar la lagrangiana $\mathcal{L} = E_c - E_p$. La gravedad terrestre viene de mgx_3, cuyo gradiente respecto de x es mge_3 de una sola derivada; la atracción mutua viene de $\phi = -Gm_Am_B/|x-y|$, y ahí sí hay que derivar una norma, por lo que se comprueba en el texto. Fíjese en que ese potencial es **uno solo para las dos partículas**: la misma función ϕ da las fuerzas $-\nabla_x\phi$ sobre A y $-\nabla_y\phi$ sobre B. Que salgan iguales y opuestas —la tercera ley de Newton— no hay que imponerlo: sale de que ϕ dependa de x e y solo a través de la diferencia $x - y$.
Una predicción que sirve como control independiente del programa: las dos partículas sienten *la misma* aceleración de la gravedad, $-ge_3$, independientemente de su masa. Luego la diferencia $x - y$ no ve la gravedad en absoluto, y si además G fuera cero, las dos caerían exactamente igual y su separación seguiría una recta.
Paso 1 (el núcleo): las dos fuerzas derivan de un potencial. Sean $x(t), y(t) \in \mathbb{R}^3$ las posiciones. Sobre A actúan la gravedad, $G_A = -m_Age_3 = -\nabla_x(m_Agx_3)$, y la atracción de B,

$$F_A = \frac{Gm_Am_B}{|y-x|^2}\,\frac{y-x}{|y-x|} = -\frac{Gm_Am_B}{|x-y|^3}(x-y) = -\nabla_x\phi, \qquad \phi = -\frac{Gm_Am_B}{|x-y|}.$$

(Comprobación, con la regla de la cadena: como $|x-y|^2 = (x-y)\cdot(x-y)$ tiene gradiente $2(x-y)$ respecto de x, se tiene $\nabla_x|x-y| = (x-y)/|x-y|$; luego $\nabla_x|x-y|^{-1} = -|x-y|^{-2}\,\nabla_x|x-y| = -|x-y|^{-3}(x-y)$, y $-\nabla_x\phi = Gm_Am_B\cdot(-|x-y|^{-3}(x-y))$, que es F_A.)
Sobre B, $G_B = -m_Bge_3$ y $F_B = -F_A = -\nabla_y\phi$.
Paso 2: montar el lagrangiano y aplicar Euler-Lagrange. Las energías son

$$E_c = \frac{1}{2}\big(m_A|x'|^2 + m_B|y'|^2\big), \qquad E_p = -\frac{Gm_Am_B}{|x-y|} + m_Agx_3 + m_Bgy_3,$$

y la lagrangiana $\mathcal{L} = E_c - E_p$. Euler-Lagrange, $-\frac{d}{dt}\nabla_{(x',y')}\mathcal{L} + \nabla_{(x,y)}\mathcal{L} = 0$, se escribe

$$-\frac{d}{dt}\begin{pmatrix} m_A x' \\ m_B y' \end{pmatrix} + \begin{pmatrix} -\dfrac{Gm_A m_B}{|x-y|^3}(x-y) \\ \dfrac{Gm_A m_B}{|x-y|^3}(x-y) \end{pmatrix} - \begin{pmatrix} m_A g e_3 \\ m_B g e_3 \end{pmatrix} = 0,$$

o, por separado y dividiendo por las masas,

$$x'' = -\frac{Gm_B}{|x-y|^3}(x-y) - ge_3, \qquad y'' = \frac{Gm_A}{|x-y|^3}(x-y) - ge_3.$$

Paso 3: pasar a primer orden para el integrador. El sistema es de segundo orden y con seis incógnitas, así que el vector de estado tiene doce componentes, $z = (x, x', y, y')$, y eso es lo único que hay que preparar antes de teclear.
Órdenes nuevas. Cuando el segundo miembro es largo, en vez de una función anónima se escribe una **función auxiliar** al final del mismo fichero, y se pasa a `ode45` con un manejador, `@dzdt`. Para que esa función vea los parámetros g, G, m_A, m_B sin pasárselos como argumentos, el fichero de partida los declara `global` en las dos funciones: una variable global es visible en toda función que la declare. Es una práctica que se usa mucho en los ficheros de la asignatura, aunque hoy se prefiere pasar los parámetros a través de la función anónima, `@(t,z) dzdt(t,z,g,G,mA,mB)`. Para dibujar en tres dimensiones, `plot3(x, y, z)` es `plot` con una coordenada más, `zlabel` rotula el tercer eje y `view(az, el)` fija el punto de vista con dos ángulos. `quiver3` dibuja un campo de vectores y `plotyy` dibuja dos curvas con dos escalas verticales distintas; las dos aparecen en figuras complementarias. `norm(z(1:3))` es el módulo de un trozo del vector de estado.
La versión siguiente detiene la integración mediante eventos al llegar al suelo: seleccionar el último nodo con altura positiva, como hace el fichero original, solo aproxima ese instante y puede integrar innecesariamente bajo tierra. También se detiene si la separación baja de 10^{-6} de la inicial; ese umbral numérico evita continuar hasta la singularidad $x = y$ y *no* representa una colisión de cuerpos de tamaño finito. Se requieren masas positivas, $g, G \geq 0$, $T > 0$ y posiciones iniciales distintas por encima del suelo.

```
function DosParticulas(mA, mB, g, G, y0, yp0, T)
% DosParticulas: dos particulas bajo la gravedad y su atraccion mutua.
% x'' = -G*mB/|x-y|^3 (x-y) - g e3, y'' = G*mA/|x-y|^3 (x-y) - g e3
% Estado z = [x; x'; y; y'] (12 componentes)
% Prueba: DosParticulas(1, 1, 9.8, 3e4, [100;0;1000], [0;1;0], 40)
global gg GG MA MB
gg = g; GG = G; MA = mA; MB = mB;

x0 = [0; 0; 1000]; xp0 = zeros(3,1);
y0 = y0(:); yp0 = yp0(:);
assert(numel(y0)==3 && numel(yp0)==3, 'Vectores de tres componentes');
assert(mA>0 && mB>0 && g>=0 && G>=0 && T>0 && y0(3)>0);
r0 = norm(x0-y0); assert(r0>0, 'Posiciones inicialmente distintas');
rmin = 1e-6*r0;
z0 = [x0; xp0; y0; yp0];
opts = odeset('RelTol',1e-9,'AbsTol',1e-11, ...
    'Events',@(t,z) parada(t,z,rmin));
[times,z,te,ze,ie] = ode45(@dzdt,[0,T],z0,opts);
```

```
if isempty(te)
    assert(times(end)>=T, 'El integrador no alcanzo T');
    fprintf('No se alcanza el suelo en [0,%g]\n',T);
elseif ie(end)==1
    fprintf('Primer contacto con el suelo: t = %.9g\n',te(end));
else
    warning('Integracion detenida por proximidad entre particulas');
end
pos = 1:numel(times);

close all
figure(1) % orbitas en 3D
plot3(x0(1), x0(2), x0(3), 'b.', 'MarkerSize', 20)
hold on
plot3(y0(1), y0(2), y0(3), 'r.', 'MarkerSize', 20)
ha = plot3(z(pos,1), z(pos,2), z(pos,3), 'b')
hb = plot3(z(pos,7), z(pos,8), z(pos,9), 'r')
legend([ha,hb], 'Particula A', 'Particula B')
xlabel('X'), ylabel('Y'), zlabel('Z'), title(' Orbitas de las dos particulas')
grid on, view(140, 30), hold off

figure(2) % proyeccion sobre OXY
plot(x0(1), x0(2), 'b.', 'MarkerSize', 20), hold on
plot(y0(1), y0(2), 'r.', 'MarkerSize', 20)
ha = plot(z(pos,1), z(pos,2), 'b'); hb = plot(z(pos,7), z(pos,8), 'r');
legend([ha,hb], 'Trayect. A', 'Trayect. B'), xlabel('X'), ylabel('Y'), hold off
shg
end

function zp = dzdt(t, z)
% Segundo miembro del sistema de 12 ecuaciones
global gg GG MA MB
dif = z(1:3) - z(7:9);
Gdif3 = GG/norm(dif)^3;
zp = zeros(size(z));
zp(1:3) = z(4:6);
zp(4:6) = -Gdif3*MB*dif;
zp(7:9) = z(10:12);
zp(10:12) = Gdif3*MA*dif;
zp(6) = zp(6) - gg; % la gravedad, solo en la tercera componente
zp(12) = zp(12) - gg;
end

function [value,isterminal,direction] = parada(t,z,rmin)
value = [min(z(3),z(9)); norm(z(1:3)-z(7:9))-rmin];
isterminal = [1;1]; direction = [-1;-1];
end
```

SALIDA ESPERADA

```
>> DosParticulas(1, 1, 9.8, 3e4, [100;0;1000], [0;1;0], 40)
(figura 1: las dos orbitas caen desde z = 1000 girando una alrededor de la otra;
 tocan el suelo hacia t = 14.3 s)
```

Con $G = 3 \cdot 10^4$ y masas unidad la atracción es del orden de $G/100^2 = 3\ \mathrm{m/s^2}$, comparable a g: las dos partículas giran una en torno a la otra mientras caen, y llegan al suelo a los 14.2857

s, que es *exactamente* la caída libre $\sqrt{2 \cdot 1000/9.8}$. Y no es una aproximación: restando las terceras componentes de las dos ecuaciones, $(x_3 - y_3)'' = -\frac{G(m_A+m_B)}{|x-y|^3}(x_3 - y_3)$, que es lineal y homogénea en $x_3 - y_3$; como las dos partículas salen de la misma altura y sin velocidad vertical, $x_3 - y_3$ y su derivada valen cero en $t = 0$, luego $x_3 \equiv y_3$ mientras existe la solución. El vector $x - y$ se mantiene horizontal, la atracción mutua no tiene nunca componente vertical, y la altura común obedece $x_3'' = -g$ sin más: toda la interacción vive en el plano OXY. Con $G = 10^3$ apenas se desvían; si se aumenta m_B a 5, el desplazamiento de A respecto del centro de masas es cinco veces el de B debido a la interacción: ambas se mueven, con amplitudes inversas a sus masas.

La atracción se vuelve muy intensa en los acercamientos. Para los datos del enunciado, $r = x - y$ cumple $r'' = -G(m_A + m_B)r/|r|^3$, y se conservan $\frac{1}{2}|r'|^2 - G(m_A + m_B)/|r|$ y $r \times r'$. Comprobar su deriva y repetir con menor tolerancia ayuda a distinguir la órbita de los errores del integrador. La distancia mínima es aproximadamente 0.0834 m, mucho menor que los 100 m iniciales: un paso adecuado lejos de ese punto puede resultar demasiado grande cerca de él.

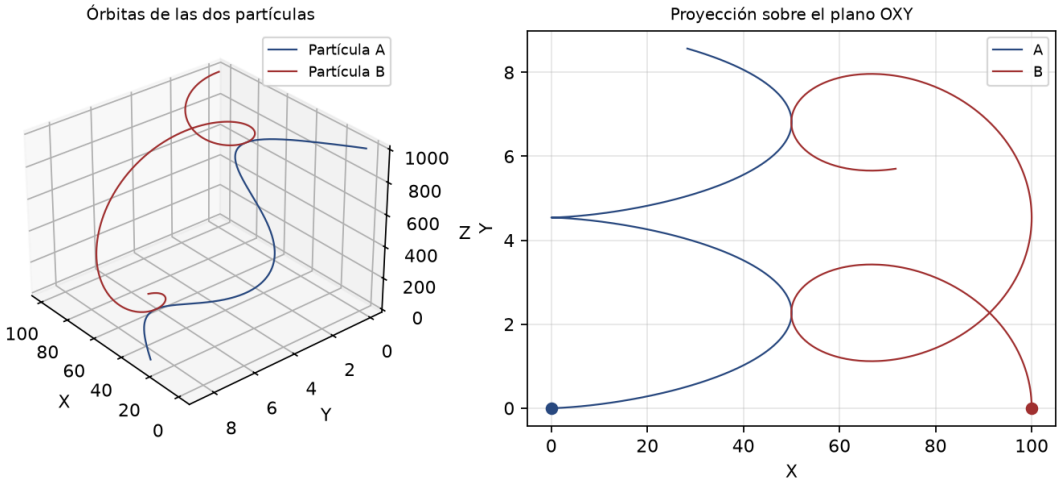


Figura 3.5: Las dos partículas con $G = 3 \cdot 10^4$: órbitas en el espacio y su proyección sobre el plano horizontal. Giran una alrededor de la otra mientras caen.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_ivp
def dos_particulas(mA, mB, g, G, y0, yp0, T):
    y0 = np.asarray(y0, dtype=float).reshape(3)
    yp0 = np.asarray(yp0, dtype=float).reshape(3)
    r0 = np.linalg.norm(np.array([0,0,1000.0])-y0)
    if not (mA>0 and mB>0 and g>=0 and G>=0 and T>0 and y0[2]>0 and r0>0):
        raise ValueError("Datos fuera del dominio indicado")
    z0 = np.r_[[0, 0, 1000.0], np.zeros(3), y0, yp0]
    def dzdt(t, z):
        dif = z[0:3] - z[6:9]; Gd3 = G / np.linalg.norm(dif) ** 3
        zp = np.zeros(12)
        zp[0:3] = z[3:6]; zp[3:6] = -Gd3 * mB * dif; zp[6:9] = z[9:12]; zp[9:12] = Gd3 *
```

```
            mA * dif
        zp[5] -= g; zp[11] -= g
        return zp
    def suelo(t,z): return min(z[2],z[8])
    def proximidad(t,z): return np.linalg.norm(z[:3]-z[6:9])-1e-6*r0
    suelo.terminal = proximidad.terminal = True
    suelo.direction = proximidad.direction = -1
    s = solve_ivp(dzdt,[0,T],z0,events=(suelo,proximidad),
                  rtol=1e-9,atol=1e-11,max_step=0.05)
    if not s.success: raise RuntimeError(s.message)
    if len(s.t_events[0]): print("Primer contacto con el suelo:",s.t[-1])
    elif len(s.t_events[1]): print("Parada por proximidad entre particulas")
    else: print("No se alcanza el suelo en [0,T]")
    return s.t, s.y.T
t, z = dos_particulas(1, 1, 9.8, 3e4, np.array([100, 0, 1000.0]), np.array([0, 1, 0.0]),
    40)
# t[-1] es el contacto, el umbral de proximidad o T, segun el mensaje.
```

En Python no hace falta `global`: la función interior `dzdt` ve las variables de la función que la contiene. Es el mismo mecanismo que en MATLAB se consigue anidando funciones o con una anónima que llame a la auxiliar con los parámetros.

Qué hemos aprendido y dónde se falla. El ejercicio enseña a montar un sistema de varias partículas desde un solo potencial y a pasarlo a primer orden. Los fallos están en tres sitios.

- *En el signo de la atracción.* Con `zp(4:6) = +Gdif3*MB*dif`, y el signo opuesto en `zp(10:12)`, las partículas se repelen: el programa corre igual, y al tocar el suelo están a unos 415 m una de otra en lugar de girar juntas. La comprobación es de una línea: `dif` es $x - y$, que apunta de B hacia A, y la fuerza sobre A tiene que apuntar hacia B, es decir, llevar el signo menos.
- *Al fiarse de un control que no mira donde está el error.* El centro de masas cumple $(m_A x + m_B y)'' = -(m_A + m_B) g e_3$, porque las fuerzas mutuas se cancelan, y con los datos del enunciado tiene que tocar el suelo en $(50,\ 0.5\,t_c,\ 0) = (50,\ 7.1429,\ 0)$, con $t_c = 14.2857$. Ese control detecta un signo mal puesto en la gravedad o una masa cambiada. No detecta el error de integración en el acercamiento, porque la cancelación de las fuerzas mutuas también es exacta en el método numérico.
- *En las tolerancias.* Con las tolerancias por defecto (en la versión en Python, `rtol`$= 10^{-3}$), el acercamiento a 0.083 m se integra mal: la energía del movimiento relativo, $\frac{1}{2}|r'|^2 - G(m_A + m_B)/|r|$, pasa de -599.5 a un valor positivo, así que la pareja ligada acaba separándose, y en el suelo A está a más de cien metros de donde la pone el cálculo fino. El centro de masas sale igual de exacto en los dos cálculos. Con `rtol`$= 10^{-6}$ el desvío ya baja a unos 6 cm. Por eso el programa fija tolerancias finas, y por eso el control que hay que mirar aquí es la energía relativa, que sí ve el error.

La orden `bvp4c`

Un problema de contorno para un sistema de primer orden $z' = F(x,z)$ en $[a,b]$, con condiciones $g(z(a),z(b)) = 0$, se resuelve con

```
sol = bvp4c(F, condcon, solinit, options)
```

donde `F = @(x,z) [...]` es el segundo miembro, como en `ode45`; `condcon = @(za, zb) [...]` es una función de los valores en los dos extremos que devuelve un vector con *tantas componentes como ecuaciones*, y que debe anularse (una componente por condición de contorno); `solinit = bvpinit(xmalla, zinicial)` es una **aproximación inicial** de la solución sobre una malla, que el método necesita porque resuelve un sistema no lineal por iteraciones y hay que decirle desde dónde empezar; y `options = bvpset('Stats','on','RelTol',1e-5)` fija opciones, como pedir estadísticas o la tolerancia. El resultado `sol` es una estructura con los campos `sol.x` (la malla final) y `sol.y` (la solución sobre ella, **una fila por componente**, al revés que `ode45`); para evaluar en una rejilla cualquiera se usa `deval(sol, xx, k)`, que devuelve la componente k en los puntos `xx`.

LA IDEA: POR QUÉ HACE FALTA UNA APROXIMACIÓN INICIAL

`ode45` construye una solución a partir del estado inicial completo. `bvp4c` busca una función que cumpla a la vez la ecuación y las condiciones en ambos extremos. Discretiza el intervalo por colocación y resuelve un sistema algebraico, en general no lineal, mediante iteraciones de tipo Newton. El punto de partida influye en la convergencia y en qué solución se encuentra, pero Newton no elige necesariamente el cero más cercano. En las vigas lineales de estas prácticas basta una aproximación nula; en Bratu probaremos aproximaciones de distintas amplitudes. Un fallo numérico puede deberse a una mala aproximación inicial o al condicionamiento: por sí solo no demuestra que no exista solución.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
from scipy.integrate import solve_bvp
# solve_bvp(F, condcon, xmalla, zinicial): F(x, z) devuelve un array con una FILA por componente
# (z tiene forma (n, puntos)); condcon(za, zb) igual que en MATLAB; zinicial de forma (n, puntos).
# sol.sol(xx) evalua la solucion (todas las componentes) en xx; sol.status == 0 si convergio.
```

Ejercicio 15: viga elástica con extremos fijos

Ejercicio 3.5. *Desplazamiento de una viga elástica*

ORIGEN: hoja de ejercicios del tema 3, ejercicio 15, con el fichero `VigaElastica.m` del curso.

Una viga ocupa $(0,L)$, tiene rigidez $k > 0$, está sometida a la fuerza vertical $f(x) = f_0 \sin(n\pi x/L)$ y sus extremos se desplazan $u(0) = \alpha$, $u(L) = \beta$. Su desplazamiento minimiza

$$\int_0^L \Big(\frac{k}{2}|v'|^2 - f(x)v\Big)dx \quad \text{sobre } v \in C^1([0,L]),\ v(0) = \alpha,\ v(L) = \beta,$$

y por tanto es solución de $-ku'' = f(x)$, $u(0) = \alpha$, $u(L) = \beta$. Escribir `VigaElastica(kappa, L, alpha, beta, f0, nforce)` que resuelva ese problema de contorno y dibuje u; ejecutarla con `VigaElastica(.5, 1, 0, 0, -1, 2)`.

Resolución.
La idea, y la decisión de método es disparar o no disparar. Un problema de contorno no es un problema de valor inicial: no se conoce el estado completo en un extremo, sino un dato en cada uno, así que no se puede integrar hacia delante sin más. Hay dos maneras de arreglarlo.
La primera es el **tiro**: inventar el dato que falta —aquí $u'(0)$—, integrar con `ode45` hasta $x = L$, ver cuánto se falla en $u(L) = \beta$ y corregir el disparo con `fzero`. Es cómoda de programar porque reutiliza lo del capítulo 2, y tiene un defecto que se verá en la viga de cuarto orden: si el sistema inicial tiene modos crecientes, puede amplificar exponencialmente los errores de integración y del dato inicial. Entonces el valor final es demasiado sensible al disparo y corregirlo con precisión resulta difícil. Esto no ocurre en todos los problemas de contorno.
La segunda es **resolver todo a la vez**, que es lo que hace `bvp4c`: propone una solución sobre una malla entera, impone la ecuación en cada punto y las dos condiciones en los extremos, y resuelve por Newton el sistema no lineal resultante. Al tratar ambos extremos a la vez evita esa propagación del error propia del tiro simple, aunque el problema de contorno o su sistema discretizado también pueden estar mal condicionados. Es el método que usa este capítulo y el siguiente, y por eso conviene aprender su patrón de una vez aquí, en el problema más sencillo posible —lineal, con solución exacta conocida— antes de usarlo donde no se pueda comprobar.
El núcleo del programa son tres líneas: `dzdx`, que es la ecuación reducida a primer orden; `condcon`, que es el vector de condiciones que debe anularse; y la llamada a `bvp4c`. La cuarta línea que importa es `solinit`, la aproximación inicial, que aquí puede tomarse nula: el sistema es lineal y su solución es única. En Bratu, en cambio, influirá en qué solución se obtiene y en si las iteraciones convergen.
Análisis teórico. Con $H(v, v') = \frac{k}{2}|v'|^2 - f(x)v$, $\partial_{v'}H = kv'$ y $\partial_v H = -f$, y la condición de Euler-Lagrange es $-(ku')' - f = 0$, es decir, $-ku'' = f$. Para $L > 0$ y $n \in \{1, 2, \dots\}$, la solución exacta se obtiene integrando dos veces: $u'' = -\frac{f_0}{k}\sin\frac{n\pi x}{L}$, luego $u = \frac{f_0 L^2}{k n^2 \pi^2}\sin\frac{n\pi x}{L} + c_1 x + c_2$, y las condiciones de contorno dan la recta $c_1 x + c_2 = \alpha + (\beta - \alpha)x/L$ porque el seno se anula en 0 y L. Con $k = 0.5$, $f_0 = -1$, $n = 2$, $L = 1$: amplitud $-1/(0.5 \cdot 4\pi^2) \approx -0.0507$.
El programa. Sigue el patrón de `bvp4c`: sistema de primer orden, condiciones en los extremos, aproximación inicial nula, resolución, evaluación con `deval` y dibujo. La opción `'Stats','on'` hace que `bvp4c` escriba cuántos puntos de malla ha usado y el error residual.

```
function VigaElastica(kappa, L, alpha, beta, f0, nforce)
% VigaElastica: -k u'' = f(x) en (0,L), u(0) = alpha, u(L) = beta,
% con f(x) = f0*sin(nforce*pi*x/L). Resuelto con bvp4c.
% Pruebas: VigaElastica(.5,1,0,0,-1,1) VigaElastica(1,1,0,0,-5,1) VigaElastica
    (1,1,0,.2,5,5)
assert(kappa>0 && L>0 && nforce>=1 && nforce==fix(nforce));
unk = -1/kappa;
coefsin = nforce*pi/L;
fsin = @(x) f0*sin(coefsin*x);
```

```
9
10 dzdx = @(x,z) [z(2); unk*fsin(x)]; % z = [u; u']
11 condcon = @(za,zb) [za(1) - alpha; zb(1) - beta]; % u(0) = alpha, u(L) = beta
12
13 xinit = linspace(0, L, 10); % malla inicial
14 zinit = [0; 0]; % aproximacion inicial: la viga sin deformar
15 solinit = bvpinit(xinit, zinit);
16 options = bvpset('Stats', 'on');
17
18 sol = bvp4c(dzdx, condcon, solinit, options); % resolucion
19
20 close all
21 xx = linspace(0, L);
22 zz = deval(sol, xx, 1); % primera componente: u
23 hold on
24 plot([0, L], [alpha, beta], 'k', 'LineWidth', 3)
25 plot(xx, zz, 'b', 'LineWidth', 3)
26 legend('Posicion sin deformar', 'Posicion deformada')
27 xlabel('Abscisa x'), ylabel('Desplazamiento u')
28 title('Desplazamiento de una viga elastica'), xlim([0,L]), grid on, hold off
29 shg
30 end
```

SALIDA ESPERADA

```
>> VigaElastica(.5, 1, 0, 0, -1, 2)
(bvp4c informa de la malla, del residuo y del numero de evaluaciones)
(figura: una onda completa de amplitud 0.05, negativa en (0, 0.5) y positiva en (0.5, 1))
```

Las estadísticas varían con la versión y las tolerancias. Aquí disponemos además de un control independiente: comparar con la solución exacta.

Conviene distinguir el **residuo**, que mide cuánto deja de cumplirse la ecuación, del **error**, que mide cuánto se aparta la aproximación de la solución verdadera. No son la misma magnitud ni tienen por qué compartir unidades. Por ejemplo, si $\tilde{u}$ cumple los datos de contorno y $r = -k\tilde{u}'' - f$ es su residuo sin normalizar, entonces

$$\|\tilde{u} - u\|_\infty \leq \frac{L^2}{8k}\|r\|_\infty.$$

Se obtiene comparando el error con las parábolas $\pm\|r\|_\infty x(L-x)/(2k)$ mediante el principio del máximo. Si los datos de contorno también tienen error, se añade el mayor error de los dos extremos. El factor $L^2/(8k)$ muestra por qué un residuo pequeño solo se traduce en un error pequeño al tener en cuenta el condicionamiento. Además, `bvp4c` controla un residuo escalado; la cifra de `Stats` no se sustituye sin más en esta cota para el residuo escalar sin normalizar.

Los controles, y hay tres, cada uno más barato que el anterior. El primero es la solución exacta, que aquí existe: $u = f_0L^2/(kn^2\pi^2)\sin(n\pi x/L)$ más la recta de los datos de contorno, con amplitud -0.050661 para los valores del enunciado. El segundo, cuando no haya exacta, es evaluar la ecuación sobre la solución calculada, que es lo que hace `Stats`. Y el tercero cuesta cero y hay que hacerlo siempre: **mirar si la solución cumple las condiciones de contorno**, porque `bvp4c` las impone pero un error al escribir `condcon` —cambiar `za` por `zb`, o `z(1)` por `z(2)`— produce un programa que converge sin protestar a la solución de *otro* problema. Es un error que el integrador no puede reconocer: solo conoce las condiciones que hemos

escrito.

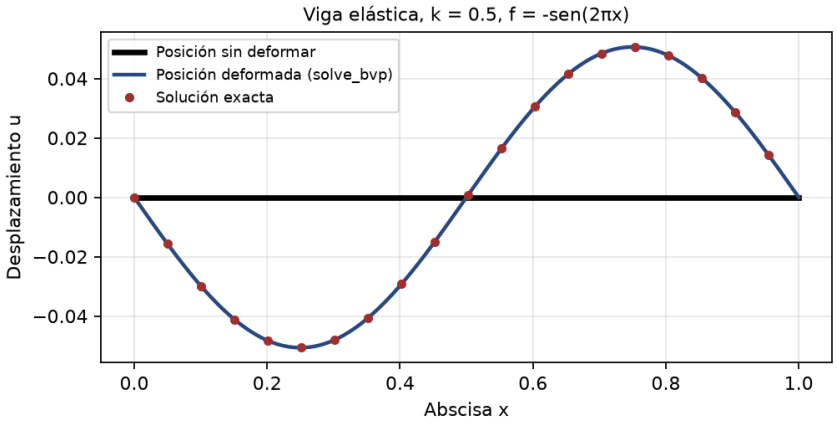


Figura 3.6: Viga con $k = 0.5$ y fuerza $-\sin(2\pi x)$: el desplazamiento es una onda de amplitud 0.0507 con la forma de la propia fuerza y su mismo signo en cada punto, porque la solución exacta es $u = f\,L^2/(kn^2\pi^2)$ y ese factor es positivo. Los puntos son la solución exacta.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np, matplotlib.pyplot as plt
from scipy.integrate import solve_bvp
def viga_elastica(kappa, L, alpha, beta, f0, nforce):
    if not (kappa>0 and L>0 and nforce>=1 and int(nforce)==nforce):
        raise ValueError("Se requiere kappa,L positivos y nforce entero positivo")
    fsin = lambda x: f0 * np.sin(nforce * np.pi * x / L)
    dzdx = lambda x, z: np.vstack([z[1], -fsin(x) / kappa]) # una fila por componente
    condcon = lambda za, zb: np.array([za[0] - alpha, zb[0] - beta])
    xi = np.linspace(0, L, 10)
    sol = solve_bvp(dzdx, condcon, xi, np.zeros((2, xi.size))) # aproximacion inicial
        nula
    if not sol.success: raise RuntimeError(sol.message)
    xx = np.linspace(0, L, 200); uu = sol.sol(xx)[0]
    plt.plot([0, L], [alpha, beta], "k", lw=3); plt.plot(xx, uu, "b", lw=2); plt.show()
    return xx, uu
viga_elastica(0.5, 1, 0, 0, -1, 2)
```

Dos limitaciones que conviene reconocer. La primera es del dibujo. El fichero original fija `axis([0,L,-0.4,0.4])`; con $k = 1$, $f_0 = -5$ y $n = 1$, la amplitud es $-5/\pi^2 = -0.5066$ y la curva queda recortada. La versión anterior deja que el eje vertical se adapte a los datos. Una meseta producida por el límite de la ventana no es una propiedad de la solución.
La segunda es de fondo y merece la pena hacerla. Si en lugar de fijar los desplazamientos en los dos extremos se dejan los dos libres —condiciones $u'(0) = u'(L) = 0$, es decir, una viga que flota sin sujeción—, el problema deja de tener solución única: sumar una constante a u no cambia nada, así que si hay una solución hay infinitas. Y además solo la hay si $\int_0^L f = 0$, que es la condición de que la fuerza total sea nula: una viga libre sometida a una fuerza neta no puede permanecer en equilibrio. Con $f = f_0 \sin(n\pi x/L)$ esa integral vale $f_0 L(1-\cos n\pi)/(n\pi)$, que, si $f_0 \neq 0$, es cero para n par y no para n impar. Al pedirle eso a `bvp4c` **no devuelve un aviso interpretable**: se detiene diciendo que ha encontrado una

jacobiana singular al resolver el sistema de colocación, tanto en el caso sin solución como en el caso con infinitas. En este ejemplo la singularidad refleja que las constantes están en el núcleo del operador. En general, ese mensaje también puede proceder de la formulación, de la aproximación inicial o del condicionamiento: no basta para concluir que el problema matemático no tiene solución única.

QUÉ DEJA. El patrón de `bvp4c`: reducir a primer orden, escribir las condiciones como un vector que debe anularse, dar una aproximación inicial y evaluar con `deval`. Y comprobar siempre con la solución exacta cuando la haya, porque proporciona una comprobación independiente del patrón numérico.

Ejercicio 16: un extremo sujeto a un muelle

Ejercicio 3.6. *Viga con un extremo sujeto a un muelle*

ORIGEN: hoja de ejercicios del tema 3, ejercicio 16. No hay fichero de partida; se escribe aquí.

La misma viga, pero ahora $u(L) = \beta$ y el extremo izquierdo está sujeto a un muelle de constante $m \geq 0$: u minimiza

$$E(v) := \int_0^L \Big(\frac{k}{2}|v'|^2 - fv\Big)dx + \frac{m}{2}|v(0)|^2 - \alpha v(0) \quad \text{sobre } v \in C^1([0,L]),\ v(L) = \beta.$$

Deducir el problema de contorno y escribir `VigaElasticaF(kappa, L, m, alpha, beta, f0, nforce)` que lo resuelva; ejecutarla con `VigaElasticaF(0.5, 1, 1, 0, 0, -0.2, 1)`.

Resolución.
La idea, en términos físicos. El funcional tiene cuatro sumandos y cada uno dice una aspecto del modelo, así que conviene leerlos antes de derivar nada: $\frac{k}{2}|v'|^2$ es la energía elástica que cuesta estirar la viga; $-fv$ es su energía potencial debida a la carga distribuida; $\frac{m}{2}|v(0)|^2$ es la energía del muelle que sujeta el extremo izquierdo; y $-\alpha v(0)$ es el potencial de una fuerza puntual aplicada en ese mismo extremo. Aunque la hoja usa la palabra «amortiguado», aquí hay una fuerza elástica proporcional al desplazamiento, no una fuerza disipativa proporcional a la velocidad. El extremo derecho está clavado, $v(L) = \beta$, y por eso allí no hay ningún término: no se le puede hacer trabajo a algo que no se mueve.
De esa lectura sale, sin cuentas, qué forma tendrá la respuesta. **Los términos que viven en el interior producen la ecuación diferencial; los que viven en un extremo producen la condición de contorno en ese extremo.** Los dos primeros sumandos dan $-ku'' = f$; los dos últimos, evaluados en 0, dan la condición en 0. Y esa condición es un *equilibrio de fuerzas* en el extremo izquierdo: tomando positivo hacia arriba, la viga tira con $ku'(0)$, el muelle ejerce $-mu(0)$ y la carga externa vale α. Su suma se anula porque el extremo está en equilibrio; es la segunda ley de Newton con aceleración cero.
Paso 1 (el núcleo): perturbar e integrar por partes. Sea $h \in C^1([0,L])$ con $h(L) = 0$. Como en el ejemplo de la cuerda con muelles elásticos, $F(\varepsilon) = E(u + \varepsilon h)$ tiene un mínimo en $\varepsilon = 0$, y

$$F'(0) = \int_0^L \big(ku'h' - fh\big)dx + mu(0)h(0) - \alpha h(0) = 0.$$

Integrando por partes el primer sumando ($\int_0^L ku'h' = [ku'h]_0^L - \int ku''h = -ku'(0)h(0) - \int ku''h$, porque $h(L) = 0$),

$$\int_0^L (-ku'' - f)h\,dx + \big(-ku'(0) + mu(0) - \alpha\big)h(0) = 0 \qquad \text{para toda } h \text{ con } h(L) = 0.$$

Paso 2: usar la arbitrariedad de h dos veces, primero dentro y luego en el borde. Con $h(0) = h(L) = 0$, el lema fundamental da $-ku'' = f$ en $(0, L)$; sabido eso, con $h(0) = 1$ queda $-ku'(0) + mu(0) = \alpha$. El problema de contorno es

$$\begin{cases} -ku'' = f(x) & \text{en } (0, L), \\ -ku'(0) + mu(0) = \alpha, \quad u(L) = \beta. \end{cases}$$

Paso 3: la solución exacta, que sirve para validar el programa. Tiene solución exacta: $u = A\sin(n\pi x/L) + c_1 x + c_2$ con $A = f_0 L^2/(kn^2\pi^2)$, y las dos condiciones dan

$$c_2 = \frac{k\beta + \alpha L + kAn\pi}{k + mL}, \qquad c_1 = \frac{\beta - c_2}{L}.$$

El denominador es positivo para $k, L > 0$, $m \geq 0$: la solución es única. También es el mínimo, pues la variación de segundo orden es $k\int |h'|^2 + m|h(0)|^2$ y solo se anula para $h = 0$ cuando $h(L) = 0$. Con los datos de la prueba, $u(0) = -0.0424$ y $u'(0) = -0.0849$, que cumplen $-0.5u'(0) + u(0) = 0$.

Qué hace cada dato. $m > 0$ es lo que hace que el extremo izquierdo esté *sujeto*: con $m = 0$ la condición se convierte en $-ku'(0) = \alpha$, un extremo libre con una fuerza aplicada, y con $m \to \infty$ y los demás datos fijos, la fórmula de c_2 da $u(0) \to 0$: se recupera un extremo clavado. El muelle interpola entre las dos condiciones clásicas, y ese es el sentido de una condición de Robin. Al aumentar α en $\Delta\alpha$, el perfil cambia en $\Delta\alpha(L - x)/(k + mL)$: sube el extremo izquierdo y el derecho permanece fijo. Si $\alpha = 0$, la condición de Robin es homogénea, pero eso no obliga a $u(0) = 0$.

```
function VigaElasticaF(kappa, L, m, alpha, beta, f0, nforce)
% VigaElasticaF: -k u'' = f en (0,L), -k u'(0) + m u(0) = alpha (Robin), u(L) = beta.
% Prueba: VigaElasticaF(0.5, 1, 1, 0, 0, -0.2, 1)
assert(kappa>0 && L>0 && m>=0 && nforce>=1 && nforce==fix(nforce));
fsin = @(x) f0*sin(nforce*pi*x/L);
dzdx = @(x,z) [z(2); -fsin(x)/kappa];
condcon = @(za,zb) [-kappa*za(2) + m*za(1) - alpha; zb(1) - beta]; % Robin en 0,
    Dirichlet en L

solinit = bvpinit(linspace(0, L, 10), [0; 0]);
sol = bvp4c(dzdx, condcon, solinit, bvpset('Stats', 'on'));

close all
xx = linspace(0, L);
uu = deval(sol, xx, 1);
hold on
plot([0, L], [uu(1), beta], 'k', 'LineWidth', 3)
plot(xx, uu, 'b', 'LineWidth', 3)
legend('Segmento entre extremos', 'Posicion deformada')
xlabel('Abscisa x'), ylabel('Desplazamiento u')
```

```
title('Viga con muelle en el extremo izquierdo'), hold off
fprintf('u(0) = %10.6f, u''(0) = %10.6f, -k u''(0) + m u(0) = %10.6f \n', ...
       uu(1), deval(sol, 0, 2), -kappa*deval(sol, 0, 2) + m*uu(1))
shg
end
```

SALIDA ESPERADA

```
>> VigaElasticaF(0.5, 1, 1, 0, 0, -0.2, 1)
u(0) = -0.042441, u'(0) = -0.084883, -k u'(0) + m u(0) = 0.000000
```

La comilla doble '' dentro de una cadena de `fprintf` escribe una comilla simple. El extremo izquierdo baja 0.042 porque el muelle cede; con m grande volvería a cero.

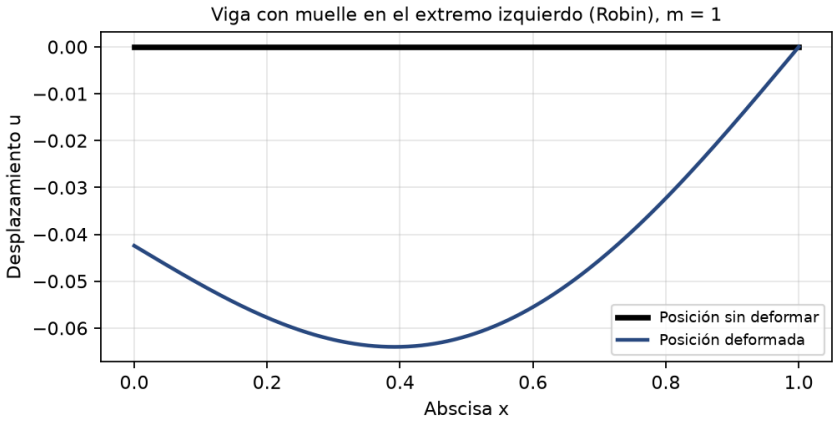


Figura 3.7: Viga con el extremo izquierdo sujeto a un muelle ($m = 1$) y el derecho fijo, bajo la fuerza $-0.2\sin(\pi x)$. El extremo izquierdo se desplaza lo que dicta la condición de Robin.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_bvp
def viga_F(kappa, L, m, alpha, beta, f0, nforce):
    if not (kappa>0 and L>0 and m>=0 and nforce>=1 and int(nforce)==nforce):
        raise ValueError("Datos fuera del dominio indicado")
    fsin = lambda x: f0 * np.sin(nforce * np.pi * x / L)
    dzdx = lambda x, z: np.vstack([z[1], -fsin(x) / kappa])
    bc = lambda za, zb: np.array([-kappa * za[1] + m * za[0] - alpha, zb[0] - beta])
    xi = np.linspace(0, L, 10); sol = solve_bvp(dzdx, bc, xi, np.zeros((2, xi.size)))
    if not sol.success: raise RuntimeError(sol.message)
    print("u(0) =", sol.sol(0)[0], " u'(0) =", sol.sol(0)[1])
viga_F(0.5, 1, 1, 0, 0, -0.2, 1)
```

Qué hemos aprendido y dónde se falla. Lo que se lleva el ejercicio es la lectura del funcional por sus sumandos: lo que vive en el interior da la ecuación, y lo que vive en un extremo da la condición de ese extremo. Los fallos están en esa condición.

- *En el signo de ku'.* El término de borde de la integración por partes es $[ku'h]_0^L$, que entra con signo menos en el extremo izquierdo y con signo más en el derecho. Por eso la condición es $-ku'(0) + mu(0) = \alpha$ aquí, y sería $ku'(L) + mu(L) = \alpha$ si el muelle estuviera a la derecha. Quien escriba `+kappa*za(2)` en `condcon` resuelve otro problema, y `bvp4c` converge sin

protestar: con los datos de la prueba sale $u(0) = +0.1273$ en lugar de -0.0424. El control físico lo delata enseguida: con la carga hacia abajo y sin fuerza en el extremo, el muelle no puede dejar ese extremo por encima de su posición de reposo.

- *Al comprobar con la misma expresión que se programó.* La línea que imprime $-ku'(0) + mu(0)$ solo sirve si se escribe con independencia de `condcon`. Si se copia de allí con el mismo signo equivocado, imprime cero igual. La comprobación independiente es la solución exacta del paso 3, deducida a mano, o el control físico anterior.

Ejercicio 17: viga no homogénea con apoyo elástico

Ejercicio 3.7. *Viga no homogénea con apoyo elástico distribuido*

ORIGEN: hoja de ejercicios del tema 3, ejercicio 17. No hay fichero de partida; se escribe aquí.

La viga tiene rigidez $k(x) = 1 + k_0 x^2$, $k_0 > 0$, está sometida a $f(x) = f_0 \sin(n\pi x/L)$, tiene una fuerza de restitución $-hu$ proporcional al desplazamiento, con $h \geq 0$, y $u(0) = \alpha$, $u(L) = \beta$. Minimiza

$$E(v) := \int_0^L \Big(\frac{k(x)}{2} |v'|^2 + \frac{h}{2} |v|^2 - f(x) v \Big) dx \quad \text{sobre } v \in C^1([0,L]),\ v(0) = \alpha,\ v(L) = \beta.$$

Determinar el problema de contorno y escribir `VigaElasticaNH(kcero, L, alpha, beta, f0, nforce, h)` que lo resuelva para h, $h/2$ y $h/4$ y dibuje las tres soluciones juntas.

Resolución.
La idea: distinguir rigidez y restitución. El término $\frac{k(x)}{2}|v'|^2$ penaliza la inclinación: cuanto mayor sea $k(x)$, más cuesta una misma pendiente. El término $\frac{h}{2}|v|^2$ penaliza el desplazamiento respecto de cero. Puede interpretarse como un conjunto de muelles repartidos bajo la viga; aunque la hoja lo llama «amortiguamiento», aquí no hay una fuerza dependiente de la velocidad ni disipación de energía.
En la prueba, la carga es negativa, los extremos están a altura cero y el principio del máximo da $u \leq 0$. Aumentar h acerca ese perfil a cero: si $h_2 > h_1$ y u_i es la solución correspondiente, la diferencia $w = u_2 - u_1$ satisface

$$-(kw')' + h_2 w = -(h_2 - h_1) u_1 \geq 0, \qquad w(0) = w(L) = 0,$$

luego $w \geq 0$: un mínimo interior negativo daría $w' = 0$, $w'' \geq 0$ y $-(kw')' + h_2 w < 0$, contradiciendo la ecuación. Así se justifica que al bajar h aumente el desplazamiento hacia abajo. Esta comparación usa el signo de la carga y los datos homogéneos; no afirma que cualquier perfil disminuya punto a punto en valor absoluto al aumentar h.
Como la rigidez crece hacia la derecha, es razonable esperar una deformación asimétrica. La localización del mínimo, sin embargo, se obtiene de la solución: no se deduce solo de mirar el funcional, y con otras cargas o condiciones de contorno puede cambiar.
Paso 1: la deducción. Con $H(x, v, v') = \frac{k(x)}{2} v'^2 + \frac{h}{2} v^2 - f(x) v$, Euler-Lagrange es $-(k(x)u')' + hu - f = 0$ con las condiciones de Dirichlet: es el ejemplo de la cuerda no homogénea con un término de reacción hu (la restitución elástica). La energía es convexa en (v, v') (suma

de cuadrados con coeficientes no negativos más un término lineal), luego el problema de contorno caracteriza el mínimo. Como sistema de primer orden,

$$z_1' = z_2, \qquad z_2' = \frac{hz_1 - f(x) - k'(x)z_2}{k(x)}, \qquad k'(x) = 2k_0 x.$$

```
function VigaElasticaNH(kcero, L, alpha, beta, f0, nforce, h)
% VigaElasticaNH: -(k(x) u')' + h u = f(x), u(0) = alpha, u(L) = beta, k(x) = 1 + kcero
    x^2,
% resuelto para h, h/2 y h/4.
% Prueba: VigaElasticaNH(1, 1, 0, 0, -1, 1, 8)
assert(kcero>=0 && L>0 && h>=0 && nforce>=1 && nforce==fix(nforce));
k = @(x) 1 + kcero*x.^2;
kp = @(x) 2*kcero*x; % k'(x)
fsin = @(x) f0*sin(nforce*pi*x/L);
condcon = @(za,zb) [za(1) - alpha; zb(1) - beta];
solinit = bvpinit(linspace(0, L, 20), [0; 0]);

close all
hold on
hh = [h, h/2, h/4];
colores = 'brg';
xx = linspace(0, L);
for j = 1:3
  hj = hh(j);
  dzdx = @(x,z) [z(2); (hj*z(1) - fsin(x) - kp(x)*z(2))/k(x)];
  sol = bvp4c(dzdx, condcon, solinit);
  plot(xx, deval(sol, xx, 1), colores(j), 'LineWidth', 1.5)
  fprintf('h = %6.3f: minimo en la malla %9.5f \n', hj, min(deval(sol, xx, 1)))
end
plot([0, L], [alpha, beta], 'k--')
legend(['h = ', num2str(h)], ['h/2 = ', num2str(h/2)], ['h/4 = ', num2str(h/4)], 'Sin
    deformar', 'Location', 'Best')
xlabel('x'), ylabel('u(x)'), title('Viga no homogenea con apoyo elastico: k(x) = 1 + k_0
    x^2')
hold off, shg
end
```

SALIDA ESPERADA

```
>> VigaElasticaNH(1, 1, 0, 0, -1, 1, 8)
h = 8.000: minimo en la malla -0.04756
h = 4.000: minimo en la malla -0.05868
h = 2.000: minimo en la malla -0.06645
```

Los valores de la salida son aproximaciones sobre los cien puntos de la malla del dibujo; no son mínimos exactos. Refinando la solución y localizando un cero de u', los mínimos son aproximadamente -0.0475682, -0.0586848 y -0.0664463, en $x = 0.4603$, 0.4574 y 0.4558, respectivamente. En esta prueba quedan a la izquierda del centro, hacia el lado menos rígido.

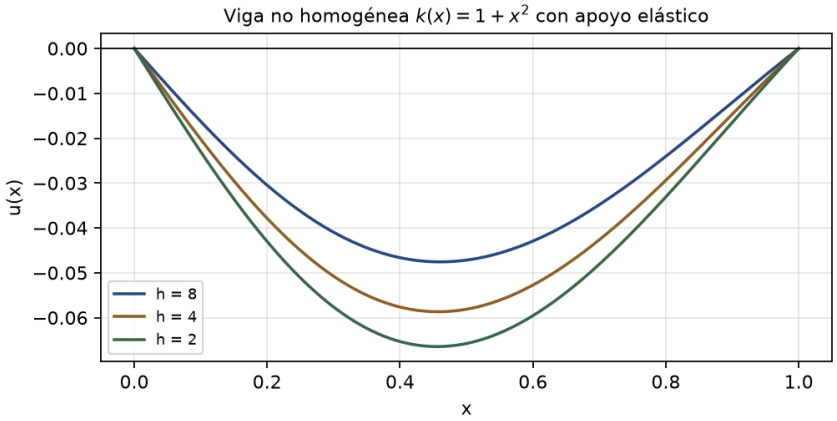


Figura 3.8: Viga no homogénea, $k(x) = 1 + x^2$, con fuerza $-\sin(\pi x)$ y tres valores del coeficiente de restitución h. Con esta carga y estos datos de contorno, la deformación hacia abajo crece al bajar h y el mínimo queda hacia el lado menos rígido.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_bvp
def viga_NH(kcero, L, alpha, beta, f0, nforce, h):
    if not (kcero>=0 and L>0 and h>=0 and nforce>=1 and int(nforce)==nforce):
        raise ValueError("Datos fuera del dominio indicado")
    k = lambda x: 1 + kcero * x ** 2; kp = lambda x: 2 * kcero * x
    fsin = lambda x: f0 * np.sin(nforce * np.pi * x / L)
    dzdx = lambda x, z: np.vstack([z[1], (h * z[0] - fsin(x) - kp(x) * z[1]) / k(x)])
    bc = lambda za, zb: np.array([za[0] - alpha, zb[0] - beta])
    xi = np.linspace(0, L, 20); sol = solve_bvp(dzdx, bc, xi, np.zeros((2, xi.size)))
    if not sol.success: raise RuntimeError(sol.message)
    return sol.sol(np.linspace(0, L, 200))[0]
for hh in (8, 4, 2):
    print(hh, viga_NH(1, 1, 0, 0, -1, 1, hh).min())
```

Qué hemos aprendido y dónde se falla. Una rigidez variable obliga a escribir la ecuación en forma de divergencia, $-(k(x)u')'$, y al pasarla a primer orden aparece el término $k'(x)z_2$. Ese término es el sitio del fallo.

- *Al olvidar $k'(x)z_2$.* Escribir la segunda componente de `dzdx` sin ese término, como `(hj*z(1)-fsin(x))/k(x)`, resuelve $-ku'' + hu = f$, que es otra ecuación. El resultado se parece mucho y engaña: con $h = 8$ el mínimo pasa de -0.04757 en $x = 0.460$ a -0.04904 en $x = 0.482$, un 3 % más hondo y casi centrado. Se pierde justo lo que el ejercicio quería ver: que la deformación se carga hacia el lado menos rígido.
- *Al probar con el caso que no lo detecta.* Con $k_0 = 0$ la rigidez es constante, $k' = 0$, y las dos ecuaciones coinciden, así que un programa con el término olvidado pasa esa prueba. Un control que sí lo ve es el balance integral de la ecuación: integrando $-(ku')' + hu = f$ en $(0, L)$,

$$k(L)u'(L) - k(0)u'(0) = \int_0^L \big(hu - f\big)\, dx.$$

  Con la ecuación buena y $h = 8$, los dos lados valen 0.39845; con la mala difieren en 0.062.

Ejercicio 18: vibraciones forzadas y resonancia

Ejercicio 3.8. *Vibraciones forzadas*

ORIGEN: hoja de ejercicios del tema 3, ejercicio 18, con el fichero `VibForzadas.m` del curso.

Se considera el problema de las vibraciones forzadas

$$y'' + \omega^2 y = A\cos(\omega_\star t), \qquad t \in [0,T], \qquad y(0) = y_{00},\ y'(0) = y_{01}.$$

La ecuación sale del principio de acción estacionaria como la del muelle, con una fuerza externa $f(t)$: $E_c = \frac{m}{2}|y'|^2$, $E_p = \frac{k}{2}|y|^2 - f(t)y$, $\mathcal{L} = E_c - E_p$, $\omega^2 = k/m$ y $f(t)/m = A\cos(\omega_\star t)$. Suponemos $\omega > 0$, $\omega_\star \geq 0$ y $T > 0$. Escribir `VibForzadas(om, omstar, A, y00, y01, T)` que la resuelva numéricamente y compare con la solución exacta.

Resolución.
La idea. Un oscilador empujado por una fuerza periódica responde de dos maneras radicalmente distintas según si la frecuencia con que se le empuja coincide o no con la suya. Si no coinciden, la respuesta es una suma de dos oscilaciones puras y está **acotada**; la componente particular tiene amplitud $|A|/|\omega^2 - \omega_\star^2|$, pero la amplitud total depende también de los datos iniciales y de la interferencia entre ambas componentes. Si coinciden las frecuencias y $A \neq 0$, aparece una oscilación con *amplitud que crece linealmente con el tiempo*: la resonancia. El denominador de la fórmula no resonante se anula, aunque para un tiempo fijo la solución tiene un límite finito.
La razón matemática es limpia y conviene decirla, porque es la que dicta qué ensayo hay que probar en cada caso: cuando $\omega = \omega_\star$ el segundo miembro *ya es solución de la ecuación homogénea*, así que probar con $D\cos(\omega t)$ no puede funcionar —da cero— y hay que multiplicar por t. Ese t de más es exactamente la amplitud que crece. La imagen física es la de aportar energía en fase con el movimiento; fuera de resonancia se alternan etapas de refuerzo y de oposición. El trabajo se mide por fuerza por velocidad, no solo por el signo de la fuerza.
Paso 1: la solución exacta, en los dos casos. Por el teorema de las EDO lineales de coeficientes constantes, el homogéneo tiene solución general $C_1\cos(\omega t) + C_2\sin(\omega t)$. Para una particular:

- Si $\omega \neq \omega_\star$, se prueba $y_p = D\cos(\omega_\star t)$: $y_p'' + \omega^2 y_p = D(\omega^2 - \omega_\star^2)\cos(\omega_\star t)$, luego $D = A/(\omega^2 - \omega_\star^2)$ y

$$y(t) = C_1\cos(\omega t) + C_2\sin(\omega t) + \frac{A}{\omega^2 - \omega_\star^2}\cos(\omega_\star t).$$

- Si $\omega = \omega_\star$ (**resonancia**), se prueba $y_p = Dt\sin(\omega t)$: $y_p' = D\sin(\omega t) + D\omega t\cos(\omega t)$, $y_p'' = 2D\omega\cos(\omega t) - D\omega^2 t\sin(\omega t)$, luego $y_p'' + \omega^2 y_p = 2D\omega\cos(\omega t)$ y $D = A/(2\omega)$:

$$y(t) = C_1\cos(\omega t) + C_2\sin(\omega t) + \frac{A}{2\omega}\,t\sin(\omega t).$$

Paso 2 (el núcleo): el coeficiente del término resonante, que es donde todo el mundo se equivoca. **Una errata de la hoja y del fichero.** La hoja escribe el término resonante como $\frac{A}{2}t\sin(\omega t)$, y el fichero `VibForzadas.m` lo programa así. El coeficiente correcto es $\frac{A}{2\omega}$; los dos coinciden solo

si $\omega = 1$. Con $\omega = 2$, la «solución exacta» del fichero se separa de la numérica en hasta 12.37 unidades sobre $[0, 50]$ cuando $A = 1$ y los datos iniciales son nulos; en $t = 50$ la diferencia entre ambas fórmulas es 6.33. El error numérico con el coeficiente corregido depende de las tolerancias. Si se usa la solución exacta para comprobar, hay que derivarla y no copiarla.
Las constantes se fijan con los datos iniciales: $C_1 = y_{00} - D$ (caso no resonante) o $C_1 = y_{00}$ (resonante), y $C_2 = y_{01}/\omega$ en los dos casos, porque $y_p'(0) = 0$.
Una misma fórmula, también cerca de resonancia. Restar dos cosenos casi iguales y multiplicar por D, que es enorme, pierde precisión. La identidad de diferencia de cosenos da una expresión equivalente que evita esa cancelación:

$$y(t) = y_{00}\cos(\omega t) + \frac{y_{01}}{\omega}\sin(\omega t) + \frac{At}{\omega+\omega_\star}\sin\frac{(\omega+\omega_\star)t}{2}\,\sigma\!\left(\frac{(\omega-\omega_\star)t}{2}\right), \qquad \sigma(s) = \begin{cases} \sin s/s, & s \neq 0, \\ 1, & s = 0. \end{cases}$$

Como $\sigma(0) = 1$, la misma fórmula da el término resonante al hacer $\omega_\star = \omega$. El programa usa esta forma; no necesita decidir numéricamente cuándo dos frecuencias están «lo bastante cerca».

```
function VibForzadas(om, omstar, A, y00, y01, T)
% VibForzadas: y'' + om^2 y = A cos(omstar t), y(0) = y00, y'(0) = y01, en [0,T].
% Compara ode45 con la solucion exacta (coeficiente resonante A/(2 om)).
% Pruebas: VibForzadas(2, 2.5, 1, 0, 0, 50) VibForzadas(2, 2.1, 1, 0, 0, 100)
    VibForzadas(2, 2, 1, 0, 0, 50)
assert(om>0 && omstar>=0 && T>0);
dzdt = @(t,z) [z(2); -om^2*z(1) + A*cos(omstar*t)];
opts = odeset('RelTol',1e-9,'AbsTol',1e-11);
[times,z] = ode45(dzdt,[0,T],[y00;y01],opts);
exacta = @(t) y00*cos(om*t) + (y01/om)*sin(om*t) ...
   + (A/(om+omstar))*t.*sin((om+omstar)*t/2).*sigma((om-omstar)*t/2);

close all
plot(times, z(:,1)), hold on
plot(times, exacta(times), 'r.', 'MarkerSize', 7)
legend('Solucion numerica', 'Solucion exacta', 'Location', 'Best')
title(['Vibraciones forzadas. A=', num2str(A), ', w=', num2str(om), ', w*=', num2str(
    omstar)])
fprintf('Error maximo frente a la exacta: %10.3e \n', max(abs(z(:,1) - exacta(times))))
shg
end

function v = sigma(s)
% sin(s)/s, con su valor limite en cero; sin toolbox adicional.
v = ones(size(s)); nz = s~=0;
v(nz) = sin(s(nz))./s(nz);
end
```

SALIDA ESPERADA

```
>> VibForzadas(2, 2.5, 1, 0, 0, 50)
(se imprime el error maximo en los nodos; repetir con menor tolerancia para comprobarlo)
```

La figura 3.9 usa datos iniciales nulos y $A = 1$. Los máximos de $|y|$ *observados en* $[0, 50]$ son aproximadamente 0.89, 2.38, 4.88 y 12.18. El último no es una cota para todo tiempo: con

$\omega_\star = 2.01$ la envolvente puede alcanzar $2/|4 - 2.01^2| \approx 49.88$. En efecto, con estos datos,

$$y(t) = \frac{2A}{\omega^2 - \omega_\star^2} \sin\frac{(\omega + \omega_\star)t}{2} \sin\frac{(\omega - \omega_\star)t}{2}.$$

El segundo seno modula lentamente la oscilación rápida. El intervalo entre máximos de la *envolvente en valor absoluto* es $2\pi/|\omega - \omega_\star|$: las pulsaciones se alargan al acercarse las frecuencias. En resonancia y sin rozamiento, si $A \neq 0$ la amplitud crece sin límite en el modelo lineal.

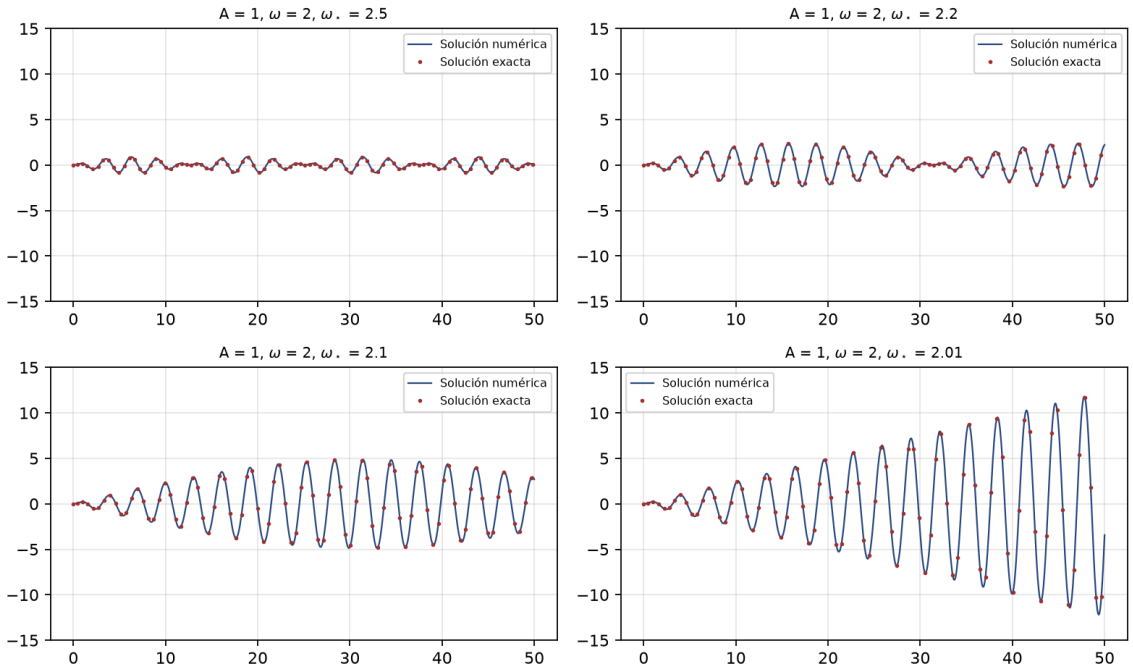


Figura 3.9: Vibraciones forzadas con $\omega = 2$ y $\omega_\star = 2.5, 2.2, 2.1, 2.01$. Las pulsaciones se alargan y crecen al acercarse a la resonancia. Los puntos son la solución exacta.

La idea: qué cambia al añadir rozamiento

El crecimiento ilimitado indica que el modelo lineal sin pérdidas acabará saliendo de su régimen de validez; no basta para predecir cómo se romperá una estructura concreta. Si se añade rozamiento, $y'' + cy' + \omega^2 y = A\cos(\omega_\star t)$ con $c > 0$, la respuesta periódica que queda tras el transitorio tiene amplitud

$$R(\omega_\star) = \frac{|A|}{\sqrt{(\omega^2 - \omega_\star^2)^2 + c^2\omega_\star^2}}.$$

Se obtiene ensayando $a\cos(\omega_\star t) + b\sin(\omega_\star t)$ y resolviendo para a, b. A frecuencia de excitación $\omega_\star = \omega$, vale $|A|/(c\omega)$: es finita. El máximo de R, cuando está en una frecuencia positiva, ocurre en $\sqrt{\omega^2 - c^2/2}$, por lo que con rozamiento el pico se desplaza hacia frecuencias menores.

Versión en Python (NumPy/Matplotlib)

```python
import numpy as np
from scipy.integrate import solve_ivp
def vib_forzadas(om, omstar, A, y00, y01, T):
    if not (om>0 and omstar>=0 and T>0): raise ValueError("Frecuencias/T no validos")
    s = solve_ivp(lambda t,z: [z[1],-om**2*z[0]+A*np.cos(omstar*t)],
                  [0,T],[y00,y01],rtol=1e-9,atol=1e-11,max_step=0.05)
    if not s.success: raise RuntimeError(s.message)
    # np.sinc(x) = sin(pi*x)/(pi*x): por eso se divide el argumento por pi.
    exacta = lambda t: (y00*np.cos(om*t)+(y01/om)*np.sin(om*t)
        + A*t/(om+omstar)*np.sin((om+omstar)*t/2)*np.sinc((om-omstar)*t/(2*np.pi)))
    print("error maximo en los nodos:", np.max(np.abs(s.y[0]-exacta(s.t))))
vib_forzadas(2,2.01,1,0,0,50); vib_forzadas(2,2,1,0,0,50)
```

Ejercicio 19: una viga empotrada, de cuarto orden

Ejercicio 3.9. *Viga delgada de sección variable*

ORIGEN: hoja de ejercicios del tema 3, ejercicio 19, con el fichero `VigaOrden4.m` del curso.

Una viga delgada $B = \{(x,y,z) : 0 < x < 1,\ y^2+z^2 < \varepsilon r(x)^2\}$, con $\varepsilon > 0$ y r regular y positiva, está horizontal y sujeta por los extremos. El radio es $\sqrt{\varepsilon}\, r(x)$. Escribiendo el desplazamiento vertical como $w(x) = u(x)/\varepsilon$, la función escalada u resuelve

$$\begin{cases} k\dfrac{d^2}{dx^2}\left(r(x)^4\dfrac{d^2u}{dx^2}\right) = -r(x)^2 & \text{en } (0,1), \\ u(0) = u(1) = u'(0) = u'(1) = 0, \end{cases}$$

con $k = E_Y/(4\rho g) > 0$, donde E_Y es el módulo de Young y ρ la densidad uniforme ($k = E_Y/(4g)$ si esta se normaliza a uno). Escribir `VigaOrden4(kappa, funr)` que lo resuelva y dibuje u para $k = 1$, $r(x) = 2 + \sin(2\pi x)$.

Resolución.
La idea, y la decisión de método está en cómo se nombran las variables. La reducción de una ecuación de orden cuatro a un sistema de primer orden parece automática —$y_1 = u$, $y_2 = u'$, $y_3 = u''$, $y_4 = u'''$— y con esa elección *el programa también funcionaría*. Pero habría que derivar r^4 dos veces a mano y escribir

$$y_4' = \frac{-r^2/k - 8r^3r'y_4 - (12r^2(r')^2 + 4r^3r'')y_3}{r^4},$$

que es larga, propensa a errores de signo y, sobre todo, exige que r sea dos veces derivable y que se conozcan esas derivadas: el argumento `funr` tendría que venir acompañado de otros dos. La elección del programa —$y_3 = r^4u''$ y $y_4 = (r^4u'')'$, es decir, tomar como incógnitas cantidades proporcionales al **momento flector** y al **esfuerzo cortante** en vez de las derivadas desnudas de u— hace que la ecuación se lea directamente como $y_4' = -r^2/k$ y que r solo tenga que evaluarse, nunca derivarse.
La elección tiene sentido físico. En una unión sin fuerza ni par concentrados se mantienen continuos el momento y el cortante, aunque cambie bruscamente la sección y pueda saltar

u''. Con cargas puntuales habría que imponer los saltos correspondientes. Estas variables también aparecen en el problema adjunto del capítulo 4: **evitan derivar los coeficientes y conservan las cantidades que intervienen en el equilibrio.**

El núcleo del programa son dos líneas —`dydx` y `condcon`— y la llamada a `bvp4c`. Obsérvese que las cuatro condiciones van todas sobre y_1 e y_2, dos en cada extremo: es el empotramiento, que fija posición y pendiente. Si un extremo permanece empotrado y en el otro se imponen $y_3 = y_4 = 0$, ese segundo extremo queda libre de momento y cortante: es una viga en voladizo. No solo importa cuántas condiciones hay en cada extremo, sino qué cantidades fijan.

La reducción. Con $y_1 = u$, $y_2 = u'$, $y_3 = r^4 u''$ e $y_4 = (r^4 u'')'$,

$$y_1' = y_2, \qquad y_2' = \frac{y_3}{r^4}, \qquad y_3' = y_4, \qquad y_4' = -\frac{r^2}{k},$$

con $y_1(0) = y_2(0) = 0$ e $y_1(1) = y_2(1) = 0$. La ventaja de esta elección es que la EDO $k(r^4 u'')'' = -r^2$ se lee directamente como $y_4' = -r^2/k$ sin derivar r dos veces. Es la misma idea que en la viga no homogénea, llevada al cuarto orden. *De dónde sale.* El modelo es el de energía mínima con la energía de flexión $\frac{k}{2}\int r^4 |u''|^2$ (una viga se resiste a *curvarse*, no a estirarse, y la rigidez de una sección circular crece con la cuarta potencia del radio) más la gravitatoria $\int r^2 u$; Euler-Lagrange con segunda derivada da la ecuación de cuarto orden, y las cuatro condiciones de empotramiento son las de Dirichlet en u y u'.

Las potencias de r se comprueban en la sección circular: su área es $\mathcal{A} = \pi\varepsilon r^2$ y su segundo momento de área es $I = \pi\varepsilon^2 r^4/4$. La ecuación física $(E_Y I w'')'' = -\rho g \mathcal{A}$, al sustituir $w = u/\varepsilon$ y dividir por $\rho g \pi \varepsilon$, da precisamente la ecuación escrita. Así se distingue el perfil de radio r del área, proporcional a r^2.

Órdenes nuevas. `funr` es un argumento que es una función: dentro se usa como cualquier otra, `funr(x)`, y se construyen a partir de ella otras anónimas, `funr2 = @(x) funr(x).^2`.

```
function VigaOrden4(kappa, funr)
% VigaOrden4: k (r^4 u'')'' = -r^2 en (0,1), u(0) = u(1) = u'(0) = u'(1) = 0.
% Cambio: y1 = u, y2 = u', y3 = r^4 u'', y4 = (r^4 u'')'.
% Pruebas: VigaOrden4(1, @(x) 2+sin(2*pi*x)) VigaOrden4(1, @(x) 2+cos(2*pi*x))
assert(kappa>0); % Se requiere ademas funr(x)>0 en todo [0,1].
funr2 = @(x) funr(x).^2;
funr2k = @(x) -funr2(x)/kappa;
funr4 = @(x) 1./funr2(x).^2; % 1/r^4

dydx = @(x,y) [y(2); y(3)*funr4(x); y(4); funr2k(x)];
condcon = @(ya,yb) [ya(1); ya(2); yb(1); yb(2)]; % u y u' nulos en los dos extremos

solinit = bvpinit(linspace(0, 1, 5), [1; 1; 1; 1]);
sol = bvp4c(dydx, condcon, solinit, bvpset('RelTol',1e-7,'AbsTol',1e-10));

xx = linspace(0, 1);
uu = deval(sol, xx, 1);
close all
plot(xx, uu, 'LineWidth', 1.1)
xlim([0,1]), grid on
title('Viga elastica empotrada, EDO de cuarto orden')
[umin,jmin] = min(uu);
fprintf('Minimo en la malla %10.4e en x = %5.3f \n',umin,xx(jmin))
```

```
shg
end
```

SALIDA ESPERADA

```
>> VigaOrden4(1, @(x) 2+sin(2*pi*x))
Minimo en la malla -1.0589e-03 en x = 0.596
```

El desplazamiento es del orden de 10^{-3}, negativo (la viga cuelga), y el mínimo está a la derecha del centro: donde r es menor (en $x = 3/4$, $r = 1$) la viga es más flexible. Con $r = 2 + \cos(2\pi x)$ y el mismo k sale $-2.624 \cdot 10^{-4}$ en $x = 1/2$: la viga es más delgada en el centro, la flecha queda simétrica y es unas 4.03 veces menor. Cambiar r modifica a la vez la rigidez, proporcional a r^4, y el peso, proporcional a r^2; la comparación no se explica solo por el grosor en el punto central.

Controles independientes. En los dos perfiles variables no hemos usado una fórmula cerrada para u, pero sí tenemos un caso de referencia: si $r \equiv R > 0$,

$$u(x) = -\frac{x^2(1-x)^2}{24kR^2}.$$

Basta derivar cuatro veces y comprobar los extremos. Esta prueba sencilla verifica también la escala $1/(kR^2)$ del programa.

Otro control es un **balance integral exacto**. Como $y_4' = -r^2/k$, integrando entre 0 y 1:

$$y_4(1) - y_4(0) = -\frac{1}{k}\int_0^1 r(x)^2\,dx,$$

y tanto con $r = 2 + \sin(2\pi x)$ como con $r = 2 + \cos(2\pi x)$ esa integral vale $4 + 0 + \frac{1}{2} = 4.5$. La diferencia debe aproximar $-4.5/k$: es el balance global de fuerzas, el cortante en un extremo menos el del otro igual al peso con su signo. Es un control integral; no detecta por sí solo errores locales que se cancelen al integrar.

El control **de simetría** vale para el segundo dato. La función $r = 2 + \cos(2\pi x)$ cumple $r(1-x) = r(x)$, y las condiciones de contorno también son simétricas, y la solución es única: para la diferencia v de dos soluciones, integrar dos veces por partes da $\int r^4|v''|^2 = 0$, luego $v'' = 0$ y los datos de contorno obligan a $v = 0$. Por ello $u(x) = u(1-x)$. Un cálculo refinado da $u(0.25) = u(0.75) = -9.5438 \cdot 10^{-5}$. Con $r = 2 + \sin(2\pi x)$ no hay simetría, y en efecto $u(0.25) = -4.485 \cdot 10^{-4}$ frente a $u(0.75) = -6.674 \cdot 10^{-4}$. Una simetría del problema que la solución numérica respeta es la comprobación más barata que existe, y la que más errores de tecleo caza.

También hay un control de signo: el desplazamiento tiene que ser negativo en todo el interior —la viga cuelga— y nulo en los dos extremos. Si saliera positivo en algún punto, habría un signo mal en la gravedad.

Y una precisión sobre el mínimo que imprime el programa. La línea `[umin,jmin]=min(uu)` busca el mínimo *sobre la malla* `linspace(0,1)`, que tiene cien puntos, y devuelve $x = 0.596$. El mínimo verdadero, refinado, está en $x = 0.598$: la diferencia es el espaciado de la malla, $1/99$. La comparación con el caso simétrico lo enseña mejor que cualquier explicación: allí la simetría da $u'(1/2) = 0$ y el mínimo está en $x = 1/2$, y el programa imprime 0.495, porque 0.5 no es uno de los cien puntos de `linspace(0,1)`. **Un extremo leído sobre una malla nunca es mejor que la malla**; si hace falta la posición, se refina resolviendo $u'(x) = 0$, que aquí es pedir el cero de la segunda componente de la solución.

Dónde se rompe. En la aproximación inicial y en el condicionamiento, que aquí van juntos. La malla inicial son cinco puntos y la aproximación es el vector constante $(1, 1, 1, 1)$, que no cumple ninguna de las cuatro condiciones de contorno. Aquí puede usarse porque el problema es *lineal* y tiene solución única; la aproximación no elige entre ramas. Esto no significa que la adaptación de malla o los errores de redondeo desaparezcan. En cuanto la ecuación deje de ser lineal, como en el problema de Bratu del ejercicio siguiente, esa despreocupación se acaba. La otra rotura es de escala: si r se acercara a cero en algún punto, el término y_3/r^4 puede amplificar los errores; en los perfiles de la figura se cumple $r \geq 1$. Con $r = 1 + \sin(2\pi x)$, que se anula en $x = 3/4$, se sale de las hipótesis: el sistema escrito es singular y ya no puede usarse este programa sin un tratamiento adicional.

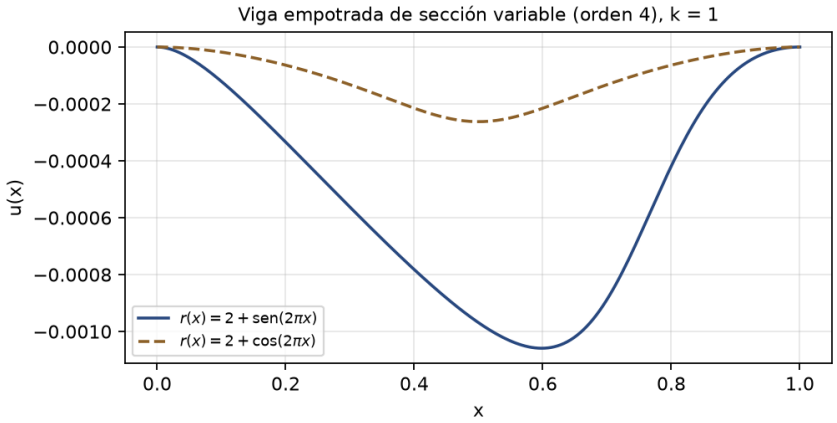


Figura 3.10: Desplazamiento escalado u de la viga empotrada de sección variable. Con $r = 2 + \sin(2\pi x)$ la deformación se carga hacia el lado delgado; con $r = 2 + \cos(2\pi x)$ es simétrica y menor.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_bvp
def viga_orden4(kappa, funr):
    if kappa<=0: raise ValueError("kappa debe ser positivo")
    def r2(x):
        r = np.broadcast_to(np.asarray(funr(x),dtype=float),np.shape(x))
        if np.any(r<=0): raise ValueError("El radio debe ser positivo")
        return r**2
    dydx = lambda x, y: np.vstack([y[1], y[2] / r2(x) ** 2, y[3], -r2(x) / kappa])
    bc = lambda ya, yb: np.array([ya[0], ya[1], yb[0], yb[1]])
    xi = np.linspace(0, 1, 5); sol = solve_bvp(dydx, bc, xi, np.ones((4, xi.size)),tol=1e
        -7,max_nodes=10000)
    if not sol.success: raise RuntimeError(sol.message)
    xx = np.linspace(0, 1, 200); return xx, sol.sol(xx)[0]
xx, uu = viga_orden4(1, lambda x: 2 + np.sin(2 * np.pi * x)); print("Minimo en la malla:"
    , uu.min(), xx[uu.argmin()])
```

Ejercicio 20: el problema de Bratu

Ejercicio 3.10. *Las dos soluciones del problema de Bratu*

ORIGEN: hoja de ejercicios del tema 3, ejercicio 20, con el fichero `Bratu.m` del curso.

El problema de Bratu en una dimensión es

$$u'' + \lambda e^u = 0 \text{ en } (0,1), \qquad u(0) = u(1) = 0.$$

Según $\lambda > 0$ tiene una, dos o ninguna solución: existe $\lambda_\star \approx 3.51$ tal que para $\lambda > \lambda_\star$ no hay solución, para $\lambda = \lambda_\star$ hay exactamente una y para $0 < \lambda < \lambda_\star$ hay dos. Las soluciones exactas son

$$u(x) = 2\log\frac{\cosh\alpha}{\cosh\big(\alpha(1-2x)\big)}, \qquad \text{con } \alpha \text{ solución de } \cosh\alpha = \frac{4\alpha}{\sqrt{2\lambda}}.$$

Escribir `Bratu(Lambda)` que calcule las dos soluciones y las dibuje junto con las exactas.

PISTAS, DE MENOR A MAYOR

1. *Las dos soluciones se obtienen con dos aproximaciones iniciales distintas: una pequeña y otra grande.*
2. *Las α se calculan a ambos lados del máximo de $\alpha/\cosh\alpha$; el extremo derecho del segundo intervalo se amplía hasta encerrar la raíz.*
3. *$\lambda_\star$ es el valor a partir del cual la ecuación en α deja de tener solución.*

Resolución.

La idea: buscar y distinguir las ramas. En los problemas lineales anteriores la aproximación inicial no elegía entre distintas soluciones. Aquí hay dos para $0 < \lambda < \lambda_\star$, y una llamada a `bvp4c` busca una sola. Dos llamadas con arranques diferentes tampoco garantizan dos respuestas diferentes: pueden converger a la misma rama o fallar.

La fórmula cerrada permite preparar el cálculo y comprobarlo después. Usaremos dos parábolas que satisfacen los extremos y tienen alturas cercanas a las de cada rama; en este ejemplo, esas alturas se obtienen de la fórmula exacta. Después compararemos los *perfiles completos* con ella. Así el programa no presenta como dos soluciones verificadas lo que podrían ser dos copias de una misma solución.

El fichero de partida usa constantes $(0.1, 0)$ y $(3, 1)$ como arranque, útiles para ciertos valores de λ, pero no universales. También fija los intervalos $[0, 1.2]$ y $[1.2, 5]$ para las raíces. Hay dos problemas: 1.2 solo aproxima el punto de máximo, y la raíz grande supera 5 cuando λ es suficientemente pequeño. La versión siguiente calcula el máximo y amplía el segundo intervalo cuando hace falta.

De dónde sale $\lambda_\star$. La ecuación $\cosh\alpha = 4\alpha/\sqrt{2\lambda}$ equivale a $\lambda = 8(\alpha/\cosh\alpha)^2$. La función $\alpha/\cosh\alpha$ vale 0 en 0, crece hasta un máximo y decrece a 0; por tanto para cada λ menor que $\lambda_\star = 8\,\text{máx}(\alpha/\cosh\alpha)^2$ hay dos valores de α, uno a cada lado del máximo, para $\lambda = \lambda_\star$ uno, y por encima ninguno. Numéricamente $\lambda_\star = 3.5138307$, que es el valor que se cita habitualmente.

El punto de máximo α_c satisface $\alpha_c \tanh\alpha_c = 1$ y es único: la función $a \mapsto a\tanh a$ crece de cero a infinito. Se obtiene $\alpha_c \approx 1.19967864$.

Por qué la fórmula recoge todas las soluciones. La concavidad estricta da un único máximo $M = u(x_0) > 0$. Con $b = \sqrt{\lambda e^M/2}$, la función $M - 2\log\cosh(b(x - x_0))$ satisface la ecuación

y los mismos datos $u(x_0) = M$, $u'(x_0) = 0$. La unicidad del problema inicial obliga a que sea u. Los dos extremos nulos implican $\cosh(bx_0) = \cosh(b(1 - x_0))$, de donde $x_0 = 1/2$. Tomando $\alpha = b/2$ se recuperan la fórmula anterior y $\lambda = 8(\alpha/\cosh\alpha)^2$. Esto justifica tanto el recuento completo de soluciones como su simetría; la simetría de una ecuación, por sí sola, no obliga a que todas sus soluciones sean simétricas.

Órdenes nuevas. `cosh` es el coseno hiperbólico. La solución de `bvp4c` se puede dibujar directamente con sus campos `sol.x` y `sol.y(1,:)`, sin pasar por `deval`. En el fichero, λ se comparte con la función auxiliar `exacta` mediante `global`. Y una función puede devolver *funciones*: `exacta` devuelve dos anónimas.

```
function Bratu(Lambda)
% Dos ramas de u'' + lambda*exp(u)=0, u(0)=u(1)=0.
% Pruebas: Bratu(0.01) Bratu(1) Bratu(3.5)
global lambda
lambda = Lambda;
[ex1,ex2] = exacta;
M1 = ex1(0.5); M2 = ex2(0.5);
dydx = @(x,y) [y(2); -lambda*exp(y(1))];
condcon = @(ya,yb) [ya(1); yb(1)];
xi = linspace(0,1,50);
solinit1 = bvpinit(xi,@(x) [4*M1*x*(1-x); 4*M1*(1-2*x)]);
solinit2 = bvpinit(xi,@(x) [4*M2*x*(1-x); 4*M2*(1-2*x)]);
options = bvpset('RelTol',1e-7,'AbsTol',1e-9,'NMax',20000);
sol1 = bvp4c(dydx,condcon,solinit1,options);
sol2 = bvp4c(dydx,condcon,solinit2,options);
xx = linspace(0,1,1001);
u1 = deval(sol1,xx,1); u2 = deval(sol2,xx,1);
err = max([max(abs(u1-ex1(xx))),max(abs(u2-ex2(xx)))]);
if err > min(1e-4,abs(M2-M1)/4)
    error('No se han verificado dos ramas: ajustar arranque o tolerancias');
end
plot(xx,u1,xx,u2), hold on
xp = linspace(0,1,31);
plot(xp,ex1(xp),'*',xp,ex2(xp),'*')
legend('Rama baja numerica','Rama alta numerica','Exacta baja','Exacta alta')
title(['Bratu, lambda = ',num2str(lambda)]), hold off, shg
fprintf('u(1/2): %8.6f y %8.6f \n',deval(sol1,.5,1),deval(sol2,.5,1))
end

function [y1,y2] = exacta
global lambda
ac = fzero(@(a) a*tanh(a)-1,[1,2]);
lcrit = 8*(ac/cosh(ac))^2;
if ~(isfinite(lambda) && lambda>0 && lambda<lcrit)
    error('Dos ramas solo para 0 < lambda < lambda_crit');
end
fun = @(a) 8*(a./cosh(a)).^2-lambda;
hi = 2*ac;
while fun(hi)>0, hi = 2*hi; end
a1 = fzero(fun,[0,ac]); a2 = fzero(fun,[ac,hi]);
y1 = @(x) 2*(log(cosh(a1))-log(cosh(a1*(1-2*x))));
y2 = @(x) 2*(log(cosh(a2))-log(cosh(a2*(1-2*x))));
end
```

SALIDA ESPERADA

```
>> Bratu(1)
u(1/2): 0.140539 y 4.091467
>> Bratu(3.5)
u(1/2): 1.085159 y 1.294585
```

Con $\lambda = 1$, $\alpha_1 = 0.379291$ y $\alpha_2 = 2.734676$: una solución pequeña y otra veintinueve veces mayor. Con $\lambda = 3.5$, cerca del crítico, las dos se acercan ($\alpha_1 = 1.1380$ y $\alpha_2 = 1.2636$, que dan 1.085 y 1.295 en el centro) y la separación numérica de las ramas se vuelve más delicada. Con $\lambda = 4$ sabemos *por el análisis anterior* que no existe solución; el programa rechaza ese dato antes de intentar calcular dos ramas.

El control, y es el mejor del capítulo porque hay fórmula. Los dos números de la salida salen también de la fórmula cerrada: $u(1/2) = 2\log\cosh\alpha$, que con $\alpha_1 = 0.379291$ da 0.140539 y con $\alpha_2 = 2.734676$ da 4.091467, las seis cifras de referencia de la salida. Por eso el programa dibuja las exactas con asteriscos encima de las numéricas: no es adorno, es la comprobación, y el ejercicio está construido para poder hacerla. Merece la pena aprovecharla, porque permite detectar también que las dos llamadas hayan convergido a la misma rama.

Un segundo control, para cuando no la haya: la solución de este problema tiene que ser **simétrica** respecto de $x = 1/2$, como acabamos de demostrar, y **positiva** en el interior, porque $u'' = -\lambda e^u < 0$ hace la solución cóncava y una función cóncava que se anula en los dos extremos es positiva entre ellos. Las dos propiedades se comprueban en una línea y no exigen conocer α.

Dónde se rompe, y hay tres sitios. El primero, $\lambda > \lambda_\star$: el análisis demuestra que no hay solución. Un fallo de `bvp4c` sería compatible con ese hecho, pero no lo demostraría. El mismo mensaje puede aparecer en un problema con solución si el arranque, las tolerancias o el condicionamiento no permiten encontrarla. En $\lambda = \lambda_\star$ la fórmula da una sola solución; el programa de dos ramas está planteado para el intervalo abierto inferior.

El segundo, λ cerca de $\lambda_\star$: las dos soluciones se juntan —en $\lambda = 3.5$ están a 0.21 una de otra en el centro, frente a 3.95 en $\lambda = 1$—, el problema se vuelve sensible a las perturbaciones, y un arranque que servía para $\lambda = 1$ puede caer en la otra rama o en ninguna. Es una **bifurcación de nodo-silla**, y su firma numérica es siempre la misma: el método se vuelve delicado justo donde la matemática se vuelve delicada.

El tercero es de lectura, y es el que más engaña. Nada impide que las dos llamadas devuelvan **la misma** solución si los dos arranques caen en la misma cuenca; entonces el programa dibuja dos curvas encima y quien lo mire deprisa creerá haber encontrado las dos. Por eso el programa compara ambas aproximaciones con sus perfiles exactos antes de dibujarlas. El umbral exige además que el error sea menor que una cuarta parte de la separación de sus alturas centrales. Si no se cumple, se detiene: no afirma haber distinguido dos ramas. Por ejemplo, muy cerca del crítico dos arranques parabólicos distintos pueden acabar ambos en la rama alta incluso aunque el integrador anuncie convergencia.

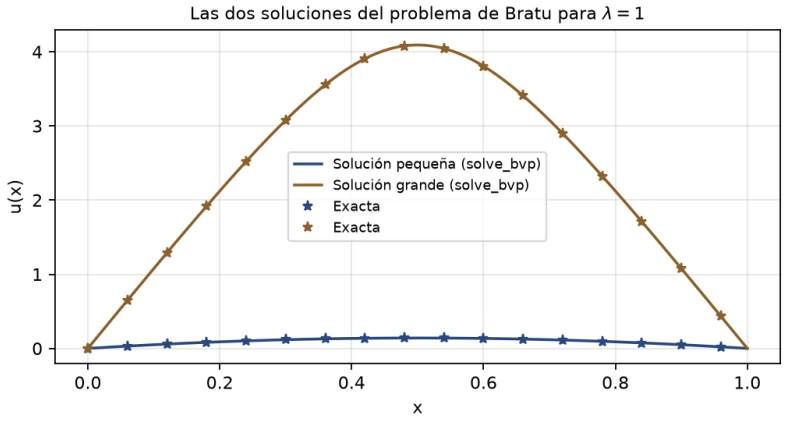


Figura 3.11: Las dos soluciones de Bratu para $\lambda = 1$, numéricas y exactas. La aproximación inicial influye en la rama encontrada; aquí ambas curvas se comparan con la fórmula exacta.

La idea: por qué dos, una o ninguna

Es el fenómeno del capítulo 3 en versión no lineal: la condición necesaria puede tener varias soluciones o ninguna. En una interpretación térmica, $u'' + \lambda e^u = 0$ describe un estado estacionario en el que la generación de calor por una reacción, proporcional a e^u, se compensa con la difusión hacia unos bordes mantenidos a temperatura de referencia $u = 0$. Es el modelo estacionario de Frank-Kamenetskii. Las ramas fría y caliente se unen en el valor crítico; por encima no hay un estado estacionario de esta forma. Para hablar de estabilidad o de cuánto tarda en crecer la temperatura hay que añadir una evolución como $u_t = u_{xx} + \lambda e^u$: la ecuación estacionaria sola no contiene esa información.

Versión en Python (NumPy/Matplotlib)

```python
import numpy as np
from scipy.integrate import solve_bvp
from scipy.optimize import brentq
def bratu(lam):
    ac = brentq(lambda a: a*np.tanh(a)-1,1,2)
    lcrit = 8*(ac/np.cosh(ac))**2
    if not (np.isfinite(lam) and 0<lam<lcrit):
        raise ValueError("Dos ramas solo para 0<lambda<lambda_crit")
    fun = lambda a: 8*(a/np.cosh(a))**2-lam
    hi = 2*ac
    while fun(hi)>0: hi *= 2
    aa = [brentq(fun,0,ac),brentq(fun,ac,hi)]
    MM = 2*np.log(np.cosh(aa)); gap = MM[1]-MM[0]
    xi = np.linspace(0,1,50); xx = np.linspace(0,1,1001)
    for a,M in zip(aa,MM):
        guess = np.vstack([4*M*xi*(1-xi),4*M*(1-2*xi)])
        sol = solve_bvp(lambda x,y: np.vstack([y[1],-lam*np.exp(y[0])]),
            lambda ya,yb: np.array([ya[0],yb[0]]),xi,guess,tol=1e-7,max_nodes=20000)
        if not sol.success: raise RuntimeError(sol.message)
        exact = 2*(np.log(np.cosh(a))-np.log(np.cosh(a*(1-2*xx))))
        err = np.max(np.abs(sol.sol(xx)[0]-exact))
```

```
        if err>min(1e-4,gap/4):
            raise RuntimeError("No se han verificado dos ramas; ajustar el arranque")
        print("u(1/2):",sol.sol(.5)[0]," alpha:",a," error:",err)
bratu(1.0)
```

QUÉ DEJA. Un problema de contorno no lineal se resuelve por iteraciones que arrancan de una aproximación inicial. Si se buscan dos soluciones, hay que comprobar que las dos respuestas son distintas y que satisfacen el problema. El análisis determina cuántas soluciones hay; la convergencia o el fallo de un cálculo, por sí solos, no sustituyen ese análisis.

Ejercicios 21, 22 y 23: autovalores de un problema de contorno

Ejercicio 3.11. *Autovalores y autofunciones de* $-y'' + f(x)y = \lambda y$

ORIGEN: hoja de ejercicios del tema 3, ejercicios 21 a 23. Los programas siguientes se desarrollan aquí.

Se considera el problema de autovalores

$$-y'' + 2\cos(2x)\, y = \lambda y \ \text{ en } (0,\pi), \qquad y(0) = y(\pi) = 0.$$

Sus autovalores se comportan asintóticamente como k^2 y los ceros de una autofunción separan los de la autofunción del autovalor siguiente. Escribir `Autofun(n)` que calcule un autovalor «próximo» a n^2 y una autofunción asociada, usando $\sin(nx)$ como aproximación inicial (ejercicio 21); generalizarla a `Autofun2(n, funcoef)` con un coeficiente real continuo $f \in C([0,\pi];\mathbb{R})$ (ejercicio 22); y a `Autofun3(n, funcoef)` con condiciones de Neumann $y'(0) = y'(\pi) = 0$ (ejercicio 23).

PISTAS, DE MENOR A MAYOR

1. *λ es una incógnita más: `bvp4c` admite parámetros desconocidos como tercer argumento de `bvpinit`, y entonces las funciones reciben (x, y, λ).*
2. *Con una incógnita más hace falta una condición más: una autofunción está definida salvo escala, así que se fija $y'(0) = 1$ (Dirichlet) o $y(0) = 1$ (Neumann).*
3. *Sin f, los autovalores serían exactamente n^2 con autofunciones $\sin(nx)$; con $f = 2\cos 2x$ cambian; el arranque no asegura por sí solo qué autovalor se encontrará.*

Resolución.
La idea: calcular también el parámetro. Hasta ahora λ era un dato y u la incógnita. Aquí buscamos ambos. Un sistema de dos ecuaciones de primer orden tiene dos constantes de integración; al añadir un parámetro desconocido hacen falta tres condiciones. Las dos del problema no fijan la escala: si y es una autofunción, también lo es $5y$. Por eso añadimos $y'(0) = 1$ en Dirichlet. Esta normalización se impone en el borde, pero no representa una condición física adicional: elige un representante de cada autofunción. Es válida porque $y(0) = y'(0) = 0$ obligaría a $y \equiv 0$ por unicidad. En Neumann se fija $y(0) = 1$ por el mismo motivo. Una normalización solo sirve si no excluye el modo buscado.
El método busca **un** autovalor a partir de una aproximación. `bvp4c` es local: el resultado depende tanto del perfil inicial como del valor inicial de λ, y puede converger a otro modo o fallar. Arrancamos de $\lambda = n^2$ y de un seno (o un coseno en Neumann), soluciones del

problema con $f = 0$. Usar un problema resuelto para preparar otro parecido reaparecerá en el capítulo 4. Después habrá que comprobar qué modo hemos obtenido.
El núcleo son tres líneas: `dydx`, que escribe $-y'' + fy = \lambda y$ como sistema; `condcon`, que tiene **tres** componentes y no dos; y el tercer argumento de `bvpinit`, que es lo que declara λ como incógnita. Quitar ese tercer argumento y dejar los otros dos produce un error de dimensiones, y es la manera más rápida de recordar por qué las condiciones son tres.
El problema. Con $f \equiv 0$ es $-y'' = \lambda y, y(0) = y(\pi) = 0$, cuyas autofunciones son los múltiplos no nulos de $\sin(nx)$, con $n \geq 1$ y $\lambda = n^2$: es el problema de Sturm-Liouville más simple, el de la cuerda vibrante. Con $f = 2\cos(2x)$ es la **ecuación de Mathieu**, y los autovalores no tienen una fórmula elemental como n^2, pero se sabe que son $\lambda_n \approx n^2$ para n grande y que las autofunciones entrelazan sus ceros.
Órdenes nuevas. `bvpinit(xmalla, yinicial, parametros)` con un tercer argumento declara parámetros desconocidos; entonces el segundo miembro y las condiciones de contorno se escriben con tres argumentos, `@(x,y,lambda)` y `@(ya,yb,lambda)`, y `condcon` debe devolver tantas componentes como ecuaciones *más* parámetros: aquí tres. El parámetro calculado está en `sol.parameters`. La aproximación inicial puede ser una función de x: `bvpinit(xmalla, @(x) [sin(n*x); n*cos(n*x)])`.

```
function [lambda, sol] = Autofun2(n, funcoef, neumann)
% Autofun2(n, funcoef): autovalor cercano a n^2 y autofuncion de
% -y'' + f(x) y = lambda y en (0,pi), con y(0) = y(pi) = 0.
% Con neumann = true, condiciones y'(0) = y'(pi) = 0 (ejercicio 23).
% Autofun(n) del ejercicio 21 es Autofun2(n, @(x) 2*cos(2*x)).
% Pruebas: Autofun2(1, @(x) 2*cos(2*x)) Autofun2(3, @(x) 2*cos(2*x)) Autofun2(1, @(x) 2*
    cos(2*x), true)
if nargin < 3, neumann = false; end
assert(isscalar(n) && isfinite(n) && n==fix(n) && n>=~neumann, ...
      'n entero: >=1 en Dirichlet, >=0 en Neumann');

dydx = @(x,y,lambda) [y(2); (funcoef(x) - lambda)*y(1)]; % -y'' + f y = lambda y
if neumann
  condcon = @(ya,yb,lambda) [ya(2); yb(2); ya(1) - 1]; % y'(0)=0, y'(pi)=0, normalizacion
       y(0)=1
  yinit = @(x) [cos(n*x); -n*sin(n*x)];
else
  condcon = @(ya,yb,lambda) [ya(1); yb(1); ya(2) - 1]; % y(0)=0, y(pi)=0, normalizacion y
      '(0)=1
  yinit = @(x) [sin(n*x)/n; cos(n*x)];
end
solinit = bvpinit(linspace(0,pi,max(50,12*n+1)),yinit,n^2);
opts = bvpset('RelTol',1e-7,'AbsTol',1e-9,'NMax',20000);
sol = bvp4c(dydx,condcon,solinit,opts);
lambda = sol.parameters;
xx = unique([sol.x, linspace(0,pi,max(400,40*(n+1)))]);
yy = deval(sol,xx,1); z = yy(2:end-1); % excluir los extremos
z = z(abs(z)>1e-8*max(abs(yy))); % evitar signos de redondeo cerca de cero
ceros = sum(z(1:end-1).*z(2:end)<0);
esperados = n-1+neumann;
fprintf('Arranque n=%i: lambda=%.6f, ceros interiores=%i\n',n,lambda,ceros)
if ceros~=esperados, warning('Revisar el modo: se esperaban %i ceros',esperados); end
plot(xx, deval(sol, xx, 1), 'LineWidth', 1.3), hold on
plot([0, pi], [0, 0], 'k'), xlabel('x'), title(['Autofuncion para \lambda = ', num2str(
```

```
    lambda)])
shg
end
```

SALIDA ESPERADA

```
>> f = @(x) 2*cos(2*x);
>> Autofun2(1, f); Autofun2(2, f); Autofun2(3, f); Autofun2(4, f);
Arranque n=1: lambda=-0.110249, ceros interiores=0
Arranque n=2: lambda=3.917025, ceros interiores=1
Arranque n=3: lambda=9.047739, ceros interiores=2
Arranque n=4: lambda=16.032970, ceros interiores=3
>> Autofun2(0, f, true); Autofun2(1, f, true); Autofun2(2, f, true);
Arranque n=0: lambda=-0.455139, ceros interiores=0
Arranque n=1: lambda=1.859108, ceros interiores=1
Arranque n=2: lambda=4.371301, ceros interiores=2
```

Para este coeficiente, los valores de Dirichlet se acercan a n^2: la diferencia $\lambda_n - n^2$ pasa de -1.11 a -0.083, 0.048 y 0.033. Para un f general, $\lambda_n \sim n^2$ significa que el *cociente* tiende a uno, no necesariamente que la diferencia tienda a cero: si $f \equiv c$, se obtiene $\lambda_n = n^2 + c$.
El primer autovalor de este ejemplo es negativo. No contradice el término $-y''$: multiplicando por y e integrando por partes, las condiciones del borde dan

$$\lambda = \frac{\int_0^\pi (y'^2 + f y^2)\, dx}{\int_0^\pi y^2\, dx}.$$

Un coeficiente negativo puede reducir este cociente. Que f sea negativo en algún punto no basta para asegurar un autovalor negativo; aquí lo confirma el cálculo de Mathieu.
El control mediante las funciones de Mathieu. El cambio $y'' + (\lambda - 2\cos 2x)y = 0$ es exactamente la **ecuación de Mathieu** en su forma estándar $y'' + (a - 2q\cos 2x)y = 0$ con $q = 1$ y $a = \lambda$. Las condiciones $y(0) = y(\pi) = 0$ seleccionan las soluciones impares, cuyos valores característicos se llaman $b_m(q)$, y las de Neumann las pares, con valores $a_m(q)$. La DLMF del NIST describe estos valores característicos (§28.2); en SciPy se calculan con `scipy.special.mathieu_b` y `mathieu_a`. Comparadas una a una:

$$b_1(1) = -0.110249, \quad b_2(1) = 3.917025, \quad b_3(1) = 9.047739, \quad b_4(1) = 16.032970,$$

$$a_0(1) = -0.455139, \quad a_1(1) = 1.859108, \quad a_2(1) = 4.371301,$$

que coinciden con las seis cifras decimales mostradas. Comparar con otro algoritmo permite detectar errores del cálculo de contorno, siempre que se traduzcan bien la ecuación, el parámetro q y las condiciones del borde; la comparación también exige esas comprobaciones.
Hay otros dos controles útiles. El primero es el **recuento de ceros**: la autofunción n de Dirichlet tiene $n - 1$ ceros interiores; con la numeración de Neumann que empieza en $n = 0$, tiene n. El programa cuenta cambios de signo excluyendo los extremos y valores próximos al cero por redondeo. Es una comprobación numérica: hay que refinar la malla si no resuelve las oscilaciones. El valor de n que imprime es el del *arranque*; el recuento ayuda a comprobar si coincide con el índice del modo obtenido.
El segundo control es la **ortogonalidad**. Restar las ecuaciones de dos autofunciones multiplicadas una por la otra da

$$(\lambda_m - \lambda_n) \int_0^\pi y_m y_n\, dx = [y_m y_n' - y_m' y_n]_0^\pi = 0.$$

El término del borde se anula tanto en Dirichlet como en Neumann. Si los autovalores son distintos, la integral es cero. Numéricamente conviene dividir su valor absoluto por $\|y_m\|_{L^2}\|y_n\|_{L^2}$, para no depender de la escala. La precisión de `trapz` depende también de la malla de integración: no tiene por qué ser la tolerancia del integrador.
Dónde se rompe. Con un coeficiente grande, como $f = 20\cos 2x$, el arranque basado en $f = 0$ puede acabar en un modo distinto del esperado. No se decide por cuál autovalor esté más cerca de n^2: importa la aproximación de la función completa. Un cálculo convergido con un número de ceros distinto del esperado no ha encontrado el modo pedido; hay que cambiar el arranque, por ejemplo continuándolo desde un coeficiente más pequeño. Con el código de Python de abajo, $f = 20\cos 2x$ y arranque $n = 1$ se obtiene aproximadamente $\lambda = 7.986069$ y dos ceros interiores: el aviso detecta que es el tercer modo de Dirichlet, no el primero.
La segunda rotura es la normalización. Fijar $y'(0) = 1$ funciona en Dirichlet porque $y(0) = 0$ obliga a $y'(0) \neq 0$; en Neumann hay que cambiarla por $y(0) = 1$, y por eso el programa tiene dos ramas. Elegir la normalización equivocada —pedir $y'(0) = 1$ con condiciones de Neumann, que ya imponen $y'(0) = 0$— da un sistema incompatible, y el método falla sin explicar que el problema está sobredeterminado.

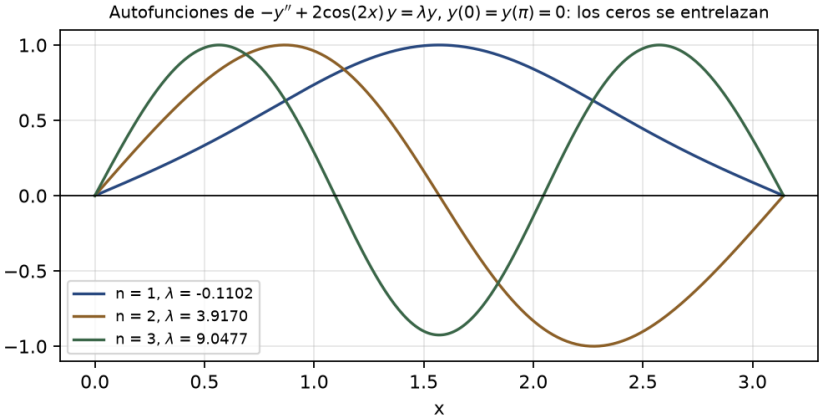


Figura 3.12: Las tres primeras autofunciones de $-y'' + 2\cos(2x)y = \lambda y$ con $y(0) = y(\pi) = 0$, escaladas a máx $|y| = 1$ para compararlas. La n-ésima tiene $n - 1$ ceros interiores. Contando también 0 y π como ceros, entre cada dos ceros consecutivos de la n-ésima hay exactamente uno de la $(n + 1)$-ésima.

Conexión: Sturm-Liouville

El problema $-y'' + f(x)y = \lambda y$ con condiciones separadas es el problema de Sturm-Liouville del libro de EDO, §5.10 «Problema de Sturm-Liouville» (página 117, con el ejemplo $y'' = \lambda y$, $y(0) = y(1) = 0$ en §5.10.1 y el «Teorema (ortogonalidad de las autofunciones)» en §5.10.2): los autovalores forman una sucesión $\lambda_1 < \lambda_2 < \cdots \to +\infty$, las autofunciones son ortogonales en $L^2(0, \pi)$ y la n-ésima de Dirichlet tiene $n - 1$ ceros interiores. La ortogonalidad se puede comprobar numéricamente integrando el producto de dos autofunciones con `trapz`. Y en el capítulo 4 reaparece la idea de L^2 como espacio de Hilbert, ahora para los controles.

Versión en Python (NumPy/Matplotlib)

```
1 import numpy as np
```

```
from scipy.integrate import solve_bvp
def autofun(n, fcoef, neumann=False):
    if not (np.isfinite(n) and n==int(n) and n>=int(not neumann)):
        raise ValueError("n entero: >=1 Dirichlet, >=0 Neumann")
    n = int(n)
    dydx = lambda x,y,p: np.vstack([y[1],(fcoef(x)-p[0])*y[0]])
    if neumann:
        bc = lambda ya,yb,p: np.array([ya[1],yb[1],ya[0]-1])
        yg = lambda x: np.vstack([np.cos(n*x),-n*np.sin(n*x)])
    else:
        bc = lambda ya,yb,p: np.array([ya[0],yb[0],ya[1]-1])
        yg = lambda x: np.vstack([np.sin(n*x)/n,np.cos(n*x)])
    xi = np.linspace(0,np.pi,max(50,12*n+1))
    s = solve_bvp(dydx,bc,xi,yg(xi),p=[float(n*n)],tol=1e-7,max_nodes=20000)
    if not s.success: raise RuntimeError(s.message)
    xx = np.union1d(s.x,np.linspace(0,np.pi,max(400,40*(n+1))))
    yy = s.sol(xx)[0]; z = yy[1:-1]
    z = z[np.abs(z)>1e-8*np.max(np.abs(yy))]
    ceros = np.count_nonzero(z[:-1]*z[1:]<0)
    print("Arranque:",n," lambda:",s.p[0]," ceros interiores:",ceros)
    if ceros!=n-1+neumann: print("Revisar: no coincide el modo esperado")
    return s
for n in (1,2,3): autofun(n,lambda x: 2*np.cos(2*x))
```

Repaso intercalado: capítulo 3, en tres rondas

A libro cerrado, en sesiones distintas. Un tercio de este capítulo y dos tercios de los anteriores y de las asignaturas en que se apoya.

Primera ronda, al terminar el capítulo.

1. Enuncia el lema fundamental y la condición de Euler-Lagrange, y di en qué paso de la demostración se usa cada hipótesis ($h(a) = h(b) = 0$, regularidad de $\nabla_{x'} f$).
2. (Capítulo 2) ¿Qué dice el teorema de las EDO autónomas escalares sobre $y' = y(K - y) - h$? Distingue los tres casos del discriminante.
3. (Capítulo 1) ¿Qué garantiza el teorema de Perron para la matriz de Google y por qué hace falta el factor de amortiguación?
4. Escribe la identidad de Beltrami y dedúcela. Para $L = E_c - E_p$ sin dependencia explícita del tiempo y E_c cuadrática en la velocidad, ¿qué cantidad mecánica conserva?
5. (EDO) ¿Cuándo tiene solución única el problema de contorno $x'' + \omega^2 x = 0$, $x(0) = x_a$, $x(\tau) = x_b$? Relaciónalo con el muelle de la sección 3.3.1.
6. ¿Cómo cambian las condiciones y la aproximación inicial al pasar de `ode45` a `bvp4c`?

Segunda ronda, una semana después.

1. Deduce las condiciones de Robin de la cuerda con muelles y comprueba la solución.
2. (Capítulo 2) Demuestra que las órbitas de Volterra están en las curvas $\lambda\psi(u) + \psi(v) = C$ y compáralo con la conservación de la energía del péndulo: ¿qué juega el papel de ψ?
3. (Capítulo 1) Escribe la ecuación $\varphi(\lambda) = 1$ de Leslie y di por qué tiene una sola raíz positiva. ¿Qué tiene en común con «$\partial_{u'} f$ inyectiva» en el problema de la distancia

mínima?
4. Resuelve la braquistócrona con Beltrami hasta $u(1+u'^2) = \text{const}$ y explica por qué sale una cicloide.
5. (EDO) Enuncia el teorema de Liapunov y úsalo para probar que el péndulo colgando es estable. ¿Qué le falta a ese argumento para dar estabilidad asintótica con rozamiento?
6. Clasifica los equilibrios del péndulo por niveles de energía: ¿qué pasa en $E = g/\ell$?

Tercera ronda, de repaso final.
1. Demuestra el teorema de convexidad y da un ejemplo de f para el que Euler-Lagrange no basta (la superficie de revolución).
2. (Capítulo 2) En el modelo SIR, ¿por qué ningún equilibrio individual es asintóticamente estable, aunque las soluciones se acerquen al conjunto $I = 0$? ¿Cómo se prueba que $I \to 0$?
3. (Capítulo 1) Programa de memoria el algoritmo de la potencia para PageRank con matrices huecas: ¿qué teorema garantiza la convergencia y qué criterio permite detenerlo con una tolerancia dada?
4. Deduce las ecuaciones de la varilla y explica por qué se desacoplan el centro de masas y el ángulo. Corrige la errata del coeficiente de $|\varphi'|^2$.
5. (Capítulo 4, anticipo) En control óptimo el funcional es $\frac{\alpha}{2}\int |y - y_d|^2 + \frac{\mu}{2}\int |v|^2$ y la incógnita es v, no y. ¿Cómo cambiaría el método «perturbar, derivar en $\sigma = 0$, integrar por partes» si la relación entre v e y es una EDO?
6. Escribe el problema de Bratu como sistema y explica cómo influye el arranque en las iteraciones. ¿Por qué dos cálculos convergidos no aseguran haber encontrado dos soluciones?

CAPÍTULO 4

Problemas de control óptimo para EDOs

ANTES DE LEER

Cuatro preguntas antes de leer, a libro cerrado.

1. En el capítulo 3 la incógnita era la trayectoria. Si ahora la trayectoria está gobernada por una EDO y lo que se elige es un dato de esa EDO (una fuente, un coeficiente), ¿qué se minimiza, sobre qué conjunto, y qué papel hace la EDO?
2. El método directo de mínimos necesita extraer una subsucesión convergente. En un espacio de funciones una sucesión acotada puede no tenerla en norma. ¿Qué noción de convergencia puede servir?
3. El método «perturbar, derivar, integrar por partes» del capítulo 3 daba una EDO de segundo orden. Aquí va a dar *dos* EDOs de primer orden acopladas, una de ellas hacia atrás en el tiempo. ¿De dónde sale la segunda?
4. Si el control tiene que estar entre dos cotas, ¿cómo cambia la condición de optimalidad? Piensa en minimizar una parábola en un intervalo cerrado.

Las respuestas están en las secciones 4.1 (las tres primeras) y 4.6.

En los problemas de control óptimo que se consideran hay siempre tres cosas:

- un problema de valores iniciales para una EDO o un sistema de EDOs, donde la solución, el **estado**, depende de uno de los datos (un coeficiente, un dato inicial, un segundo miembro), que es el **control**;
- un conjunto de **controles admisibles**;
- una función de **coste** definida en ese conjunto, con valores reales que dependen del control y del estado.

En todos los casos, el problema consiste en **minimizar el coste en el conjunto de controles admisibles**. Es la última etapa del esquema de modelado: no solo describir y resolver el sistema, sino *actuar* sobre él eligiendo algunos de los datos para conseguir un comportamiento deseable. Los modelos de los capítulos anteriores reaparecen como ecuaciones de estado: la población repoblada desde fuera, la pesca con esfuerzo variable, el tumor bajo terapia, el

ecosistema con un catalizador. Y la técnica es la del capítulo 3, con dos añadidos: se demuestra la existencia del mínimo, con un argumento de compacidad débil en L^2, y la condición de optimalidad se escribe con una función auxiliar, el **estado adjunto**, que resuelve una EDO hacia atrás.

CONEXIÓN: LO QUE SE USA DE ANÁLISIS Y DE EDO, REENUNCIADO DONDE HACE FALTA

De EDO: la fórmula de variación de constantes y el lema de Gronwall (apéndice A, teoremas sobre sistemas lineales y de Gronwall), y la reducción de una EDO de segundo orden a un sistema. De Análisis funcional, tres hechos sobre el espacio $L^2(0,T)$ que se reenuncian en la sección 4.1 en el punto en que se usan: que es un espacio de Hilbert, que sus bolas son débilmente compactas, y el teorema de la convergencia dominada. El primero está en la introducción del libro de EDO del autor (§1.3 «Espacios de Banach y de Hilbert», con §1.3.3 «Sobre el espacio L^2»), y la jerarquía de las convergencias (uniforme, en L^p, puntual, en medida) en §1.4; los otros dos **no** están en ese libro, y por eso se enuncian aquí con su idea de demostración.

4.1 Problemas básicos (I): EDO lineal sin restricciones con observación en $(0,T)$

LA PREGUNTA

Un gestor de una reserva regula una población mediante una repoblación o una captura a ritmo $v(t)$. Quiere que siga una trayectoria deseada $y_d(t)$ durante el año, pero la intervención cuesta recursos. ¿Existe una política de intervención que equilibre las dos cosas de la mejor manera posible, es única, y cómo se calcula sin probar todas las políticas?

Se supone que la **ecuación de estado** es

$$\begin{cases} y' = ay + v, & t \in (0,T), \\ y(0) = y_0, \end{cases} \tag{4.1}$$

y se considera el **problema de control básico**

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{\alpha}{2}\displaystyle\int_0^T |y - y_d|^2\,dt + \dfrac{\mu}{2}\int_0^T |v|^2\,dt \\ \text{Sujeto a: } v \in L^2(0,T),\ (y,v) \text{ verifica (4.1).} \end{cases} \tag{4.2}$$

Los datos son $\alpha > 0$, $\mu > 0$, $T > 0$, $a \in \mathbb{R}$, $y_0 \in \mathbb{R}$ e $y_d \in C^0([0,T])$. Por ejemplo, $y = y(t)$ es una población con una entrada o salida externa de individuos a ritmo $v = v(t)$: es el modelo exponencial del capítulo 2 con un término fuente. Se dice que y es el **estado**, v el **control**, $L^2(0,T)$ el conjunto de controles admisibles y J el **funcional de coste**. El primer término de J mide cuánto se aparta el estado del deseado a lo largo de $(0,T)$ (**observación en** $(0,T)$); el segundo, cuánto cuesta el control. Los pesos α y μ fijan el compromiso.

Por qué el modelo es este, término a término. La ecuación de estado $y' = ay + v$ es el modelo exponencial del capítulo 2 con una entrada: ay es lo que la población hace sola

(nacimientos menos muertes, con $a < 0$ si se extingue sin ayuda) y $v(t)$ son los individuos que el gestor añade o quita por unidad de tiempo. Que v entre *sumando* y no multiplicando es una decisión: sumando es una repoblación o una captura, que no cambia la biología; multiplicando ($y' = ay + vy$) sería alterar la tasa, por ejemplo alimentando, y ese caso es el control bilineal de la sección 4.8. En el coste, cada pieza tiene una razón. Que la desviación se mida con $|y - y_d|^2$ y no con $|y - y_d|$: el cuadrado hace que duplicar una desviación multiplique por cuatro su penalización. Es una elección de modelado; además resulta derivable y convexo. La unicidad quedará asegurada por el término estrictamente convexo que penaliza el control. Que se integre en $(0, T)$: importa toda la trayectoria; si solo importara el final, la integral se sustituye por $|y(T)|^2$, sección 4.2. Que el coste del control sea $\frac{\mu}{2} \int |v|^2$ y no $\int |v|$, que sería lo natural si cada individuo costara lo mismo: el cuadrado da convexidad estricta y una cota de la norma L^2 de los controles minimizantes. Esa cota no es una cota puntual: L^2 permite picos altos en intervalos cortos. Un coste lineal puede conducir a controles *bang-bang* cuando se imponen cotas, pero no lo exige siempre; un coste $\int |v|$ tampoco es lineal en controles que cambian de signo. Y que no haya cotas sobre v: es la simplificación más visible. La sección 4.6 introduce cotas sobre la intervención; si además se exige $y \geq 0$, eso es una restricción sobre el estado que no garantizan por sí solas unas cotas arbitrarias de v. Lo que se pierde, en resumen: la no linealidad de la biología (sección 4.5), las cotas físicas del control, y un coste con sentido económico literal. Lo que se gana: un problema que se resuelve entero.

EL CAMINO QUE NO SE TOMÓ: POR QUÉ EL CONTROL VIVE EN L^2 Y NO EN C^0

La alternativa. Buscar el control entre las funciones continuas, o C^1, como en el capítulo 3.
Su ventaja. La ecuación de estado se resuelve en sentido clásico y no hace falta hablar de soluciones «en casi todo punto».
La dificultad. Probar la existencia. Para probar que el mínimo existe se toma una sucesión minimizante y hay que extraer una subsucesión convergente; en C^0 con la norma del máximo haría falta controlar también la equicontinuidad, cosa que el coste no proporciona. En $L^2(0, T)$, que es un espacio de Hilbert, toda sucesión acotada tiene una subsucesión débilmente convergente, y la cota la da el propio coste, $\|v\|^2 \leq \frac{2}{\mu} J(v)$. Es exactamente lo que faltaba en el contraejemplo del capítulo 3 (el $\arctan(nt)$ que tendía a un escalón). El precio es extender la noción de solución de (4.1) a controles que solo son de cuadrado integrable, y eso es la definición siguiente. Esto no afirma que el óptimo sea discontinuo: en el problema básico resultará incluso C^1 por la ecuación adjunta. Trabajar primero en L^2 permite probar su existencia sin imponer esa regularidad de antemano. Con restricciones pueden aparecer esquinas.

Hacen falta controles en el espacio de Hilbert $L^2(0, T)$, luego hay que extender el concepto de solución de un problema de Cauchy como (4.1).

Definición. Sea $v \in L^2(0, T)$. Se llama **solución maximal** de (4.1) a una función $y = y_v$ **absolutamente continua** en $[0, T]$ que cumple $y(0) = y_0$ y $y' = ay + v$ en casi todo punto. Equivalentemente, es una función continua que satisface

$$y(t) = y_0 + \int_0^t \big(ay(s) + v(s)\big)\, ds \qquad (0 \leq t \leq T).$$

CONEXIÓN: $L^2(0, T)$ Y «CASI TODO PUNTO»

$L^2(0, T)$ es el espacio de las funciones $v : (0, T) \to \mathbb{R}$ medibles con $\int_0^T |v|^2\, dt < \infty$, identificando las que coinciden salvo en un conjunto de medida nula, con el producto escalar $(u, v) =$

$\int_0^T uv\,dt$ y la norma $\|v\| = (\int |v|^2)^{1/2}$. Es un espacio de Hilbert: completo respecto de esa norma. Una función de L^2 puede ser discontinua, como un control que se enciende y se apaga; además sus valores en conjuntos de medida nula no están fijados. Por eso la ecuación se exige en casi todo punto.
La continuidad absoluta es precisamente lo que permite recuperar una función integrando su derivada: $y(t) = y(0) + \int_0^t y'(s)\,ds$, con $y' \in L^1$. La continuidad y la derivabilidad en casi todo punto, por sí solas, no bastan. La función de Cantor es continua, tiene derivada cero casi en todas partes y no es constante; sin esta hipótesis adicional, $y' = 0$, $y(0) = 0$ no tendría la unicidad que vamos a usar. Esta precisión corrige la definición del material de partida. Para repasar el espacio L^2 y comparar su norma con la del máximo, véase el libro de EDO, §1.3.3 «Sobre el espacio L^2» (página 14).

Teorema (lema de existencia y unicidad para la ecuación de estado)**.** Para cada $v \in L^2(0, T)$ e $y_0 \in \mathbb{R}$, existe una única solución maximal $y = y_v$ de (4.1), dada por

$$y(t) = e^{at} y_0 + \int_0^t e^{a(t-s)} v(s)\,ds \qquad \text{para todo } t \in [0, T]. \tag{4.3}$$

La idea, en una frase. No hay nada que buscar: la fórmula (4.3) es la variación de constantes de siempre, y lo único que hay que comprobar es que sigue funcionando cuando el segundo miembro v es apenas de cuadrado integrable, sin ninguna continuidad. La respuesta es que sí: la ecuación integra a v, y una integral indefinida de una función integrable es absolutamente continua.

Paso 1: existencia. Hay que comprobar que la función dada por (4.3) verifica las propiedades pedidas. Escribimos $y(t) = e^{at}\big(y_0 + \int_0^t e^{-as} v(s)\,ds\big)$. Como $\int_0^T |v| \le \sqrt{T}\|v\|$ por Cauchy-Schwarz y e^{-as} está acotada en $[0, T]$, la integral es absolutamente continua. Multiplicarla por e^{at} conserva esa propiedad. Es derivable en casi todo punto, con derivada $ae^{at}y_0 + v(t) + a\int_0^t e^{a(t-s)}v(s)\,ds = ay(t) + v(t)$, por el teorema fundamental del cálculo para integrales de Lebesgue (válido en los puntos de Lebesgue de v, que son casi todos). Y $y(0) = y_0$. Es la fórmula de variación de constantes del apéndice (teorema sobre sistemas lineales), extendida a segundos miembros de L^2.
Paso 2: unicidad, por Gronwall (el núcleo). Si y y z son dos soluciones maximales, $w := y - z$ es absolutamente continua, con $w' = aw$ en casi todo punto y $w(0) = 0$. Ahora sí podemos integrar la derivada:

$$w(t) = a\int_0^t w(s)\,ds \qquad \text{para todo } t \in [0, T],$$

luego $|w(t)| \le \int_0^t |a||w(s)|\,ds$, y por el lema de Gronwall (teorema de Gronwall del apéndice, con $C = 0$), $w \equiv 0$. ∎

Qué hace cada hipótesis, y dónde está esto demostrado con más detalle. La diferencia w satisface $|w(t)| \le |a|\int_0^t |w|$, y Gronwall con constante inicial nula fuerza $w \equiv 0$: es el mecanismo estándar de unicidad, y conviene retenerlo porque es el que se repite en cuanto la ecuación deja de ser lineal. La linealidad de la ecuación es lo que hace que aquí la desigualdad salga con constante $|a|$ exacta; en el caso no lineal de la sección 4.5 el mismo argumento se aplica con la constante de Lipschitz en su lugar. El lema de Gronwall está demostrado, con su intuición y su aplicación, en el libro de EDO,

> §3.1 «Lema de Gronwall» del libro de EDO del autor, y la fórmula de variación de constantes para segundos miembros continuos, en libro de EDO, §4.5 «El SDO lineal no homogéneo. Método de Lagrange de variación de las constantes»; lo único que aporta esta demostración sobre aquello es que la fórmula sobrevive al pasar de v continua a $v \in L^2$, y que la derivada se cumple entonces solo en casi todo punto.

Observación. El funcional J es **cuadrático**: se escribe $J(v) = \frac{1}{2}a(v,v) + L(v) + c$ con $a(\cdot,\cdot)$ una forma bilineal continua sobre $L^2(0,T)$, L una forma lineal continua y $c \in \mathbb{R}$. En efecto, el estado se descompone como $y_v = Y_0 + z_v$, donde Y_0 es la solución de (4.1) con $v \equiv 0$ (el sistema sin controlar, $Y_0 = e^{at}y_0$) y z_v es la solución con $y_0 = 0$, que es *lineal* en v por (4.3). Entonces

$$J(v) = \frac{\mu}{2}\|v\|^2 + \frac{\alpha}{2}\|z_v\|^2 + \alpha\,(z_v, Y_0 - y_d) + \frac{\alpha}{2}\|Y_0 - y_d\|^2,$$

con normas y productos en $L^2(0,T)$: los dos primeros términos son la parte cuadrática, el tercero la lineal y el cuarto la constante. De esta expresión se deduce que J es **estrictamente convexo** en $L^2(0,T)$:

$$J\big((1-\theta)v_1 + \theta v_2\big) < (1-\theta)J(v_1) + \theta J(v_2) \qquad \text{para } v_1 \neq v_2,\ \theta \in (0,1), \tag{4.4}$$

porque $\frac{\mu}{2}\|v\|^2$ es estrictamente convexo ($\mu > 0$), $\frac{\alpha}{2}\|z_v\|^2$ es convexo (la composición de la aplicación lineal $v \mapsto z_v$ con una norma al cuadrado) y los otros dos términos son afines.

De dónde sale la convexidad, y qué pasa cuando el coste no es cuadrático. La observación demuestra la convexidad estricta desarrollando el cuadrado, y eso tapa la razón de fondo. Son tres hechos, y ninguno usa que la observación sea un cuadrado.

1. *El estado depende del control de forma afín.* Por (4.3), $y_v = Y_0 + z_v$ con z_v lineal en v, luego $y_{(1-\theta)v_1+\theta v_2} = (1-\theta)y_{v_1} + \theta y_{v_2}$: el estado de una mezcla de controles es la misma mezcla de los estados.
2. *Una función convexa compuesta con una aplicación afín es convexa.* Si Φ es convexa como función del estado, $\Phi\big(y_{(1-\theta)v_1+\theta v_2}\big) = \Phi\big((1-\theta)y_{v_1} + \theta y_{v_2}\big) \leq (1-\theta)\Phi(y_{v_1}) + \theta\Phi(y_{v_2})$. La igualdad es el punto 1 y la desigualdad es la convexidad de Φ. El término de observación es convexo en v en cuanto lo es en y.
3. *El coste del control es estrictamente convexo.* Desarrollando los dos lados con el producto escalar se obtiene la identidad

$$\|(1-\theta)v_1 + \theta v_2\|^2 = (1-\theta)\|v_1\|^2 + \theta\|v_2\|^2 - \theta(1-\theta)\|v_1 - v_2\|^2,$$

   que dice que la norma al cuadrado queda por debajo de la cuerda, y con un margen que se lee en la fórmula: $\theta(1-\theta)\|v_1 - v_2\|^2 > 0$ en cuanto $v_1 \neq v_2$.

Convexo más estrictamente convexo es estrictamente convexo. En una frase: *la convexidad la aporta la observación, y la estrictez la paga siempre el término* $\frac{\mu}{2}\|v\|^2$. Por eso el argumento vale, sin cambiar nada, para un término final como $\frac{\alpha}{4}|y(T) - y_d|^4$ (la función $r \mapsto r^4$ es convexa y $v \mapsto y_v(T)$ es afín) o para $\alpha\int_0^T |y - y_d|\,dt$, que ni siquiera es derivable.

Un ejemplo en el que falla. Lo que no se puede quitar es la convexidad de la observación. Sean $a = 0$, $y_0 = 0$ y el coste

$$J(v) = \frac{\alpha}{4}\big(y(T)^2 - 1\big)^2 + \frac{\mu}{2}\int_0^T |v|^2\,dt,$$

que premia por igual llegar a $y(T)=1$ o a $y(T)=-1$: la función $r\mapsto (r^2-1)^2$ es un pozo doble, con dos mínimos y una joroba en $r=0$. Como $y(T)=\int_0^T v$, el control nulo da $J(0)=\alpha/4$, y los constantes $v=\pm 1/T$ llevan a $y(T)=\pm 1$ con $J(\pm 1/T)=\mu/(2T)$. Si $\mu<\alpha T/2$, el control nulo, que es el punto medio del segmento entre $1/T$ y $-1/T$, cuesta más que la media de los extremos: J no es convexo. Además $y_{-v}=-y_v$, luego $J(-v)=J(v)$ y los óptimos van por parejas.

Se puede llegar al final a mano. Sea $s:=y(T)=\int_0^T v$. Cauchy-Schwarz contra la función 1 da $s^2\le T\int_0^T |v|^2$, con igualdad solo si v es constante: entre todos los controles que llevan al mismo s, el más barato es el constante $v=s/T$. El problema se reduce a minimizar en $\mathbb{R}$

$$g(s)=\frac{\alpha}{4}(s^2-1)^2+\frac{\mu}{2T}\,s^2,\qquad g'(s)=s\Big(\alpha(s^2-1)+\frac{\mu}{T}\Big).$$

Si $\mu<\alpha T$, g tiene un máximo local en $s=0$ y dos mínimos en $s=\pm\sqrt{1-\mu/(\alpha T)}$: hay *exactamente dos* controles óptimos, $\hat v=\pm\frac{1}{T}\sqrt{1-\mu/(\alpha T)}$. Si $\mu\ge\alpha T$, en cambio, $g(s)-g(0)=\frac{\alpha}{4}s^4+\frac{s^2}{2}\big(\frac{\mu}{T}-\alpha\big)\ge 0$ y el único óptimo es $\hat v=0$. Con $\alpha=T=1$ y $\mu=\frac14$: $J(0)=0.25$, $J(\pm1)=0.125$, y los dos óptimos son $\hat v=\pm\sqrt{3}/2\approx\pm0.866$, con $J=7/64\approx 0.109$. La lectura: un control suficientemente caro devuelve la unicidad a costa de no hacer nada; pero la unicidad para *todo* $\mu>0$, que es la que da el teorema siguiente, sale de que la observación sea convexa.

En relación con el problema (4.2), las cuestiones que interesan son tres:

(A) La **existencia y unicidad** de solución. A toda solución, es decir, a todo control $\hat v\in L^2(0,T)$ con $J(\hat v)\le J(v)$ para todo $v\in L^2(0,T)$, se le llama **control óptimo**.

(B) La **caracterización**: aceptando que existe, deducir propiedades que deben cumplir los controles óptimos, las **condiciones de optimalidad**, análogas a las condiciones necesarias de mínimo del Cálculo de Variaciones.

(C) El **cálculo**: hay dos maneras esencialmente distintas, usar las condiciones de optimalidad o recurrir a algoritmos de descenso; en ambos casos se genera una sucesión de controles, cuya convergencia hay que estudiar.

(A) Existencia y unicidad

La existencia combina una cota de los controles con compacidad débil y semicontinuidad del coste; la unicidad procede de su convexidad estricta.

Teorema. En las condiciones anteriores, el problema (4.2) posee exactamente una solución $\hat v\in L^2(0,T)$.

CONEXIÓN: LOS TRES HECHOS DE ANÁLISIS QUE USA LA DEMOSTRACIÓN

Compacidad débil en un Hilbert. Toda sucesión acotada en un espacio de Hilbert H tiene una subsucesión que converge *débilmente*: existe $\hat v\in H$ con $(v_n,v)\to(\hat v,v)$ para todo $v\in H$. Es el sustituto de Bolzano-Weierstrass en dimensión infinita, donde las bolas no son compactas para la norma. **Semicontinuidad de la norma.** Si $v_n\rightharpoonup\hat v$ débilmente, entonces $\|\hat v\|\le\liminf\|v_n\|$, y $(v_n,\hat v)\to\|\hat v\|^2$ (tomar $v=\hat v$ en la definición). **Convergencia dominada.** Si $f_n\to f$ en casi todo punto y $|f_n|\le g$ con g integrable, entonces $\int f_n\to\int f$.
Dónde están. El libro de EDO del autor presenta en §1.3 el espacio de Hilbert L^2 y en §1.4 la jerarquía de las convergencias, pero **no** la convergencia débil ni el teorema de la convergencia dominada; no se puede remitir allí, así que se da aquí lo que falta. La convergencia dominada es el teorema de Lebesgue de la teoría de la integral, y se admite. La compacidad débil se

prueba así en $L^2(0,T)$, que es separable: sea $\{w_k\}$ una sucesión densa (por ejemplo, polinomios con coeficientes racionales) y $\|v_n\| \le M$. Para cada k la sucesión numérica (v_n, w_k) está acotada por $M\|w_k\|$; por el procedimiento diagonal hay una subsucesión, que se sigue llamando v_n, con (v_n, w_k) convergente para todo k. Para w cualquiera y $\varepsilon > 0$ se elige w_k con $\|w - w_k\| < \varepsilon$; entonces $|(v_n - v_m, w)| \le |(v_n - v_m, w_k)| + 2M\varepsilon$, luego (v_n, w) es de Cauchy y converge a un número $\ell(w)$. La aplicación $w \mapsto \ell(w)$ es lineal y $|\ell(w)| \le M\|w\|$, y el teorema de representación de Riesz (toda forma lineal continua en un Hilbert es el producto escalar por un vector fijo) da $\hat{v}$ con $\ell(w) = (\hat{v}, w)$: eso es $v_n \rightharpoonup \hat{v}$. La semicontinuidad es una línea: $\|\hat{v}\|^2 = \lim(v_n, \hat{v}) \le \liminf \|v_n\|\,\|\hat{v}\|$. Y la imagen de por qué una sucesión puede converger débilmente sin converger en norma: las funciones $\operatorname{sen}(nt)$ tienden débilmente a cero (contra cualquier w fija, la integral oscila y se cancela) mientras que $\|\operatorname{sen}(nt)\|^2 = \int_0^T \operatorname{sen}^2(nt)\,dt \to T/2$, que no es cero.

LA IDEA: EL MISMO ARGUMENTO EN $\mathbb{R}^n$, Y QUÉ CAMBIA AQUÍ

Para ver que una función continua $F : \mathbb{R}^n \to \mathbb{R}$ con $F(v) \ge c|v|^2$, $c > 0$, alcanza su mínimo se hace esto: se toma una sucesión v_n con $F(v_n) \to \inf F$; como $F(v_n)$ está acotada, $|v_n|$ está acotada; por Bolzano-Weierstrass hay una subsucesión que converge a un $\hat{v}$; y por continuidad $F(\hat{v}) = \lim F(v_n) = \inf F$. Aquí F es J, el espacio es $L^2(0,T)$, y fallan dos cosas: en dimensión infinita una sucesión acotada puede no tener ninguna subsucesión convergente en norma (las $\operatorname{sen}(nt)$ de la caja anterior), y J no tiene por qué ser continua para la convergencia débil, que es la única que se tiene. La demostración repara las dos cosas: la compacidad débil da un candidato $\hat{v}$ (paso 2), y después se prueba *a mano* que $J(\hat{v})$ no puede ser mayor que el ínfimo, usando que la parte del coste que depende del estado *sí* pasa bien al límite (paso 3, porque el estado se obtiene integrando el control, y la integral suaviza) y que la norma al cuadrado no sube al pasar al límite débil (paso 4). Léase primero la lista de los cuatro pasos y solo después las cuentas de cada uno.

Existencia, por el método directo del Cálculo de Variaciones. El funcional J está acotado inferiormente, $J(v) \ge 0$, luego existe $\beta := \inf_{v \in L^2(0,T)} J(v) \ge 0$. Por definición de ínfimo existe una **sucesión minimizante** $\{v_n\} \subset L^2(0,T)$ con $J(v_n) \to \beta$. El objetivo es extraer una subsucesión que converja en $L^2(0,T)$ hacia una solución $\hat{v}$ de (4.2), en cuatro pasos.

1. La sucesión está acotada. La sucesión numérica $J(v_n)$ converge, luego está acotada: $J(v_n) \le C$. De la definición de J, $\|v\|^2 \le \frac{2}{\mu} J(v)$ (se descarta el primer término, que es no negativo), luego $\|v_n\|^2 \le 2C/\mu$: acotada en $L^2(0,T)$. **Aquí se usa** $\mu > 0$: sin el término de coste del control, la sucesión podría no estar acotada.

2. Subsucesión débilmente convergente: la compacidad rebajada. Como $L^2(0,T)$ es de Hilbert, existe una subsucesión, que se sigue llamando $\{v_n\}$, que converge débilmente a un $\hat{v} \in L^2(0,T)$: $(v_n, v) \to (\hat{v}, v)$ para todo v.

3. Los estados convergen fuertemente: la ecuación de estado alisa (el núcleo de la prueba). Sean y_n e $\hat{y}$ los estados de v_n y $\hat{v}$. Por (4.3), para cada t,

$$y_n(t) - \hat{y}(t) = \int_0^t e^{a(t-s)} \big(v_n(s) - \hat{v}(s)\big) ds = \big(v_n - \hat{v},\ e^{a(t-\cdot)} \mathbf{1}_{[0,t]}\big),$$

que se lee así: la diferencia de los estados en el instante t es la integral de la diferencia de los controles contra la función $s \mapsto e^{a(t-s)}$ en $[0,t]$ (y cero en $(t,T]$, eso es la $\mathbf{1}_{[0,t]}$), es decir, el producto escalar en $L^2(0,T)$ de $v_n - \hat{v}$ con una función fija que no depende de n. La convergencia débil dice exactamente que esos productos escalares tienden a cero; luego $y_n(t) \to \hat{y}(t)$ para todo $t \in [0,T]$. Además, por Cauchy-Schwarz,

$$|y_n(t) - \hat{y}(t)| \le \Big(\int_0^t e^{2a(t-s)}ds\Big)^{1/2}\Big(\int_0^t |v_n - \hat{v}|^2 ds\Big)^{1/2} \le T^{1/2}e^{|a|T}\,\|v_n - \hat{v}\| \le C,$$

uniformemente en n y t. Convergencia puntual y acotación uniforme: por el teorema de la convergencia dominada (con $g \equiv C^2$ en $(0,T)$),

$$\|y_n - \hat{y}\|_{L^2(0,T)} \to 0, \tag{4.5}$$

y en consecuencia

$$\int_0^T |y_n - y_d|^2\,dt \to \int_0^T |\hat{y} - y_d|^2\,dt. \tag{4.6}$$

4. Los controles convergen fuertemente: se recupera la norma perdida. La idea: se escribe el cuadrado de la distancia $\|v_n - \hat{v}\|^2$ como una suma de cosas de las que ya se sabe adónde tienden, y se ve que el límite es ≤ 0. Para todo n, desarrollando el cuadrado,

$$\frac{\mu}{2}\|v_n - \hat{v}\|^2 = \frac{\mu}{2}\|v_n\|^2 - \mu(v_n,\hat{v}) + \frac{\mu}{2}\|\hat{v}\|^2 = \frac{\mu}{2}\|v_n\|^2 - \frac{\mu}{2}\|\hat{v}\|^2 + \mu\big(\|\hat{v}\|^2 - (v_n,\hat{v})\big),$$

y escribiendo $\frac{\mu}{2}\|v_n\|^2 = J(v_n) - \frac{\alpha}{2}\int |y_n - y_d|^2$ y lo mismo para $\hat{v}$,

$$\frac{\mu}{2}\|v_n - \hat{v}\|^2 = J(v_n) - J(\hat{v}) + \frac{\alpha}{2}\Big(\int_0^T |\hat{y} - y_d|^2 - \int_0^T |y_n - y_d|^2\Big) + \mu\big(\|\hat{v}\|^2 - (v_n,\hat{v})\big).$$

Tomando límite: $J(v_n) \to \beta$; el paréntesis con α tiende a 0 por (4.6); y el último término tiende a 0 por la convergencia débil. Queda $\limsup \frac{\mu}{2}\|v_n - \hat{v}\|^2 = \beta - J(\hat{v}) \le 0$, porque β es el ínfimo. Luego $\|v_n - \hat{v}\| \to 0$, y en particular $\int |v_n|^2 \to \int |\hat{v}|^2$ (4.7). De (4.6) y (4.7), $J(v_n) \to J(\hat{v})$, es decir, $J(\hat{v}) = \beta$: $\hat{v}$ **es un control óptimo**.

Unicidad, por convexidad estricta. Sean $\hat{v}$ y $\hat{w}$ dos controles óptimos con $\hat{v} \neq \hat{w}$, de modo que $J(\hat{v}) = J(\hat{w}) = \beta$. Sea $v = \frac{1}{2}(\hat{v} + \hat{w})$. Por (4.4),

$$J(v) < \tfrac{1}{2}J(\hat{v}) + \tfrac{1}{2}J(\hat{w}) = \beta,$$

absurdo, porque β es el ínfimo. ■

Dónde está el peso. En el paso 3, y merece decirse en una frase: *la ecuación de estado convierte convergencia débil en convergencia fuerte*. La razón es que el estado se obtiene del control por una integral, y una integral contra una función fija es exactamente lo que la convergencia débil sabe manejar; además la integración amortigua las oscilaciones rápidas. La convergencia débil también permite concentración, por ejemplo $\sqrt{n}\,\mathbf{1}_{(0,1/n)}$ para n grande: norma uno y límite débil cero. La cota integral del paso 3 trata ambos comportamientos. Sin ese paso no habría (4.6), sin (4.6) no habría paso 4, y sin el paso 4 solo se sabría que $\hat{v}$ es el límite débil de una sucesión minimizante, que no basta para concluir que sea óptimo.

Qué hace cada hipótesis. La hipótesis $\mu > 0$ se usa en el paso 1 —está señalada en el propio texto— y es la que da la *coercividad*: sin ella los controles podrían crecer sin límite mientras el coste baja, la sucesión minimizante no estaría acotada y el paso 2 no tendría de qué extraer subsucesión. Es la misma penalización que daba convexidad estricta al funcional. La *convexidad estricta* de J, que viene también de $\mu > 0$, es toda la unicidad: si J fuese solo convexa el mínimo podría alcanzarse en un segmento entero de controles, todos igual de buenos. Y que T sea *finito* entra en el paso 3, en la constante $T^{1/2}e^{|a|T}$: en horizonte infinito esa cota se pierde y hay que rehacer el argumento con pesos.

LA IDEA: EL MÉTODO DIRECTO

Es el argumento de Weierstrass (una función continua en un compacto alcanza su mínimo) en un espacio donde no hay compactos con interior. Se sustituye la compacidad por la débil, que sí tienen las bolas de un Hilbert, y la continuidad por algo más débil, que J no suba al pasar al límite débil (semicontinuidad inferior). Lo que hace que todo funcione es que J es una suma de una norma al cuadrado, que es débilmente semicontinua, y de un término que depende del estado, y el estado depende del control a través de una integral, que *mejora* la convergencia de débil a fuerte (paso 3). Esa mejora, «integrar compacta», es la misma que en Arzelà-Ascoli y la que el capítulo 1 usó con Bolzano-Weierstrass en $\mathbb{R}^n$: aquí la dimensión es infinita y hay que trabajar más.

EXPLÍCATE. *En el paso 4, ¿por qué es $\beta - J(\hat{v}) \leq 0$ y no $= 0$? ¿Qué se sabía de $J(\hat{v})$ antes de ese paso, y qué se sabe después?*

(B) Caracterización: el sistema de optimalidad

Teorema. En las condiciones anteriores, $\hat{v} \in L^2(0,T)$ es solución de (4.2), con $\hat{y}$ el estado asociado, si y solo si existe $\hat{\varphi}$ tal que la terna $(\hat{y}, \hat{\varphi}, \hat{v})$ satisface el **sistema de optimalidad**

$$\begin{cases} \hat{y}' = a\hat{y} + \hat{v}, & t \in (0,T), \\ \hat{y}(0) = y_0, & \\ -\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y} - y_d), & t \in (0,T), \\ \hat{\varphi}(T) = 0, & \\ \hat{v} = -\dfrac{1}{\mu}\hat{\varphi}. & \end{cases} \tag{4.8}$$

La función $\hat{\varphi}$ se llama **estado adjunto** asociado a $\hat{v}$.

LA IDEA: EL TEOREMA EN DIMENSIÓN FINITA, CON MATRICES

Si el estado fuera un vector $y = Y_0 + Mv \in \mathbb{R}^k$ que depende linealmente de un control $v \in \mathbb{R}^m$ a través de una matriz M, el coste sería $J(v) = \frac{\alpha}{2}|Y_0 + Mv - y_d|^2 + \frac{\mu}{2}|v|^2$, y su gradiente, $\nabla J(v) = \alpha M^t(y - y_d) + \mu v$. El mínimo cumple $\nabla J = 0$, es decir, $\mu\hat{v} = -\varphi$ con $\varphi := \alpha M^t(\hat{y} - y_d)$. Ese φ es el estado adjunto: la *traspuesta* de «cómo el control mueve al estado» aplicada a «cuánto se desvía el estado». En el problema

de verdad, M es el operador que a v le asocia la solución de la EDO con dato cero, $z_v(t) = \int_0^t e^{a(t-s)}v(s)\,ds$, y calcular su traspuesta es lo que hace la integración por partes del paso 2: la EDO adjunta, con el signo de la derivada cambiado y el dato en T, es «M^t» escrito como ecuación diferencial. Con esto en la cabeza, la demostración es la misma cuenta: derivada del coste igual a cero, y traspuesta.

Paso 1: perturbar y derivar — la parábola en ε. $\hat{v}$ es control óptimo si y solo si $J(\hat{v}) \le J(v)$ para todo $v \in L^2(0,T)$. Escribiendo $v = \hat{v} + \varepsilon w$ con $w \in L^2(0,T)$ y $\varepsilon \in \mathbb{R}$ (todo v se escribe así), equivale a $J(\hat{v}) \le J(\hat{v}+\varepsilon w)$ para todo w y todo ε. Para cada w se define $F_w(\varepsilon) := J(\hat{v}+\varepsilon w)$, y $\hat{v}$ es óptimo si y solo si

$$F_w(0) \le F_w(\varepsilon) \qquad \text{para todo } \varepsilon \in \mathbb{R} \text{ y todo } w \in L^2(0,T). \tag{4.9}$$

El estado asociado a $\hat{v}+\varepsilon w$ es $\hat{y} + \varepsilon z_w$, donde z_w es la solución de (4.1) con control w y dato inicial cero (linealidad de (4.3)). Por tanto

$$F_w(\varepsilon) = \frac{\alpha}{2}\int_0^T |\hat{y} + \varepsilon z_w - y_d|^2\,dt + \frac{\mu}{2}\int_0^T |\hat{v}+\varepsilon w|^2\,dt,$$

que es un **polinomio de segundo grado en** ε con coeficiente principal $\frac{\alpha}{2}\|z_w\|^2 + \frac{\mu}{2}\|w\|^2 \ge 0$: una parábola convexa. Una parábola convexa tiene su mínimo en 0 si y solo si su derivada en 0 es cero. Luego (4.9) equivale a

$$F_w'(0) = 0 \qquad \text{para todo } w \in L^2(0,T). \tag{4.10}$$

(Aquí está la diferencia con el capítulo 3: allí $F'(0)=0$ era solo necesaria; aquí, por ser F_w una parábola, es también suficiente.) Derivando,

$$F_w'(\varepsilon) = \alpha\int_0^T (\hat{y} - y_d + \varepsilon z_w) z_w\,dt + \mu\int_0^T (\hat{v}+\varepsilon w) w\,dt,$$

luego $\hat{v}$ es óptimo si y solo si

$$\alpha\int_0^T (\hat{y}-y_d) z_w\,dt + \mu\int_0^T \hat{v} w\,dt = 0 \qquad \text{para todo } w \in L^2(0,T). \tag{4.11}$$

En palabras: al mover el control en la dirección w, el coste cambia en dos partes, lo que cambia la desviación del estado (primera integral, donde z_w es cuánto se mueve el estado) y lo que cambia el gasto en control (segunda integral); en el óptimo, la suma es cero para toda dirección. El problema es que z_w depende de w de forma indirecta, a través de una EDO, y no se puede «sacar w factor común» todavía.

Paso 2: el estado adjunto, para sacar w factor común — el núcleo de la prueba. La primera integral de (4.11) contiene z_w, que depende de w a través de una EDO; hay que reescribirla como una integral *directamente contra* w. Para ello se introduce $\hat{\varphi}$, solución de

$$\begin{cases} -\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y}-y_d), & t \in (0,T), \\ \hat{\varphi}(T) = 0, \end{cases} \tag{4.12}$$

que es una EDO lineal con dato *final*: se resuelve hacia atrás desde T, y existe y es única por el mismo lema de la ecuación de estado lineal con el cambio $t \to T-t$.

Usando (4.12), después integrando por partes, y después usando la ecuación de z_w ($z_w' = a z_w + w$, $z_w(0) = 0$),

$$\alpha \int_0^T (\hat{y} - y_d) z_w \, dt = \int_0^T (-\hat{\varphi}' - a\hat{\varphi}) z_w \, dt = -\big[\hat{\varphi} z_w\big]_0^T + \int_0^T \hat{\varphi} z_w' \, dt - \int_0^T a\hat{\varphi} z_w \, dt$$
$$= -\hat{\varphi}(T) z_w(T) + \hat{\varphi}(0) z_w(0) + \int_0^T \hat{\varphi}\,(z_w' - a z_w)\, dt = \int_0^T \hat{\varphi}\, w \, dt,$$

donde los dos términos de frontera se anulan por $\hat{\varphi}(T) = 0$ y $z_w(0) = 0$: **ese es el motivo de que el adjunto lleve dato final y no inicial**. Volviendo a (4.11), $\hat{v}$ es óptimo si y solo si

$$\int_0^T (\hat{\varphi} + \mu\hat{v})\, w \, dt = 0 \qquad \text{para todo } w \in L^2(0,T),$$

y tomando $w = \hat{\varphi} + \mu\hat{v}$ se obtiene $\|\hat{\varphi} + \mu\hat{v}\|^2 = 0$, es decir, $\hat{\varphi} + \mu\hat{v} = 0$ casi en todas partes. Podemos elegir el representante $\hat{v} = -\frac{1}{\mu}\hat{\varphi}$, que es C^1 porque la fuente del adjunto es continua. (Es el lema fundamental en versión L^2, y aquí no hace falta invocarlo: como w recorre todo L^2, basta tomarla igual al propio integrando, que es lo que se acaba de hacer.) Como todos los pasos son equivalencias, queda demostrada la equivalencia entre (4.2) y (4.8). ■

El mecanismo, en tres movimientos, y por qué se repite en todo el capítulo. Conviene quedarse con el esqueleto, porque las demostraciones de las secciones siguientes —observación final, sistemas, segundo orden, no lineal, restricciones— son *esta misma* con adaptaciones, y quien vea el esqueleto no tiene que aprenderlas una a una.
Primer movimiento: la parábola en ε. Se empuja el control óptimo en una dirección w con intensidad ε y se mira el coste como función de ese único número. En el caso lineal y cuadrático sale una parábola, y una parábola con las ramas hacia arriba tiene su mínimo donde su derivada se anula. Aquí está la ventaja sobre el capítulo 3: allí «derivada nula» era necesaria y punto; aquí, por ser parábola, es necesaria *y* suficiente, y por eso el teorema es un «si y solo si» en vez de una condición.
Segundo movimiento: la integración por partes. La derivada nula deja una integral con dos sumandos, y uno de ellos contiene z_w, que es la respuesta del estado a la perturbación: una cantidad que depende de w a través de una ecuación diferencial y de la que no se puede despejar nada. Todo el trabajo consiste en convertir esa integral en otra escrita *directamente contra w*, y la herramienta es la integración por partes. Lo que aparece como multiplicador de w es el estado adjunto.
Tercer movimiento: el lema fundamental en L^2. Con w ya libre y sola, «la integral es cero para toda w» obliga al integrando a ser cero. En L^2 este paso es incluso más fácil que en el capítulo 3, porque se puede tomar w igual al propio integrando y leer que su norma al cuadrado es nula.

Qué mide el adjunto y por qué lleva dato final. Su fórmula explícita es

$$\hat{\varphi}(t) = \alpha \int_t^T e^{a(s-t)} \big(\hat{y}(s) - y_d(s)\big)\, ds.$$

Si en el instante t se aumenta el estado en una cantidad pequeña δ y se mantiene el control posterior, la variación del estado en $s \geq t$ es $e^{a(s-t)}\delta$. Al derivar el coste de

seguimiento, su variación de primer orden es precisamente $\hat{\varphi}(t)\delta$. Puede ser positiva o negativa: el empujón puede empeorar o mejorar el seguimiento. Esta sensibilidad se llama *precio sombra*. Depende de la desviación futura, por eso la integral recorre $[t, T]$; aquí se anula en T porque no hay coste final. La variación completa del coste al cambiar el control en una dirección w es $\int_0^T (\hat{\varphi} + \mu\hat{v})w$. La condición de mínimo equilibra así la variación del seguimiento y la del esfuerzo, sin confundirlas.

Qué hace cada hipótesis. El parámetro $\mu > 0$ aparece dos veces y es la hipótesis que asegura una penalización de al menos $\mu\|w\|^2/2$ en toda dirección $w \neq 0$, y es lo que permite despejar $\hat{v} = -\frac{1}{\mu}\hat{\varphi}$ al final. El parámetro $\alpha \geq 0$ solo pondera el término de seguimiento; con $\alpha = 0$ el sistema de optimalidad sigue siendo válido y da $\hat{\varphi} \equiv 0$, $\hat{v} \equiv 0$: si no importa llegar a ninguna parte, lo óptimo es no hacer nada. Y la *linealidad* de la ecuación de estado es la que hace que el estado de $\hat{v} + \varepsilon w$ sea exactamente $\hat{y} + \varepsilon z_w$, sin términos en ε^2; sin ella F_w deja de ser una parábola y la equivalencia se degrada a condición necesaria, que es justamente lo que ocurre en el problema no lineal de la sección 4.5.

LA IDEA: QUÉ ES EL ESTADO ADJUNTO

$\hat{\varphi}(t)$ mide cuánto cambia el coste si en el instante t se empuja un poco el estado: es el *precio sombra* del estado, en la jerga económica, y por eso se determina hacia atrás, desde el final, donde no queda tiempo para que un empujón influya ($\hat{\varphi}(T) = 0$). La condición $\hat{v} = -\hat{\varphi}/\mu$ dice que en cada instante el control óptimo equilibra el coste marginal del control, $\mu\hat{v}$, con el beneficio marginal de mover el estado, $-\hat{\varphi}$. La integración por partes es el truco que convierte «cómo depende el coste del control a través de la EDO» en una función explícita; en Cálculo de Variaciones ese truco daba la ecuación de Euler-Lagrange, y aquí da el sistema estado-adjunto, que es la misma ecuación de Euler-Lagrange escrita como sistema de primer orden en este modelo: al sustituir $v = y' - ay$, el integrando es $F(t, y, y') = \alpha(y - y_d)^2/2 + \mu(y' - ay)^2/2$ y $\partial_{y'} F = \mu v = -\varphi$. El extremo final libre da $\varphi(T) = 0$.

Observación. En la prueba, para la *suficiencia* (el sistema de optimalidad implica que $\hat{v}$ es óptimo) se ha usado de manera crucial que (4.1) es lineal y J cuadrático, de donde (4.9) y (4.10) son equivalentes. En los problemas no lineales de la sección 4.5 solo quedará la necesidad. El teorema de optimalidad del control básico, que dice que el control óptimo, el estado y el adjunto verifican (4.8), es el caso lineal y cuadrático del **Principio del Máximo de Pontryaguin**.

DE DÓNDE VIENE: PONTRYAGUIN Y EL OTRO CAMINO

A mediados de los años cincuenta, Lev Pontryaguin, Vladímir Boltyanskii y Revaz Gamkrelidze desarrollaron el principio del máximo: una condición de optimalidad a lo largo de la trayectoria. En la misma década, Richard Bellman desarrollaba la programación dinámica, que busca el coste óptimo como función del tiempo y del estado inicial. En su formulación continua aparece la ecuación de Hamilton-Jacobi-Bellman. Cuando esa función valor es suficientemente regular, su derivada respecto del estado, evaluada sobre la trayectoria óptima, coincide con el adjunto usado aquí: ambas descripciones expresan la sensibilidad del coste óptimo. El caso lineal y cuadrático tuvo un desarrollo fundamental en el artículo de Kalman de 1960, *Contributions to*

the Theory of Optimal Control.

El sistema de optimalidad resuelto a mano, en el caso más simple. Antes de las técnicas de cálculo conviene ver una vez el sistema (4.8) resuelto entero. Sean $a = 0$ (la población ni crece ni decrece sola) e $y_d = 0$ (se quiere eliminar una plaga). El sistema es

$$\hat{y}' = \hat{v}, \qquad \hat{y}(0) = y_0, \qquad -\hat{\varphi}' = \alpha\hat{y}, \qquad \hat{\varphi}(T) = 0, \qquad \hat{v} = -\frac{1}{\mu}\hat{\varphi}.$$

Derivando la primera ecuación y usando las otras, $\hat{y}'' = \hat{v}' = -\hat{\varphi}'/\mu = \frac{\alpha}{\mu}\hat{y}$. Con $k := \sqrt{\alpha/\mu}$, el par estado-adjunto se ha convertido en una sola EDO de segundo orden, $\hat{y}'' = k^2\hat{y}$, con una condición en cada extremo: $\hat{y}(0) = y_0$ y $\hat{y}'(T) = \hat{v}(T) = -\hat{\varphi}(T)/\mu = 0$. Es la ecuación de Euler-Lagrange de la caja «qué es el estado adjunto», con su condición natural en el extremo libre. Sus soluciones son las combinaciones de e^{kt} y e^{-kt} (EDO lineal de coeficientes constantes con raíces $\pm k$, teorema de las EDO lineales de coeficientes constantes del apéndice), y conviene escribirlas como $c_1 \cosh k(T-t) + c_2 \sinh k(T-t)$ para que la condición en T se lea sola: la derivada en T vale $-kc_2$, luego $c_2 = 0$, y $\hat{y}(0) = y_0$ fija c_1:

$$\hat{y}(t) = y_0 \frac{\cosh k(T-t)}{\cosh kT}, \qquad \hat{v}(t) = -y_0 k \frac{\sinh k(T-t)}{\cosh kT}, \qquad \hat{\varphi} = -\mu\hat{v}.$$

El coste óptimo sale sin integrar a mano. Con $\hat{v} = \hat{y}'$ y $\alpha = \mu k^2$ es $J(\hat{v}) = \frac{\mu}{2}\int_0^T (\hat{y}'^2 + k^2\hat{y}^2)\,dt$. Integrando por partes y usando $\hat{y}'' = k^2\hat{y}$, se obtiene $\int_0^T \hat{y}'^2 = [\hat{y}\hat{y}']_0^T - k^2\int_0^T \hat{y}^2$. Como $\hat{y}'(T) = 0$ y $\hat{y}'(0) = -y_0 k \tanh kT$,

$$J(\hat{v}) = \frac{\mu}{2}\left[\hat{y}\hat{y}'\right]_0^T = \frac{\mu}{2}\, y_0^2 k \tanh kT = \frac{1}{2}\sqrt{\alpha\mu}\; y_0^2 \tanh kT.$$

Qué se lee en estas fórmulas:

- *La forma.* El estado baja desde y_0 y llega a T con tangente horizontal. El control empuja con más fuerza al principio, $|\hat{v}(0)| = y_0 k \tanh kT$, y se apaga en T. Es el precio sombra en acción: cerca del final, mover el estado ya casi no cambia la desviación futura, y no compensa pagar por ello.
- *La escala.* $k = \sqrt{\alpha/\mu}$ es la inversa de un tiempo. Si el horizonte es largo (kT grande), lejos del final es $\hat{y} \approx y_0 e^{-kt}$, y $1/k$ es el tiempo en que el control óptimo reduce la plaga en un factor e. Duplicar el precio μ del control alarga ese tiempo en un factor $\sqrt{2}$, no 2.
- *Los límites.* Si $\mu \to 0$, $k \to \infty$ y $J \to 0$: con control gratis, la plaga se elimina casi al instante. Si $\mu \to \infty$, $\tanh kT \approx kT$ y $J \to \frac{\alpha}{2} y_0^2 T$, que es el coste de no hacer nada ($y \equiv y_0$).
- *Con números.* Para $\alpha = \mu = T = y_0 = 1$: $k = 1$, $\hat{y}(T) = 1/\cosh 1 \approx 0.648$, $\hat{v}(0) = -\tanh 1 \approx -0.762$ y $J(\hat{v}) \approx 0.381$, frente a $J(0) = 0.5$.

Y la imagen que conviene guardar: el estado se escribe hacia delante desde su dato inicial y el adjunto hacia atrás desde su dato final; aquí las dos condiciones han acabado en los dos extremos de una misma EDO de segundo orden, y el problema es un problema de contorno como los del capítulo 3.

(C) Cálculo del control óptimo

El sistema (4.8) proporciona un primer método de cálculo: partiendo de un control inicial v^0, se itera según el esquema

$$v^\ell \;\mapsto\; y^\ell \;\mapsto\; \varphi^\ell \;\mapsto\; v^{\ell+1},$$

con

$$\begin{cases} (y^\ell)' = ay^\ell + v^\ell, \\ y^\ell(0) = y_0, \end{cases} \qquad \begin{cases} -(\varphi^\ell)' = a\varphi^\ell + \alpha(y^\ell - y_d), \\ \varphi^\ell(T) = 0, \end{cases} \qquad v^{\ell+1} = -\frac{1}{\mu}\varphi^\ell.$$

Es un **método de aproximaciones sucesivas**: se considera la aplicación $\Lambda : L^2(0,T) \to L^2(0,T)$ que a cada v le asocia primero el estado y, después el adjunto φ y finalmente $\Lambda(v) := -\frac{1}{\mu}\varphi$; el algoritmo es $v^{\ell+1} = \Lambda(v^\ell)$, y el control óptimo es el punto fijo de Λ. Lo que hay que demostrar, y se demuestra entero a continuación, es que **si μ es suficientemente grande, Λ es contractiva** y la sucesión converge a $\hat{v}$ fuertemente en $L^2(0,T)$.

Por qué es contractiva para μ grande. Es el ejercicio 4.1 del curso en el caso escalar. La idea es que cada una de las tres flechas $v \mapsto y \mapsto \varphi \mapsto v$ multiplica la distancia entre dos entradas por, a lo sumo, una constante, y la última flecha divide además por μ: si μ supera al producto de las constantes, la vuelta completa acorta distancias.
Paso 1: la primera flecha, del control al estado. Sean v_1, v_2 dos controles y y_i, φ_i sus estados y adjuntos. Por linealidad, $y_1 - y_2$ es la solución con control $v_1 - v_2$ y dato cero, y por (4.3) y Cauchy-Schwarz, $|y_1(t) - y_2(t)| \le \sqrt{T}e^{|a|T}\|v_1 - v_2\|$ para todo t, luego $\|y_1 - y_2\| \le Te^{|a|T}\|v_1 - v_2\|$.
Paso 2: la segunda flecha, del estado al adjunto. Del mismo modo, $\varphi_1 - \varphi_2$ resuelve la ecuación adjunta con fuente $\alpha(y_1 - y_2)$ y dato final cero, y $\|\varphi_1 - \varphi_2\| \le \alpha Te^{|a|T}\|y_1 - y_2\|$.
Paso 3: la tercera flecha divide por μ, y ahí está todo (el núcleo). Por tanto

$$\|\Lambda(v_1) - \Lambda(v_2)\| = \frac{1}{\mu}\|\varphi_1 - \varphi_2\| \le \frac{\alpha T^2 e^{2|a|T}}{\mu}\,\|v_1 - v_2\|,$$

y Λ es contractiva si $\mu > \alpha T^2 e^{2|a|T}$. Entonces el teorema del punto fijo de Banach en el espacio completo $L^2(0,T)$ da la convergencia. La constante es pesimista: la práctica 27 muestra que con $a = -0.5$, $T = 10$, $\alpha = 1$ la iteración converge para $\mu = 20$ y diverge para $\mu = 1$. ■

Qué significa la condición, y qué se rompe al perderla. La desigualdad $\mu > \alpha T^2 e^{2|a|T}$ se lee sin cuentas: la vuelta completa amplifica por α (lo que pesa la desviación del estado), por T^2 (dos integraciones sobre un horizonte de longitud T) y por $e^{2|a|T}$ (lo que la propia dinámica amplifica), y contrae por μ (lo que cuesta el control). El bucle acorta distancias cuando el freno del control gana a las tres amplificaciones. Es una cota suficiente, no el factor exacto de la aplicación. Un horizonte largo o una dinámica que amplifica perturbaciones puede dificultar la contracción; aumentar μ disminuye directamente la sensibilidad de la última flecha.
Conviene subrayar qué es lo que se pierde cuando la condición falla, porque no es la solución. El sistema de optimalidad sigue teniendo solución única para *todo* $\mu > 0$:

eso lo garantizan los dos teoremas anteriores. Si la desigualdad suficiente falla, solo se pierde esta garantía: la iteración todavía puede converger, como muestra la figura. Si de hecho diverge, puede resolverse el sistema por otra vía, como el problema de contorno acoplado.

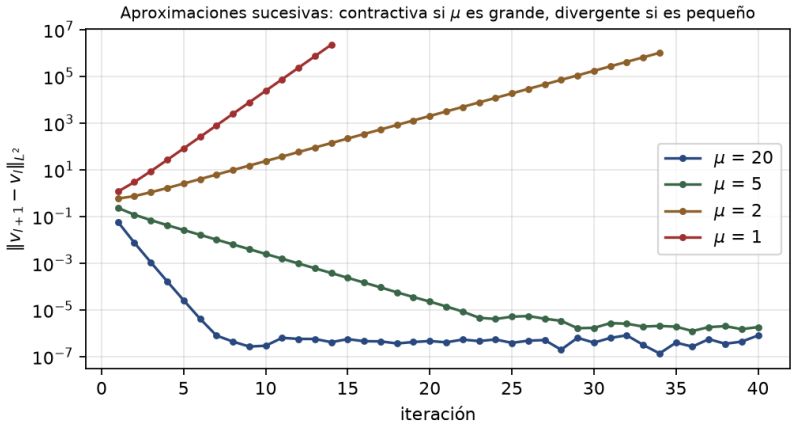


Figura 4.1: Aproximaciones sucesivas $v^\ell \mapsto y^\ell \mapsto \varphi^\ell \mapsto v^{\ell+1}$ para $a = -0.5$, $T = 10$, $y_0 = 3$, $\alpha = 1$, $v^0 = 0$ y $y_d(t) = 5e^{-t/2} - 2e^{-t}$. Con $\mu = 20$ y $\mu = 5$ la diferencia entre iterados disminuye inicialmente de forma geométrica y después alcanza el suelo de error numérico del cálculo. Con $\mu = 2$ y $\mu = 1$ crece. Las mesetas finales de las curvas convergentes reflejan también el error de resolver e interpolar las EDOs en cada paso.

Otro método proviene de que (4.8) da $\hat{v}$ explícitamente en función de $\hat{\varphi}$. Se sustituye en la EDO de $\hat{y}$ y se obtiene el **problema de contorno**

$$\begin{cases} \hat{y}' = a\hat{y} - \dfrac{1}{\mu}\hat{\varphi}, & t \in (0,T), \\ -\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y} - y_d), & t \in (0,T), \\ \hat{y}(0) = y_0, \quad \hat{\varphi}(T) = 0 : \end{cases}$$

un SDO de primer orden en $(\hat{y}, \hat{\varphi})$ con condiciones en $t = 0$ y en $t = T$. Se resuelve numéricamente con `bvp4c`, como en el capítulo 3, y después se toma $\hat{v} = -\frac{1}{\mu}\hat{\varphi}$. Es lo que hace la práctica 27. El sistema lineal de contorno tiene solución única para cualquier $\mu > 0$; no necesita la contractividad de la iteración anterior. El cálculo numérico sigue exigiendo comprobar residuos, tolerancias y condicionamiento, como en el capítulo 3.

EL CAMINO QUE NO SE TOMÓ: LOS ALGORITMOS DE DESCENSO

La alternativa. La tercera vía de (C): calcular el gradiente de J, que por (4.11) se identifica en L^2 con $\nabla J(v) = \varphi + \mu v$, y hacer $v^{\ell+1} = v^\ell - \rho\,(\varphi^\ell + \mu v^\ell)$ con un paso $\rho > 0$: el método del gradiente.
Su ventaja. En este problema converge para todo $\mu > 0$ con un paso suficientemente pequeño. No basta decir que J es convexa: también usamos el control de su gradiente. Una condición suficiente es $0 < \rho < 2\mu/L^2$, con $L = \mu + \alpha T^2 e^{2|a|T}$. En efecto, para $d = v_1 - v_2$ y $g = \nabla J(v_1) - \nabla J(v_2)$, la fórmula del adjunto da $(g,d) = \mu\|d\|^2 + \alpha\|z_d\|^2 \geq \mu\|d\|^2$ y $\|g\| \leq L\|d\|$.

Entonces

$$\|d - \rho g\|^2 \leq (1 - 2\rho\mu + \rho^2 L^2)\|d\|^2,$$

que prueba la contracción. La cota del paso es conservadora. El método se adapta también a estados no lineales y a EDP, pero esas extensiones necesitan su propio análisis y no heredan automáticamente la convergencia al mínimo global.

Por qué se presentan el punto fijo y el problema de contorno. Porque son los que se ven en el sistema de optimalidad sin nada más, y porque en el caso lineal escalar el problema de contorno es exacto y barato. El gradiente es el mismo cálculo (un estado hacia delante, un adjunto hacia atrás) con un paso distinto —son los llamados «algoritmos de descenso»—, y las prácticas 24 a 26 usan minimizadores genéricos de MATLAB que hacen algo parecido sin gradiente.

Qué se ha ganado en esta sección. La pregunta del gestor tenía tres partes, y las tres tienen respuesta. *Existe* una política óptima: el coste acota los controles, las bolas de L^2 son débilmente compactas y la ecuación de estado convierte la convergencia débil en fuerte. Es *única*, por la convexidad estricta, que la observación aporta si es convexa y el término $\frac{\mu}{2}\|v\|^2$ hace estricta. Y *se calcula sin probar todas las políticas*: el óptimo es la única solución del sistema de optimalidad (4.8), un estado que avanza desde $t = 0$ y un adjunto que retrocede desde $t = T$, acoplados por $\hat{v} = -\hat{\varphi}/\mu$. En el caso $a = 0$, $y_d = 0$ se resuelve con un coseno hiperbólico, y en general con el punto fijo (si μ es grande), con el problema de contorno o con el gradiente. Quedan abiertas cuatro variantes, que son las secciones siguientes: que solo importe el estado final (sección 4.2), que estado y control sean vectores (sección 4.3), que la ecuación sea de segundo orden o no lineal (secciones 4.4 y 4.5) y que el control tenga cotas (sección 4.6).

4.2 Problemas básicos (II): EDO lineal sin restricciones con observación final

LA PREGUNTA

Al gestor de la reserva ya no le importa la trayectoria durante el año: solo quiere que al final haya el menor número posible de individuos de una plaga, gastando lo menos posible. La observación pasa de ser una integral a ser un solo número. ¿Qué cambia en la condición de optimalidad, y qué no?

Se considera otro problema de control asociado al mismo problema de Cauchy (4.1):

$$\begin{cases} \text{Minimizar } I(v) := \dfrac{\alpha}{2}|y(T)|^2 + \dfrac{\mu}{2}\displaystyle\int_0^T |v|^2\,dt \\ \text{Sujeto a: } v \in L^2(0,T),\ (y,v) \text{ verifica (4.1)}, \end{cases} \tag{4.13}$$

con $\alpha, \mu > 0$. El cambio de modelo es solo el primer término: $\frac{\alpha}{2}|y(T)|^2$ dice que lo único que importa es cuántos individuos quedan al final, cuadrado para penalizar más los fracasos grandes, y no importa nada lo que pase por el camino. Se pierde con ello el daño intermedio (una plaga que arrasa en marzo y desaparece en junio tiene coste de observación cero) y a cambio el problema se resuelve a mano, como se verá. De nuevo hay existencia y unicidad

de control óptimo, con la prueba del teorema de existencia y unicidad del control básico (el paso 3 da ahora $y_n(T) \to \hat{y}(T)$ directamente por la convergencia débil, sin necesidad de la convergencia dominada).

La existencia con observación final, paso a paso, y qué usa de verdad. Conviene reescribir los cuatro pasos para (4.13), porque así se ve qué pieza depende de que la observación sea un cuadrado y cuál no. Sean $\beta := \inf I \geq 0$ y (v_n) una sucesión minimizante.

1. *Acotación.* El término de observación es ≥ 0, luego $\frac{\mu}{2}\|v_n\|^2 \leq I(v_n)$, que está acotado. Usa $\mu > 0$ y que la observación está acotada inferiormente.
2. *Compacidad débil.* Una subsucesión converge débilmente a un $\hat{v} \in L^2(0,T)$. No usa nada del coste.
3. *El estado final pasa al límite (el núcleo).* Por (4.3) en $t = T$, $y_n(T) = e^{aT}y_0 + (v_n, g)$ con $g(s) = e^{a(T-s)}$, una función fija. La convergencia débil dice exactamente que $(v_n, g) \to (\hat{v}, g)$, luego $y_n(T) \to \hat{y}(T)$, y por continuidad $\frac{\alpha}{2}|y_n(T)|^2 \to \frac{\alpha}{2}|\hat{y}(T)|^2$. Usa que la observación sea continua como función de $y(T)$.
4. *La norma no sube al pasar al límite.* $\|\hat{v}\| \leq \liminf \|v_n\|$ (semicontinuidad de la norma, en la caja de los tres hechos de Análisis). Sumando con el paso 3, $I(\hat{v}) \leq \liminf I(v_n) = \beta$, luego $I(\hat{v}) = \beta$. No usa nada de la observación.

Ninguno de los cuatro pasos usa la forma cuadrática. La misma prueba da un óptimo para cualquier término final $\Phi(y(T))$ con Φ continua y acotada inferiormente. La unicidad pide además que Φ sea convexa, por el argumento «de dónde sale la convexidad» de la sección anterior: el pozo doble de allí, con $\mu < \alpha T$, tiene dos óptimos, mientras que con $\frac{\alpha}{4}|y(T) - y_d|^4$ hay exactamente uno. Tampoco la caracterización necesita la parábola: basta que $F_w(\varepsilon)$ sea convexa, y lo es si Φ es convexa (convexa compuesta con afín, más una parábola). Si además Φ es derivable, «derivada nula en 0» sigue siendo necesaria y suficiente, y el adjunto conserva su ecuación: solo cambia su dato final, que pasa a ser $\Phi'(\hat{y}(T))$. El apéndice C (sección C.4) lo hace entero con la cuarta potencia, hasta reducir el problema a una ecuación de tercer grado.

La caracterización para (4.13) es:

Teorema. $\hat{v} \in L^2(0,T)$ es solución de (4.13), con $\hat{y}$ el estado asociado, si y solo si existe $\hat{\varphi}$ tal que $(\hat{y}, \hat{\varphi}, \hat{v})$ satisface

$$\begin{cases} \hat{y}' = a\hat{y} + \hat{v}, & t \in (0,T), \\ \hat{y}(0) = y_0, & \\ -\hat{\varphi}' = a\hat{\varphi}, & t \in (0,T), \\ \hat{\varphi}(T) = \alpha\hat{y}(T), & \\ \hat{v} = -\dfrac{1}{\mu}\hat{\varphi}. & \end{cases} \tag{4.14}$$

La idea: qué cambia y qué no. El mecanismo de tres movimientos es idéntico. Lo único distinto es *dónde vive el objetivo*: ya no se persigue una trayectoria a lo largo de todo el intervalo, sino un valor en el instante final. Y esa diferencia se traslada íntegra al adjunto. Al derivar el coste, la parte de seguimiento ya no aporta una integral sino un *único término*, $\alpha\hat{y}(T)z_w(T)$, evaluado en T. Como no hay integral que integrar por partes, la ecuación adjunta se queda *sin fuente*: es homogénea. Y como hay que fabricar el valor $\alpha\hat{y}(T)$ en el instante final, ese valor pasa a ser el dato final del adjunto en lugar del cero de antes. En la lectura de precios sombra la moraleja es clara: cuando lo único que importa es dónde se acaba, el precio del estado en T

es exactamente la desviación final multiplicada por α, y ese precio se propaga hacia atrás sin que nada lo alimente por el camino.

Es análoga a la del teorema de optimalidad del control básico.
Paso 1: perturbar y derivar — la parábola en ε. $\hat{v}$ es solución si y solo si $I(\hat{v}) \le I(\hat{v}+\varepsilon w)$ para todo $\varepsilon \in \mathbb{R}$ y $w \in L^2(0,T)$. Con $F_w(\varepsilon) := I(\hat{v}+\varepsilon w)$ y el estado $\hat{y}+\varepsilon z_w$,

$$F_w(\varepsilon) = \frac{\alpha}{2}|\hat{y}(T)+\varepsilon z_w(T)|^2 + \frac{\mu}{2}\int_0^T |\hat{v}+\varepsilon w|^2\,dt,$$

otra vez una parábola convexa en ε, luego $\hat{v}$ es óptimo si y solo si $F_w'(0)=0$ para todo w, es decir,

$$\alpha\hat{y}(T)z_w(T) + \mu\int_0^T \hat{v}w\,dt = 0 \qquad \text{para todo } w \in L^2(0,T). \tag{4.15}$$

Paso 2: el adjunto convierte el término final en una integral contra w (el núcleo). Sea $\hat{\varphi}$ la solución del **problema adjunto** $-\hat{\varphi}' = a\hat{\varphi}$, $\hat{\varphi}(T) = \alpha\hat{y}(T)$. Entonces, por el teorema fundamental del cálculo y las ecuaciones de $\hat{\varphi}$ y z_w,

$$\begin{aligned}\alpha\hat{y}(T)z_w(T) = \hat{\varphi}(T)z_w(T) = \big[\hat{\varphi}z_w\big]_0^T &= \int_0^T \frac{d}{dt}(\hat{\varphi}z_w)\,dt = \int_0^T (\hat{\varphi}'z_w + \hat{\varphi}z_w')\,dt \\ &= \int_0^T \big(-a\hat{\varphi}z_w + \hat{\varphi}(az_w+w)\big)dt = \int_0^T \hat{\varphi}w\,dt,\end{aligned}$$

usando $z_w(0)=0$. Aquí la integración por partes se ha usado al revés que en el teorema anterior: allí se partía de una integral y se sacaba un término de frontera; aquí se parte de un término de frontera y se fabrica una integral, escribiéndolo como $\int_0^T \frac{d}{dt}(\hat{\varphi}z_w)$. Es la misma identidad leída de derecha a izquierda.
Paso 3: el lema fundamental en L^2. Combinando con (4.15), $\int_0^T(\hat{\varphi}+\mu\hat{v})w\,dt = 0$ para todo w, equivalente a $\hat{v} = -\frac{1}{\mu}\hat{\varphi}$. ■

Qué hace cada hipótesis. La única diferencia con el teorema de optimalidad del control básico está en el funcional, y se ve dónde repercute: al no haber término integral de seguimiento, la ecuación adjunta pierde la fuente $\alpha(\hat{y}-y_d)$ y se vuelve homogénea, de modo que $\hat{\varphi}(t) = \alpha\hat{y}(T)e^{a(T-t)}$ se puede escribir en cerrado. Y $\mu > 0$ sigue haciendo lo mismo que antes: garantiza que la parábola tiene ramas hacia arriba y permite despejar el control. Si α se anulara, el adjunto sería idénticamente nulo y el control óptimo también, que es lo esperable cuando no se persigue nada.

La idea: la observación se traslada al dato final del adjunto

Lo que no cambia: la ecuación de estado, la relación $\hat{v} = -\hat{\varphi}/\mu$, y el mecanismo de la integración por partes. Lo que cambia: la ecuación adjunta ya no tiene fuente (no hay nada que observar durante el intervalo) y en cambio tiene un dato final dado por el gradiente de la observación final. Regla general que la sección 4.7 formaliza: *lo que se observa dentro del intervalo aparece como fuente del adjunto; lo que se observa al final aparece como su dato final*. Aquí el adjunto se resuelve a mano, $\hat{\varphi}(t) = \alpha\hat{y}(T)e^{a(T-t)}$, y el control

óptimo es $\hat{v}(t) = -\frac{\alpha}{\mu}\hat{y}(T)e^{a(T-t)}$: una exponencial cuya amplitud depende de $\hat{y}(T)$, que a su vez depende de $\hat{v}$; ese acoplamiento se cierra sustituyendo en (4.3), y sale una ecuación escalar para $\hat{y}(T)$. El ejercicio del control con observación final lo hace.

El sistema de optimalidad permite calcular el control. En este caso, el ejercicio siguiente lo resuelve mediante una fórmula explícita que también sirve para comprobar un cálculo numérico.

Ejercicio 4.1. *El control con observación final, resuelto a mano*

ORIGEN: ejercicio propuesto en el curso al final de la sección 4.2 («deducir métodos numéricos»); aquí se desarrolla una solución explícita.

Para el problema (4.13), calcular explícitamente $\hat{v}$, $\hat{y}$ e $\hat{y}(T)$ en función de los datos, y comprobar que $\hat{y}(T) \to 0$ cuando $\mu \to 0$.

PISTAS, DE MENOR A MAYOR

1. *Resuelve el adjunto hacia atrás: es $\varphi' = -a\varphi$ con $\varphi(T)$ dado.*
2. *Sustituye $\hat{v} = -\hat{\varphi}/\mu$ en (4.3) y evalúa en $t = T$: queda una ecuación lineal en la incógnita $\hat{y}(T)$.*
3. *La integral $\int_0^T e^{2a(T-s)}\,ds$ se calcula, con cuidado en el caso $a = 0$.*

Resolución.

La idea, antes de escribir nada. El sistema de optimalidad tiene mala fama porque acopla dos ecuaciones que van en sentidos contrarios, y parece que hay que resolverlas a la vez. Aquí no hace falta, y la razón es que *el adjunto no tiene fuente*: su ecuación es $-\hat{\varphi}' = a\hat{\varphi}$, una exponencial, y lo único que no se conoce de ella es el dato con el que arranca en T, que es $\alpha\hat{y}(T)$. Llámese $Y = \hat{y}(T)$ al estado final desconocido. Todo lo demás —el adjunto, el control, el estado— queda escrito en función de Y, y al evaluar el estado en T se obtiene una ecuación que dice $Y = (\text{algo conocido}) - (\text{constante})\,Y$. **Un problema de contorno acoplado se ha convertido en una ecuación de primer grado con una incógnita**, y esa es toda la gracia del ejercicio. La imagen que lo predice: con observación final el precio sombra solo se paga al llegar, así que no hay nada que lo alimente por el camino y su forma está fijada de antemano; solo falta su escala.

Paso 1: el adjunto, que es explícito porque no tiene fuente. $-\hat{\varphi}' = a\hat{\varphi}$ con $\hat{\varphi}(T) = \alpha\hat{y}(T) =: \alpha Y$ da $\hat{\varphi}(t) = \alpha Y e^{a(T-t)}$, luego $\hat{v}(t) = -\frac{\alpha Y}{\mu}e^{a(T-t)}$. Obsérvese que la *forma* del control ya está decidida —una exponencial— y lo único que falta es su amplitud.

Paso 2 (el núcleo): cerrar el acoplamiento evaluando en T. Por (4.3) en $t = T$,

$$Y = \hat{y}(T) = e^{aT}y_0 + \int_0^T e^{a(T-s)}\hat{v}(s)\,ds = e^{aT}y_0 - \frac{\alpha Y}{\mu}\int_0^T e^{2a(T-s)}\,ds = e^{aT}y_0 - \frac{\alpha Y}{\mu}K,$$

$$K := \int_0^T e^{2a(T-s)}ds = \begin{cases} \dfrac{e^{2aT}-1}{2a} & a \neq 0, \\ T & a = 0. \end{cases}$$

Despejando, $Y(1 + \alpha K/\mu) = e^{aT}y_0$, es decir,

$$\hat{y}(T) = \frac{e^{aT}y_0}{1+\alpha K/\mu} = \frac{\mu\, e^{aT}y_0}{\mu + \alpha K}, \qquad \hat{v}(t) = -\frac{\alpha\, e^{aT}y_0}{\mu+\alpha K}\,e^{a(T-t)}.$$

También falta dar el estado durante el intervalo. Si $K_s := \int_0^s e^{2a(s-r)}\,dr$, de modo que $K = K_T$, al sustituir el control en (4.3) resulta

$$\hat{y}(t) = e^{at} y_0 \, \frac{\mu + \alpha K_{T-t}}{\mu + \alpha K_T}.$$

Para simplificar se usa $K_T = e^{2a(T-t)} K_t + K_{T-t}$, que separa la integral en $[0,t]$ y $[t,T]$. La fórmula da $\hat{y}(0) = y_0$ y el valor final calculado; si $a = 0$, se reduce a $\hat{y}(t) = y_0(\mu + \alpha(T-t))/(\mu + \alpha T)$.

Paso 3: los límites de control barato y caro. Con $\mu \to 0$ es $\hat{y}(T) \to 0$: cuando el control es gratis, se lleva el estado final exactamente a cero, y el control tiende a $-\frac{e^{aT}y_0}{K} e^{a(T-t)}$, que es el control de *norma mínima* que anula $y(T)$ (control exacto a cero). Para verlo, sea $k(t) = e^{a(T-t)}$. Cualquier control que anule $y(T)$ satisface $(k, v) = -e^{aT} y_0$. Cauchy-Schwarz da $\|v\|^2 \geq e^{2aT} y_0^2 / \|k\|^2 = e^{2aT} y_0^2 / K$; la igualdad se alcanza precisamente cuando v es el múltiplo de k que acaba de obtenerse. Cuando $\mu \to \infty$, $\hat{v} \to 0$ e $\hat{y}(T) \to e^{aT} y_0$: no se controla. Un método numérico es innecesario aquí; en el caso de un sistema (sección 4.3) la misma cuenta lleva a un sistema lineal $n \times n$ para $\hat{y}(T)$, la *matriz de gramiano de controlabilidad*.

Qué hace cada hipótesis. Con $\mu = 0$ ya no podemos dividir por μ en la condición de optimalidad. El problema sigue teniendo mínimo: como $K > 0$, se puede conseguir $Y = 0$, y todos los controles que lo consigan son óptimos. Por ejemplo, se puede añadir al control anterior cualquier $w \in L^2$ con $(k, w) = 0$. Se pierde la unicidad; el límite $\mu \to 0^+$ selecciona entre ellos el de menor norma. Si $\alpha = 0$ y $\mu > 0$, el mínimo es $v = 0$. Finalmente, $T > 0$ hace que $K > 0$ y que haya tiempo para modificar el estado; para $\mu > 0$ el denominador $\mu + \alpha K$ es positivo en cualquier caso.

Con números, y qué mide K. Sean $a = 0$ y $T = y_0 = \alpha = \mu = 1$. Entonces $K = 1$, $\hat{y}(T) = \frac{1}{2}$, $\hat{v} \equiv -\frac{1}{2}$, $\hat{y}(t) = 1 - \frac{t}{2}$, e $I(\hat{v}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$. Compárese con los dos extremos: no hacer nada deja $y(T) = 1$ y cuesta $\frac{1}{2}$, todo en desviación; llevar el estado exactamente a cero con el control más barato que lo consigue, $v \equiv -1$, cuesta también $\frac{1}{2}$, todo en control. El óptimo reparte: llega a la mitad y paga la cuarta parte. La fórmula $\hat{y}(T) = e^{aT} y_0 \cdot \frac{\mu}{\mu + \alpha K}$ se lee así: el estado final sin control, multiplicado por un factor entre 0 y 1 que compara el precio del control, μ, con αK. Y K tiene significado propio: es $\|g\|^2$ para $g(s) = e^{a(T-s)}$, y como $y_v(T) - e^{aT} y_0 = (v, g)$, Cauchy-Schwarz dice que un control de norma 1 mueve el estado final como mucho $\sqrt{K}$. K mide el *alcance* del control. Cuando la dinámica amplifica ($a > 0$), un empujón temprano crece hasta T y el mismo esfuerzo llega más lejos: con $T = y_0 = \alpha = \mu = 1$, la fracción del estado final sin control que queda es $\frac{\mu}{\mu + \alpha K} \approx 0.70$ para $a = -1$, 0.5 para $a = 0$ y 0.24 para $a = 1$.

Dónde se falla.

- *Copiando el dato final de (4.8).* Con $\hat{\varphi}(T) = 0$ y la ecuación adjunta homogénea sale $\hat{\varphi} \equiv 0$ y $\hat{v} \equiv 0$, que cumple todo lo que se ha escrito y no es el óptimo. Con observación final el dato del adjunto es $\alpha \hat{y}(T)$, y solo se anula si $\hat{y}(T) = 0$.
- *Cambiando el signo de $\hat{v} = -\hat{\varphi}/\mu$.* Sale $Y(1 - \alpha K/\mu) = e^{aT} y_0$. Con $a = 0$ y $T = y_0 = \alpha = 1$, para $\mu = 2$ da $Y = 2$ (el «control» aleja el estado y dobla el valor final) y para $\mu = 1$ no hay solución. La comprobación que lo detecta: $\hat{y}(T)$ tiene que tener el signo de $e^{aT} y_0$ y ser menor en valor absoluto.
- *En el caso $a = 0$, en el ordenador.* La expresión `(exp(2*a*T)-1)/(2*a)` con `a=0` es $0/0$, que MATLAB devuelve como `NaN`; y con a muy pequeño pierde cifras por cancelación. Se

separa el caso ($K = T$) o se usa `expm1(2*a*T)/(2*a)`, que calcula $e^x - 1$ sin cancelar.
- *Con una sola exponencial en K.* La exponencial aparece dos veces, una en el control y otra en la fórmula (4.3), y por eso $K = \int_0^T e^{2a(T-s)}ds$. Quien escriba $\int_0^T e^{a(T-s)}ds$ obtiene, con $a = T = 1$, $e - 1 \approx 1.72$ en vez de $\frac{e^2-1}{2} \approx 3.19$.

QUÉ DEJA. Con observación final el adjunto es explícito, y el problema entero se reduce a una ecuación escalar para $\hat{y}(T)$. Y los dos límites, $\mu \to 0$ y $\mu \to \infty$, son la comprobación de que el modelo hace lo que promete: controlar del todo o no controlar.

4.3 Sistema de EDOs lineales sin restricciones

LA PREGUNTA

Un avión tiene decenas de variables de estado y solo unos pocos mandos. Un ecosistema tiene muchas especies y se actúa sobre una. ¿Se puede gobernar un sistema de n variables con $m < n$ controles, y qué forma toma el sistema de optimalidad cuando el estado y el control son vectores?

Se considera el problema de Cauchy no escalar

$$\begin{cases} y' = Ay + Bv, & t \in (0,T), \\ y(0) = y_0, \end{cases} \tag{4.16}$$

con $A \in \mathcal{L}(\mathbb{R}^n;\mathbb{R}^n)$, $B \in \mathcal{L}(\mathbb{R}^m;\mathbb{R}^n)$ e $y_0 \in \mathbb{R}^n$. El control $v = (v_1,\dots,v_m)^t$ pertenece a $L^2(0,T;\mathbb{R}^m)$ y el estado es $y = (y_1,\dots,y_n)^t$. Los casos interesantes aparecen cuando n es grande y m pequeño: se controlan «muchas componentes» con «pocos controles escalares»; la matriz B dice sobre qué componentes actúa cada control.

Definición. Sea $v \in L^2(0,T;\mathbb{R}^m)$. Se llama solución maximal de (4.16) a toda función $y : [0,T] \to \mathbb{R}^n$ absolutamente continua componente a componente que verifica $y'(t) = Ay(t) + Bv(t)$ para casi todo $t \in (0,T)$ e $y(0) = y_0$.

Para cada $v \in L^2(0,T;\mathbb{R}^m)$ existe una única solución maximal, dada por

$$y(t) = e^{At}y_0 + \int_0^t e^{A(t-s)}Bv(s)\,ds \qquad \text{para todo } t \in [0,T], \tag{4.17}$$

con la exponencial de matriz del apéndice (teorema sobre la exponencial de una matriz). El argumento es el del lema de la ecuación de estado lineal: la fórmula verifica las propiedades, y si y, z son dos soluciones, $y - z$ resuelve un SDO lineal homogéneo con dato nulo y, por Gronwall (aplicado a $|y - z|$ con la norma de A), $y = z$.

El problema de control es

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{\alpha}{2}\displaystyle\int_0^T |Cy - z_d|^2_{\mathbb{R}^k}\,dt + \frac{\mu}{2}\int_0^T |v|^2_{\mathbb{R}^m}\,dt \\ \text{Sujeto a: } v \in L^2(0,T;\mathbb{R}^m),\ (y,v) \text{ verifica (4.16)}, \end{cases} \tag{4.18}$$

con α, μ, T como antes, $C \in \mathcal{L}(\mathbb{R}^n;\mathbb{R}^k)$, $k \geq 1$, y $z_d \in \mathbb{R}^k$ (o una función continua de t con valores en $\mathbb{R}^k$). La matriz C es el **operador de observación**: dice qué combinaciones del estado

se miden y se comparan con lo deseado; con $C = I$ se observa todo el estado. Se usa $(\cdot,\cdot)_{\mathbb{R}^\ell}$ y $|\cdot|_{\mathbb{R}^\ell}$ para el producto escalar y la norma euclídeos.

LA IDEA: UN CASO CONCRETO PARA TENER DELANTE

Con $n = 2$, $m = 1$ y $k = 2$: dos variables acopladas, un solo control que entra en la primera ecuación, y se observan las dos. Es la práctica 28, con

$$A = \begin{pmatrix} -0.05 & 0.03 \\ -0.11 & 0.115 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad C = I,$$

de modo que $Bv = (v, 0)^t$ (el control solo empuja a y_1) y $B^t\varphi = \varphi_1$ (el control óptimo solo mira la primera componente del adjunto). Todo lo que sigue se puede leer con este ejemplo: A cambia por A^t en la adjunta, y el control óptimo es $\hat{v} = -\hat{\varphi}_1/\mu$. Respecto del caso escalar, lo único nuevo es que «multiplicar por a» se convierte en «multiplicar por una matriz», y al pasar una matriz al otro lado de un producto escalar aparece su traspuesta.

(A) Existencia y unicidad

Teorema. El problema (4.18) posee exactamente una solución $\hat{v} \in L^2(0, T; \mathbb{R}^m)$.

La demostración es totalmente análoga a la del teorema de existencia y unicidad del control básico y usa de manera fundamental la linealidad de (4.16), que J es cuadrático y que $\mu > 0$. Los únicos cambios respecto de aquella son que en el paso 3 la cota es $|y_n(t) - \hat{y}(t)| \le \sqrt{T}\,\|e^{A\cdot}\|_\infty \|B\|\,\|v_n - \hat{v}\|$, con la norma de operador. Multiplicar por C conserva la convergencia fuerte de los estados, pues $\|C(y_n - \hat{y})\|_{L^2} \le \|C\|\,\|y_n - \hat{y}\|_{L^2}$. La convexidad estricta sigue viniendo del término $\frac{\mu}{2}\|v\|^2$, ahora con la norma de $L^2(0, T; \mathbb{R}^m)$ en vez de la escalar.

(B) Sistema de optimalidad

Teorema. En las condiciones anteriores, $\hat{v} \in L^2(0, T; \mathbb{R}^m)$ es solución de (4.18), con $\hat{y}$ el estado asociado, si y solo si existe $\hat{\varphi}$ tal que $(\hat{y}, \hat{\varphi}, \hat{v})$ satisface

$$\begin{cases} \hat{y}' = A\hat{y} + B\hat{v}, & t \in (0, T), \\ \hat{y}(0) = y_0, & \\ -\hat{\varphi}' = A^t\hat{\varphi} + \alpha C^t(C\hat{y} - z_d), & t \in (0, T), \\ \hat{\varphi}(T) = 0, & \\ \hat{v} = -\dfrac{1}{\mu} B^t\hat{\varphi}. & \end{cases} \tag{4.19}$$

La idea: la traspuesta hace todo el trabajo. En el caso escalar el adjunto cambiaba el signo de la derivada y poco más. En el vectorial se ve por qué: lo que la integración por partes hace es *trasponer* el operador que lleva el control al estado, y trasponer, en dimensión finita, es exactamente lo que la matriz A^t significa. Por eso la ecuación adjunta lleva A^t y no A, y por eso el control óptimo sale con B^t: cada flecha del problema directo aparece dada la vuelta en el adjunto. La imagen es la de un sistema de tuberías recorrido al revés: en el problema

directo la información va del control a los estados que toca, y en el adjunto el precio vuelve desde los estados observados hasta el control que puede moverlos.

Hay además un detalle técnico que conviene anunciar, porque es el que hace la cuenta legible. En vez de integrar por partes se deriva directamente el producto escalar $(z_w, \hat{\varphi})$. Es lo mismo —integrar por partes *es* integrar la derivada de un producto— pero con las dos ecuaciones a mano se ve al instante lo que va a pasar: los términos con A salen uno como $(Az_w, \hat{\varphi})$ y otro como $(z_w, A^t\hat{\varphi})$, que son el mismo número, y se cancelan. La ecuación adjunta está construida precisamente para que se cancelen.

Es la del teorema de optimalidad del control básico con las adaptaciones vectoriales. **Paso 1: perturbar y derivar — la parábola en ε.** $\hat{v}$ es solución si y solo si $J(\hat{v}) \leq J(\hat{v} + \varepsilon w)$ para todo $\varepsilon \in \mathbb{R}$ y $w \in L^2(0,T;\mathbb{R}^m)$. Con $F_w(\varepsilon) := J(\hat{v} + \varepsilon w)$ y el estado $\hat{y} + \varepsilon z_w$, donde z_w resuelve

$$\begin{cases} z' = Az + Bw, & t \in (0,T), \\ z(0) = 0, \end{cases} \tag{4.20}$$

se tiene

$$F_w(\varepsilon) = \frac{\alpha}{2}\int_0^T |C(\hat{y} + \varepsilon z_w) - z_d|^2_{\mathbb{R}^k}\, dt + \frac{\mu}{2}\int_0^T |\hat{v} + \varepsilon w|^2_{\mathbb{R}^m}\, dt,$$

parábola convexa en ε, y $\hat{v}$ es óptimo si y solo si $F_w'(0) = 0$ para todo w:

$$\alpha\int_0^T (C\hat{y} - z_d, Cz_w)_{\mathbb{R}^k}\, dt + \mu\int_0^T (\hat{v}, w)_{\mathbb{R}^m}\, dt = 0,$$

es decir, pasando C al otro lado del producto escalar con su traspuesta,

$$\alpha\int_0^T \big(C^t(C\hat{y} - z_d), z_w\big)_{\mathbb{R}^n}\, dt + \mu\int_0^T (\hat{v}, w)_{\mathbb{R}^m}\, dt = 0 \qquad \text{para todo } w. \tag{4.21}$$

Paso 2: el adjunto, y la cancelación de los términos con A (el núcleo). Sea $\hat{\varphi}$ el estado adjunto, solución de

$$\begin{cases} -\hat{\varphi}' = A^t\hat{\varphi} + \alpha C^t(C\hat{y} - z_d), & t \in (0,T), \\ \hat{\varphi}(T) = 0. \end{cases} \tag{4.22}$$

Ahora hay que convertir la integral con z_w en una integral con w. En vez de integrar por partes se hace algo equivalente y más limpio: se deriva el producto escalar $(z_w, \hat{\varphi})$, se sustituyen las dos EDO, y se observa que los términos con A se cancelan entre sí porque $(Az, \varphi) = (z, A^t\varphi)$. Usando los SDOs de z_w y $\hat{\varphi}$, la derivada del producto escalar es

$$\begin{aligned} \frac{d}{dt}(z_w, \hat{\varphi})_{\mathbb{R}^n} &= (z_w', \hat{\varphi}) + (z_w, \hat{\varphi}') = (Az_w + Bw, \hat{\varphi}) + \big(z_w, -A^t\hat{\varphi} - \alpha C^t(C\hat{y} - z_d)\big) \\ &= (w, B^t\hat{\varphi})_{\mathbb{R}^m} - \alpha\big(z_w, C^t(C\hat{y} - z_d)\big)_{\mathbb{R}^n}, \end{aligned}$$

porque $(Az_w, \hat{\varphi}) = (z_w, A^t\hat{\varphi})$ se cancela con $-(z_w, A^t\hat{\varphi})$, y $(Bw, \hat{\varphi}) = (w, B^t\hat{\varphi})$. Integrando en $[0,T]$ y usando $z_w(0) = \hat{\varphi}(T) = 0$, el lado izquierdo integra a cero, y

$$\alpha\int_0^T \big(C^t(C\hat{y} - z_d), z_w\big)_{\mathbb{R}^n}\, dt = \int_0^T \big(B^t\hat{\varphi}, w\big)_{\mathbb{R}^m}\, dt.$$

Paso 3: el lema fundamental en L^2. Por tanto $\hat{v}$ es solución si y solo si $\int_0^T (B^t\hat{\varphi} + \mu\hat{v}, w)_{\mathbb{R}^m}\, dt = 0$ para todo w, es decir, si y solo si $B^t\hat{\varphi} + \mu\hat{v} = 0$ casi en todas partes. ■

Qué hace cada matriz. Las tres matrices entran en sitios distintos y cada una deja su traspuesta en un lugar distinto del sistema de optimalidad. La matriz A es la dinámica, y aparece traspuesta en la ecuación adjunta: el precio se propaga hacia atrás por las mismas conexiones que el estado usa hacia adelante, pero en sentido contrario. La matriz B dice qué componentes del estado puede tocar el control, y aparece traspuesta en la fórmula $\hat{v} = -\frac{1}{\mu}B^t\hat{\varphi}$: el control solo reacciona al precio de aquello que puede mover. Si la columna j de B es nula, el control v_j no actúa sobre el estado y la fila j de B^t da $\hat{v}_j = 0$: gastar en ese control no aporta nada. La matriz C dice qué se observa, y aparece traspuesta en la fuente del adjunto: solo lo que se mide genera precio. Cuando B y C son la identidad y $A = a$ escalar se recupera exactamente el teorema de optimalidad del control básico. Y $\mu > 0$ hace, una vez más, lo mismo: da la convexidad estricta de la parábola y permite despejar el control.

LA IDEA: LAS TRASPUESTAS

Todo lo que en el caso escalar era «multiplicar por a» se convierte en «multiplicar por A» en la ecuación de estado y por A^t en la adjunta: la adjunta es, literalmente, la ecuación del operador adjunto, de ahí su nombre. B^t aparece en el control óptimo: el control solo «ve» la componente del adjunto en las direcciones sobre las que actúa. Y C^t aparece en la fuente del adjunto: solo lo que se observa genera precio sombra. Con $m < n$, $B^t\hat{\varphi}$ es un vector de m componentes, y por eso se puede gobernar un sistema de n variables buscando la mejor intervención con los m mandos disponibles. Esto no asegura poder alcanzar cualquier destino: si $B = 0$, el óptimo es $v = 0$ y no se puede modificar el estado. Que además se pueda llevar el estado a cualquier sitio es la *controlabilidad*, que depende de A y B (criterio de Kalman) y que queda fuera de este curso.

QUÉ DEJA. El mecanismo es siempre el mismo: parábola en ε, derivada nula, adjunto definido para que la integración por partes convierta z_w en w, y control obtenido aplicando B^t al adjunto y multiplicando por $-1/\mu$. Lo que se observa por dentro va a la fuente del adjunto; lo que se observa al final va a su dato final.

(C) Cálculo

Con ayuda de (4.19) se tiene el método de aproximaciones sucesivas $v^\ell \mapsto y^\ell \mapsto \varphi^\ell \mapsto v^{\ell+1}$, con

$$\begin{cases} (y^\ell)' = Ay^\ell + Bv^\ell, \\ y^\ell(0) = y_0, \end{cases} \qquad \begin{cases} -(\varphi^\ell)' = A^t\varphi^\ell + \alpha C^t(Cy^\ell - z_d), \\ \varphi^\ell(T) = 0, \end{cases} \qquad v^{\ell+1} = -\frac{1}{\mu}B^t\varphi^\ell.$$

(Si se observa todo el estado, $C = I$, la fuente del adjunto es $\alpha(y^\ell - y_d)$; con observación C es $\alpha C^t(Cy^\ell - z_d)$, como en (4.19).) El ejercicio 4.1 del curso pide probar que, si μ es suficientemente grande, la aplicación $v^\ell \mapsto v^{\ell+1}$ es contractiva y los v^ℓ convergen en $L^2(0,T;\mathbb{R}^m)$ hacia $\hat{v}$: es la cuenta de la sección 4.1. Si $E = \max_{0\le t\le T}\|e^{At}\|$, las dos integraciones dan

$\|y_1 - y_2\| \le TE\|B\|\,\|v_1 - v_2\|$ y $\|\varphi_1 - \varphi_2\| \le \alpha TE\|C\|^2\|y_1 - y_2\|$. Por tanto,

$$\|\Lambda(v_1) - \Lambda(v_2)\| \le \frac{\alpha T^2 E^2 \|B\|^2 \|C\|^2}{\mu}\|v_1 - v_2\|.$$

Basta que el coeficiente de la derecha sea menor que uno. También, como el control está dado explícitamente en (4.19), se puede sustituir y resolver un problema de contorno para un SDO de $2n$ ecuaciones, las n de $\hat{y}$ y las n de $\hat{\varphi}$, y tomar después $\hat{v} = -\frac{1}{\mu}B^t\hat{\varphi}$: es la práctica 28.

Ejercicio 4.2. *Observación final en un sistema*

ORIGEN: ejercicio 4.2 del curso.

Se considera

$$\begin{cases} \text{Minimizar } I(v) := \dfrac{\alpha}{2}|Cy(T) - z_T|^2_{\mathbb{R}^k} + \dfrac{\mu}{2}\displaystyle\int_0^T |v|^2_{\mathbb{R}^m}\,dt \\ \text{Sujeto a: } v \in L^2(0,T;\mathbb{R}^m),\ (y,v) \text{ verifica (4.16)}, \end{cases}$$

con $\alpha, \mu > 0$, $C \in \mathcal{L}(\mathbb{R}^n;\mathbb{R}^k)$ y $z_T \in \mathbb{R}^k$. Demostrar la existencia y unicidad de control óptimo, obtener el sistema de optimalidad y dar un método numérico con convergencia para μ grande.

Resolución.

La idea. Es el ejercicio anterior con matrices, y la moraleja de fondo es que *no hace falta ninguna idea nueva*: los mismos tres movimientos valen palabra por palabra, y lo único que hay que vigilar es dónde aparece cada traspuesta. La razón de que valgan es que el mapa control $\to$ estado final sigue siendo **afín y continuo** —ahora la fórmula de variación de las constantes lleva $e^{A(T-s)}B$ en vez de $e^{a(T-s)}$—, y ese es exactamente el ingrediente que la convergencia débil respeta. Dicho con la imagen del capítulo: cada componente de $y(T)$ es un producto escalar del control contra una función fija; el producto escalar contra algo fijo es lo único que sobrevive al límite débil, y por eso la existencia sale sin pelear.

Y las traspuestas no son un adorno de notación: C^t traduce el error, que vive en $\mathbb{R}^k$, al espacio del estado, donde vive el adjunto; A^t es la traspuesta de la dinámica, y aparece al integrar por partes; y B^t traduce el adjunto al espacio de los controles, que es donde tiene que estar $\hat{v}$. **Cada traspuesta lleva un vector del sitio donde está al sitio donde hace falta**. Comprobar dimensiones detecta productos imposibles; para decidir entre A y A^t, ambas cuadradas, hay que usar también la identidad del producto escalar.

Paso 1: existencia y unicidad, por el método directo. $I \ge 0$; sucesión minimizante acotada por $\|v_n\|^2 \le 2I(v_n)/\mu$; subsucesión débilmente convergente a $\hat{v}$; por (4.17), $y_n(T) - \hat{y}(T) = \int_0^T e^{A(T-s)}B(v_n - \hat{v})(s)\,ds \to 0$ por la convergencia débil (cada componente es un producto escalar en L^2 contra una fila fija de $e^{A(T-s)}B$); luego $|Cy_n(T) - z_T|^2 \to |C\hat{y}(T) - z_T|^2$, y el paso 4 es idéntico: $v_n \to \hat{v}$ en norma e $I(\hat{v}) = \text{ínf } I$. Unicidad por convexidad estricta, porque I es cuadrático con $\frac{\mu}{2}\|v\|^2$.

Paso 2 (el núcleo): el sistema de optimalidad, y la traspuesta que lo hace funcionar. $F_w(\varepsilon) = \frac{\alpha}{2}|C(\hat{y}(T) + \varepsilon z_w(T)) - z_T|^2 + \frac{\mu}{2}\|\hat{v} + \varepsilon w\|^2$, y $F_w'(0) = 0$ es

$$\alpha\big(C^t(C\hat{y}(T) - z_T), z_w(T)\big)_{\mathbb{R}^n} + \mu\int_0^T (\hat{v}, w)\,dt = 0 \qquad \text{para todo } w.$$

Con $\hat\varphi$ solución de $-\hat\varphi' = A^t\hat\varphi$, $\hat\varphi(T) = \alpha C^t(C\hat y(T) - z_T)$, la derivada de $(z_w, \hat\varphi)$ es $(w, B^t\hat\varphi)$ (el término con A se cancela y no hay fuente), e integrando, $(\hat\varphi(T), z_w(T)) = \int_0^T (B^t\hat\varphi, w)\,dt$. Queda $\int_0^T (B^t\hat\varphi + \mu\hat v, w)\,dt = 0$ para todo w, y el sistema de optimalidad es

$$\hat y' = A\hat y + B\hat v, \quad \hat y(0) = y_0; \qquad \hat v = -\frac{1}{\mu}B^t\hat\varphi;$$
$$-\hat\varphi' = A^t\hat\varphi, \quad \hat\varphi(T) = \alpha C^t(C\hat y(T) - z_T).$$

Paso 3: el método numérico, y por qué μ grande basta. $v^\ell \mapsto y^\ell \mapsto \varphi^\ell \mapsto v^{\ell+1} = -\frac{1}{\mu}B^t\varphi^\ell$, con el adjunto resuelto hacia atrás desde $\varphi^\ell(T) = \alpha C^t(Cy^\ell(T) - z_T)$. Es contractivo para μ grande. Con E como antes, $|y_1(T) - y_2(T)| \le \sqrt{T}E\|B\|\,\|v_1 - v_2\|$ y $\|\varphi_1 - \varphi_2\| \le \sqrt{T}E\alpha\|C\|^2|y_1(T) - y_2(T)|$. Así,

$$\|\Lambda(v_1) - \Lambda(v_2)\| \le \frac{\alpha T E^2\|B\|^2\|C\|^2}{\mu}\|v_1 - v_2\|,$$

y basta que el coeficiente sea menor que uno. Como el adjunto es explícito, $\hat\varphi(t) = e^{A^t(T-t)}\hat\varphi(T)$, el problema también se reduce, como en el ejercicio del control con observación final, a un sistema lineal para $Y = \hat y(T)$:

$$\left(I_n + \frac{\alpha}{\mu}WC^tC\right)Y = e^{AT}y_0 + \frac{\alpha}{\mu}WC^tz_T, \qquad W := \int_0^T e^{A(T-s)}BB^te^{A^t(T-s)}\,ds.$$

Se obtiene sustituyendo $\hat v(t) = -(\alpha/\mu)B^te^{A^t(T-t)}C^t(CY - z_T)$ en (4.17). La matriz W es el gramiano de controlabilidad. El sistema para Y tiene solución única por la unicidad del problema de control, incluso cuando W es singular.

Qué hace cada hipótesis. $\mu > 0$ aparece dos veces y por dos motivos distintos: da la cota del paso 1 (sin ella la sucesión minimizante podría irse al infinito) y da el $1/\mu$ que hace contractiva la iteración del paso 3. La existencia sigue siendo válida con $\alpha = 0$: el término con μ conserva la coercividad y el óptimo es $v = 0$. Al tomar $\alpha > 0$ se penaliza además el error final. Y de C no se ha pedido nada: puede ser cualquier matriz, incluso de rango bajo, porque lo que se observa es $Cy(T)$ y el adjunto recoge el error ya traducido por C^t. Si C tiene núcleo, las variaciones del estado final en ese núcleo no cambian directamente el término de observación. Aun así, esas componentes pueden moverse por el acoplamiento de la dinámica o porque el mismo control modifica varias componentes del estado.

Un caso con números: un carrito que hay que parar en el origen. Sea $y = (y_1, y_2)$ la posición y la velocidad de un carrito empujado por una fuerza v por unidad de masa:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad C = (1, 0), \qquad z_T = 0,$$

con $T = \alpha = \mu = 1$ e $y_0 = (1, 0)^t$: parte en reposo a distancia 1 y solo importa lo cerca del origen que acabe. Como A es nilpotente, las exponenciales son polinómicas: $e^{As} = \left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)$ y $e^{A^ts} = \left(\begin{smallmatrix} 1 & 0 \\ s & 1 \end{smallmatrix}\right)$. Con $Y_1 = \hat y_1(T)$, el dato final del adjunto es $\hat\varphi(T) = \alpha C^tCY = (Y_1, 0)^t$, y

$$\hat\varphi(t) = e^{A^t(T-t)}\hat\varphi(T) = \begin{pmatrix} Y_1 \\ (T-t)Y_1 \end{pmatrix}, \qquad \hat v(t) = -B^t\hat\varphi(t) = -(1-t)\,Y_1.$$

La segunda componente del adjunto tiene una lectura directa: un cambio δ de la velocidad en el instante t desplaza la posición final en $(T-t)\delta$, y por eso su precio es el de la posición multiplicado por el tiempo que queda. La posición final es $Y_1 = 1 + \int_0^1 (1-s)\hat{v}(s)\,ds = 1 - \frac{Y_1}{3}$, luego $Y_1 = \frac{3}{4}$, y el control óptimo es $\hat{v}(t) = -\frac{3}{4}(1-t)$: frena con fuerza al principio y se apaga al final, porque una fuerza temprana actúa durante más tiempo. El coste es $I(\hat{v}) = \frac{1}{2}\cdot\frac{9}{16} + \frac{1}{2}\cdot\frac{9}{16}\cdot\frac{1}{3} = \frac{3}{8} = 0.375$, frente a $\frac{1}{2}$ sin control. El sistema con el gramiano da lo mismo: $W = \left(\begin{smallmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{smallmatrix}\right)$, y su primera fila es $Y_1(1+\frac{1}{3}) = 1$. (Minimizar directamente sobre 200 valores de v a trozos da también 0.375.)

Qué hemos aprendido y dónde se falla. Lo que se lleva el ejercicio es que cada componente del adjunto es el precio de una componente del estado, y que A^t es la regla con que los precios se transmiten hacia atrás. Los fallos, con este mismo carrito:

- *Escribir A en vez de A^t en la adjunta.* Las dos son 2×2 y las dimensiones no avisan. Con A, el adjunto sería $e^{A(T-t)}\hat{\varphi}(T) = (Y_1, 0)^t$, luego $B^t\hat{\varphi} = 0$ y el «control óptimo» sería $v \equiv 0$: no hacer nada, con coste $\frac{1}{2}$. El error dice que la velocidad no tiene precio, cuando es justo lo que el control puede mover.
- *Cambiar el signo del dato final*, $\hat{\varphi}(T) = \alpha C^t(z_T - CY)$. El control empuja hacia fuera y sale $Y_1 = 1 + \frac{Y_1}{3}$, es decir, $Y_1 = \frac{3}{2}$: el carrito acaba más lejos de lo que empezó. La comprobación es la del caso escalar: el control óptimo tiene que dejar el estado observado más cerca del objetivo que el sistema libre.
- *Olvidar la C^t.* Con $C = (1,0)$, el producto $C(CY - z_T)$ es un vector fila y no se puede usar como dato de un adjunto que es columna. Este fallo sí lo detectan las dimensiones: en MATLAB, `A'*phi` con `phi` fila da error. Pero no al sumar: una fila más una columna no da error, sino que MATLAB la expande a una matriz 2×2 sin avisar.

4.4 EDO lineal de segundo orden

LA PREGUNTA

Un oscilador (un amortiguador, un péndulo pequeño) se gobierna con una fuerza $v(t)$. Su ecuación es de segundo orden. ¿Hace falta una teoría nueva, o basta con lo anterior y un cambio de variable, y cómo quedan entonces las condiciones del adjunto?

Se considera el problema de valores iniciales para una EDO lineal de segundo orden

$$\begin{cases} y'' = a(t)y + b(t)y' + v, & t \in (0,T), \\ y(0) = y_{00},\ y'(0) = y_{01}, \end{cases} \tag{4.24}$$

con $a, b \in C^0([0,T])$ y $y_{00}, y_{01} \in \mathbb{R}$. Para $v \in L^2$, se entiende $y \in C^1$ con y' absolutamente continua y la EDO válida casi en todas partes. Es el oscilador del capítulo 3 con una fuerza exterior v por unidad de masa, donde $a(t)y$ es la fuerza recuperadora (o repulsiva) y $b(t)y'$ el amortiguamiento cuando $b < 0$; los dos datos iniciales son la posición y la velocidad. Se elige la fuerza $v(t)$ que hace seguir al oscilador una trayectoria y_d gastando poco, y el problema de control

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{\alpha}{2}\displaystyle\int_0^T |y - y_d|^2\,dt + \dfrac{\mu}{2}\int_0^T |v|^2\,dt \\ \text{Sujeto a: } v \in L^2(0,T),\ (y,v) \text{ verifica (4.24)}, \end{cases} \tag{4.25}$$

con $\alpha, \mu > 0$ e $y_d \in L^2(0,T)$ dada. El problema (4.24) se transforma con el cambio de variable del apéndice (sección A.1) en un SDO de dimensión 2: poniendo $(y_1, y_2) = (y, y')$,

$$\begin{cases} Y' = AY + Bv, \\ Y(0) = Y_0, \end{cases} \qquad Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad Y_0 = \begin{pmatrix} y_{00} \\ y_{01} \end{pmatrix}, \tag{4.26}$$

con $n = 2$ y $m = 1$, y A que ahora depende de t (la teoría de la sección anterior vale igual, con el operador de evolución $U(t,s) = \Phi(t)\Phi(s)^{-1}$ en lugar de $e^{A(t-s)}$, donde $\Phi' = A(t)\Phi$ y $\Phi(0) = I$). El problema equivalente es

$$\begin{gathered} \text{Minimizar } J(v) := \frac{\alpha}{2}\int_0^T |y_1 - y_d|^2\,dt + \frac{\mu}{2}\int_0^T |v|^2\,dt, \\ \text{sobre } v \in L^2(0,T), \qquad (Y,v) \text{ verifica (4.26)}, \end{gathered} \tag{4.27}$$

que es (4.18) con $C = (1,0)$: se observa solo la primera componente. Por tanto posee solución única.

Teorema. En las condiciones precedentes, $\hat{v} \in L^2(0,T)$ es solución de (4.25), con $\hat{y}$ el estado asociado, si y solo si existe $\hat{\varphi} \in C^1([0,T])$, con $\hat{\varphi}' + b\hat{\varphi}$ absolutamente continua, tal que $(\hat{y}, \hat{\varphi}, \hat{v})$ satisface

$$\begin{cases} \hat{y}'' = a\hat{y} + b\hat{y}' + \hat{v}, & t \in (0,T), \\ \hat{y}(0) = y_{00},\ \hat{y}'(0) = y_{01}, & \\ (\hat{\varphi}' + b\hat{\varphi})' = a\hat{\varphi} + \alpha(\hat{y} - y_d), & \text{casi todo } t \in (0,T), \\ \hat{\varphi}(T) = 0,\ \hat{\varphi}'(T) = 0, & \\ \hat{v} = -\dfrac{1}{\mu}\hat{\varphi}. & \end{cases}$$

La derivada se deja sobre el producto porque b solo se supone continua. Si $b \in C^1$, la adjunta se escribe

$$\hat{\varphi}'' = (a - b')\hat{\varphi} - b\hat{\varphi}' + \alpha(\hat{y} - y_d).$$

Para b constante desaparece b'; si además $b = 0$, se recupera la ecuación $\hat{\varphi}'' = a\hat{\varphi} + \alpha(\hat{y} - y_d)$. La deducción siguiente cubre los tres casos.

El plan, antes de las cuentas: el teorema de sistemas da un adjunto *vectorial* de dos componentes; la segunda es la única que ve el control (porque $B = (0,1)^t$), y se la llama $\hat{\varphi}$; la primera se elimina despejándola de una ecuación y derivando, y así el sistema de dos ecuaciones de primer orden en el adjunto se convierte en una de segundo orden. Vale la pena hacerlo primero con $a = b = 0$, la ecuación $y'' = v$ de un cuerpo empujado por una fuerza: el adjunto vectorial cumple $-\hat{\varphi}_1' = \alpha(\hat{y} - y_d)$ y $-\hat{\varphi}_2' = \hat{\varphi}_1$, luego $\hat{\varphi}_2'' = -\hat{\varphi}_1' = \alpha(\hat{y} - y_d)$: la adjunta es $\hat{\varphi}'' = \alpha(\hat{y} - y_d)$ con $\hat{\varphi}(T) = \hat{\varphi}'(T) = 0$, y $\hat{v} = -\hat{\varphi}/\mu$. Es lo que dice el teorema con $a = b = 0$. El caso general es la misma cuenta arrastrando a y b.

Merece la pena anticipar en qué se traduce cada pieza, porque es lo que da sentido al resultado. Como el control entra solo en la segunda componente ($B = (0,1)^t$), el control óptimo resulta proporcional a $\hat{\varphi}_2$ y no a $\hat{\varphi}_1$: es el precio de la *velocidad*, no el de la posición, lo que se empareja directamente con la fuerza en la primera variación del coste. Y como se observa solo la primera componente ($C = (1,0)$), la desviación alimenta la ecuación de $\hat{\varphi}_1$ y llega a $\hat{\varphi}_2$ por acoplamiento. Por eso al final la ecuación adjunta sale de segundo orden en $\hat{\varphi}_2$.

Paso 1: escribir el problema como sistema y aplicar el teorema vectorial. Se aplica el teorema de optimalidad para sistemas al sistema (4.26) con $C = (1,0)$, $z_d = y_d$. El adjunto $\hat{\Phi} = (\hat{\varphi}_1, \hat{\varphi}_2)^t$ verifica $-\hat{\Phi}' = A^t\hat{\Phi} + \alpha C^t(C\hat{Y} - y_d)$ con $\hat{\Phi}(T) = 0$, y como

$$A^t = \begin{pmatrix} 0 & a \\ 1 & b \end{pmatrix}, \qquad C^t(C\hat{Y} - y_d) = \begin{pmatrix} \hat{y}_1 - y_d \\ 0 \end{pmatrix},$$

se tiene, componente a componente,

$$-\hat{\varphi}_1' = a\hat{\varphi}_2 + \alpha(\hat{y} - y_d), \qquad -\hat{\varphi}_2' = \hat{\varphi}_1 + b\hat{\varphi}_2, \qquad \hat{\varphi}_1(T) = \hat{\varphi}_2(T) = 0,$$

y el control es $\hat{v} = -\frac{1}{\mu}B^t\hat{\Phi} = -\frac{1}{\mu}\hat{\varphi}_2$.

Paso 2: eliminar $\hat{\varphi}_1$ para volver al segundo orden (el núcleo). Llamando $\hat{\varphi} := \hat{\varphi}_2$, la segunda ecuación da $\hat{\varphi}_1 = -\hat{\varphi}' - b\hat{\varphi}$, y derivando y sustituyendo en la primera,

$$-\hat{\varphi}_1' = (\hat{\varphi}' + b\hat{\varphi})' = a\hat{\varphi} + \alpha(\hat{y} - y_d).$$

Esta identidad es válida aunque b no sea derivable: la combinación $\hat{\varphi}' + b\hat{\varphi} = -\hat{\varphi}_1$ sí es absolutamente continua. Si $b \in C^1$, la regla del producto da la fórmula expandida anterior.

Paso 3: traducir las condiciones finales. $\hat{\varphi}(T) = \hat{\varphi}_2(T) = 0$ y $\hat{\varphi}'(T) = -\hat{\varphi}_1(T) - b\hat{\varphi}(T) = 0$. ■

Qué hace cada hipótesis. La continuidad de a, b permite resolver el sistema vectorial; no hace falta derivarlos. Exigir $b \in C^1$ solo sirve para expandir la derivada del producto en la ecuación escalar. Que el control entre solo por la segunda componente —la fuerza actúa sobre la aceleración, no sobre la posición— es lo que hace que $\hat{v}$ dependa solo de $\hat{\varphi}_2$; si el modelo permitiera empujar directamente la posición, B tendría otra forma y aparecerían las dos componentes.

LA IDEA: CUATRO CONDICIONES, DOS EN CADA EXTREMO

Un sistema de segundo orden en $\hat{y}$ y otro en $\hat{\varphi}$ son cuatro ecuaciones de primer orden y necesitan cuatro condiciones: las dos iniciales del estado y las dos finales del adjunto. Es exactamente el problema de contorno que `bvp4c` resolvería; `ControlOrden2.m`, en cambio, lo integra con `ode45` desde $t = 0$ tomando $\varphi(0) = \varphi'(0) = 0$, que *no* es el sistema de optimalidad (las condiciones del adjunto van en T). La práctica 28 lo comenta. Para b constante, el término de la adjunta es $-b\hat{\varphi}'$ frente a $b\hat{y}'$ en el estado; con b variable hay que conservar también la derivada del producto.

Ejercicio 4.3. *Observación final para la EDO de segundo orden*

ORIGEN: ejercicio 4.4 del curso.

Para (4.24) y el coste $I(v) = \frac{\alpha}{2}|y(T) - y_T|^2 + \frac{\mu}{2}\int_0^T |v|^2\,dt$ con $y_T \in \mathbb{R}$, probar la existencia y unicidad de solución y obtener la condición de optimalidad.

Resolución.
La idea. Todo el ejercicio consiste en una *traducción*, y conviene decirlo antes de empezar porque es lo único que cuesta. La teoría general está escrita para sistemas de primer orden, y una EDO de segundo orden se convierte en uno de ellos con el truco de siempre, $Y = (y, y')$. Aplicada la teoría, el adjunto sale **vectorial**, con dos componentes, y hay que volver a bajarlo a una sola ecuación escalar. La pregunta que organiza la cuenta es: *¿cuál de las dos componentes del adjunto es la que se empareja con el control?* La respuesta es la segunda, porque el control entra en la segunda ecuación del sistema, y por eso $\hat{v} = -\hat{\varphi}_2/\mu$.
Y de ahí sale lo que a primera vista desconcierta. El coste observa la *posición* $y(T)$, así que la condición terminal cae sobre $\hat{\varphi}_1$; pero el adjunto que sobrevive es $\hat{\varphi}_2$, y lo que $\hat{\varphi}_1(T)$ dice sobre él es una condición sobre su **derivada** en T. **Observar la posición se traduce en una condición sobre la velocidad del adjunto**: es el reflejo exacto de que en el estado el control mueve la aceleración y no la posición.
Paso 1: existencia y unicidad, por traducción. Por el ejercicio de observación final en un sistema aplicado al sistema (4.26). *Paso 2 (el núcleo): el adjunto vectorial y su bajada a escalar.* Adjunto sin fuente, $-\hat{\Phi}' = A^t\hat{\Phi}$, con $\hat{\Phi}(T) = (\alpha(\hat{y}(T) - y_T), 0)^t$, y $\hat{v} = -\hat{\varphi}_2/\mu$. En componentes, $-\hat{\varphi}_1' = a\hat{\varphi}_2$, $-\hat{\varphi}_2' = \hat{\varphi}_1 + b\hat{\varphi}_2$, con $\hat{\varphi}_1(T) = \alpha(\hat{y}(T) - y_T)$ y $\hat{\varphi}_2(T) = 0$. Con $\hat{\varphi} = \hat{\varphi}_2$, se obtiene $(\hat{\varphi}' + b\hat{\varphi})' = a\hat{\varphi}$, y las condiciones finales son $\hat{\varphi}(T) = 0$ y $\hat{\varphi}'(T) = -\hat{\varphi}_1(T) - b\hat{\varphi}(T) = -\alpha(\hat{y}(T) - y_T)$ —aquí está la traducción anunciada, y es el único punto delicado—:

$$\hat{y}'' = a\hat{y} + b\hat{y}' + \hat{v}, \quad \hat{y}(0) = y_{00},\ \hat{y}'(0) = y_{01};$$
$$(\hat{\varphi}' + b\hat{\varphi})' = a\hat{\varphi}, \quad \hat{\varphi}(T) = 0,\ \hat{\varphi}'(T) = -\alpha(\hat{y}(T) - y_T); \qquad \hat{v} = -\frac{1}{\mu}\hat{\varphi}.$$

Paso 3: el caso más simple, resuelto entero ($a = b = 0$). Es el carrito del ejercicio de observación final en un sistema, ahora con objetivo y_T y datos cualesquiera. Sea $Y = \hat{y}(T)$. La adjunta es $\hat{\varphi}'' = 0$ con $\hat{\varphi}(T) = 0$ y $\hat{\varphi}'(T) = -\alpha(Y - y_T)$, luego

$$\hat{\varphi}(t) = \alpha(Y - y_T)(T - t), \qquad \hat{v}(t) = -\frac{\alpha}{\mu}(Y - y_T)(T - t)\ :$$

una fuerza que decrece linealmente y se anula en T, la misma forma que salió allí. Para cerrar el acoplamiento se usa la fórmula de Taylor con resto integral, $\hat{y}(T) = y_{00} + y_{01}T + \int_0^T (T-s)\hat{v}(s)\,ds$, que dice que la fuerza aplicada en el instante s desplaza la posición final en proporción al tiempo que le queda, $T - s$. Como $\int_0^T (T-s)^2 ds = T^3/3$,

$$Y\Big(1 + \frac{\alpha T^3}{3\mu}\Big) = y_{00} + y_{01}T + \frac{\alpha T^3}{3\mu}\, y_T.$$

Se lee como una media ponderada: Y queda entre el punto al que llegaría el carrito sin fuerza, $y_{00} + y_{01}T$, y el objetivo y_T, con pesos 1 y $\alpha T^3/(3\mu)$. El horizonte entra al cubo, porque un T mayor da a la vez más tiempo para empujar y más brazo de palanca. Con $y_{00} = 1$, $y_{01} = 0$, $y_T = 0$ y $T = \alpha = \mu = 1$ sale $Y = \frac{3}{4}$, como en la versión de sistema.

Qué hemos aprendido y dónde se falla. La traducción entre la forma vectorial y la escalar es mecánica, pero cada condición tiene un sitio exacto. Los fallos, con los mismos números:

- *Poner el dato en $\hat{\varphi}(T)$ en vez de en $\hat{\varphi}'(T)$.* Con $\hat{\varphi}(T) = \alpha(Y - y_T)$ y $\hat{\varphi}'(T) = 0$, la adjunta $\hat{\varphi}'' = 0$ da un adjunto constante y una fuerza constante, $v \equiv -Y$; entonces $Y = 1 - \frac{Y}{2}$, luego $Y = \frac{2}{3}$, $v \equiv -\frac{2}{3}$ y coste $\frac{4}{9} \approx 0.444$, peor que el 0.375 del óptimo. Es un control admisible, pero empuja igual al final, cuando ya no rinde, que al principio.
- *Equivocar el signo de $\hat{\varphi}'(T)$.* El control empuja hacia fuera y sale $Y = \frac{3}{2}$, igual que al cambiar el signo en la versión de sistema.
- *Integrar la adjunta desde $t = 0$ con $\hat{\varphi}(0) = \hat{\varphi}'(0) = 0$,* como hace `ControlOrden2.m`. Aquí da $\hat{\varphi} \equiv 0$ y $v \equiv 0$. Las dos condiciones del adjunto van en T.

QUÉ DEJA. La observación final de la posición se convierte en una condición sobre la *derivada* del adjunto en T, porque el adjunto escalar $\hat{\varphi}$ es la segunda componente del vectorial. Traducir entre la formulación vectorial y la escalar es la única dificultad de las EDOs de segundo orden.

4.5 EDO no lineal sin restricciones

LA PREGUNTA

Modelos como la logística con pesca variable o un tumor bajo terapia son no lineales. El método del capítulo sigue funcionando para deducir condiciones necesarias, pero ya no garantiza unicidad ni suficiencia. ¿Qué hipótesis permiten asegurar la existencia y qué parte del razonamiento se conserva?

La ecuación de estado es ahora

$$\begin{cases} y' = g(t, y, v), & t \in (0, T), \\ y(0) = y_0, \end{cases} \tag{4.28}$$

con $y_0 \in \mathbb{R}$ y $g \in C^1(\mathbb{R}^3)$ verificando

$$\Big|\frac{\partial g}{\partial t}\Big| + \Big|\frac{\partial g}{\partial y}\Big| + \Big|\frac{\partial g}{\partial v}\Big| \le C \qquad \text{en } \mathbb{R}^3 :$$

derivadas acotadas, luego g globalmente lipschitziana en y y en v. Conviene decir qué modelos caben aquí y cuáles no, literalmente. La logística con pesca variable, $g = \lambda y - \kappa y^2 -$

vy, tiene $\partial_y g = \lambda - 2\kappa y - v$, que no está acotada en todo $\mathbb{R}^3$: no cumple la hipótesis tal como está escrita. Con $\lambda, \kappa > 0$, $0 \le y_0 \le \lambda/\kappa$ y controles acotados $0 \le v \le M$, el intervalo $[0, \lambda/\kappa]$ es invariante: en sus extremos el campo apunta hacia dentro o es cero. En esa región se puede conservar g y prolongarla suavemente fuera de ella con derivadas acotadas. Esto justifica trabajar con esos controles y estados; no autoriza a aplicar sin más el teorema a todos los controles de L^2, que no tienen una cota puntual común.

Definición. Sea $v \in L^2(0,T)$. Solución de (4.28) en $[0,T]$ es toda función y absolutamente continua con $y'(t) = g(t, y(t), v(t))$ para casi todo $t \in (0,T)$ e $y(0) = y_0$.

En estas condiciones, para cada $v \in L^2(0,T)$ existe una única solución y_v en todo $[0,T]$. La demostración es algo más complicada que en el caso lineal: un argumento posible es reescribir (4.28) como una ecuación de punto fijo $y = \Lambda(y)$ con $\Lambda(y)(t) = y_0 + \int_0^t g(s, y(s), v(s))\, ds$ en $C^0([0,T])$, y probar que una potencia Λ^ℓ es contractiva (es la demostración de Picard del libro de EDO del autor, con la iteración de la constante de Lipschitz que produce el factor $L^\ell T^\ell / \ell!$).

Se considera el problema de control

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{\alpha}{2} \displaystyle\int_0^T |y - y_d|^2\, dt + \dfrac{\mu}{2} \int_0^T |v|^2\, dt \\ \text{Sujeto a: } v \in L^2(0,T),\ (y, v) \text{ verifica (4.28).} \end{cases} \tag{4.29}$$

Se mantienen $\alpha, \mu > 0$ e $y_d \in L^2(0,T)$. La existencia global de cada estado no implica por sí sola que exista un control óptimo. Para asegurar esto último hace falta una hipótesis que permita pasar al límite en la dependencia respecto del control.

Teorema. Bajo las hipótesis anteriores, si además g es afín respecto del control,

$$g(t, y, v) = g_0(t, y) + g_1(t, y) v,$$

el problema (4.29) posee al menos una solución. La unicidad no queda garantizada.

Dónde está la dificultad. Si $v_n \rightharpoonup \hat{v}$ en L^2, no se puede pasar sin más al límite dentro de una función no lineal de v_n. Por ejemplo, controles que alternan 1 y -1 cada vez más deprisa convergen débilmente a cero, pero $\cos(v_n) = \cos(1)$ no converge a $\cos(0)$. La continuidad de g, incluso con derivadas acotadas, no resuelve esa dificultad. La dependencia afín sí permite separar un coeficiente que converge uniformemente del control que converge débilmente.

1. Compacidad de controles y estados. Una sucesión minimizante (v_n) está acotada por $\mu \|v_n\|^2 \le 2J(v_n)$; se extrae $v_n \rightharpoonup \hat{v}$. La cota $|g(t,y,v)| \le c + C(|y| + |v|)$ para $0 \le t \le T$ y Gronwall dan una cota uniforme de $|y_n|$. Por la misma desigualdad, $\|y_n'\|_{L^2}$ está acotada. Además,

$$|y_n(t) - y_n(s)| \le \|y_n'\|_{L^2} |t - s|^{1/2}.$$

Los estados están uniformemente acotados y son equicontinuos; Arzelà–Ascoli permite extraer una subsucesión que converge uniformemente a una función $\hat{y}$.

2. Identificar el estado límite. Se usa la ecuación integral. Los términos $g_0(t, y_n)$ y $g_1(t, y_n)$ convergen uniformemente a los correspondientes con $\hat{y}$. Para el término que contiene el control se escribe

$$g_1(t, y_n) v_n - g_1(t, \hat{y}) \hat{v} = \big(g_1(t, y_n) - g_1(t, \hat{y})\big) v_n + g_1(t, \hat{y})(v_n - \hat{v}).$$

La integral del primer sumando tiende a cero por la convergencia uniforme y la cota de $\|v_n\|_{L^1}$. La del segundo sobre $[0,t]$ tiende a cero por la convergencia débil, al probarla contra $\mathbf{1}_{[0,t]}g_1(\cdot,\hat{y}) \in L^2$. Así, $\hat{y}$ satisface la ecuación integral con $\hat{v}$ y es su estado.
3. Pasar al coste. El término de seguimiento converge por la convergencia uniforme de los estados, mientras que $\|\hat{v}\|^2 \leq \liminf \|v_n\|^2$. Por tanto, $J(\hat{v}) \leq \liminf J(v_n) = \inf J$, y $\hat{v}$ es óptimo.

La dependencia afín es una condición suficiente para este argumento. Para una dependencia general en v se necesitan otras condiciones, como las de convexidad de los teoremas de existencia de control óptimo; no basta la regularidad inicial. Las condiciones necesarias que siguen, en cambio, se deducen para el g general de clase C^1, siempre que exista un óptimo.

Teorema. Sea $\hat{v}$ una solución de (4.29) y $\hat{y}$ el estado asociado. Entonces existe $\hat{\varphi}$, el estado adjunto, tal que $(\hat{y},\hat{\varphi},\hat{v})$ verifica el sistema de optimalidad

$$\begin{cases} \hat{y}' = g(t,\hat{y},\hat{v}), & t \in (0,T), \\ \hat{y}(0) = y_0, & \\ -\hat{\varphi}' = \dfrac{\partial g}{\partial y}(t,\hat{y},\hat{v})\,\hat{\varphi} + \alpha(\hat{y}-y_d), & t \in (0,T), \\ \hat{\varphi}(T) = 0, & \\ \mu\hat{v} + \dfrac{\partial g}{\partial v}(t,\hat{y},\hat{v})\,\hat{\varphi} = 0. & \end{cases} \tag{4.30}$$

La última condición puede escribirse como $\hat{v} = -(1/\mu)\partial_v g(t,\hat{y},\hat{v})\hat{\varphi}$; en general es implícita, porque $\hat{v}$ también aparece a la derecha. Para $g = ay + v$ vuelve a ser $\hat{v} = -\hat{\varphi}/\mu$.

La intuición falsa, y lo que de verdad se pierde. Al ver que el sistema de optimalidad tiene la misma forma que en el caso lineal, es tentador pensar que el teorema dice lo mismo. No lo dice, y la diferencia está en una sola palabra del enunciado: aquí pone «*entonces existe*», y en el caso lineal ponía «*si y solo si*». La razón es concreta y está en el paso 2: cuando g no es lineal, el estado no depende linealmente del control, la función $\varepsilon \mapsto J(\hat{v}+\varepsilon w)$ deja de ser una parábola, y «derivada nula» vuelve a ser lo que era en el capítulo 3: una condición *necesaria* y nada más. Resolver el sistema de optimalidad puede devolver un mínimo, un máximo o un punto de silla, exactamente igual que resolver $\nabla J = 0$ en $\mathbb{R}^n$. Y por el mismo motivo se pierde la garantía de unicidad: nada impide que haya varios controles óptimos.

Y lo que se conserva, que es el mecanismo. Los tres movimientos siguen ahí, y en el mismo orden. Lo único que cambia es que *todo se linealiza en torno al óptimo*: donde antes había una constante a ahora hay $\partial g/\partial y$ evaluada en el óptimo, y donde había un 1 multiplicando al control hay $\partial g/\partial v$. La ecuación que gobierna cómo responde el estado a una perturbación pequeña ya no es la ecuación de estado, sino su *linealizada* alrededor de $(\hat{y},\hat{v})$; y la ecuación adjunta es la traspuesta de esa linealizada. La lectura de precios sombra sobrevive intacta: el precio del estado se propaga hacia atrás por la dinámica *tal como se ve desde el óptimo*.

Idea de la demostración, con los pasos rellenados.
1. Que la derivada exista: dependencia diferenciable respecto del parámetro. Fijados $v, w \in L^2(0,T)$, la función $\varepsilon \mapsto J(v+\varepsilon w)$ está bien definida y es continuamente diferenciable en $\mathbb{R}$: el estado $y_{v+\varepsilon w}$ depende diferenciablemente de ε por el teorema de dependencia

diferenciable respecto de parámetros (consecuencia de que $g \in C^1$ con derivadas acotadas), y J es un polinomio de segundo grado en el estado y en el control.
2. Perturbar y derivar — pero ya no es una parábola. Si $\hat{v}$ es óptimo, $J(\hat{v}) \le J(\hat{v}+\varepsilon w)$ para todo ε y w, luego por el teorema de Fermat

$$\frac{d}{d\varepsilon} J(\hat{v}+\varepsilon w)\Big|_{\varepsilon=0} = 0 \qquad \text{para todo } w \in L^2(0,T).$$

Ahora la función de ε ya no es una parábola: la condición es solo *necesaria*.
3. La ecuación linealizada: cómo responde el estado (el núcleo). Sea $\hat{y}$ el estado de $\hat{v}$ y $z_w := \frac{d}{d\varepsilon} y_{\hat{v}+\varepsilon w}|_{\varepsilon=0}$ la derivada del estado respecto del control, que se obtiene derivando (4.28) respecto de ε en $\varepsilon = 0$ (regla de la cadena):

$$\begin{cases} z' = \dfrac{\partial g}{\partial y}(t,\hat{y},\hat{v})\, z + \dfrac{\partial g}{\partial v}(t,\hat{y},\hat{v})\, w, & t \in (0,T), \\ z(0) = 0, \end{cases}$$

la **ecuación linealizada**. Derivando J,

$$\alpha \int_0^T (\hat{y}-y_d) z_w \, dt + \mu \int_0^T \hat{v} w \, dt = 0 \qquad \text{para todo } w \in L^2(0,T).$$

4. La integración por partes, con el coeficiente linealizado. Con $\hat{\varphi}$ como en (4.30), la misma integración por partes de siempre, ahora con el coeficiente $\frac{\partial g}{\partial y}$ en lugar de a:

$$\alpha \int_0^T (\hat{y}-y_d) z_w \, dt = \int_0^T \Big(-\hat{\varphi}' - \frac{\partial g}{\partial y}\hat{\varphi}\Big) z_w \, dt = \int_0^T \hat{\varphi}\Big(z_w' - \frac{\partial g}{\partial y} z_w\Big) dt = \int_0^T \hat{\varphi}\, \frac{\partial g}{\partial v}\, w \, dt,$$

con los términos de frontera nulos por $\hat{\varphi}(T) = z_w(0) = 0$. Queda $\int_0^T \big(\mu\hat{v} + \frac{\partial g}{\partial v}\hat{\varphi}\big) w \, dt = 0$ para todo w, de donde $\mu\hat{v} + \frac{\partial g}{\partial v}(t,\hat{y},\hat{v})\hat{\varphi} = 0$ casi en todas partes. ■

Qué hace cada hipótesis. Que g sea *de clase C^1 con derivadas acotadas* es lo que sostiene los pasos 1 y 3: la acotación da la existencia global del estado por Gronwall —igual que en el lema de la ecuación de estado lineal, con la constante de Lipschitz en el papel de $|a|$— y la regularidad C^1 es la que permite derivar respecto de ε y escribir la ecuación linealizada. Con menor regularidad haría falta otro argumento para justificar esa derivación. La hipótesis $\mu > 0$ sigue dando la coercividad que la existencia necesita, pero ya **no basta para la unicidad**: la convexidad de J venía de que el estado dependía linealmente del control, y eso es justo lo que se ha perdido.

La idea: qué se pierde y qué queda

Se pierde la garantía de unicidad, porque J puede dejar de ser convexo como función de v, y se pierde la suficiencia: el sistema de optimalidad puede tener soluciones que no son mínimos. La existencia se conserva bajo la hipótesis adicional del teorema anterior. Queda el método para deducir condiciones necesarias, con dos cambios mecánicos: en el adjunto, a se sustituye por $\partial g/\partial y$ evaluado en la trayectoria óptima (la ecuación adjunta es siempre lineal, es la traspuesta de la linealizada), y en la condición del

control aparece $\partial g/\partial v$. Es el mismo esquema que en el capítulo 2: la teoría lineal se aplica al sistema linealizado en la solución.

El hamiltoniano. Con la función

$$H(t,y,\varphi,v) := \frac{\alpha}{2}|y-y_d|^2 + \frac{\mu}{2}|v|^2 + \varphi\, g(t,y,v),$$

llamada **hamiltoniano**, el sistema de optimalidad (4.30) se reescribe

$$\begin{cases} \hat{y}' = \dfrac{\partial H}{\partial \varphi}(t,\hat{y},\hat{\varphi},\hat{v}), & t\in(0,T), \quad \hat{y}(0)=y_0, \\ -\hat{\varphi}' = \dfrac{\partial H}{\partial y}(t,\hat{y},\hat{\varphi},\hat{v}), & t\in(0,T), \quad \hat{\varphi}(T)=0, \\ \dfrac{\partial H}{\partial v}(t,\hat{y},\hat{\varphi},\hat{v}) = 0. \end{cases} \tag{4.31}$$

Se comprueba: $\partial_\varphi H = g$; $\partial_y H = \alpha(y-y_d) + \varphi\partial_y g$; $\partial_v H = \mu v + \varphi\partial_v g$. Las dos primeras tienen la forma canónica de las ecuaciones de Hamilton para el par (y,φ); la relación de signos con el momento mecánico se precisa en la sección 4.7. La tercera dice que **el control óptimo hace estacionario el hamiltoniano en casi todo instante**. La sección 4.7 lo convierte en regla general.

Ejercicio 4.4. *Unicidad e iteración para μ grande en el caso no lineal*

ORIGEN: ejercicio 4.5 del curso.

Suponer además $g \in C^2$ con derivadas segundas acotadas y que existe un control óptimo (por ejemplo, bajo la hipótesis afín del teorema de existencia). Probar que, si μ es suficientemente grande, ese óptimo es único y se puede obtener mediante una iteración de (4.30) que converge en $L^2(0,T)$, comenzando en $v^0 = 0$.

PISTAS, DE MENOR A MAYOR

1. *El método: $v^\ell \mapsto y^\ell$ (ecuación de estado no lineal) $\mapsto \varphi^\ell$ (adjunta, lineal) $\mapsto v^{\ell+1} = -\frac{1}{\mu}\frac{\partial g}{\partial v}(t,y^\ell,v^\ell)\varphi^\ell$.*
2. *Las cotas: g lipschitziana en (y,v) y Gronwall dan $\|y_1-y_2\| \le c\|v_1-v_2\|$; las derivadas segundas acotadas permiten estimar la diferencia de los coeficientes adjuntos. Trabaja en una bola fija de L^2 y comprueba que la iteración permanece en ella.*
3. *Si Λ es contractiva, tiene un único punto fijo, y todo control óptimo es un punto fijo; hay que asegurar que todos los óptimos están en esa bola.*

Resolución.
La idea, y cabe en una frase. La condición de optimalidad $\mu\hat{v} + \partial_v g\,\hat{\varphi} = 0$ se lee como un **punto fijo**: despejando, $\hat{v} = -\frac{1}{\mu}\partial_v g\,\hat{\varphi}$, donde el miembro de la derecha depende de $\hat{v}$ —a través del estado y del adjunto— pero *dividido por* μ. En una bola fija de controles, las cotas de estado y adjunto son uniformes. Si μ es grande, ese factor $1/\mu$ hace contractiva la aplicación y permite que conserve la bola.
El coste da la bola que hace falta: todo óptimo satisface $\mu\|\hat{v}\|^2 \le 2J(\hat{v}) \le 2J(0)$. Aquí $J(0)$ no depende de μ. Por tanto, fijado $R>0$, todos los óptimos pertenecen a $\mathcal{B}_R := \{v\in L^2 : \|v\|\le R\}$ cuando $\mu \ge 2J(0)/R^2$.

Paso 1: la aplicación cuyo punto fijo se busca. Sea $\Lambda(v) := -\frac{1}{\mu}\partial_v g(t, y_v, v)\varphi_v$, con y_v el estado y φ_v el adjunto de (4.30) calculado con (y_v, v). Para $v \in \mathcal{B}_R$, Gronwall y la fórmula integral del adjunto dan constantes M_R, P_R independientes de μ tales que

$$\|y_v\|_\infty \le M_R, \qquad \|\varphi_v\|_\infty \le P_R := \alpha e^{CT}\big(TM_R + \sqrt{T}\|y_d\|\big).$$

Como $|\partial_v g| \le C$, se tiene $\|\Lambda(v)\| \le C\sqrt{T}P_R/\mu$. Si $\mu \ge C\sqrt{T}P_R/R$, Λ lleva la bola cerrada $\mathcal{B}_R$ a sí misma; por tanto los iterados que empiezan en $v^0 = 0$ permanecen en ella.
Paso 2 (el núcleo): las tres cotas encadenadas. Para v_1, v_2: por la lipschitzianidad de g y Gronwall, $|y_1(t) - y_2(t)| \le \int_0^t C(|y_1 - y_2| + |v_1 - v_2|)\,ds$ da $\|y_1 - y_2\| \le c_1\|v_1 - v_2\|$. El adjunto $\varphi_1 - \varphi_2$ resuelve una EDO lineal hacia atrás con coeficiente acotado y fuente $\alpha(y_1 - y_2) + (\partial_y g_1 - \partial_y g_2)\varphi_2$; como $\|\varphi_2\|_\infty \le P_R$ y las derivadas segundas están acotadas, $\|\partial_y g_1 - \partial_y g_2\| \le L(\|y_1 - y_2\| + \|v_1 - v_2\|)$. Otra aplicación de Gronwall da $\|\varphi_1 - \varphi_2\| \le c_2\|v_1 - v_2\|$. Entonces

$$\|\Lambda(v_1) - \Lambda(v_2)\| \le \frac{1}{\mu}\Big(\|\partial_v g_1\|_\infty\|\varphi_1 - \varphi_2\| + \|\partial_v g_1 - \partial_v g_2\|\,\|\varphi_2\|_\infty\Big) \le \frac{c_3}{\mu}\|v_1 - v_2\|,$$

donde c_1, c_2, c_3 dependen de R y de los datos, pero no de μ. Es contractiva en $\mathcal{B}_R$ si $\mu > c_3$.
Paso 3: de la contracción a la unicidad del óptimo. Se elige μ que cumpla las tres cotas anteriores. La bola cerrada es completa, Λ la conserva y es contractiva; Banach da un único punto fijo en ella y convergencia de los iterados desde $v^0 = 0$. Todo control óptimo está en la bola y satisface (4.30), luego es ese punto fijo. Como se ha supuesto existencia, el límite es el único óptimo.

QUÉ DEJA. Para que «μ grande» dé convergencia hacen falta tres comprobaciones: todos los óptimos están en una bola fija, la iteración conserva esa bola y su constante de Lipschitz es menor que uno. La regularidad adicional controla cómo cambian los coeficientes del adjunto.

Ejercicio 4.5. *Observación final en el caso no lineal*

ORIGEN: ejercicio 4.6 del curso.

Para (4.28) y el coste $I(v) = \frac{\alpha}{2}|y(T)|^2 + \frac{\mu}{2}\int_0^T |v|^2\,dt$, aceptando la existencia de control óptimo, obtener el sistema de optimalidad y un método numérico convergente para μ grande bajo la hipótesis adicional C^2 del ejercicio anterior.

PISTAS, DE MENOR A MAYOR

1. *Combina la sección 4.2 (dato final del adjunto) con el teorema de optimalidad en el caso no lineal (coeficientes linealizados).*

Resolución.
La idea. Este ejercicio es la suma de dos que ya están hechos, y no hace falta ni una idea nueva: la no linealidad aporta los *coeficientes* de la ecuación adjunta —que ahora son las derivadas parciales de g evaluadas a lo largo del óptimo, en vez de constantes— y la observación final aporta el *dato terminal* $\hat\varphi(T) = \alpha\hat y(T)$, en vez de cero. La regla general del capítulo lo resume: **lo que se paga durante el trayecto alimenta la ecuación del adjunto; lo**

que se paga al llegar fija con qué valor arranca su viaje hacia atrás. Aquí solo se paga al llegar, luego el adjunto va sin fuente y con dato terminal proporcional a $\hat{y}(T)$, que puede ser cero.

Paso 1 (el núcleo): la condición necesaria y el adjunto que borra z_w. Con z_w la solución de la ecuación linealizada, la condición necesaria es $\alpha\hat{y}(T)z_w(T) + \mu\int_0^T \hat{v}w\,dt = 0$ para todo w. Con $\hat{\varphi}$ solución de $-\hat{\varphi}' = \partial_y g(t,\hat{y},\hat{v})\hat{\varphi}$, $\hat{\varphi}(T) = \alpha\hat{y}(T)$, se tiene $\frac{d}{dt}(\hat{\varphi}z_w) = \hat{\varphi}\partial_v g\, w$, e integrando, $\alpha\hat{y}(T)z_w(T) = \int_0^T \hat{\varphi}\partial_v g\, w\,dt$. El sistema de optimalidad es

$$\hat{y}' = g(t,\hat{y},\hat{v}),\ \hat{y}(0) = y_0; \qquad -\hat{\varphi}' = \partial_y g(t,\hat{y},\hat{v})\hat{\varphi},\ \hat{\varphi}(T) = \alpha\hat{y}(T);$$
$$\mu\hat{v} + \partial_v g(t,\hat{y},\hat{v})\hat{\varphi} = 0,$$

Paso 2: el método numérico. Es el del ejercicio anterior con este adjunto, contractivo para μ grande en una bola fija como la anterior. Ahora $\|\varphi_v\|_\infty \le \alpha M_R e^{CT}$, y la diferencia de adjuntos tiene además un dato final $\alpha(y_1(T) - y_2(T))$, controlado por Gronwall. Con esas cotas se asegura que la bola contiene todos los óptimos, que la iteración la conserva y que es contractiva.

Qué cambia respecto del caso lineal. El adjunto escalar todavía tiene una fórmula exponencial:

$$\hat{\varphi}(t) = \alpha\hat{y}(T)\exp\left(\int_t^T \partial_y g(s,\hat{y}(s),\hat{v}(s))\,ds\right).$$

Lo que ya no se conoce de antemano es el coeficiente: depende de la trayectoria y del control que se buscan. La fórmula permite verificar signos e integrar el adjunto una vez conocidos ambos; resolver todo el sistema acoplado suele requerir un método numérico.

4.6 EDO lineal con restricciones de desigualdad

LA PREGUNTA

Un programa de repoblación solo puede introducir individuos y tiene una capacidad máxima por unidad de tiempo: $0 \le v(t) \le M$. La condición «derivada igual a cero» ya no puede ser la respuesta, porque el mínimo puede estar en el borde. ¿Qué la sustituye, y cómo se calcula el control óptimo con las cotas?

Interesa ahora una situación más realista, en la que v está sometido a restricciones:

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{\alpha}{2}\displaystyle\int_0^T |y - y_d|^2\,dt + \dfrac{\mu}{2}\int_0^T |v|^2\,dt \\ \text{Sujeto a: } v \in \mathcal{U}_{\text{ad}},\ (y,v) \text{ verifica (4.1)}, \end{cases} \tag{4.33}$$

donde

$$\mathcal{U}_{\text{ad}} = \{v \in L^2(0,T) : v_* \le v \le v^* \text{ c.p.d.}\} \tag{4.34}$$

es el conjunto de **controles admisibles**, con $v_*, v^* \in \mathbb{R}$, $v_* < v^*$, y «c.p.d.» significa «en casi todo punto».

(A) Existencia y unicidad

Teorema. El problema (4.33) posee exactamente una solución $\hat{v}$.

La idea, en una frase. Restringir el conjunto de controles solo puede estropear una cosa del argumento anterior: que el límite de la sucesión minimizante siga siendo admisible. Todos los demás pasos no miraban nunca de dónde salían los v_n. Y ese único punto delicado tiene una respuesta con nombre: un convexo cerrado es también *débilmente* cerrado, así que la convergencia débil —que era la que se había pagado por recuperar compacidad— no se escapa del conjunto admisible.

Paso 1: el único punto delicado — que el límite débil siga siendo admisible (el núcleo). Para la existencia se razona como en el teorema de existencia y unicidad del control básico. Hay que tener cuidado en lo siguiente: si los $v_n \in \mathcal{U}_{\text{ad}}$ convergen débilmente a $\hat{v}$, entonces $\hat{v} \in \mathcal{U}_{\text{ad}}$. Esto ocurre siempre que $\mathcal{U}_{\text{ad}}$ es **convexo y cerrado** en $L^2(0,T)$, porque un convexo cerrado es débilmente cerrado (teorema de Mazur). En el caso de (4.34) se comprueba directamente: si $\hat{v} > v^*$ en un conjunto E de medida positiva, tomando $w = \mathbf{1}_E$ en la definición de convergencia débil, $\int_E v_n \to \int_E \hat{v} > v^*|E|$, mientras que $\int_E v_n \leq v^*|E|$ para todo n: contradicción; igual con v_*. Y $\mathcal{U}_{\text{ad}}$ es convexo: una combinación convexa de funciones entre las cotas está entre las cotas.
Paso 2: lo demás se repite sin tocar nada. Los pasos 1 a 4 del teorema de existencia y unicidad del control básico valen literalmente, porque nunca usaron que los v_n recorrieran todo el espacio: solo usaron que estuvieran acotados, y una restricción solo puede ayudar en eso.
Paso 3: unicidad. No tiene dificultad adicional: J es estrictamente convexo y el punto medio de dos óptimos distintos estaría en $\mathcal{U}_{\text{ad}}$ (convexo) con coste menor. ■

Qué hace cada hipótesis. Las dos propiedades que se le piden a $\mathcal{U}_{\text{ad}}$ no son adorno, y cada una bloquea un fallo distinto. Si $\mathcal{U}_{\text{ad}}$ no fuera *cerrado* —por ejemplo el intervalo abierto (v_*, v^*)— el ínfimo podría alcanzarse solo en el borde, fuera del conjunto, y no habría solución: es el mismo fenómeno que hace que $\text{ínf}\{x : x \in (0,1)\}$ no se alcance. Si no fuera *convexo*, se pierden las dos cosas a la vez: el teorema de Mazur deja de aplicarse y el límite débil puede salirse del conjunto. Por ejemplo, si $\mathcal{U}_{\text{ad}} = \{v \in L^2 : v(t) \in \{-1, 1\} \text{ c.p.d.}\}$, una sucesión que alterna signos cada vez más rápido converge débilmente a 0, que no es admisible; y además el argumento de unicidad del paso 3 se cae, porque el punto medio de dos controles admisibles podría no ser admisible.

(B) Caracterización

Conviene introducir el **operador de proyección** $P_{ad} : \mathbb{R} \to [v_*, v^*]$,

$$P_{ad}(z) = \begin{cases} v_* & \text{si } z < v_*, \\ z & \text{si } z \in [v_*, v^*], \\ v^* & \text{si } z > v^*, \end{cases}$$

que a cada número le asigna el punto más cercano del intervalo.

Teorema. Sea $\hat{v} \in \mathcal{U}_{\text{ad}}$ e $\hat{y}$ su estado. Entonces $\hat{v}$ es la única solución de (4.33) si y solo si existe $\hat{\varphi}$ tal que $(\hat{y}, \hat{\varphi}, \hat{v})$ verifica

$$\begin{cases} \hat{y}' = a\hat{y} + \hat{v}, & t \in (0,T), \\ \hat{y}(0) = y_0, & \\ -\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y} - y_d), & t \in (0,T), \\ \hat{\varphi}(T) = 0, & \\ \hat{v} = P_{ad}\Big(-\dfrac{1}{\mu}\hat{\varphi}\Big). & \end{cases} \tag{4.35}$$

La idea: cuando no se puede empujar en las dos direcciones. El mecanismo de tres movimientos sigue siendo el mismo, pero el primero cambia, y cambia por una razón geométrica que conviene ver antes que la cuenta. Hasta ahora el óptimo estaba en el interior del conjunto de controles y se podía empujar hacia los dos lados; por eso la derivada tenía que ser cero, ni positiva ni negativa. Con restricciones el óptimo puede estar *pegado a una cota*, y desde ahí solo se puede empujar hacia adentro. La condición se rebaja entonces de «la derivada es cero» a «la derivada es ≥ 0 en toda dirección admisible»: no se puede mejorar avanzando hacia el interior, aunque hacia afuera sí se mejorase si estuviera permitido. Eso es una *desigualdad variacional*, y es exactamente lo que en $\mathbb{R}^n$ ocurre al minimizar sobre un convexo, donde $\nabla J = 0$ se sustituye por $\nabla J \cdot (v - \hat{v}) \geq 0$.

Y el desenlace tiene una lectura muy concreta. Como en cada instante el problema es minimizar una parábola sobre un intervalo, la respuesta es la de siempre: si el vértice cae dentro, ahí está el mínimo; si cae fuera, el mínimo está en el extremo más cercano. Eso es la proyección, y por eso la fórmula $\hat{v} = -\hat{\varphi}/\mu$ del caso libre se convierte aquí en $\hat{v} = P_{ad}(-\hat{\varphi}/\mu)$: la misma decisión de antes, recortada a lo que es posible.

Es análoga a la del teorema de optimalidad del control básico, con una diferencia en el paso 1.

Paso 1: perturbar solo hacia dentro — de la igualdad a la desigualdad variacional. Ya no se puede perturbar en todas las direcciones. Si $\hat{v} \in \mathcal{U}_{\text{ad}}$ y $v \in \mathcal{U}_{\text{ad}}$, entonces $\hat{v} + \varepsilon(v - \hat{v}) \in \mathcal{U}_{\text{ad}}$ para $\varepsilon \in [0,1]$ (convexidad), sin que esto esté garantizado para $\varepsilon < 0$. La función $F_{v-\hat{v}}(\varepsilon) = J(\hat{v} + \varepsilon(v - \hat{v}))$ es la misma parábola convexa de antes, y $\hat{v}$ es óptimo si y solo si $F_{v-\hat{v}}(0) \leq F_{v-\hat{v}}(\varepsilon)$ para $\varepsilon \in [0,1]$ y todo $v \in \mathcal{U}_{\text{ad}}$, que para una parábola convexa equivale a

$$F'_{v-\hat{v}}(0) = \frac{d}{d\varepsilon} J\big(\hat{v} + \varepsilon(v - \hat{v})\big)\Big|_{\varepsilon=0} \geq 0 \quad \text{para todo } v \in \mathcal{U}_{\text{ad}} :$$

la **desigualdad variacional**.

Paso 2: el adjunto, la misma cuenta de antes con una sola dirección permitida. Lo único que cambia respecto del caso libre es que la dirección ya no es arbitraria: hay que tomar $w = v - \hat{v}$, con $v \in \mathcal{U}_{\text{ad}}$. Calculando la derivada como en (4.11) con esa w, y con $\hat{\varphi}$ el adjunto de (4.12), la integración por partes da

$$\int_0^T (\hat{\varphi} + \mu\hat{v})(v - \hat{v})\, dt \geq 0 \qquad \text{para todo } v \in \mathcal{U}_{\text{ad}}.$$

Paso 3: de la desigualdad integral a una decisión instante a instante (el núcleo). Falta ver que esto equivale a $\hat{v} = P_{ad}(-\hat{\varphi}/\mu)$. Como la desigualdad vale para toda $v \in \mathcal{U}_{\text{ad}}$,

se puede elegir v punto a punto, y equivale a la condición *puntual*: para casi todo t,

$$\big(\hat\varphi(t)+\mu\hat v(t)\big)\big(s-\hat v(t)\big)\ge 0 \qquad \text{para todo } s\in[v_*,v^*].$$

(Basta comprobar los dos extremos $s=v_*$ y $s=v^*$. Si una de esas desigualdades fallara en un conjunto de medida positiva, se toma v igual a ese extremo allí y a $\hat v$ fuera: la integral sería negativa. Las demás s son combinaciones convexas de los extremos.) Dividiendo por $\mu>0$ y llamando $z=-\hat\varphi(t)/\mu$, la condición es $(\hat v(t)-z)(s-\hat v(t))\ge 0$ para todo $s\in[v_*,v^*]$, que es exactamente la caracterización de la proyección de z sobre el intervalo: si z está dentro, obliga a $\hat v(t)=z$ (tomar $s=z$); si $z>v^*$, obliga a $\hat v(t)=v^*$ (con $\hat v(t)<v^*$ se tomaría $s=v^*$ y saldría negativo); si $z<v_*$, a $\hat v(t)=v_*$. Luego $\hat v(t)=P_{ad}(-\hat\varphi(t)/\mu)$. Todos los pasos son equivalencias. ■

Qué hace cada hipótesis. La *convexidad* de $\mathcal{U}_{\text{ad}}$ es la que permite escribir la perturbación como $\hat v+\varepsilon(v-\hat v)$ con $\varepsilon\in[0,1]$ y saber que sigue siendo admisible; sin ella no habría direcciones admisibles garantizadas y el paso 1 no arrancaría. Que $\mathcal{U}_{\text{ad}}$ esté definido por cotas *puntuales* $v_*\le v(t)\le v^*$ es lo que sostiene el paso 3: es lo que permite elegir v instante a instante y romper una condición integral en una familia de condiciones escalares. Con una restricción global —por ejemplo $\int_0^T|v|^2\le K$ con $K>0$, un presupuesto total de energía— la condición inicial es integral. Puede expresarse con un multiplicador de Lagrange escalar común a todo el intervalo, que acopla las decisiones en distintos instantes. Y $\mu>0$ vuelve a hacer lo mismo: es lo que permite dividir para pasar de $\hat\varphi+\mu\hat v$ a $z=-\hat\varphi/\mu$, y lo que garantiza que la parábola de cada instante tiene mínimo.

LA IDEA: MINIMIZAR UNA PARÁBOLA EN UN INTERVALO

Sin restricciones, el mínimo de $\frac{\mu}{2}s^2+\hat\varphi s$ está en $s=-\hat\varphi/\mu$. Con $s\in[v_*,v^*]$, el mínimo es ese punto si cae dentro y el extremo más cercano si cae fuera: la proyección. El teorema dice que el problema con restricciones se resuelve *instante a instante* de esa manera, con $\hat\varphi(t)$ como pendiente: es el principio del máximo de Pontryaguin en su forma con controles acotados. En el capítulo 3 no hubo restricciones de este tipo; en programación matemática, la desigualdad variacional $J'(\hat v)[v-\hat v]\ge 0$ caracteriza el mínimo de un funcional convexo diferenciable sobre un conjunto convexo, y con $\mathcal{U}_{\text{ad}}$ dado por cotas se convierte en las condiciones de Karush-Kuhn-Tucker: el multiplicador de la cota activa es $|\hat\varphi+\mu\hat v|$.

Vemos pues que, en este caso, al igual que para (4.2), gracias a la linealidad de la EDO y al carácter cuadrático del coste, la existencia de $\hat\varphi$ con (4.35) es condición necesaria y suficiente de optimalidad. Numéricamente, se sustituye $\hat v=P_{ad}(-\hat\varphi/\mu)$ en la ecuación de estado y se resuelve el problema de contorno, que ahora es *no lineal* por la proyección; `bvp4c` puede tratarlo, como en la práctica 27 bis. La fórmula recorta $-\hat\varphi/\mu$ con el adjunto del problema restringido. No basta con recortar el control óptimo libre: al hacerlo cambia el estado y, con él, el adjunto, por lo que hay que resolver de nuevo el sistema acoplado.

Recortar no es optimizar: un caso resuelto a mano. Tómese el problema del caso más simple de la sección 4.1 ($a=0$, $y_d=0$, $y_0=\alpha=\mu=T=1$), cuyo control libre es

$v(t) = -\operatorname{sinh}(1-t)/\cosh 1$, y añádase la cota $v \geq -\frac{1}{2}$: no se puede retirar población a más ritmo que $\frac{1}{2}$. El control libre arranca en -0.762, así que la cota muerde al principio.

- *El recorte del control libre*, $\max\{v(t), -\frac{1}{2}\}$, satura hasta el instante en que $\operatorname{sinh}(1-t) = \frac{1}{2}\cosh 1$, es decir, hasta $t \approx 0.290$, y después coincide con el libre.
- *El óptimo verdadero* es $\hat{v} = P_{ad}(-\hat{\varphi}/\mu)$ con *su* adjunto. Satura en un intervalo $[0, \tau]$ y después deja de saturar. En $[\tau, 1]$ el sistema es el del caso libre, así que $\hat{y} = c\cosh(1-t)$ y $\hat{v} = -c\operatorname{sinh}(1-t)$, con $\hat{v}(1) = 0$. Hay dos incógnitas, c y τ, y dos condiciones en τ. El control es continuo, porque el adjunto lo es y la proyección también: $c\operatorname{sinh}(1-\tau) = \frac{1}{2}$. El estado es continuo y en $[0, \tau]$ vale $1 - \frac{t}{2}$: $c\cosh(1-\tau) = 1 - \frac{\tau}{2}$. Dividiendo, $\frac{1}{2}\coth(1-\tau) = 1 - \frac{\tau}{2}$, cuya raíz es $\tau \approx 0.316$, con $c \approx 0.677$. (En $[0, \tau]$ el adjunto crece hacia atrás, porque $-\hat{\varphi}' = \hat{y} > 0$, y $-\hat{\varphi}/\mu$ se queda por debajo de $-\frac{1}{2}$: la saturación es coherente.)

Los dos controles son admisibles y no son el mismo: el óptimo satura hasta 0.316 y el recorte suelta la cota en 0.290. Los costes son 0.38445 y 0.38452 (el libre, sin cota, 0.38080). La razón se lee con el precio sombra: al principio no se pudo empujar todo lo que se quería, la población es mayor que en el problema libre, su precio $\hat{\varphi}$ es más alto, y compensa seguir empujando al máximo durante más tiempo. Que la diferencia de coste sea pequeña es lo esperable: cerca del óptimo el coste es plano, y un error de primer orden en el control cuesta de segundo orden. Pero el control es otro, y la práctica 27 bis lo muestra con otros datos: allí el libre cruza la cota en $t = 0.2618$ y el óptimo restringido la suelta en $t = 0.3178$.

4.7 Hamiltoniano

LA PREGUNTA

Cada sección de este capítulo ha repetido la misma deducción con pequeños cambios. ¿Hay una regla que dé el sistema de optimalidad de cualquier problema de control de una vez, sin repetir la integración por partes, y qué relación tiene con la Mecánica del capítulo 3?

Con carácter general, el sistema de optimalidad de un problema de control para un SDO se puede obtener con el **hamiltoniano**. La regla que sigue es formal: presupone que el estado y el coste están bien definidos y que la regularidad permite las variaciones e integraciones por partes indicadas. Como antes, se busca una condición necesaria para un óptimo existente; no se prueba aquí un teorema de existencia general.

1. Se considera el problema de valores iniciales

$$\begin{cases} y' = g(t, y, v), & t \in (0, T), \\ y(0) = y_0, \end{cases} \tag{4.36}$$

   con $g : [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ e $y_0 \in \mathbb{R}^n$.
2. Se considera el problema de control

$$\begin{cases} \text{Minimizar } J(v) := \displaystyle\int_0^T G(t, y, v)\, dt + \Psi(y(T)) \\ \text{Sujeto a: } v \in L^2(0, T; \mathbb{R}^m),\ (y, v) \text{ verifica (4.36)}, \end{cases} \tag{4.37}$$

   con $G : [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ y $\Psi : \mathbb{R}^n \to \mathbb{R}$ dadas: un coste integral general más un coste final general.
3. Por definición, el hamiltoniano del problema es

$$H(t, y, \varphi, v) := G(t, y, v) + g(t, y, v) \cdot \varphi,$$

   con $\varphi \in \mathbb{R}^n$.

LA IDEA: DE DÓNDE SALE H, Y POR QUÉ ESA Y NO OTRA COMBINACIÓN

El hamiltoniano parece caído del cielo, y no lo es: es la regla de los **multiplicadores de Lagrange** aplicada a una restricción que resulta ser una ecuación diferencial. Minimizar $J(v)$ *sujeto a* que (y, v) cumpla $y' = g(t, y, v)$ es minimizar con una restricción de igualdad, y lo que se hace en $\mathbb{R}^n$ con una restricción $c(x) = 0$ —formar $F(x) + \lambda c(x)$ y buscar puntos críticos— se hace aquí formando el **funcional aumentado**

$$\mathcal{L}(y, v, \varphi) := \int_0^T \Big[G(t, y, v) + \varphi \cdot \big(g(t, y, v) - y' \big) \Big] dt + \Psi(y(T)),$$

donde el multiplicador $\varphi(t)$ es ahora una *función*, porque hay una restricción por cada instante t. Mírese el corchete: es $G + g \cdot \varphi$, que es exactamente H, menos el término $\varphi \cdot y'$. Ese es el hamiltoniano: **todo lo que hay dentro de la integral salvo el término que lleva la derivada**. Y el multiplicador φ tiene, por eso mismo, la lectura de sensibilidad que ya apareció en las secciones anteriores: el producto $\varphi(t) \cdot \delta y$ mide, a

primer orden, el cambio del coste futuro al perturbar el estado y mantener el control posterior. Puede tener cualquier signo.
Derivar $\mathcal{L}$ da las ecuaciones del estado y el adjunto y una condición de estacionariedad respecto del control:

- **Derivar en** φ devuelve la restricción, $y' = g$: la primera línea es la ecuación de estado, tal cual, hacia delante desde y_0.
- **Derivar en** y exige antes quitarle la derivada a y, integrando por partes $-\int \varphi \cdot y' = -\left[\varphi \cdot y\right]_0^T + \int \varphi' \cdot y$. El integrando pasa a ser $G + g \cdot \varphi + \varphi' \cdot y$, y anular su derivada respecto de y da $\varphi' + \left(\partial g/\partial y\right)^t \varphi + \partial G/\partial y = 0$, que es la segunda línea. En la primera variación, el término de frontera es $-\varphi(T) \cdot \delta y(T)$ y se combina con $\nabla\Psi(y(T)) \cdot \delta y(T)$: de ahí, y solo de ahí, sale la condición final $\varphi(T) = \nabla\Psi(\hat{y}(T))$.
- **Derivar en** v da $\partial H/\partial v = 0$, que sin restricciones es la tercera condición de estacionariedad.

La idea: cómo se lee (4.38) en voz alta, y qué es lo que de verdad se ha ganado

Las tres líneas dicen, por orden: *propaga el estado hacia delante*, *propaga el precio hacia atrás*, *y en cada instante elige el control más barato*. La primera es la física del problema y no ha cambiado. La segunda es la misma dinámica *traspuesta* y recorrida al revés, con una fuente que mide cuánto le importa al coste estar donde se está: el coste final fija su dato en T y el coste acumulado aporta la fuente. No se trata necesariamente de un descuento económico. La tercera es la que hay que mirar despacio, porque es el motivo de todo el capítulo: **la minimización sobre v es puntual y finito-dimensional**. En cada instante t, con $\hat{y}(t)$ y $\hat{\varphi}(t)$ ya conocidos, hay que resolver un problema de mínimos en $\mathbb{R}^m$, no en un espacio de funciones. El problema original vivía en $L^2(0,T;\mathbb{R}^m)$, de dimensión infinita; el sistema de optimalidad lo ha convertido en dos EDO acopladas más una familia de minimizaciones pequeñas. Esa es la ganancia, y es la razón de que el principio del máximo se pueda programar.

La idea: la traspuesta, que es donde se falla

En la ecuación adjunta aparece $\left(\partial g/\partial y\right)^t$, no $\partial g/\partial y$. Con $n = 1$ las dos son lo mismo y no se nota, y por eso el error sobrevive intacto hasta el primer sistema, donde ya cambia el resultado. La forma de acordarse no es memorizarlo, sino recordar de dónde salió: en el paso 2 de la demostración de la sección 4.3, la traspuesta es exactamente lo que permite pasar la matriz de un lado al otro del producto escalar, $(Az, \varphi) = (z, A^t\varphi)$, para que los dos términos con la dinámica se cancelen. Sin traspuesta no se cancelan, y el adjunto deja de servir para lo único que se le pide.
Conviene además tener presente que la regla es *formal*: no demuestra que el mínimo exista ni que estas condiciones sean suficientes. Lo demostrado son los teoremas de las secciones anteriores; el hamiltoniano es la forma rápida de recordarlos y de escribir el sistema de un problema nuevo, y lo sano es comprobar siempre que reproduce el caso conocido más parecido antes de fiarse de él en el caso nuevo.

La condición de *mínimo* del hamiltoniano requiere un paso más que anular una derivada. La idea del principio de Pontryaguin es cambiar el control a un valor w durante un intervalo

corto de longitud ε. En un punto de Lebesgue apropiado, la variación del coste, tras usar el adjunto, es

$$\varepsilon\big[H(t,\hat y(t),\hat\varphi(t),w)-H(t,\hat y(t),\hat\varphi(t),\hat v(t))\big]+o(\varepsilon).$$

Como el óptimo no mejora con esa sustitución, el corchete debe ser no negativo. Estas *variaciones de aguja* comparan valores del control que pueden estar lejos entre sí; son más informativas que una derivada respecto de su amplitud. Bajo las hipótesis que justifican este argumento, el sistema de condiciones necesarias para (4.37) es

$$\begin{cases} \hat y' = \dfrac{\partial H}{\partial \varphi} = g(t,\hat y,\hat v), & t\in(0,T), \quad \hat y(0)=y_0,\\[2mm] -\hat\varphi' = \dfrac{\partial H}{\partial y} = \Big(\dfrac{\partial g}{\partial y}(t,\hat y,\hat v)\Big)^t\hat\varphi + \dfrac{\partial G}{\partial y}(t,\hat y,\hat v), & t\in(0,T), \quad \hat\varphi(T)=\nabla\Psi(\hat y(T)),\\[2mm] H(t,\hat y,\hat\varphi,\hat v)\le H(t,\hat y,\hat\varphi,v) & \text{para todo } v\in\mathbb{R}^m,\ \text{c.p.d. } t. \end{cases} \tag{4.38}$$

Si H es diferenciable en v, la última desigualdad *implica* $\partial_v H(t,\hat y,\hat\varphi,\hat v)=0$ casi en todas partes. La recíproca necesita, por ejemplo, convexidad de H respecto de v: una derivada nula por sí sola puede señalar un máximo o una silla.

Con restricciones *puntuales* $\mathcal{U}_{\text{ad}}=\{v\in L^2(0,T;\mathbb{R}^m): v(t)\in U \text{ c.p.d.}\}$, donde $U\subset\mathbb{R}^m$ es cerrado y convexo, se minimiza sobre $w\in U$:

$$H(t,\hat y,\hat\varphi,\hat v)\le H(t,\hat y,\hat\varphi,w) \qquad \text{para todo } w\in U, \quad \text{c.p.d. } t.$$

Esto implica $\partial_v H(t,\hat y,\hat\varphi,\hat v)\cdot(w-\hat v)\ge 0$; de nuevo, la equivalencia exige convexidad de H en el control. Para un conjunto convexo general $\mathcal{U}_{\text{ad}}\subset L^2$, que puede acoplar distintos instantes, la variación admisible da la condición integral

$$\int_0^T \partial_v H(t,\hat y,\hat\varphi,\hat v)\cdot(v-\hat v)\,dt\ge 0 \qquad \text{para todo } v\in\mathcal{U}_{\text{ad}}.$$

No se puede convertir en minimizaciones independientes sin aprovechar la estructura de las restricciones.

El coste terminal Ψ no entra en H: su gradiente fija el dato final del adjunto. Se comprueba con las secciones anteriores: con $G=\frac{\alpha}{2}|y-y_d|^2+\frac{\mu}{2}|v|^2$, $g=ay+v$ y $\Psi=0$ sale (4.8); con $G=\frac{\mu}{2}|v|^2$ y $\Psi=\frac{\alpha}{2}|y|^2$ sale (4.14); con $g=Ay+Bv$ y $G=\frac{\alpha}{2}|Cy-z_d|^2+\frac{\mu}{2}|v|^2$ sale (4.19), porque $\partial_y(g\cdot\varphi)=A^t\varphi$ y $\partial_v(g\cdot\varphi)=B^t\varphi$.

Conviene hacer una vez la primera comprobación entera, porque es la que enseña a usar la regla. Con $H=\frac{\alpha}{2}|y-y_d|^2+\frac{\mu}{2}|v|^2+(ay+v)\varphi$:

- $\partial_\varphi H=ay+v$, y la primera ecuación de (4.38) es la de estado;
- $-\hat\varphi'=\partial_y H=\alpha(\hat y-y_d)+a\hat\varphi$, con $\hat\varphi(T)=\nabla\Psi=0$ porque no hay coste final: es la adjunta de (4.8);
- H es una parábola en v con coeficiente $\frac{\mu}{2}>0$, y su mínimo está donde $\partial_v H=\mu v+\hat\varphi=0$, es decir, en $\hat v=-\hat\varphi/\mu$.

En (4.14) solo cambian dos cosas: sin término integral de observación, $\partial_y H=a\varphi$ y la adjunta pierde la fuente; y con $\Psi=\frac{\alpha}{2}|y|^2$, el dato final es $\nabla\Psi(\hat y(T))=\alpha\hat y(T)$. Tres derivadas parciales sustituyen a la integración por partes de cada sección.

La idea: por qué «hamiltoniano», y la relación con el capítulo 3

Las dos primeras ecuaciones tienen la forma canónica de Hamilton, pero hay que cuidar los signos al identificar el momento. Un problema variacional mín $\int f(t,x,x')\,dt$ se escribe como control tomando $g = v$ y $G = f(t,y,v)$: el control es la velocidad. Entonces $H = f + \varphi \cdot v$, y la estacionariedad da

$$\varphi = -\partial_v f = -p, \qquad -\varphi' = \partial_y f,$$

donde $p = \partial_{x'} f$ es el momento variacional. Al combinar ambas igualdades se recupera $\frac{d}{dt}\partial_{x'} f = \partial_x f$, la ecuación de Euler–Lagrange. Cuando se puede eliminar la velocidad mediante la transformación de Legendre, el valor de H es $f - p \cdot x' = -H_{\text{mec}}$; el signo del momento cambia a la vez. Por tanto la conexión es exacta, aunque φ no coincide con p bajo la convención usada aquí. También se traducen las condiciones de los extremos: si el estado final está fijado, $\delta y(T) = 0$ y no se impone la condición de extremo libre $\varphi(T) = \nabla\Psi(y(T))$.

El camino que no se tomó: la programación dinámica de Bellman

La alternativa. Definir la *función valor* $V(t,x)$ como el coste óptimo del problema que empieza en el instante t en el estado x. Cuando V es diferenciable, satisface la ecuación en derivadas parciales de Hamilton-Jacobi-Bellman, $-\partial_t V = \text{ínf}_v\, H(t,x,\nabla_x V, v)$, con $V(T,x) = \Psi(x)$.
Su ventaja. Puede dar una ley de realimentación $\hat{v} = \hat{v}(t,x)$, aplicable a los estados de su dominio. Una solución C^1 de HJB con el dato final correcto, junto con una ley admisible que alcance el mínimo y cuyo estado exista, permite verificar optimalidad global. Las cotas puntuales se incorporan al conjunto sobre el que se minimiza. Esta vía conduce, por ejemplo, a la ecuación de Riccati del regulador lineal-cuadrático.
Qué la aparta aquí. V depende de $n+1$ variables: una malla para la EDP puede crecer muy deprisa con la dimensión, mientras que el sistema del principio del máximo tiene $2n$ EDO acopladas. Resolverlo tampoco es automático, pero resulta más próximo a las herramientas del capítulo. Además, V puede no ser diferenciable; entonces la formulación clásica necesita ampliarse. Cuando la regularidad lo permite, $\hat{\varphi}(t) = \nabla_x V(t, \hat{y}(t))$: las dos teorías se encuentran.

4.8 Nuevos problemas más complicados

La pregunta

Un oncólogo tiene dos fármacos, uno que mata células tumorales y otro que corta el riego sanguíneo del tumor, con dosis máximas y un límite de toxicidad acumulada. Quiere el tumor más pequeño posible al final del tratamiento. ¿Cómo se escribe eso como problema de control, y qué tienen estos problemas que la teoría del capítulo no cubre?

Se enuncian varios problemas de control de naturaleza más compleja, de interés en las aplicaciones. Los resultados teóricos que se pueden probar para ellos son de nivel más elevado y se omiten; quien quiera verlos tiene los libros de Lee y Markus y de Trélat. Lo que sí se puede hacer con lo aprendido es identificar en cada uno el estado, el control, el conjunto admisible y el coste, y escribir formalmente el hamiltoniano.

4.8.1 Sistema de EDOs no lineales con restricciones

Un modelo de la interacción entre el VIH y las células T, con dos tratamientos:

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{1}{2}\displaystyle\int_0^T \left(\alpha_1|v_1|^2 + \alpha_2|v_2|^2\right)dt - \int_0^T Y\,dt \\ \text{Sujeto a: } v = (v_1, v_2) \in \mathcal{U}_{\text{ad}},\ (Y, V, v) \text{ verifica (4.40)}, \end{cases} \tag{4.39}$$

con $\mathcal{U}_{\text{ad}} = \{v \in (L^2(0,T))^2 : a_i \le v_i \le b_i \text{ c.p.d.}\}$ y

$$\begin{cases} Y' = s_1 - \dfrac{s_2 V}{B_1 + V} - \mu Y - kYV + v_1 Y, & t \in (0,T), \\ V' = \dfrac{g(1-v_2)V}{B_2 + V} - cYV, & t \in (0,T), \\ Y(0) = Y_0,\ V(0) = V_0. \end{cases} \tag{4.40}$$

Aquí $Y(t)$ representa células T CD4$^+$ no infectadas y $V(t)$ la carga viral. Los anticuerpos son moléculas, no células: no son la variable de este modelo. Cada término del sistema es una frase, como en el capítulo 2. En Y': s_1 es la producción basal de células T, representada por una constante; $-\frac{s_2 V}{B_1+V}$ es una reducción de esa producción cuya magnitud crece con V pero se *satura* en s_2 cuando $V \gg B_1$; B_1 es el valor de V para el que la reducción vale $s_2/2$. $-\mu Y$ representa la muerte natural; aislando ese término, la disminución es exponencial y su semivida es $\log 2/\mu$ cuando $\mu > 0$; $-kYV$ es la pérdida de células T no infectadas por infección, modelada mediante encuentros; y $+v_1 Y$ añade v_1 a la tasa de crecimiento por individuo: el tratamiento entra de forma *bilineal*. En V': $\frac{g(1-v_2)V}{B_2+V}$ es la aportación de virus al compartimento observado, saturada y reducida por el tratamiento en el factor $1 - v_2$; con $v_2 = 1$ se anula esa aportación en el modelo; y $-cYV$ es el virus que las células inmunitarias contribuyen a eliminar mediante la respuesta inmunitaria, representada aquí por encuentros. El control v_1 es un tratamiento inmunoestimulante y v_2 un antirretroviral. El coste penaliza la intensidad de los tratamientos al cuadrado, como aproximación a su coste o efectos adversos, y premia (signo menos) la integral temporal de células T no infectadas. Lo que el modelo deja fuera: la latencia del virus, las mutaciones que lo hacen resistente, y que las dosis reales se administran a horas fijas y no de forma continua. Lo nuevo respecto de la teoría: dos controles con cotas, un sistema no lineal, y un coste que no es definido positivo (el término $-\int Y$).

Para la interpretación anterior se toman $B_1, B_2 > 0$, los demás parámetros biológicos no negativos, $s_1 \ge s_2$, $Y_0, V_0 \ge 0$, $0 \le a_i \le b_i$ y $b_2 \le 1$. Estas condiciones hacen invariante el cuadrante no negativo: en $Y = 0$ se tiene $Y' \ge s_1 - s_2 \ge 0$ y en $V = 0$ se tiene $V' = 0$. Además, $Y' \le s_1 + \max\{b_1 - \mu, 0\}Y$ y $V' \le g$, de modo que los estados tienen una cota uniforme en todo intervalo finito. Con $\alpha_i > 0$, el coste queda acotado inferiormente; la compacidad de los estados y la dependencia afín en los controles permiten aplicar el método directo. La unicidad no se deduce de este argumento. El modelo se estudia en H. R. Joshi, *Optimal Control Problems in PDE and ODE Systems* (2002), capítulo sobre VIH.

4.8.2 EDO con restricciones de otro tipo

En algunos casos hay que imponer restricciones no solo a los controles sino también a los **estados**. Dos ejemplos. El primero:

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{1}{2}\displaystyle\int_0^1 |v|^2\,dt \\ \text{Sujeto a: } v \in \mathcal{U}_{\text{ad}},\ x \in X_{ad},\ (x,v) \text{ verifica (4.42)}, \end{cases} \tag{4.41}$$

con $\mathcal{U}_{\text{ad}}$ con cotas como antes, $X_{ad} = \{x \in C^1([0,1]) : x'(0) = -x'(1) = 1,\ x(t) \le 1/9\ \forall t\}$ y

$$\begin{cases} x'' = v, & t \in (0,1), \\ x(0) = x(1) = 0. \end{cases} \tag{4.42}$$

Puede interpretarse como una posición controlada por la aceleración $v(t)$: parte con velocidad $+1$, termina con velocidad -1 y no puede superar la altura $1/9$. Para controles de L^2, se exige también que x' sea absolutamente continua. Las cotas del control deben ser compatibles con los datos: por ejemplo, $\int_0^1 v(t)\,dt = x'(1) - x'(0) = -2$ es una condición necesaria de admisibilidad. Una restricción de estado como $x(t) \le 1/9$ puede introducir un multiplicador que sea una medida, concentrada donde la restricción está activa. El adjunto sigue siendo una función, pero puede tener variación acotada y saltos asociados a átomos de esa medida. Es una ampliación del principio de Pontryaguin que queda fuera de este desarrollo. Por eso estos dos problemas se plantean y se leen término a término, como modelos, pero no se resuelven.

El segundo combina un tratamiento antiangiogénico con uno citotóxico:

$$\begin{cases} \text{Minimizar } J(v) := x_1(T) + \alpha\displaystyle\int_0^T |v_1|^2\,dt \\ \text{Sujeto a: } v = (v_1, v_2) \in \mathcal{U}_{\text{ad}},\ x \in X_{ad},\ (x,v) \text{ verifica (4.44)}, \end{cases} \tag{4.43}$$

con $\mathcal{U}_{\text{ad}} = \{v : 0 \le v_i \le b_i \text{ c.p.d.}\}$, $X_{ad} = \{x \in (C^0([0,T]))^3 : x_i \ge 0,\ x_2(t) \le B_2,\ x_3(t) \le B_3\ \forall t\}$ y

$$\begin{cases} x_1' = -a x_1 \log\dfrac{x_1}{x_2} - b x_1 v_2, & t \in (0,T), \\ x_2' = c x_1 - \mu x_2 - d x_1^{2/3} x_2 - c x_2 v_1 - f x_1 v_2, & t \in (0,T), \\ x_3' = v_1, & t \in (0,T), \\ x_i(0) = x_{i0},\ 1 \le i \le 3. \end{cases} \tag{4.44}$$

Es la respuesta a la pregunta del oncólogo: x_1 es el número de células tumorales, que sigue una ley de Gompertz (el $x\log(\theta/x)$ de la práctica 7, con la capacidad de soporte x_2 variable), x_2 es el número de células endoteliales que forman los vasos que alimentan el tumor, y x_3 es la dosis acumulada del fármaco antiangiogénico v_1, limitada por B_3: la toxicidad. Término a término: en x_1', $-a x_1 \log(x_1/x_2)$ es crecimiento de Gompertz hacia la capacidad x_2 (positivo si $x_1 < x_2$, negativo si la supera). Es una tendencia de crecimiento o decrecimiento, no una cota rígida $x_1 \le x_2$, y $-b x_1 v_2$ es el citotóxico, que elimina células con tasa per cápita $b v_2$. En x_2', $c x_1$ es el estímulo que el tumor manda a los vasos para que crezcan, proporcional a su tamaño; $-\mu x_2$ es la muerte natural de las células endoteliales; $-d x_1^{2/3} x_2$ es la inhibición que el propio tumor ejerce sobre los vasos; el exponente $2/3$ corresponde a una escala de superficie frente a volumen si se supone una geometría de forma aproximadamente constante; $-c x_2 v_1$ es el

antiangiogénico, que mata vasos; y $-fx_1v_2$ recoge que el citotóxico también daña los vasos. Y $x_3' = v_1$ no es biología: es un contador de la dosis total, puesto ahí para poder imponer un límite de dosis acumulada, $x_3(t) \le B_3$, como aproximación a una restricción de toxicidad. Con $x_{30} = 0$ se tiene $x_3(t) = \int_0^t v_1(s)\,ds$. El control v_2 es el citotóxico. El coste es el tamaño final del tumor más el coste de la terapia antiangiogénica. Lo que se pierde: la resistencia adquirida, la heterogeneidad del tumor y las dosis discretas. El fichero `TumorQuimio.m` del tema 2 ilustra la simulación de un modelo tumoral con tratamiento fijado, como paso previo a plantear su optimización.

El logaritmo exige trabajar con $x_1, x_2 > 0$; las desigualdades $x_i \ge 0$ de X_{ad} no bastan por sí solas para definir el campo en la frontera. Se consideran datos iniciales y trayectorias dentro de ese dominio, además de las cotas indicadas. Tampoco toda elección de controles dentro de sus cotas garantiza que el estado las respete: la admisibilidad se comprueba sobre la trayectoria completa.

4.8.3 Control óptimo bilineal

Se llama **bilineal** al caso en que el control multiplica al estado. Un ejemplo, relacionado con la producción de una fábrica:

$$\begin{cases} \text{Minimizar } J(v) := -\displaystyle\int_0^T e^{-at}(1-v)x\,dt \\ \text{Sujeto a: } v \in \mathcal{U}_{\text{ad}} = \{v \in L^2(0,T) : 0 \le v \le 1 \text{ c.p.d.}\},\ (x,v) \text{ verifica (4.46)}, \end{cases} \tag{4.45}$$

con

$$\begin{cases} x' = \gamma v x, \quad t \in (0,T), \\ x(0) = x_0. \end{cases} \tag{4.46}$$

Se suponen $x_0 > 0$, $\gamma > 0$ y $a \ge 0$. Aquí $x(t)$ es la producción y $v(t)$ la fracción que se reinvierte: el crecimiento es proporcional a la inversión. Lo que no se reinvierte, $(1-v)x$, se consume; el objetivo es maximizar ese consumo descontado a tasa a. El hamiltoniano es

$$H = -e^{-at}(1-v)x + \varphi\gamma v x = x\big(v\sigma - e^{-at}\big), \qquad \sigma(t) := e^{-at} + \gamma\varphi(t).$$

Como $x(t) = x_0 \exp(\gamma \int_0^t v) > 0$, minimizar H en $0 \le v \le 1$ equivale a minimizar $v\sigma$: reinvertir todo si $\sigma < 0$ y consumir todo si $\sigma > 0$.

LA IDEA: UNA RECTA DECIDE ENTRE INVERTIR Y CONSUMIR

En el caso lineal cuadrático, la dependencia de H en v era una parábola y el mínimo podía estar en el interior. Aquí es una recta: su pendiente decide qué extremo conviene. Cuando la pendiente es cero, todos los valores empatan en esa condición. Por tanto, la regla *bang-bang* requiere comprobar dónde se anula σ; el hecho de que el coste sea lineal no basta para descartar tramos singulares en otros problemas.
En este ejemplo se puede determinar el cambio. La ecuación adjunta da

$$\varphi' = e^{-at}(1-v) - \gamma v\varphi, \quad \varphi(T) = 0, \qquad \sigma' = (\gamma - a)e^{-at} - \gamma v\sigma.$$

La primera es $-\varphi' = \partial_x H = -e^{-at}(1-v) + \gamma v\varphi$. La segunda sale derivando $\sigma = e^{-at} + \gamma\varphi$ y sustituyendo la primera: $\sigma' = -ae^{-at} + \gamma e^{-at}(1-v) - \gamma^2 v\varphi = (\gamma - a)e^{-at} - \gamma v(e^{-at} + \gamma\varphi)$, y el último paréntesis es σ. Se tiene $\sigma(T) = e^{-aT} > 0$, porque

$\varphi(T) = 0$: al final conviene consumir, porque ya no queda tiempo para que la inversión rinda. Si $\gamma \le a$, σ permanece positiva y se consume desde el principio. La razón cabe en una línea: con $E(t) = \exp\big(\gamma \int_0^t v\big)$, la segunda ecuación da $(E\sigma)' = E\,(\sigma' + \gamma v\sigma) = (\gamma - a)e^{-at}E \le 0$, así que $E\sigma$ no crece y, como es positiva en T, es positiva en todo $[0, T]$. Dicho en palabras: si lo que rinde reinvertir no supera lo que descuenta el paso del tiempo, no hay ningún instante en que compense. Si $\gamma > a$, en cada cero se tiene $\sigma' = (\gamma - a)e^{-at} > 0$; solo puede cruzar de negativo a positivo, así que hay a lo sumo un cambio y ningún intervalo singular.
¿Cuánto tiempo debe quedar para que compense invertir? En el tramo final $v = 0$,

$$\sigma(t) = e^{-at}\left[1 - \frac{\gamma}{a}\big(1 - e^{-a(T-t)}\big)\right] \quad (a > 0),$$

y para $a = 0$ se obtiene $\sigma(t) = 1 - \gamma(T - t)$. Por ello, si $\gamma > a$, el horizonte mínimo para invertir es

$$h_* = \begin{cases} \dfrac{1}{a}\log\dfrac{\gamma}{\gamma - a}, & a > 0, \\ 1/\gamma, & a = 0. \end{cases}$$

Si $T \le h_*$, se consume todo desde el inicio. Si $T > h_*$, se reinvierte todo hasta $\tau = T - h_*$ y se consume todo después. El tramo de inversión existe solo cuando queda tiempo suficiente para recuperar lo invertido.

Con números, y una comprobación sin adjunto. Sean $\gamma = 0.5$ y $T = 10$. Sin descuento, $h_* = 1/\gamma = 2$: se reinvierte todo hasta $\tau = 8$ y se consume en las dos últimas unidades de tiempo. Se lee así: reinvertir una cantidad c aumenta la producción en γc por unidad de tiempo, y consumiendo ese aumento se tarda $1/\gamma$ en recuperar c; con menos tiempo por delante, reinvertir no compensa. Con descuento $a = 0.1$, $h_* = 10\log 1.25 \approx 2.23$, y el cambio se adelanta a $\tau \approx 7.77$: el consumo futuro vale menos y conviene empezar a consumir antes. La regla se puede comprobar sin adjunto dentro de las políticas «reinvertir hasta τ y consumir después». Con ellas $x(\tau) = x_0 e^{\gamma\tau}$, la producción queda constante desde τ, y el consumo total descontado es

$$x_0 e^{\gamma\tau}\int_\tau^T e^{-at}\,dt, \qquad \text{que para } a = 0 \text{ vale } x_0 e^{\gamma\tau}(T - \tau).$$

Su derivada respecto de τ es $x_0 e^{\gamma\tau}\big(\gamma(T - \tau) - 1\big)$ cuando $a = 0$, y se anula exactamente en $T - \tau = 1/\gamma$. Con $a > 0$ la derivada se anula donde $e^{-a(T-\tau)} = 1 - a/\gamma$, que es $T - \tau = h_*$. El consumo óptimo es 109.2 x_0 sin descuento y 44.7 x_0 con él, frente a 10 x_0 y 6.3 x_0 si se consume todo desde el principio. Lo que el cálculo elemental no da es que ninguna otra política, con más cambios o con valores intermedios de v, lo mejore; eso lo certifica la cota siguiente.

Aquí también se puede certificar que la regla da un mínimo global. El adjunto construido con la regla satisface $\varphi' = e^{-at} - \min\{0, \sigma\}$ y $\varphi(T) = 0$: donde $\sigma < 0$ la regla pone $v = 1$ y la ecuación adjunta da $\varphi' = -\gamma\varphi = e^{-at} - \sigma$; donde $\sigma > 0$ pone $v = 0$ y da $\varphi' = e^{-at}$. Ese φ queda fijo. Se toma ahora *cualquier* control admisible v, con su estado x ($x' = \gamma v x$), y se integra $(\varphi x)' = \varphi' x + \gamma v \varphi x$ entre 0 y T: como $\varphi(T) = 0$, sale $-\varphi(0)x_0 = \int_0^T x\big(e^{-at} - \min\{0, \sigma\} + \gamma v\varphi\big)\,dt$. Al sumarle $J(v) = \int_0^T x\big(-e^{-at} + ve^{-at}\big)\,dt$, los términos $\pm e^{-at}x$ se cancelan y $ve^{-at} +$

$\gamma v\varphi = v\sigma$:

$$J(v) - \varphi(0)x_0 = \int_0^T x(t)\big[v(t)\sigma(t) - \min\{0, \sigma(t)\}\big]\,dt \geq 0.$$

El integrando no es negativo porque $x > 0$ y, para $0 \leq v \leq 1$, $v\sigma \geq \min\{0, \sigma\}$: si $\sigma \geq 0$, $v\sigma \geq 0$; si $\sigma < 0$, $v\sigma \geq \sigma$ porque $v \leq 1$. La regla anterior alcanza la igualdad, porque pone $v = 1$ donde $\sigma < 0$ y $v = 0$ donde $\sigma > 0$: proporciona una cota inferior común a todos los controles y un control que la alcanza. Con los datos de antes ($\gamma = 0.5$, $T = 10$, $a = 0.1$), $\varphi(0)x_0 = -44.7\,x_0$, que es menos el consumo óptimo calculado arriba.

Otro problema bilineal, sobre un ecosistema presa-depredador:

$$\begin{cases} \text{Minimizar } J(v) := x(T) + y(T) \\ \text{Sujeto a: } v \in \mathcal{U}_{\text{ad}} = \{v \in L^2(0,T) : 0 \leq v \leq 1 \text{ c.p.d.}\},\ (x, y, v) \text{ verifica (4.48)}, \end{cases} \tag{4.47}$$

con

$$\begin{cases} x' = r_1 x - \mu_1 xyv, & t \in (0,T), \\ y' = -r_2 y + \mu_2 xyv, & t \in (0,T), \\ x(0) = x_0,\ y(0) = y_0. \end{cases} \tag{4.48}$$

Se toman $r_1, r_2, \mu_1, \mu_2, x_0, y_0 > 0$. Aquí x e y son las presas y los depredadores del modelo de Volterra del capítulo 2, y v es un **catalizador**: un control que activa la interacción entre las especies (con $v = 0$ no se encuentran, con $v = 1$ es Volterra). Se quiere la menor población total al final.

El mismo interruptor, y qué dice aquí. Este segundo problema no tiene coste distribuido: todo el coste está en el instante final, $\Psi(x, y) = x + y$. El hamiltoniano se reduce entonces a la parte de la dinámica,

$$H = \varphi_1(r_1 x - \mu_1 xyv) + \varphi_2(-r_2 y + \mu_2 xyv) = \varphi_1 r_1 x - \varphi_2 r_2 y + xy\,v\,(\mu_2\varphi_2 - \mu_1\varphi_1),$$

otra vez lineal en v. A lo largo de una trayectoria, integrar $x'/x = r_1 - \mu_1 yv$ e $y'/y = -r_2 + \mu_2 xv$ da expresiones exponenciales y conserva ambas poblaciones positivas. La función de conmutación es $\sigma = \mu_2\varphi_2 - \mu_1\varphi_1$, y se activa la interacción ($v = 1$) cuando $\sigma < 0$. La diferencia con el problema anterior está en el dato final del adjunto: como el coste vive en T, ya no es cero sino el gradiente del coste final, $\varphi(T) = \nabla\Psi = (1, 1)$, de donde $\sigma(T) = \mu_2 - \mu_1$. En el último tramo, por tanto, conviene dejar que las especies se encuentren si $\mu_2 < \mu_1$, es decir, si el encuentro quita más presas de las que produce depredadores, y conviene aislarlas si es al revés. Si $\mu_1 = \mu_2 = \mu$, el valor final no decide, pero las ecuaciones adjuntas dan $\sigma'(T) = \mu(r_1 + r_2) > 0$: justo antes de T, $\sigma < 0$ y conviene activar la interacción. Estas conclusiones describen el último tramo; no determinan por sí solas toda la historia óptima.

La idea: qué han añadido estos problemas

Estos modelos combinan controles que multiplican al estado, varios controles acotados, costes lineales y restricciones sobre el estado. Algunos encajan en variantes de la teoría anterior; las restricciones de estado requieren, además, multiplicadores que pueden ser medidas y adjuntos con saltos. En los costes lineales, el signo de la función de conmutación decide entre los extremos cuando no es cero. Se conserva el esquema: identificar estado, control, admisibles y coste; escribir el hamiltoniano y estudiar su mínimo en el control, junto con las condiciones adicionales que exijan las restricciones.

Las prácticas 24 a 26 usan otra vía para problemas pequeños: tratar el coste como una función de pocos parámetros y minimizarla directamente con un algoritmo genérico, sin adjunto. Es lo que se hace cuando el control es un número o unos pocos.

4.9 El problema inverso: identificar parámetros por mínimos cuadrados

LA PREGUNTA

En todo el capítulo los coeficientes de la ecuación de estado han venido dados y lo que se buscaba era una función $v(t)$. En la práctica pasa justo lo contrario: se tienen unas cuantas medidas de un experimento y no se conocen los coeficientes del modelo. ¿Es eso un problema de control óptimo, y si lo es, cuál es el control, qué se conserva de toda la teoría anterior y qué deja de valer?

El planteamiento. Sea el problema de Cauchy

$$\begin{cases} y' = f(t, y, p) & \text{en } (0, T), \\ y(0) = y_0(p), \end{cases} \tag{4.49}$$

donde ahora $p = (p_1, \dots, p_k) \in \mathcal{P}_{\text{ad}} \subset \mathbb{R}^k$ es un vector de **parámetros desconocidos** —constantes de reacción, tasas de contagio, un dato inicial que no se midió— y $\mathcal{P}_{\text{ad}}$ el conjunto de valores admisibles, casi siempre un producto de intervalos, porque una tasa no puede ser negativa. El dato inicial puede ser conocido, en cuyo caso y_0 es constante, o depender de los parámetros. Se supone que el estado existe de forma única hasta T para cada parámetro admisible. Se dispone de m observaciones en $0 < t_1 < \cdots < t_m \le T$,

$$z_i \approx C y(t_i; p), \qquad i = 1, \dots, m,$$

donde C es la matriz de observación, que dice qué componentes se miden —en una epidemia se puede que solo se registren casos con síntomas—. El problema inverso es

$$\text{Minimizar } J(p) := \sum_{i=1}^{m} \left| C y(t_i; p) - z_i \right|^2 \quad \text{sujeto a } p \in \mathcal{P}_{\text{ad}},\ y \text{ solución de (4.49)}, \tag{4.50}$$

y se llama **de mínimos cuadrados** por la forma del funcional. Se le dice *inverso* porque el problema *directo* es el de siempre —dados los parámetros, calcular la solución— y aquí se recorre la flecha al revés.

LA IDEA: SÍ ES UN PROBLEMA DE CONTROL, Y EN QUÉ SE DISTINGUE

La estructura de la sección 4.1 se conserva: una ecuación de estado, algo que se elige, un conjunto admisible y un coste. Cambian tres aspectos:

- **El control es finito-dimensional.** Se busca en $\mathbb{R}^k$. Si $\mathcal{P}_{\text{ad}}$ es compacto y $p \mapsto y(\cdot; p)$ es continua en $C([0, T])$, Weierstrass asegura que el mínimo se alcanza. La dependencia continua se obtiene bajo las hipótesis habituales de la EDO, siempre que las

soluciones existan en el intervalo común $[0, T]$; la compacidad de los parámetros no evita por sí sola una explosión del estado.

- **Aquí no se ha añadido regularización.** Podría incorporarse, por ejemplo, $\frac{\mu}{2}|p - p_{\text{previo}}|^2$ para penalizar alejarse de una estimación previa. En un problema no lineal, ese término no garantiza por sí solo convexidad ni unicidad.
- **El coste puede no ser convexo.** Incluso en $y' = py$, la solución $y_0 e^{pt}$ depende de forma no lineal del parámetro. Por eso pueden aparecer varios mínimos locales y dependencia del arranque numérico. La unicidad del caso lineal cuadrático anterior se apoyaba en que el estado dependía de forma afín del control; tampoco estaba garantizada para todos los problemas no lineales del capítulo.

El caso más simple, a mano: una tasa de crecimiento. Antes de la maquinaria general conviene resolver el problema con un parámetro. Sean $y' = py$, $y(0) = y_0 > 0$ conocido, y p la tasa que se quiere identificar; el estado es $y(t; p) = y_0 e^{pt}$.

- *Una observación $z_1 > 0$ en t_1.* El coste $J(p) = (y_0 e^{pt_1} - z_1)^2$ se anula en $p = \frac{1}{t_1} \log \frac{z_1}{y_0}$, y solo allí, porque e^{pt_1} es estrictamente creciente en p. Un dato determina un parámetro, y la fórmula dice cómo: la tasa es el logaritmo del factor de crecimiento observado, por unidad de tiempo.
- *Varias observaciones.* Ya no hay fórmula cerrada. La tentación es tomar logaritmos: $\log z_i \approx \log y_0 + pt_i$ es una recta en t_i de ordenada conocida, y su ajuste por mínimos cuadrados es la regresión del ejercicio de la recta de regresión del capítulo 1, que da $p_{\log} = \sum t_i \log(z_i/y_0) / \sum t_i^2$. Pero eso minimiza $\sum (pt_i - \log(z_i/y_0))^2$, que mide errores *relativos*, y no (4.50), que mide errores absolutos. Con $y_0 = 1$, $t_i = 1, 2, 3$ y datos $z_i = 3$, 7, 21 (próximos a los de $p = 1$, que serían 2.72, 7.39 y 20.09), el ajuste logarítmico da $p_{\log} \approx 1.0089$ y el mínimo de (4.50) está en $p \approx 1.0128$, con $J = 0.4145$ frente a $J(p_{\log}) = 0.4775$. El segundo se fija más en la observación mayor: su residuo en $t = 3$ es -0.13, frente a -0.37. Ninguno está mal; responden a preguntas distintas, y el enunciado dice cuál es la que se plantea.
- *La forma del coste.* J no es convexa: $J''(0) = -384 < 0$. Para p muy negativo el modelo predice casi cero en todos los instantes, J se acerca a $\sum z_i^2 = 499$ y es casi plana: en $p = -2$ la derivada vale -1.6, y un método de descenso que arranque allí apenas se mueve. El ajuste logarítmico, que es una línea de cálculo, da un buen punto de partida.

El gradiente, y por qué el adjunto sigue siendo la herramienta. Escribiendo la versión con observación distribuida, $J(p) = \frac{1}{2} \int_0^T |Cy - z|^2 \, dt$, el cálculo sigue el argumento de la sección 4.3. Se suponen f y y_0 de clase C^1 en las variables que se derivan. Derivando (4.49) respecto de p_j, la **sensibilidad** $w_j = \partial y / \partial p_j$ resuelve

$$w_j' = \frac{\partial f}{\partial y}(t, y, p)\, w_j + \frac{\partial f}{\partial p_j}(t, y, p), \qquad w_j(0) = \frac{\partial y_0}{\partial p_j}(p),$$

y $J'(p)_j = \int_0^T C^t (Cy - z) \cdot w_j$. Introduciendo el adjunto

$$-\varphi' = \Big(\frac{\partial f}{\partial y}\Big)^t \varphi + C^t (Cy - z), \qquad \varphi(T) = 0,$$

y derivando el producto $\varphi \cdot w_j$,

$$\frac{d}{dt}(\varphi \cdot w_j) = -C^t(Cy - z) \cdot w_j + \varphi \cdot \frac{\partial f}{\partial p_j},$$

que integrado entre 0 y T, con $\varphi(T) = 0$, da

$$\boxed{\frac{\partial J}{\partial p_j}(p) = \varphi(0) \cdot \frac{\partial y_0}{\partial p_j}(p) + \int_0^T \varphi \cdot \frac{\partial f}{\partial p_j}(t, y, p)\, dt}, \quad j = 1, \dots, k.$$

Si el dato inicial es fijo, el término de borde desaparece. Si el parámetro es justamente un dato inicial y f no depende de él, toda su derivada se obtiene de $\varphi(0)$.

Lectura de la fórmula, y una comprobación a mano. Los dos sumandos siguen la misma pauta: precio del estado por cuánto mueve el parámetro al estado. $\varphi(0)$ es el precio del estado inicial y $\partial y_0/\partial p_j$ dice cuánto lo cambia el parámetro; $\varphi(t)$ es el precio del estado en el instante t y $\partial f/\partial p_j$ dice cuánto empuja el parámetro a la ecuación en ese instante. La integral suma los empujones de todos los instantes, cada uno pagado a su precio. Se comprueba en el caso más simple, $y' = py$ con $y(0) = 1$ y $z = 0$, donde todo se puede hacer a mano. Directamente, $J(p) = \frac{1}{2}\int_0^T e^{2pt}\, dt$ y $J'(p) = \int_0^T te^{2pt}\, dt$. Por el adjunto: aquí $\partial f/\partial y = p$ y $\partial f/\partial p = y = e^{pt}$, la adjunta es $-\varphi' = p\varphi + e^{pt}$ con $\varphi(T) = 0$, y su solución es $\varphi(t) = e^{-pt}\int_t^T e^{2ps}\, ds$. La fórmula recuadrada da

$$J'(p) = \int_0^T \varphi(t)\, e^{pt}\, dt = \int_0^T \int_t^T e^{2ps}\, ds\, dt = \int_0^T s\, e^{2ps}\, ds,$$

porque al cambiar el orden de integración cada s aparece para los $t \in [0, s]$, un intervalo de longitud s. Es lo mismo que la cuenta directa. El adjunto no ha hecho nada que no se pudiera hacer a mano; lo que aporta es que con k parámetros hace las k derivadas con una sola integración hacia atrás.

LA IDEA: UN ADJUNTO PARA TODAS LAS DERIVADAS

El adjunto φ no depende de j: se integra el estado una vez y el adjunto una vez; después se calculan las k componentes del gradiente con las integrales anteriores. Las diferencias finitas hacia delante requieren, en cambio, $k + 1$ integraciones del estado y un paso de perturbación que introduce error. El ahorro aumenta cuando hay muchos parámetros, aunque evaluar las k integrales también tiene coste. Es la misma idea de propagar sensibilidades hacia atrás que aparece en asimilación de datos y en la retropropagación de redes neuronales, con sus correspondientes ecuaciones o discretizaciones.

Con observaciones discretas, el adjunto satisface $-\varphi' = f_y^{\mathsf{T}}\varphi$ entre las medidas y, al recorrer el tiempo hacia atrás, recibe los saltos

$$\varphi(t_i^-) = \varphi(t_i^+) + 2C^t\big(Cy(t_i; p) - z_i\big).$$

Se empieza con cero después de la última observación. Si hay una medida en T, se aplica allí el salto: $\varphi(T^-) = 2C^t(Cy(T; p) - z_m)$; si no la hay, $\varphi(T) = 0$. Con este adjunto, la fórmula del gradiente mantiene el término inicial y la integral. El factor 2 procede de que (4.50) no lleva $1/2$: al añadirlo, tanto los saltos como el gradiente se

reducen a la mitad, sin cambiar los minimizadores.

Identificabilidad: la pregunta que hay que hacerse antes de calcular. Que el mínimo exista no significa que sirva. Lo que hace falta es que la aplicación

$$p \longmapsto \big(Cy(t_1;p),\dots,Cy(t_m;p)\big)$$

sea **inyectiva**, al menos localmente: si dos juegos de parámetros distintos producen las mismas observaciones, ningún método los va a distinguir, y el problema está mal planteado por mucho que el ordenador devuelva un número. Una comprobación local utiliza la matriz de sensibilidades: si cada observación tiene q componentes, S tiene mq filas y k columnas, agrupadas en bloques $S_{ij} = \partial(Cy(t_i;p))/\partial p_j \in \mathbb{R}^q$. Bajo diferenciabilidad, que S tenga rango k es una condición suficiente de inyectividad local, no necesaria. Si existe una dirección d con $|Sd|$ muy pequeño frente a $|d|$, variar los parámetros en esa dirección apenas cambia las predicciones a primer orden. En torno a un buen ajuste, el coste puede presentar un valle estrecho: el ruido o las tolerancias numéricas pueden alterar mucho los parámetros estimados sin cambiar apenas el residuo. Esto es mala condición, y no implica necesariamente que haya varios mínimos exactos.

LA IDEA: DÓNDE OBSERVAR ES PARTE DEL MODELADO

La identificabilidad depende del modelo, de las componentes observadas y de cuándo se miden. En el ejercicio de identificación de A y B, si la solución tiende a un equilibrio positivo, el forzamiento $e^{-t}\cos(2\pi t)$ se extingue y el límite satisface $Ay - By^2 = 0$: el nivel final revela A/B. Observar solo esa meseta límite no separa A de B. A tiempos grandes pero finitos puede quedar información sobre ambos, aunque sea demasiado pequeña frente al ruido. Incluir el transitorio ayuda a distinguir el ritmo de cambio del nivel final. Elegir los instantes de medida forma parte del problema de modelado.

Cuántas observaciones, y qué mide el mínimo. Hay mq ecuaciones escalares para k parámetros; si $mq > k$, el sistema está sobredeterminado y en general los datos no se ajustan exactamente. Con datos sintéticos generados por un parámetro admisible, el mínimo teórico es cero. Recuperar los parámetros conocidos es una comprobación útil cuando las observaciones los identifican; también hay que comprobar el residuo y las tolerancias del integrador y del optimizador.

Con datos reales suele quedar un residuo, que combina ruido, discrepancia del modelo y posible error de optimización. Para observaciones escalares, $\sqrt{J_{\text{mín}}/m}$ es la raíz del error cuadrático medio, en las unidades de la medida; se puede comparar con su incertidumbre. Si las componentes tienen escalas o precisiones distintas, conviene ponderar los residuos. Un error grande invita a revisar datos, ajuste y modelo; por sí solo no identifica al culpable.

EL CAMINO QUE NO SE TOMÓ: MINIMIZAR SIN DERIVADAS O CON ELLAS

La alternativa. Gauss–Newton aprovecha la suma de cuadrados: linealiza el residuo y resuelve un problema lineal de mínimos cuadrados en cada paso. Puede converger cuadráticamente cerca de una solución con residuo nulo y jacobiano de rango completo, bajo regularidad suficiente; no es una garantía para cualquier ajuste.
Su ventaja. Usar el jacobiano suele ahorrar evaluaciones y resulta especialmente útil con

muchos parámetros. `lsqnonlin` ofrece algoritmos para mínimos cuadrados no lineales, como región de confianza y Levenberg–Marquardt; el jacobiano puede suministrarse o aproximarse numéricamente.
Qué la aparta aquí. `lsqnonlin` requiere la *Optimization Toolbox*. Las prácticas usan `fminsearch`, disponible en MATLAB básico: aplica Nelder–Mead sin derivadas y permite practicar con pocos parámetros. La sencillez no garantiza convergencia a un mínimo, ni siquiera local. Conviene comparar arranques, comprobar el residuo y verificar que las tolerancias de la EDO permiten resolver las diferencias de coste.

Un modelo donde todo esto se ve: el SEAIR. Es una ampliación del modelo SIR del capítulo 2 con dos compartimentos más. La población se reparte en susceptibles S, infectados en una etapa inicial E —aún no clasificados como asintomáticos o sintomáticos— , infectados asintomáticos A, infectados con síntomas I y recuperados R, y los contagios se producen por cualquiera de las tres interacciones (S,E), (S,A) y (S,I):

$$\begin{cases} S' = -\beta_1 SE - \beta_2 SA - \beta_3 SI, \\ E' = \beta_1 SE + \beta_2 SA + \beta_3 SI - (\lambda_1 + \lambda_2)E, \\ A' = \lambda_1 E - \delta_1 A, \\ I' = \lambda_2 E - \delta_2 I, \\ R' = \delta_1 A + \delta_2 I, \end{cases} \tag{4.51}$$

con los coeficientes en $[0,1]$, según las unidades y cotas del ejercicio, y datos iniciales no negativos. Sumando las cinco ecuaciones sale $N' = 0$: el sistema es cerrado y la población total se conserva, como en el SIR de la sección 2.2.1 y al contrario que en el de la sección 2.2.2. Cada coeficiente tiene su lectura: λ_1 y λ_2 reparten a los incubados entre asintomáticos y sintomáticos, de modo que, si $\lambda_1 + \lambda_2 > 0$, $1/(\lambda_1 + \lambda_2)$ es el tiempo medio en E y $\lambda_2/(\lambda_1 + \lambda_2)$ la fracción de casos que dan la cara; δ_1 y δ_2 son las tasas de recuperación de cada tipo, y sus inversas las duraciones medias cuando esas tasas son positivas. El término $\beta_1 SE$ supone que quienes están en E ya pueden contagiar.

Supongamos que solo se dispone de observaciones de I; los casos asintomáticos también podrían medirse mediante pruebas, pero aquí no se incluyen esos datos. El ejercicio plantea: dados $(\beta_2, \beta_3, \lambda_1, \lambda_2, \delta_1)$, hallar (β_1, δ_2) de modo que la I del modelo se ajuste a las observaciones. Es (4.50) con $k = 2$, $Cy = I$ y $\mathcal{P}_{\text{ad}} = [0,1]^2$, y puede abordarse con `fminsearch`, controlando explícitamente las cotas porque este método no las impone; la M-función completa, con su subfunción de simulación, está resuelta en el apéndice C.

LA IDEA: QUÉ PUEDE DECIR LA CURVA SOBRE LOS DOS PARÁMETROS

β_1 afecta al contagio desde E y δ_2 a la salida de I, pero ambos influyen en la forma completa de la curva. Cuando $\lambda_2 E \approx \delta_2 I$, la población sintomática cambia lentamente; eso puede producir una meseta transitoria, no un equilibrio endémico positivo de este sistema cerrado con tasas de recuperación positivas.
Estas lecturas ayudan a elegir datos, pero no prueban que los dos parámetros sean identificables. Por ejemplo, si $E(0) = A(0) = I(0) = 0$, entonces $I(t) = 0$ para todos sus valores: esa observación no distingue ninguno. Con una epidemia en marcha, se estudian las sensibilidades de $I(t_i)$ respecto de ambos parámetros. Tampoco β_1 y β_2 son indistinguibles por el mero hecho de que ambos favorezcan contagios: multiplican

a E y a A, que pueden evolucionar de forma diferente. Es la información concreta de las observaciones la que debe permitir separarlos.

4.10 Prácticas de MATLAB del tema 4

LA PREGUNTA

Hay dos maneras de calcular un control óptimo con un ordenador: tratar el coste como una función de unos pocos números y minimizarla a ciegas, probando valores, o resolver el sistema de optimalidad como un problema de contorno. La hoja del tema 4 hace las dos. ¿Cuándo conviene cada una, y cómo se comprueba que el resultado es el control óptimo y no otro?

Esta sección resuelve los ejercicios 24 a 28 de la hoja del tema 4. Los tres primeros son **problemas inversos** de la sección 4.9: identificar un coeficiente o un dato inicial a partir de observaciones, minimizando una función de una o dos variables con los minimizadores genéricos de MATLAB, que se presentan en el 24. Los dos últimos resuelven el sistema de optimalidad con `bvp4c`, que ya se conoce del tema 3. La hoja viene acompañada de `ControlEDO.m`, `ControlEDOr.m` y `ControlOrden2.m`; los demás ficheros se escriben aquí.

Los minimizadores `fminbnd` y `fminsearch`

`fminbnd(J, a, b)` busca un mínimo local de una función escalar `J` de *una* variable en el intervalo $[a, b]$, sin derivadas, combinando búsqueda áurea e interpolación parabólica. `fminsearch(J, x0)` minimiza una función de *varias* variables (que recibe un vector) partiendo de `x0`, también sin derivadas, con el método del símplex de Nelder-Mead: mueve un triángulo (en dos variables) por el paisaje de J hasta reducir el tamaño del símplex y las diferencias de coste. Los dos devuelven el punto y, como segunda salida, el valor: `[x, Jx] = fminsearch(J, x0)`. Las opciones se fijan con `optimset`: `optimset('Display','iter','TolX',1e-8)` muestra cada iteración y afina la tolerancia. Como no usan derivadas, sirven aunque `J` se calcule integrando una EDO, que es el caso de las tres prácticas. Su carácter local no significa que encuentren el mínimo más cercano al arranque ni asegura el global. Se revisan el estado de salida, el residuo y la dependencia del punto de partida; en `fminbnd` se comprueban además los extremos del intervalo si pueden contener el mínimo.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
from scipy.optimize import minimize_scalar, minimize
# fminbnd(J, a, b) -> minimize_scalar(J, bounds=(a, b), method="bounded").x
# fminsearch(J, x0) -> minimize(J, x0, method="Nelder-Mead").x (J recibe un array)
```

Ejercicio 24: identificar un coeficiente de una ecuación de Ricatti

Ejercicio 4.6. *Problema inverso para la ecuación de Ricatti*

ORIGEN: hoja de ejercicios del tema 4, ejercicio 24. No hay fichero de partida; se escribe aquí.

Se considera el problema de Cauchy

$$\begin{cases} y' = e^t \cos t + Ay - By^2 & \text{en } (0,T), \\ y(0) = y_0. \end{cases}$$

Conocidos y_0 y A, se desea calcular $B > 0$ de forma que $y(T) = y_{obs}$, un valor observado. Se plantea

$$\text{Minimizar } J(B) := \big(y(T) - y_{obs}\big)^2 \quad \text{sujeto a } B \in [0,+\infty),\ y \text{ solución de } (\mathrm{P})_B.$$

Escribir `[B] = InvRicatti(ycero, A, Tfin, yobs, B0)` que calcule B con B_0 como aproximación inicial. Comprobación: para $y_0 = 0.5$, $A = 0.1$, $B = 0.5$, es $y(2.5) \approx -2.838971$.

Resolución.
La idea. Cada valor de B define una EDO; resolverla hasta T permite evaluar $J(B)$. Al haber un solo parámetro, se puede practicar con un minimizador sin derivadas. El adjunto también sería posible, pero su ahorro resulta mayor con muchos parámetros. Después de discretizar un control funcional se puede optimizar igualmente; lo costoso sería explorar una rejilla de todas las combinaciones, no cualquier método directo.
El núcleo del programa es convertir «integrar la EDO y comparar su valor final» en una función escalar. Hay dos comprobaciones inseparables de esa evaluación: que $B \geq 0$ y que la solución llegue a T. Una Ricatti puede explotar en tiempo finito; entonces $y(T)$ no existe y el último valor calculado no lo sustituye.
Órdenes nuevas. Con una salida, `sol = ode45(f, [0, T], y0)` devuelve una estructura. `deval(sol, T)` usa la interpolación continua del propio integrador; no exige que T coincida con un nodo interno, pero debe estar en el intervalo resuelto. `sol.x(end)` permite comprobar hasta dónde ha llegado. El código devuelve `Inf` para rechazar candidatos inválidos y exige un arranque con coste finito; es una protección de la evaluación, no una garantía de convergencia del optimizador.

```
function [B] = InvRicatti(ycero, A, Tfin, yobs, B0)
% Ajusta B >= 0. Tfin debe ser positivo y B0 debe dar un estado hasta Tfin.
assert(isscalar(Tfin) && isfinite(Tfin) && Tfin > 0)
assert(isscalar(B0) && isfinite(B0) && B0 >= 0)
optEDO = odeset('RelTol',1e-10,'AbsTol',1e-12);
if ~isfinite(coste(B0))
    error('El estado del arranque no llega a Tfin; cambia B0.')
end
opciones = optimset('Display','iter','TolX',1e-8,'TolFun',1e-14);
[B, JB, flag] = fminsearch(@coste, B0, opciones);
if flag <= 0 || ~isfinite(JB)
    error('El ajuste no ha terminado correctamente.')
end
fprintf('B = %.6f, J(B) = %.3e\n', B, JB)

    function valor = coste(b)
        valor = Inf;
        if ~isfinite(b) || b < 0, return, end
        sol = ode45(@(t,y) exp(t)*cos(t)+A*y-b*y.^2, ...
                    [0,Tfin], ycero, optEDO);
        if sol.x(end) < Tfin || any(~isfinite(sol.y(:))), return, end
```

```
        valor = (deval(sol,Tfin)-yobs)^2;
    end
end
```

Comprobación con datos conocidos. Se llama *experimento gemelo* a generar observaciones con un parámetro conocido y tratar de recuperarlo. Es una comprobación útil si los datos identifican el parámetro; puede complementarse con sensibilidades o con otro método de integración. Para estos datos, una integración de mayor precisión da $y(2.5) \approx -2.839028519$ cuando $B = 0.5$. Ajustar el valor -2.838971 de la hoja da, en cambio, $B \approx 0.499978521$. La discrepancia del dato, unas $5.75 \cdot 10^{-5}$ unidades, explica esa diferencia; no conocemos las tolerancias con que se obtuvo el número de la hoja.
Aquí se puede cuantificar la sensibilidad sin adivinarla. Derivando la EDO respecto de B,

$$w' = (A - 2By)w - y^2, \qquad w(0) = 0, \qquad w = \frac{\partial y}{\partial B}.$$

La fórmula del factor integrante da $w(T) < 0$ si y no es idénticamente cero en $[0, T]$. Por tanto, mientras los estados existan, $y(T; B)$ decrece estrictamente en este ejemplo y puede alcanzar un valor observado a lo sumo una vez. Cerca de $B = 0.5$, $w(T) \approx -2.677868$, de modo que un error pequeño del dato produce $\Delta B \approx \Delta y_{obs}/w(T)$: aquí unos $-2.15 \cdot 10^{-5}$. En otros problemas, un $|w(T)|$ pequeño puede amplificar mucho el error. Conviene ajustar tanto las tolerancias del integrador como las del minimizador.
La figura 4.2 muestra el coste únicamente donde la solución llega a T. Para estos datos, el umbral numérico es $B_* \approx 1.947949$: al acercarse desde abajo, $y(T)$ tiende a $-\infty$. A partir del umbral, la solución explota en T o antes y $J(B)$ ya no está definido. La zona sombreada no representa un coste grande.

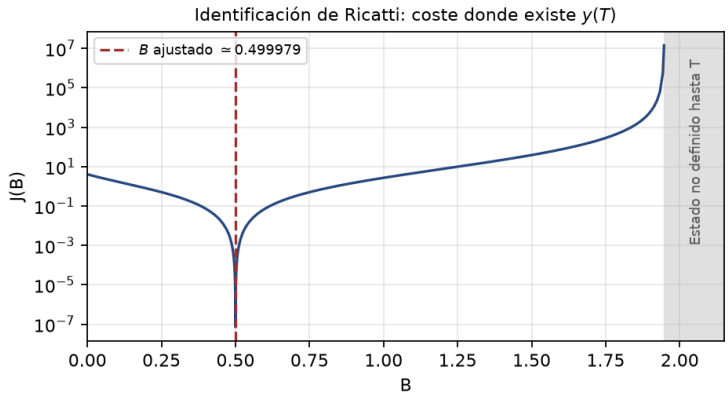


Figura 4.2: Coste de identificación de Ricatti para $y_0 = 0.5$, $A = 0.1$, $T = 2.5$ y $y_{obs} = -2.838971$, en escala logarítmica. La línea discontinua señala el parámetro ajustado, $B \approx 0.499979$. El coste crece sin cota al aproximarse a B_* por la izquierda; en la zona gris no hay solución de estado hasta T.

Una alternativa escalar es `fminbnd` sobre un intervalo donde el estado exista, por ejemplo $[0, 1.9]$ para estos datos. No basta darle $[0, 5]$ y confiar en que evite la parte inválida: el método no conoce el dominio de la EDO. Si se inicia el código anterior con $B_0 = 2$, debe rechazar el arranque en lugar de anunciar un ajuste basado en una integración incompleta.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_ivp
from scipy.optimize import minimize

def inv_ricatti(ycero, A, Tfin, yobs, B0):
    if not np.isfinite(Tfin) or Tfin <= 0:
        raise ValueError("Tfin debe ser positivo")
    def coste(b):
        B = b[0]
        if not np.isfinite(B) or B < 0:
            return np.inf
        sol = solve_ivp(lambda t, y: np.exp(t)*np.cos(t)+A*y-B*y*y,
                        [0, Tfin], [ycero], rtol=1e-10, atol=1e-12)
        if not sol.success or sol.t[-1] < Tfin:
            return np.inf
        return (sol.y[0, -1]-yobs)**2
    if not np.isfinite(coste([B0])):
        raise ValueError("El estado del arranque no llega a Tfin")
    r = minimize(coste, [B0], method="Nelder-Mead",
                 options={"xatol": 1e-10, "fatol": 1e-14})
    if not r.success or not np.isfinite(r.fun):
        raise RuntimeError(r.message)
    print("B =", r.x[0], "J =", r.fun)
    return r.x[0]

inv_ricatti(0.5, 0.1, 2.5, -2.838971, 1.0) # B aproximadamente 0.499979
```

Ejercicio 25: dos coeficientes a partir de una serie de observaciones

Ejercicio 4.7. *Identificar A y B con varias observaciones*

ORIGEN: hoja de ejercicios del tema 4, ejercicio 25. No hay fichero de partida; se escribe aquí.

En $y' = e^{-t}\cos(2\pi t) + Ay - By^2$ los coeficientes A y B son desconocidos, y se conocen los valores $y(t_i) = y_i$, $i = 0, \ldots, m$. Escribir `[Coefs] = InvRicatti2(tobs, yobs, coefs0)` que calcule A y B desde una aproximación inicial, y un programa auxiliar que genere datos `tobs`, `yobs` correspondientes de verdad a una solución.

Resolución.
La idea. Se busca el par (A, B) que ajusta la trayectoria a los puntos medidos: $J(A, B) = \sum_i |y(t_i; A, B) - y_i|^2$. Es una regresión no lineal en la que cada evaluación requiere integrar una EDO. Aquí se usa Nelder–Mead; también podrían emplearse las sensibilidades o el adjunto de la sección 4.9.
Se comprueban las mismas cuestiones que en la práctica anterior: el arranque, el residuo, las tolerancias y que cada integración llegue al último tiempo observado. Un residuo mayor del esperado puede deberse a un mínimo local, a tolerancias insuficientes o a un error del programa; su valor por sí solo no distingue esas causas. El enunciado toma $y(t_0) = y_0$ como dato inicial exacto. Si esa primera observación también tuviera ruido, podría tratarse el dato

inicial como otro parámetro que ajustar.

Paso 1: la orden que hace posible comparar con las observaciones. Si el segundo argumento de `ode45` es un vector con más de dos componentes, la solución se devuelve exactamente en esos instantes, aunque internamente use otros pasos. Es lo que hace falta para comparar con observaciones.

```
function [Coefs] = InvRicatti2(tobs, yobs, coefs0)
% Ajusta A y B usando y(tobs(1)) = yobs(1) como dato inicial exacto.
tobs = tobs(:).'; yobs = yobs(:).';
assert(numel(tobs) >= 2 && numel(tobs) == numel(yobs))
assert(all(isfinite([tobs,yobs])) && all(diff(tobs) > 0))
assert(numel(coefs0) == 2 && all(isfinite(coefs0(:))))
optEDO = odeset('RelTol',1e-10,'AbsTol',1e-12,'MaxStep',0.05);
if ~isfinite(coste(coefs0)), error('El estado del arranque no llega al final.'), end
[Coefs, JC, flag] = fminsearch(@coste, coefs0, ...
                       optimset('TolX',1e-8,'TolFun',1e-14));
if flag <= 0 || ~isfinite(JC), error('El ajuste no ha terminado correctamente.'), end
fprintf('A = %.6f, B = %.6f, J = %.3e\n', Coefs(1), Coefs(2), JC)

    function valor = coste(c)
        valor = Inf;
        if any(~isfinite(c)), return, end
        f = @(t,y) exp(-t)*cos(2*pi*t)+c(1)*y-c(2)*y.^2;
        sol = ode45(f,[tobs(1),tobs(end)],yobs(1),optEDO);
        if sol.x(end) < tobs(end) || any(~isfinite(sol.y(:))), return, end
        residuo = deval(sol,tobs)-yobs;
        valor = sum(residuo.^2);
    end
end
```

El programa auxiliar genera datos sintéticos con coeficientes conocidos:

```
% DatosRicatti2: genera observaciones con A = 0.3, B = 0.8, y(0) = 1, y llama a
    InvRicatti2
Aex = 0.3; Bex = 0.8;
tobs = 0:0.5:5;
f = @(t,y) exp(-t).*cos(2*pi*t) + Aex*y - Bex*y.^2;
optEDO = odeset('RelTol',1e-10,'AbsTol',1e-12,'MaxStep',0.05);
[~, yobs] = ode45(f, tobs, 1, optEDO); % aproximaciones en los tiempos tobs
yobs = yobs'; % en fila, como tobs
Coefs = InvRicatti2(tobs, yobs, [1, 1]);
```

Los primeros valores aproximados son 1, 0.8055, 0.6914, 0.6214, 0.5665 y el último es 0.4349. Con estos datos y el arranque $(1, 1)$ se recuperan aproximadamente $(0.3, 0.8)$; las últimas cifras dependen de las tolerancias y del algoritmo.

Hay once observaciones y dos coeficientes, aunque la primera observación se usa también como condición inicial y su residuo es siempre cero. El mínimo teórico de los datos exactos generados por el modelo es cero; numéricamente debe compararse con la precisión usada al generar y ajustar los datos. Si se observa casi exclusivamente el régimen final, las sensibilidades respecto de A y B pueden hacerse casi dependientes: se estima mejor el cociente A/B que cada coeficiente por separado, como se explicó en la sección 4.9.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
```

```
from scipy.integrate import solve_ivp
from scipy.optimize import minimize

def ysim(c, y0, tobs):
    if not np.all(np.isfinite(c)):
        return None
    f = lambda t,y: np.exp(-t)*np.cos(2*np.pi*t)+c[0]*y-c[1]*y*y
    sol = solve_ivp(f, [tobs[0],tobs[-1]], [y0], t_eval=tobs,
                    rtol=1e-10, atol=1e-12, max_step=0.05)
    if not sol.success or len(sol.t) != len(tobs):
        return None
    return sol.y[0]

def inv_ricatti2(tobs, yobs, coefs0):
    tobs = np.asarray(tobs, dtype=float).ravel()
    yobs = np.asarray(yobs, dtype=float).ravel()
    if (len(tobs) < 2 or len(tobs) != len(yobs)
            or not np.all(np.isfinite(np.r_[tobs,yobs]))
            or not np.all(np.diff(tobs) > 0)):
        raise ValueError("Datos finitos, de igual longitud y tiempos crecientes")
    def coste(c):
        y = ysim(c, yobs[0], tobs)
        return np.inf if y is None else np.sum((y-yobs)**2)
    if not np.isfinite(coste(coefs0)):
        raise ValueError("El estado del arranque no llega al final")
    r = minimize(coste, coefs0, method="Nelder-Mead",
                 options={"xatol":1e-8,"fatol":1e-14})
    if not r.success or not np.isfinite(r.fun):
        raise RuntimeError(r.message)
    print("A, B =", r.x, "J =", r.fun)
    return r.x

tobs = np.linspace(0, 5, 11)
yobs = ysim([0.3, 0.8], 1.0, tobs)
coefs = inv_ricatti2(tobs, yobs, [1.0, 1.0])
```

Qué hemos aprendido y dónde se falla. Un problema inverso con una EDO se programa como una función de coste que integra y compara, y un minimizador genérico que la llama. Lo delicado no es el minimizador, sino lo que la función de coste hace con la EDO. Los fallos, medidos con la versión de Python y los mismos datos:

- *Tolerancias del integrador.* Con las tolerancias por defecto (relativa 10^{-3}, la misma que `ode45` por defecto), el ajuste devuelve $A \approx 0.2961$ y $B \approx 0.7960$ con $J \approx 2 \cdot 10^{-5}$: el error del integrador se confunde con un error del modelo y se reparte entre los parámetros. Las tolerancias del ajuste tienen que ser más finas que la precisión que se espera de los parámetros.
- *El dato inicial tomado como exacto.* Si la primera observación trae un error de 0.01 (un uno por ciento), el ajuste da $A \approx 0.312$ y $B \approx 0.822$, errores del cuatro y del tres por ciento, con $J \approx 6 \cdot 10^{-6}$. El residuo parece excelente y no avisa. Si hay motivo para desconfiar de $y(t_0)$, se añade como tercer parámetro.
- *Un arranque que hace explotar el estado.* Con $B < 0$, el término $-By^2$ empuja hacia arriba y la solución explota: con $(A, B) = (1, -1)$ la integración se detiene cerca de $t = 0.65$. La comprobación de que la integración llega al último instante es la que impide evaluar el

coste con un estado que no existe; con ella el programa se niega y hay que cambiar el arranque. Desde arranques razonables, como $(0,0)$, $(2,5)$ o $(0.3,3)$, se llega a $(0.3,0.8)$.
- *Observar solo el régimen final.* Con ruido de desviación 0.002 y treinta repeticiones, las once observaciones en $[0,5]$ dan desviaciones de 0.004 en A y 0.006 en B. Con diez observaciones en $[3,5]$ (más el dato inicial) pasan a 0.013 y 0.024, mientras que el cociente A/B se sigue estimando con un error relativo del uno y medio por ciento. Es, con números, la meseta que solo revela A/B de la sección 4.9.

QUÉ DEJA. El mismo esquema sirve para varios coeficientes: integrar, evaluar los residuos y minimizar su suma de cuadrados. El experimento con datos conocidos comprueba la implementación; la información de las observaciones y las tolerancias determinan qué precisión se puede esperar en los parámetros.

Ejercicio 26: recuperar el dato inicial de Lotka-Volterra

Ejercicio 4.8. *Problema inverso para el dato inicial de un sistema presa-depredador*

ORIGEN: hoja de ejercicios del tema 4, ejercicio 26. No hay fichero de partida; se escribe aquí.

Se considera el problema de Lotka-Volterra

$$u' = 5u - 3uv, \qquad v' = -v + 2uv, \qquad u(0) = u_0,\ v(0) = v_0,$$

con dato inicial desconocido, y se conocen los valores $u(t_i) = u_i$, $i = 1, \dots, m$, con $t_i > 0$. Escribir `[usol, vsol] = InvLotkaVolterra(tobs, uobs, u0, v0)` que calcule (u_0, v_0) minimizando $J(u_0, v_0) = \sum_i (u(t_i) - u_i)^2$ sobre $u_0, v_0 \geq 0$, desde la aproximación (u_0, v_0) dada.

Resolución.
La idea. Ahora las incógnitas son las condiciones iniciales. La evaluación del coste sigue el mismo esquema: integrar desde cada candidato (u_0, v_0) y comparar las presas calculadas con las observaciones. Se conserva la restricción $u_0, v_0 \geq 0$.
¿Cuánta información contiene observar solo las presas? Si se conociera la curva continua $u(t) > 0$ sin error, la primera ecuación permitiría reconstruir

$$v(t) = \frac{1}{3}\left(5 - \frac{u'(t)}{u(t)}\right).$$

Conocer ambas poblaciones en un instante determina la trayectoria por unicidad. Pero esta fórmula no convierte unas pocas observaciones ruidosas en una curva y su derivada exactas. Los instantes deben aportar información suficiente para distinguir los dos datos iniciales; la periodicidad puede crear ambigüedades de fase con un muestreo escaso. Además, si $u_0 = 0$, entonces $u(t) = 0$ para cualquier v_0: observar solo las presas no permite recuperar los depredadores en ese caso.
En el experimento siguiente se toman doce observaciones de una trayectoria positiva. El ajuste se comprueba recuperando el dato inicial usado para generarlas; para datos nuevos conviene examinar también las sensibilidades, como en la sección 4.9.

```
function [usol, vsol] = InvLotkaVolterra(tobs, uobs, u0, v0)
tobs = tobs(:).'; uobs = uobs(:).';
assert(numel(tobs) >= 1 && numel(tobs) == numel(uobs))
assert(all(isfinite([tobs,uobs])) && all(tobs > 0) && all(diff(tobs) > 0))
f = @(t,y) [5*y(1)-3*y(1)*y(2); -y(2)+2*y(1)*y(2)];
optEDO = odeset('RelTol',1e-10,'AbsTol',1e-12,'MaxStep',0.01);
if ~isfinite(coste([u0,v0])), error('El arranque no es admisible.'), end
[c, Jc, flag] = fminsearch(@coste,[u0,v0], ...
                      optimset('TolX',1e-8,'TolFun',1e-14));
if flag <= 0 || ~isfinite(Jc), error('El ajuste no ha terminado correctamente.'), end
usol = c(1); vsol = c(2);
fprintf('u0 = %.6f, v0 = %.6f, J = %.3e\n',usol,vsol,Jc)
[t,y] = ode45(f,[0,tobs(end)],[usol;vsol],optEDO);
plot(t,y(:,1),'b',t,y(:,2),'r--',tobs,uobs,'ko','LineWidth',1.2)
legend('u(t) recuperada','v(t)','observaciones de u'), xlabel('t'), shg

    function valor = coste(c)
        valor = Inf;
        if any(~isfinite(c)) || any(c < 0), return, end
        sol = ode45(f,[0,tobs(end)],c(:),optEDO);
        if sol.x(end) < tobs(end) || any(~isfinite(sol.y(:))), return, end
        valor = sum((deval(sol,tobs,1)-uobs).^2);
    end
end
```

El programa auxiliar genera los datos con $u_0 = 1$ y $v_0 = 0.8$:

```
% DatosLotka: genera observaciones de u con (u0, v0) = (1, 0.8) y llama a
    InvLotkaVolterra
f = @(t,y) [5*y(1) - 3*y(1)*y(2); -y(2) + 2*y(1)*y(2)];
tobs = 0.25:0.25:3;
optEDO = odeset('RelTol',1e-10,'AbsTol',1e-12,'MaxStep',0.01);
uobs = deval(ode45(f, [0, 3], [1; 0.8], optEDO), tobs, 1);
fprintf('>>> Valores exactos: u0 = %12.6f, v0 = %12.6f \n', 1, 0.8)
[usol, vsol] = InvLotkaVolterra(tobs, uobs, 0.8, 0.6);
```

Con este arranque se recupera aproximadamente $(1, 0.8)$ y el residuo es pequeño frente a la escala de las observaciones. Es una verificación de este ajuste sintético, no una garantía de identificación para cualquier colección de datos.

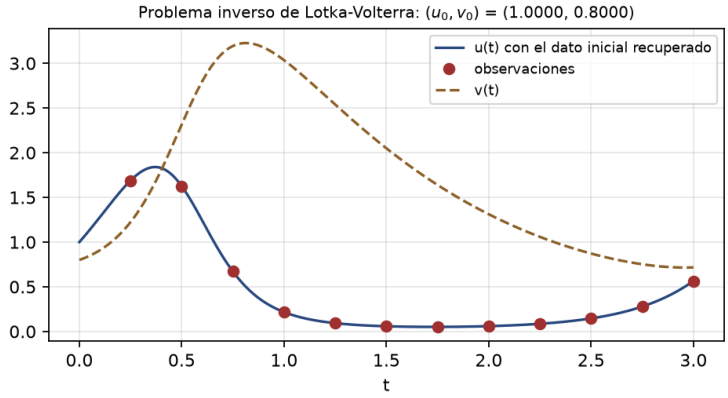


Figura 4.3: Presas $u(t)$ con el dato inicial recuperado, las doce observaciones y los depredadores $v(t)$ correspondientes al ajuste del experimento sintético.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```python
import numpy as np
from scipy.integrate import solve_ivp
from scipy.optimize import minimize
f = lambda t,y: [5*y[0]-3*y[0]*y[1], -y[1]+2*y[0]*y[1]]

def usim(c, tobs):
    if not np.all(np.isfinite(c)) or np.any(np.asarray(c) < 0):
        return None
    sol = solve_ivp(f,[0,tobs[-1]],c,t_eval=tobs,
                    rtol=1e-10,atol=1e-12,max_step=0.01)
    if not sol.success or len(sol.t) != len(tobs):
        return None
    return sol.y[0]

def inv_lotka(tobs, uobs, c0):
    tobs = np.asarray(tobs,dtype=float).ravel()
    uobs = np.asarray(uobs,dtype=float).ravel()
    if (len(tobs) == 0 or len(tobs) != len(uobs)
            or not np.all(np.isfinite(np.r_[tobs,uobs]))
            or not np.all(tobs > 0) or not np.all(np.diff(tobs) > 0)):
        raise ValueError("Datos finitos y tiempos positivos y crecientes")
    def coste(c):
        u = usim(c,tobs)
        return np.inf if u is None else np.sum((u-uobs)**2)
    if not np.isfinite(coste(c0)):
        raise ValueError("El arranque no es admisible")
    r = minimize(coste,c0,method="Nelder-Mead",
                 options={"xatol":1e-8,"fatol":1e-14})
    if not r.success or not np.isfinite(r.fun):
        raise RuntimeError(r.message)
    print("u0, v0 =",r.x,"J =",r.fun)
    return r.x

tobs = np.linspace(0.25,3,12)
uobs = usim([1.0,0.8],tobs)
```

```
inicial = inv_lotka(tobs,uobs,[0.8,0.6])
```

Qué hemos aprendido y dónde se falla. El dato inicial se identifica como cualquier otro parámetro, pero lo que se sabe de v_0 llega solo a través de u, y depende de cuándo se mira. Con la versión de Python:

- *Dónde se observa.* Con ruido de desviación 0.01 en las presas y veinte repeticiones, las doce observaciones en $[0.25, 3]$, que recorren casi un periodo entero de la órbita (unos 3.21), dan u_0 y v_0 con desviaciones de 0.006 y 0.002. Doce observaciones en $[1.25, 2.25]$, cuando las presas están en su mínimo (entre 0.05 y 0.09), dan 0.06 y 0.02: diez veces peor. Allí u apenas cambia con el dato inicial y el ruido pesa más que la señal; el menor valor singular de la matriz de sensibilidades baja de 1.85 a 0.14.
- *Olvidar la restricción.* `fminsearch` no admite cotas, y es la línea `any(c < 0)` del coste la que le impide probar poblaciones negativas. Y el caso $u_0 = 0$ no se puede recuperar: $u \equiv 0$ para cualquier v_0.
- *Comparar en los instantes del integrador.* Con un intervalo $[0, t_f]$ como segundo argumento, el integrador devuelve los instantes que elige él, en número variable: en Python, para el dato $(1, 0.8)$ y las tolerancias por defecto, once instantes en $[0, 3]$, menos que las doce observaciones. La comparación se hace en `tobs`, con `deval` o pasando `tobs` como vector de tiempos.

Ejercicio 27: el sistema de optimalidad como problema de contorno

Ejercicio 4.9. *Control óptimo de una EDO lineal con* `bvp4c`

ORIGEN: hoja de ejercicios del tema 4, ejercicio 27, con los ficheros `ControlEDO.m` y `ControlEDOr.m` del curso.

Para el problema (4.2) con datos a, T, y_0 y $\alpha, \mu > 0$ con $\alpha + \mu = 2$, resolver el problema de contorno

$$\begin{cases} y' = ay - \dfrac{1}{\mu}\varphi, \\ -\varphi' = a\varphi + \alpha(y - y_d), \\ y(0) = y_0, \quad \varphi(T) = 0, \end{cases}$$

siendo y_d la solución de $y' = ay + e^{-t}$, $y(0) = y_0$. Representar el estado calculado y el deseado, y el control calculado comparado con e^{-t}. Después, con `ControlEDOr.m`, lo mismo con la restricción $0 \le v \le M$.

Resolución.

La idea. El sistema de optimalidad caracteriza el control resolviendo un problema de contorno en (y, φ). En el caso lineal cuadrático, las condiciones necesarias son también suficientes y hay una única solución. `bvp4c` calcula las dos funciones conjuntamente, mediante discretización, iteraciones y adaptación de malla; no realiza la iteración alternada sobre controles de la sección 4.1. En problemas no lineales puede haber varias soluciones del sistema de optimalidad y el arranque puede determinar la rama que se encuentra.

El núcleo del programa son `dydt`, con las ecuaciones de estado y adjunto, y `condcon`, que reparte las condiciones entre los extremos: $y(0) = y_0$ y $\varphi(T) = 0$. El signo $-a$ de la ecuación

adjunta viene de la integración por partes que la define. Su dato final es cero porque el estado final es libre y el coste no tiene término terminal; con un coste terminal, el dato sería su gradiente. Poner ambas condiciones en $t = 0$ resolvería otro problema, que es el error señalado más adelante en `ControlOrden2.m`.

El estado deseado. $y' = ay + e^{-t}$ tiene solución general Ce^{at} más una particular. Si $a \neq -1$, se prueba De^{-t}: $-De^{-t} = aDe^{-t} + e^{-t}$ da $D = -1/(a+1)$, y con $y(0) = y_0$,

$$y_d(t) = \Big(y_0 + \frac{1}{a+1}\Big)e^{at} - \frac{1}{a+1}e^{-t}.$$

Si $a = -1$ la fuente es solución del homogéneo y se prueba Dte^{-t}: $y_d = (y_0 + t)e^{-t}$. Es lo que hace el `if a==-1` del fichero.

Qué se espera, y esto es a la vez el control del programa. El control $v = e^{-t}$ reproduce y_d exactamente —sin ninguna desviación— con coste $\frac{\mu}{2}\int_0^T e^{-2t}\,dt = \frac{\mu}{4}(1 - e^{-2T}) \approx 0.025$ para $\mu = 0.1$, $T = 10$. Ese número es una **cota superior calculada a mano**: el óptimo tiene que costar a lo sumo eso, porque e^{-t} es admisible. Y cuesta menos, $J(\hat{v}) = 0.023062$, con un estado que se aparta de y_d en 0.0410 como máximo y un control que arranca en 0.7216 en vez de 1. Es el compromiso: se ahorra control a cambio de un poco de desviación.

Comparar con un candidato admisible proporciona una comprobación necesaria: si el coste calculado es mayor por una diferencia que supera el error numérico, el resultado no puede ser óptimo. Ganar esa comparación no demuestra por sí solo optimalidad. Aquí la comparación muestra cuánto se ahorra respecto de seguir exactamente el objetivo: unos 0.002, alrededor del ocho por ciento del coste de e^{-t}.

La dependencia de μ ofrece otra comprobación. Con $\mu = 1$ el estado se aparta 0.2461 y el control arranca en 0.29; con $\mu = 0.1$, 0.0410 y 0.72; con $\mu = 0.01$, 0.0057 y 0.90. Al abaratar el control —μ pequeño— el óptimo se permite gastar más y seguir mejor al deseado, acercándose a e^{-t}; al encarecerlo, renuncia. La monotonía que garantiza la optimización se refiere a las normas L^2 del control y del error de seguimiento al variar μ/α; no obliga a que cada valor puntual o cada error máximo sea monótono. Los números anteriores describen este ejemplo.

```
function ControlEDO
% ControlEDO: minimizar alpha/2 ||y - yd||^2 + mu/2 ||v||^2 con y' = a y + v, y(0) = y0,
% resolviendo el sistema de optimalidad como problema de contorno con bvp4c.
a = -0.5; T = 10; y0 = 3;
mu = 0.1; alpha = 2 - mu; unmu = 1/mu;

% Estado deseado: solucion de y' = a y + exp(-t), y(0) = y0
if a == -1
  yd = @(t) (y0 + t).*exp(-t);
else
  una = 1/(a+1); C = y0 + una;
  yd = @(t) C*exp(a*t) - una*exp(-t);
end

% Sistema de optimalidad: z = [y; phi]
dydt = @(t,z) [a*z(1) - unmu*z(2); -a*z(2) - alpha*(z(1) - yd(t))];
condcon = @(za,zb) [za(1) - y0; zb(2)]; % y(0) = y0, phi(T) = 0
solinit = bvpinit(linspace(0, T, 30), [y0; 0]);
options = bvpset('Stats', 'on', 'RelTol', 1e-7, 'AbsTol', 1e-10);
```

```
sol = bvp4c(dydt, condcon, solinit, options);

% Representacion
tt = linspace(0, T);
yy = deval(sol, tt, 1);
yphi = deval(sol, tt, 2);
vcalc = -yphi*unmu;
subplot(2,1,1)
plot(tt, yy, 'b', 'LineWidth', 2), hold on
plot(tt, yd(tt), 'ro')
legend('Estado asociado al control optimo', 'Estado deseado')
subplot(2,1,2)
plot(tt, vcalc, 'g', 'LineWidth', 2), hold on
plot(tt, exp(-tt), 'mo')
legend('Control optimo calculado', 'Control optimo si fuera mu=0')
shg
end
```

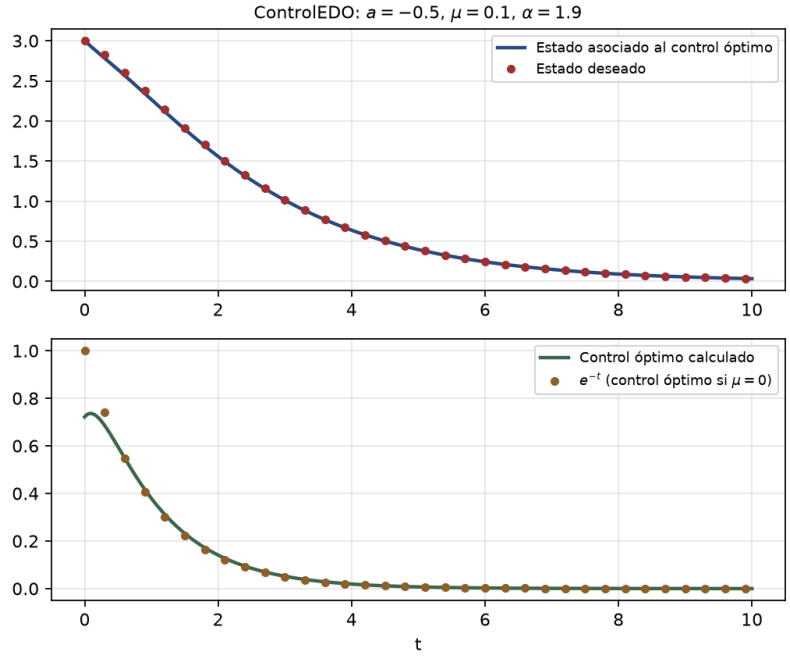


Figura 4.4: Arriba, el estado con el control óptimo y el deseado; abajo, el control óptimo frente a e^{-t}, que sería óptimo con $\mu = 0$. Con $\mu = 0.1$ el óptimo gasta menos control y se separa un poco del estado deseado.

Con restricciones: `ControlEDOr`. Con $\mathcal{U}_{\text{ad}} = \{0 \le v \le M\}$, $M = 0.7$, el criterio de optimalidad con restricciones dice $\hat{v} = P_{ad}(-\hat{\varphi}/\mu)$ con $P_{ad}(z) = \text{máx}(\text{mín}(z, M), 0)$, y la ecuación de estado del problema de contorno es $y' = ay + P_{ad}(-\varphi/\mu)$. **Una errata del fichero.** `ControlEDOr.m` escribe esa ecuación como `a*y(1) - Pfun(unmu*y(2))`, es decir, $y' = ay - P_{ad}(\varphi/\mu)$, que corresponde a un control $-P_{ad}(\varphi/\mu) \in [-M, 0]$, mientras que en el dibujo calcula el control como `Pfun(-yphi*unmu)`, el correcto. Con el signo del fichero, el control que entra en la ecuación es cero: se obtiene $y(t) = y_0 e^{at}$ y un adjunto negativo

antes de T. Pero el control que se *dibuja*, $P_{ad}(-\varphi/\mu)$, es positivo y llega a M; por tanto, no corresponde al estado calculado. Comprobar $y' = ay + v$ con el control representado descubre la incoherencia. Con el signo correcto, el control satura en $M = 0.7$ hasta aproximadamente $t = 0.3178$ y después baja cerca del control sin restricciones, con una diferencia de unos 0.022 justo al abandonar el techo. Conviene detenerse en esto, porque es la trampa del dibujo. El control restringido **no** es la proyección del control sin restricciones sobre $[0, M]$: es $P_{ad}(-\hat{\varphi}/\mu)$ con *su* adjunto, y ese adjunto resuelve un problema distinto, porque el estado ya no es el mismo. Se comprueba en los propios números: el control sin restricciones arranca en 0.7216, todavía *sube* hasta 0.7362 en $t = 0.088$ y no baja de 0.7 hasta $t = 0.2618$, mientras que el restringido no suelta el techo hasta $t = 0.3178$. Son dos instantes distintos, y en la figura 4.5 se ve que la curva gris cruza la línea horizontal antes que la verde. Las líneas que cambian son dos:

```
M = 0.7;
Pfun = @(z) max(min(z, M), 0);  % proyeccion sobre [0, M]
dydt = @(t,z) [a*z(1) + Pfun(-unmu*z(2)); -a*z(2) - alpha*(z(1) - yd(t))];  % v = P(-phi/
    mu)
% ... el resto como en ControlEDO, y en el dibujo:
vcalc = Pfun(-yphi*unmu);
```

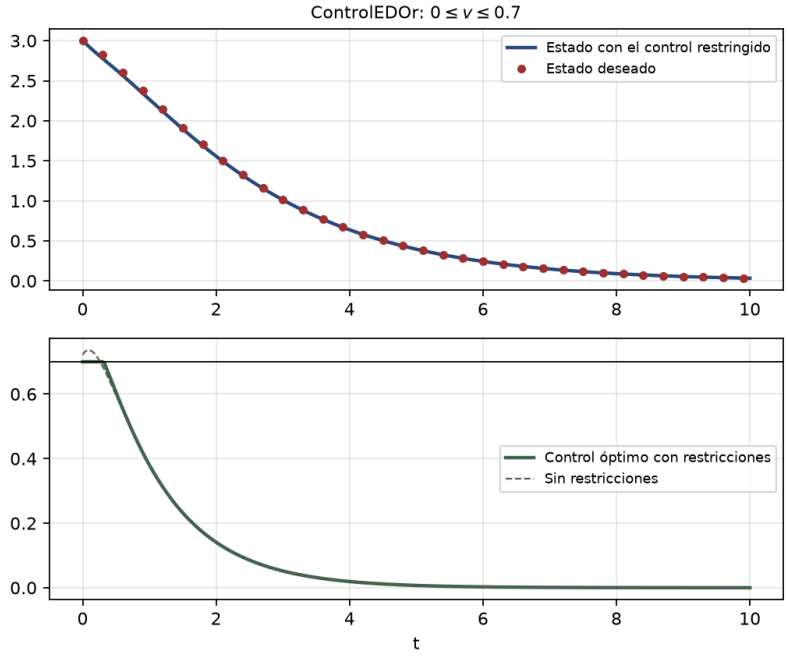


Figura 4.5: El mismo problema con $0 \le v \le 0.7$: el control óptimo satura al principio y después queda casi pegado al control sin restricciones (gris). Casi, y no exactamente: el gris sube por encima de 0.72 y cruza la banda en $t = 0.2618$, mientras que el verde no la abandona hasta $t = 0.3178$, porque cada uno es el óptimo de un problema distinto.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_bvp
```

```
a, T, y0, mu, alpha, M = -0.5, 10.0, 3.0, 0.1, 1.9, 0.7
if a == -1:
    yd = lambda t: (y0+t)*np.exp(-t)
else:
    una = 1/(a+1)
    yd = lambda t: (y0+una)*np.exp(a*t)-una*np.exp(-t)
P = lambda z: np.maximum(np.minimum(z, M), 0.0)
def resolver(proyectar):
    v_de_phi = (lambda ph: P(-ph / mu)) if proyectar else (lambda ph: -ph / mu)
    dydt = lambda t, z: np.vstack([a * z[0] + v_de_phi(z[1]), -a * z[1] - alpha * (z[0] -
         yd(t))])
    bc = lambda za, zb: np.array([za[0] - y0, zb[1]])
    xi = np.linspace(0, T, 30)
    sol = solve_bvp(dydt, bc, xi, np.vstack([y0*np.ones(30),np.zeros(30)]),
                    tol=1e-8, max_nodes=10000)
    if not sol.success:
        raise RuntimeError(sol.message)
    return sol
tt = np.linspace(0, T, 401)
for proyectar in (False, True):
    s = resolver(proyectar); y, ph = s.sol(tt)
    print("con restricciones" if proyectar else "sin restricciones", ": max|y-yd| =", np.
        max(np.abs(y - yd(tt))))
```

Dónde se rompe. En tres sitios, y los tres son propios del control óptimo.
El primero es **el signo y la coherencia del par estado–control**. La versión incorrecta de `ControlEDOr.m` calcula el estado del control nulo, cuyo coste es aproximadamente 0.6332, pero dibuja otro control. Para el problema restringido no se puede comparar con e^{-t}, que supera $M = 0.7$ al principio. Sí se puede usar $0.7e^{-t}$: es admisible, cuesta aproximadamente 0.06923, y el óptimo restringido cuesta 0.02309. Esta comparación descarta el control nulo como óptimo sin salir del conjunto admisible.
El segundo es **la proyección dentro del problema de contorno**. La función P_{ad} es continua y Lipschitz, pero no derivable en las cotas. Esos puntos pueden dificultar las iteraciones y la estimación del error del método de contorno; conviene revisar la convergencia y refinar la malla cerca de las conmutaciones. Un control *bang-bang* con saltos exige atención adicional. Suavizar una proyección cambia el problema y requiere comprobar el efecto de esa aproximación.
El tercero es de horizonte, y es más sutil de lo que parece. La condición $\varphi(T) = 0$ hace que cerca de T el control afloje, porque ya no queda futuro que compensar. Con $T = 10$ ese efecto de borde ocupa una fracción pequeña del intervalo y casi no se nota; con $T = 1$ ocupa el intervalo entero, y aunque el control arranca prácticamente igual —0.7131 frente a 0.7216—, el estado llega a apartarse 0.1067 de y_d en vez de 0.0410. **Cambiar T no es refinar un cálculo: es plantear otro problema**, cuya respuesta puede parecerse en el arranque y diferenciarse mucho en lo que mide el funcional. Por eso el ejercicio siguiente elige su T con un argumento —el período de la fuente— y no por comodidad.

QUÉ DEJA. El sistema de optimalidad lineal cuadrático permite calcular estado y adjunto como un problema de contorno. Con cotas, la proyección se aplica al adjunto del problema restringido: $v = P_{ad}(-\varphi/\mu)$. Se comprueban las ecuaciones, las condiciones en los extremos y el coste frente a candidatos que cumplan las mismas restricciones.

Ejercicio 28: un sistema de dos ecuaciones

Ejercicio 4.10. *Control óptimo de un sistema lineal* 2×2 *con un solo control*

ORIGEN: hoja de ejercicios del tema 4, ejercicio 28. No hay fichero de partida; se escribe aquí.

Se considera

$$\begin{cases} y_1' = Ay_1 + By_2 + v, & t \in (0,T), \\ y_2' = Cy_1 + Dy_2, \\ y_1(0) = y_{01},\ y_2(0) = y_{02}, \end{cases}$$

es decir, $y' = My + Qv$ con $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, $Q = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, y el coste (4.18) con observación completa (la matriz de observación es I; aquí C designa un coeficiente de M). Resolver el problema de contorno

$$\hat{y}' = M\hat{y} - \frac{1}{\mu} QQ^t\hat{\varphi}, \qquad -\hat{\varphi}' = M^t\hat{\varphi} + \alpha(\hat{y} - y_d), \qquad \hat{y}(0) = y_0,\ \hat{\varphi}(T) = 0,$$

siendo y_d la solución del sistema con fuente $\frac{1}{10}\cos(-t/50)$ en la primera ecuación, para $A = -0.05$, $B = 0.03$, $C = -0.11$, $D = 0.115$, $y_{01} = 1$, $y_{02} = 1.5$, $\alpha + \mu = 2$. Representar el estado calculado y el deseado, y el control comparado con la fuente.

PISTAS, DE MENOR A MAYOR

1. *Es el criterio de optimalidad para sistemas con $n = 2$, $m = 1$, $B = Q$: el control es $\hat{v} = -\frac{1}{\mu}Q^t\hat{\varphi} = -\hat{\varphi}_1/\mu$, solo la primera componente del adjunto.*
2. *Cuatro ecuaciones de primer orden, cuatro condiciones: $y(0)$ y $\varphi(T)$.*
3. *Para dibujar y_d se puede integrar su ecuación con `ode45` y evaluar la estructura devuelta con `deval`. Para calcular el control conviene trabajar directamente con el error $e = y - y_d$.*
4. *El enunciado no fija T. Tomamos $T = 100$, aproximadamente un tercio del período 100π de la fuente, para ilustrar su variación.*

Resolución.

La idea, y lo que este ejercicio añade a los anteriores es una sola cosa: hay menos controles que incógnitas. El estado tiene dos componentes y el control es un solo escalar, que entra únicamente en la primera ecuación. Que aun así se pueda gobernar y_2 es la pregunta de fondo, y la respuesta está en la matriz: $C = -0.11 \neq 0$ acopla las dos, de modo que actuar sobre y_1 mueve y_2 por rebote. Si C fuera cero, y_2 evolucionaría sola y ningún control podría tocarla; el problema seguiría teniendo solución, pero el término de y_2 en el coste sería una constante y el óptimo lo ignoraría.

Ahí está toda la modelización de este ejercicio, y por eso la matriz $Q = (1,0)^t$ merece más atención que el código. QQ^t es la proyección sobre la primera componente: dice *por dónde* se puede empujar. La traspuesta M^t del adjunto dice por dónde se propaga la información hacia atrás, y sus coeficientes son los de M con B y C intercambiados. La lectura es simétrica y vale la pena retenerla: M **dice cómo se propaga el efecto de una acción hacia el futuro;** M^t**, cómo se propaga hacia el pasado la exigencia de un objetivo**.

El horizonte $T = 100$ es una elección para este ejemplo, no un umbral necesario para observar la dinámica. Cambiarlo modifica el problema de optimización, como en el experimento anterior.

Calcular el error directamente. Aquí conocemos la ecuación que genera el objetivo: $y_d' = My_d + Qf$, con $f(t) = \frac{1}{10}\cos(-t/50)$ y el mismo dato inicial. Restándola de la ecuación de estado obtenemos

$$e' = Me + Q(v - f), \qquad -\varphi' = M^t\varphi + \alpha e, \qquad e(0) = 0, \quad \varphi(T) = 0, \quad v = -Q^t\varphi/\mu.$$

Es el mismo problema de contorno con cuatro incógnitas. La ventaja es que calcula el error pequeño sin restar dos estados que pueden ser grandes. Después se recupera $y = y_d + e$ para dibujarlo. Las líneas `dzdt` y `condcon` expresan estas ecuaciones y sus condiciones; el patrón matricial se extiende a otras dimensiones adaptando los bloques e índices del programa.

Las traspuestas. $M^t = \begin{pmatrix} A & C \\ B & D \end{pmatrix}$ y $QQ^t = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$: el control solo entra en la primera ecuación, y el adjunto acopla las dos componentes a través de C y B intercambiados. La incógnita $z = (e_1, e_2, \varphi_1, \varphi_2)$ tiene cuatro componentes.

```
function ControlSDO
% ControlSDO: y' = M y + Q v, coste alpha/2 ||y - yd||^2 + mu/2 ||v||^2, sistema de
    optimalidad
% con bvp4c, usando el error e = y - yd. Datos: A=-0.05, B=0.03, C=-0.11, D=0.115, y0
    =(1,1.5); T=100 elegido.
A = -0.05; B = 0.03; C = -0.11; D = 0.115;
M = [A, B; C, D]; Q = [1; 0]; y0 = [1; 1.5];
T = 100; mu = 0.1; alpha = 2 - mu;

% Estado deseado: solucion con fuente (1/10) cos(-t/50) en la primera ecuacion
fuente = @(t) 0.1*cos(-t/50);
sold = ode45(@(t,y) M*y + Q*fuente(t), [0, T], y0, ...
    odeset('RelTol',1e-10,'AbsTol',1e-12));
if sold.x(end) < T, error('El estado deseado no alcanzo T'); end
yd = @(t) deval(sold, t); % columna (y_d1; y_d2)

% Sistema de optimalidad: z = [e1; e2; phi1; phi2]
dzdt = @(t,z) [M*z(1:2) - (Q*Q')*z(3:4)/mu - Q*fuente(t); ...
              -M'*z(3:4) - alpha*z(1:2)];
condcon = @(za,zb) [za(1:2); zb(3:4)]; % e(0) = 0, phi(T) = 0
solinit = bvpinit(linspace(0, T, 100), zeros(4,1));
sol = bvp4c(dzdt, condcon, solinit, ...
    bvpset('RelTol',1e-8,'AbsTol',1e-10,'NMax',50000));

tt = linspace(0, T, 400);
zz = deval(sol, tt);
vcalc = -(Q'*zz(3:4,:))/mu; % v = -phi1/mu
ydd = yd(tt); ycalc = ydd + zz(1:2,:);
close all
subplot(2,1,1)
plot(tt, ycalc(1,:), 'b', tt, ycalc(2,:), 'g', 'LineWidth', 2), hold on
plot(tt, ydd(1,:), 'bo', tt, ydd(2,:), 'gs', 'MarkerSize', 3)
legend('y_1 calculado', 'y_2 calculado', 'y_{d,1}', 'y_{d,2}', 'Location', 'Best')
subplot(2,1,2)
plot(tt, vcalc, 'm', 'LineWidth', 2), hold on
plot(tt, fuente(tt), 'ro', 'MarkerSize', 3)
legend('Control optimo calculado', 'Fuente (1/10) cos(-t/50)', 'Location', 'Best')
fprintf('max |y - yd| = %8.4f %8.4f \n', max(abs(zz(1:2,:)), [], 2))
shg
end
```

Órdenes nuevas. `max(A, [], 2)` calcula el máximo por filas (el segundo argumento vacío es obligatorio para poder pasar la dimensión), frente a `max(A)`, que lo hace por columnas. La función anónima `yd` evalúa con `deval` el estado deseado para las figuras.

Para $T = 100$, $\mu = 0.1$ y $\alpha = 1.9$, el máximo del error es aproximadamente 0.00945 en la primera componente y 0.000292 en la segunda. El control comienza cerca de la fuente: 0.1004 frente a 0.1. Su mayor diferencia, unos 0.0416, se da en T: la condición $\varphi(T) = 0$ impone $v(T) = 0$, mientras que la fuente vale aproximadamente -0.0416. Es el efecto del extremo final que ya apareció en el ejercicio anterior.

Comprobaciones. Al reducir μ/α, seguir el objetivo gana peso frente al coste del control. La fuente permite alcanzarlo exactamente, lo que proporciona un caso conocido para comprobar el programa; no es por sí solo una prueba de que un problema inverso esté bien planteado.

Tampoco hay una regla general que obligue al error de la segunda componente a ser menor: su tamaño depende del acoplamiento y de los datos. Se comprueban las dos ecuaciones de estado, las dos del adjunto y las condiciones $e(0) = 0$, $\varphi(T) = 0$.
También se compara el coste con el de la fuente, que es un control admisible:

$$J(f) = \frac{\mu}{2} \int_0^T f(t)^2 \, dt \approx 0.020270, \qquad J(\hat{v}) \approx 0.020250.$$

El óptimo ahorra una pequeña parte del esfuerzo a cambio de permitir una desviación del objetivo, que aquí es *exactamente* alcanzable. Si el coste calculado fuera mayor que el del candidato por encima del error numérico, no sería óptimo. Ganar esta comparación es una comprobación necesaria; la suficiencia viene del sistema de optimalidad y la convexidad del problema.
Qué comprueba poner $C = 0$. Entonces la segunda ecuación queda desacoplada: y_2 e $y_{d,2}$ resuelven la misma ecuación con el mismo dato inicial. Por tanto, $e_2 = 0$ exactamente, aunque el control no pueda influir sobre esa componente. El segundo término del coste desaparece por construcción. Para comprobar la capacidad de actuar sobre y_2 habría que proponer otro objetivo, no uno que ya satisface su evolución libre.
Horizontes largos y escalas numéricas. Los autovalores de M son aproximadamente -0.0267 y 0.0917. El modo inestable amplifica los errores absolutos y hace crecer los estados; en $T = 100$, y_d llega aproximadamente a $(-204.5, -968.4)$. Calcular e evita obtener errores de tamaño 10^{-2} restando esas cantidades, y desacopla el cálculo del control de la integración previa de y_d.
La matriz del problema de contorno en (e, φ) es

$$\mathcal{H} = \begin{pmatrix} M & -QQ^t/\mu \\ -\alpha I & -M^t \end{pmatrix}.$$

Sus autovalores dependen también de los bloques de acoplamiento: con estos datos son aproximadamente ± 4.3570 y ± 0.1592, no $\pm$ los de M. Los modos que crecen en sentidos opuestos dificultan resolver todo como un problema de valores iniciales; el método de contorno impone condiciones en ambos extremos. No existe una frontera universal que haga fiable $T = 100$ e inválido $T = 300$. Con la formulación en e, una comprobación independiente del sistema lineal da resultados consistentes también para $T = 300, 600, 1000$. En cada ejecución hay que revisar los avisos de convergencia, las ecuaciones y condiciones satisfechas y la estabilidad del resultado al ajustar tolerancias. Si se reconstruye y a tiempos largos, su integración y representación siguen teniendo sus propias limitaciones numéricas.

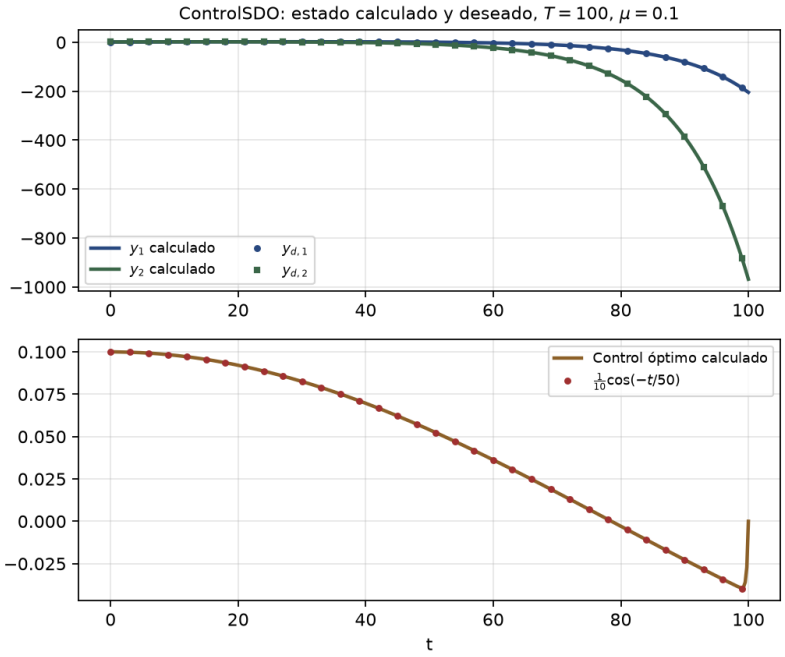


Figura 4.6: Sistema 2×2 con un control en la primera componente y $T = 100$. Arriba, estado calculado y deseado, indistinguibles a esta escala; abajo, control óptimo y fuente. La diferencia cerca de T refleja la condición final del adjunto.

Sobre `ControlOrden2.m`. El curso incluye también un fichero para la EDO de segundo orden de la sección 4.4, que integra el sistema de cuatro ecuaciones con `ode45` desde $t = 0$ imponiendo $\varphi(0) = \varphi'(0) = 0$. Eso no es el sistema de optimalidad, cuyas condiciones para φ van en $t = T$; es una simulación del sistema acoplado con un arranque arbitrario, y el control que dibuja no es el óptimo. Para resolverlo bien se procede como en este ejercicio: `bvp4c` con $z = (y, y', \varphi, \varphi')$, condiciones `[za(1)-ycero; za(2)-yuno; zb(3); zb(4)]` y, para los coeficientes constantes del fichero, la ecuación $\varphi'' = a\varphi - b\varphi' + \alpha(y - y_d)$ del criterio de optimalidad de segundo orden.

VERSIÓN EN PYTHON (NUMPY/MATPLOTLIB)

```
import numpy as np
from scipy.integrate import solve_ivp, solve_bvp
M = np.array([[-0.05, 0.03], [-0.11, 0.115]]); Q = np.array([1.0, 0.0]); y0 = np.array
    ([1.0, 1.5])
T, mu, alpha = 100.0, 0.1, 1.9
fuente = lambda t: 0.1 * np.cos(-t / 50)
sd = solve_ivp(lambda t, y: M @ y + Q * fuente(t), [0, T], y0,
               dense_output=True, rtol=1e-10, atol=1e-12)
if not sd.success or sd.t[-1] < T:
    raise RuntimeError("El estado deseado no alcanzo T")
def rhs(t, z):
    e, ph = z[0:2], z[2:4]
    return np.vstack([M @ e - np.outer(Q, Q) @ ph / mu - Q[:,None]*fuente(t),
                      -(M.T @ ph) - alpha * e])
bc = lambda za, zb: np.r_[za[:2], zb[2:4]]
xi = np.linspace(0, T, 100)
```

```
s = solve_bvp(rhs, bc, xi, np.zeros((4,xi.size)), tol=1e-8, max_nodes=50000)
if not s.success:
    raise RuntimeError(s.message)
tt = np.linspace(0, T, 401); z = s.sol(tt); v = -(Q @ z[2:4]) / mu
y = sd.sol(tt) + z[:2] # reconstruccion para dibujar
print("max|y - yd| =", np.max(np.abs(z[:2]), axis=1))
```

Repaso intercalado: capítulo 4, en tres rondas

A libro cerrado, en sesiones distintas. Un tercio de este capítulo y dos tercios de los anteriores.

Primera ronda, al terminar el capítulo.

1. Escribe el problema de control básico y su sistema de optimalidad, y di qué es cada una de las cinco líneas.
2. (Capítulo 3) Enuncia la condición de Euler-Lagrange y deduce de ella el sistema de optimalidad del control con $g = v$: ¿qué es φ?
3. (Capítulo 2) ¿Qué teorema decide el destino de $y' = \lambda y - \kappa y^2 - h$ sin resolver, y cuál es el máximo h sostenible?
4. ¿Por qué el adjunto lleva dato final y no inicial? Señala la línea de la demostración.
5. (Capítulo 1) ¿Qué es R_0 y qué es $\hat{\lambda}$, y en qué se parecen como umbrales?
6. ¿Qué tres cosas hacen falta en L^2 para probar la existencia del control óptimo?

Segunda ronda, una semana después.

1. Demuestra la existencia y unicidad del control óptimo para (4.2), paso a paso, y di dónde se usa $\mu > 0$.
2. (Capítulo 3) Resuelve la cuerda con muelles y compara sus condiciones de Robin con la condición de proyección de la sección 4.6. ¿Qué variaciones están permitidas en cada caso y por qué aparecen una igualdad en el extremo libre y una desigualdad variacional con restricciones?
3. (Capítulo 2) Deduce el modelo de Volterra adimensionalizado y da la cantidad conservada. ¿Qué papel juega en el ejercicio de identificación del dato inicial de Lotka–Volterra?
4. Escribe el hamiltoniano de (4.2) y comprueba que (4.31) devuelve (4.8).
5. (Capítulo 1) Demuestra que el método de la potencia para la matriz de Google converge con tasa a lo sumo α, y compáralo con la contractividad del punto fijo $v \mapsto \Lambda(v)$.
6. Con restricciones $v_* \le v \le v^*$, deduce $\hat{v} = P_{ad}(-\hat{\varphi}/\mu)$ desde la desigualdad variacional.

Tercera ronda, de repaso final.

1. Plantea como problema de control la pesca con esfuerzo variable $\mu(t)$ del capítulo 2 (estado logístico, coste $-\int \mu p$ más un término en μ^2), escribe el hamiltoniano y el sistema de optimalidad.
2. (Capítulo 3) Demuestra el teorema de convexidad y explica por qué en control lineal cuadrático la condición necesaria es también suficiente sin invocarlo.
3. (Capítulo 2) En el modelo SIR, si se añade una vacunación $v(t)$ que saca susceptibles, escribe la ecuación de estado y el adjunto para el coste $\int I + \frac{\mu}{2}\int v^2$.

4. Escribe de memoria el programa que resuelve el sistema de optimalidad de (4.2) con `bvp4c`, y el que identifica un parámetro con `fminsearch`. ¿Cuándo se usa cada uno?
5. (Capítulo 1) Un problema inverso para Leslie: dadas las poblaciones totales en veinte períodos, identificar b_1. Escribe el coste y di con qué orden de MATLAB lo minimizarías.
6. En el problema bilineal de invertir o consumir, calcula el hamiltoniano y explica por qué el control óptimo es bang-bang.

APÉNDICE A

Complementos sobre ecuaciones y sistemas diferenciales ordinarios

Este apéndice reúne resultados de Ecuaciones Diferenciales Ordinarias que se usan como apoyo en el libro. Conserva las tres secciones del apéndice del curso y añade una cuarta dedicada al lema de Gronwall. Las referencias a *Notas de Ecuaciones Diferenciales Ordinarias* del autor (en adelante, «libro de EDO») permiten ampliar las demostraciones; aquí se desarrollan las de las propiedades de las órbitas, la prolongación dentro de compactos y Gronwall.

Cómo usarlo: no se lee de corrido. Se lee cada resultado la primera vez que un capítulo lo invoca, y se vuelve a él cuando haga falta. El orden es el natural: primero lo que vale para cualquier sistema, después lo propio de los sistemas autónomos, la estabilidad y, al final, la desigualdad de Gronwall que se usa en control.

La notación del libro de EDO, traducida. En sus capítulos 1 a 5 la variable independiente se llama x y la solución $y(x)$; a partir del capítulo 6 se llama t, como aquí. La hipótesis estándar es la misma, $f \in C(\Omega;\mathbb{R}^N) \cap \mathrm{Lip}_{\mathrm{loc}}(y,\Omega)$. La solución maximal se escribe $(I(x_0,y_0), \varphi(\cdot;x_0,y_0))$ y, en los sistemas autónomos, $I(y_0) := I(0,y_0)$ y $\varphi(t;0,y_0)$, que aquí se abrevia $\varphi(t;y_0)$. La matriz fundamental es F («m.f.»). La estabilidad se define allí para la solución nula φ_0 de un sistema $y' = f(t,y)$ con $f(t,0) = 0$ (su hipótesis (H)), y una observación de su sección 6.1, «reducción al equilibrio nulo», la traslada a cualquier equilibrio $\bar{y}$ con el cambio $z = y - \bar{y}$: así se enuncia aquí directamente. La derivada respecto del sistema es $\dot{V}(y) = \sum_i \partial_{y_i} V(y)\, f_i(y)$, que aquí se escribe $\nabla V \cdot f$. Las órbitas, semiórbitas y conjuntos límite se escriben igual: $\gamma(y_0)$, $\gamma^{\pm}(y_0)$, $\Lambda^{\pm}(y_0)$.

A.1 Sistemas no autónomos

LA PREGUNTA

Un físico escribe la ecuación del movimiento de un cuerpo y le da una posición y una velocidad iniciales. Antes de calcular nada quiere saber dos cosas: si el problema que ha escrito tiene solución, y si es la única. ¿Qué condición sobre la ecuación se lo garantiza, y durante cuánto tiempo?

A.1.1 Sistemas de primer orden en forma normal

Definición (localmente Lipschitz)**.** Sea $\Omega \subset \mathbb{R}^{N+1}$ un abierto no vacío, cuyos puntos se denotan (t, y) con $y = (y_1, \dots, y_N)$. Una función f es **(globalmente) lipschitziana respecto de** y **en** Ω, y se escribe $f \in \mathrm{Lip}(y; \Omega)$, si existe una constante $L > 0$, la constante de Lipschitz, tal que

$$|f(t, y_1) - f(t, y_2)| \leq L|y_1 - y_2| \qquad \text{para todo } (t, y_1), (t, y_2) \in \Omega.$$

Es **localmente lipschitziana respecto de** y **en** Ω, $f \in \mathrm{Lip}_{\mathrm{loc}}(y, \Omega)$, si lo es en un entorno de cada punto: para todo $(t_0, y_0) \in \Omega$ existen $r, L > 0$, que dependen del punto, con la desigualdad anterior para todo $(t, y_1), (t, y_2)$ en la bola $B((t_0, y_0); r)$.

Proposición. Si $f : \Omega \to \mathbb{R}^N$ es continua y existen las derivadas parciales de todas sus componentes respecto de las y_j y son continuas en Ω, entonces $f \in \mathrm{Lip}_{\mathrm{loc}}(y, \Omega)$.

Es el criterio que se usa en muchos de los modelos: los polinomios y las funciones con logaritmos o senos son de clase C^1 en sus dominios, luego localmente lipschitzianos. La hipótesis temporal de continuidad no cubre directamente un control meramente L^2; el capítulo 4 trata ese caso mediante soluciones absolutamente continuas y la ecuación integral. La demostración del criterio C^1 usa el teorema del valor medio aplicado en una bola convexa; está en el libro de EDO, §2.2 «Condición de Lipschitzianidad».

Teorema (solución maximal)**.** Sean $\Omega \subset \mathbb{R}^{N+1}$ un abierto no vacío y $f \in C^0(\Omega; \mathbb{R}^N) \cap \mathrm{Lip}_{\mathrm{loc}}(y, \Omega)$. Entonces, para todo $(t_0, y_0) \in \Omega$ existe una única solución maximal $\varphi \in C^1(I(t_0, y_0); \mathbb{R}^N)$ del problema de Cauchy

$$\begin{cases} y' = f(t, y), \\ y(t_0) = y_0. \end{cases}$$

Además, el intervalo $I(t_0, y_0)$ donde está definida la solución maximal es abierto.

La solución es maximal en el sentido de que, si I es un intervalo que contiene a t_0 y $\psi \in C^1(I; \mathbb{R}^N)$ es solución en I, entonces $I \subset I(t_0, y_0)$ y $\varphi = \psi$ en I: cualquier otra solución es un trozo de la maximal. Es usual escribir la solución maximal como función también de los datos iniciales, $\varphi = \varphi(t; t_0, y_0)$.

CONEXIÓN: DÓNDE ESTÁ EN EL LIBRO DE EDO, Y QUÉ HAY DETRÁS

El teorema es el de Picard más el de prolongación, y en el libro de EDO ocupa dos capítulos. La existencia y unicidad *local* es el «Teorema (de Picard)» del libro de EDO, §2.3 «Intuición física de Picard» (página 35): con $f \in C(\Omega; \mathbb{R}^N) \cap \mathrm{Lip}_{\mathrm{loc}}(y, \Omega)$, para cada $(x_0, y_0) \in \Omega$ existe $\delta > 0$ tal que en $I_\delta = [x_0 - \delta, x_0 + \delta]$ hay una y solo una solución del problema de Cauchy; se demuestra

por el punto fijo en el espacio de Banach de las funciones continuas con la norma del máximo (§2.1, con la comparación entre la vía de Picard y la de Peano). La unicidad *global* es el teorema de §3.2: dos soluciones locales con el mismo dato coinciden en la intersección de sus intervalos (dos pruebas, una por conexión y otra por Gronwall). El enunciado de arriba corresponde al «Teorema (de existencia y unicidad de solución maximal)» de §3.3 «Prolongación de soluciones. Existencia y unicidad de solución maximal» (página 47), que también prueba que $I(x_0, y_0)$ es abierto. La *caracterización de las soluciones maximales* de §3.4 (página 49) dice que una solución maximal escapa de todo compacto de Ω: si $\beta = \sup I < +\infty$, o bien $|\varphi(x)|$ no está acotada cuando $x \to \beta^-$, o bien $(x, \varphi(x))$ se acerca a la frontera de Ω. Esa caracterización es la que se usa una y otra vez en el capítulo 2, en la forma de la proposición de prolongación en un compacto, cuya demostración se incluye más adelante.

A.1.2 Ecuación de orden superior en forma normal

En lugar de un sistema de primer orden se puede considerar una EDO de orden n,

$$\begin{cases} y^{(n)} = g\big(t, y, y', y'', \dots, y^{(n-1)}\big), \\ y(t_0) = y_0^0,\ y'(t_0) = y_0^1,\ \dots,\ y^{(n-1)}(t_0) = y_0^{n-1}, \end{cases}$$

con g continua y localmente Lipschitz en el conjunto de variables $(y, y', \dots, y^{(n-1)})$. Este caso se reduce al anterior con el cambio $y_1 = y$, $y_2 = y'$, $\dots$, $y_n = y^{(n-1)}$, que lo transforma en el sistema de primer orden

$$\begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{pmatrix} = \begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ g(t, y_1, y_2, \dots, y_n) \end{pmatrix}, \qquad \begin{pmatrix} y_1(t_0) \\ y_2(t_0) \\ \vdots \\ y_n(t_0) \end{pmatrix} = \begin{pmatrix} y_0^0 \\ y_0^1 \\ \vdots \\ y_0^{n-1} \end{pmatrix}.$$

Es el cambio que se hace en el péndulo (capítulo 3, con $\psi = \varphi'$), en las vibraciones forzadas, en la viga de orden cuatro y en el control de una EDO de segundo orden (capítulo 4): cada vez que `ode45` o `bvp4c` necesitan un sistema de primer orden, es esto. La existencia y unicidad para la ecuación de orden n es el «Teorema (de Picard para una EDO de orden n)» del libro de EDO, §2.4 «Picard para EDOs de orden n» (página 37), con $g \in C(O) \cap \mathrm{Lip}_{\mathrm{loc}}(y, y', \dots, y^{(n-1)}; O)$.

A.1.3 Sistemas lineales

Un caso especialmente interesante y simple es el de un SDO **lineal**,

$$\begin{cases} y' = A(t)y + b(t), \\ y(t_0) = y_0, \end{cases}$$

con $A \in C^0(I; \mathcal{L}(\mathbb{R}^N))$ y $b \in C^0(I; \mathbb{R}^N)$, siendo I un intervalo no degenerado de $\mathbb{R}$, no necesariamente abierto. El sistema es **homogéneo** si $b \equiv 0$. La función $f(t, y) = A(t)y + b(t)$ es localmente lipschitziana en $I \times \mathbb{R}^N$ (de hecho globalmente en cada compacto de tiempos, con constante máx $\|A(t)\|$).

Teorema.

a) Las soluciones maximales del sistema lineal están definidas en todo el intervalo I. (En adelante, «solución» de un sistema lineal quiere decir la maximal.)

b) El conjunto de soluciones del sistema homogéneo $y' = A(t)y$ es un espacio vectorial de dimensión N.
c) El conjunto de soluciones del sistema no homogéneo $y' = A(t)y + b(t)$ es una variedad lineal afín de dimensión N cuyo espacio vectorial soporte es el conjunto de soluciones del homogéneo. Por tanto, si φ_p es una solución particular del no homogéneo, cualquier otra solución es $\varphi = \varphi_p + \varphi_h$ con φ_h solución del homogéneo.

Observación. Si $\varphi_1, \dots, \varphi_N$ es una base del espacio de soluciones del homogéneo, cualquier solución se escribe $\varphi(t) = c_1\varphi_1(t) + \cdots + c_N\varphi_N(t)$ con $c_i \in \mathbb{R}$, o en forma matricial $\varphi(t) = F(t)c$, siendo $F(t)$ la matriz que tiene por columnas las φ_i y c el vector columna de los c_i. Se dice que F es una **matriz fundamental** del sistema.

Proposición. Una función matricial $F \in C^1(I; \mathcal{L}(\mathbb{R}^N))$ es una matriz fundamental (sus columnas son N soluciones linealmente independientes del homogéneo) si y solo si verifica la EDO matricial $F' = A(t)F$ en I y existe $t_0 \in I$ con $\det F(t_0) \neq 0$. En ese caso, $\det F(t) \neq 0$ para todo $t \in I$.

Si se conoce una matriz fundamental, es decir, si se sabe resolver el homogéneo, la solución del no homogéneo se calcula por el **método de variación de constantes**: las soluciones son de la forma $\varphi(t) = F(t)c(t)$ con $c \in C^1(I; \mathbb{R}^N)$ verificando $F(t)c'(t) = b(t)$ para todo $t \in I$. Es la fórmula que en el capítulo 4 da la solución $y(t) = e^{at}y_0 + \int_0^t e^{a(t-s)}v(s)\,ds$ de $y' = ay + v$: aquí $F(t) = e^{at}$ y $c'(t) = e^{-at}v(t)$.

CONEXIÓN: EL CAPÍTULO DE SISTEMAS LINEALES DEL LIBRO DE EDO

Todo esto está en el capítulo 4 del libro de EDO, «Sistemas diferenciales ordinarios lineales»: la dimensión N del espacio de soluciones, en §4.2 «Motivación del caso lineal»; la matriz fundamental y sus tres propiedades (toda solución es $F(x)c$; otra solución matricial es $F(x)C$, y es m.f. si y solo si $\det C \neq 0$; la solución del problema de Cauchy es $\varphi(x; x_0, y_0) = F(x)F(x_0)^{-1}y_0$), en el teorema de §4.3 «El SDO lineal homogéneo. Matriz fundamental» (página 59); la fórmula de Jacobi-Liouville para $\det F$, en §4.4 (es lo que hay detrás de «$\det F(t) \neq 0$ para todo t»); y el método de variación de las constantes, en §4.5, con la imagen de la balsa que flota con la corriente. Y, más adelante, el capítulo 5, «Problemas de contorno para un SDO lineal», que es lo que hace `bvp4c` en el capítulo 3: un problema de contorno se ve allí (§5.2) como la intersección de la variedad de soluciones con las condiciones en los dos extremos, y el teorema de alternativa de §5.4 dice que puede no tener solución o tener infinitas; el ejemplo de §5.5, $y'' + y = 0$ con $y(0) = y(\pi) = 0$, es exactamente el muelle entre dos posiciones dadas del capítulo 3.

A.1.4 Comparación de SDOs

Para saber si la solución maximal está definida en todo $\mathbb{R}$ se suele recurrir a argumentos de comparación. Hay dos resultados: el primero compara dos soluciones del mismo SDO con distintos datos iniciales, y de él se deduce la continuidad de la solución respecto de los datos; el segundo compara las soluciones de dos EDOs distintas con el mismo dato. Se usa la notación $S(t_0, y_0) := \{(I, \varphi) : \varphi \text{ es solución del problema de Cauchy en el intervalo } I\}$.

Teorema. Sean $\Omega \subset \mathbb{R}^{N+1}$ abierto no vacío, $f \in C^0(\Omega; \mathbb{R}^N) \cap \mathrm{Lip}_{\mathrm{loc}}(y, \Omega)$ y $(t_0, y_0), (t_0, z_0) \in \Omega$. Si $(I, \varphi) \in S(t_0, y_0)$ y $(J, \psi) \in S(t_0, z_0)$, entonces para cada $t_f \in I \cap J$ con $t_f \geq t_0$ existe una constante $L > 0$, que depende de t_0, t_f, y_0 y z_0, tal que

$$|\varphi(t) - \psi(t)| \leq |y_0 - z_0|\, e^{L(t-t_0)} \qquad \text{para todo } t \in [t_0, t_f].$$

Si f es globalmente lipschitziana en Ω con constante L, la desigualdad vale con esa L para todo $t_f \in I \cap J$.

La demostración aplica el lema de Gronwall de la sección A.4. Para deducir dependencia continua, se toma una misma constante de Lipschitz en un entorno de la trayectoria sobre un intervalo compacto de tiempos. Así, los datos suficientemente próximos permanecen próximos en ese intervalo. Es, con x por t, el primer teorema y el corolario del libro de EDO, §3.5 «Comparación de soluciones» (página 50). La versión fuerte, que la solución maximal $\varphi(t; t_0, y_0)$ es continua en sus tres variables y que su dominio Θ es abierto, es el «Teorema (Dependencia continua respecto de los datos iniciales)» de §6.1 (página 126).

Teorema (comparación, caso escalar)**.** Sean $\Omega \subset \mathbb{R}^2$ abierto no vacío, $f, g \in C^0(\Omega) \cap \mathrm{Lip}_{\mathrm{loc}}(y, \Omega)$, $(t_0, y_0) \in \Omega$, y sean (I, φ) y (J, ψ) soluciones respectivas de

$$\begin{cases} y' = f(t,y), \\ y(t_0) = y_0, \end{cases} \qquad \text{y} \qquad \begin{cases} y' = g(t,y), \\ y(t_0) = y_0. \end{cases}$$

Si $f \leq g$ en Ω, entonces

$$\varphi(t) \leq \psi(t) \ \text{ para } t \in I \cap J \cap [t_0, +\infty) \qquad \text{y} \qquad \varphi(t) \geq \psi(t) \ \text{ para } t \in I \cap J \cap (-\infty, t_0].$$

LA IDEA: LA CARRERA

Dos corredores salen del mismo punto y el segundo tiene siempre, en cada posición y en cada instante, una velocidad mayor o igual: nunca se queda detrás. Hacia atrás en el tiempo, al revés. Es el argumento que en el capítulo 2 acota la competición simplificada por exponenciales ($u' = u(1-v) < u$ con datos positivos, luego $u < u_0 e^t$ para $t > 0$) y la competición general por logísticas, y con ello prueba que las soluciones viven para todo $t \geq 0$. Es el segundo teorema del libro de EDO, §3.5 «Comparación de soluciones» (página 51), con la demostración por reducción al absurdo en el primer instante en que se cruzan; allí se advierte que solo vale para $N = 1$.

A.2 Sistemas autónomos

LA PREGUNTA

Los modelos del capítulo 2 no dependen del reloj: la ley que rige una población hoy es la misma que mañana. ¿Qué gana un sistema por esa propiedad, qué significa entonces «la órbita» de una solución y por qué dos órbitas no pueden cruzarse nunca?

A.2.1 Sistemas de primer orden en forma normal

Sea $D \subset \mathbb{R}^N$ abierto no vacío y $f : D \to \mathbb{R}^N$ localmente lipschitziana en D: para cada $y_0 \in D$ existen $r, L > 0$ con $|f(y_1) - f(y_2)| \leq L|y_1 - y_2|$ para $y_1, y_2 \in B(y_0; r)$. Se escribe $f \in \mathrm{Lip}_{\mathrm{loc}}(D; \mathbb{R}^N)$ y se considera el SDO **autónomo**

$$y' = f(y).$$

Proposición. Si $\varphi \in C^1(I)$ es una solución maximal del SDO autónomo, para todo $\tau \in \mathbb{R}$ la función $\varphi_\tau(t) := \varphi(t + \tau)$, $t \in I - \tau$, es también solución maximal.

La comprobación es la regla de la cadena, con derivada interior 1: $\varphi'_\tau(t) = \varphi'(t + \tau) = f(\varphi(t + \tau)) = f(\varphi_\tau(t))$, y no se puede prolongar porque φ no se podía. En consecuencia, si $I(t_0, y_0)$ es el intervalo de la solución maximal $\varphi(t; t_0, y_0)$ con dato $y(t_0) = y_0$:

Proposición. Para todo $t_0 \in \mathbb{R}$ y todo $y_0 \in D$,

$$I(t_0, y_0) = I(0, y_0) + t_0 \qquad \text{y} \qquad \varphi(t; t_0, y_0) = \varphi(t - t_0; 0, y_0) \quad \text{para } t \in I(t_0, y_0).$$

Gracias a esto, en el caso autónomo se puede tomar siempre $t_0 = 0$, y se escribe $I(y_0) := I(0, y_0)$ y $\varphi(t; y_0) := \varphi(t; 0, y_0)$. Es lo que hace el capítulo 2 en todos sus problemas de Cauchy. Las dos proposiciones están en el libro de EDO, §6.2 «Sistemas autónomos» (página 130), donde además se prueba que en un sistema autónomo estabilidad y estabilidad uniforme coinciden, y se usan de nuevo al principio de §8.3 «Órbitas cíclicas y conjuntos límite. Propiedades» (página 168).

A.2.2 Sistema lineal homogéneo

Los SDOs lineales autónomos son $y' = Ay$ con $A \in \mathbb{R}^{N \times N}$.

Teorema. La función $F(t) := e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k$, $t \in \mathbb{R}$, es una matriz fundamental de $y' = Ay$. En consecuencia, la solución con $y(0) = y_0$ es $y(t) = e^{tA} y_0$.

Está en el libro de EDO, §4.6 «SDO lineal de coeficientes constantes. Exponencial de una matriz» (página 67): la definición de e^A con la convergencia de la serie en norma matricial, $e^{A+B} = e^A e^B$ si $AB = BA$, la proposición de que $F_A(x) = e^{xA}$ es de clase C^1 con $F'_A = AF_A$ y es matriz fundamental, y el corolario $\varphi(x; x_0, y_0) = e^{(x-x_0)A} y_0 + \int_{x_0}^{x} e^{(x-s)A} b(s)\, ds$ para el no homogéneo. Es el resultado con el que el capítulo 1 compara $v(k) = L^k v(0)$ con $y(t) = e^{tA} y_0$, y el que el capítulo 4 usa para escribir $y(t) = e^{At} y_0 + \int_0^t e^{A(t-s)} B v(s)\, ds$.

A.2.3 EDOs lineales de orden n con coeficientes constantes

Teorema. Sean $a_0, \ldots, a_{n-1} \in \mathbb{R}$ y considérese la EDO lineal de orden n

$$y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0,$$

y sea λ una raíz del polinomio característico $P(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1 \lambda + a_0$ de multiplicidad $m \geq 1$.

a) Si $\lambda \in \mathbb{R}$, entonces $e^{\lambda t}, te^{\lambda t}, \ldots, t^{m-1} e^{\lambda t}$ son m soluciones independientes.

b) Si $\lambda = \alpha + i\beta \in \mathbb{C}$ con $\beta \neq 0$, entonces las $2m$ soluciones independientes asociadas a $\alpha \pm i\beta$ son

$$e^{\alpha t} \cos(\beta t), \ldots, t^{m-1} e^{\alpha t} \cos(\beta t),$$
$$e^{\alpha t} \operatorname{sen}(\beta t), \ldots, t^{m-1} e^{\alpha t} \operatorname{sen}(\beta t).$$

Es lo que resuelve el muelle $x'' + \frac{k}{m}x = 0$ (raíces $\pm i\sqrt{k/m}$, soluciones cos y sen) y el péndulo linealizado en el capítulo 3, y lo que da la solución exacta de las vibraciones forzadas de la práctica 18 (junto con una solución particular para el término forzante). Es el teorema del libro de EDO, §4.9 «EDOs lineales de orden n de coeficientes constantes» (página 74), escrito allí con la ecuación en la forma $y^{(n)} = \sum_{k=1}^{n} a_k y^{(n-k)}$, el polinomio $p(\lambda) = \lambda^n - \sum_k a_k \lambda^{n-k}$ y las funciones $\varphi_{jk}(x) = x^k e^{\lambda_j x}$; el sistema fundamental de soluciones y el wronskiano de la ecuación de orden n están en §4.8.

A.3 Estabilidad de SDOs autónomos

LA PREGUNTA

Un lápiz de pie sobre la punta y un lápiz colgando de un hilo están los dos en equilibrio. Todo el mundo sabe cuál de los dos aguanta un soplo. ¿Cómo se convierte esa diferencia en una definición, y qué hay que calcular para decidirla sin soplar, con solo la ecuación?

Definición. Una solución φ del sistema autónomo $y' = f(y)$ es un **punto de equilibrio** (o punto crítico, o equilibrio) si es constante. Los equilibrios son las funciones $\varphi(t) = \bar{y}$ con $\bar{y} \in D$ solución de $f(\bar{y}) = 0$.

En lo que sigue $\bar{y}$ es un equilibrio.

Definición. $\bar{y}$ es un equilibrio **estable** si para todo $\varepsilon > 0$ existe $\delta > 0$ con $B(\bar{y};\delta) \subset D$ tal que, para todo $y_0 \in B(\bar{y};\delta)$, se tiene $I(y_0) \supset [0,+\infty)$ y

$$|\varphi(t;y_0) - \bar{y}| \leq \varepsilon \qquad \text{para todo } t \in [0,+\infty).$$

Definición. $\bar{y}$ es un equilibrio **atractivo** si existe $\gamma > 0$ con $B(\bar{y};\gamma) \subset D$ tal que, para todo $y_0 \in B(\bar{y};\gamma)$, $I(y_0) \supset [0,+\infty)$ y $\lim_{t\to+\infty} \varphi(t;y_0) = \bar{y}$.

Observación. Existen sistemas autónomos con equilibrios estables no atractivos (el $(1,1)$ de Volterra en el capítulo 2, rodeado de órbitas cerradas) y con equilibrios atractivos que no son estables (hay datos arbitrariamente próximos cuyas soluciones salen de un entorno fijo antes de volver).

Definición. $\bar{y}$ es **asintóticamente estable** si es estable y atractivo.

Definición. $\bar{y}$ es **exponencialmente asintóticamente estable** si existen $\gamma, M, \lambda > 0$ con $B(\bar{y};\gamma) \subset D$ tales que, para todo $y_0 \in B(\bar{y};\gamma)$, $I(y_0) \supset [0,+\infty)$ y

$$|\varphi(t;y_0) - \bar{y}| \leq M|y_0 - \bar{y}|e^{-\lambda t} \qquad \text{para todo } t \in [0,+\infty).$$

Observación. Exponencialmente asintóticamente estable implica asintóticamente estable, pero la implicación contraria es falsa en general (el equilibrio 0 de $I' = -\beta I^2$ en el ejercicio del SIS con $R_0 = 1$ decae como $1/t$, considerando aquí estabilidad relativa al dominio físico $I \geq 0$). Además del comportamiento cuando $t \to +\infty$ puede interesar el de $t \to -\infty$; se reduce al anterior con el cambio $t \to -t$.

LA IDEA: LAS TRES DEFINICIONES EN EL LÁPIZ

Estable es «si empiezo cerca, me quedo cerca para siempre»: la bola de radio ε se puede elegir tan pequeña como se quiera, a costa de arrancar en una bola δ más pequeña. Atractivo es «si empiezo cerca, acabo volviendo». Asintóticamente estable es las dos cosas. Exponencialmente es «vuelvo, y a un ritmo geométrico que no se degrada». El lápiz colgando sin rozamiento es estable (oscila cerca) pero no atractivo (no para); con rozamiento es asintóticamente estable. El lápiz de punta no es ni una cosa ni otra. Las definiciones están en el libro de EDO, §6.1 «Definiciones de estabilidad» (páginas 127 y 128), enunciadas para la solución nula φ_0 y con sus versiones *uniformes* (el δ no depende de t_0), que en los sistemas autónomos coinciden con estas (§6.2). La definición exponencial está en §6.3 (página 143); el factor $|y_0 - \bar{y}|$ mide el tamaño de la perturbación inicial. Las primeras definiciones están comentadas con dibujos en §6.1.1 «Intuición de las definiciones».

Los resultados siguientes permiten decidir si un equilibrio es estable o atractivo sin recurrir a la definición.

A.3.1 Estabilidad de SDOs lineales de coeficientes constantes

Teorema. Sea $A \in \mathcal{L}(\mathbb{R}^N)$. Se tiene:

a) El equilibrio nulo es atractivo para $y' = Ay$ si y solo si todos los autovalores de A tienen parte real negativa. En ese caso es, de hecho, exponencialmente asintóticamente estable.
b) El equilibrio nulo es estable para $y' = Ay$ si y solo si los autovalores de A tienen parte real negativa o nula y, para los de parte real nula, la multiplicidad geométrica (la dimensión del espacio de autovectores) coincide con la multiplicidad algebraica.

Proposición. Sean $A \in \mathcal{L}(\mathbb{R}^N)$, $b \in \mathbb{R}^N$ e $\bar{y} \in \mathbb{R}^N$ con $A\bar{y} + b = 0$. Entonces $\bar{y}$ es estable (resp. atractivo, exponencialmente asintóticamente estable) para $y' = Ay + b$ si y solo si el equilibrio nulo lo es para $y' = Ay$.

Observación. Para SDOs lineales de coeficientes constantes, la atractividad implica estabilidad y además equivale a la estabilidad asintótica exponencial. Esto es falso en general para SDOs no lineales.

CONEXIÓN: LA FORMA DE JORDAN, OTRA VEZ

La demostración es la misma cuenta que el capítulo 1 hizo con L^k: $e^{tA} = Se^{tJ}S^{-1}$ y un bloque de Jordan de autovalor μ produce entradas $t^j e^{\mu t}$, que tienden a cero si y solo si $\operatorname{Re}\mu < 0$, y permanecen acotadas con $\operatorname{Re}\mu = 0$ solo si $j = 0$, es decir, si el bloque es 1×1, que es la condición sobre las multiplicidades. En el libro de EDO es el teorema de §6.3 «Sistemas lineales» (página 135), enunciado para cualquier $A(t)$ con la matriz fundamental y la norma espectral ($\sup_{t \geq t_0} \|F(t)\|_s < \infty$ para la estabilidad, $\|F(t)\|_s \to 0$ para la asintótica), y el teorema de §6.3.1 «Sistemas lineales con coeficientes constantes» (página 143), que lo traduce a los autovalores y a las cajas de Jordan exactamente como aquí.

A.3.2 Estabilidad de algunos SDOs autónomos no lineales

Una primera manera de analizar la estabilidad de un sistema no lineal consiste en averiguar qué le ocurre a un sistema lineal «parecido», aproximándolo cerca del equilibrio. Si f es de clase C^1 en D y $\bar{y}$ es un equilibrio, el desarrollo de Taylor de orden 1 da

$$f(y) = Df(\bar{y})(y-\bar{y}) + E(y), \qquad \lim_{y\to\bar{y}} \frac{E(y)}{|y-\bar{y}|} = 0,$$

lo que permite aproximar $y' = f(y)$ por el sistema lineal $y' = Df(\bar{y})(y-\bar{y})$, de matriz $A = Df(\bar{y})$.

Teorema (estabilidad por linealización)**.** Sea $\bar{y}$ un equilibrio de $y' = f(y)$ con f diferenciable en $\bar{y}$.

a) Si existe al menos un autovalor de $Df(\bar{y})$ con parte real positiva, $\bar{y}$ es inestable.
b) Si todos los autovalores de $Df(\bar{y})$ tienen parte real negativa, $\bar{y}$ es exponencialmente asintóticamente estable.

Este resultado **no dice nada** cuando todos los autovalores tienen parte real ≤ 0 y alguno la tiene nula. Es un caso que se presenta con frecuencia: en el capítulo 2, el $(1,1)$ de Volterra (autovalores $\pm\sqrt{\lambda}\,i$) y el continuo de equilibrios del SIR; en el capítulo 3, los equilibrios bajos del péndulo. Ahí el estudio se hace con la **teoría de Liapunov**.

CONEXIÓN: EL PRIMER MÉTODO DE LYAPUNOV

Es el «primer método de Lyapunov» del libro de EDO, §6.4 «Primer método de Lyapunov» (páginas 146 y 147), en tres piezas: el «Teorema (estabilidad en primera aproximación)» para $y' = A(t)y + g(t,y)$ con $|g(t,y)|/|y| \to 0$ uniformemente en t; el «Corolario (estabilidad por linealización)», que es el apartado b) de arriba con $f \in C^2$; y el «Teorema (inestabilidad por linealización)», que es el apartado a). En §6.4.1 está el caso en que f deriva de un potencial. El libro previo del autor lo citó desde el capítulo 1 al considerar la aproximación lineal de otros modelos de población cerca de un equilibrio. En tiempo discreto también se estudia la jacobiana, que no tiene por qué ser una matriz de Leslie: el criterio de atracción usa autovalores de módulo menor que 1, en lugar de parte real negativa.

Definición. Sean $\bar{y} \in \mathbb{R}^N$ y $\rho > 0$. Una función $V : \bar{B}(\bar{y};\rho) \to \mathbb{R}$ es **definida positiva** en $\bar{B}(\bar{y};\rho)$ si $V(\bar{y}) = 0$ y $V(y) > 0$ para todo $y \in \bar{B}(\bar{y};\rho) \setminus \{\bar{y}\}$. Es **definida negativa** si $-V$ es definida positiva.

Definición. Dada $V \in C^1(\bar{B}(\bar{y};\rho))$ con $\bar{B}(\bar{y};\rho) \subset D$, la **derivada de** V **respecto del sistema** $y' = f(y)$ es la función $\dot{V}(y) := \nabla V(y) \cdot f(y)$.

Observación. Si $\varphi \in C^1(I)$ es solución del sistema con imagen en $\bar{B}(\bar{y};\rho)$, entonces $\frac{d}{dt}V(\varphi(t)) = \nabla V(\varphi(t)) \cdot \varphi'(t) = \dot{V}(\varphi(t))$ para todo $t \in I$. Por tanto, si $\dot{V} \leq 0$, V decrece a lo largo de las trayectorias. Lo notable es que $\dot{V}$ se calcula **sin conocer las soluciones**: solo con V y f.

Definición. Sea $\bar{B}(\bar{y};\rho) \subset D$ y $V \in C^1(\bar{B}(\bar{y};\rho))$. V es una **función de Liapunov** en $\bar{B}(\bar{y};\rho)$ si es definida positiva allí y $\dot{V}(y) \leq 0$ para todo $y \in \bar{B}(\bar{y};\rho)$.

Teorema (Liapunov)**.** Sea $\bar{y}$ un punto crítico de $y' = f(y)$. Si existe una función de Liapunov V en $\bar{B}(\bar{y};\rho) \subset D$, entonces $\bar{y}$ es un equilibrio estable. Si además $\dot{V}$ es definida negativa en $\bar{B}(\bar{y};\rho)$, entonces $\bar{y}$ es asintóticamente estable.

LA IDEA: LA ENERGÍA QUE NO SUBE

V es una energía generalizada: vale cero en el equilibrio, positiva alrededor, y las soluciones no la aumentan. Una solución que empieza con energía pequeña no puede alejarse, porque alejarse costaría energía que no tiene: eso es la estabilidad. Si además la energía baja estrictamente fuera del equilibrio, la solución acaba en él. En Volterra, $V = \lambda\psi(u) + \psi(v)$ tiene $\dot{V} = 0$ en las variables adimensionales: energía exactamente conservada, órbitas en sus curvas de nivel, estable sin ser atractivo. En el péndulo con rozamiento, la energía mecánica tiene $\dot{V} = -c\varphi'^2 \leq 0$, que no es definida negativa (se anula en $\varphi' = 0$): Liapunov da solo estabilidad, y hace falta LaSalle para la atractividad. En el libro de EDO, las definiciones son las del libro de EDO, §7.2 «Funciones de Lyapunov», y el teorema es el «Teorema (Condiciones suficientes de estabilidad de Liapunov)» de §7.3 (página 155), que añade un apartado c): si $c_1|y|^2 \leq V(y) \leq c_2|y|^2$ y $\dot{V}(y) \leq -c_3|y|^2$, el equilibrio es exponencialmente asintóticamente estable. El teorema de Tchetaev de inestabilidad, que aquí no se usa, está en §7.5.

Los resultados anteriores se refieren a soluciones que se acercan a un punto. Hay veces en que oscilan y terminan aproximándose a un conjunto no unitario. Para eso hacen falta las órbitas y los conjuntos límite, y el teorema de LaSalle, que mejora al de Liapunov pidiendo menos para la atractividad.

Definición. Sea $y_0 \in D$. La **trayectoria** $\tau(y_0)$, la **semitrayectoria positiva** $\tau^+(y_0)$, la **órbita** $\gamma(y_0)$ y la **semiórbita positiva** $\gamma^+(y_0)$ asociadas a y_0 son

$$\begin{aligned}
\tau(y_0) &:= \{(t, \varphi(t; y_0)) : t \in I(y_0)\} \subset \mathbb{R}^{N+1},\\
\tau^+(y_0) &:= \{(t, \varphi(t; y_0)) : t \in I(y_0),\ t \geq 0\},\\
\gamma(y_0) &:= \{\varphi(t; y_0) : t \in I(y_0)\} \subset \mathbb{R}^N,\\
\gamma^+(y_0) &:= \{\varphi(t; y_0) : t \in I(y_0),\ t \geq 0\}.
\end{aligned}$$

La trayectoria es el grafo de la solución; la órbita es su imagen en el **espacio de fases** $\mathbb{R}^N$, la curva que dibuja el punto sin decir cuándo pasa por cada sitio. Las figuras del capítulo 2 en el plano (u, v) son órbitas; las de $u(t)$ y $v(t)$ contra el tiempo son trayectorias proyectadas.

Definición. Una trayectoria $\tau(y_0)$ es **periódica** si $I(y_0) = \mathbb{R}$ y existe $T > 0$ con $\varphi(t + T; y_0) = \varphi(t; y_0)$ para todo t. Se dice que T es un período y que la órbita $\gamma(y_0)$ es **cíclica**.

Proposición.

a) Dadas dos órbitas $\gamma(y_0)$ y $\gamma(y_1)$, se tiene $\gamma(y_0) \cap \gamma(y_1) \neq \emptyset$ si y solo si $\gamma(y_0) = \gamma(y_1)$.
b) Si existen $t_1 \neq t_2$ en $I(y_0)$ con $\varphi(t_1; y_0) = \varphi(t_2; y_0)$, entonces $\gamma(y_0)$ es cíclica y $|t_2 - t_1|$ es un período.

La idea, en una frase. Los dos apartados son el mismo hecho leído dos veces: en un sistema autónomo el reloj no importa, así que dos trayectorias que pasan por el mismo punto —o una misma trayectoria que pasa dos veces por un punto— son, salvo un desfase temporal, la misma. La imagen es la del raíl de una montaña rusa: la vía es una sola, y lo único que distingue a dos trenes es a qué hora salieron.

Paso 1: dos órbitas que se tocan son la misma (apartado a), el núcleo). Si $\varphi(t_1; y_0) = \varphi(t_2; y_1) =: p$, la solución que sale de p en el instante 0 es, por la proposición de invariancia por traslación temporal, tanto $\varphi(t + t_1; y_0)$ como $\varphi(t + t_2; y_1)$; por unicidad son la misma función, y sus imágenes, que son las dos órbitas, coinciden.
Paso 2: volver al mismo punto obliga a repetir la órbita (apartado b). Las funciones $t \mapsto \varphi(t + t_1; y_0)$ y $t \mapsto \varphi(t + t_2; y_0)$ son soluciones maximales con el mismo valor en $t = 0$. Por unicidad, coinciden sus funciones y sus intervalos de definición: $I(y_0) - t_1 = I(y_0) - t_2$. Un intervalo no vacío que coincide con su traslación por $t_2 - t_1 \neq 0$ solo puede ser $\mathbb{R}$. La identidad de las funciones da entonces $\varphi(t + t_2 - t_1; y_0) = \varphi(t; y_0)$ para todo $t \in \mathbb{R}$: el período es $|t_2 - t_1|$. ■

Qué hace cada hipótesis. Los dos pasos descansan en lo mismo, y en nada más: la *unicidad* de solución del problema de Cauchy, más la invarianza por traslación temporal, que es lo que aporta el ser *autónomo*. Si el sistema no fuera autónomo el resultado es falso, y se ve en el espacio de fases extendido: dos soluciones distintas de $y' = f(t, y)$ pueden pasar por el mismo punto y del espacio de estados en instantes distintos —el campo ha cambiado entre medias— y sus proyecciones se cruzan sin problema. Lo que sigue siendo cierto en el caso no autónomo es que las *gráficas* en el espacio-tiempo no se cortan; lo que se pierde al proyectar es que tampoco se corten las trayectorias.

El apartado a) es la frase «dos órbitas distintas no se cortan» que el capítulo 2 usa para la monotonía en el teorema de las EDO autónomas y para la positividad en Volterra y en competición. Está en el libro de EDO, §8.2 «Órbitas» (página 163: órbita, punto crítico, y los dos ejemplos de un SDO lineal y de un ciclo límite) y en §8.3 «Órbitas cíclicas y conjuntos límite. Propiedades» (página 169: el apartado a) es la proposición «$y_1 \in \gamma(y_0)$ si y solo si $\gamma(y_1) = \gamma(y_0)$», y el apartado b) es la proposición de la órbita que se corta a sí misma).

Definición. Un conjunto $\Gamma \subset D$ es **invariante** (resp. **positivamente invariante**) para el sistema si $y_0 \in \Gamma \Rightarrow \gamma(y_0) \subset \Gamma$ (resp. $\gamma^+(y_0) \subset \Gamma$).

Definición. p es un **punto límite positivo** asociado a $y_0 \in D$ si existe una sucesión $t_n \in I(y_0)$ con $t_n \to \sup I(y_0)$ tal que $\varphi(t_n; y_0) \to p$. El conjunto de los puntos límite positivos de y_0 se denota $\Lambda^+(y_0)$.

Proposición. Sea $K \subset D$ compacto y $y_0 \in K$ con $\gamma^+(y_0) \subset K$. Entonces:
a) $I(y_0) \supset [0, +\infty)$;
b) $\Lambda^+(y_0)$ es un subconjunto compacto, invariante y no vacío de K.

Observación. En las condiciones anteriores, como $\Lambda^+(y_0)$ es invariante, si se reduce a un punto $\bar{y}$ entonces $\bar{y}$ es un punto crítico. Dicho de otro modo: **si existe** $\lim_{t \to +\infty} \varphi(t; y_0) = \bar{y} \in D$**, entonces** $\bar{y}$ **es un punto crítico.**

La idea. Los tres enunciados son consecuencias de una sola frase: *una solución que no puede escapar de un compacto tampoco puede acabarse*. La única forma que tiene una solución maximal de morir en tiempo finito es irse al infinito o salirse del dominio, y las dos cosas son «salir de todo compacto»; encerrada en K no le queda ninguna manera de terminar, así que dura para siempre. A partir de ahí, el conjunto de sus puntos límite hereda las buenas propiedades del compacto —no es vacío por Bolzano-Weierstrass, es cerrado por construcción— y la

invariancia sale de que el sistema es autónomo. La observación final es el remate útil: si ese conjunto de destinos se reduce a un único punto, la solución que sale de ese punto tiene que quedarse ahí, y una solución que no se mueve es un equilibrio.

Paso 1: encerrada en un compacto, la solución no puede acabarse (apartado a, el núcleo). Es la caracterización de las soluciones maximales del libro de EDO, §3.4 «Caracterización de soluciones prolongables y maximales»: si $\sup I(y_0) = T$ fuera finito, la solución tendría que salir de todo compacto cuando $t \to T$, y no sale de K.
Paso 2: el conjunto límite hereda la compacidad y es invariante (apartado b). No vacío y compacto: $\varphi(t_n; y_0)$ está en el compacto K y tiene subsucesiones convergentes (Bolzano-Weierstrass), y el conjunto de límites es cerrado dentro de K. Invariante: si $p = \lim \varphi(t_n; y_0)$, por la dependencia continua respecto de los datos (teorema de dependencia respecto de los datos), $\varphi(t; p) = \lim \varphi(t; \varphi(t_n; y_0)) = \lim \varphi(t + t_n; y_0)$, que también es un punto límite. Esta igualdad vale para cada $t \in I(p)$, incluidos los negativos: para n grande, $t_n + t \geq 0$. Por tanto toda la órbita de p permanece en K; la prolongación en un compacto la extiende a ambos sentidos del tiempo.
Paso 3: un destino único es un equilibrio (la observación). Si $\Lambda^+(y_0) = \{\bar{y}\}$ es invariante, la órbita de $\bar{y}$ está contenida en $\{\bar{y}\}$: la solución que sale de $\bar{y}$ es constante, luego $f(\bar{y}) = 0$. ■

Qué hace cada hipótesis. La *compacidad* de K hace todo el trabajo, y lo hace dos veces: en el paso 1 da la vida infinita y en el paso 2 da la existencia de subsucesiones convergentes. Sin ella el conjunto límite puede ser vacío —basta con que la solución se vaya a infinito— y el enunciado se queda sin objeto. Que K esté contenido en el *dominio* es igual de necesario: una solución puede acabarse en tiempo finito acercándose a la frontera del abierto donde f está definida, sin irse a infinito. Y el paso 3 usa además que el sistema es *autónomo*, a través de la invariancia; en un sistema no autónomo el límite de una solución no tiene por qué ser un equilibrio, porque el campo cambia con el tiempo.

Los conjuntos invariantes y los puntos límite, con la proposición de que $\Lambda^+(y_0)$ es compacto, invariante y no vacío cuando la semiórbita positiva vive en un compacto, están en el libro de EDO, §8.3 «Órbitas cíclicas y conjuntos límite. Propiedades» (páginas 170 y 171).

Proposición. Sea $D = \mathbb{R}^N$ y supongamos que, para todo $M > 0$, $\sup\{|\varphi(t; y_0)| : t \in I(y_0) \cap [0, M]\} < +\infty$. Entonces $I(y_0) \supset [0, +\infty)$.

Es la consecuencia práctica: para que una solución viva para siempre basta que no se escape al infinito en tiempo finito, y eso se prueba con cotas como las del teorema de comparación. Es lo que cierra la demostración del teorema de las cotas de la competición simplificada.

Teorema (LaSalle)**.** Sea $K \subset D$ compacto y $V \in C^1(D)$ con $\dot{V}(y) \leq 0$ para todo $y \in K$. Supongamos que existe $y_0 \in K$ con $\gamma^+(y_0) \subset K$, sea $E := \{y \in K : \dot{V}(y) = 0\}$ y sea M el mayor subconjunto invariante contenido en E. Entonces:
1. $I(y_0) \supset [0, +\infty)$;
2. $\Lambda^+(y_0) \subset M$ (y por tanto $M \neq \emptyset$);
3. $\lim_{t \to +\infty} \operatorname{dist}(\varphi(t; y_0), M) = 0$.

Corolario. Sea $\bar{y} \in D$ un equilibrio y $V \in C^1(\bar{B}(\bar{y}; \rho))$ una función de Liapunov, con $\bar{B}(\bar{y}; \rho) \subset$

D, tal que $\{\bar{y}\}$ es el mayor conjunto invariante contenido en $E := \{y \in \bar{B}(\bar{y};\rho) : \dot{V}(y) = 0\}$. Entonces $\bar{y}$ es asintóticamente estable.

LA IDEA: LA ENERGÍA QUE NO BAJA EN TODAS PARTES

Liapunov exigía $\dot{V} < 0$ fuera del equilibrio. LaSalle permite que $\dot{V}$ se anule en un conjunto E, y dice que la solución acaba en la parte de E donde puede *quedarse*: el mayor conjunto invariante de E. En el péndulo con rozamiento $\dot{V} = -c\varphi'^2$ se anula en el eje $\varphi' = 0$; pero una solución no puede quedarse en ese eje salvo en los equilibrios (si $\varphi' = 0$ en un intervalo, entonces $\varphi'' = 0$ y sin $\varphi = 0$). En un entorno del equilibrio inferior que no contenga otros equilibrios, M se reduce a ese punto y el corolario da su estabilidad asintótica. En el SIR reducido a (I, R), con $V = I + R$ se tiene $\dot{V} = \beta IS \geq 0$; con $V = -(I + R)$, $\dot{V} = -\beta IS \leq 0$ y $E = \{I = 0\} \cup \{S = 0\}$, que es la vía del repaso del capítulo 2. Está, con este mismo enunciado, en el libro de EDO, §8.4 «Principio de invariancia y Teorema de LaSalle» (página 172). En esa misma sección, página 174, figura como «Corolario (caso unidimensional)» el teorema de las EDO autónomas escalares con el que abre el capítulo 2, deducido de todo esto en lugar de demostrado directamente.

Para terminar, un resultado sobre el comportamiento a largo plazo de sistemas **planos**, de dimensión 2, que permite deducir periodicidad cuando la semiórbita permanece en un compacto sin puntos críticos.

Teorema (Poincaré-Bendixson)**.** Supongamos $N = 2$. Sea K un subconjunto compacto de D que no contiene puntos críticos, y supongamos que existe $y_0 \in K$ con $\gamma^+(y_0) \subset K$. Entonces $\Lambda^+(y_0)$ es una órbita cíclica no degenerada (no se reduce a un punto) y, o bien $\gamma(y_0)$ coincide con $\Lambda^+(y_0)$, o bien $\gamma^+(y_0)$ converge en espiral a $\Lambda^+(y_0)$.

La hipótesis de ausencia de puntos críticos es esencial. Si el conjunto límite los contiene, puede incluir también órbitas que conectan equilibrios distintos (heteroclínicas) o que salen y regresan al mismo (homoclínicas). Por tanto, el teorema no clasifica toda órbita plana como equilibrio, ciclo o escape al infinito; tampoco cubre una trayectoria que se aproxima a la frontera de D.

En Volterra, cada curva de nivel no trivial $\Gamma(C)$ es compacta, no contiene puntos críticos y está recorrida por la órbita: esta es periódica. La conservación de V impide acercarse desde otro nivel a un ciclo. En competición, descartar ciclos por sí solo no basta para deducir convergencia a un equilibrio; hay que usar además las propiedades del modelo establecidas en el capítulo 2.

CONEXIÓN: POR QUÉ SOLO EN EL PLANO

El teorema usa el teorema de la curva de Jordan: una órbita cerrada en el plano separa un interior de un exterior, y una órbita que entra en el interior no puede salir sin cortarla. En dimensión tres una órbita no separa nada, y aparecen los atractores extraños; el fichero `Lorentz.m` del tema 2 integra el sistema de Lorenz en $\mathbb{R}^3$ y dibuja uno. En el libro de EDO es el teorema del libro de EDO, §8.5 «El caso de los sistemas planos. Teorema de Poincaré-Bendixson» (página 175), con las definiciones de segmento transversal y de acercamiento en espiral que aquí se han dado en palabras.

A.4 El lema de Gronwall

El capítulo 4 usa este lema para probar unicidad de la ecuación de estado y controlar la dependencia respecto del control, y es también lo que hay detrás de la dependencia continua del teorema de dependencia respecto de los datos. El enunciado y la demostración son los del libro de EDO, §3.1 «Lema de Gronwall» (páginas 41 a 44), que allí se escribe $u(x) \le h + \int_{x_0}^{x} k(s)u(s)\,ds \Rightarrow u(x) \le h \exp \int_{x_0}^{x} k(s)\,ds$, con la intuición del interés compuesto en §3.1.1 y la aplicación a dos soluciones con datos distintos en §3.1.2.

Teorema (lema de Gronwall)**.** Sean $u, \alpha : [t_0, t_1] \to \mathbb{R}$ continuas, $\alpha \ge 0$, y $C \ge 0$ una constante, tales que

$$u(t) \le C + \int_{t_0}^{t} \alpha(s)u(s)\,ds \qquad \text{para todo } t \in [t_0, t_1].$$

Entonces

$$u(t) \le C \exp\Big(\int_{t_0}^{t} \alpha(s)\,ds\Big) \qquad \text{para todo } t \in [t_0, t_1].$$

En particular, si $C = 0$ y $u \ge 0$, entonces $u \equiv 0$.

La idea, y por qué el truco es el que es. La hipótesis dice algo sobre u pero lo dice *a través de una integral de* u, y de una integral no se despeja. El movimiento entero de la demostración consiste en dejar de mirar u y mirar en su lugar esa integral, que se llama w: a diferencia de u, la función w es derivable, y su derivada se controla usando la propia hipótesis $u \le w$. Con eso la desigualdad integral, que no se sabía manejar, se ha convertido en una desigualdad diferencial $w' \le \alpha w$, que es exactamente lo que un factor integrante sabe resolver. Y como $u \le w$, la cota que se obtenga para w vale también para u.

Paso 1: mirar la integral en vez de la función (el núcleo). Sea $w(t) := C + \int_{t_0}^{t} \alpha(s)u(s)\,ds$, que es de clase C^1 con $w'(t) = \alpha(t)u(t)$, y por hipótesis $u \le w$. Como $\alpha \ge 0$, $w'(t) = \alpha(t)u(t) \le \alpha(t)w(t)$: aquí se ha usado la hipótesis, y se ha usado en el único sitio donde hace falta.
Paso 2: el factor integrante convierte la desigualdad en monotonía. Multiplicando por el factor integrante positivo $\exp(-\int_{t_0}^{t} \alpha)$,

$$\frac{d}{dt}\Big[w(t)\exp\Big(-\int_{t_0}^{t}\alpha(s)\,ds\Big)\Big] = \big(w'(t) - \alpha(t)w(t)\big)\exp\Big(-\int_{t_0}^{t}\alpha\Big) \le 0,$$

luego la función entre corchetes es decreciente y vale $w(t_0) = C$ en t_0: $w(t)\exp(-\int_{t_0}^{t}\alpha) \le C$.
Paso 3: volver de w **a** u**.** Es decir, $u(t) \le w(t) \le C\exp(\int_{t_0}^{t}\alpha)$. Si $C = 0$ queda $u \le 0$, y con $u \ge 0$, $u \equiv 0$. ■

Qué hace cada hipótesis. La condición $\alpha \ge 0$ es la que permite pasar de $u \le w$ a $\alpha u \le \alpha w$ en el paso 1: con α negativa la desigualdad se invertiría al multiplicar y el argumento se cae. Y el *sentido del intervalo* importa: la conclusión vale para $t \ge t_0$, porque «decreciente» solo acota hacia adelante; hacia atrás la desigualdad se invierte, igual que ocurría con las cotas de la competición del capítulo 2. La versión con $C = 0$, que es la que usan las pruebas de unicidad, es el caso extremo y el más útil: dice que una cantidad no negativa que empieza en cero y crece a lo sumo en proporción a sí

misma no puede despegar nunca. Este lema, con su intuición y sus aplicaciones, está desarrollado en el libro de EDO, §3.1 «Lema de Gronwall» del libro de EDO del autor.

LA IDEA: UNA DESIGUALDAD QUE SE RESUELVE COMO UNA IGUALDAD

Si fuera $u = C + \int \alpha u$, derivando saldría $u' = \alpha u$ y $u = Ce^{\int \alpha}$. Gronwall dice que con $\leq$ en la hipótesis se conserva el $\leq$ en la conclusión: una función que crece como mucho en proporción a sí misma crece como mucho exponencialmente. Aplicado a $w = y - z$, diferencia de dos soluciones de $y' = ay + v$ con el mismo dato, se tiene $|w(t)| \leq \int_0^t |a||w(s)|\,ds$, con $C = 0$: luego $w \equiv 0$, que es la unicidad del capítulo 4. Y con datos distintos, $|w(t)| \leq |y_0 - z_0| + \int |a||w|$, que da $|w(t)| \leq |y_0 - z_0|e^{|a|t}$: la dependencia continua.

APÉNDICE B

Qué cae en el examen

Este apéndice orienta la preparación de Modelización Matemática en la Universidad de Sevilla a partir de las instrucciones y pruebas recogidas en el apéndice C. El resto del libro puede estudiarse sin esta guía.

B.1 Cómo se califica en 2026-27

Según los proyectos docentes de 2026-27, idénticos en los cuatro grupos, **todo examen —parcial o de convocatoria— tiene una parte teórica y una parte práctica de programación, de** 5 **puntos cada una**.

- **En convocatoria** se aprueba con 5 **y al menos** 2 **sobre** 5 **en cada parte**; si falta alguna de las dos condiciones, la nota es el mínimo entre 4 y la del examen. La programación no es opcional: sin ese 2 se suspende aunque la teoría esté perfecta.
- **Por curso** hay dos pruebas parciales de seguimiento. Cada una necesita al menos un 4, con 2 sobre 5 en cada parte, y la media de las dos tiene que llegar a 5. Quien apruebe solo una puede presentarse en la primera convocatoria únicamente a la parte suspensa.

B.2 El archivo de exámenes

Los exámenes de convocatorias anteriores están resueltos en el apéndice C: la primera convocatoria de 2026 (de la que solo se conserva un ejercicio), el examen final del 2 de junio de 2025 —el único completo—, la primera prueba de seguimiento del 5 de abril de 2024, la segunda convocatoria del 13 de julio de 2023, los primeros parciales del 29 de marzo de 2023 y del 28 de marzo de 2022, y una colección de ejercicios sueltos de los cursos 2015-16 y 2016-17. Allí van también las instrucciones del examen y un simulacro de repaso elaborado para el libro.

El formato observado. El examen final completo de 2025 contiene **cuatro bloques de cinco puntos**: dos de teoría y dos de prácticas con MATLAB. En esa prueba, la parte práctica

representa la mitad de la nota. Los primeros parciales del archivo reparten teoría $2 + 3$ y prácticas 5.

Preguntas recurrentes. En la parte teórica del archivo aparecen estas tareas:

1. **Adimensionalizar**: se da un modelo con cuatro o cinco constantes y se pide un cambio de unidades que lo deje con una o dos. Aparece en 2025, 2024, 2023 (las dos convocatorias), 2022 y en los cuadernos de 2016-17. Se explica paso a paso, empezando por la logística y con los errores habituales, en el modelo de Volterra (sección 2.2.4); la simbiosis (sección 2.2.7) lo repite con otra escritura.
2. **Equilibrios y primera aproximación**: calcularlos y clasificarlos, diciendo cuándo el método no decide. Se hace con el teorema de linealización del apéndice A; el modelo de Volterra (sección 2.2.4) muestra el caso en que no decide (autovalores imaginarios puros), y la competición general y la simbiosis (secciones 2.2.6 y 2.2.7), los demás.
3. **Invariancia de los ejes y positividad**: probar que las soluciones con dato positivo se mantienen positivas. Hay dos vías: escribir cada ecuación como lineal en su propia incógnita, como en el SIR con mortalidad (sección 2.2.2), o usar que los ejes son órbitas, como en la simbiosis (sección 2.2.7).
4. **Prolongación de la solución maximal** a $[0, +\infty)$ mediante cotas a priori, y los límites cuando $t \to +\infty$. El esquema completo, con su resumen, está en el teorema de las cotas de la competición (sección 2.2.5), y se repite en el SIR con mortalidad (sección 2.2.2).

Y en la parte de prácticas: resolver un sistema con `ode45`, dibujar trayectorias y órbita, localizar un instante concreto con `fzero` o `fminbnd`, comparar con una solución exacta y medir el error, y **un problema inverso por mínimos cuadrados** con `fminsearch`, que sale en 2017 y en 2025. La teoría, con el caso de una tasa de crecimiento resuelto a mano, está en la sección 4.9, y las prácticas 24 a 26 lo programan.

Las demostraciones. El libro no marca una lista de teoremas para memorizar. En las pruebas reales recogidas se pide aplicar resultados y demostrar propiedades de modelos concretos: existencia y unicidad, convexidad de un funcional o condiciones de optimalidad. Las demostraciones del texto sirven para comprender esos argumentos y reconocer sus hipótesis; la muestra de exámenes no permite excluir preguntas futuras.

B.3 La parte práctica, en el aula de informática

Las instrucciones de examen reproducidas en el apéndice C establecen:

- Se hace **con MATLAB y solo con MATLAB**. No se puede usar ningún otro programa.
- **La ayuda de MATLAB sí está permitida**, y conviene acostumbrarse a ella: quien sabe leer `help polyfit` no necesita recordar el orden de los argumentos.
- Se crea en el escritorio una carpeta con los apellidos, **sin espacios ni acentos** (`GarciaGarcia` vale; `García García` no), y se selecciona como carpeta actual antes de ejecutar nada.
- Cada programa que se entrega empieza con una línea de comentario con el nombre.
- Al terminar se comprime la carpeta con su mismo nombre.

Conviene practicar desde el comienzo con nombres de archivo válidos y una carpeta de trabajo bien definida, como se explica en la sesión cero.

B.4 Qué hacer con ello

Las prácticas incluyen código y resultados de referencia. Para preparar la prueba, puede resolverse una selección de las hojas a libro cerrado, con tiempo limitado y usando MATLAB; después se compara el procedimiento y se corrigen los errores.

Los resultados numéricos dependen del método, las tolerancias y el problema. Una discrepancia se comprueba reduciendo las tolerancias y contrastando el residuo, los invariantes o una solución exacta, cuando exista. No toda diferencia en las últimas cifras es un error, ni toda diferencia pequeña es aceptable.

APÉNDICE C

Exámenes resueltos

Este apéndice reúne un simulacro y exámenes de convocatoria con sus resoluciones. Las fichas indican la fecha conocida, la puntuación y qué partes se conservan; los exámenes reales se ordenan de más reciente a más antiguo. Esa muestra sirve para reconocer tipos de problemas, no para asegurar el contenido de una convocatoria futura.

Cómo usar las resoluciones. Para comprobar lo que ya sabes, intenta la prueba cronometrada y sin consultar la solución, y después corrige cada apartado. Para aprender un método nuevo, puedes estudiar primero un ejemplo resuelto: explica por qué es válido cada paso y resuelve después otro problema sin mirarlo. Ambos usos son útiles; conviene distinguir aprender un procedimiento de comprobar que puedes reconstruirlo.

La parte de MATLAB. En el examen completo del 2 de junio de 2025 hay cuatro bloques de cinco puntos: dos de teoría y dos de prácticas. Aquí las prácticas llevan M-funciones completas y comentadas. Para ensayarlas, guárdalas en una carpeta de trabajo y compara la salida con la interpretación matemática, además de comprobar que el programa termina.

Las fuentes. Se conservan enunciados impresos y cuadernos que los acompañan con resoluciones manuscritas. Las soluciones que siguen desarrollan los razonamientos desde los datos de los enunciados. Cuando falta una parte del examen o no consta su fecha exacta, la ficha lo indica. El simulacro está identificado expresamente como una propuesta del libro.

C.1 Instrucciones oficiales

Resumen del documento conservado «Instrucciones y consejos para la realización de los exámenes». Describe el procedimiento del curso; los datos concretos de entrega son los indicados para cada convocatoria.

Generales. Se espera en la puerta del aula y se siguen las indicaciones sobre dónde sentarse. No se puede llevar ningún dispositivo electrónico; los móviles, apagados y en la mochila, aparte. Nombre y apellidos con letra clara arriba a la derecha de todas las hojas. No presuponer, tras una lectura rápida, lo que se pregunta: leer varias veces. La puntuación

indicada en cada pregunta es la *máxima*: una respuesta escasa, escueta o vaga recibe una nota escasa, escueta o vaga; que no se entienda la letra, que el desarrollo esté desordenado o que los resultados importantes no estén resaltados resta. Al terminar, salir con la menor distracción posible.

Específicas del examen de MATLAB.

1. Crear en el **escritorio** una carpeta con los **apellidos**, sin espacios, acentos, eñes ni puntos (`GarciaGarcia` sí; `Garcia Garcia`, `MunozBaza` con eñe, `Gomez.Valle`, `CarlosBaras` no), y seleccionarla como carpeta de trabajo en MATLAB.
2. No se puede usar ningún otro material del ordenador ni ningún otro programa aparte de MATLAB; hacerlo invalida el examen. **Sí se puede usar la ayuda de MATLAB.**
3. En todos los programas entregados debe aparecer el nombre en líneas de comentario, **al comienzo** del programa, no al final.
4. Al terminar, comprimir la carpeta con su mismo nombre y depositarla mediante el procedimiento de entrega del aula. El documento describe la transferencia con FileZilla; el servidor y la cuenta se toman de las indicaciones de la convocatoria. Comprobar antes de irse que el archivo comprimido se ha entregado.

La idea: usar la ayuda con una pregunta concreta

La ayuda de MATLAB permite consultar los argumentos de `bvp4c` o de `fzero`; lo que hay que decidir es qué problema se está resolviendo y qué función corresponde usar. Por ejemplo, `fzero` busca un cero de una función escalar: antes de llamarla hay que expresar la cantidad cuyo cero interesa. En papel, el mismo criterio de claridad pide que las hipótesis, el razonamiento y la conclusión puedan distinguirse al leer.

C.2 Simulacro de examen

Este simulacro es una propuesta de repaso del libro, no una convocatoria real. Tiene dos preguntas teóricas y dos ejercicios de MATLAB, con cinco puntos y 75 minutos para cada parte: diez puntos y dos horas y media en total. Su escala difiere de los veinte puntos del examen final de junio de 2025. Para usarlo como ensayo, respeta los tiempos y corrígelo al terminar; las M-funciones se guardan en una carpeta de trabajo, con el nombre al comienzo.

Enunciado

Parte teórica (5 puntos, 75 minutos).

Pregunta 1 (2.5 puntos). Se considera el modelo SIS con población total constante N, tasa de contagio $\beta > 0$ y tasa de recuperación $\gamma > 0$.

a) (0.5) Escribir el sistema para (I, S) y reducirlo a una única EDO para I.

b) (1) Enunciar el teorema sobre el comportamiento de las soluciones de una EDO autónoma escalar $y' = f(y)$, y usarlo para describir el comportamiento de $I(t; I_0)$ para $I_0 \in (0, N)$ en los casos $R_0 \leq 1$ y $R_0 > 1$, con $R_0 = N\beta/\gamma$.

c) (1) Demostrar, en el caso $f(y_0) > 0$ del teorema, que φ es estrictamente creciente y que, si existe un punto crítico mayor que y_0, la solución está definida en $[0, +\infty)$ y converge al menor de ellos.

Pregunta 2 (2.5 puntos). Sea A la matriz de conectividad de una red de n páginas sin nodos sin salida y P su matriz de transición.

a) (1) Probar que P es estocástica por columnas, que 1 es autovalor y que $\rho(P) = 1$.

b) (1) Definir la matriz de Google $\mathcal{P}$ con factor $\alpha \in (0,1)$ y $w = v = \frac{1}{n}e$, explicar por qué se le aplica el teorema de Perron, y deducir la fórmula $x^{k+1} = \alpha P x^k + \gamma_k v$ con $\gamma_k = 1 - \|\alpha P x^k\|_1$.

c) (0.5) Para la red de tres páginas con enlaces $1 \to 2$, $2 \to 3$, $3 \to 1$ y $3 \to 2$, escribir A y P y calcular el PageRank sin amortiguación.

Parte práctica de MATLAB (5 puntos, 75 minutos).

Ejercicio 1 (2.5 puntos). Una población sigue el modelo logístico con pesca constante, $p' = \lambda p - \kappa p^2 - h$, con $\lambda = 0.1$ y $\kappa = 0.01$. Escribir una M-función `PescaConstante(h, p0, T)` que:

a) calcule los equilibrios (si existen) y los escriba en pantalla;

b) resuelva numéricamente el modelo en $[0, T]$ desde $p(0) = p_0$, deteniéndose si la población se anula, y dibuje la solución junto con los equilibrios;

c) calcule, si existe, el instante en que la población cae por debajo de la mitad del dato inicial.

Ejecutarla con $h = 0.2$, $p_0 = 3$, $T = 40$ y con $h = 0.2$, $p_0 = 2.5$, $T = 40$, y explicar en un comentario lo que se observa a la vista de la teoría.

Ejercicio 2 (2.5 puntos). Una cuerda elástica de rigidez $\mu = 2$ ocupa $(0,1)$, está sometida a la fuerza vertical $F(x) = -5\sin(\pi x)$ y tiene el extremo izquierdo fijo, $u(0) = 0$, y el derecho sujeto a un muelle de constante $k = 4$. Su desplazamiento minimiza $\int_0^1 \left(\frac{\mu}{2}u'^2 - Fu\right)dx + \frac{k}{2}u(1)^2$.

a) Escribir en un comentario el problema de contorno que verifica u.

b) Escribir una M-función `CuerdaMuelle(mu, k, F0)` que lo resuelva con `bvp4c` para $F(x) = F_0\sin(\pi x)$ y dibuje u.

c) Comprobar el resultado con la solución exacta, y escribir el desplazamiento del extremo derecho en pantalla.

Resolución

Pregunta 1.

a) $I' = \beta IS - \gamma I$, $S' = -\beta IS + \gamma I$. Sumando, $(I+S)' = 0$, luego $S = N - I$ y

$$I' = \beta I(N - I) - \gamma I = \beta I\big((N - \gamma/\beta) - I\big) =: f(I).$$

b) **Teorema.** Sea φ la solución maximal de $y' = f(y)$, $y(0) = y_0$, con $f \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$, definida en $I(y_0)$. Si $f(y_0) = 0$, y_0 es un equilibrio y $\varphi \equiv y_0$ en $\mathbb{R}$. Si $f(y_0) > 0$, φ es estrictamente creciente; si existe un punto crítico mayor que y_0, $I(y_0) \supset [0, +\infty)$ y $\varphi(t) \to \bar{y}$, el menor de ellos; si no existe, $\varphi(t) \to +\infty$ cuando $t \to \sup I(y_0)$, finito o no. Si $f(y_0) < 0$, lo simétrico con decreciente y el mayor punto crítico menor que y_0.

Aplicación. Los ceros de f son $I = 0$ e $I = N - \gamma/\beta$, y f es una parábola hacia abajo con esos ceros. Si $R_0 \le 1$, es $N - \gamma/\beta \le 0$: el único cero en $[0, N]$ es 0 y $f < 0$ en $(0, N)$: en $f(I) = \beta I \cdot ((N - \gamma/\beta) - I)$ el primer factor es positivo y el segundo negativo. Luego para $I_0 \in (0, N)$, $f(I_0) < 0$, la solución es estrictamente decreciente, existe el punto crítico $0 < I_0$, está definida en $[0, +\infty)$ y tiende a 0: la epidemia desaparece. Si $R_0 > 1$, $\bar{I}_2 = N - \gamma/\beta \in (0, N)$; para $I_0 < \bar{I}_2$ es $f(I_0) > 0$ y la solución crece hacia el menor crítico por encima, $\bar{I}_2$; para $I_0 = \bar{I}_2$ es constante; para $I_0 > \bar{I}_2$ es $f(I_0) < 0$ y decrece hacia el mayor crítico por debajo, $\bar{I}_2$. En todos los casos $I \to N - \gamma/\beta$: endémica.

c) Como en la sección 2.1: dos órbitas no se cortan y los equilibrios son órbitas de un punto, luego $\varphi(t)$ nunca es crítico y $\varphi'(t) = f(\varphi(t)) \neq 0$; por continuidad y $\varphi'(0) > 0$, $\varphi' > 0$ en todo $I(y_0)$ (Bolzano). Si $\tilde{y} > y_0$ es crítico, φ no puede tomar el valor $\tilde{y}$, luego $y_0 \leq \varphi(t) < \tilde{y}$ para $t \geq 0$: la semiórbita positiva está en el compacto $[y_0, \tilde{y}]$, y por la proposición del compacto (apéndice A) $I(y_0) \supset [0, +\infty)$. Monótona y acotada, tiene límite $L \leq \tilde{y}$, y un límite es un punto crítico (observación del apéndice A). Es el menor por encima de y_0: si hubiera otro entre y_0 y L, la solución tendría que cruzarlo. (Se le llama L y no β a propósito: en esta pregunta β es la tasa de contagio del modelo SIS y no conviene gastar la letra dos veces.)

Qué buscaba. La pregunta evalúa el teorema central del capítulo 2 y su uso en el modelo más simple; el atajo es reconocer que f es una parábola y leer los signos sin resolver nada. *Error típico*: analizar el caso $R_0 < 1$ diciendo que $\bar{I}_2$ es inestable «pero negativo», sin concluir a dónde va I.

Pregunta 2.

a) Cada columna de A tiene $s_j > 0$ unos y P se obtiene dividiendo la columna j por s_j: entradas ≥ 0 y columnas de suma 1. Con $e = (1, \ldots, 1)^t$, «columnas de suma 1» es $e^t P = e^t$, luego $P^t e = e$ y 1 es autovalor de P^t, que tiene el mismo polinomio característico que P. Por Gershgorin aplicado a P^t (discos centrados en p_{jj} de radio $1 - p_{jj}$, contenidos en el disco unidad), todo autovalor tiene módulo ≤ 1. Luego $\rho(P) = 1$.

b) $\mathcal{P} = \alpha P + (1-\alpha)\frac{1}{n} e e^t$ (sin nodos sin salida, $d = 0$). Cada entrada es $\geq (1-\alpha)/n > 0$: $\mathcal{P}$ es estrictamente positiva y se le aplica el teorema de Perron: 1 es autovalor simple estrictamente dominante, con un único autovector estocástico x_∞, y, desde un vector estocástico x^0, el método de la potencia $x^{k+1} = \mathcal{P} x^k$ converge a él. Para x^k estocástico, $e^t x^k = 1$, luego $\mathcal{P} x^k = \alpha P x^k + (1-\alpha) v$ con $v = e/n$; y como P es estocástica, $\|Px^k\|_1 = 1$, de modo que $\gamma_k := 1 - \|\alpha P x^k\|_1 = 1 - \alpha$ en este caso. (Con nodos sin salida, $\gamma_k = \alpha(1 - \|Px^k\|_1) + (1-\alpha)$ recoge además la probabilidad de estar en un nodo sin salida, y la fórmula es la misma.)

c) Enlaces $1 \to 2$, $2 \to 3$, $3 \to 1$, $3 \to 2$: $a_{ij} = 1$ si $j \to i$, luego

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad P = \begin{pmatrix} 0 & 0 & 1/2 \\ 1 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}.$$

$Px = x$: $x_1 = x_3/2$, $x_2 = x_1 + x_3/2$, $x_3 = x_2$. De la primera y la tercera, $x_1 = x_2/2$; la segunda, $x_2 = x_2/2 + x_2/2$, se cumple. Con $x_2 = 2$: $x = (1, 2, 2)$, y normalizando, $x_P = (1/5, 2/5, 2/5)$. La página 1, que solo recibe medio voto de la 3, es la menos relevante; las páginas 2 y 3 empatan aunque la 2 reciba dos enlaces y la 3 uno, porque el de la 3 viene entero de la 2.

Qué buscaba. Las dos ideas del PageRank: por qué el problema es de autovectores y por qué la amortiguación hace aplicable a Perron; y una cuenta pequeña. *Error típico*: escribir A con las flechas al revés (fila origen, columna destino); la regla es «columna origen, fila destino».

Ejercicio 1 de MATLAB.

Teoría que hay que tener delante. Es la sección 2.1.4: equilibrios $p = \frac{\lambda \pm \sqrt{\lambda^2 - 4\kappa h}}{2\kappa}$; con $h = 0.2$, $\lambda^2 - 4\kappa h = 0.002 > 0$ y $\bar{p}_1 \approx 2.7639$, $\bar{p}_2 \approx 7.2361$. Desde $p_0 = 3 > \bar{p}_1$ la población crece hacia 7.24 y nunca cae a la mitad; desde $p_0 = 2.5 < \bar{p}_1$ se extingue en tiempo finito y cae a 1.25 en un instante que hay que calcular.

```
function PescaConstante(h, p0, T)
% Nombre Apellido Apellido
% PescaConstante(h, p0, T): p' = lambda p - kappa p^2 - h, lambda = 0.1, kappa = 0.01.
% Pruebas: PescaConstante(0.2, 3, 40) PescaConstante(0.2, 2.5, 40)
```

```
lambda = 0.1; kappa = 0.01;
if h < 0 || p0 <= 0 || T <= 0
  error('Se requiere h >= 0, p0 > 0 y T > 0')
end

% (a) Equilibrios; tolerancia para reconocer el caso critico
hcrit = lambda^2/(4*kappa);
tol = 100*eps(hcrit);
disc = lambda^2 - 4*kappa*h;
if h > hcrit + tol
  disp('No hay equilibrios: la poblacion se extingue en tiempo finito para todo p0')
  eq = [];
elseif abs(h-hcrit) <= tol
  eq = lambda/(2*kappa);
  fprintf('Equilibrio critico p = %8.4f: atrae por arriba, repele por abajo\n', eq)
else
  eq = (lambda + [-1, 1]*sqrt(disc))/(2*kappa);
  fprintf('Equilibrios: p1 = %8.4f (inestable), p2 = %8.4f (estable) \n', eq)
end

% (b) Resolucion; se detiene si p llega a 0 (evento de ode45)
f = @(t,p) lambda*p - kappa*p.^2 - h;
opciones = odeset('Events', @(t,p) deal(p, 1, -1)); % valor p, terminal, direccion
    decreciente
[times, p, te] = ode45(f, [0, T], p0, opciones);
if isempty(te) && times(end) < T
  error('Integracion incompleta: no se ha detectado la extincion')
end
close all
plot(times, p, 'b', 'LineWidth', 1.5), hold on
for j = 1:numel(eq)
  plot([0, T], [eq(j), eq(j)], 'r--')
end
plot(0, p0, 'r.', 'MarkerSize', 20)
xlabel('t'), ylabel('p(t)'), title(['Pesca constante h = ', num2str(h), ', p(0) = ',
    num2str(p0)])
grid on

% (c) Instante en que p cae por debajo de p0/2
if min(p) <= p0/2
  tmitad = fzero(@(t) interp1(times, p, t) - p0/2, [times(1), times(end)]);
  plot(tmitad, p0/2, 'k.', 'MarkerSize', 20)
  fprintf('La poblacion cae a la mitad en t = %8.4f \n', tmitad)
else
  disp('La poblacion no cae a la mitad del dato inicial en [0,T]')
end
if ~isempty(te)
  fprintf('La poblacion se extingue en t = %8.4f \n', te(end))
end
shg
end
```

Órdenes que no habían aparecido y se explican en el examen si se usan. `odeset('Events', g)` hace que `ode45` vigile los ceros de la función `g(t,p)` y se detenga en ellos; `g` devuelve tres cosas, el valor cuyo cero se busca (p), si el evento es terminal (1) y en qué dirección se cruza (-1: decreciente); `deal` reparte tres valores en tres salidas. La tercera salida de `ode45`,

`te`, contiene los instantes detectados: permite distinguir una extinción de una integración interrumpida por un fallo numérico. Recortar después los valores negativos sirve para el dibujo, pero no localiza por sí solo el instante de extinción; además, prolongar la solución negativa puede hacer fallar al integrador.

El caso crítico $h = \lambda^2/(4\kappa) = 0{,}25$ tiene un solo equilibrio, $p = 5$: $f(p) = -\kappa(p-5)^2$, por lo que atrae desde arriba y repele desde abajo. La tolerancia `100*eps(hcrit)` evita que el redondeo numérico lo clasifique como dos raíces separadas. Para los datos del enunciado, la salida es la siguiente (el instante se muestra redondeado):

```
>> PescaConstante(0.2, 3, 40)
Equilibrios: p1 = 2.7639 (inestable), p2 = 7.2361 (estable)
La poblacion no cae a la mitad del dato inicial en [0,T]
>> PescaConstante(0.2, 2.5, 40)
Equilibrios: p1 = 2.7639 (inestable), p2 = 7.2361 (estable)
La poblacion cae a la mitad en t = 33.8
```

Una comprobación independiente viene de separar variables. Para $p_0 < \bar{p}_1$,

$$t(p) = \frac{1}{\kappa(\bar{p}_2 - \bar{p}_1)} \log\left(\frac{(p-\bar{p}_1)(p_0-\bar{p}_2)}{(p-\bar{p}_2)(p_0-\bar{p}_1)}\right).$$

Con $p_0 = 2{,}5$, da $t(1{,}25) \approx 33{,}8216$ y $t(0) \approx 43{,}0409$; además, $p(40) \approx 0{,}5272$. En $[0, 40]$ todavía no se ha extinguido; para observarlo puede usarse $T = 60$. El cero que encuentra `fzero` se calcula sobre una interpolación de la solución numérica: su precisión depende de esa interpolación y de las tolerancias de integración. Cerca del equilibrio inestable, $f'(\bar{p}_1) > 0$, las perturbaciones pequeñas crecen inicialmente.

El comentario que pide el enunciado: con p_0 a un lado y a otro de $\bar{p}_1 = 2.76$ el destino es opuesto, crecimiento hacia $\bar{p}_2$ o extinción en tiempo finito, como predice el teorema de las EDO autónomas con los signos de f; y $\bar{p}_1$ es el umbral de la pesca constante.

Qué buscaba. Saber usar la teoría antes de programar (los equilibrios y su estabilidad), integrar con `ode45`, dibujar, y contestar a una pregunta con `fzero` sobre la solución interpolada. *Error típico*: llamar a `fzero` sin comprobar que el nivel se cruza; o no detener la integración y dejar que p se haga muy negativa, con lo que el dibujo no se entiende.

Ejercicio 2 de MATLAB.

a) El problema de contorno. Como en el ejemplo de la cuerda sujeta a muelles, se toma una variación h con $h(0) = 0$ y $h(1)$ libre. Si J es el funcional del enunciado, $\frac{d}{d\varepsilon} J(u+\varepsilon h)\Big|_{\varepsilon=0} = \int_0^1 (\mu u' h' - Fh)\,dx + ku(1)h(1) = 0$; integrando por partes, $\int_0^1 (-\mu u'' - F)h\,dx + (\mu u'(1) + ku(1))h(1) = 0$; con $h(1) = 0$ sale la EDO y después la condición en 1:

$$-\mu u'' = F(x) \text{ en } (0,1), \qquad u(0) = 0, \qquad \mu u'(1) + ku(1) = 0.$$

La solución del contorno es además el mínimo único: al expandir el funcional, la primera variación se anula y queda $J(u+h) - J(u) = \frac{\mu}{2}\int_0^1 (h')^2\,dx + \frac{k}{2}h(1)^2 > 0$ para toda variación $h \neq 0$ con $h(0) = 0$.

Solución exacta para comprobar: $u'' = -\frac{F_0}{\mu}\sin(\pi x)$ da $u = \frac{F_0}{\mu\pi^2}\sin(\pi x) + c_1 x + c_2$; $u(0) = 0$ da $c_2 = 0$; y $\mu u'(1) + ku(1) = \mu(-\frac{F_0}{\mu\pi} + c_1) + kc_1 = 0$ da $c_1 = \frac{F_0}{\pi(\mu+k)}$. Con $\mu = 2$, $k = 4$, $F_0 = -5$: $u(1) = c_1 = -5/(6\pi) \approx -0.2653$.

```
1 function CuerdaMuelle(mu, k, F0)
2  % Nombre Apellido Apellido
```

```
% CuerdaMuelle: -mu u'' = F0 sin(pi x) en (0,1), u(0) = 0, mu u'(1) + k u(1) = 0 (Robin).
% Prueba: CuerdaMuelle(2, 4, -5)
F = @(x) F0*sin(pi*x);
dzdx = @(x,z) [z(2); -F(x)/mu]; % z = [u; u']
condcon = @(za,zb) [za(1); mu*zb(2) + k*zb(1)]; % Dirichlet en 0, Robin en 1
sol = bvp4c(dzdx, condcon, bvpinit(linspace(0, 1, 10), [0; 0]));

xx = linspace(0, 1);
uu = deval(sol, xx, 1);
c1 = F0/(pi*(mu + k));
exacta = F0/(mu*pi^2)*sin(pi*xx) + c1*xx; % solucion exacta
close all
plot(xx, uu, 'b', 'LineWidth', 2), hold on
plot(xx(1:8:end), exacta(1:8:end), 'ro')
plot([0, 1], [0, 0], 'k')
legend('bvp4c', 'Exacta', 'Sin deformar'), xlabel('x'), ylabel('u(x)')
fprintf('Desplazamiento del extremo derecho: u(1) = %10.6f (exacto %10.6f) \n', uu(end),
    c1)
fprintf('Error maximo frente a la exacta: %10.2e \n', max(abs(uu - exacta)))
shg
end
```

```
>> CuerdaMuelle(2, 4, -5)
Desplazamiento del extremo derecho: u(1) = -0.265258 (exacto -0.265258)
```

El programa también muestra el error máximo sobre los puntos de `xx`; su valor numérico depende de las tolerancias de `bvp4c` y de la malla que construya. La fórmula exacta permite comprobarlo en cada ejecución, sin suponer de antemano una precisión cercana a la de máquina.

Qué buscaba. Deducir una condición de Robin desde el funcional (el mecanismo del capítulo 3), traducirla a `condcon`, y comprobar con la exacta. *Error típico*: escribir la condición de Robin con el signo de u' cambiado (en el extremo izquierdo es $-\mu u'(0) + ku(0)$, en el derecho $\mu u'(1) + ku(1)$: el signo viene de la integración por partes), o poner la condición de Robin en `za` en vez de `zb`.

Cómo corregirse

Cada apartado se puntúa como indica el enunciado, con el criterio de las instrucciones oficiales: una respuesta que no escribe el enunciado del teorema con sus hipótesis, o que programa algo que funciona pero no dice por qué tiene que salir lo que sale, no tiene la nota máxima. Los fallos se anotan con el nombre del ejercicio y la fecha, y se vuelve a ellos en la siguiente sesión de prácticas cronometradas.

C.3 Primera convocatoria de 2026: un modelo SIR con mortalidad

De esta convocatoria se conserva solo el ejercicio 1, la primera parte de teoría. La fuente es un cuaderno con el enunciado impreso y su resolución manuscrita. Lleva el rótulo «Primera convocatoria 2026», pero no consta la fecha exacta ni el resto del examen.

Ficha

Asignatura	Modelización matemática
Convocatoria	Primera, curso 2025-26
Fecha	No consta en la fuente
Lo que se conserva	Solo el ejercicio 1: «Teoría: Primera Parte, 5 puntos»
Resto del examen	No consta en esta fuente

Una advertencia de notación. El enunciado llama $\mathcal{R}_0 = \beta_f/\gamma$ al número básico de reproducción y, a la vez, $R_0 = R(0)$ al número inicial de recuperados. Son dos cosas distintas con la misma letra, y conviene tenerlo presente al leer el apartado (d). En la sección 2.1.2 del capítulo 2 el libro escribe R_0, sin caligrafía, para el número básico de reproducción del modelo SIS: es el mismo objeto que el $\mathcal{R}_0$ de aquí. Se conserva la letra caligráfica del examen para no confundirla con $R(0)$.

Enunciado

Ejercicio 1 [Teoría: Primera Parte, 5 puntos]. Se considera el problema de Cauchy para un modelo SIR donde una proporción de los infectados no se recuperan y mueren por la enfermedad:

$$\begin{cases} S' = -\dfrac{\beta_f}{S+I+R}\, SI, \\ I' = \dfrac{\beta_f}{S+I+R}\, SI - \gamma I, \\ R' = m\gamma I, \\ S(0) = S_0, \quad I(0) = I_0, \quad R(0) = R_0, \end{cases} \tag{C.1}$$

donde $\beta_f, \gamma \in (0,\infty)$, $m \in (0,1)$, $S_0, I_0, R_0 \in [0,\infty)$ y $N_0 := S_0 + I_0 + R_0 > 0$. La constante $1-m$ está relacionada con los infectados que no se recuperan y mueren. Se usa la notación $N = S + I + R$ para la población total en el instante t (que ahora **no es constante**). En estas condiciones, se pide:

a) Calcular los dominios maximales de existencia y unicidad del SDO en (C.1) y demostrar que el problema de Cauchy (C.1) posee una única solución maximal $(S(\cdot), I(\cdot), R(\cdot))$ definida en el intervalo $J(S_0, I_0, R_0)$ (con $S_0, I_0, R_0 \in [0,\infty)$ y $N_0 := S_0 + I_0 + R_0 > 0$). Calcular los puntos críticos de (C.1).
   En lo que sigue se supone además que $S_0, I_0 > 0$ y $R_0 \geq 0$. Entonces,

b) Demostrar que

$$S(t) > 0, \qquad I(t) > 0, \qquad \forall t \in J(S_0, I_0, R_0),$$

   que S y N son estrictamente decrecientes en $J(S_0, I_0, R_0)$, que R es estrictamente creciente en $J(S_0, I_0, R_0)$ y que $N(t) > 0$ para cualquier $t \in J(S_0, I_0, R_0)$.

c) Como consecuencia, demostrar que $J(S_0, I_0, R_0) \supset [0,\infty)$ y que existen los siguientes límites

$$\lim_{t\to\infty} S(t) = S_\infty \geq 0, \qquad \lim_{t\to\infty} R(t) = R_\infty > 0,$$
$$\lim_{t\to\infty} N(t) = N_\infty > 0 \qquad \text{y} \qquad \lim_{t\to\infty} I(t) = 0.$$

d) Si $m = 1$ y $\dfrac{S_0}{N_0} > \dfrac{1}{\mathcal{R}_0}$, con $\mathcal{R}_0 = \dfrac{\beta_f}{\gamma} \in (0, \infty)$ (número básico de reproducción), probar que existe un único tiempo $T_{max} \in (0, \infty)$ tal que $I(t)$ es estrictamente creciente en $(0, T_{max})$ y estrictamente decreciente en (T_{max}, ∞).

Qué modelo es este, antes de tocar nada

Merece la pena situarlo, porque los cuatro apartados son variaciones sobre el SIR de la sección 2.2.1 y todo el trabajo consiste en ver qué se rompe y qué no. Hay **dos** diferencias con aquel, y las dos importan.

Primera: cómo se cuentan los contagios. En el capítulo 2 el término de contagio es βSI, que se llama *incidencia de masa*. Detrás está la hipótesis de que cada individuo se cruza con un número de personas *proporcional al tamaño de la población*: si la ciudad duplica sus habitantes, cada uno tiene el doble de encuentros. Aquí el término es

$$\frac{\beta_f}{N} SI = \beta_f S \cdot \frac{I}{N},$$

que se llama *incidencia de frecuencia*, y la hipótesis es la contraria: cada individuo tiene un número *fijo* de contactos por unidad de tiempo, y una fracción I/N de ellos son infecciosos. Por eso la constante β_f tiene unidades de $1/\text{tiempo}$, mientras que β las tenía de $1/(\text{individuo} \cdot \text{tiempo})$. Para una población que no cambia de tamaño las dos formulaciones son la misma con constantes distintas ($\beta = \beta_f/N$); en cuanto la población varía, dejan de serlo, y aquí varía.

Segunda: la población no se conserva. Solo una proporción m de los infectados llega a R; la proporción $1 - m$ restante muere y sale del sistema. Eso es lo que hace $R' = m\gamma I$ en vez de $R' = \gamma I$, y es lo que hace que N no sea constante.

LA IDEA: EL HILO DE LOS CUATRO APARTADOS, EN CUATRO FRASES

(a) El campo tiene un denominador, así que lo primero es decir dónde está definido: en el semiespacio abierto $N > 0$ que contiene los datos iniciales. La región biológica exige además $S, I, R \geq 0$. (b) S e I aparecen como *factores* en sus propias ecuaciones, luego cada una de ellas es una ecuación lineal escalar disfrazada, y una ecuación lineal escalar homogénea conserva el signo de su dato inicial. (c) Hay una combinación de N y R que se conserva; de ella salen todas las cotas, las cotas encierran la órbita en un compacto y el compacto da la existencia global y los límites. (d) El signo de I' es el signo de una función estrictamente decreciente del tiempo, luego cambia una vez y solo una: ese cambio es el pico de la epidemia.

Resolución del apartado (a)

La idea. Un problema de Cauchy no se plantea en el vacío: se plantea en un abierto donde el campo esté definido y sea localmente lipschitziano. Aquí el campo es una función racional cuyo único denominador es $S + I + R$, así que el trabajo entero de este apartado es *mirar dónde se anula ese denominador* y quedarse con la componente conexa donde vive el dato inicial.

Paso 1: dónde está definido el campo, y cuáles son los dominios maximales. El segundo miembro de (C.1) es

$$f(S, I, R) = \Big(-\frac{\beta_f SI}{S+I+R},\ \frac{\beta_f SI}{S+I+R} - \gamma I,\ m\gamma I\Big),$$

una función racional de (S, I, R) con denominador $S + I + R$. Luego f está definida y es de clase C^∞ exactamente en

$$\mathbb{R}^3 \setminus \{(S, I, R) : S + I + R = 0\},$$

el complementario de un plano. Ese abierto tiene **dos componentes conexas**, los dos semiespacios abiertos

$$\Omega^+ = \{(S, I, R) \in \mathbb{R}^3 : S + I + R > 0\}, \qquad \Omega^- = \{(S, I, R) \in \mathbb{R}^3 : S + I + R < 0\},$$

y esos dos semiespacios son los **dominios maximales de existencia y unicidad** del sistema: son los abiertos conexos más grandes en los que f está definida y es de clase C^1. El plural del enunciado («los dominios maximales») no es un descuido: hay dos.

El que interesa es $\Omega := \Omega^+$, por dos razones que conviene no confundir. La matemática: el dato inicial cumple $N_0 = S_0 + I_0 + R_0 > 0$, luego está en Ω^+ y no en Ω^-. La del modelo: N es una población total y una población total negativa no significa nada.

Paso 2 (el núcleo): Picard en Ω. Por ser $f \in C^1(\Omega)$, f es localmente lipschitziana respecto de (S, I, R) en Ω (condición suficiente de Lipschitz local). Es además autónoma, luego continua en $\mathbb{R} \times \Omega$. Estamos por tanto en las hipótesis del teorema de existencia y unicidad de solución maximal (existencia y unicidad maximal, libro de EDO, §2.3 «Existencia y unicidad de solución maximal»): para cada dato inicial de Ω existe una **única** solución maximal, definida en un intervalo *abierto* $J(S_0, I_0, R_0)$ que contiene a 0, y con valores en Ω.

Esta última coletilla es la que se olvida y la que va a hacer falta después: por construcción, mientras la solución existe, sus valores están en Ω, es decir, $N(t) > 0$. No es un teorema, es la definición de solución maximal en Ω.

Paso 3: los puntos críticos. Son los $(\bar S, \bar I, \bar R) \in \Omega$ con $f(\bar S, \bar I, \bar R) = 0$. La tercera componente lo decide todo: $m\gamma\bar I = 0$ con $m, \gamma > 0$ obliga a $\bar I = 0$. Y con $\bar I = 0$ las otras dos componentes se anulan solas, porque las dos llevan I como factor. En Ω^+ resulta

$$\boxed{\text{Puntos críticos} = \{(\bar S, 0, \bar R) : \bar S + \bar R > 0\} = \{I = 0\} \cap \Omega.}$$

En Ω^- se obtiene análogamente $\{(\bar S, 0, \bar R) : \bar S + \bar R < 0\}$. En la componente del dato inicial, es **la parte del plano $I = 0$ dentro de Ω**. Restringido a la región con sentido biológico, $\bar S, \bar R \geq 0$: todos los pares no negativos salvo el origen, que además no está en Ω. La lectura es la de siempre: sin infecciosos no pasa nada.

La idea: qué permite concluir la linealización

Los equilibrios no son aislados: hay otros arbitrariamente cerca. Ninguno puede ser atractivo como punto, porque la solución que parte de uno de esos vecinos permanece allí. La matriz jacobiana en $(\bar{S}, 0, \bar{R})$ tiene autovalores

$$0, \quad 0, \quad \frac{\beta_f \bar{S}}{\bar{S} + \bar{R}} - \gamma.$$

Si el tercero es positivo, el teorema de inestabilidad por linealización sí permite concluir que el equilibrio es inestable. Si es negativo o nulo, los dos ceros impiden decidir la estabilidad mediante ese criterio. Para responder a los apartados siguientes se estudia directamente cada solución.

Qué hace cada hipótesis. $\gamma > 0$ y $m > 0$ se usan en el paso 3 para poder despejar $\bar{I} = 0$ de la tercera ecuación; si fuera $m = 0$ la tercera ecuación no diría nada, pero las dos primeras seguirían dando lo mismo (de la primera, $\beta_f \bar{S}\bar{I}/\bar{N} = 0$, y llevándolo a la segunda, $\gamma \bar{I} = 0$). La hipótesis $N_0 > 0$ es la que elige la componente conexa. Con $N_0 = 0$ el campo no está definido; un dato matemático con $N_0 < 0$ pertenecería a Ω^- y también tendría solución local única, aunque no representaría una población.

Resolución del apartado (b)

La idea, y es una sola. Mírense las dos primeras ecuaciones con los ojos entornados:

$$S' = \underbrace{\left(-\frac{\beta_f I}{N}\right)}_{\text{una función de } t} \cdot S, \qquad I' = \underbrace{\left(\frac{\beta_f S}{N} - \gamma\right)}_{\text{una función de } t} \cdot I.$$

A lo largo de la solución, los paréntesis son funciones continuas del tiempo, y punto. Es decir: *cada una de esas dos ecuaciones es una EDO lineal escalar homogénea disfrazada*, y la solución de $x' = a(t)x$ es $x(t) = x(0)e^{\int a}$, que, si $x(0) \neq 0$, nunca se anula y conserva ese signo. Toda la positividad del modelo está en esa observación, y no usa nada de epidemias.

Paso 1 (el núcleo): $I(t) > 0$ en todo J. Sea (S, I, R) la solución maximal del apartado anterior y sea

$$a(t) := \frac{\beta_f S(t)}{N(t)} - \gamma, \qquad t \in J(S_0, I_0, R_0).$$

La función a es continua en J: S y N lo son por ser componentes de una solución, y $N(t) \neq 0$ porque la solución toma valores en Ω (paso 2 del apartado anterior). La segunda ecuación dice entonces $I'(t) = a(t)I(t)$ en J, y de ahí, integrando I'/I o multiplicando por el factor integrante $e^{-\int_0^t a}$,

$$I(t) = I_0 \exp\Big(\int_0^t a(s)\, ds\Big), \qquad t \in J.$$

Como $I_0 > 0$ y la exponencial es positiva, $I(t) > 0$ en todo J, hacia adelante y hacia atrás.

Paso 2: $S(t) > 0$ en todo J. Idéntico, con $b(t) := -\beta_f I(t)/N(t)$, que es continua en J por lo mismo:

$$S(t) = S_0 \exp\Big(\int_0^t b(s)\, ds\Big) > 0, \qquad t \in J.$$

Paso 3: los tres signos. Ya con $S > 0$, $I > 0$ y $N > 0$ en J:

- $S' = -\dfrac{\beta_f SI}{N} < 0$: S es estrictamente decreciente en J.
- $R' = m\gamma I > 0$: R es estrictamente creciente en J.
- Sumando las tres ecuaciones, los dos términos de contagio se cancelan y queda

$$N' = S' + I' + R' = -\gamma I + m\gamma I = -(1-m)\gamma I < 0, \tag{C.2}$$

porque $m < 1$: N es estrictamente decreciente en J. Esta identidad es el corazón del modelo y se va a usar otra vez en (c): *la población total solo baja, y baja al ritmo al que mueren los infectados*.

Paso 4: $N(t) > 0$. No hay nada que probar: la solución maximal toma sus valores en $\Omega = \{N > 0\}$ por definición. Conviene decirlo así de claro, porque la pregunta interesante es otra y es la del apartado siguiente: *si N solo baja, ¿puede llegar a cero en tiempo finito y alcanzar el borde del dominio?* Que N sea decreciente no lo impide; lo que lo impide es la cota de (c).

Otra vía (la de las órbitas invariantes, que es la que piden otras convocatorias). El mismo resultado sale del argumento geométrico que el profesor pide con otras palabras en casi todos los exámenes («probar que las rectas $u = 0$ y $v = 0$ son invariantes»). Los dos conjuntos

$$\{S = 0\} \cap \Omega \qquad \text{y} \qquad \{I = 0\} \cap \Omega$$

son **invariantes**: si $I = 0$ el sistema queda $S' = R' = 0$ y cada punto es un equilibrio; si $S = 0$ queda $S' = 0$, $I' = -\gamma I$, $R' = m\gamma I$, cuya solución es $S \equiv 0$, $I(t) = I_0 e^{-\gamma t}$, $R(t) = R_0 + m(I_0 - I(t))$, y en ningún momento sale de $\{S = 0\}$. Como por unicidad dos órbitas distintas no se cortan (propiedad de las órbitas), la órbita que parte de (S_0, I_0, R_0) con $S_0, I_0 > 0$ no puede tocar ninguna de las dos, y por continuidad se queda en $\{S > 0,\ I > 0\}$ durante todo su intervalo de existencia. Es la misma verdad contada al revés: la fórmula exponencial del paso 1 *es* la razón por la que esas órbitas no se cortan.

Qué hace cada hipótesis.

- $I_0 > 0$ y $S_0 > 0$ son imprescindibles: con $I_0 = 0$ la solución es un equilibrio y con $S_0 = 0$ es la órbita semitrivial de arriba. En los dos casos no se cumplen todas las conclusiones de (b): con $I_0 = 0$ las tres componentes son constantes; con $S_0 = 0 < I_0$, S es constante, pero R crece y N decrece.
- $m < 1$ es lo único que da $N' < 0$. Si $m = 1$ —que es el caso del apartado (d)— (C.2) da $N' \equiv 0$ y N es **constante**, no decreciente. Merece la pena señalarlo: el enunciado declara $m \in (0, 1)$ al principio y luego toma $m = 1$ en (d), así que ese apartado se sale del rango declarado y hay que leerlo como el caso límite. Se vuelve sobre ello allí.
- Nada de esto usa el signo de R, y por eso vale en todo J, también hacia atrás en el tiempo. En (c) sí hará falta $R \geq 0$, y por eso allí las conclusiones son solo hacia adelante.

Resolución del apartado (c)

La idea. Una solución maximal solo puede dejar de existir en tiempo finito por dos motivos: porque se va al infinito, o porque se acerca al borde del dominio —aquí, al plano $N = 0$, donde el campo deja de estar definido. Hay que prohibir las dos cosas, y las dos se prohíben con *una sola* observación: el sistema tiene una cantidad conservada, una combinación fija de N y R que no se mueve nunca. De ella salen todas las cotas de golpe.

Paso 1 (el núcleo): la cantidad conservada. Compárense

$$N' = -(1-m)\gamma I \qquad \text{y} \qquad R' = m\gamma I.$$

Las dos derivadas son múltiplos de la misma función γI, con coeficientes $-(1-m)$ y m. Multiplicando la primera por m y la segunda por $(1-m)$ y sumando, la parte de I se cancela:

$$\big(mN + (1-m)R\big)' = -m(1-m)\gamma I + (1-m)m\gamma I = 0.$$

Por tanto, en todo J,

$$\boxed{m\,N(t) + (1-m)\,R(t) = m\,N_0 + (1-m)\,R_0 =: K > 0.} \tag{C.3}$$

($K > 0$ porque $m > 0$, $N_0 > 0$, $1-m > 0$ y $R_0 \geq 0$.)

> LA IDEA: QUÉ SE ESTÁ CONSERVANDO
>
> Cada individuo que abandona la clase I hace dos cosas a la vez: con probabilidad m suma uno a R, y con probabilidad $1-m$ resta uno a N. Los dos flujos son proporcionales, con razón fija $m : (1-m)$, así que ninguno de los dos por separado se conserva pero la mezcla ponderada sí. Dicho en la forma más gráfica: por cada m personas que se curan, muere $1-m$, y $mN + (1-m)R$ es justo la contabilidad que hace que las dos cosas se compensen. Es el sustituto exacto de «$S + I + R = N$ constante» del SIR del capítulo 2: cuando la población deja de conservarse, lo que se conserva es esto.

Paso 2: las cotas a priori, para $t \geq 0$. Sea $t \in J \cap [0,\infty)$. De (b), R es creciente, luego $R(t) \geq R_0 \geq 0$; y $S(t) > 0$, $I(t) > 0$. Entonces

$$N(t) = S(t) + I(t) + R(t) > R(t) \geq 0.$$

Ahora se lee (C.3) dos veces, en las dos direcciones:

- Sustituyendo $R \leq N$ en $K = mN + (1-m)R$ se obtiene $K \leq mN + (1-m)N = N$, es decir
$$\boxed{N(t) \;\geq\; K = mN_0 + (1-m)R_0 \;>\; 0}$$
Esta es la cota que salva el ejercicio: la población total no solo es positiva, está separada de cero por una constante que se conoce de antemano, y por tanto el denominador del campo no puede degenerar.
- Sustituyendo $N \geq K$ en $(1-m)R = K - mN$ se obtiene $(1-m)R \leq K - mK = (1-m)K$, es decir $R(t) \leq K$. Y de $mN = K - (1-m)R \leq K$ se obtiene $N(t) \leq K/m$.

Junto con $0 < S(t) \leq S_0$ (S decreciente) y $0 < I(t) = N - S - R \leq N \leq K/m$, queda todo acotado.

Paso 3: el compacto y la existencia global. Los cinco renglones anteriores dicen que la semiórbita positiva vive en

$$\mathcal{K} := \Big\{(S,I,R):\ 0 \le S \le S_0,\ 0 \le I \le \tfrac{K}{m},\ 0 \le R \le K,\ S+I+R \ge K\Big\},$$

que es cerrado y acotado en $\mathbb{R}^3$, luego compacto, y que está *dentro* de Ω porque en él $N \ge K > 0$. La proposición del compacto (una solución maximal cuya semiórbita positiva permanece en un compacto contenido en el dominio está definida en todo $[0,+\infty)$; libro de EDO, §3.4 «Caracterización de las soluciones maximales») da entonces

$$J(S_0,I_0,R_0) \supset [0,+\infty).$$

Paso 4: los cuatro límites.

- S es estrictamente decreciente y está acotada inferiormente por 0, luego converge: $S(t) \to S_\infty \ge 0$.
- R es estrictamente creciente y está acotada superiormente por K, luego converge: $R(t) \to R_\infty$. Y $R_\infty > 0$: como R es estrictamente creciente, $R_\infty > R(0) = R_0 \ge 0$.
- N sale gratis de la cantidad conservada: $N = (K-(1-m)R)/m \to N_\infty = (K-(1-m)R_\infty)/m$, y $N_\infty \ge K > 0$ por el paso 2.
- Y entonces $I = N - S - R$ también converge, a $I_\infty := N_\infty - S_\infty - R_\infty \ge 0$. Falta ver que $I_\infty = 0$, y ese es el único punto con contenido. De $R' = m\gamma I$, integrando entre 0 y t y haciendo $t \to \infty$,

$$m\gamma \int_0^\infty I(s)\,ds = R_\infty - R_0 < \infty.$$

O sea, I es integrable en $[0,\infty)$. Y una función que tiene límite en $+\infty$ y es integrable tiene que tener límite cero: si fuese $I_\infty = c > 0$, habría un t_0 con $I(s) > c/2$ para $s \ge t_0$ y la integral sería infinita. Luego $I_\infty = 0$.

LA IDEA: POR QUÉ LA EPIDEMIA SE APAGA, SIN FÓRMULA NINGUNA

No se ha resuelto nada. El argumento es puramente contable: cada infeccioso que se cura suma a R, y R no puede crecer indefinidamente porque hay una cantidad conservada que se lo impide. Si I tendiera a una constante positiva, R crecería sin límite y rompería el techo. Por eso I tiende a cero. No se anula en tiempo finito: por (b), $I(t) > 0$ en todo J. Es el mismo argumento del SIR del capítulo 2, con la diferencia de que allí el techo era la población total N —constante— y aquí es la constante K de (C.3).

Qué hace cada hipótesis.

- $m > 0$ es lo que hace $K \ge mN_0 > 0$. Si fuera $m = 0$ (todos los infectados mueren, no hay recuperados) la constante sería $K = R_0$ y la cota $N \ge R_0$ no diría nada cuando además $R_0 = 0$: habría que trabajar de otra manera. Es el sitio exacto en que se usa que m no es cero.
- $R_0 \ge 0$ se usa en el paso 2 para $R(t) \ge 0$ y con ello $N \ge R$. Es también la razón de que todo esto sea **solo hacia adelante**: hacia atrás R puede hacerse negativa, S crece y ninguna de las cotas sobrevive. El enunciado pide $J \supset [0,\infty)$ y no dice nada de $t < 0$, y hace bien.
- El compacto tiene que estar *dentro* del dominio, no solo ser compacto. Si la cota $N \ge K$ no existiera, $\mathcal{K}$ tocaría el plano $N = 0$, que es donde el campo deja de estar definido, y la proposición del compacto no se podría aplicar.

Resolución del apartado (d)

Antes de empezar, una observación sobre el enunciado. El apartado toma $m = 1$, que está fuera del intervalo $m \in (0, 1)$ declarado al principio. Es el **caso límite** en el que no muere nadie: con $m = 1$ la ecuación (C.2) da $N' \equiv 0$, la población total es constante e igual a N_0, y el sistema es el SIR clásico de la sección 2.2.1 con incidencia de frecuencia. De las conclusiones de (b) sobrevive todo menos una: S sigue siendo estrictamente decreciente, R estrictamente creciente, $S > 0$ e $I > 0$, pero N ya no decrece, es constante. Las conclusiones de (c) también se mantienen: $0 \le S, I, R \le N_0$ y $N = N_0 > 0$ dan directamente un compacto dentro de Ω. No se divide por $1 - m$, que ahora es cero. La monotonía de S, R y la identidad $I = N_0 - S - R$ dan sus límites, y $R' = \gamma I$ permite concluir de nuevo $I \to 0$.

La idea. Con $N \equiv N_0$, la ecuación de los infecciosos es

$$I'(t) = I(t)\, g(t), \qquad g(t) := \frac{\beta_f}{N_0} S(t) - \gamma.$$

Como $I > 0$, **el signo de I' es exactamente el signo de g**. Y g es una función estrictamente decreciente del tiempo, porque S lo es. Una función estrictamente decreciente cambia de signo como mucho una vez, y solo puede hacerlo de $+$ a $-$. Así que el ejercicio se reduce a dos comprobaciones: que g empieza siendo positiva (eso es la hipótesis) y que acaba siendo negativa (eso es lo único que hay que demostrar).

Paso 1: g es continua y estrictamente decreciente en $[0, \infty)$. Por (c), $[0, \infty) \subset J$, así que g está definida en todo $[0, \infty)$. Es continua porque S lo es, y estrictamente decreciente porque S lo es y $\beta_f / N_0 > 0$.

Paso 2: $g(0) > 0$ es la hipótesis, traducida. Con $\mathcal{R}_0 = \beta_f / \gamma$,

$$g(0) = \frac{\beta_f}{N_0} S_0 - \gamma = \gamma\Big(\mathcal{R}_0 \frac{S_0}{N_0} - 1\Big),$$

y por tanto $g(0) > 0$ si y solo si $S_0 / N_0 > 1/\mathcal{R}_0$, que es justo lo que se supone. La hipótesis del enunciado no es un tecnicismo: es la condición de que *haya brote*.

Paso 3 (el núcleo): g termina siendo negativa. Es lo único que cuesta, y se hace por reducción al absurdo. Supóngase que $g(t) > 0$ para todo $t \ge 0$, es decir,

$$S(t) > \frac{\gamma N_0}{\beta_f} = \frac{N_0}{\mathcal{R}_0} \qquad \text{para todo } t \ge 0.$$

Entonces $I' = gI > 0$ en $[0, \infty)$, luego I es creciente y $I(t) \ge I_0 > 0$. Llevando las dos desigualdades a la ecuación de S:

$$S'(t) = -\frac{\beta_f}{N_0} S(t) I(t) \;\le\; -\frac{\beta_f}{N_0} \cdot \frac{N_0}{\mathcal{R}_0} \cdot I_0 = -\frac{\beta_f}{\mathcal{R}_0} I_0 = -\gamma I_0 < 0,$$

una cota *uniforme* en t. Integrando, $S(t) \le S_0 - \gamma I_0 t$ para todo $t \ge 0$, que se hace negativo en cuanto $t > S_0/(\gamma I_0)$ y contradice $S(t) > 0$. Luego existe algún $t \ge 0$ con $g(t) \le 0$.

Obsérvese dónde está el peso: no basta con decir «S decrece, luego g acabará bajando de cero», porque una función decreciente y positiva puede tener límite positivo. Lo que cierra el argumento es que, mientras $g > 0$, los infecciosos no bajan de I_0 y eso obliga a S a caer a *velocidad mínima fija*, no cada vez más despacio.

Paso 4: el instante es único, y el signo cambia una sola vez. g es continua, estrictamente decreciente, $g(0) > 0$ y $g(t_1) \leq 0$ para algún t_1. Por Bolzano existe $T_{max} \in (0, t_1]$ con $g(T_{max}) = 0$, y por ser g estrictamente decreciente ese cero es único y además

$$g > 0 \text{ en } [0, T_{max}), \qquad g < 0 \text{ en } (T_{max}, \infty).$$

Como el signo de I' es el de g, se concluye que I es estrictamente creciente en $(0, T_{max})$ y estrictamente decreciente en (T_{max}, ∞), y $T_{max} \in (0, \infty)$ es el único instante con esa propiedad. ■

LA IDEA: EL PICO NO ES EL FINAL, ES LA INMUNIDAD DE GRUPO

T_{max} es el instante en que $S(T_{max}) = N_0/\mathcal{R}_0$, es decir, en que la fracción de susceptibles cae hasta $1/\mathcal{R}_0$. Antes de ese momento cada infeccioso contagia, por término medio, a más de uno en las condiciones de ese instante, y los infecciosos aumentan; después, a menos de uno, y los infecciosos disminuyen. Es la definición operativa de *inmunidad de grupo*: no hace falta que se contagie todo el mundo, basta con que queden pocos susceptibles para que cada enfermo no llegue a reemplazarse. El pico llega **con la epidemia todavía en marcha**: en T_{max} el número de infecciosos es máximo, no cero, y después siguen produciéndose contagios. Un ejemplo de curva con un solo pico aparece a la izquierda de la figura 2.14, que es la que dibuja *I contra el tiempo*. En la figura 2.4 la misma campana aparece tumbada, porque allí los ejes son I y R: como R es estrictamente creciente, hace de reloj, y cada órbita se abomba hacia la derecha y vuelve al eje. Conviene no confundirlas, y tener presente que en esa segunda figura tres de las diez órbitas no se abomban en absoluto, porque arrancan por debajo del umbral y no llegan a tener pico.

Otra vía (la del primer integral, que además da la altura del pico). Con $m = 1$ y $N \equiv N_0$ se puede eliminar el tiempo dividiendo las dos primeras ecuaciones:

$$\frac{dI}{dS} = \frac{\frac{\beta_f}{N_0} SI - \gamma I}{-\frac{\beta_f}{N_0} SI} = -1 + \frac{\gamma N_0}{\beta_f S} = -1 + \frac{N_0}{\mathcal{R}_0} \cdot \frac{1}{S},$$

que se integra directamente: a lo largo de toda la órbita,

$$I + S - \frac{N_0}{\mathcal{R}_0} \ln S = I_0 + S_0 - \frac{N_0}{\mathcal{R}_0} \ln S_0 =: C.$$

La órbita es la gráfica de I como función de S, y su máximo está donde $dI/dS = 0$, esto es, en $S = N_0/\mathcal{R}_0$: el mismo umbral del paso 3, obtenido sin hablar del tiempo. Sustituyendo, la altura del pico es

$$I_{max} = C - \frac{N_0}{\mathcal{R}_0} + \frac{N_0}{\mathcal{R}_0} \ln \frac{N_0}{\mathcal{R}_0},$$

una fórmula cerrada que el argumento del tiempo no daba. Lo que esta vía *no* da es la existencia de un instante T_{max} finito: para eso hay que volver al tiempo, porque la órbita podría, en principio, tardar infinito en llegar al umbral. Las dos vías se complementan y conviene tener las dos.

Qué buscaba este examen

- **Apartado (a).** Que se plantee el problema de Cauchy en el abierto correcto. La dificultad está en el denominador: el campo es C^∞ solo fuera del plano $N = 0$. Hay que identificar las dos componentes de ese abierto.
- **Apartado (b).** La técnica que se repite en todas las convocatorias: leer una ecuación como lineal escalar en su propia incógnita y sacar la positividad de la fórmula exponencial, o equivalentemente de que las órbitas no se cortan.
- **Apartado (c).** Encontrar la cantidad conservada del modelo cuando la población ya no se conserva, y con ella las cotas a priori que meten la órbita en un compacto. Es el esquema *positividad* $\Rightarrow$ *cotas* $\Rightarrow$ *compacto* $\Rightarrow$ *prolongación* $\Rightarrow$ *límites* que aparece en casi todos los exámenes de la asignatura.
- **Apartado (d).** Reducir una pregunta sobre la forma de una curva a un estudio de signos de una función monótona, y saber por qué el caso «g nunca baja de cero» no puede darse.

El atajo que ahorra tiempo es hacer (b), (c) y (d) *en ese orden y apoyándose unos en otros*, como pide el enunciado con su «como consecuencia»: la positividad de (b) es lo que permite las cotas de (c), y la prolongación de (c) es lo que permite hablar de $t \to \infty$ en (d). Intentar (d) sin (c) obliga a repetir el trabajo.

Errores típicos

- **Tomar $\Omega = \mathbb{R}^3$ o $\Omega = \{S, I, R > 0\}$.** Lo primero es falso (el campo no está definido en el plano $N = 0$); lo segundo no es un dominio maximal, es un trozo de uno, y además excluye los datos admitidos con alguna componente nula. Picard puede aplicarse en cualquier abierto adecuado que contenga el dato; la maximalidad del dominio y la invariancia son cuestiones distintas.
- **Justificar $S, I > 0$ «porque son poblaciones».** Eso es la razón por la que el modelo sería malo si fallara, no una demostración de que no falla. Lo que se demuestra es precisamente que el modelo respeta esa exigencia.
- **Dar por hecho $N(t) > 0$ y creer que se ha demostrado algo.** Es cierto por definición de solución maximal en Ω; lo que hay que demostrar es que N está *separada de cero*, que es otra cosa.
- **Decir que N es decreciente y positiva, luego converge, luego ya está.** Converge, sí, pero a algo que podría ser cero, y entonces el compacto tocaría el borde del dominio. La cota $N \geq K$ es la que permite aplicar aquí el criterio del compacto.
- **Afirmar que I es monótona en el caso general.** No lo es: (d) demuestra que sube y luego baja. Que $I \to 0$ se prueba con la integrabilidad, no con monotonía.
- **En (d), olvidar el paso 3.** Es el único paso con contenido; sin él solo se ha probado que *si g* se anula, el instante es único.
- **Confundir R con $\mathcal{R}_0$.** Con esta notación pasa, y una vez que pasa el apartado (d) se vuelve ininteligible.

C.4 Examen final del 2 de junio de 2025 (primera convocatoria de 2024-25)

Se conserva el enunciado completo de esta convocatoria: cuatro bloques de cinco puntos, dos de teoría y dos de MATLAB. Las soluciones siguientes desarrollan cada uno de ellos.

Ficha

Asignatura	Modelización matemática, curso 2024/25
Convocatoria	Primera (junio)
Fecha	2 de junio de 2025
Estructura	Cuatro bloques de 5 puntos, 20 en total
	1. Teoría, primera parte
	2. Teoría, segunda parte
	3. Práctica, primera parte (MATLAB)
	4. Práctica, segunda parte (MATLAB)
Otros datos	En la cabecera se pide también el grupo de prácticas

Enunciado

Ejercicio 1 [Teoría: Primera Parte, 5 puntos]. Se considera el modelo de dos especies en simbiosis con término logístico:

$$\begin{cases} \bar{u}' = \bar{u}(-a + b\bar{v}), \\ \bar{v}' = \bar{v}(-c + d\bar{u} - e\bar{v}), \end{cases}$$

con $a, b, c, d, e > 0$. Se pide:

a) Mediante un cambio de unidades apropiado, probar que el anterior sistema se puede escribir bajo la forma

$$\begin{cases} u' = \lambda u(-1 + v), \\ v' = v(-1 + u - \beta v), \end{cases} \tag{C.4}$$

con $\lambda, \beta > 0$.

b) Calcular los puntos críticos del sistema (C.4) y estudiar la estabilidad de los que están situados en el primer cuadrante $\{(u, v) : u \geq 0,\ v \geq 0\}$.

c) Consideramos el problema de Cauchy

$$\begin{cases} y' = y(-1 - \beta y), \\ y(0) = y_0. \end{cases} \tag{C.5}$$

Calcular los puntos críticos de (C.5) y estudiar su estabilidad. Calcular las órbitas $\gamma(y_0)$ para cada $y_0 \geq 0$.

d) Deducir que las rectas $u = 0$ y $v = 0$ son invariantes para el sistema (C.4) y que la solución maximal de (C.4) con condición inicial $u(0) = u_0$, $v(0) = v_0$, para datos $u_0, v_0 > 0$, satisface

$$u(t) > 0, \qquad v(t) > 0 \qquad \forall t \in I(u_0, v_0),$$

siendo $I(u_0, v_0)$ el intervalo de definición de la solución maximal.

e) Estudiar las zonas de crecimiento y de decrecimiento de las órbitas $\gamma(u_0, v_0)$ del sistema (C.4) con $u_0 > 0$ y $v_0 > 0$. ¿Pueden existir órbitas cíclicas $\gamma(u_0, v_0)$ cuando $u_0 > 0$ y $v_0 > 0$?

Ejercicio 2 [Teoría: Segunda Parte, 5 puntos]. Consideremos el problema de Cauchy

$$\begin{cases} y' = ay + v, \quad t \in (0, T) \\ y(0) = y_0 \end{cases} \tag{C.6}$$

y el problema de control óptimo

$$\begin{cases} \text{Minimizar } J(v) := \dfrac{\alpha}{4}\left|y(T) - y_d\right|^4 + \dfrac{\mu}{2}\displaystyle\int_0^T |v(t)|^2\, dt \\ \text{Sujeto a } v \in L^2(0, T), \quad (y, v) \text{ satisface (C.6)}, \end{cases} \tag{C.7}$$

con $\alpha, \mu \in (0, \infty)$, $T > 0$ y $a, y_0, y_d \in \mathbb{R}$. Se pide:

a) Demostrar que $J : v \in L^2(0, T) \mapsto J(v) \in \mathbb{R}$ está bien definido y es un funcional estrictamente convexo en $L^2(0, T)$.

b) Demostrar que el problema de mínimos (C.7) tiene una única solución $\tilde{v} \in L^2(0, T)$.

Ejercicio 3 [Práctica: Primera Parte, 5 puntos]. Analizamos una reacción química consecutiva $A \xrightarrow{k_1} B \xrightarrow{k_2} C$, donde $A = A(t)$, $B = B(t)$ y $C = C(t)$ son las concentraciones respectivas de cada reactivo en cada instante t y k_1, k_2 son las correspondientes velocidades de reacción. Teniendo en cuenta que la destrucción de A es proporcional a su concentración, que la variación de B depende de la cantidad creada desde A y de la cantidad que se destruye de B, y que para C es irrelevante la cantidad que provenga de A, se obtiene el sistema

$$\begin{cases} \dfrac{dA}{dt} = -k_1 A, \\ \dfrac{dB}{dt} = k_1 A - k_2 B, \\ \dfrac{dC}{dt} = k_2 B, \end{cases} \tag{C.8}$$

para $t \in (0, T)$, donde k_1 y k_2 son constantes dadas. Se supone que inicialmente hay una cantidad A_0 del primer reactivo y 0 de B y de C. Se pide escribir una M-función

```
ReacConsecut(k1, k2, A0, T)
```

que a partir de los datos k_1, k_2, A_0 y T haga lo siguiente:

a) Resolver numéricamente el modelo en el intervalo $[0, T]$.

b) Representar en la misma gráfica las trayectorias de A, B y C obtenidas.

c) Representar la órbita (A, C) y señalar su punto inicial y su punto final.

d) Mostrar el resultado de las siguientes cuestiones en la ventana de comandos:

   1) Obtener el punto de la trayectoria en que hay la misma cantidad del reactivo A y del producto C, determinando dicha cantidad.
   2) Hallar la cantidad máxima de B que se alcanza a lo largo de la reacción y el instante en que se obtiene.
   3) **Solo cuando** $k_1 \neq k_2$, la solución exacta para la concentración de B es

$$B = \frac{k_1}{k_2 - k_1} A_0 \left(e^{-k_1 t} - e^{-k_2 t}\right).$$

   Obtener el error en norma euclídea cometido al resolver el problema numéricamente en este caso.

Añadir en cada caso leyendas explicativas de las gráficas. Ejecutar el programa en los siguientes casos: $k_1 = 1\,\mathrm{s}^{-1}$, $k_2 = 2\,\mathrm{s}^{-1}$, $A_0 = 1\,\mathrm{M}$, $T = 7\,\mathrm{s}$; y $k_1 = 1\,\mathrm{s}^{-1}$, $k_2 = 1\,\mathrm{s}^{-1}$, $A_0 = 1\,\mathrm{M}$, $T = 7\,\mathrm{s}$.

Ejercicio 4 [Práctica: Segunda Parte, 5 puntos]. Consideramos el siguiente problema de valores iniciales correspondiente a un modelo de tipo SEAIR, donde consideramos poblaciones susceptibles de enfermar ($S = S(t)$), en proceso de incubación ($E = E(t)$), infectados asintomáticos ($A = A(t)$), infectados con síntomas ($I = I(t)$) y recuperados ($R(t)$). El mecanismo de la enfermedad sigue el siguiente esquema: los contagios se pueden producir por interacción (S, E), (S, A) o (S, I). Se trata de un sistema cerrado (la población total no varía con el tiempo) que viene descrito por las siguientes ecuaciones:

$$\begin{cases} S'(t) = -\beta_1 SE - \beta_2 SA - \beta_3 SI, \\ E'(t) = \beta_1 SE + \beta_2 SA + \beta_3 SI - (\lambda_1 + \lambda_2)E, \\ A'(t) = \lambda_1 E - \delta_1 A, \\ I'(t) = \lambda_2 E - \delta_2 I, \\ R'(t) = \delta_1 A + \delta_2 I, \\ S(0) = S_0,\ E(0) = E_0,\ A(0) = A_0,\ I(0) = I_0,\ R(0) = R_0, \end{cases} \tag{C.9}$$

donde β_i $(i = 1, 2, 3)$, λ_j y δ_j $(j = 1, 2)$ son parámetros en el intervalo $[0, 1]$.

Dados los parámetros $(\beta_2, \beta_3, \lambda_1, \lambda_2, \delta_1)$, queremos calcular el valor de los parámetros (β_1, δ_2) para que la población $I(t)$ verifique

$$I(t_i) = I_i, \qquad i = 1, \ldots, m, \quad (m \geq 1). \tag{C.10}$$

Para ello se pide escribir una M-function

```
[beta1,delta2] = InvSEAIR(T,tobs,iobs,beta10,delta20)
```

que realice las siguientes operaciones:

a) Calcular `beta1` y `delta2`, valores aproximados de los parámetros β_1 y δ_2 respectivamente, de tal forma que la solución $I(t)$ del problema (C.9) verifique la condición (C.10), siendo `tobs` e `iobs` dos vectores de la misma longitud que contienen los valores observados y `beta10`, `delta20` $\in [0, 1]$ sugerencias de aproximaciones iniciales de β_1 y δ_2. Para ello, formúlese un problema de mínimos cuadrados y resuélvase numéricamente el correspondiente problema de mínimo. Los valores óptimos de `beta1` y `delta2` deben aparecer como argumento de salida de la M-function.

b) Para los datos $T = 60$, $[S_0, E_0, A_0, I_0, R_0] = [800, 10, 5, 3, 182]$,

$$\texttt{tobs} = [1, 5, 9, 12, 15, 30],$$

$$\texttt{iobs} = [14.1927,\ 62.2850,\ 92.4531,\ 107.0014,\ 116.6409,\ 126.7263],$$

$$[\beta_2, \beta_3, \lambda_1, \lambda_2, \delta_1] = [0.012,\ 0.005,\ 0.08,\ 0.02,\ 0.03],$$

`beta10` $= 0.1$, `delta20` $= 0$, se pide: representar la solución del problema (C.9) correspondiente a los valores óptimos de `beta1` y `delta2` calculados en el apartado anterior junto con los valores observados (`tobs`, `iobs`), añadiendo a la gráfica título e informaciones explicativas necesarias junto con los valores calculados; y representar en otra gráfica la solución del problema (C.9) correspondiente a los valores de `beta10` y `delta20` junto con los valores observados, con título e informaciones explicativas.

Resolución del ejercicio 1

Qué modelo es. Es una **simbiosis obligada** con autolimitación en la segunda especie. Léanse los signos: si $\bar{v} = 0$, la primera especie cumple $\bar{u}' = -a\bar{u}$ y se extingue sola; si $\bar{u} = 0$, la segunda cumple $\bar{v}' = -\bar{v}(c + e\bar{v})$ y también se extingue. Ninguna de las dos sobrevive sin la otra —eso es lo que quiere decir *obligada*—, y cada una se beneficia de la presencia de la contraria ($+b\bar{v}$ en la primera, $+d\bar{u}$ en la segunda). El término $-e\bar{v}^2$ es el de siempre: la segunda especie compite consigo misma por un recurso limitado. Es el contraste con la competición de la sección 2.2.5: aquí los términos de interacción benefician a ambas especies. Invertir el tiempo cambiaría también los signos de mortalidad y autolimitación, por lo que no basta para transformar aquel modelo en este.

Apartado (a): la adimensionalización

La idea. Un cambio de unidades no puede cambiar la física del problema: solo puede cambiar las etiquetas de los ejes. Lo que sí hace es *revelar cuántos parámetros hay de verdad*. Aquí hay cinco constantes, a, b, c, d, e, pero se dispone de tres unidades libres —una para medir $\bar{u}$, otra para medir $\bar{v}$ y otra para medir el tiempo—, y en este sistema pueden usarse para fijar tres coeficientes. Quedan dos combinaciones adimensionales. El cambio explícito que sigue justifica ese recuento; disponer de tres escalas no garantiza por sí solo que eliminen tres parámetros independientes.

Paso 1: escribir el cambio con incógnitas. Se busca

$$\bar{u}(t) = A\,u(s), \qquad \bar{v}(t) = B\,v(s), \qquad s = \frac{t}{\tau},$$

con $A, B, \tau > 0$ a determinar: A y B son las unidades en que se miden las dos poblaciones y τ la unidad de tiempo. Por la regla de la cadena, $\dfrac{d\bar{u}}{dt} = \dfrac{A}{\tau}\dfrac{du}{ds}$, y lo mismo para $\bar{v}$.

Paso 2 (el núcleo): imponer la forma que se quiere y leer las condiciones. Sustituyendo en la primera ecuación,

$$\frac{A}{\tau}u' = A\,u\,(-a + bBv) \quad\Longrightarrow\quad u' = \tau a\,u\Big(-1 + \frac{bB}{a}v\Big).$$

Para que esto sea $u' = \lambda u(-1 + v)$ hace falta exactamente

$$\frac{bB}{a} = 1 \quad\Longrightarrow\quad \boxed{B = \frac{a}{b}}, \qquad \lambda = \tau a.$$

En la segunda,

$$\frac{B}{\tau}v' = B\,v\,(-c + dAu - eBv) \quad\Longrightarrow\quad v' = \tau c\,v\Big(-1 + \frac{dA}{c}u - \frac{eB}{c}v\Big),$$

y para que sea $v' = v(-1 + u - \beta v)$ hacen falta

$$\tau c = 1 \implies \boxed{\tau = \frac{1}{c}}, \qquad \frac{dA}{c} = 1 \implies \boxed{A = \frac{c}{d}}, \qquad \beta = \frac{eB}{c} = \frac{ea}{bc}.$$

Con $\tau = 1/c$ queda además $\lambda = \tau a = a/c$.

Paso 3: comprobar que sirve. Las tres constantes son positivas ($A = c/d > 0$, $B = a/b > 0$, $\tau = 1/c > 0$), luego el cambio es admisible, y los dos parámetros que quedan también:

$$\boxed{\lambda = \frac{a}{c} > 0, \qquad \beta = \frac{ae}{bc} > 0.}$$

Escrito con la notación del profesor, el cambio es $u(s) = \frac{d}{c}\,\bar{u}\left(\frac{s}{c}\right)$ y $v(s) = \frac{b}{a}\,\bar{v}\left(\frac{s}{c}\right)$.

LA IDEA: QUÉ SIGNIFICAN λ Y β

$1/a$ y $1/c$ son las escalas de tiempo asociadas a la mortalidad de cada especie. El cociente $\lambda = a/c$ compara sus tasas: si $\lambda > 1$, la primera decae más rápidamente cuando falta la interacción. En la segunda ecuación, $\beta = ea/(bc) = eB/c$ mide la autolimitación a la escala $B = a/b$, comparada con la tasa c. El parámetro d queda absorbido en $A = c/d$: sigue siendo necesario para volver a las poblaciones originales, aunque no aparezca en las ecuaciones adimensionales.

Dos conjuntos de parámetros con los mismos λ, β tienen las mismas ecuaciones adimensionales. Para comparar sus trayectorias hay que transformar también los datos iniciales: $u_0 = \bar{u}_0/A$ y $v_0 = \bar{v}_0/B$. Cambiar una tasa física no equivale, en general, a cambiar solo la unidad de tiempo.

Apartado (b): puntos críticos y estabilidad

Paso 1: resolver el sistema $u' = v' = 0$. De $\lambda u(-1+v) = 0$ salen dos casos, $u = 0$ o $v = 1$; de $v(-1+u-\beta v) = 0$, dos más, $v = 0$ o $u = 1+\beta v$. Se cruzan:

- $u = 0$ y $v = 0$: el origen $(0,0)$.
- $u = 0$ y $u = 1+\beta v$: da $v = -1/\beta$, es decir $(0, -1/\beta)$, que está fuera del primer cuadrante.
- $v = 1$ y $v = 0$: imposible.
- $v = 1$ y $u = 1+\beta v$: el punto $(1+\beta,\, 1)$, que sí está en el primer cuadrante.

Luego hay tres puntos críticos, y los del primer cuadrante son

$$P_0 = (0,0) \qquad \text{y} \qquad P_* = (1+\beta,\, 1).$$

Paso 2: la matriz jacobiana. Con $f(u,v) = \left(\lambda(-u+uv),\ -v+uv-\beta v^2\right)$,

$$Df(u,v) = \begin{pmatrix} \lambda(-1+v) & \lambda u \\ v & -1+u-2\beta v \end{pmatrix}.$$

Paso 3: el origen es exponencialmente asintóticamente estable.

$$Df(0,0) = \begin{pmatrix} -\lambda & 0 \\ 0 & -1 \end{pmatrix},$$

diagonal, con autovalores $-\lambda < 0$ y $-1 < 0$. Los dos tienen parte real estrictamente negativa, luego por el método de la primera aproximación (teorema de linealización, libro de EDO, §6.4 «Primer método de Lyapunov») P_0 es **exponencialmente asintóticamente estable**.

Paso 4 (el núcleo): P_* es un punto de silla. En $u = 1+\beta$, $v = 1$:

$$Df(P_*) = \begin{pmatrix} \lambda(-1+1) & \lambda(1+\beta) \\ 1 & -1+(1+\beta)-2\beta \end{pmatrix} = \begin{pmatrix} 0 & \lambda(1+\beta) \\ 1 & -\beta \end{pmatrix}.$$

No hace falta calcular los autovalores: basta con el determinante,

$$\det Df(P_*) = 0 \cdot (-\beta) - \lambda(1+\beta) \cdot 1 = -\lambda(1+\beta) < 0.$$

El determinante es el producto de los dos autovalores. Si es negativo, los autovalores son reales (una pareja compleja conjugada tiene producto $|\mu|^2 > 0$) y de signos opuestos: **punto de silla**, y en particular **inestable**, porque hay un autovalor con parte real positiva. Explícitamente, el polinomio característico es $\mu^2 + \beta\mu - \lambda(1+\beta) = 0$, con raíces $\mu_\pm = \frac{-\beta \pm \sqrt{\beta^2 + 4\lambda(1+\beta)}}{2}$, y el discriminante es positivo, luego $\mu_- < 0 < \mu_+$.

LA IDEA: UN EQUILIBRIO DE COEXISTENCIA QUE NO ATRAE A SUS VECINOS

Cerca del origen, el beneficio mutuo no compensa la mortalidad y las dos especies se extinguen. En cambio, P_* es una silla: hay trayectorias que se acercan a él por su variedad estable y otras que se alejan por la dirección inestable. Por eso su existencia no garantiza una coexistencia estable.
El apartado (e) identifica una región donde ambas poblaciones decrecen hacia el origen y otra donde crecen sin quedar acotadas. Esa coexistencia de extinción y crecimiento recuerda al efecto Allee, pero ahora el umbral depende de las dos poblaciones. La autolimitación de v no basta para acotar este sistema. Añadirla también en u cambiaría el modelo: su efecto dependería de su intensidad frente al beneficio mutuo.

Apartado (c): la EDO escalar y sus órbitas

Por qué está aquí este apartado. No es un ejercicio suelto: (C.5) es *exactamente* lo que queda de (C.4) sobre la recta $u = 0$, porque allí $v' = v(-1 + 0 - \beta v) = v(-1 - \beta v)$. Resolver (c) es describir la dinámica sobre uno de los dos ejes, y por eso (d) empieza con «deducir».

Paso 1: puntos críticos y estabilidad. Con $f(y) = y(-1-\beta y) = -y - \beta y^2$, los ceros son $y = 0$ e $y = -1/\beta$. Como $f'(y) = -1 - 2\beta y$:

$$f'(0) = -1 < 0 \implies y = 0 \text{ exponencialmente asintóticamente estable;}$$

$$f'(-1/\beta) = -1 + 2 = 1 > 0 \implies y = -1/\beta \text{ inestable.}$$

Paso 2: las órbitas para $y_0 \geq 0$. Si $y_0 = 0$ la solución es constante y $\gamma(0) = \{0\}$. Si $y_0 > 0$, entonces $f(y_0) = -y_0(1+\beta y_0) < 0$, y el teorema de las EDO autónomas escalares (teorema de las EDO autónomas) dice: la solución es estrictamente decreciente; como existe un punto crítico por debajo de y_0 —el 0— la solución está definida en $[0, +\infty)$ y converge al mayor de los puntos críticos menores que y_0, que es 0. Hacia atrás en el tiempo la solución crece, y como no hay ningún punto crítico por encima de y_0, tiende a $+\infty$.

Paso 3: la órbita, calculada. La ecuación se integra explícitamente y conviene hacerlo, porque enseña algo. Separando variables y descomponiendo en fracciones simples,

$$\frac{1}{y(1+\beta y)} = \frac{1}{y} - \frac{\beta}{1+\beta y} \implies \ln\frac{y}{1+\beta y} = -t + \text{cte},$$

de donde, imponiendo $y(0) = y_0$ y despejando,

$$y(t) = \frac{Ce^{-t}}{1 - \beta C e^{-t}}, \qquad C = \frac{y_0}{1+\beta y_0} \in \left(0, \frac{1}{\beta}\right).$$

El denominador se anula cuando $e^{-t} = 1/(\beta C) > 1$, es decir en $t^* = \ln(\beta C) < 0$: la explosión ocurre en *tiempo negativo*. Luego $I(y_0) = (t^*, +\infty)$, la solución recorre todos los valores desde $+\infty$ (cuando $t \to t^{*+}$) hasta 0 (cuando $t \to +\infty$), sin alcanzarlos, y por tanto

$$\boxed{\gamma(y_0) = (0, +\infty) \text{ para todo } y_0 > 0, \qquad \gamma(0) = \{0\}.}$$

LA IDEA: HAY UNA SOLA ÓRBITA POSITIVA, Y ESO ES LO QUE SE VA A USAR

La conclusión que parece un detalle es la importante: **todos** los datos iniciales positivos están sobre la *misma* órbita, la semirrecta $(0, +\infty)$ entera. Tiene que ser así, y se puede ver sin cuentas: en dimensión uno una órbita es un intervalo, sus extremos solo pueden ser puntos críticos o $\pm\infty$, y entre 0 y $+\infty$ no hay ningún punto crítico. Lo que hace falta en (d) es precisamente esto: la semirrecta $\{0\} \times (0, +\infty)$ del plano de fases es una órbita completa del sistema, y una órbita es una barrera que ninguna otra puede cruzar.

Apartado (d): invariancia de los ejes y positividad

Paso 1: la recta $u = 0$ es invariante, y sus órbitas son las de (c). Si en algún instante $u = 0$, la primera ecuación da $u' = \lambda \cdot 0 \cdot (-1 + v) = 0$. Más precisamente: la función $(0, y(t))$, con y la solución de (C.5), resuelve (C.4) con dato $(0, y_0)$, porque la primera componente cumple $0' = \lambda \cdot 0 \cdot (-1 + y)$ y la segunda es (C.5). Por unicidad, esa *es* la solución con ese dato. Luego la recta $u = 0$ está formada por órbitas del sistema; en particular, con $v_0 > 0$, la semirrecta $\{0\} \times (0, +\infty)$ es una órbita entera, por el apartado (c).

Paso 2: la recta $v = 0$ también. Si $v_0 = 0$, la función $(u_0 e^{-\lambda t}, 0)$ resuelve el sistema: la segunda componente da $0' = 0 \cdot (-1 + u - 0)$ y la primera, $u' = \lambda u(-1 + 0) = -\lambda u$. Por unicidad es la solución. Las órbitas contenidas en $v = 0$ son, por tanto, el origen y las dos semirrectas $(0, +\infty) \times \{0\}$ y $(-\infty, 0) \times \{0\}$.

Paso 3 (el núcleo): las órbitas no se cortan. Sea (u, v) la solución maximal con $u_0, v_0 > 0$. Si existiera $t_1 \in I(u_0, v_0)$ con $u(t_1) = 0$, el punto $(0, v(t_1))$ estaría a la vez en la órbita de (u_0, v_0) y en una de las órbitas contenidas en la recta $u = 0$; por la propiedad de las órbitas, dos órbitas que se cortan son la misma, luego (u_0, v_0) estaría en la recta $u = 0$, y no lo está. Lo mismo con v. Así que $u(t) \neq 0$ y $v(t) \neq 0$ en todo $I(u_0, v_0)$, y como son continuas y parten de valores positivos, el teorema de Bolzano da

$$u(t) > 0, \qquad v(t) > 0 \qquad \text{para todo } t \in I(u_0, v_0).$$

Otra vía (la fórmula exponencial, que evita hablar de órbitas). Igual que en la convocatoria de 2026: a lo largo de la solución, $u' = \alpha(t)u$ con $\alpha(t) = \lambda(-1 + v(t))$ continua, y $v' = \theta(t)v$ con $\theta(t) = -1 + u(t) - \beta v(t)$ continua. Entonces

$$u(t) = u_0 \exp\Big(\int_0^t \alpha\Big) > 0, \qquad v(t) = v_0 \exp\Big(\int_0^t \theta\Big) > 0.$$

Es más corto y no necesita el apartado (c); pero el enunciado pide expresamente *deducir* la positividad de la invariancia, así que en el examen conviene dar la vía de las órbitas y mencionar esta.

Apartado (e): zonas de crecimiento y órbitas cíclicas

Paso 1: las dos isoclinas nulas. En el cuadrante abierto $u > 0$, $v > 0$:

$$u' = 0 \iff v = 1, \qquad v' = 0 \iff u = 1 + \beta v.$$

La primera es una recta horizontal; la segunda, una recta de pendiente $1/\beta$ en el plano (u, v). Se cortan justamente en $P_* = (1 + \beta, 1)$, como tenía que ser. Y los signos:

$$u' > 0 \iff v > 1, \qquad v' > 0 \iff u > 1 + \beta v.$$

Paso 2: las cuatro zonas. Las dos rectas parten el cuadrante abierto en cuatro regiones, y en cada una las dos flechas tienen un signo fijo:

Región	u'	v'	Descripción
R_1: $v > 1$ y $u > 1 + \beta v$	$+$	$+$	las dos poblaciones crecen
R_2: $v > 1$ y $u < 1 + \beta v$	$+$	$-$	crece u, decrece v
R_3: $v < 1$ y $u > 1 + \beta v$	$-$	$+$	decrece u, crece v
R_4: $v < 1$ y $u < 1 + \beta v$	$-$	$-$	las dos decrecen, hacia el origen

Paso 3 (el núcleo): R_1 y R_4 son positivamente invariantes. Este es el hecho del que cuelga todo lo demás, y se comprueba mirando qué hace el campo en las fronteras.

- Frontera de R_4 con $v = 1$ (y $u < 1 + \beta$): allí $v' = v(-1 + u - \beta) = -1 + u - \beta < 0$, el campo apunta hacia abajo, hacia $v < 1$: se entra en R_4 y no se sale.
- Frontera de R_4 con $u = 1 + \beta v$ (y $v < 1$): allí $u' = \lambda u(v - 1) < 0$, el campo apunta hacia la izquierda, hacia $u < 1 + \beta v$: otra vez hacia dentro.

El mismo cálculo con los signos cambiados muestra que R_1 también es positivamente invariante. Dentro de R_4 las dos componentes decrecen estrictamente y permanecen en el rectángulo compacto $[0, u_0] \times [0, v_0]$, contenido en el dominio $\mathbb{R}^2$ del campo. La solución existe para todo $t \geq 0$; las dos componentes convergen, y el límite es un punto crítico (límite finito de una solución autónoma): tiene que ser el origen. Dentro de R_1 las dos crecen estrictamente; si estuvieran acotadas convergerían a un punto crítico de $\overline{R_1}$, y el único es la esquina P_*, imposible porque se sale de un punto estrictamente mayor. Al menos una componente no está acotada. De hecho, ambas tienden a $+\infty$ al acercarse al extremo derecho del intervalo maximal: si u estuviera acotada, $v < (u - 1)/\beta$ también lo estaría. Si solo v estuviera acotada, la ecuación $u' = \lambda(v - 1)u$ impediría una explosión en tiempo finito; y en tiempo infinito, $u \to \infty$ haría finalmente $v'/v = u - 1 - \beta v \geq 1$, contradiciendo esa cota de v.

Paso 4: no hay órbitas cíclicas. Supóngase que $\gamma(u_0, v_0)$ es cíclica con $u_0, v_0 > 0$. Una órbita cíclica es acotada y vuelve sobre sí misma, así que

- *no puede visitar* R_1: si entrara, se quedaría (invariancia positiva) y además u crecería estrictamente para siempre, luego no podría volver al punto de partida;
- *no puede visitar* R_4: mismo argumento con u decreciendo;
- *no puede pasar por* P_*, porque es un punto crítico y una órbita cíclica no contiene puntos críticos.

Solo le quedan R_2 y R_3 —y las partes de las isoclinas que las separan de las otras dos—. Pero $\overline{R_2} \cap \overline{R_3} = \{P_*\}$: las dos regiones solo se tocan en el equilibrio, por el que no puede pasar. Así que la órbita entera estaría contenida en $\overline{R_2}$ o entera en $\overline{R_3}$. En el primer caso $u' > 0$ en todo

el interior y u sería estrictamente creciente; en el segundo, estrictamente decreciente. En los dos, u no puede volver a su valor inicial. Contradicción:

$$\boxed{\text{no existen órbitas cíclicas con } u_0 > 0,\ v_0 > 0.}$$

Otra vía (el criterio de Dulac, que lo despacha en dos líneas). Como alternativa al argumento por regiones, el criterio de Dulac afirma: si en un abierto simplemente conexo del plano existe $B > 0$ de clase C^1 tal que $\operatorname{div}(Bf)$ tiene signo constante y no nulo, no hay órbitas cíclicas dentro (porque el teorema de Green aplicado al interior de una órbita cerrada daría una integral no nula igual a cero). Aquí, con $B(u,v) = 1/(uv)$ en el cuadrante abierto,

$$\frac{\partial}{\partial u}\Big(\frac{\lambda(-1+v)}{v}\Big) + \frac{\partial}{\partial v}\Big(\frac{-1+u-\beta v}{u}\Big) = 0 - \frac{\beta}{u} < 0.$$

LA IDEA: POR QUÉ LA SIMBIOSIS NO PUEDE OSCILAR Y EL PRESA-DEPREDADOR SÍ

En el modelo de Volterra de la sección 2.2.4 las órbitas giran, y la razón es que las dos especies se empujan en sentidos opuestos: más presas hacen crecer a los depredadores, y más depredadores hacen bajar a las presas. Ese desfase es lo que produce el ciclo. En la simbiosis las dos se empujan en el *mismo* sentido: más u hace crecer a v y más v hace crecer a u. En este sistema, el argumento por regiones y el cálculo de Dulac demuestran que no hay ciclos: una trayectoria no puede recorrer sucesivamente las cuatro zonas y volver al punto inicial. Las regiones R_4 y R_1 son invariantes hacia delante, una con destino al origen y otra con crecimiento no acotado.

Resolución del ejercicio 2

Qué problema es. Es el problema de **observación final** de la sección 4.2 con dos retoques: se quiere llegar cerca de un objetivo y_d y no cerca de cero, y el precio del fallo final se mide con una **cuarta** potencia en vez de con un cuadrado. El segundo retoque es el que da carácter al ejercicio: con $\frac{\alpha}{2}|y(T)|^2$ el funcional es cuadrático y todo se hace con la teoría del capítulo tal cual; con $\frac{\alpha}{4}|y(T)-y_d|^4$ ya no lo es, y hay que ver qué propiedades del cuadrado se han usado. El término cuártico conserva la convexidad, la continuidad en $y(T)$ y la no negatividad, que permite obtener directamente la cota coerciva del apartado (b).

Apartado (a): J está bien definido y es estrictamente convexo

Paso 1: el estado existe y $y(T)$ es un número. Para cada $v \in L^2(0,T)$, el problema (C.6) tiene una única solución absolutamente continua, que satisface la ecuación casi en todo punto y viene dada por variación de constantes (lema de la ecuación de estado):

$$y(t) = e^{at}y_0 + \int_0^t e^{a(t-s)}v(s)\,ds, \qquad t \in [0,T]. \tag{C.11}$$

Que la integral tenga sentido con v solo de cuadrado integrable es la desigualdad de Cauchy-Schwarz: $\int_0^T |v| \le \sqrt{T}\,\|v\|_2 < \infty$. Y de (C.11),

$$|y(T)| \le e^{|a|T}|y_0| + e^{|a|T}\int_0^T |v(s)|\,ds \le e^{|a|T}\big(|y_0| + \sqrt{T}\,\|v\|_2\big) < \infty.$$

Luego los dos sumandos de $J(v)$ son números reales finitos y $J : L^2(0,T) \to \mathbb{R}$ está bien definido. (Además $J \geq 0$, cosa que hará falta en (b).)

Paso 2 (el núcleo): $v \mapsto y(T)$ es afín y continua. Escríbase (C.11) en $t = T$:

$$y_v(T) = \underbrace{e^{aT}y_0}_{\text{constante}} + \underbrace{\int_0^T e^{a(T-s)}v(s)\,ds}_{=:\Lambda(v)}, \qquad \Lambda(v) = \langle v, g\rangle_{L^2}, \quad g(s) := e^{a(T-s)}. \tag{C.12}$$

Λ es una aplicación **lineal** de $L^2(0,T)$ en $\mathbb{R}$, y es continua porque $|\Lambda(v)| \leq \|g\|_2\|v\|_2$ (Cauchy-Schwarz otra vez), con $g \in L^2(0,T)$ por ser continua en un intervalo acotado. Por tanto $v \mapsto y_v(T) - y_d$ es **afín** y continua.

Esta observación es la que sostiene todo el ejercicio, y conviene decirla en voz alta: la ecuación de estado es lineal, así que el mapa control $\to$ estado final no es más que un producto escalar con una exponencial fija, más una constante. Todo lo que se pregunta sobre J es, en realidad, una pregunta sobre la función $w \mapsto \frac{\alpha}{4}|w|^4$ compuesta con una recta.

Paso 3: el primer término es convexo. La función real $h(w) = \frac{\alpha}{4}|w|^4 = \frac{\alpha}{4}w^4$ es de clase C^2 con $h''(w) = 3\alpha w^2 \geq 0$, luego es convexa en $\mathbb{R}$. Y una función convexa compuesta con una aplicación afín es convexa: si $\Phi(v) = y_v(T) - y_d$ es afín,

$$\Phi(\theta v + (1-\theta)w) = \theta\Phi(v) + (1-\theta)\Phi(w),$$
$$h(\Phi(\theta v + (1-\theta)w)) \leq \theta h(\Phi(v)) + (1-\theta)h(\Phi(w)).$$

Luego $v \mapsto \frac{\alpha}{4}|y_v(T) - y_d|^4$ es convexo en $L^2(0,T)$. *No* es estrictamente convexo, y no puede serlo: Λ tiene un núcleo enorme (todas las v ortogonales a g), y sobre una recta contenida en ese núcleo el primer término es constante.

Paso 4: el segundo término es estrictamente convexo, y eso basta. Para $v \neq w$ en $L^2(0,T)$ y $\theta \in (0,1)$, la identidad del paralelogramo da

$$\|\theta v + (1-\theta)w\|_2^2 = \theta\|v\|_2^2 + (1-\theta)\|w\|_2^2 - \theta(1-\theta)\|v-w\|_2^2 < \theta\|v\|_2^2 + (1-\theta)\|w\|_2^2,$$

porque $\theta(1-\theta) > 0$ y $\|v-w\|_2 > 0$. Multiplicando por $\mu/2 > 0$, el segundo término es estrictamente convexo. Sumando: convexo más estrictamente convexo es estrictamente convexo, luego

$$\boxed{J \text{ es estrictamente convexo en } L^2(0,T).}$$

La idea: de dónde sale la convexidad estricta, y qué pasaría sin el término μ

El primer término no distingue dos controles que produzcan el mismo estado final; y hay infinitos que lo hacen, porque para llegar a un sitio en el instante T hay infinitas maneras. Sin el término $\frac{\mu}{2}\int|v|^2$ el problema tendría infinitas soluciones y ninguna razón para preferir una. Ese término es el que dice «de todas las maneras de llegar, la más barata», y es el que rompe el empate: es **estrictamente** convexo porque la norma de un espacio de Hilbert al cuadrado lo es, y esa es toda la razón por la que el control óptimo es único. El parámetro μ se llama a veces de *regularización* por esto mismo.

Apartado (b): existencia y unicidad

La idea. La unicidad es gratis con (a). La existencia es el método directo del cálculo de variaciones, en su versión más limpia: se toma una sucesión que se acerque al ínfimo, se ve que está acotada (porque J crece al alejarse), se extrae de ella una subsucesión débilmente convergente (porque L^2 es Hilbert) y se comprueba que el límite no puede tener un valor mayor que el ínfimo. La única dificultad real es que la convergencia débil no basta para pasar al límite en cualquier término; aquí sí basta, y el motivo es (C.12).

Paso 1: unicidad. Si $\tilde{v}_1 \neq \tilde{v}_2$ fueran dos mínimos con $J(\tilde{v}_1) = J(\tilde{v}_2) = m$, la convexidad estricta con $\theta = 1/2$ daría

$$J\Big(\frac{\tilde{v}_1 + \tilde{v}_2}{2}\Big) < \frac{J(\tilde{v}_1) + J(\tilde{v}_2)}{2} = m,$$

contra la definición de m como mínimo. Luego hay a lo sumo uno.

Paso 2: coercividad y sucesión minimizante acotada. Como $J \geq 0$, el ínfimo $m := \inf_{L^2} J$ existe y $0 \leq m \leq J(0) < \infty$. Tómese (v_n) con $J(v_n) \to m$. De

$$\frac{\mu}{2}\|v_n\|_2^2 \leq J(v_n) \leq m + 1 \quad \text{para } n \text{ grande}$$

se sigue que (v_n) está acotada en $L^2(0,T)$.

Paso 3 (el núcleo): la convergencia débil basta. $L^2(0,T)$ es un espacio de Hilbert, luego reflexivo, y toda sucesión acotada tiene una subsucesión débilmente convergente: existe $\tilde{v} \in L^2(0,T)$ y una subsucesión (que se sigue llamando v_n) con $v_n \rightharpoonup \tilde{v}$. Ahora se pasa al límite en los dos términos, y cada uno por su motivo:

- *El primero converge, no solo semiconverge.* Por (C.12), $y_n(T) = e^{aT}y_0 + \langle v_n, g\rangle$, y la definición misma de convergencia débil es que $\langle v_n, g\rangle \to \langle \tilde{v}, g\rangle$ para toda $g \in L^2$. Luego $y_n(T) \to \tilde{y}(T)$ en $\mathbb{R}$, y por continuidad de $w \mapsto \frac{\alpha}{4}|w - y_d|^4$,

$$\frac{\alpha}{4}|y_n(T) - y_d|^4 \longrightarrow \frac{\alpha}{4}|\tilde{y}(T) - y_d|^4.$$

  Aquí está el punto fino de todo el argumento: **el estado final es un funcional afín continuo del control**; su parte variable, Λ, es lineal continua y por eso respeta la convergencia débil.
- *Para el segundo basta una desigualdad con el límite inferior.* La norma es débilmente semicontinua inferiormente: $\|\tilde{v}\|_2 \leq \liminf\|v_n\|_2$. (Se ve en una línea: $\|\tilde{v}\|_2^2 = \lim\langle v_n, \tilde{v}\rangle \leq \liminf\|v_n\|_2\|\tilde{v}\|_2$.)

Sumando las dos cosas,

$$J(\tilde{v}) \leq \liminf_n J(v_n) = m.$$

Y como m es el ínfimo, $J(\tilde{v}) = m$: $\tilde{v}$ es solución. Con el paso 1, es la única. ■

Qué hace cada hipótesis.

- $\mu > 0$ da la coercividad (paso 2) y la unicidad (paso 1). Con $\mu = 0$ se pierden las dos propiedades. En este problema, si $\mu = 0$, hay infinitos controles óptimos: basta elegir $\langle v, g\rangle = y_d - e^{aT}y_0$. Como $g \neq 0$, esa igualdad es alcanzable; pueden añadirse libremente controles ortogonales a g.
- $\alpha > 0$ conserva la convexidad y la no negatividad del término terminal; con $\alpha = 0$ la solución sería $\tilde{v} = 0$.
- La misma prueba sirve con $|y(T) - y_d|^p$, $p \geq 1$, porque conserva esas propiedades. Para un término $F(y(T))$ convexo y continuo que pueda tomar valores negativos, se modifica la cota del paso 2: una minorante afín $F(r) \geq \eta r + \zeta$ da $J(v) \geq \frac{\mu}{2}\|v\|_2^2 - C\|v\|_2 - D$, que también es coerciva.

- La linealidad de la ecuación de estado hace afín el mapa $v \mapsto y(T)$ y permite los dos pasos de esta prueba: convexidad y continuidad débil del término terminal. Con otra ecuación de estado habría que comprobar de nuevo esas propiedades.

Complemento: el sistema de optimalidad y el cálculo del control

El examen se detenía en la existencia, pero el problema se termina, y merece la pena porque el resultado es sorprendentemente concreto. Repítase la deducción del sistema de optimalidad con observación final con el nuevo término terminal. Sea $\tilde{v}$ el óptimo, $\tilde{y}$ su estado, $w \in L^2(0,T)$ arbitraria y z la solución de $z' = az + w$, $z(0) = 0$ (la derivada del estado en la dirección w). Derivando $\varepsilon \mapsto J(\tilde{v} + \varepsilon w)$ en $\varepsilon = 0$, y usando que $\frac{d}{dw}\frac{1}{4}|w|^4 = |w|^2 w$,

$$0 = J'(\tilde{v})\, w = \alpha|\tilde{y}(T) - y_d|^2\big(\tilde{y}(T) - y_d\big)z(T) + \mu \int_0^T \tilde{v}\, w\, dt.$$

Introdúzcase el estado adjunto $\tilde{\varphi}$ como solución de

$$-\tilde{\varphi}' = a\tilde{\varphi} \text{ en } (0,T), \qquad \tilde{\varphi}(T) = \alpha|\tilde{y}(T) - y_d|^2\big(\tilde{y}(T) - y_d\big).$$

Entonces $(\tilde{\varphi}z)' = \tilde{\varphi}'z + \tilde{\varphi}z' = -a\tilde{\varphi}z + \tilde{\varphi}(az + w) = \tilde{\varphi}w$, e integrando en $(0,T)$ con $z(0) = 0$ queda $\tilde{\varphi}(T)z(T) = \int_0^T \tilde{\varphi}w$. Sustituyendo arriba, $\int_0^T (\tilde{\varphi} + \mu\tilde{v})w = 0$ para toda $w \in L^2$, y por el lema fundamental,

$$\boxed{\tilde{v} = -\frac{1}{\mu}\tilde{\varphi}.}$$

Es el mismo sistema de optimalidad de (4.14), con la única diferencia de la condición terminal del adjunto: $\alpha\tilde{y}(T)$ se convierte en $\alpha|\tilde{y}(T) - y_d|^2(\tilde{y}(T) - y_d)$, que ya no es lineal en $\tilde{y}(T)$.

Y todavía se puede llegar hasta el final. De $-\tilde{\varphi}' = a\tilde{\varphi}$ sale $\tilde{\varphi}(t) = \tilde{\varphi}(T)e^{a(T-t)}$, luego $\tilde{v}(t) = -\frac{\tilde{\varphi}(T)}{\mu}e^{a(T-t)}$. Llevándolo a (C.11) en $t = T$ y llamando $z := \tilde{y}(T) - y_d$,

$$z = e^{aT}y_0 - y_d - \frac{\alpha z^3}{\mu}\int_0^T e^{2a(T-s)}\, ds = e^{aT}y_0 - y_d - \frac{\alpha c}{\mu}z^3, \qquad c := \int_0^T e^{2a(T-s)}ds = \frac{e^{2aT} - 1}{2a}$$

(y $c = T$ si $a = 0$). Es decir, **todo el problema se reduce a una ecuación de tercer grado**:

$$\frac{\alpha c}{\mu}\, z^3 + z = e^{aT}y_0 - y_d.$$

Como $c > 0$, la función $z \mapsto \frac{\alpha c}{\mu}z^3 + z$ es estrictamente creciente y suprayectiva de $\mathbb{R}$ en $\mathbb{R}$, así que esa ecuación tiene *exactamente una* raíz real. Lo cual vuelve a demostrar, ahora con las manos, la unicidad del apartado (b). Con esa z,

$$\tilde{v}(t) = -\frac{\alpha z^3}{\mu}\, e^{a(T-t)}.$$

LA IDEA: POR QUÉ EL CONTROL ÓPTIMO TIENE ESTE PERFIL EXPONENCIAL

$\tilde{\varphi}(t) = \tilde{\varphi}(T)e^{a(T-t)}$ es el precio sombra del estado: lo que cuesta, medido en unidades del objetivo final, mover el estado un poquito en el instante t. Como el estado evoluciona con $y' = ay$, un empujón dado en t se amplifica por $e^{a(T-t)}$ hasta el instante final, así que **cuanto antes se actúa, más rinde** (si $a > 0$). Y el control óptimo es proporcional

> a ese precio con signo cambiado: se gasta donde el gasto rinde. Si $a < 0$, al revés: el sistema olvida lo que se hizo pronto y conviene actuar al final. La cuarta potencia solo cambia *cuánto* se gasta en total, a través de la raíz de la cúbica, no *cómo* se reparte en el tiempo.

Resolución del ejercicio 3 (MATLAB)

La teoría que hay que tener delante antes de teclear. Este modelo se resuelve entero a mano, y quien lo haga primero programa mucho mejor después, porque sabe qué tiene que salir.

La primera ecuación es la desintegración de la sección 2.1.1: $A(t) = A_0 e^{-k_1 t}$. La segunda es lineal de primer orden con segundo miembro conocido, $B' + k_2 B = k_1 A_0 e^{-k_1 t}$; multiplicando por el factor integrante $e^{k_2 t}$, $(Be^{k_2 t})' = k_1 A_0 e^{(k_2 - k_1)t}$, y como $B(0) = 0$:

$$k_1 \neq k_2: \quad B(t) = \frac{k_1 A_0}{k_2 - k_1}\left(e^{-k_1 t} - e^{-k_2 t}\right), \qquad k_1 = k_2 = k: \quad B(t) = k A_0\, t\, e^{-kt}.$$

La primera fórmula es la del enunciado; la segunda es su límite cuando $k_2 \to k_1$ (una indeterminación $0/0$ que L'Hôpital resuelve), y es la que hace falta en la segunda ejecución que se pide. Por último, sumando las tres ecuaciones se obtiene $(A + B + C)' = 0$, es decir, la **conservación de la materia**: $A + B + C \equiv A_0$, de donde $C = A_0 - A - B$.

Con eso, las tres cantidades que pide el apartado (d) se conocen de antemano y sirven para comprobar el programa:

- *Dónde $A = C$.* Con $k_1 = 1$, $k_2 = 2$, $A_0 = 1$: $A = e^{-t}$, $B = e^{-t} - e^{-2t}$ y $C = 1 - 2e^{-t} + e^{-2t} = (1 - e^{-t})^2$. Llamando $x = e^{-t}$, la ecuación $A = C$ es $x = (1 - x)^2$, es decir $x^2 - 3x + 1 = 0$, cuya raíz en $(0, 1)$ es $x = \frac{3 - \sqrt{5}}{2} = 0.381966$. Luego

  $$t^* = \ln \frac{3 + \sqrt{5}}{2} = 0.962424, \qquad A(t^*) = C(t^*) = \frac{3 - \sqrt{5}}{2} = 0.381966.$$

  (Es el inverso del cuadrado de la razón áurea; aparece porque la ecuación es cuadrática con esos coeficientes, no por ningún motivo profundo.) Con $k_1 = k_2 = 1$: $A = e^{-t}$, $B = te^{-t}$, $C = 1 - (1 + t)e^{-t}$, y $A = C$ equivale a $(2 + t)e^{-t} = 1$, que es trascendente y hay que resolver numéricamente: $t^* = 1.146193$, $A = C = 0.317844$.
- *El máximo de B.* Derivando la fórmula de B para $k_1 \neq k_2$ e igualando a cero, $k_1 e^{-k_1 t} = k_2 e^{-k_2 t}$, de donde

  $$t_B = \frac{\ln(k_2/k_1)}{k_2 - k_1}.$$

  Con $k_1 = 1$, $k_2 = 2$: $t_B = \ln 2 = 0.693147$ y $B_{max} = \frac{1}{2} - \frac{1}{4} = 0.25$. Con $k_1 = k_2 = 1$, de $B' = (1 - t)e^{-t}$ sale $t_B = 1$ y $B_{max} = e^{-1} = 0.367879$; obsérvese que es el límite de la fórmula anterior cuando $k_2 \to k_1$.
- *El error.* Es una comprobación del integrador, no una propiedad del modelo. Lo que se compara es el vector de valores numéricos de B con el vector de valores exactos, en la misma malla.

La idea: por qué B sube y baja, y C tiene forma de ese

B es un embalse con una entrada y una salida. Al principio no hay nada dentro, la salida $k_2 B$ es nula y la entrada $k_1 A$ es máxima: el embalse se llena. Según A se agota, la entrada se seca, y llega un momento en que la salida supera a la entrada: es exactamente $k_1 A = k_2 B$, la condición del máximo. A partir de ahí B baja. Y C, que solo recibe, crece siempre, primero despacio (porque B es pequeña), luego deprisa (cuando B está en su máximo) y otra vez despacio (cuando B se agota): de ahí la forma de ese. La órbita (A, C) se recorre por tanto **de derecha a izquierda**, desde $(A_0, 0)$ hacia $(0, A_0)$, y es una curva monótona; se puede indicar ese sentido en la gráfica mediante los puntos inicial y final.

```
function ReacConsecut(k1, k2, A0, T)
% Nombre Apellido Apellido
% ReacConsecut(k1, k2, A0, T)
% Reaccion quimica consecutiva A --k1--> B --k2--> C:
% A' = -k1*A, B' = k1*A - k2*B, C' = k2*B, A(0)=A0, B(0)=C(0)=0.
% Casos pedidos: ReacConsecut(1, 2, 1, 7) y ReacConsecut(1, 1, 1, 7)

close all
if k1 <= 0 || k2 <= 0 || A0 <= 0 || T <= 0
   error('Se requieren k1, k2, A0 y T positivos')
end

% ---------- (a) Resolucion numerica en [0,T] --------------------------
F = @(t, y) [-k1*y(1); k1*y(1) - k2*y(2); k2*y(2)]; % y = [A; B; C]
y0 = [A0; 0; 0];
opciones = odeset('RelTol', 1e-8, 'AbsTol', 1e-10);
sol = ode45(F, [0 T], y0, opciones); % estructura, para usar deval
if sol.x(end) < T
   error('Integracion incompleta: no se puede evaluar todo [0,T]')
end
t = linspace(0, T, 1000); % malla fina donde evaluar y dibujar
Y = deval(sol, t);
A = Y(1,:); B = Y(2,:); C = Y(3,:);

% ---------- (b) Las tres trayectorias en la misma grafica -------------
figure(1)
hA = plot(t, A, 'b-', 'LineWidth', 1.6); hold on
hB = plot(t, B, 'r-', 'LineWidth', 1.6);
hC = plot(t, C, 'g-', 'LineWidth', 1.6);
xlabel('t (s)'), ylabel('concentracion (M)'), grid on
title(['Reaccion consecutiva A -> B -> C con k1 = ', num2str(k1), ...
       ', k2 = ', num2str(k2), ', A0 = ', num2str(A0)])

% ---------- (c) La orbita (A,C), con inicio y final -------------------
figure(2)
plot(A, C, 'k-', 'LineWidth', 1.6), hold on
plot(A(1), C(1), 'go', 'MarkerSize', 9, 'MarkerFaceColor', 'g')
plot(A(end), C(end), 'ro', 'MarkerSize', 9, 'MarkerFaceColor', 'r')
text(A(1), C(1), [' inicio (', num2str(A0), ', 0)'])
text(A(end), C(end), ' final')
xlabel('A (M)'), ylabel('C (M)'), grid on
```

```
title('Orbita (A,C): se recorre de derecha a izquierda')
legend('orbita', 'punto inicial', 'punto final', 'Location', 'north')

% ---------- (d1) Instante en que hay tanto A como C -------------------
difAC = @(s) deval(sol, s, 1) - deval(sol, s, 3); % A(s) - C(s)
if difAC(0)*difAC(T) <= 0
   tAC = fzero(difAC, [0, T]);
   vAC = deval(sol, tAC, 1);
   fprintf('A = C en t = %8.4f s, con A = C = %8.4f M \n', tAC, vAC)
   figure(1), hAC = plot(tAC, vAC, 'k.', 'MarkerSize', 22);
else
   disp('A y C no llegan a igualarse dentro de [0,T]')
   hAC = [];
end

% ---------- (d2) Cantidad maxima de B y cuando se alcanza -------------
% B tiene un solo pico: se busca en todo el intervalo.
tB = fminbnd(@(s) -deval(sol, s, 2), 0, T);
% fminbnd busca en el interior: tambien se comparan ambos extremos.
candidatos = [0, tB, T];
[BM, jmax] = max(deval(sol, candidatos, 2));
tB = candidatos(jmax);
fprintf('B maximo = %8.4f M, alcanzado en t = %8.4f s \n', BM, tB)
figure(1), hBM = plot(tB, BM, 'r.', 'MarkerSize', 22);

% ---------- (d3) Error frente a la solucion exacta de B ---------------
if k2 ~= k1
   d = abs(k2-k1);
   % Formula equivalente; evita restar exponenciales casi iguales.
   Bex = k1*A0*exp(-min(k1,k2)*t).*(-expm1(-d*t))/d;
   etiq = 'B exacta';
else
   disp('k1 = k2: la formula del enunciado da 0/0 y no se puede usar.')
   disp('Se usa su limite k2 -> k1, que es B = k1*A0*t*exp(-k1*t).')
   Bex = k1*A0*t.*exp(-k1*t);
   etiq = 'B exacta (caso limite)';
end
fprintf('Error ||B_num - B_exacta||_2 = %10.3e (malla de %i puntos) \n', ...
       norm(B - Bex), numel(t))
fprintf('Error maximo puntual = %10.3e \n', max(abs(B - Bex)))
figure(1), hE = plot(t(1:50:end), Bex(1:50:end), 'ko');
man = [hA hB hC];
nombres = {'A(t)', 'B(t)', 'C(t)'};
if ~isempty(hAC)
   man = [man hAC];
   nombres = [nombres {'A = C'}];
end
legend([man hBM hE], [nombres {'B maximo', etiq}], 'Location', 'east')
shg
end
```

Valores con los que comprobar la ejecución. Para $k_1 = 1, k_2 = 2, A_0 = 1, T = 7$, deben obtenerse aproximadamente $t_{A=C} = 0{,}9624$, $A = C = 0{,}3820$, $t_B = 0{,}6931$ y $B_{\text{máx}} = 0{,}2500$. Para $k_1 = k_2 = 1$, los valores son $t_{A=C} = 1{,}1462$, $A = C = 0{,}3178$, $t_B = 1$ y $B_{\text{máx}} = 0{,}3679$.

Son referencias analíticas; las cifras impresas por el programa dependen de su aproximación numérica.

Qué mide cada error. El programa calcula sobre la misma malla el vector $e_i = B_{\text{num}}(t_i) - B_{\text{exacta}}(t_i)$ y muestra $\|e\|_2$ y $\max_i |e_i|$. Sus valores deben medirse en la ejecución, no fijarse como datos del problema.

- Las tolerancias controlan el integrador; reducirlas suele mejorar la precisión, pero no garantizan una cifra concreta del error global. Evaluar con `deval` en más puntos no mejora por sí solo la solución ya calculada.
- $\|e\|_2 = \sqrt{n}\,(\|e\|_2/\sqrt{n})$: la norma euclídea incluye un factor $\sqrt{n}$ frente al error cuadrático medio. En mallas uniformes suficientemente finas, si el perfil de error se mantiene, $\|e\|_2/\sqrt{n}$ aproxima $\left(T^{-1}\int_0^T |e(t)|^2\,dt\right)^{1/2}$. No es exactamente independiente de la malla. El máximo muestreado también puede cambiar al incluir puntos nuevos, aunque no lleva ese factor $\sqrt{n}$.

Se incluyen los extremos al buscar el máximo de B en $[0, T]$: si $T < t_B$, el máximo observado está en T. Para $k_1 \neq k_2$, con $d = |k_2 - k_1|$, la fórmula del enunciado se puede escribir

$$B(t) = \frac{k_1 A_0}{d} e^{-\min(k_1,k_2)t}(1 - e^{-dt}).$$

El programa usa `-expm1(-d*t)` para calcular $1 - e^{-dt}$ sin restar dos números casi iguales cuando las tasas son próximas. Si coinciden, utiliza el límite ya calculado.

Órdenes que no habían aparecido antes. `clf` no se usa: basta `close all` al principio. `'MarkerFaceColor'` rellena el interior del marcador, para que un círculo se vea como un punto grueso. `odeset('RelTol',r,'AbsTol',a)` fija las tolerancias de `ode45`, igual que `bvpset` las de `bvp4c` en la práctica de la viga elástica. Y el uso de manejadores `hA = plot(...)` con `legend([hA hB ...],...)` —que ya apareció en el ejercicio de regresión— permite asociar cada nombre con su curva. Si A y C no se igualan en $[0, T]$, se omiten tanto el marcador como su nombre: quitar solo uno desplazaría las etiquetas.

Resolución del ejercicio 4 (MATLAB)

Qué se pide, en una frase. Los dos ejercicios de prácticas anteriores del libro resolvían un modelo con parámetros conocidos. Este hace lo contrario: se conocen unos pocos valores medidos de una de las poblaciones y hay que **averiguar los parámetros** que los producen. Es un *problema inverso*, y es lo que más se repite en la parte práctica de esta asignatura: los ejercicios Riccati inverso y Lotka–Volterra inverso del capítulo 4 son el mismo esquema.

La formulación de mínimos cuadrados, que el enunciado pide explícitamente. Sea $p = (\beta_1, \delta_2)$. Para cada $p \in [0, 1]^2$, el campo polinómico da existencia y unicidad local. Las tasas son no negativas, el campo apunta hacia dentro en cada cara del ortante no negativo y $S + E + A + I + R = N_0$ se conserva. La solución permanece por tanto en un simplejo compacto y existe para todo $t \geq 0$. Así queda bien definida $I(t; p)$ en los tiempos observados. Las igualdades pueden no ser compatibles exactamente con el modelo; tener seis datos y dos parámetros no lo decide por sí solo. Los mínimos cuadrados buscan la mejor aproximación:

$$\boxed{\min_{p\in[0,1]^2} \Phi(p), \qquad \Phi(p) := \sum_{i=1}^{m} \big|I(t_i; p) - I_i\big|^2.}$$

LA IDEA: UNA REGRESIÓN CUYO MODELO SE EVALÚA RESOLVIENDO EDO

En una regresión lineal, la dependencia de la predicción respecto de los parámetros es lineal, el funcional es cuadrático y las ecuaciones normales caracterizan sus mínimos. Aquí la regresión es no lineal: $I(t_i; p)$ depende de p a través de un sistema de EDO que el programa resuelve numéricamente. Cada evaluación de Φ cuesta una integración numérica completa del sistema, y el mínimo hay que buscarlo iterando. Eso tiene tres consecuencias prácticas que conviene tener presentes: el algoritmo puede encontrar resultados distintos según el punto de partida, no ofrece una garantía de mínimo global, y conviene dar tolerancias finas al integrador, porque el ruido numérico de `ode45` se traduce en un Φ rugoso que despista al minimizador.

El programa. Se minimiza con `fminsearch`, que es el minimizador sin derivadas que ya usan los ejercicios Riccati inverso y Lotka–Volterra inverso.

```
function [beta1, delta2] = InvSEAIR(T, tobs, iobs, beta10, delta20)
% Nombre Apellido Apellido
% [beta1, delta2] = InvSEAIR(T, tobs, iobs, beta10, delta20)
% Problema inverso para un modelo SEAIR: calcula beta1 y delta2 de modo que
% la componente I(t) del modelo pase lo mas cerca posible de los valores
% observados iobs en los instantes tobs (minimos cuadrados).
% Ejecucion pedida:
% tobs = [1 5 9 12 15 30];
% iobs = [14.1927 62.2850 92.4531 107.0014 116.6409 126.7263];
% [b1, d2] = InvSEAIR(60, tobs, iobs, 0.1, 0)

% ---------- Datos fijos del enunciado ---------------------------------
y0 = [800; 10; 5; 3; 182]; % [S0; E0; A0; I0; R0]
fijo = [0.012, 0.005, 0.08, 0.02, 0.03]; % [beta2 beta3 lambda1 lambda2 delta1]
tobs = tobs(:)'; iobs = iobs(:)'; % filas, para poder restar
if T <= 0 || isempty(tobs) || numel(tobs) ~= numel(iobs) || ...
   any(~isfinite([tobs iobs])) || any(tobs < 0 | tobs > T) || ...
   any([beta10 delta20] < 0 | [beta10 delta20] > 1)
   error('Revise tiempos, observaciones y parametros iniciales en [0,1]')
end

% ---------- (a) El problema de minimos cuadrados ----------------------
% p = [beta1, delta2]; Phi(p) = suma de los cuadrados de los residuos
Phi = @(p) CosteSEAIR(p, tobs, iobs, T, y0, fijo);

opciones = optimset('Display', 'iter', 'TolX', 1e-10, 'TolFun', 1e-14);
[p, valor, flag] = fminsearch(Phi, [beta10, delta20], opciones);
if ~isfinite(valor) || any(p < 0 | p > 1)
   error('No se ha encontrado un ajuste admisible')
end
if flag <= 0
   warning('Se alcanzo el limite de iteraciones o evaluaciones; revise el ajuste')
end

beta1 = p(1); delta2 = p(2);
fprintf('\n beta1 = %10.6f \n delta2 = %10.6f \n', beta1, delta2)
fprintf(' Suma de cuadrados de partida: %10.3e \n', Phi([beta10, delta20]))
fprintf(' Suma de cuadrados al final: %10.3e \n', Phi(p))

```

```
% ---------- (b) Las dos graficas ------------------------------------
close all
tt = linspace(0, T, 500);

figure(1) % con los parametros ajustados
plot(tt, Isim(p, tt, T, y0, fijo), 'b-', 'LineWidth', 1.8), hold on
plot(tobs, iobs, 'ro', 'MarkerSize', 9, 'MarkerFaceColor', 'r')
xlabel('t (dias)'), ylabel('I(t): infectados con sintomas'), grid on
title(['SEAIR ajustado: \beta_1 = ', num2str(beta1), ...
       ', \delta_2 = ', num2str(delta2)])
legend('I(t) con los parametros ajustados', 'valores observados', ...
       'Location', 'southeast')

figure(2) % con los valores de partida
plot(tt, Isim([beta10, delta20], tt, T, y0, fijo), 'b--', 'LineWidth', 1.8), hold on
plot(tobs, iobs, 'ro', 'MarkerSize', 9, 'MarkerFaceColor', 'r')
xlabel('t (dias)'), ylabel('I(t): infectados con sintomas'), grid on
title(['SEAIR de partida: \beta_1 = ', num2str(beta10), ...
       ', \delta_2 = ', num2str(delta20)])
legend('I(t) con beta10 y delta20', 'valores observados', 'Location', 'southeast')
shg
end

% =========================================================================
function valor = CosteSEAIR(p, tobs, iobs, T, y0, fijo)
% fminsearch es libre; esta funcion excluye los parametros no admisibles.
if any(~isfinite(p)) || any(p < 0 | p > 1)
   valor = Inf;
   return
end
valor = sum((Isim(p, tobs, T, y0, fijo) - iobs).^2);
end

function Iv = Isim(p, tt, T, y0, fijo)
% Resuelve el SEAIR con beta1 = p(1) y delta2 = p(2), y devuelve la cuarta
% componente del estado (los infectados con sintomas) en los instantes tt.
b1 = p(1); d2 = p(2);
b2 = fijo(1); b3 = fijo(2);
l1 = fijo(3); l2 = fijo(4); d1 = fijo(5);

F = @(t, y) [ -(b1*y(1)*y(2) + b2*y(1)*y(3) + b3*y(1)*y(4));
              (b1*y(1)*y(2) + b2*y(1)*y(3) + b3*y(1)*y(4)) - (l1 + l2)*y(2);
              l1*y(2) - d1*y(3);
              l2*y(2) - d2*y(4);
              d1*y(3) + d2*y(4) ]; % y = [S; E; A; I; R]

sol = ode45(F, [0 T], y0, odeset('RelTol', 1e-8, 'AbsTol', 1e-10));
if sol.x(end) < T || any(~isfinite(sol.y(:)))
   error('Integracion incompleta o no finita: no se evalua el ajuste')
end
Iv = deval(sol, tt, 4);
end
```

Una referencia numérica para comprobar el ajuste. Con los datos del enunciado se obtiene un ajuste cercano a

$$\beta_1 = 0{,}01800325, \qquad \delta_2 = 0{,}01000001, \qquad \Phi \approx 4{,}9 \cdot 10^{-9},$$

frente a $\Phi(0{,}1, 0) \approx 1{,}202 \cdot 10^3$. Son valores numéricos de referencia; los últimos dígitos dependen de la integración y de la parada del optimizador.

LA IDEA: QUÉ PERMITE CONCLUIR UN RESIDUO PEQUEÑO

Con seis observaciones, $\sqrt{\Phi/6} \approx 2{,}9 \cdot 10^{-5}$ es el tamaño cuadrático medio de los residuos, del orden del redondeo de los datos impresos. El ajuste reproduce muy bien estas observaciones. Eso no prueba que se hayan generado con el modelo, que los parámetros sean únicos ni que el modelo describa igual de bien observaciones nuevas. Tampoco hay que sustituir el parámetro ajustado por 0,018 y llamarlo exacto: con $(\beta_1, \delta_2) = (0{,}018, 0{,}01)$, una integración precisa da $I(1) \approx 14{,}191964$, que no redondea al 14,1927 del enunciado. Un residuo mayor puede deberse a datos, modelo, tolerancias o búsqueda del mínimo; conviene comprobar cada parte antes de atribuirlo a un error de programación.

Tres cosas que conviene saber de este programa.

- **Se comparan las cosas correctas.** La componente observada es I, la *cuarta* del estado, y por eso el `4` de `deval(sol, tt, 4)`. Confundirla con A (la tercera) es el error más frecuente aquí y produce un ajuste que converge a algo y no se parece a nada.
- `fminsearch` **no incorpora las cotas.** Por eso `CosteSEAIR` devuelve `Inf` fuera de $[0, 1]^2$, sin integrar un modelo con tasas no admisibles. La salida se comprueba antes de utilizarla. Es una forma sencilla de mantener este ejemplo dentro de las funciones básicas de MATLAB; no añade una garantía de convergencia o de mínimo global. Si se dispone de Optimization Toolbox, `lsqnonlin` permite imponer directamente las cotas y trabajar con el vector de residuos.
- **Las tolerancias del integrador importan más que de costumbre.** Con las de `ode45` por defecto, dos evaluaciones de Φ en puntos muy próximos pueden diferir por el ruido del propio integrador, y `fminsearch` se para antes de tiempo o da tumbos. De ahí el `odeset('RelTol',1e-8,'AbsTol',1e-10)`: no es un capricho, es una forma de reducir el error de integración en la función que se minimiza; no convierte una parada del algoritmo en un certificado de optimalidad.

LA IDEA: DISTINGUIR AJUSTE E IDENTIFICACIÓN

La curva ajustada tiene un pico cerca de $t = 25{,}68$, con $I \approx 127{,}75$, y después desciende: $I(40) \approx 119{,}62$ e $I(60) \approx 100{,}25$. El último dato, $t = 30$, está cerca del pico; no demuestra una meseta permanente. Con las tasas de recuperación positivas del ajuste, un equilibrio debe tener $A = I = 0$, porque $R' = \delta_1 A + \delta_2 I$; después, las ecuaciones de A e I obligan a $E = 0$.

Los dos parámetros dejan efectos distintos, pero no basta con verlos en una curva para afirmar que quedan identificados. Por ejemplo, $I'(0) = \lambda_2 E_0 - \delta_2 I_0$: δ_2 cambia directamente la pendiente inicial, mientras que β_1 influye a través de la evolución posterior de E.

> Para comprobar la separación local de sus efectos puede estudiarse la matriz de sensibilidades $\left(\partial I(t_i; p)/\partial p_j\right)$. Si sus dos columnas son independientes, los datos distinguen perturbaciones pequeñas de los parámetros cerca del ajuste; eso no descarta otros ajustes alejados. Repetir la búsqueda desde varios puntos y retirar algún dato son comprobaciones útiles de estabilidad numérica, pero ni un cambio grande prueba por sí solo falta de identificación, ni un cambio pequeño prueba identificación global.

Qué buscaba este examen

- **Ejercicio 1.** El recorrido completo del capítulo 2 sobre un modelo nuevo: adimensionalizar contando cuántos parámetros deben sobrevivir, calcular equilibrios, clasificarlos por linealización, usar el teorema escalar autónomo para describir la dinámica sobre un eje, deducir la positividad de la invariancia de los ejes, y leer el plano de fases por regiones de signo constante. Es un recorrido útil para repasar las técnicas de ese capítulo.
- **Ejercicio 2.** Si se ha entendido *qué* propiedades del funcional cuadrático se usan en la teoría del capítulo 4. La cuarta potencia está puesta a propósito para que no sirva repetir de memoria la demostración del libro.
- **Ejercicio 3.** Programar un modelo que se sabe resolver a mano, y responder con `fzero` y `fminbnd` a preguntas sobre la solución numérica; más una trampa fina, el caso $k_1 = k_2$ en el que la fórmula que da el propio enunciado se rompe.
- **Ejercicio 4.** El problema inverso: formularlo como mínimos cuadrados, escribir la función que hay que minimizar sabiendo que cada evaluación es una integración, y minimizar con `fminsearch`.

En el ejercicio 1, el recuento de parámetros orienta el cambio de escala y las ecuaciones permiten comprobarlo. En (b), el signo negativo del determinante basta para clasificar la silla sin calcular sus dos autovalores.

Errores típicos

- **Adimensionalizar cambiando solo las variables y no el tiempo.** Con dos unidades no se pueden matar tres constantes; el cambio $s = t/\tau$ es imprescindible y es el que suele olvidarse.
- **Clasificar el equilibrio** $(1 + \beta, 1)$ **como estable** porque la traza es negativa. La traza negativa solo garantiza estabilidad cuando el determinante es positivo; aquí es negativo y hay un autovalor positivo.
- **Contestar en (c) que la órbita de** $y_0 > 0$ **es** $[0, y_0]$**.** La órbita es el *conjunto de valores que recorre la solución en todo su intervalo maximal*, hacia adelante y hacia atrás; hacia atrás la solución explota, así que la órbita es $(0, +\infty)$ entera, y el 0 no está en ella porque no se alcanza.
- **En (d), decir «los ejes son invariantes» sin decir por qué eso da la positividad.** El eslabón que falta es que dos órbitas distintas no se cortan.
- **En (e), responder que no hay ciclos «porque el sistema no es de Volterra».** Hay que dar un argumento; el de las dos regiones invariantes vale y es corto.
- **En el ejercicio 2, intentar demostrar que el primer término es estrictamente convexo.** No lo es, y perder el tiempo intentándolo hace fallar el apartado. Lo estricto lo pone el término

$\frac{\mu}{2}\int |v|^2$.

- **Confundir semicontinuidad inferior débil con continuidad débil.** La norma no es débilmente continua (hay sucesiones débilmente convergentes cuya norma no converge a la del límite); solo es semicontinua inferiormente, y esa desigualdad va en el sentido que hace falta.
- **En el ejercicio 3, ejecutar el segundo caso sin proteger la división por $k_2 - k_1$.** MATLAB devuelve `NaN` y el programa sigue como si nada, imprimiendo un error `NaN`.
- **En el ejercicio 4, resolver el sistema una sola vez y ajustar después.** El sistema se resuelve dentro de la función objetivo, para cada parámetro que prueba el algoritmo.

C.5 Primera prueba de seguimiento del 5 de abril de 2024

Esta prueba de seguimiento reparte los diez puntos entre dos preguntas de teoría, de 2 y 3 puntos, y una práctica de 5.

Ficha

Asignatura	Modelización matemática, curso 2023/24
Prueba	Primera prueba de seguimiento
Fecha	5 de abril de 2024
Estructura	Tres ejercicios: $2 + 3 + 5 = 10$ puntos
	1. Teoría (2 p.): logística adimensionalizada y con emigración
	2. Teoría (3 p.): el modelo SIR con población constante
	3. Prácticas (5 p.): la M-función `Especies`

Una errata del enunciado. En el ejercicio 2 se lee «las funciones $R = R(t)$ e $I = I(t)$ representan el número infectados y recuperados en el instante t», con los papeles cruzados respecto de lo que dicen las ecuaciones. Las ecuaciones mandan: $R' = \gamma I$ hace de R el compartimento que solo recibe (los *recuperados*) y de I el de los *infectados*. Se resuelve con esa lectura.

Enunciado

Ejercicio 1 [Teoría: 2 puntos]. Supongamos que la evolución de una población viene regida por la ecuación logística

$$u' = \lambda u - \mu u^2$$

con $\lambda, \mu > 0$. Se pide:

a) Demuestra que se pueden escoger $p, r > 0$ tales que la función $\bar{u}(s) = p\,u(rs)$ verifica $\bar{u}' = \bar{u}(1 - \bar{u})$.

b) Supongamos que añadimos a la ecuación anterior (con $\bar{u}$ reemplazada por u) un término de emigración, de forma que la ecuación se escribe como

$$u' = u(1 - u) + a - bu, \tag{C.13}$$

con $a, b > 0$. Describe razonadamente el comportamiento de las soluciones para $t \geq 0$.

Ejercicio 2 [Teoría: 3 puntos]. Se considera el modelo de epidemias SIR con población constante N, el cual recordamos viene dado por

$$\begin{cases} R' = \gamma I, \\ I' = I\big(\beta(N - R - I) - \gamma\big), \\ I(0) = I_0, \quad R(0) = R_0, \end{cases} \tag{C.14}$$

con $N, \gamma, \beta > 0$, $I_0 > 0$, $0 \leq R_0$, $I_0 + R_0 \leq N$. Las funciones $I = I(t)$ y $R = R(t)$ representan el número de infectados y de recuperados en el instante t. El número de susceptibles viene dado por $N - R(t) - I(t)$.

Describir justificadamente el comportamiento para $t \geq 0$ de las soluciones del problema anterior, incluyendo las propiedades de crecimiento y decrecimiento de las funciones I, R.

Ejercicio 3 [Prácticas: 5 puntos]. Consideramos el siguiente modelo de interacción de dos especies:

$$\begin{cases} u' = r\,u\Big(1 - \dfrac{u}{100}\Big) - \alpha\,u\,v, \\ v' = -r\,v\Big(1 - \dfrac{v}{100}\Big) + \beta\,u\,v, \\ u(0) = u_0, \quad v(0) = v_0, \end{cases} \tag{C.15}$$

donde $u(t)$ y $v(t)$ representan la cantidad de individuos en el instante t y u_0, v_0 sus valores en el instante inicial $t = 0$, respectivamente. Escribir una M-función

```
function Especies(r, alpha, beta, u0, v0, T)
```

que sea capaz de hacer lo siguiente:

a) Resolver numéricamente el modelo (C.15) en el intervalo $[0, T]$ para unos valores de r, α, β, u_0 y v_0 dados.
b) Representar en la misma gráfica las trayectorias de $u(t)$ y $v(t)$ obtenidas. Añadir un título y las leyendas adecuadas para su comprensión.
c) Obtener una aproximación numérica del primer instante de tiempo t^* tal que $u(t^*) = v(t^*)$. Mostrar dicho resultado en la ventana gráfica del apartado anterior.
d) Representar gráficamente la órbita de la solución. Añadir a la gráfica las anotaciones necesarias para su correcta comprensión. Si $\beta = \alpha$, obtener una aproximación numérica del periodo de la órbita que se describe. Mostrar dicho resultado en la misma ventana gráfica.

Ejecutar el programa para los valores $r = 0.03$, $\alpha = 0.004$, $u_0 = 10$, $v_0 = 30$, $T = 1000$ y $\beta = \alpha$, $\alpha/2$.

Resolución del ejercicio 1

Apartado (a): la logística adimensionalizada

La idea. Es la misma cuenta de unidades que en el examen de 2025, en su versión más pequeña: dos constantes, dos unidades libres (una para la población y otra para el tiempo), cero constantes supervivientes. El enunciado incluso da el cambio hecho, $\bar{u}(s) = p\,u(rs)$, y solo pide encontrar p y r.

Paso 1: derivar y sustituir. Por la regla de la cadena,

$$\bar{u}'(s) = p\,r\,u'(rs) = p\,r\big[\lambda u(rs) - \mu u(rs)^2\big].$$

Como $u(rs) = \bar{u}(s)/p$,

$$\bar{u}'(s) = p\,r\Big[\lambda\frac{\bar{u}}{p} - \mu\frac{\bar{u}^2}{p^2}\Big] = r\lambda\,\bar{u} - \frac{r\mu}{p}\,\bar{u}^2.$$

Paso 2: identificar coeficientes. Se quiere $\bar{u}' = \bar{u} - \bar{u}^2$, luego

$$r\lambda = 1 \implies \boxed{r = \frac{1}{\lambda}}, \qquad \frac{r\mu}{p} = 1 \implies \boxed{p = r\mu = \frac{\mu}{\lambda}},$$

y los dos son positivos porque $\lambda, \mu > 0$. Hecho.

LA IDEA: QUÉ SON p Y r, CON NOMBRES

$r = 1/\lambda$ es el **tiempo característico** de la población: el tiempo que tardaría en multiplicarse por e si creciera sin freno. Medir el tiempo en esa unidad es la razón de que la tasa de crecimiento se vuelva 1. Y $1/p = \lambda/\mu$ es la **capacidad de carga**: el equilibrio no trivial de la logística original. Medir la población en unidades de su propia capacidad de carga es la razón de que el segundo equilibrio se vuelva 1. La ecuación $\bar{u}' = \bar{u}(1-\bar{u})$ es, por tanto, *la* logística: cualquier otra es esta con otras etiquetas en los ejes, y todo lo que se demuestre para ella vale para todas.

Apartado (b): la logística con entrada constante

Qué dice el término nuevo. La ecuación es $u' = u(1-u) + a - bu$ con $a, b > 0$. El sumando $+a$ es un aporte *constante*, independiente del tamaño de la población: llegan a individuos por unidad de tiempo, haya los que haya. El sumando $-bu$ es una pérdida *proporcional*: se marcha una fracción fija b de la población por unidad de tiempo. (El enunciado llama al conjunto «término de emigración»; estrictamente, $+a$ es inmigración y $-bu$ emigración. Lo que importa para el análisis es el signo de cada uno, y son los que están escritos.) Es el modelo hermano del de pesca de la sección 2.1.4, donde el término constante iba con signo menos.

Paso 1: la función y sus ceros. Con

$$f(u) = u(1-u) + a - bu = -u^2 + (1-b)u + a,$$

que es una parábola hacia abajo. Sus ceros resuelven $u^2 - (1-b)u - a = 0$:

$$\bar{u}_\pm = \frac{(1-b) \pm \sqrt{(1-b)^2 + 4a}}{2}, \qquad D := (1-b)^2 + 4a.$$

Paso 2 (el núcleo): hay exactamente un equilibrio no negativo, y siempre. Aquí está todo el ejercicio, y se ve sin resolver nada: como $a > 0$, el producto de las dos raíces es $-a < 0$, luego **una es positiva y la otra negativa**, cualesquiera que sean a y b. Además $D > 0$, así que son reales y distintas. Por tanto

$$\bar{u}_- < 0 < \bar{u}_+,$$

y el único punto crítico en $[0, +\infty)$ es $\bar{u}_+$. Obsérvese lo que *no* pasa: $u = 0$ ya no es equilibrio, porque $f(0) = a > 0$. La población no se puede extinguir.

Paso 3: estabilidad. Con $f'(u) = (1 - b) - 2u$,

$$f'(\bar{u}_+) = (1-b) - \left[(1-b) + \sqrt{D}\right] = -\sqrt{D} < 0, \qquad f'(\bar{u}_-) = +\sqrt{D} > 0.$$

Luego $\bar{u}_+$ es exponencialmente asintóticamente estable y $\bar{u}_-$ inestable (teorema de linealización). Como $\bar{u}_- < 0$, la parte que interesa es solo $\bar{u}_+$.

Paso 4: el comportamiento para $t \geq 0$, con el teorema escalar. Sea $u_0 \geq 0$ el dato inicial y aplíquese el teorema de las EDO autónomas escalares. Como la parábola es positiva entre sus raíces y negativa fuera:

- Si $0 \leq u_0 < \bar{u}_+$, entonces $f(u_0) > 0$ (porque $\bar{u}_- < 0 \leq u_0 < \bar{u}_+$): la solución es **estrictamente creciente**, existe un punto crítico por encima —$\bar{u}_+$—, luego está definida en $[0, +\infty)$ y $u(t) \nearrow \bar{u}_+$.
- Si $u_0 = \bar{u}_+$, la solución es constante.
- Si $u_0 > \bar{u}_+$, entonces $f(u_0) < 0$: la solución es **estrictamente decreciente**, existe un punto crítico por debajo —$\bar{u}_+$ otra vez—, luego está definida en $[0, +\infty)$ y $u(t) \searrow \bar{u}_+$.

En resumen: *para todo dato inicial no negativo, la solución está definida en* $[0, +\infty)$*, es monótona y converge a* $\bar{u}_+$. El equilibrio $\bar{u}_+$ es globalmente asintóticamente estable en $[0, +\infty)$.

LA IDEA: POR QUÉ UN GOTEO CONSTANTE DE RECIÉN LLEGADOS BORRA LA EXTINCIÓN

En la logística pura, $u = 0$ es un equilibrio: sin nadie no nace nadie. Basta un aporte constante, por pequeño que sea, para que deje de serlo, y con él desaparece toda posibilidad de extinción para datos no negativos: si se parte de cero, la población empieza a crecer; si se parte de un valor positivo, permanece positiva. Compárese con el modelo de pesca con cuota fija de la sección 2.1.4, donde el término constante tiene el signo contrario: allí, si la cuota es demasiado alta, la parábola no corta al eje, no hay ningún equilibrio y la población se extingue *en tiempo finito*. El mismo término, con los dos signos, produce los dos extremos: con signo menos crea un umbral de extinción, y con signo más lo borra. Y obsérvese que b, por grande que sea, no cambia esta conclusión. Con $b > 1$ lo que se hace negativo es el *coeficiente lineal* $f'(0) = 1 - b$, no el crecimiento: es $f(0) = a > 0$ pase lo que pase, así que la población sigue creciendo cerca de cero y la extinción sigue siendo imposible. Lo único que hace b es bajar el nivel de equilibrio, y lo baja siempre: $\bar{u}_+ = \frac{(1-b)+\sqrt{(1-b)^2+4a}}{2}$ es estrictamente decreciente en b y estrictamente positivo para todo $b > 0$ (con $a = 0.5$ pasa de 1.288 con $b = 0.1$ a 0.121 con $b = 5$, sin llegar nunca a cero).

Resolución del ejercicio 2

Cómo está escrito el modelo. Es el SIR de la sección 2.2.1, con la población total eliminada: como $S + I + R \equiv N$, se puede escribir $S = N - R - I$ y quedarse con dos ecuaciones en vez de tres. Conviene recuperar S desde el principio, porque el modelo se lee mucho mejor con ella. Definiendo $S(t) := N - R(t) - I(t)$,

$$S' = -R' - I' = -\gamma I - I(\beta S - \gamma) = -\beta S I,$$

de modo que el sistema es, exactamente,

$$S' = -\beta S I, \qquad I' = I(\beta S - \gamma), \qquad R' = \gamma I, \tag{C.16}$$

con $S_0 = N - I_0 - R_0 \geq 0$ por hipótesis. La pregunta pide describir el comportamiento; lo que sigue es el guion completo, en el orden en que cada pieza necesita a la anterior.

Paso 1: existencia y unicidad. El segundo miembro de (C.14) es polinómico en (I, R), luego de clase C^∞ en todo $\mathbb{R}^2$ y localmente lipschitziano (criterio C^1). Por el teorema de existencia y unicidad maximal hay una única solución maximal definida en un intervalo abierto $I(I_0, R_0) \ni 0$. Los puntos críticos son los de $I = 0$ con R arbitrario: una recta entera de equilibrios, como en el SIR del capítulo. La linealización tiene autovalores 0 y $\beta(N - R) - \gamma$: si el segundo es positivo, sí permite concluir inestabilidad. Para describir toda la evolución de estas soluciones usaremos positividad, cotas y monotonías.

Paso 2 (el núcleo): positividad, por la fórmula exponencial. Las tres ecuaciones de (C.16) tienen la incógnita como factor, salvo la de R. A lo largo de la solución:

$$I(t) = I_0 \exp\Big(\int_0^t (\beta S - \gamma)\Big) > 0, \qquad S(t) = S_0 \exp\Big(-\beta \int_0^t I\Big) \geq 0,$$

las dos válidas en todo el intervalo maximal, porque los integrandos son continuos. La primera da $I(t) > 0$ *siempre*, que es lo que hace falta para todo lo demás. La segunda da $S(t) > 0$ si $S_0 > 0$ y $S \equiv 0$ si $S_0 = 0$.

Paso 3: monotonías. Con $I > 0$ y $S \geq 0$, en todo el intervalo:

- $R' = \gamma I > 0$: R **es estrictamente creciente, siempre**. Es la conclusión más fácil y la más útil: los recuperados no se pierden.
- $S' = -\beta S I \leq 0$, y < 0 si $S_0 > 0$: S **es estrictamente decreciente** (nadie vuelve a ser susceptible) si $S_0 > 0$; si $S_0 = 0$, es constante.
- $I' = I(\beta S - \gamma)$: el signo de I' es el de $\beta S - \gamma$, y como S solo baja, ese signo cambia como mucho una vez, de $+$ a $-$. I **no es monótona en general**: sube y baja.

Paso 4: prolongación a $[0, +\infty)$. Para $t \geq 0$ las tres funciones son no negativas y suman N, luego $0 < I \leq N$, $R_0 \leq R \leq N$ y $0 \leq S \leq S_0$. La semiórbita positiva vive en el triángulo compacto $\{(I, R) : I \geq 0,\ R \geq R_0,\ I + R \leq N\}$, contenido en el dominio $\mathbb{R}^2$; por la propiedad de prolongación en compactos, $I(I_0, R_0) \supset [0, +\infty)$.

Paso 5: los límites. R es creciente y acotada por N, luego $R(t) \to R_\infty \leq N$; S es decreciente y acotada inferiormente por 0, luego $S(t) \to S_\infty \geq 0$; y entonces $I = N - S - R \to I_\infty$. Falta ver que $I_\infty = 0$, y sale del propio R: de $R' = \gamma I$,

$$\gamma \int_0^\infty I(t)\, dt = R_\infty - R_0 < \infty,$$

así que I es integrable en $[0, \infty)$; una función con límite e integrable tiene límite cero. Luego $I_\infty = 0$ y $S_\infty + R_\infty = N$: **los infectados tienden a cero**, aunque $I(t) > 0$ en todo instante

finito. El punto límite $(0, R_\infty)$ es uno de los equilibrios de la recta $I = 0$, como tenía que ser (propiedad del límite de una órbita).

Paso 6: el umbral, que es la parte que hay que discutir. Escríbase $\mathcal{R}_{\text{ef}}(0) := \beta S_0/\gamma$, el factor de reproducción efectivo inicial; el básico, para una población enteramente susceptible, sería $\beta N/\gamma$. Por el paso 3, el signo de I' es el de $\beta S(t) - \gamma$, con S estrictamente decreciente si $S_0 > 0$.

- **Si** $\mathcal{R}_{\text{ef}}(0) \leq 1$, o sea $\beta S_0 \leq \gamma$: entonces $\beta S(t) - \gamma < 0$ para todo $t > 0$, luego I es estrictamente decreciente en $(0, +\infty)$ y baja a 0 sin más. *No hay brote*: la epidemia se apaga desde el primer día.
- **Si** $\mathcal{R}_{\text{ef}}(0) > 1$: entonces $I'(0) > 0$ e I empieza creciendo. Y S acaba bajando del umbral γ/β, porque si no lo hiciera se tendría $I' > 0$ para todo $t \geq 0$, luego $I \geq I_0$, y entonces $S' = -\beta SI \leq -\beta(\gamma/\beta)I_0 = -\gamma I_0 < 0$ daría $S(t) \leq S_0 - \gamma I_0 t \to -\infty$, imposible. Como S es continua y estrictamente decreciente, existe un **único** $T_{max} > 0$ con $\beta S(T_{max}) = \gamma$, y

  $$I \text{ estrictamente creciente en } (0, T_{max}), \qquad I \text{ estrictamente decreciente en } (T_{max}, +\infty).$$

  Hay brote, con un único pico.

Paso 7: cuántos se libran. Dividiendo las ecuaciones de S y R,

$$\frac{dS}{dR} = \frac{-\beta SI}{\gamma I} = -\frac{\beta}{\gamma} S \implies S = S_0\, e^{-\beta(R-R_0)/\gamma}$$

a lo largo de toda la órbita. Como $R(t) \to R_\infty \leq N$, el exponente tiene un límite finito, y $S_\infty = S_0 e^{-\beta(R_\infty - R_0)/\gamma} > 0$ si $S_0 > 0$: **no se contagia todo el mundo**, ni siquiera con $\mathcal{R}_{\text{ef}}(0)$ grande. Y pasando al límite con $R_\infty = N - S_\infty$ se obtiene la ecuación del tamaño final,

$$S_\infty = S_0\, e^{-\beta(N - S_\infty - R_0)/\gamma},$$

trascendente en S_∞, que se resuelve numéricamente en el ejercicio de programación del SIR.

Caso degenerado, que conviene no olvidar. Si $S_0 = 0$ (es decir, $I_0 + R_0 = N$), entonces $S \equiv 0$, la ecuación de I se reduce a $I' = -\gamma I$ y todo es explícito: $I(t) = I_0 e^{-\gamma t}$, $R(t) = N - I(t) \nearrow N$. No hay brote posible porque no queda a quién contagiar.

LA IDEA: POR QUÉ EL UMBRAL ESTÁ EN S Y NO EN I

Lo llamativo del resultado es que la condición de brote no dice nada sobre cuántos enfermos hay al principio, solo sobre cuántos sanos. Con un solo infectado y muchos susceptibles hay epidemia; con muchos infectados y pocos susceptibles, no. La razón está escrita en $I' = I(\beta S - \gamma)$: el crecimiento es *por individuo infectado*, y lo que decide si cada uno se reemplaza antes de curarse es cuánta gente susceptible encuentra. Para un mismo S_0, el signo de $I'(0)$ no depende del valor positivo de I_0. Pero I_0 sí influye en la evolución posterior: cambia la rapidez con que se agotan los susceptibles y, con ello, el pico y el tamaño final de la epidemia. En la convocatoria de 2026 aparece el mismo criterio con S/N, porque allí se usa una tasa de contagio normalizada por la población.

Resolución del ejercicio 3 (MATLAB)

Qué modelo es, antes de programar. La primera ecuación es la de una presa logística de capacidad 100 a la que un depredador se come a ritmo αuv. La segunda es más rara y conviene leerla despacio: $-rv(1-v/100) = -rv + \frac{r}{100}v^2$, que es *negativa* cuando $0 < v < 100$. Sin presas y con un dato en ese intervalo, el depredador tiende a extinguirse. Pero $v = 100$ es un equilibrio y, por encima de 100, el término cuadrático positivo provoca crecimiento: no es una autolimitación logística del depredador. En las dos ejecuciones pedidas, v permanece por debajo de 100.

El equilibrio de coexistencia, y por qué $\beta = \alpha$ es especial. Anulando los corchetes,

$$u' = 0: \quad v = \frac{r}{\alpha}\Big(1 - \frac{u}{100}\Big), \qquad v' = 0: \quad u = \frac{r}{\beta}\Big(1 - \frac{v}{100}\Big).$$

Con los datos del ejercicio, las dos rectas se cortan en un único punto $(\bar{u}, \bar{v})$ del cuadrante abierto. Con $r = 0.03$ y $\alpha = 0.004$ sale $\bar{u} = \bar{v} = 6.9767$ cuando $\beta = \alpha$, y $(\bar{u}, \bar{v}) = (14.0329,\ 6.4475)$ cuando $\beta = \alpha/2$. La traza de la matriz jacobiana en ese punto es

$$\mathrm{tr} = \Big(-\frac{r}{100}\bar{u}\Big) + \Big(\frac{r}{100}\bar{v}\Big) = \frac{r}{100}(\bar{v} - \bar{u}),$$

que **se anula exactamente cuando** $\bar{u} = \bar{v}$, y eso ocurre cuando $\beta = \alpha$, porque entonces las dos rectas son simétricas respecto de la diagonal. Con $\beta = \alpha$ los autovalores son imaginarios puros ($\pm 0.0278\, i$ con estos datos) y la linealización no decide; con $\beta = \alpha/2$ la traza es negativa y el equilibrio es un foco atractivo: las órbitas suficientemente próximas convergen a él en espiral. Esta es una conclusión local; la evolución desde $(10, 30)$ se comprueba a continuación.

LA IDEA: POR QUÉ CON $\beta = \alpha$ LA ÓRBITA SE CIERRA DE VERDAD, Y NO POR CASUALIDAD

Que los autovalores sean imaginarios puros no basta para concluir que las órbitas son cerradas: un foco muy débil puede tener esa misma linealización. Aquí una **simetría** permite obtener el periodo a partir de dos cruces. Con $\beta = \alpha$ el sistema es

$$u' = u\Big(r - \frac{r}{100}u - \alpha v\Big), \qquad v' = -v\Big(r - \frac{r}{100}v - \alpha u\Big),$$

y las dos ecuaciones son la misma salvo por el signo y por el intercambio $u \leftrightarrow v$. Dicho con precisión: si $(u(t), v(t))$ es solución, también lo es $\big(v(-t), u(-t)\big)$. El sistema es *reversible* respecto de la reflexión en la diagonal $u = v$.

Supóngase que una órbita no constante cruza la diagonal en el instante t_1, es decir $u(t_1) = v(t_1)$. La solución reflejada $\big(v(2t_1 - t), u(2t_1 - t)\big)$ toma en $t = t_1$ el mismo valor que la original, luego por unicidad *es* la original: la órbita es simétrica respecto de la diagonal. Si además la cruza una segunda vez, en t_2, el mismo argumento en t_2 da otra simetría, y componer dos reflexiones da una traslación: $2(t_2 - t_1)$ es un periodo. Si son cruces consecutivos, es el periodo mínimo. La reversibilidad por sí sola no garantiza que una trayectoria llegue a cruzar dos veces.

Esto es lo que enlaza el apartado (c) con el (d), que parecían independientes: **el instante t^* en que $u = v$ es un punto de simetría de la órbita, y el periodo es el doble de la distancia entre dos cruces consecutivos**. Con los datos del enunciado, los cruces son $t^* \approx 157{,}053612$ y $t_2 \approx 323{,}489583$, de donde $P \approx 332{,}871942$. Para comprobar la

vuelta a $(10, 30)$ se usa el periodo conservado en memoria, no el redondeado de la pantalla: sustituirlo por 332,87 ya produce una distancia de aproximadamente 0,0021. Para estos datos también puede justificarse que la órbita es cerrada sin depender del dibujo. La función

$$H(u, v) = (uv)^{3/37}(100 - u - v)$$

es constante a lo largo de las soluciones cuando $r = 0{,}03$ y $\alpha = \beta = 0{,}004$: basta derivarla y sustituir las dos ecuaciones. En el triángulo $u, v > 0$, $u + v < 100$, es positiva, se anula en el borde y tiene un único punto crítico, el equilibrio de coexistencia, donde alcanza su máximo. El nivel que pasa por $(10, 30)$ queda por debajo de ese máximo y es compacto y regular. Sus componentes son curvas cerradas sin equilibrios; la trayectoria recorre una de ellas periódicamente.

```
function Especies(r, alpha, beta, u0, v0, T)
% Nombre Apellido Apellido
% Especies(r, alpha, beta, u0, v0, T)
% u' = r*u*(1 - u/100) - alpha*u*v
% v' = -r*v*(1 - v/100) + beta *u*v
% Ejecuciones pedidas:
% Especies(0.03, 0.004, 0.004, 10, 30, 1000) % beta = alpha
% Especies(0.03, 0.004, 0.002, 10, 30, 1000) % beta = alpha/2
% (alpha y beta son nombres de funciones de MATLAB; usarlos como argumentos
% los oculta dentro de esta funcion, lo cual aqui no estorba.)

close all
if any(~isfinite([r alpha beta u0 v0 T])) || ...
   any([r alpha beta T] <= 0) || any([u0 v0] < 0)
   error('Se requieren tasas y T positivos y poblaciones no negativas')
end

% ---------- (a) Resolucion numerica ----------------------------------
F = @(t, y) [ r*y(1)*(1 - y(1)/100) - alpha*y(1)*y(2);
              -r*y(2)*(1 - y(2)/100) + beta *y(1)*y(2) ];
sol = ode45(F, [0 T], [u0; v0], odeset('RelTol', 1e-9, 'AbsTol', 1e-11));
if sol.x(end) < T || any(~isfinite(sol.y(:)))
   error('Integracion incompleta o no finita: no se representa todo [0,T]')
end
estacionario = all(F(0, [u0; v0]) == 0);
t = linspace(0, T, 4000);
Y = deval(sol, t); u = Y(1,:); v = Y(2,:);

% ---------- (b) Las dos trayectorias ---------------------------------
figure(1)
hu = plot(t, u, 'b-', 'LineWidth', 1.5); hold on
hv = plot(t, v, 'r-', 'LineWidth', 1.5);
xlabel('t'), ylabel('numero de individuos'), grid on
title(['Dos especies: r = ', num2str(r), ', \alpha = ', num2str(alpha), ...
       ', \beta = ', num2str(beta), ', (u_0,v_0) = (', num2str(u0), ',', num2str(v0), ')'])

% ---------- (c) Primer instante con u = v ----------------------------
dif = @(s) deval(sol, s, 1) - deval(sol, s, 2);
d = u - v;
```

```
cambio = find(d(1:end-1).*d(2:end) < 0);
raices = t(d == 0); % incluye t=0 si u0=v0
for j = cambio
   raices(end+1) = fzero(dif, [t(j), t(j+1)]);
end
raices = unique(sort(raices));
if isempty(raices)
   disp('No se detecta un cruce en [0,T]; revise T y la malla')
   hE = [];
else
   tigual = raices(1);
   vigual = deval(sol, tigual, 1);
   fprintf('Primer instante con u = v: t* = %10.4f, u(t*) = v(t*) = %8.4f \n', ...
          tigual, vigual)
   hE = plot(tigual, vigual, 'k.', 'MarkerSize', 24);
   text(tigual, vigual, [' t^* = ', num2str(tigual)])
end
nombres = {'u(t): presa', 'v(t): depredador'};
if ~isempty(hE)
   nombres{end+1} = 'primer t con u = v';
end
legend([hu hv hE], nombres, 'Location', 'northeast')

% ---------- (d) La orbita, y el periodo si beta = alpha --------------
figure(2)
plot(u, v, 'k-', 'LineWidth', 1.4), hold on
plot(u0, v0, 'go', 'MarkerSize', 9, 'MarkerFaceColor', 'g')
text(u0, v0, ' dato inicial')
M = max([u, v]);
plot([0 M], [0 M], 'b:') % la diagonal u = v
xlabel('u (presa)'), ylabel('v (depredador)'), grid on, axis equal
title('Orbita en el plano de fases')
legend('orbita', 'dato inicial', 'diagonal u = v', 'Location', 'northeast')

if estacionario
   disp('Solucion constante: no tiene un periodo minimo positivo')
elseif beta == alpha
   if numel(raices) >= 2
      t2 = raices(2);
      P = 2*(t2 - tigual); % la orbita es simetrica respecto de la diagonal
      fprintf('Periodo aproximado de la orbita: P = %10.4f \n', P)
      if P <= sol.x(end)
         fprintf('Comprobacion ||y(P) - y(0)|| = %10.3e \n', ...
                norm(deval(sol, P) - [u0; v0]))
      else
         disp('Para comprobar la vuelta a y(0), amplie T hasta superar P')
      end
      text(u0, v0, [' periodo P = ', num2str(P)])
   else
      disp('No hay dos cruces de la diagonal en [0,T]: hay que alargar T')
   end
else
   disp('beta ~= alpha: no se aplica la formula de simetria para el periodo.')
end
shg
end
```

Valores de referencia. En la primera ejecución se obtiene $t^* \approx 157{,}0536$, $u(t^*) = v(t^*) \approx 0{,}8220$ y $P \approx 332{,}8719$. En la segunda, $t^* \approx 129{,}7848$ y $u(t^*) = v(t^*) \approx 1{,}3684$. El error de retorno debe medirse en la ejecución; depende de la integración y de la localización numérica de los cruces.

Lo que se ve en las dos ejecuciones. Con $\beta = \alpha$ las dos poblaciones oscilan indefinidamente entre 0.20 y 31.84, con el mismo mínimo y el mismo máximo las dos —consecuencia directa de la simetría respecto de la diagonal—, y la órbita es una curva cerrada. Con $\beta = \alpha/2$ el depredador aprovecha peor a la presa, la simetría se rompe y la órbita **espirala hacia dentro** en dirección al equilibrio $(14.03,\ 6.45)$: el sistema se amortigua. En $T = 1000$ todavía no ha llegado, pero la tendencia se ve en el dibujo —la distancia al equilibrio baja de 23,90 a 5,77—. El signo de la traza caracteriza la estabilidad local del equilibrio, pero una aproximación observada en un intervalo finito no prueba por sí sola la convergencia.

En la segunda ejecución, el máximo del depredador en $[0, 1000]$ es precisamente el dato inicial $v_0 = 30$: un máximo en un intervalo cerrado puede estar en un extremo. La tabla siguiente recoge otra cosa, los **máximos locales interiores** de cada oscilación, obtenidos localizando los ceros de las derivadas:

Máximos interiores	primero	segundo	tercero
Presa u	50.18 ($t = 294.7$)	38.29 ($t = 577.3$)	30.96 ($t = 836.9$)
Depredador v	21.68 ($t = 336.6$)	16.86 ($t = 622.5$)	13.80 ($t = 885.3$)

El máximo global de la presa es aproximadamente 50,18, y el primer máximo interior del depredador es 21,68. Comparar picos sucesivos ayuda a ver la amortiguación; describir el rango completo exige incluir también los extremos del intervalo.

Dos comentarios sobre el programa.

- **La igualdad de parámetros es una hipótesis matemática.** La llamada usa exactamente el mismo valor para `alpha` y `beta`. Sustituir la igualdad por `abs(beta-alpha)<1e-12` no haría reversible un sistema con parámetros distintos, aunque sus órbitas pudieran parecer casi cerradas. La búsqueda por cambios de signo también depende de la malla: conviene refinarla y comprobar que los primeros cruces y el periodo se estabilizan. La ausencia de cambios detectados no es una prueba de que no exista un cruce.
- **Otra forma de medir el periodo**, para quien no vea la simetría: buscar el primer $t > 0$ en que la solución vuelve cerca del dato inicial, minimizando $\|y(t) - y(0)\|$ con `fminbnd` en una ventana que se localiza primero a ojo sobre la gráfica. El resultado depende de la ventana y de las tolerancias. También se pueden comparar máximos consecutivos de u; en una oscilación amortiguada sus distancias no definen un periodo exacto.

Qué buscaba este examen

- **Ejercicio 1.** La adimensionalización más sencilla del curso, y después la lectura completa de una EDO autónoma escalar por el signo de su segundo miembro. El punto fino de (b) es ver que el signo del producto de las raíces basta para decidir que hay un único equilibrio no negativo, sin discutir casos según a y b.
- **Ejercicio 2.** El guion entero del SIR, de memoria y en orden: Picard, positividad, monotonías, compacto, límites y umbral. El examen de 2026 permite comparar este razonamiento con una variante del modelo.
- **Ejercicio 3.** Integrar un sistema, dibujar trayectorias y órbita, y contestar con `fzero` a dos preguntas sobre la solución numérica, una de ellas —el periodo— que exige darse cuenta de por qué la órbita se cierra justamente cuando $\beta = \alpha$.

Errores típicos

- **En 1(a), olvidar que hay dos unidades libres** y buscar solo el cambio en u. Sin reescalar el tiempo no se pueden matar las dos constantes.
- **En 1(b), decir que $u = 0$ es un equilibrio** por costumbre de la logística. No lo es: $f(0) = a > 0$.
- **En 1(b), discutir casos según el signo de $1 - b$.** No hace falta: el producto de las raíces es $-a < 0$ pase lo que pase.
- **En 2, afirmar que I es decreciente** porque «la epidemia acaba». Solo lo es si $\mathcal{R}_{\text{ef}}(0) \leq 1$; en el otro caso sube primero.
- **En 2, deducir $I \to 0$ de que I está acotada.** Estar acotada no da límite. El argumento correcto pasa por la integrabilidad de I, que viene de que R sea creciente y acotada.
- **En 2, olvidar el caso $S_0 = 0$**, que el enunciado permite expresamente al escribir $I_0 + R_0 \leq N$ con igualdad posible.
- **En 3, calcular el periodo como $t_2 - t_1$** en vez de como el doble. Entre dos cruces consecutivos de la diagonal transcurre *medio* periodo.
- **En 3, aplicar la fórmula de simetría cuando $\beta \neq \alpha$.** Falta la hipótesis que la justifica; dos cruces no bastan entonces para obtener un periodo.

C.6 Segunda convocatoria del 13 de julio de 2023

Los tres exámenes que siguen —esta convocatoria y los dos primeros parciales— proceden de un mismo cuaderno, en el que el enunciado impreso está pegado arriba y la resolución a mano debajo. El cuaderno conserva la parte de teoría de cada prueba, que suma cinco puntos; no contiene sus ejercicios de prácticas.

Ficha

Asignatura	Modelización matemática, curso 2022/23
Convocatoria	Segunda (julio)
Fecha	13 de julio de 2023
Lo que se conserva	Los dos ejercicios de teoría, 2,5 puntos cada uno
Lo que falta	La parte de prácticas

Enunciado

Ejercicio 1 [Teoría: 2,5 puntos]. Se considera la siguiente variación del modelo logístico para describir el tamaño $y = y(t)$ de una población en cada instante $t \geq 0$

$$\begin{cases} y' = a\big(1 - e^{-by}\big) - cy^2, \\ y(0) = y_0, \end{cases} \tag{C.17}$$

donde a, b, c son tres parámetros positivos, mientras que y_0 es un valor no negativo que representa el tamaño inicial de la población. Se pide:

a) Realizar una adimensionalización que permita reducir los parámetros del sistema de manera que este adquiera la forma

$$\begin{cases} y' = 1 - e^{-y} - \lambda y^2, \\ y(0) = y_0, \end{cases}$$

para una constante λ positiva.

b) Probar que el sistema presenta únicamente dos puntos de equilibrio (uno de ellos el trivial $y = 0$).

c) Estudiar la estabilidad de los puntos de equilibrio y describir el comportamiento asintótico del modelo en función del dato inicial y_0.

Ejercicio 2 [Teoría: 2,5 puntos]. Se considera la siguiente variación del modelo de Volterra

$$(P) \quad \begin{cases} u' = u\big(\log(1+v) - 1\big), \\ v' = \mu v(1-u), \\ u(0) = u_0, \quad v(0) = v_0, \end{cases} \tag{C.18}$$

donde $u = u(t)$ representa el tamaño de la población de depredadores y $v = v(t)$ el tamaño de la población de presas en el instante $t \geq 0$, siendo u_0, v_0 los tamaños iniciales de ambas poblaciones, y μ un parámetro positivo. Se pide:

a) Calcular los puntos de equilibrio de (P). Aplicar el método de la primera aproximación y concluir la estabilidad de los equilibrios cuando sea posible.

b) Probar que si u_0 y v_0 son ambos no negativos, entonces se tiene $u(t) \geq 0$ y $v(t) \geq 0$ para todo $t \geq 0$.

c) Dibujar (esbozar) los campos de vectores tangentes a las órbitas de (P).

d) Demostrar que si denotamos por $F = F(v)$ una primitiva de

$$\frac{\log(1+v) - 1}{v}, \qquad v \in (0, +\infty),$$

entonces para cada $u_0 > 0$, $v_0 > 0$, existe una constante C de manera que la solución de (P) satisface

$$\mu\big(u(t) - \log(u(t)) - 1\big) + F(v(t)) = C, \qquad \forall t \in I(u_0, v_0),$$

donde $I(u_0, v_0)$ representa el intervalo maximal de definición.

Resolución del ejercicio 1

Qué modelo es. Es la logística con el término de nacimientos *saturado*. En la logística clásica $y' = \lambda y - \mu y^2$ el primer sumando crece sin techo con la población; aquí es $a(1 - e^{-by})$, que para y pequeño vale aproximadamente aby —igual que antes— pero se estanca en a cuando y crece. La lectura biológica es que hay un límite físico a cuántos individuos nuevos aparecen por unidad de tiempo, por mucha población que haya. El segundo sumando, $-cy^2$, es la mortalidad por hacinamiento de siempre.

Apartado (a): la adimensionalización

Se busca $\hat{y}(t) = p\, y(st)$ con $p, s > 0$. Derivando,

$$\hat{y}'(t) = ps\, y'(st) = ps\Big[a\big(1 - e^{-b\, y(st)}\big) - c\, y(st)^2\Big] = psa\Big(1 - e^{-\frac{b}{p}\hat{y}}\Big) - \frac{sc}{p}\,\hat{y}^2,$$

donde se ha usado $y(st) = \hat{y}(t)/p$. Comparando con $\hat{y}' = 1 - e^{-\hat{y}} - \lambda\hat{y}^2$:

$$\frac{b}{p} = 1 \implies \boxed{p = b}, \qquad psa = 1 \implies \boxed{s = \frac{1}{ab}}, \qquad \lambda = \frac{sc}{p} = \boxed{\frac{c}{ab^2}}.$$

Los tres son positivos, luego el cambio es admisible y $\lambda > 0$. Tres constantes, dos unidades libres, un parámetro superviviente: el recuento de siempre.

Apartado (b): hay exactamente dos equilibrios

La idea, y es toda la gracia del apartado. Con

$$f(y) = 1 - e^{-y} - \lambda y^2$$

no hace falta despejar los ceros de la ecuación trascendente $f(y) = 0$. Para contar sus soluciones basta una observación: f **es estrictamente cóncava**. Una función estrictamente cóncava corta al eje a lo sumo dos veces, y aquí se exhiben las dos.

Paso 1: $y = 0$ **es equilibrio.** $f(0) = 1 - 1 - 0 = 0$.

Paso 2 (el núcleo): f **es estrictamente cóncava en todo** $\mathbb{R}$**.**

$$f'(y) = e^{-y} - 2\lambda y, \qquad f''(y) = -e^{-y} - 2\lambda < 0 \quad \text{para todo } y \in \mathbb{R}.$$

La segunda derivada es negativa *siempre*, sin condiciones. Y una función con $f'' < 0$ tiene f' estrictamente decreciente, luego f no puede tener tres ceros: entre dos ceros consecutivos habría un punto con derivada nula por Rolle, y con f' estrictamente decreciente eso solo puede pasar una vez.

Paso 3: no hay ceros negativos. Como f' es estrictamente decreciente y $f'(0) = 1 > 0$, se tiene $f'(y) > 1 > 0$ para todo $y < 0$; luego f es estrictamente creciente en $(-\infty, 0]$ y por tanto $f(y) < f(0) = 0$ para $y < 0$. Ningún cero negativo.

Paso 4: hay exactamente un cero positivo. Como $f'(0) = 1 > 0$, f es estrictamente creciente a la derecha de 0, luego $f > 0$ en un intervalo $(0, \delta)$. Y

$$f(y) = 1 - e^{-y} - \lambda y^2 \le 1 - \lambda y^2 \xrightarrow[y \to +\infty]{} -\infty,$$

así que f acaba siendo negativa. Por Bolzano hay un cero $\bar{y} > 0$, y por el paso 2 no puede haber otro. En total, **exactamente dos equilibrios**: 0 y $\bar{y} > 0$. Además $\bar{y} < 1/\sqrt{\lambda}$, porque $1 - e^{-\bar{y}} < 1$ obliga a $\lambda\bar{y}^2 < 1$.

Apartado (c): estabilidad y comportamiento asintótico

Paso 1: el origen es inestable. $f'(0) = 1 > 0$, luego $y = 0$ es un equilibrio inestable (teorema de linealización en dimensión uno; el criterio del signo de f' del capítulo 2).

Paso 2: $\bar{y}$ es exponencialmente asintóticamente estable. Por los pasos anteriores, $f > 0$ en $(0, \bar{y})$ y $f < 0$ en $(\bar{y}, +\infty)$: la función cruza el eje *bajando*, luego $f'(\bar{y}) \leq 0$. Y no puede ser cero: si $f'(\bar{y}) = 0$, entonces $\bar{y}$ sería un máximo local de f (por concavidad), pero $f(\bar{y}) = 0$ y $f > 0$ inmediatamente a la izquierda, lo que contradice que sea máximo. Luego $f'(\bar{y}) < 0$ y el equilibrio es exponencialmente asintóticamente estable.

Paso 3: el comportamiento, por el teorema escalar autónomo (teorema de las EDO autónomas escalares):

- $y_0 = 0$: solución constante 0.
- $0 < y_0 < \bar{y}$: $f(y_0) > 0$, solución estrictamente creciente, con el punto crítico $\bar{y}$ por encima; definida en $[0, +\infty)$ y $y(t) \nearrow \bar{y}$.
- $y_0 = \bar{y}$: constante.
- $y_0 > \bar{y}$: $f(y_0) < 0$, estrictamente decreciente, con $\bar{y}$ por debajo; definida en $[0, +\infty)$ y $y(t) \searrow \bar{y}$.

En resumen: **toda población inicial estrictamente positiva converge, de forma monótona, al mismo tamaño** $\bar{y}$, y solo la población exactamente nula se queda en cero.

LA IDEA: POR QUÉ SATURAR LOS NACIMIENTOS NO CAMBIA NADA CUALITATIVAMENTE

El retrato de fases es idéntico al de la logística: un equilibrio inestable en el cero, uno estable positivo, y convergencia monótona hacia él desde cualquier dato positivo. Cambiar λy por $a(1 - e^{-by})$ ha cambiado *el número* $\bar{y}$, pero no este retrato de fases. Ambos modelos coinciden en las monotonías y el destino de las soluciones positivas. Sus velocidades y tamaños de equilibrio sí difieren; elegir entre ellos para describir una población requiere datos.

Lo que sí cambiaría la historia es tocar el comportamiento *cerca de cero*, y merece la pena decir con precisión qué haría falta, porque la respuesta rápida —«que los nacimientos sean $O(y^2)$»— no basta. Si se sustituye $1 - e^{-y}$ por algo que sea $O(y^2)$, por ejemplo $f(y) = (1 - e^{-y})^2 - \lambda y^2$, los dos sumandos son cuadráticos cerca de cero y el signo lo decide el cociente de sus coeficientes: con $\lambda < 1$ el origen sigue repeliendo y sigue habiendo un único equilibrio positivo estable —la misma historia de antes, con otro número—, y con $\lambda \geq 1$ es $f < 0$ en todo $(0, +\infty)$ y hay extinción desde cualquier dato positivo, **sin umbral**. No aparece aquí un efecto Allee fuerte, que separaría extinción y persistencia mediante un umbral positivo. Con $\lambda < 1$ sí puede hablarse de efecto Allee débil: cerca de cero, el crecimiento por individuo $f(y)/y \sim (1 - \lambda)y$ aumenta con la población, aunque sigue siendo positivo.

En el modelo clásico con umbral de la sección 2.1.2, $y' = ry(y/a - 1)(1 - y/K)$ con $0 < a < K$, el crecimiento es negativo por debajo de a y positivo entre a y K. La condición $f'(0) < 0$ explica el decrecimiento cerca de cero, pero no basta por sí sola para producir los otros dos equilibrios. Tampoco una mortalidad cuadrática excluye cualquier umbral. Por ejemplo,

$$f(y) = \frac{3y^3}{1 + y^2} - y^2 = y^2 \left(\frac{3y}{1 + y^2} - 1 \right)$$

tiene nacimientos no negativos y mortalidad y^2, pero decrece para $0 < y < (3-\sqrt{5})/2$, crece hasta $(3+\sqrt{5})/2$ y vuelve a decrecer después. Aquí $f'(0) = 0$ y hay, aun así, un umbral de extinción. Lo que decide el retrato es el signo del crecimiento neto, no solo el orden de un sumando.

Resolución del ejercicio 2

Dónde vive el sistema. Antes de nada: el campo tiene un $\log(1+v)$, así que solo está definido en

$$\Omega = \{(u,v) \in \mathbb{R}^2 : v > -1\},$$

un semiplano abierto, y ahí es de clase C^∞. Picard da una única solución maximal por cada dato de Ω. Conviene decirlo, porque el enunciado no lo dice y sin ello ni siquiera está claro qué se está resolviendo.

Qué modelo es. Es el de Volterra de la sección 2.2.4 con la respuesta del depredador cambiada. En aquel, $u' = u(-c+dv)$: el beneficio del depredador crece proporcionalmente al número de presas. Aquí es $\log(1+v)-1$, que crece pero cada vez más despacio: duplicar las presas no duplica el beneficio por depredador. La respuesta logarítmica no tiene un techo finito; es sublineal, no saturada. La ecuación de las presas, en cambio, es la de siempre.

Apartado (a): equilibrios y linealización

Paso 1: los equilibrios. De $u(\log(1+v)-1) = 0$ salen $u = 0$ o $\log(1+v) = 1$, es decir $v = e-1$. De $\mu v(1-u) = 0$ salen $v = 0$ o $u = 1$. Cruzando, y descartando los cruces imposibles:

$$(0,0) \qquad \text{y} \qquad (1, e-1).$$

Paso 2: la matriz jacobiana. Con $F(u,v) = \big(u(\log(1+v)-1),\ \mu v - \mu u v\big)$,

$$DF(u,v) = \begin{pmatrix} \log(1+v)-1 & \dfrac{u}{1+v} \\ -\mu v & \mu(1-u) \end{pmatrix}.$$

Paso 3: el origen es un punto de silla.

$$DF(0,0) = \begin{pmatrix} -1 & 0 \\ 0 & \mu \end{pmatrix},$$

con autovalores $-1 < 0$ y $\mu > 0$: hay uno con parte real positiva, luego el origen es **inestable** (y es una silla). Tiene sentido: sin depredadores las presas se disparan.

Paso 4 (el núcleo): en $(1, e-1)$ la linealización no decide.

$$DF(1,e-1) = \begin{pmatrix} \log e - 1 & \dfrac{1}{e} \\ -\mu(e-1) & 0 \end{pmatrix} = \begin{pmatrix} 0 & \dfrac{1}{e} \\ -\mu(e-1) & 0 \end{pmatrix},$$

cuyo polinomio característico es $\lambda^2 + \dfrac{\mu(e-1)}{e} = 0$, con raíces

$$\lambda = \pm i\sqrt{\frac{\mu(e-1)}{e}},$$

imaginarias puras. El teorema de la primera aproximación exige que todos los autovalores tengan parte real estrictamente negativa (para concluir estabilidad) o que alguno la tenga estrictamente positiva (para concluir inestabilidad); con parte real cero *no dice nada*, y no por falta de refinamiento: es que el resultado depende de los términos no lineales, que la linealización ha tirado. Hay que responder exactamente eso, y esperar al apartado (d), que es el que lo resuelve.

Apartado (b): positividad

Igual que en los otros exámenes, por invariancia de los ejes.

- *El eje* $u = 0$. La función $\big(0,\ v_0 e^{\mu t}\big)$ resuelve el sistema con dato $(0, v_0)$: la primera componente da $0' = 0$ y la segunda, $v' = \mu v(1-0) = \mu v$. Por unicidad es *la* solución, luego el eje $u = 0$ está formado por órbitas.
- *El eje* $v = 0$. La función $\big(u_0 e^{-t},\ 0\big)$ resuelve el sistema con dato $(u_0, 0)$: $u' = u(\log 1 - 1) = -u$ y $v' = 0$. Otra vez, órbitas.

Si $u_0, v_0 > 0$, la órbita correspondiente no puede cortar ninguna de esas (unicidad de las órbitas), luego $u(t) > 0$ y $v(t) > 0$ mientras la solución exista. Si $u_0 = 0$ o $v_0 = 0$, la solución es una de las órbitas de arriba y también es no negativa. En los dos casos, ambas componentes son no negativas mientras exista la solución.

Falta justificar que existe para todo $t \geq 0$. De $v' \leq \mu v$ se deduce $v(t) \leq v_0 e^{\mu t}$. En cualquier intervalo $0 \leq t \leq T$,

$$u' \leq u \log(1 + v_0 e^{\mu T}), \qquad u(t) \leq u_0 \exp\big(T \log(1 + v_0 e^{\mu T})\big).$$

Si el extremo maximal derecho fuese finito, estas cotas mantendrían la solución en un compacto del semiplano $v > -1$, y el criterio de prolongación permitiría continuarla. Por tanto, la solución es global hacia delante y la no negatividad vale para todo $t \geq 0$, como pide el enunciado.

Apartado (c): el campo de direcciones

En el cuadrante abierto $u > 0$, $v > 0$ las isoclinas nulas son las rectas $v = e - 1$ (donde $u' = 0$) y $u = 1$ (donde $v' = 0$), y los signos son:

$$u' > 0 \iff v > e - 1, \qquad v' > 0 \iff u < 1.$$

Eso parte el cuadrante en cuatro regiones, y las flechas se leen de un vistazo:

Región	u'	v'	Hacia dónde apunta el campo
$u < 1, v < e-1$	$-$	$+$	arriba y a la izquierda
$u < 1, v > e-1$	$+$	$+$	arriba y a la derecha
$u > 1, v > e-1$	$+$	$-$	abajo y a la derecha
$u > 1, v < e-1$	$-$	$-$	abajo y a la izquierda

Recorriendo la tabla se ve que el campo **gira en el sentido de las agujas del reloj alrededor de** $(1, e-1)$: desde abajo a la izquierda se sube, luego se va a la derecha por arriba, luego se baja por la derecha y luego se vuelve a la izquierda por abajo. Es el mismo giro que en el Volterra de la sección 2.2.4. El siguiente esbozo recoge los signos; la longitud de las flechas no representa la rapidez.

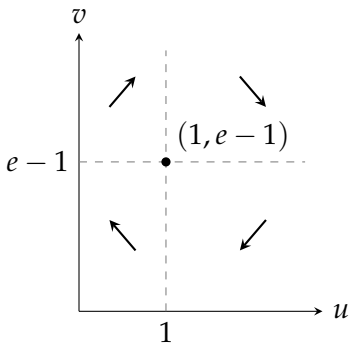


Figura C.1: Sentido del campo en las cuatro regiones de signo. Los depredadores están en el eje horizontal y las presas en el vertical.

Conviene comprobar el sentido en un punto y no fiarse del recorrido mental, que es donde se equivoca todo el mundo: en $(1.5,\, e-1)$, justo a la derecha del centro, es $u' = u(\log e - 1) = 0$ y $v' = \mu(e-1)(1-1.5) < 0$, de modo que ahí el campo apunta **hacia abajo**. A la derecha del centro hacia abajo es giro horario, y basta ese punto para fijar el sentido de toda la figura. La lectura del modelo es la de siempre: con muchos depredadores ($u > 1$) las presas caen, y son las presas las que están en el eje vertical.

Apartado (d): la integral primera, y lo que resuelve

La idea. Se pide comprobar que una cierta función no cambia a lo largo de las soluciones. Comprobarlo es derivar y ver que sale cero; encontrarla, en cambio, es lo que tiene mérito, y se hace separando variables en la órbita:

$$\frac{du}{dv} = \frac{u'}{v'} = \frac{u(\log(1+v)-1)}{\mu v(1-u)} \implies \mu\,\frac{1-u}{u}\,du = \frac{\log(1+v)-1}{v}\,dv,$$

y las dos partes se integran por separado. La de la izquierda da $\mu(\log u - u)$; la de la derecha, $F(v)$. Igualando y cambiando el signo se obtiene la expresión del enunciado.

La comprobación, que es lo que se pide. Sea

$$H(u,v) := \mu\big(u - \log u - 1\big) + F(v),$$

bien definida para $u > 0$, $v > 0$ (que es donde vive la solución por el apartado (b) y la observación de que $u_0, v_0 > 0$). Derivando a lo largo de una solución, con $F'(v) = \dfrac{\log(1+v)-1}{v}$:

$$\frac{d}{dt}H(u(t),v(t)) = \mu\Big(1-\frac{1}{u}\Big)u' + \frac{\log(1+v)-1}{v}\,v'.$$

Sustituyendo $u' = u(\log(1+v)-1)$ y $v' = \mu v(1-u)$:

$$\begin{aligned}\frac{dH}{dt} &= \mu\Big(1-\frac{1}{u}\Big)u\big(\log(1+v)-1\big) + \frac{\log(1+v)-1}{v}\,\mu v(1-u)\\ &= \mu\big(\log(1+v)-1\big)\Big[(u-1)+(1-u)\Big] = 0.\end{aligned}$$

El corchete se anula *idénticamente*: ahí está todo. Luego H es constante a lo largo de cada solución, y llamando $C := H(u_0, v_0)$ se obtiene lo pedido. ■

LA IDEA: LA INTEGRAL PRIMERA CONTESTA LO QUE LA LINEALIZACIÓN NO SUPO

El apartado (a) se quedó sin poder decidir la estabilidad de $(1, e-1)$, y el (d) la decide. Obsérvese dónde están los puntos críticos de H:

$$\partial_u H = \mu\Big(1 - \frac{1}{u}\Big) = 0 \iff u = 1, \qquad \partial_v H = \frac{\log(1+v) - 1}{v} = 0 \iff v = e - 1.$$

¡El único punto crítico de H es justamente el equilibrio! Y es un **mínimo estricto**: $\partial^2_{uu} H = \mu/u^2 > 0$, $\partial_{uv} H = 0$ y $\partial^2_{vv} H = \frac{d}{dv}\frac{\log(1+v)-1}{v}$, que en $v = e-1$ vale $\frac{1}{e(e-1)} > 0$ (el numerador se anula ahí, así que solo sobrevive un término al derivar el cociente). Entonces $V := H - H(1, e-1)$ es una función de Liapunov: es positiva alrededor del equilibrio, se anula solo en él, y $\dot V = 0$. Por el teorema de Liapunov (teorema de Liapunov, libro de EDO, §7.3 «») el equilibrio es **estable**. Y no es asintóticamente estable, porque una solución que empieza en el nivel $V = c > 0$ se queda en ese nivel para siempre y no puede acercarse al equilibrio, donde $V = 0$. Las órbitas cercanas son las curvas de nivel de H alrededor de un mínimo estricto, es decir, **curvas cerradas**: el sistema oscila, exactamente como el Volterra clásico. Cambiar la respuesta del depredador por una respuesta logarítmica no ha roto los ciclos cercanos al equilibrio.

Qué buscaba este examen

- **Ejercicio 1.** Contar equilibrios de una ecuación trascendente *sin resolverla*, usando la concavidad; y después el teorema escalar autónomo.
- **Ejercicio 2.** Las cuatro herramientas del capítulo 2 sobre sistemas planos, en orden: equilibrios y linealización, invariancia de los ejes, campo de direcciones por regiones de signo, e integral primera. Y sobre todo, saber que la linealización *puede no decidir* y qué se hace entonces.

Errores típicos

- **Buscar una fórmula explícita para** $1 - e^{-y} = \lambda y^2$**.** El apartado pide contar los equilibrios; la concavidad basta.
- **Concluir que hay dos equilibrios «porque el dibujo lo sugiere».** El dibujo no es una demostración; la concavidad sí.
- **Decir que** $(1, e-1)$ **es estable porque los autovalores son imaginarios puros.** Eso es exactamente lo que el teorema *no* permite concluir. Un foco muy débil tiene la misma linealización.
- **Olvidar que el campo solo está definido para** $v > -1$**.**
- **En (d), derivar** F **mal.** F es una primitiva de $(\log(1+v) - 1)/v$, así que $F'(v)$ es esa expresión y no hay que volver a derivarla.

C.7 Primer parcial del 29 de marzo de 2023

Ficha

Asignatura	Modelización matemática, curso 2022/23
Prueba	Primer parcial
Fecha	29 de marzo de 2023
Lo que se conserva	Los dos ejercicios de teoría: $2+3=5$ puntos
Lo que falta	La parte de prácticas

Enunciado

Ejercicio 1 [Teoría: 2 puntos]. Sea L la matriz de Leslie

$$L=\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & 0 & 0 \\ 0 & b_2 & 0 \end{pmatrix}, \qquad \text{con } a_1,a_2,a_3\in[0,+\infty),\ a_3\neq 0,\ \text{y } b_1,b_2\in(0,1].$$

Admitiendo que todos los autovalores de L son reales, se pretende dar una demostración simple de que el mayor autovalor λ_{max} es positivo y estrictamente dominante. Para ello se pide:

a) Calcular el polinomio característico de L, $p(\lambda)=\det(\lambda I-L)$.
b) Demostrar que p admite una raíz positiva, y por tanto que $\lambda_{max}>0$.
c) Comprobar que $p(\lambda)+p(-\lambda)<0$ para todo $\lambda\in\mathbb{R}$.
d) A partir del apartado anterior, concluir que λ_{max} es estrictamente dominante.

Ejercicio 2 [Teoría: 3 puntos]. Para describir el comportamiento de un ecosistema de dos poblaciones, una formada por una especie de herbívoros y otra por una familia de plantas, se puede utilizar la siguiente variación del modelo presa-depredador

$$(P)\quad \begin{cases} N' = aN - bP\big(1-e^{-rN}\big), \\ P' = -cP + dP\big(1-e^{-rN}\big), \\ N(0)=N_0, \quad P(0)=P_0, \end{cases} \tag{C.19}$$

donde $N(t)$ representa la abundancia de plantas (presa) y $P(t)$ la de herbívoros (depredador) en el instante t. Aquí N_0 y P_0 son constantes no negativas que definen los valores iniciales de N y P respectivamente, mientras que a, b, c, d y r son constantes positivas. Esta variación del modelo presa-depredador está justificada porque para N próximo a cero se tiene que $1-e^{-rN}$ es próximo a rN. Para estudiar el modelo (P), se pide:

a) Realizar una adimensionalización que permita reducir los parámetros del sistema de manera que este adquiera la forma

$$(P_a)\quad \begin{cases} N' = \mu N - P\big(1-e^{-N}\big), \\ P' = -\nu P + P\big(1-e^{-N}\big), \\ N(0)=N_0, \quad P(0)=P_0, \end{cases} \tag{C.20}$$

con μ y ν dos constantes positivas. Se denotará por $I(N_0,P_0)$ el intervalo maximal de definición de la solución de (P_a).

b) Calcular todos los puntos de equilibrio para (P_a), y determinar condiciones sobre los parámetros μ y ν para que exista un equilibrio (N^*, P^*) con sentido biológico no trivial (de coexistencia). En lo que sigue se supondrá la existencia de tal equilibrio.
c) Estudiar la estabilidad del punto de equilibrio (N^*, P^*) con ayuda del método de la primera aproximación, y comparar el resultado obtenido con el comportamiento del punto de equilibrio no trivial del sistema de presa-depredador de Volterra.
d) Demostrar que $P(t) \geq 0$ y $N(t) \geq 0$ para todo t en $I(N_0, P_0)$.
e) Utilizando el apartado anterior, probar que $P(t) \leq P_0 e^t$ y $N(t) \leq N_0 e^{\mu t}$ para todo t en $I(N_0, P_0) \cap [0, +\infty)$, y concluir que $[0, +\infty) \subset I(N_0, P_0)$.

Resolución del ejercicio 1

De qué va el ejercicio. La sección 1.1.3.4 estudia el autovalor positivo de una matriz de Leslie. La condición $a_n > 0$ garantiza irreducibilidad; para deducir dominancia estricta mediante Perron–Frobenius hace falta, además, aperiodicidad. Aquí se obtiene la dominancia para $n = 3$ a partir del polinomio característico y de la hipótesis adicional de que todos los autovalores son reales.

Paso 1, apartado (a): el polinomio característico.

$$\lambda I - L = \begin{pmatrix} \lambda - a_1 & -a_2 & -a_3 \\ -b_1 & \lambda & 0 \\ 0 & -b_2 & \lambda \end{pmatrix}.$$

Desarrollando por la primera fila,

$$\begin{aligned} p(\lambda) &= (\lambda - a_1) \begin{vmatrix} \lambda & 0 \\ -b_2 & \lambda \end{vmatrix} + a_2 \begin{vmatrix} -b_1 & 0 \\ 0 & \lambda \end{vmatrix} - a_3 \begin{vmatrix} -b_1 & \lambda \\ 0 & -b_2 \end{vmatrix} \\ &= (\lambda - a_1)\lambda^2 - a_2 b_1 \lambda - a_3 b_1 b_2, \end{aligned}$$

es decir

$$\boxed{p(\lambda) = \lambda^3 - a_1\lambda^2 - a_2 b_1 \lambda - a_3 b_1 b_2.}$$

Los tres coeficientes que acompañan a las potencias menores son **menores o iguales que cero**, y el último es estrictamente negativo. Eso, y nada más, es lo que se va a usar.

Paso 2, apartado (b): hay una raíz positiva. Evaluando en cero,

$$p(0) = -a_3 b_1 b_2 < 0,$$

porque $a_3 \neq 0$ y $a_3 \geq 0$ obligan a $a_3 > 0$, y $b_1, b_2 > 0$. Y como p es un polinomio mónico de grado impar, $p(\lambda) \to +\infty$ cuando $\lambda \to +\infty$. Por el teorema de Bolzano existe $\lambda^\sharp > 0$ con $p(\lambda^\sharp) = 0$. Como λ_{max} es el mayor autovalor y $\lambda^\sharp$ es uno de ellos, $\lambda_{max} \geq \lambda^\sharp > 0$.

Aquí es donde se usa $a_3 \neq 0$, y es la hipótesis clave: si a_3 fuese cero, $p(0) = 0$ y el cero sería autovalor. En el lenguaje del capítulo 1, $a_n > 0$ es exactamente la condición que hace irreducible a la matriz de Leslie.

Paso 3, apartado (c): la identidad que lo decide todo. Se calcula directamente:

$$p(\lambda) = \lambda^3 - a_1\lambda^2 - a_2b_1\lambda - a_3b_1b_2, \qquad p(-\lambda) = -\lambda^3 - a_1\lambda^2 + a_2b_1\lambda - a_3b_1b_2.$$

Los términos impares se cancelan al sumar y los pares se duplican:

$$\boxed{p(\lambda) + p(-\lambda) = -2a_1\lambda^2 - 2a_3b_1b_2 \;\leq\; -2a_3b_1b_2 < 0}$$

para todo $\lambda \in \mathbb{R}$, porque $a_1 \geq 0$ y $a_3b_1b_2 > 0$. Obsérvese que el resultado *no* depende de a_2: ese término se ha ido con los impares.

Paso 4 (el núcleo), apartado (d): λ_{max} es estrictamente dominante. Hay que probar que $|\mu| < \lambda_{max}$ para todo autovalor $\mu \neq \lambda_{max}$. Como por hipótesis todos son reales, se separan dos casos.

Autovalores no negativos. Si $\mu \geq 0$ y $\mu \neq \lambda_{max}$, entonces $|\mu| = \mu < \lambda_{max}$ por ser λ_{max} el mayor. Nada que hacer.

Autovalores negativos. Sea $\mu < 0$ un autovalor, de modo que $|\mu| = -\mu > 0$. Se afirma que $-\mu < \lambda_{max}$. En efecto, supóngase lo contrario, $-\mu \geq \lambda_{max}$. Como λ_{max} es la mayor raíz real de p y p es mónico de grado tres con todas sus raíces reales, se tiene $p(\lambda) \geq 0$ para $\lambda \geq \lambda_{max}$ (el polinomio ya no vuelve a bajar del eje). Luego $p(-\mu) \geq 0$. Pero, aplicando (c) con $\lambda = \mu$ y usando que $p(\mu) = 0$,

$$p(-\mu) = p(\mu) + p(-\mu) < 0,$$

contradicción. Por tanto $-\mu < \lambda_{max}$, es decir $|\mu| < \lambda_{max}$.

Además, λ_{max} es una raíz simple. Se ve con las fórmulas de Vieta. Llamando $\lambda_1, \lambda_2, \lambda_3$ a las tres raíces reales,

$$\lambda_1\lambda_2\lambda_3 = a_3b_1b_2 > 0, \qquad \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 = -a_2b_1 \leq 0.$$

Si las tres fuesen positivas, la segunda suma sería estrictamente positiva, contra la segunda igualdad. Y ninguna puede ser cero, porque el producto no lo es. Como el producto es positivo y no todas son positivas, la única posibilidad es **una positiva y dos negativas**. Luego λ_{max} es la única raíz positiva, y en particular simple. ■

LA IDEA: QUÉ HACE CADA HIPÓTESIS, Y POR QUÉ EL TRUCO DE $p(\lambda) + p(-\lambda)$ FUNCIONA

La suma $p(\lambda) + p(-\lambda)$ se queda solo con la **parte par** del polinomio, que aquí es $-2a_1\lambda^2 - 2a_3b_1b_2$: dos términos que ya vienen con signo menos porque en la matriz de Leslie las fecundidades a_i y las supervivencias b_i son *no negativas*, y en $\det(\lambda I - L)$ entran restando. Ese es todo el mecanismo: *la positividad de los datos biológicos se traduce en signos fijos de los coeficientes, y los signos fijos impiden que exista un autovalor negativo tan grande en módulo como el positivo*. Dicho en términos del modelo: los modos oscilantes crecen más despacio, en módulo, que el modo positivo dominante. Su amplitud absoluta podría crecer si su autovalor tiene módulo mayor que uno; lo que tiende a cero es su contribución relativa al dividir por λ_{max}^k. El modo dominante fija la tasa de crecimiento a largo plazo y la distribución estable por edades (sección 1.1.3). La hipótesis $a_1 \geq 0$ se usa en (c); la $a_3 > 0$, en (b) y en la simplicidad; y la hipótesis de que los autovalores son reales es lo que permite prescindir de Perron-Frobenius, porque con autovalores complejos «el mayor» ni siquiera está definido.

Resolución del ejercicio 2

Apartado (a): la adimensionalización

Se busca $\hat{N}(t) = \rho\, N(st)$, $\hat{P}(t) = q\, P(st)$ con $\rho, q, s > 0$. Sustituyendo $N(st) = \hat{N}/\rho$ y $P(st) = \hat{P}/q$ en las dos ecuaciones:

$$\hat{N}' = \rho s\Big[a\frac{\hat{N}}{\rho} - b\frac{\hat{P}}{q}\big(1 - e^{-\frac{r}{\rho}\hat{N}}\big)\Big] = sa\,\hat{N} - \frac{\rho s b}{q}\,\hat{P}\big(1 - e^{-\frac{r}{\rho}\hat{N}}\big),$$

$$\hat{P}' = qs\Big[-c\frac{\hat{P}}{q} + d\frac{\hat{P}}{q}\big(1 - e^{-\frac{r}{\rho}\hat{N}}\big)\Big] = -sc\,\hat{P} + sd\,\hat{P}\big(1 - e^{-\frac{r}{\rho}\hat{N}}\big).$$

Se normalizan tres coeficientes comparando con (P_a):

$$\frac{r}{\rho} = 1 \Rightarrow \boxed{\rho = r}, \qquad sd = 1 \Rightarrow \boxed{s = \frac{1}{d}}, \qquad \frac{\rho s b}{q} = 1 \Rightarrow \boxed{q = \frac{rb}{d}},$$

$$\boxed{\mu = sa = \frac{a}{d} > 0}, \qquad \boxed{\nu = sc = \frac{c}{d} > 0}.$$

Cinco constantes, tres unidades, dos supervivientes: el recuento de siempre. La lectura de los dos que quedan es transparente y va a hacer falta en (b): $\mu = a/d$ compara lo deprisa que crecen las plantas con lo deprisa que los herbívoros convierten comida en herbívoros, y $\nu = c/d$ compara la mortalidad del herbívoro con esa misma capacidad de conversión.

Apartado (b): equilibrios y la condición de coexistencia

Paso 1. De la segunda ecuación, $P\big(-\nu + 1 - e^{-N}\big) = 0$, salen dos casos.

- $P = 0$. Entonces la primera da $\mu N = 0$, luego $N = 0$: el equilibrio trivial $(0, 0)$.
- $1 - e^{-N} = \nu$. Despejando, $e^{-N} = 1 - \nu$, que **solo tiene solución si** $\nu < 1$, y entonces

$$N^* = -\log(1 - \nu) = \log\frac{1}{1 - \nu} > 0.$$

Llevándolo a la primera ecuación, $\mu N^* = P\,(1 - e^{-N^*}) = P\nu$, de donde

$$P^* = \frac{\mu N^*}{\nu} = -\frac{\mu \log(1 - \nu)}{\nu} > 0.$$

Paso 2: la condición. Existe equilibrio de coexistencia (con las dos coordenadas estrictamente positivas) **si y solo si** $\nu < 1$, es decir, si y solo si $c < d$; y entonces es único. Sobre μ no hay ninguna condición: cualquier $\mu > 0$ vale, y solo cambia la altura P^*.

LA IDEA: POR QUÉ LA CONDICIÓN ES EXACTAMENTE $c < d$

La tasa de crecimiento per cápita del depredador es $-\nu + 1 - e^{-N}$, y el segundo sumando está *acotado por uno* por muchas plantas que haya: es la saturación del enunciado. Así que el mejor crecimiento al que puede aspirar un herbívoro es $-\nu + 1$. Si $\nu \geq 1$, ese número es ≤ 0 y la tasa es estrictamente negativa para todo valor finito de N: no puede haber un equilibrio con ambas poblaciones positivas. Esto no es, por sí solo, una descripción de todos los límites posibles de las soluciones. En las constantes originales, $\nu = c/d < 1$ es $c < d$: la mortalidad tiene que ser menor que la capacidad

> máxima de convertir plantas en herbívoros. Esta condición no aparece en el Volterra clásico, y la razón es justamente que allí la respuesta rN no está acotada: con bastantes presas, un depredador de Volterra siempre crece.

Apartado (c): la estabilidad del equilibrio de coexistencia

Paso 1: la matriz jacobiana. Con $F(N,P) = \big(\mu N - P(1-e^{-N}),\ -\nu P + P(1-e^{-N})\big)$,

$$DF(N,P) = \begin{pmatrix} \mu - Pe^{-N} & -(1-e^{-N}) \\ Pe^{-N} & -\nu + 1 - e^{-N} \end{pmatrix}.$$

En (N^*,P^*) se usa $1-e^{-N^*} = \nu$, o sea $e^{-N^*} = 1-\nu$, y la esquina inferior derecha se anula:

$$DF(N^*,P^*) = \begin{pmatrix} \mu - \alpha & -\nu \\ \alpha & 0 \end{pmatrix}, \qquad \alpha := P^*(1-\nu) > 0.$$

Paso 2: traza y determinante.

$$\operatorname{tr} = \mu - \alpha, \qquad \det = 0 - (-\nu)\alpha = \nu\alpha > 0.$$

Con determinante positivo, el signo de la traza decide: si $\operatorname{tr} > 0$ hay al menos un autovalor con parte real positiva y el equilibrio es inestable.

Paso 3 (el núcleo): la traza es siempre estrictamente positiva. Sustituyendo $P^* = \mu N^*/\nu$ y $N^* = -\log(1-\nu)$,

$$\operatorname{tr} = \mu - \frac{\mu N^*}{\nu}(1-\nu) = \mu\Big[1 + \frac{(1-\nu)\log(1-\nu)}{\nu}\Big].$$

Hay que ver que el corchete es positivo para todo $\nu \in (0,1)$. Con el cambio $x = 1-\nu \in (0,1)$, el corchete es $1 + \dfrac{x\log x}{1-x}$, y como $1-x>0$, su signo es el de

$$g(x) := x\log x - x + 1.$$

Ahora bien, $g(1) = 0$ y $g'(x) = \log x + 1 - 1 = \log x < 0$ en $(0,1)$: g es estrictamente decreciente, luego $g(x) > g(1) = 0$ para todo $x \in (0,1)$. Por tanto el corchete es positivo y

$$\boxed{\operatorname{tr} DF(N^*,P^*) > 0 \implies (N^*,P^*) \text{ es inestable.}}$$

Según el signo de $\operatorname{tr}^2 - 4\det$, los autovalores son dos reales positivos (nodo repulsor) o dos complejos conjugados con parte real positiva (foco repulsor); en los dos casos el equilibrio repele, y en el segundo las órbitas se alejan girando.

Paso 4: la comparación con Volterra, que es lo que el apartado pide de verdad. En el modelo de Volterra, escrito ahora con las presas en la primera componente, $u' = u(a - bv)$, $v' = v(-c + du)$, la matriz jacobiana en el equilibrio $(c/d, a/b)$ es

$$\begin{pmatrix} 0 & -bc/d \\ da/b & 0 \end{pmatrix},$$

con **traza cero** y determinante positivo: autovalores imaginarios puros. La linealización no decide, y la integral primera enseña que en realidad es un centro, rodeado de órbitas cerradas: el equilibrio es estable pero no asintóticamente estable, y las poblaciones oscilan indefinidamente con la amplitud que fije el dato inicial.

Aquí, en cambio, la traza es *estrictamente* positiva. La diferencia es enorme y no es una cuestión de grado: sustituir la respuesta lineal rN por la saturada $1 - e^{-rN}$ **desestabiliza localmente** la coexistencia. Si el equilibrio es un foco, las órbitas cercanas se alejan girando; si es un nodo, no hay ese giro local. El cálculo no determina por sí solo el destino de las trayectorias alejadas.

La idea: por qué la respuesta cóncava desestabiliza este equilibrio

Escribamos $h(N) = 1 - e^{-N}$. En el equilibrio se cumple $\mu N^* = P^* h(N^*)$, de modo que la entrada superior izquierda del jacobiano es

$$\mu - P^* h'(N^*) = P^* \left(\frac{h(N^*)}{N^*} - h'(N^*) \right) > 0.$$

La desigualdad se debe a la concavidad estricta de h y a $h(0) = 0$: la pendiente de la tangente es menor que la de la secante desde el origen. Ante un pequeño aumento de plantas, el incremento de consumo no compensa el incremento de crecimiento. Esto ocurre incluso si $h(N^*)$ está lejos de su techo. En Volterra, la respuesta es lineal y esas dos pendientes coinciden, por lo que la entrada diagonal es cero. Eso no deja inmóvil al sistema: la otra población también responde, y ambos efectos producen los ciclos del modelo clásico.

Apartado (d): las dos poblaciones no se hacen negativas

Paso 1: las dos ecuaciones son lineales disfrazadas. Con P es inmediato:

$$P' = P\big(-\nu + 1 - e^{-N}\big) \implies P(t) = P_0 \exp\Big(\int_0^t \big(-\nu + 1 - e^{-N(s)}\big) ds \Big) \geq 0,$$

y > 0 si $P_0 > 0$. Con N hay que ver primero que también lleva N como factor, y lo lleva: la función

$$g(N) := \frac{1 - e^{-N}}{N} \quad (\text{con } g(0) := 1)$$

es analítica en todo $\mathbb{R}$ —la singularidad en 0 es evitable, como se ve en el desarrollo $1 - e^{-N} = N - \frac{N^2}{2} + \cdots$—, y con ella

$$N' = \mu N - P\, N\, g(N) = N\big(\mu - P\, g(N)\big) \implies N(t) = N_0 \exp\Big(\int_0^t \big(\mu - P(s) g(N(s))\big) ds \Big) \geq 0.$$

Las dos exponenciales son positivas y los datos iniciales no negativos: queda demostrado en todo $I(N_0, P_0)$.

Otra vía (la de las órbitas, que es la que suele darse). Los dos ejes son invariantes: sobre $\{N = 0\}$ el sistema queda $N' = 0$, $P' = -\nu P$, cuya solución es $(0, P_0 e^{-\nu t})$; sobre $\{P = 0\}$ queda $P' = 0$, $N' = \mu N$, con solución $(N_0 e^{\mu t}, 0)$. Como las órbitas no se cortan, una que parta del cuadrante abierto no puede tocarlos. El argumento es el mismo de siempre; se da aquí la versión con la fórmula exponencial porque tiene un ingrediente que conviene ver: *hay que darse cuenta de que* $1 - e^{-N}$ *es divisible por* N, y eso no es evidente a primera vista.

Apartado (e): las cotas exponenciales y la existencia global

Paso 1: la cota de P. Para $t \geq 0$ ya se sabe, por (d), que $N(t) \geq 0$, luego $e^{-N} \leq 1$ y $1 - e^{-N} \leq 1$. Entonces

$$P' = P\big(-\nu + 1 - e^{-N}\big) \leq P \cdot 1 = P \implies \big(Pe^{-t}\big)' = e^{-t}(P' - P) \leq 0,$$

así que Pe^{-t} es decreciente en $[0, \infty) \cap I$ y $P(t)e^{-t} \leq P(0) = P_0$, es decir $P(t) \leq P_0 e^{t}$.

Paso 2: la cota de N. También por (d), $P \geq 0$ y $1 - e^{-N} \geq 0$, luego el término sustraendo es no negativo:

$$N' = \mu N - P(1 - e^{-N}) \leq \mu N \implies \big(Ne^{-\mu t}\big)' \leq 0 \implies N(t) \leq N_0 e^{\mu t}.$$

Paso 3 (el núcleo): de las cotas a la existencia global. Supóngase que $T^* := \sup I(N_0, P_0) < \infty$. En $[0, T^*)$ los dos pasos anteriores dan

$$0 \leq N(t) \leq N_0 e^{\mu T^*}, \qquad 0 \leq P(t) \leq P_0 e^{T^*},$$

de modo que la semiórbita positiva está contenida en el rectángulo compacto $[0, N_0 e^{\mu T^*}] \times [0, P_0 e^{T^*}]$, que está dentro del dominio $\mathbb{R}^2$. La propiedad de prolongación en compactos permitiría continuarla más allá de T^*, contra su maximalidad. Luego $[0, +\infty) \subset I(N_0, P_0)$. ■

LA IDEA: UNA COTA QUE CRECE BASTA, NO HACE FALTA QUE SEA UNIFORME

Merece la pena señalar lo que *no* se ha demostrado: las cotas $P_0 e^{t}$ y $N_0 e^{\mu t}$ crecen sin límite, así que no dicen que las poblaciones estén acotadas. Y no hace falta que lo estén. Para prolongar la solución basta con que *en cada intervalo finito* se quede dentro de un compacto, y eso es justo lo que dan estas cotas. Compárese con la convocatoria de 2026, donde la cota era uniforme ($N \geq K$ y $N \leq K/m$ para todo t) y los límites se obtenían combinando esa cota con las monotonías. Una cota uniforme sola tampoco garantiza que exista límite. Aquí se ha probado existencia global; la descripción asintótica requeriría otros argumentos.

Qué buscaba este examen

- **Ejercicio 1.** Reobtener a mano, para 3×3, lo que el teorema de Leslie da en general. Lo que se evalúa es si se sabe sacar información de los *signos* de los coeficientes de un polinomio característico, sin resolverlo.
- **Ejercicio 2.** La cadena entera de un modelo de dos poblaciones: adimensionalizar, equilibrios con la condición de existencia, estabilidad por linealización, positividad y prolongación. Y una pregunta conceptual, la comparación con Volterra, que es la única que no se contesta con una cuenta.

Errores típicos

- **Equivocar el signo del término independiente de p.** $\det(\lambda I - L)$ tiene $-a_3b_1b_2$; con el signo cambiado el apartado (b) se cae.
- **En (d) del ejercicio 1, concluir la dominancia solo de que $-\lambda_{max}$ no es autovalor.** Eso descarta un autovalor concreto, no todos los de módulo grande. Hay que argumentar con el signo de p a la derecha de λ_{max}.
- **Olvidar la condición $\nu < 1$** en el apartado (b) del ejercicio 2 y escribir $N^* = -\log(1-\nu)$ sin decir cuándo tiene sentido.
- **Calcular los autovalores en (c)** en vez de mirar traza y determinante. Se puede, pero se tarda el triple y la conclusión es la misma.
- **Dar por evidente que $1 - e^{-N}$ se anula con N como factor de primer orden.** Es cierto y es el paso que hace funcionar (d), pero hay que decirlo.
- **En (e), creer que hace falta una cota uniforme en todo el tiempo.** Las cotas en cada intervalo finito bastan para descartar explosión finita.

C.8 Primer parcial del 28 de marzo de 2022

Ficha

Asignatura	Modelización matemática, curso 2021/22
Prueba	Primer parcial
Fecha	28 de marzo de 2022
Lo que se conserva	Los dos ejercicios de teoría: $1 + 4 = 5$ puntos
Lo que falta	La parte de prácticas

Enunciado

Ejercicio 1 [Teoría: 1 punto]. Dado $E > 0$ la edad máxima de una población, describir el modelo de Leslie para aproximar el número de individuos de dicha población dividida en n clases de edad. Razonar que dicha aproximación se reduce a multiplicar repetidamente por la matriz de Leslie.

Ejercicio 2 [Cuestión teórica: 4 puntos]. Se considera el siguiente modelo SIR modificado con una tasa de crecimiento para la clase de susceptibles:

$$\begin{cases} S' = \alpha S - \beta SI, \\ I' = \beta SI - \gamma I, \\ R' = \gamma I. \end{cases} \tag{C.21}$$

Se pide:

a) (1 punto) Describir el modelo, comentando el significado de cada parámetro $\alpha, \beta, \gamma > 0$.

b) (1 punto) Realizar una adimensionalización adecuada para reducir los parámetros del sistema de la siguiente manera:

$$\begin{cases} S' = S - SI, \\ I' = \lambda(SI - I), \\ R' = I, \end{cases} \qquad \text{donde } \lambda = \gamma/\alpha.$$

c) (1 punto) Probar, usando los resultados conocidos sobre el sistema presa-depredador aplicados al subsistema (S, I), que la solución maximal de dicho subsistema $(S(t), I(t))$ asociada a cualquier dato inicial (S_0, I_0) con $S_0 > 0$, $I_0 > 0$ está definida para todo $t \in \mathbb{R}$, y que el punto crítico de dicho subsistema $(\bar{S} = 1, \bar{I} = 1)$ es estable pero no asintóticamente estable.

d) (1 punto) Probar que en el caso del apartado anterior, la solución $R(t)$ está definida hasta tiempo infinito, y que existe $\lim_{t\to+\infty} R(t) = +\infty$.

ALTO. Si no has hecho los dos ejercicios, cronometrado y a libro cerrado, no sigas leyendo.

Resolución del ejercicio 1

Es una pregunta de teoría pura, y lo que se evalúa es que la deducción esté completa y en orden. Es la de la sección 1.1.2, resumida.

Las hipótesis de partida. Se cuenta solo la población **femenina**, porque son las hembras las que producen descendencia y así el modelo se cierra con una sola variable por clase. Se supone conocida la edad máxima $E > 0$ que puede alcanzar un individuo, y se fija un entero $n \geq 1$, el número de clases.

La partición en clases. Se define el periodo

$$\tau := \frac{E}{n}$$

y se reparte la población en n clases de edad,

$$C_i = \{\text{individuos de edad en } [(i-1)\tau,\ i\tau)\}, \qquad i = 1, \ldots, n.$$

Se llama $v_i(k)$ al número de individuos de la clase C_i en el instante $t_k = k\tau$, y $v(k) = (v_1(k), \ldots, v_n(k))^t$ al vector de distribución por clases.

Por qué el paso de tiempo tiene que ser τ, y no otro. Esta es la decisión que hace funcionar todo el modelo y la que hay que justificar. Si se observa la población cada τ unidades de tiempo —exactamente la anchura de una clase—, entonces todo individuo que sobreviva de t_k a t_{k+1} ha envejecido justo lo necesario para pasar de C_i a C_{i+1}: ni se queda en su clase ni salta dos. Con otro paso, un individuo podría quedarse en la misma clase o saltarse una; la matriz de transición ya no tendría esta forma de avance de una sola clase.

Los parámetros vitales. Se suponen constantes en el tiempo: las tasas vitales no dependen del instante ni del tamaño de la población.

- $b_i \in (0, 1]$, para $i = 1, \ldots, n-1$: la fracción de individuos de C_i que sobrevive un periodo y pasa a C_{i+1}.
- $a_i \geq 0$, para $i = 1, \ldots, n$: el número medio de hijas que una hembra de la clase C_i produce durante un periodo y que están vivas para ser contadas en C_1 en la observación siguiente.

Las ecuaciones. Con esas definiciones, la contabilidad de un periodo es inmediata:

- La clase C_1 del instante siguiente son los recién nacidos, aportados por todas las clases: $v_1(k+1) = a_1 v_1(k) + a_2 v_2(k) + \cdots + a_n v_n(k)$.

- La clase C_{i+1} del instante siguiente son los supervivientes de C_i: $v_{i+1}(k+1) = b_i v_i(k)$, para $i = 1, \dots, n-1$.

No hay una clase posterior a C_n: la hipótesis de edad máxima excluye supervivientes de esa clase en la observación siguiente. Todo junto, $v(k+1) = Lv(k)$ con

$$L = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ b_1 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & b_{n-1} & 0 \end{pmatrix},$$

la **matriz de Leslie**.

Por qué se reduce a multiplicar repetidamente. Porque $v(k+1) = Lv(k)$ es una relación de recurrencia lineal de primer orden con coeficientes constantes. Por inducción sobre k: es cierto para $k = 0$ ($v(0) = L^0 v(0)$), y si $v(k) = L^k v(0)$ entonces $v(k+1) = Lv(k) = L \cdot L^k v(0) = L^{k+1} v(0)$. Luego

$$\boxed{v(k) = L^k v(0) \quad \text{para todo } k \geq 0.}$$

Conocer la población en los instantes de observación $k\tau$ se reduce a aplicar L^k al dato inicial. Los autovalores y autovectores permiten estudiar su comportamiento a largo plazo (sección 1.1.3).

Resolución del ejercicio 2

Apartado (a): qué dice el modelo

Es el SIR de la sección 2.2.1 con un solo cambio: el término $+\alpha S$ en la primera ecuación.

- $\alpha > 0$ es la **tasa de crecimiento por individuo de los susceptibles**. El aporte αS puede interpretarse como nacimientos que se incorporan a esta clase; el modelo supone que ese aporte es proporcional a S, no a la población total.
- $\beta > 0$ es el **coeficiente de transmisión**, con la hipótesis de mezcla homogénea: el número de contagios por unidad de tiempo es proporcional al producto SI, es decir, al número de encuentros posibles entre un sano y un enfermo.
- $\gamma > 0$ es la **tasa de recuperación**: el flujo de infectados a recuperados es γI, y $1/\gamma$ es la duración media de la infección en este modelo. Los recuperados no vuelven a ser susceptibles.

Lo que cambia respecto del SIR clásico, y hay que decirlo. La población total ya no se conserva:

$$N' = S' + I' + R' = \alpha S,$$

que es estrictamente positiva mientras haya susceptibles: **la población crece**. En ausencia de enfermedad ($I \equiv 0$) los susceptibles siguen $S' = \alpha S$ y crecen exponencialmente. El modelo no incorpora límites de recursos ni mortalidad natural. La reposición de susceptibles permite que la infección persista, como se demuestra en (c) y (d).

Apartado (b): la adimensionalización

Se busca $\hat{S}(t) = p\,S(st)$, $\hat{I}(t) = q\,I(st)$, $\hat{R}(t) = r\,R(st)$. Sustituyendo $S(st) = \hat{S}/p$, etc.:

$$\hat{S}' = ps\Big[\alpha\frac{\hat{S}}{p} - \beta\frac{\hat{S}}{p}\frac{\hat{I}}{q}\Big] = s\alpha\hat{S} - \frac{s\beta}{q}\hat{S}\hat{I},$$

$$\hat{I}' = qs\Big[\beta\frac{\hat{S}}{p}\frac{\hat{I}}{q} - \gamma\frac{\hat{I}}{q}\Big] = \frac{s\beta}{p}\hat{S}\hat{I} - s\gamma\hat{I}, \qquad \hat{R}' = rs\,\gamma\,\frac{\hat{I}}{q} = \frac{rs\gamma}{q}\hat{I}.$$

Comparando con el sistema pedido, $\hat{S}' = \hat{S} - \hat{S}\hat{I}$, $\hat{I}' = \lambda(\hat{S}\hat{I} - \hat{I})$, $\hat{R}' = \hat{I}$, con $\lambda = \gamma/\alpha$:

$$s\alpha = 1 \Rightarrow \boxed{s = \frac{1}{\alpha}}, \qquad \frac{s\beta}{q} = 1 \Rightarrow \boxed{q = \frac{\beta}{\alpha}}, \qquad \frac{s\beta}{p} = \lambda = \frac{\gamma}{\alpha} \Rightarrow \boxed{p = \frac{\beta}{\gamma}},$$

y las dos condiciones que quedan se cumplen solas, lo que es la comprobación de que el cambio es coherente: $s\gamma = \gamma/\alpha = \lambda$, y $\dfrac{rs\gamma}{q} = 1$ da $\boxed{r = \dfrac{q}{s\gamma} = \dfrac{\beta}{\gamma}}$.

Por qué queda un parámetro. Las tasas α y γ tienen unidades de inverso de tiempo. Al cambiar la escala temporal, ambas se multiplican por el mismo factor s; las escalas de población no alteran los coeficientes de los términos lineales. Por eso

$$\frac{s\gamma}{s\alpha} = \frac{\gamma}{\alpha} = \lambda$$

es un *invariante de escala*. Elegir $s = 1/\alpha$ normaliza la primera tasa y deja la segunda igual a λ: el cociente entre la tasa de recuperación y la de crecimiento de los susceptibles. El número de parámetros que queda depende de las escalas independientes, no solo de contar constantes.

Apartado (c): el subsistema (S, I) es exactamente un Volterra

Paso 1 (el núcleo): reconocer el modelo. Las dos primeras ecuaciones del sistema adimensionalizado no contienen R, así que forman un sistema cerrado:

$$S' = S(1 - I), \qquad I' = \lambda I(S - 1).$$

Compárese con el modelo de Volterra de la sección 2.2.4,

$$u' = u(a - bv), \qquad v' = v(-c + du),$$

y se ve que **son el mismo sistema** con

$$u = S, \quad v = I, \quad a = b = 1, \quad c = d = \lambda.$$

Los susceptibles ocupan el lugar de la presa y los infectados el del depredador en las ecuaciones. La interpretación epidemiológica es otra: el contagio transfiere individuos de S a I. El equilibrio no trivial de Volterra es $(c/d,\, a/b) = (1, 1)$, que es el que da el enunciado.

Paso 2: lo que se sabe de Volterra, aplicado aquí. La sección 2.2.4 demuestra tres cosas, y las tres se usan:

1. *Los ejes son invariantes y el cuadrante abierto también.* Con $S_0 > 0$ e $I_0 > 0$, la solución cumple $S(t) > 0$, $I(t) > 0$ mientras exista.

2. *Hay una integral primera.* Adaptando la del capítulo con los valores $a = b = 1$, $c = d = \lambda$:

$$H(S, I) := \lambda\big(S - \log S\big) + \big(I - \log I\big)$$

es constante a lo largo de las soluciones. Se comprueba en una línea:

$$\frac{d}{dt}H = \lambda\Big(1 - \frac{1}{S}\Big)S(1 - I) + \Big(1 - \frac{1}{I}\Big)\lambda I(S - 1) = \lambda(S - 1)(1 - I) + \lambda(I - 1)(S - 1) = 0.$$

3. *Los niveles por encima de $H(1,1)$ son curvas cerradas alrededor de $(1,1)$.* En efecto, H es estrictamente convexa: su hessiana es $\mathrm{diag}(\lambda/S^2, 1/I^2) > 0$, y su único punto crítico es $(1,1)$. Además, H tiende a infinito al acercarse a los ejes o al escapar hacia infinito, como se comprueba a continuación. Cada nivel no mínimo es, por tanto, la frontera regular de un conjunto convexo compacto; el nivel mínimo se reduce al equilibrio.

Paso 3: definición para todo $t \in \mathbb{R}$. La órbita que pasa por (S_0, I_0) está contenida en la curva de nivel $\{H = H(S_0, I_0)\}$, que es un compacto contenido en el cuadrante abierto —está acotada y no toca los ejes, porque $H \to +\infty$ tanto cuando $S \to 0^+$ o $I \to 0^+$ como cuando $S \to \infty$ o $I \to \infty$. Una solución cuya órbita vive en un compacto contenido en el dominio está definida en todo $\mathbb{R}$, hacia adelante y hacia atrás (criterio de prolongación en compactos aplicada en los dos sentidos). Luego $I(S_0, I_0) = \mathbb{R}$.

Paso 4: estable pero no asintóticamente estable. La función $V := H - H(1,1)$ es continua, se anula solo en $(1,1)$, es positiva alrededor (por ser $(1,1)$ un mínimo estricto) y cumple $\dot{V} \equiv 0$. Por el teorema de Liapunov (teorema de Liapunov, libro de EDO, §7.3 «»), $(1,1)$ es **estable**. Y **no** es asintóticamente estable: una solución que parte de un punto con $V = c > 0$ se queda en el conjunto $\{V = c\}$ para siempre. Si convergiera al equilibrio, la continuidad de V obligaría a que $c = V(S(t), I(t))$ tendiera a $V(1,1) = 0$, una contradicción. Las soluciones no constantes recorren periódicamente las curvas de nivel: el campo es continuo y no se anula en cada una de estas curvas compactas. Dan vueltas sin acercarse al centro. ■

Apartado (d): los recuperados crecen sin límite

Paso 1: R se obtiene por una cuadratura. La tercera ecuación es $R' = I$, con I ya conocida del apartado anterior y definida en todo $\mathbb{R}$. Luego

$$R(t) = R_0 + \int_0^t I(s)\,ds,$$

que está definida para todo $t \in \mathbb{R}$ por ser I continua. Aquí está la clave estructural de todo el ejercicio: R **no interviene en las dos primeras ecuaciones**, así que el sistema tiene una estructura triangular: se resuelve primero el sistema plano y después una integral. Por eso ha bastado con estudiar (S, I).

Paso 2 (el núcleo): I está separado de cero. La órbita permanece en el compacto $K = \{H = H(S_0, I_0)\}$ del apartado (c), contenido en el cuadrante abierto. La segunda coordenada alcanza en K un mínimo positivo:

$$I_{min} := \min_{(S,I)\in K} I > 0.$$

Por tanto $I(t) \geq I_{min}$ para todo t. Esto incluye el dato de equilibrio, para el que $I \equiv 1$. La periodicidad también permite obtener la cota, pero basta el confinamiento en K: ser positiva sin esta cota uniforme no sería suficiente.

Paso 3: la conclusión. Entonces $R'(t) = I(t) \geq I_{min} > 0$ para todo t, luego

$$R(t) \geq R_0 + I_{min}\, t \xrightarrow[t\to+\infty]{} +\infty,$$

y como R es creciente, el límite existe y vale $+\infty$. ■

LA IDEA: A QUÉ RITMO EXACTO CRECEN LOS RECUPERADOS

Para una solución no constante, de $S'/S = 1 - I$, integrando a lo largo de un periodo completo,

$$0 = \log S(T) - \log S(0) = \int_0^T \big(1 - I(t)\big)\,dt \quad \Longrightarrow \quad \frac{1}{T}\int_0^T I(t)\,dt = 1.$$

Es decir, **el valor medio de I a lo largo de un periodo es exactamente su valor de equilibrio**, $\bar{I} = 1$ (y el mismo argumento con $I'/I = \lambda(S-1)$ da que el valor medio de S es 1). También se obtiene una identidad exacta, válida incluso en el equilibrio:

$$R(t) = R_0 + t + \log\frac{S_0}{S(t)}.$$

Como S está acotado y separado de cero, el término logarítmico está acotado. Por tanto $R(t) \sim t$ cuando $t \to +\infty$: crece como una recta de pendiente uno, con una corrección acotada que oscila si la solución no es constante. En este modelo, la reposición de susceptibles permite que la infección persista; (S, I) oscila salvo cuando parte del equilibrio. En el SIR cerrado del capítulo 2, R converge a un límite finito porque no puede superar la población total constante; esto no exige que desaparezcan todos los susceptibles.

Qué buscaba este examen

- **Ejercicio 1.** Teoría de libro, contada entera y en orden, con la justificación de por qué el paso de tiempo es la anchura de la clase, que es lo que más se olvida.
- **Ejercicio 2.** Reconocer un modelo conocido debajo de otro. Toda la dificultad está en el apartado (c): quien ve que (S, I) es el sistema de Volterra tiene el ejercicio hecho, y quien no lo ve se pone a estudiar un sistema de tres ecuaciones desde cero.

Errores típicos

- **Contar machos y hembras** en la descripción del modelo de Leslie, o no decir por qué el paso de tiempo coincide con la anchura de las clases.
- **Decir que la población total es constante** en el SIR modificado. No lo es: $N' = \alpha S > 0$.
- **Intentar el apartado (c) por linealización.** La matriz jacobiana en $(1, 1)$ tiene autovalores imaginarios puros y el teorema no decide. Hay que ir a la integral primera.
- **Confundir estable con asintóticamente estable.** Aquí la distinción es justamente lo que se pregunta, y la respuesta está en que $\dot{V} = 0$ y no $\dot{V} < 0$.
- **En (d), argumentar solo que R es creciente.** Creciente no implica que tienda a infinito; hace falta una cota inferior positiva de la derivada, y la da el confinamiento en un compacto del cuadrante abierto.

C.9 Ejercicios sueltos de las convocatorias de 2016 y 2017

Los dos cuadernos reúnen ejercicios de 2016 y 2017 sin identificar la convocatoria de cada uno. Se presentan agrupados por tema: modelos de poblaciones, cálculo de variaciones, control óptimo, mecánica lagrangiana y prácticas. Complementan los exámenes anteriores con distintas condiciones de frontera y formas de modelizar.

Cada ejercicio lleva primero el enunciado y, después de la raya, la resolución. Pueden intentarse por separado antes de consultar la solución.

C.9.1 Modelos de poblaciones

Simbiosis simplificada (cuaderno de 2017, ejercicio 1)

Enunciado. Se considera el modelo de simbiosis simplificado

$$\begin{cases} u' = u(-a + bv), \\ v' = v(-c + du), \end{cases} \tag{C.22}$$

donde $a, b, c, d > 0$.

a) Mediante un cambio de variable adecuado, reducir el sistema (C.22) a este otro:

$$\begin{cases} u' = u(-1 + v), \\ v' = \lambda v(-1 + u), \end{cases} \tag{C.23}$$

con $\lambda > 0$.

b) Determinar los puntos críticos y sus propiedades de estabilidad de forma razonada.

c) Calcular las órbitas de (C.23).

d) Indicar el comportamiento cuando $t \to \pm\infty$ de las órbitas correspondientes a distintas elecciones de los datos iniciales (u_0, v_0) con $u_0, v_0 > 0$.

Resolución. Es el hermano pequeño del ejercicio 1 del examen de junio de 2025: aquel tenía además un término logístico $-e\bar{v}$ en la segunda ecuación, y este no. Conviene hacer los dos y comparar, porque la diferencia entre ellos es exactamente lo que aporta la autolimitación.

a) El cambio de unidades. Con $\hat{u}(s) = p\,u(rs)$ y $\hat{v}(s) = q\,v(rs)$,

$$\hat{u}'(s) = pr\,u(rs)\big(-a + b\,v(rs)\big) = r\,\hat{u}\Big(-a + \frac{b}{q}\hat{v}\Big), \qquad \hat{v}'(s) = r\,\hat{v}\Big(-c + \frac{d}{p}\hat{u}\Big).$$

Para que la primera sea $\hat{u}' = \hat{u}(-1 + \hat{v})$ hacen falta $b/q = a$ y $ra = 1$, es decir

$$\boxed{q = \frac{b}{a}}, \qquad \boxed{r = \frac{1}{a}},$$

y para que la segunda sea $\hat{v}' = \lambda\hat{v}(-1 + \hat{u})$, $d/p = c$ y $\lambda = rc$:

$$\boxed{p = \frac{d}{c}}, \qquad \boxed{\lambda = \frac{c}{a} > 0}.$$

Cuatro constantes, tres unidades libres, un parámetro superviviente. Y $\lambda = c/a$ es el cociente de las dos tasas de decrecimiento: compara la rapidez con que decaen las especies cuando están aisladas.

b) Equilibrios y estabilidad. De $u(-1+v) = 0$ y $\lambda v(-1+u) = 0$ salen $(0,0)$ y $(1,1)$. La matriz jacobiana es

$$Df(u,v) = \begin{pmatrix} -1+v & u \\ \lambda v & \lambda(-1+u) \end{pmatrix}.$$

En el origen, $Df(0,0) = \operatorname{diag}(-1,-\lambda)$, con los dos autovalores negativos: $(0,0)$ es **exponencialmente asintóticamente estable**. En $(1,1)$,

$$Df(1,1) = \begin{pmatrix} 0 & 1 \\ \lambda & 0 \end{pmatrix}, \qquad \det = -\lambda < 0,$$

luego los autovalores son $\pm\sqrt{\lambda}$, reales y de signos opuestos: **punto de silla, inestable**.

c) Las órbitas. Los ejes son invariantes, con las órbitas semitriviales $\{(0,v) : v > 0\}$ (donde $v(t) = v_0 e^{-\lambda t}$) y $\{(u,0) : u > 0\}$ (donde $u(t) = u_0 e^{-t}$), más los dos equilibrios. En el cuadrante abierto se elimina el tiempo:

$$\frac{du}{dv} = \frac{u(-1+v)}{\lambda v(-1+u)} \implies \lambda\,\frac{u-1}{u}\,du = \frac{v-1}{v}\,dv \implies \lambda\big(u - \log u\big) - \big(v - \log v\big) = C.$$

Escribiendo $\Psi(s) := s - \log s - 1$, que es la función de siempre en estos modelos —convexa, con mínimo $\Psi(1) = 0$ y $\Psi \to +\infty$ en 0^+ y en $+\infty$—, las órbitas del cuadrante abierto están contenidas en los conjuntos de nivel

$$\boxed{\Lambda(C) = \{(u,v) : u,v > 0,\ \lambda\Psi(u) - \Psi(v) = C\}, \qquad C \in \mathbb{R}.}$$

Cada órbita no constante es una componente conexa del nivel tras retirar el equilibrio, si lo contiene. El nivel $C = 0$ contiene $(1,1)$ y cuatro ramas, no una sola órbita.

d) El comportamiento. En Volterra la integral primera es $\lambda\Psi(u) + \Psi(v)$: sus niveles positivos son óvalos cerrados. Aquí aparece una diferencia y los niveles tienen ramas no acotadas. Para distinguir sus destinos, basta identificar las dos curvas del nivel $C = 0$ que pasan por $(1,1)$:

- La curva decreciente $v = S(u)$ une los sectores $u < 1 < v$ y $v < 1 < u$. Sus dos ramas se acercan al equilibrio al avanzar el tiempo: forman la variedad estable.
- La curva creciente $v = U(u)$ une los sectores $u, v < 1$ y $u, v > 1$. Sus ramas salen del equilibrio: forman la variedad inestable.

Ambas se obtienen de $\Psi(v) = \lambda\Psi(u)$ eligiendo, para cada u, la solución v en el sector indicado. La monotonía de Ψ a cada lado de 1 hace única esa elección. Las curvas son fronteras porque dos órbitas distintas no pueden cruzarse.

Hacia delante:

- Si $v_0 < S(u_0)$, entonces $(u(t), v(t)) \to (0,0)$ cuando $t \to +\infty$.
- Si $v_0 = S(u_0)$, la solución converge a $(1,1)$; es constante si parte de ese punto.
- Si $v_0 > S(u_0)$, ambas componentes tienden a $+\infty$ cuando $t \uparrow T_+$, donde $T_+ < \infty$ es el extremo derecho del intervalo maximal.

El signo de $u' = u(v-1)$ y de $v' = \lambda v(u-1)$ da los sentidos de recorrido. Salvo en la curva estable, la órbita acaba entrando en $u, v < 1$ o en $u, v > 1$; ahí ambas componentes son monótonas y el nivel determina el límite. *La explosión tarda un tiempo finito*: sobre la rama

donde ambas crecen sin cota, $\Psi(s) \sim s$ implica $v/u \to \lambda$. Por tanto, a partir de cierto instante $u' = u(v-1) \geq (\lambda/4)u^2$; integrar $(1/u)' \leq -\lambda/4$ impide que la solución continúe para todo tiempo positivo. Aunque el enunciado diga $t \to +\infty$, en este caso el límite correcto es $t \uparrow T_+$.

Hacia atrás: con $s = -t$, el sistema satisface $\frac{d}{ds}u(-s) \leq u(-s)$ y $\frac{d}{ds}v(-s) \leq \lambda v(-s)$. Las cotas exponenciales garantizan existencia para todo $t \leq 0$. Recorriendo los mismos niveles en sentido inverso se obtiene

dato inicial	límite cuando $t \to -\infty$
$v_0 < U(u_0)$	$u \to +\infty, \quad v \to 0$
$v_0 = U(u_0)$	$(u, v) \to (1, 1)$
$v_0 > U(u_0)$	$u \to 0, \quad v \to +\infty.$

Por ejemplo, si $\lambda = 1$ y $u_0 = v_0 = 2$, la diagonal es invariante y $u(t) = v(t) = 2/(2 - e^t)$: ambas poblaciones explotan en $T_+ = \log 2$, y hacia atrás tienden a 1, no a 0.

Un matiz sobre la monotonía, que es fácil dar por bueno y es falso. El destino sí lo decide el lado de la variedad estable, pero *las dos poblaciones no tienen por qué decrecer las dos, ni crecer las dos, desde el primer instante*: eso solo pasa dentro de las regiones $u, v < 1$ y $u, v > 1$. Si el dato cae en una de las dos regiones mixtas, una sube mientras la otra baja hasta que la órbita entra en la región que le toca. Dos ejemplos con $\lambda = 1$, comprobados integrando: desde $(0.2,\ 1.5)$ es $u'(0) = +0.10 > 0$ —la primera población *crece* al principio— y sin embargo el destino es $(0, 0)$; y desde $(0.5,\ 3)$ es $v'(0) = -1.5 < 0$ y el destino es el infinito. La frase correcta es que *a partir del momento en que la órbita entra en una de las dos regiones invariantes* —la de arriba a la derecha o la de abajo a la izquierda—, las dos componentes se mueven ya en el mismo sentido y no vuelven a salir de ahí.

LA IDEA: POR QUÉ LA SIMBIOSIS NO OSCILA Y EL PRESA-DEPREDADOR SÍ, VISTO EN LA INTEGRAL PRIMERA

Las dos integrales primeras se parecen tanto que la diferencia se puede señalar con el dedo:

$$\text{Volterra:} \quad \lambda\Psi(u) + \Psi(v) = C, \qquad \text{simbiosis:} \quad \lambda\Psi(u) - \Psi(v) = C.$$

Un signo. Pero $\Psi \geq 0$ con mínimo en 1, así que la *suma* solo puede ser pequeña si las dos coordenadas están cerca de 1: los conjuntos de nivel son óvalos alrededor de $(1, 1)$, y se recorren dando vueltas. La *diferencia*, en cambio, puede mantenerse constante creciendo las dos a la vez o decreciendo las dos a la vez: sus conjuntos de nivel se escapan por dos esquinas opuestas. Ese único signo es la razón matemática de que una relación de beneficio mutuo no produzca ciclos y una relación de depredación sí. Con el término logístico del examen de 2025 aparece autolimitación en la segunda especie y cambia la geometría de las órbitas. El equilibrio de coexistencia sigue siendo una silla; su estudio global requiere los argumentos específicos de aquel ejercicio.

La ley de Gompertz con extracción constante (cuaderno de 2016, ejercicio 2)

Enunciado. Se considera una población gobernada por la ley de Gompertz, sobre la que actúa un agente exterior:

$$\begin{cases} y' = \lambda y \log\Big(\dfrac{K}{y}\Big) - h, \quad t \geq 0, \\ y(0) = y_0, \end{cases} \tag{C.24}$$

donde $\lambda, K, y_0 > 0$ y $h \geq 0$. Se pide:

a) Para $h = 0$, determinar de manera razonada los puntos críticos y sus propiedades de estabilidad. Calcular las soluciones de (C.24) para los distintos valores de y_0 y determinar el comportamiento de las trayectorias cuando $t \to +\infty$.
b) Realizar un estudio análogo para $h > 0$.

Resolución.

Qué es la ley de Gompertz, en una frase. Es la tercera de las tres leyes de crecimiento del capítulo 2, junto a la exponencial y la logística. La exponencial dice que la tasa per cápita y'/y es constante; la logística, que decrece linealmente con la población; la de Gompertz, que decrece *logarítmicamente*:

$$\frac{y'}{y} = \lambda \log \frac{K}{y} = \lambda\big(\log K - \log y\big).$$

Su rasgo distintivo se ve en la propia ecuación: en la rama creciente, $y'' = f'(y)f(y)$ cambia de signo en $y = K/e$. Ahí se alcanza el ritmo máximo de crecimiento; en la logística, el punto correspondiente está en $K/2$.

Dónde está definido el problema. En $(0, +\infty)$: el logaritmo lo exige. En $y = 0$ el segundo miembro se puede extender por continuidad con valor $-h$ (porque $y \log(K/y) \to 0$), pero **no es localmente lipschitziano ahí**: su derivada $\lambda(\log(K/y) - 1)$ se va a $+\infty$. Conviene decirlo, porque significa que $y = 0$ no es un punto donde se pueda aplicar Picard, y con $h = 0$ es además el sitio hacia el que se va la solución cuando $t \to -\infty$ si $0 < y_0 < K$; si $y_0 > K$, crece sin cota hacia atrás.

Apartado (a): sin extracción.

Puntos críticos y estabilidad. Con $f(y) = \lambda y \log(K/y)$ en $(0, +\infty)$, el único cero es $y = K$ (el otro factor, y, no se anula en el dominio). Y

$$f'(y) = \lambda\Big(\log \frac{K}{y} + y \cdot \frac{y}{K} \cdot \Big(-\frac{K}{y^2}\Big)\Big) = \lambda\Big(\log \frac{K}{y} - 1\Big), \qquad f'(K) = -\lambda < 0,$$

luego K es exponencialmente asintóticamente estable.

La solución explícita, que aquí sí existe. El cambio $z = \log y$ linealiza la ecuación, y esa es la propiedad que hace especial a Gompertz:

$$z' = \frac{y'}{y} = \lambda(\log K - z) \implies z' + \lambda z = \lambda \log K,$$

una EDO lineal de primer orden. Con el factor integrante $e^{\lambda t}$, $z(t) = \log K + Ce^{-\lambda t}$, y de $z(0) = \log y_0$ sale $C = \log(y_0/K)$. Deshaciendo el cambio,

$$\boxed{y(t) = K\exp\Big(\log\frac{y_0}{K}\,e^{-\lambda t}\Big) = K\Big(\frac{y_0}{K}\Big)^{e^{-\lambda t}}.}$$

Es una **doble exponencial**, la firma de Gompertz. Como $e^{-\lambda t} \to 0$, el exponente tiende a 0 y

$$\lim_{t\to+\infty} y(t) = K \qquad \text{para todo } y_0 > 0,$$

de forma creciente si $y_0 < K$ y decreciente si $y_0 > K$. Hacia atrás, el exponente tiende a $\pm\infty$ y la solución se acerca a 0 o se va a $+\infty$, sin alcanzarlos: la solución maximal está definida en todo $\mathbb{R}$.

Apartado (b): con extracción constante $h > 0$. Ahora $f(y) = \lambda g(y) - h$ con $g(y) := y\log(K/y)$, y todo se decide estudiando g.

Paso 1 (el núcleo): la forma de g. Se tiene $g'(y) = \log(K/y) - 1$ y $g''(y) = -1/y < 0$, luego g es **estrictamente cóncava** en $(0, +\infty)$, con un único máximo donde $g' = 0$, es decir en $y = K/e$, y

$$g\Big(\frac{K}{e}\Big) = \frac{K}{e}\log e = \frac{K}{e}.$$

Además $g(y) \to 0$ cuando $y \to 0^+$, $g(K) = 0$ y $g \to -\infty$ cuando $y \to +\infty$. Es una campana asimétrica que sube desde 0, alcanza K/e en $y = K/e$ y baja cruzando el eje en $y = K$.

Paso 2: cuántos equilibrios hay. Los equilibrios resuelven $g(y) = h/\lambda$, y por la forma de g basta contar sus cortes con una recta horizontal:

- Si $h > \dfrac{\lambda K}{e}$: **ninguno**. La recta horizontal pasa por encima del máximo.
- Si $h = \dfrac{\lambda K}{e}$: **uno**, $y = K/e$, tangencial.
- Si $0 < h < \dfrac{\lambda K}{e}$: **dos**, $0 < \bar{y}_1 < K/e < \bar{y}_2 < K$, sin fórmula elemental en general (la ecuación es trascendente).

Paso 3: estabilidad. Como $f' = \lambda g'$ y g' es positiva a la izquierda de K/e y negativa a la derecha,

$$f'(\bar{y}_1) > 0 \;\Rightarrow\; \bar{y}_1 \text{ inestable}, \qquad f'(\bar{y}_2) < 0 \;\Rightarrow\; \bar{y}_2 \text{ exponencialmente asintóticamente estable.}$$

En el caso tangencial $h = \lambda K/e$ es $f'(K/e) = 0$ y el criterio no decide; se mira el signo de f, que es ≤ 0 en todo el dominio con igualdad solo en K/e: el equilibrio es **semiestable**: atrae por la derecha y repele por la izquierda; no es estable en el dominio completo.

Paso 4: el comportamiento, con el teorema escalar autónomo.

- $h > \lambda K/e$: $f < 0$ en todo $(0, +\infty)$, la solución es estrictamente decreciente y no hay ningún equilibrio por debajo. Como $f(y) \to -h < 0$ cuando $y \to 0^+$, la solución **tiende a cero en el extremo finito de su intervalo maximal**: se interpreta como extinción, y el modelo se detiene ahí.
- $h = \lambda K/e$: si $y_0 > K/e$, la solución decrece hasta K/e; si $y_0 = K/e$ es constante; si $y_0 < K/e$, extinción en tiempo finito.
- $0 < h < \lambda K/e$: si $y_0 > \bar{y}_1$, la solución converge a $\bar{y}_2$ (creciendo si $y_0 < \bar{y}_2$, decreciendo si $y_0 > \bar{y}_2$); si $y_0 = \bar{y}_1$ es constante; y si $y_0 < \bar{y}_1$, extinción en tiempo finito.

LA IDEA: LA CUOTA MÁXIMA SOSTENIBLE, Y POR QUÉ ESTÁ EN K/e

El número $\lambda K/e$ es la **máxima extracción sostenible**: por encima de ella no hay ningún nivel de población en el que lo que se saca se reponga, y la extinción es inevitable salga de donde salga. Es el mismo fenómeno del modelo de pesca con cuota fija de la sección 2.1.4, con dos números distintos: allí la producción biológica máxima era $\lambda^2/(4\kappa)$ y se alcanzaba con la población al **50 %** de la capacidad de carga; aquí es $\lambda K/e$ y se alcanza al $1/e \approx 36.8$ %. La diferencia entre esos dos porcentajes no es un detalle: es la que separa una recomendación pesquera de otra, y sale solo de haber elegido una curva de crecimiento u otra. Y por debajo del máximo aparece siempre el mismo peligro: dos equilibrios, uno inestable ($\bar{y}_1$) que hace de umbral y uno estable ($\bar{y}_2$) donde se asienta la población. Explotar cerca del máximo acerca los dos equilibrios y deja el sistema sin margen: un mal año que empuje la población por debajo de $\bar{y}_1$ la condena.

La ley de Gompertz con extracción proporcional (cuaderno de 2017, ejercicio 2)

Enunciado. Se considera una población gobernada por la ley de Gompertz, con acción exterior:

$$\begin{cases} y' = \lambda y \log\Big(\dfrac{K}{y}\Big) - hy, \quad t \geq 0, \\ y(0) = y_0, \end{cases} \tag{C.25}$$

donde $\lambda, K, y_0 > 0$ y $h \geq 0$. Se pide:

a) Determinar de manera razonada los puntos críticos y sus propiedades de estabilidad.
b) Calcular las soluciones de (C.25) para los distintos valores de y_0 y determinar el comportamiento de las trayectorias cuando $t \to +\infty$.

Resolución. Parece el mismo ejercicio del año anterior con una letra cambiada, y es un problema completamente distinto. Conviene hacerlos seguidos, porque la comparación es lo que de verdad enseña.

a) Equilibrios y estabilidad. Ahora el término de extracción lleva y como factor, así que *toda* la ecuación lo lleva:

$$f(y) = y\Big(\lambda \log \frac{K}{y} - h\Big), \qquad y > 0.$$

El paréntesis se anula cuando $\log(K/y) = h/\lambda$, es decir

$$\boxed{\bar{y} = K\, e^{-h/\lambda}}$$

y ese es el **único** equilibrio en $(0, +\infty)$, exista el $h \geq 0$ que exista. Además, de $f'(y) = \lambda(\log(K/y) - 1) - h$ se obtiene

$$f'(\bar{y}) = \lambda\Big(\frac{h}{\lambda} - 1\Big) - h = -\lambda < 0,$$

y el resultado es notable: **el exponente de la aproximación lineal es $-\lambda$, independiente de h**. La extracción baja el equilibrio sin cambiar ese exponente; la distancia inicial al nuevo equilibrio y la amplitud del transitorio sí pueden cambiar.

b) La solución explícita y el límite. El mismo cambio $z = \log y$ vuelve a linearizar:

$$z' = \lambda(\log K - z) - h \implies z' + \lambda z = \lambda \log K - h,$$

cuya solución es $z(t) = \log K - \frac{h}{\lambda} + Ce^{-\lambda t}$, con $C = \log(y_0/K) + h/\lambda$ al imponer el dato inicial. Deshaciendo el cambio,

$$\boxed{y(t) = K\,e^{-h/\lambda} \exp\Big(\Big(\log \frac{y_0}{K} + \frac{h}{\lambda}\Big)e^{-\lambda t}\Big),}$$

definida para todo $t \in \mathbb{R}$ y con

$$\lim_{t\to+\infty} y(t) = K\,e^{-h/\lambda} = \bar{y} \qquad \text{para todo } y_0 > 0,$$

monótonamente (creciendo si $y_0 < \bar{y}$, decreciendo si $y_0 > \bar{y}$). En este modelo, todo esfuerzo proporcional finito deja un equilibrio positivo; no aparece el umbral de extinción de la cuota fija. Eso no significa que el rendimiento sostenible crezca sin límite.

LA IDEA: UNA TASA PROPORCIONAL Y UNA CUOTA FIJA PRODUCEN UMBRALES DISTINTOS

Los dos ejercicios parten de la misma ley de crecimiento, pero el parámetro h tiene unidades distintas: población por tiempo en $-h$ e inversa de tiempo en $-hy$.

- **Cuota fija $-h$.** Se saca lo mismo haya muchos o pocos. Cuando la población baja, la extracción *no* baja con ella, y por eso puede llevársela entera: aparecen un umbral y una cuota máxima, y por encima de esta la extinción es segura.
- **Esfuerzo proporcional $-hy$.** Se extrae a una tasa per cápita fija. Cuando la población baja, la extracción baja también, automáticamente. Por eso el cero no es alcanzable: cerca de cero, la extracción es casi nula y el crecimiento la supera.

Y hay un premio de consolación que hace la comparación redonda. El rendimiento sostenido de la política proporcional es $h\bar{y} = hKe^{-h/\lambda}$, que como función de h se maximiza en $h = \lambda$ y vale entonces

$$\lambda K e^{-1} = \frac{\lambda K}{e},$$

exactamente el mismo máximo que se obtenía con la cuota fija, y alcanzado con la misma población de equilibrio $\bar{y} = K/e$. Las dos políticas comparten ese rendimiento máximo en equilibrio, pero la proporcional no tiene el umbral de extinción de la cuota fija en este modelo. La conclusión depende de la ley de Gompertz y sus hipótesis; no compara por sí sola todos los costes o riesgos de una explotación real.

C.9.2 Cálculo de variaciones: el Principio de Lagrange con extremos libres

Los dos ejercicios de este grupo tienen la misma estructura y evalúan lo mismo: que se sepa deducir las condiciones de optimalidad *desde cero*, y no aplicar una fórmula memorizada. La pieza que se pone a prueba es siempre la misma, y conviene tenerla escrita antes de empezar.

LA IDEA: EL MECANISMO, EN TRES MOVIMIENTOS, PARA NO VOLVER A DEDUCIRLO

Todo problema de este tipo se resuelve con los mismos tres movimientos, que son los de la sección 3.1:

1. **Se perturba.** Se toma una h admisible —en C^1 y respetando las condiciones que sí estén impuestas— y se mira la función de una variable real $F(\sigma) := \mathcal{F}(x + \sigma h)$. Si x es mínimo, F tiene un mínimo en $\sigma = 0$, luego $F'(0) = 0$. La única sutileza es que $x + \sigma h$ siga siendo admisible para σ pequeño, y eso se justifica con el argumento de compacidad de siempre: la imagen $t \mapsto (x(t), x'(t))$ es un compacto dentro del abierto admisible, y queda a distancia positiva de su frontera.
2. **Se integra por partes.** $F'(0)$ contiene un término con h y otro con h', y hay que juntarlos: h y h' *no* son independientes. Integrando por partes el término de h' aparecen dos cosas: una integral con h sola y unos **términos de frontera**.
3. **Se usa dos veces la arbitrariedad de h.** Primero con las h que se anulan en los dos extremos: los términos de frontera desaparecen y el lema fundamental (lema fundamental del Cálculo de Variaciones) da la **ecuación de Euler-Lagrange**. Después, sabiendo ya que la integral se anula, con las h que *no* se anulan en el extremo libre: lo que queda son los términos de frontera igualados a cero, y esas son las **condiciones naturales**.

La regla para estos problemas escalares es esta: en un extremo fijado, la variación se anula; en un extremo libre, su valor es arbitrario y obliga a anular el coeficiente de frontera. Así se distingue la condición impuesta de la natural. Si hay un coste en el extremo, $G(x(b))$, la condición natural allí es $f_{x'}(b) + G'(x(b)) = 0$: el término de frontera de la integración por partes más la derivada del coste.

Un funcional cuártico con coste final (cuaderno de 2016, ejercicio 5)

Enunciado. Se considera el problema

$$\begin{cases} \text{mín} \displaystyle\int_0^1 \left(|x'|^4 + x^2 - x\right) dt + |x(1)|^2 \\ x \in C^1([0,1]), \quad x(0) = 0. \end{cases} \tag{C.26}$$

Obtener las condiciones de optimalidad que se deducen del Principio de Lagrange.

Resolución. Hay una condición impuesta en $t = 0$ y ninguna en $t = 1$: en 1 habrá condición natural, y con coste final.

Paso 1: la derivada. Las variaciones admisibles son las $h \in C^1([0,1])$ con $h(0) = 0$, y $h(1)$ es libre. Con $\mathcal{F}(x) = \int_0^1 (x'^4 + x^2 - x)dt + x(1)^2$,

$$\frac{d}{d\sigma}\Big|_{\sigma=0} \mathcal{F}(x + \sigma h) = \int_0^1 \left(4x'^3 h' + 2xh - h\right)dt + 2x(1)h(1) = 0.$$

Paso 2 (el núcleo): integrar por partes el término de h'. Antes hay que justificar qué se deriva. Solo sabemos que $x \in C^1$, por lo que introducimos el momento $P = 4x'^3$, continuo. Con

$h(0) = h(1) = 0$, la identidad de la primera variación y el lema de Du Bois-Reymond de la sección 3.1 garantizan que $P \in C^1$. Ahora sí podemos integrar por partes:

$$\int_0^1 Ph'\,dt = \big[Ph\big]_0^1 - \int_0^1 P'h\,dt = 4x'(1)^3 h(1) - \int_0^1 P'h\,dt,$$

donde el término en $t = 0$ se ha ido porque $h(0) = 0$ (aquí es donde se usa la condición impuesta). Sustituyendo,

$$\int_0^1 \big(-P' + 2x - 1\big)h\,dt + \big(4x'(1)^3 + 2x(1)\big)h(1) = 0 \qquad \text{para toda } h \in C^1,\ h(0) = 0.$$

Paso 3: usar dos veces la arbitrariedad. Tomando además $h(1) = 0$, el corchete desaparece y el lema fundamental da la ecuación en el interior. Con eso ya sabido, la integral se anula para toda h, y queda el corchete multiplicado por $h(1)$, que es libre: el corchete tiene que ser cero. Sistema de optimalidad:

$$\boxed{\begin{cases} \dfrac{d}{dt}\big(4x'(t)^3\big) = 2x(t) - 1, \quad t \in (0,1), \\ x(0) = 0, \\ 2\,x'(1)^3 + x(1) = 0. \end{cases}}$$

Comprobación con la fórmula general. Con $f(t,x,p) = p^4 + x^2 - x$ es $f_p = 4p^3$ y $f_x = 2x - 1$, de modo que Euler-Lagrange $\frac{d}{dt}f_{x'} - f_x = 0$ da $P' = 2x - 1$: lo mismo. Donde $x' \neq 0$, se puede despejar x' como función derivable de P y escribir $12x'^2x'' = 2x - 1$. En un cero de x' esa regularidad no está garantizada; por eso se conserva la forma con P'. Y con $G(s) = s^2$, la condición natural $f_{x'}(1) + G'(x(1)) = 0$ es $4x'(1)^3 + 2x(1) = 0$: lo mismo.

La idea: qué dice la condición natural, y por qué es un equilibrio de precios

La ecuación de Euler-Lagrange gobierna el interior; la condición natural gobierna el final. Léase así: $f_{x'}(1) = 4x'(1)^3$ es lo que cuesta, en el último instante, seguir moviéndose; $G'(x(1)) = 2x(1)$ es lo que cuesta acabar donde se acaba. La condición dice que esos dos costes marginales se compensan exactamente. Con coste final $\beta x(1)^2$, la condición sería $4x'(1)^3 + 2\beta x(1) = 0$: cambiar el peso modifica ese equilibrio entre el momento y el valor final. Sin coste final quedaría $x'(1) = 0$, como en el ejercicio de los dos extremos libres. Es la misma lógica que en el capítulo 4 hace que el estado adjunto valga cero en T cuando no hay coste terminal.

Una coda que el examen no pedía. El funcional es **convexo**: $p \mapsto p^4$ es convexa, $x \mapsto x^2 - x$ es convexa, y $s \mapsto s^2$ también. Por el teorema de suficiencia por convexidad, cualquier solución del sistema de optimalidad es un mínimo global; y como $p \mapsto p^4$ es estrictamente convexa, ese mínimo es único. Es decir, en este problema las condiciones necesarias son además suficientes, cosa que no ocurre en el ejercicio siguiente.

Un funcional no convexo con los dos extremos libres (cuaderno de 2016, ejercicio 7)

Enunciado. Se considera el problema

$$\begin{cases} \text{mín} \displaystyle\int_0^1 \big(|x'|^2 + \operatorname{sen}(x)\big)\, dt \\ x \in C^1([0,1]). \end{cases} \tag{C.27}$$

Obtener las condiciones de optimalidad que se deducen del Principio de Lagrange. ¿Qué se puede decir cuando se minimiza $\displaystyle\int_0^1 \big(|x'|^2 + \operatorname{sen}(x)\big)dt + |x(1) - x(0)|^2$?

Resolución. Ahora **no hay ninguna condición impuesta**: h es cualquier función de $C^1([0,1])$, y por tanto habrá condición natural en los dos extremos.

Primera parte. La derivada es

$$\int_0^1 \big(2x'h' + \cos(x)h\big)dt = 0,$$

Con variaciones nulas en los extremos, Du Bois-Reymond da $2x' \in C^1$; aquí sí se deduce $x \in C^2$. Integrando por partes el primer sumando,

$$\big[2x'h\big]_0^1 + \int_0^1 \big(\cos(x) - 2x''\big)h\, dt = 0.$$

Con $h(0) = h(1) = 0$ y el lema fundamental sale la ecuación; después, con $h(0)$ y $h(1)$ libres e independientes, se anula cada término de frontera por separado:

$$\boxed{\begin{cases} 2x''(t) = \cos\big(x(t)\big), & t \in (0,1), \\ x'(0) = 0, \quad x'(1) = 0. \end{cases}}$$

(El signo importa y es fácil equivocarlo: Euler-Lagrange es $\frac{d}{dt}f_{x'} - f_x = 0$, y con $f = x'^2 + \operatorname{sen} x$ eso es $2x'' - \cos x = 0$, no $2x'' + \cos x = 0$.)

Y aquí este problema se puede resolver del todo, cosa rara. Obsérvese que

$$\int_0^1 \big(x'^2 + \operatorname{sen} x\big)dt \;\geq\; \int_0^1 (0 - 1)\, dt = -1,$$

porque $x'^2 \geq 0$ y $\operatorname{sen} x \geq -1$, y que la igualdad exige las dos cosas a la vez: $x' \equiv 0$ y $\operatorname{sen} x \equiv -1$. Luego los mínimos globales son exactamente las funciones constantes

$$x \equiv \frac{3\pi}{2} + 2k\pi, \qquad k \in \mathbb{Z},$$

con valor mínimo -1. Se comprueba que cumplen el sistema: son constantes, luego $x'' = 0$ y $x'(0) = x'(1) = 0$, y $\cos(3\pi/2) = 0$. **Hay infinitos mínimos**, y eso no contradice nada: el funcional no es convexo, porque el seno no lo es, y el teorema de suficiencia por convexidad no se puede aplicar.

Segunda parte: el término $|x(1) - x(0)|^2$. Repitiendo la cuenta, la derivada gana un sumando:

$$\big[2x'h\big]_0^1 + \int_0^1 \big(\cos x - 2x''\big)h\, dt + 2\big(x(1) - x(0)\big)\big(h(1) - h(0)\big) = 0.$$

Con $h(0) = h(1) = 0$ sale la misma ecuación interior, $2x'' = \cos x$: **un término que solo mira a la frontera no cambia la ecuación interior**. Y agrupando después los coeficientes de $h(1)$ y de $h(0)$, que siguen siendo libres e independientes:

$$2x'(1) + 2\big(x(1) - x(0)\big) = 0, \qquad -2x'(0) - 2\big(x(1) - x(0)\big) = 0,$$

es decir

$$\boxed{x'(0) = x'(1) = x(0) - x(1).}$$

La trayectoria sí puede cambiar al cambiar su frontera. Las dos condiciones naturales, que antes eran independientes ($x'(0) = 0$ y $x'(1) = 0$), quedan ahora **acopladas**: las dos pendientes tienen que ser iguales entre sí, y su valor común es el desnivel entre los extremos, cambiado de signo. Los mínimos globales, sin embargo, no cambian: las constantes $3\pi/2 + 2k\pi$ tienen $x'(0) = x'(1) = 0$ y $x(0) - x(1) = 0$, y siguen dando el valor -1, que sigue siendo el ínfimo porque el término añadido es ≥ 0.

LA IDEA: UN COSTE QUE CASTIGA EL DESNIVEL ES UN COSTE QUE EMPAREJA LOS EXTREMOS

El sumando $|x(1) - x(0)|^2$ penaliza que la curva no vuelva a donde empezó. Su efecto sobre las condiciones de optimalidad es *unir* los dos extremos, que antes no se hablaban: ahora las pendientes satisfacen una misma relación con el desnivel. Si se impone exactamente $x(0) = x(1)$, las variaciones cumplen $h(0) = h(1)$; el término de frontera $2[x'(1) - x'(0)]h(0)$ da entonces $x'(0) = x'(1)$. Con una penalización $\rho|x(1) - x(0)|^2$, una cota uniforme del coste fuerza el desnivel a tender a cero al crecer ρ: como la integral es al menos -1, un coste $\leq M$ implica $|x(1) - x(0)| \leq \sqrt{(M+1)/\rho}$. En este ejercicio concreto, los mínimos ya son constantes para cualquier $\rho \geq 0$.

C.9.3 Control óptimo: tres sistemas de optimalidad

Los tres ejercicios de este grupo piden lo mismo con tres estados distintos, y por eso se resuelven juntos. Conviene aislar primero el mecanismo, que es el del capítulo 4 y no cambia. Como en ese capítulo, se toman $T > 0$, pesos no negativos en la observación y $\mu > 0$; las ecuaciones con controles de L^2 se entienden casi en todo punto.

LA IDEA: EL MECANISMO, Y LA REGLA QUE DICE DÓNDE ENTRA CADA TÉRMINO DEL COSTE

En los tres problemas, el control óptimo $\hat{v}$ se caracteriza así:

1. Se perturba, $J(\hat{v} + \varepsilon w)$, y se deriva en $\varepsilon = 0$. Como el estado depende de forma afín del control (los datos iniciales están fijados), su derivada en la dirección w es la solución z_w de la misma ecuación con segundo miembro w y datos iniciales nulos.
2. En $J'(\hat{v})w$ aparece z_w, que no se conoce, y hay que quitarla de en medio. Para eso se inventa el **estado adjunto** $\hat{\varphi}$: la función que, al integrar por partes $\int \hat{\varphi}$ (ecuación de z_w), convierte todos los términos con z_w en un único $\int \hat{\varphi} w$.
3. Queda $J'(\hat{v})w = \int (\hat{\varphi} + \mu\hat{v})w = 0$ para toda w, y el lema fundamental en L^2 da $\hat{v} = -\hat{\varphi}/\mu$.

Lo que cambia de un problema a otro es solo **qué ecuación cumple** $\hat{\varphi}$, y ahí hay una regla que vale para los tres ejercicios y también para el examen de 2025: *cada término del coste entra en el problema adjunto en el sitio que le corresponde por dónde vive*. Un término

integral en $(0,T)$ se convierte en un **segundo miembro** de la ecuación adjunta en $(0,T)$; un término evaluado en T se convierte en la **condición terminal** del adjunto. Si solo hay término integral y el estado final es libre, las condiciones terminales son homogéneas: $\hat\varphi(T)=0$ en primer orden y $\hat\varphi(T)=\hat\varphi'(T)=0$ en el segundo problema. Si solo hay término final, la ecuación adjunta es homogénea. Con esa regla, los tres sistemas de optimalidad se escriben sin volver a deducirlos, y la deducción queda para justificarlos.
Y hay un detalle que se olvida y que distingue el segundo ejercicio de los otros dos: **la ecuación adjunta usa el adjunto formal del operador de estado**. Para una matriz se usa A^t; para una derivada primera, el adjunto formal de $\frac{d}{dt}$ es $-\frac{d}{dt}$ (una integración por partes, un cambio de signo), y por eso la ecuación adjunta se escribe $-\hat\varphi'=\dots$ y va hacia atrás; pero para una derivada **segunda** son dos integraciones por partes y dos cambios de signo, que se cancelan: la ecuación adjunta de y'' vuelve a ser φ'', con el mismo signo. Lo único que sigue yendo hacia atrás son las condiciones, que se imponen en T.

Un sistema de dos ecuaciones (cuaderno de 2016, ejercicio 6)

Enunciado. Se considera el problema de control óptimo

$$\begin{cases} \text{mín } J(v)=\dfrac{\alpha}{2}\displaystyle\int_0^T |y-y_d|^2dt+\frac{\mu}{2}\int_0^T |v|^2dt \\ \text{Sujeto a } v=(v_1,v_2)\in \big(L^2(0,T)\big)^2, \text{ con } (y,v) \text{ solución de:} \\ \qquad \begin{pmatrix} y_1' \\ y_2' \end{pmatrix}=\begin{pmatrix} 1 & -a \\ a & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}+\begin{pmatrix} v_1(t) \\ v_2(t) \end{pmatrix}, \qquad \begin{pmatrix} y_1(0) \\ y_2(0) \end{pmatrix}=\begin{pmatrix} y_{01} \\ y_{02} \end{pmatrix}. \end{cases} \tag{C.28}$$

Hallar el sistema de optimalidad asociado.

Resolución. Es el problema de la sección 4.3 con $A=\left(\begin{smallmatrix} 1 & -a \\ a & 1 \end{smallmatrix}\right)$, $B=I$ y $C=I$.

Paso 1: la derivada del funcional. Sea $\hat v$ el óptimo, $\hat y$ su estado, $w\in(L^2)^2$ y z_w la solución de $z'=Az+w$, $z(0)=0$. Entonces

$$J'(\hat v)\,w=\alpha\int_0^T\langle \hat y-y_d,\,z_w\rangle\,dt+\mu\int_0^T\langle \hat v,\,w\rangle\,dt=0.$$

Paso 2 (el núcleo): el adjunto y la traspuesta. Defínase $\hat\varphi$ como solución de

$$-\hat\varphi'=A^t\hat\varphi+\alpha(\hat y-y_d)\ \text{ en } (0,T), \qquad \hat\varphi(T)=0.$$

Entonces, derivando el producto escalar y usando la propiedad que define la traspuesta, $\langle A^t\varphi,z\rangle=\langle\varphi,Az\rangle$:

$$\frac{d}{dt}\langle\hat\varphi,z_w\rangle=\langle -A^t\hat\varphi-\alpha(\hat y-y_d),z_w\rangle+\langle\hat\varphi,Az_w+w\rangle=-\alpha\langle\hat y-y_d,z_w\rangle+\langle\hat\varphi,w\rangle.$$

Integrando entre 0 y T, el miembro de la izquierda vale $\langle \hat{\varphi}(T), z_w(T)\rangle - \langle \hat{\varphi}(0), z_w(0)\rangle = 0$ porque $\hat{\varphi}(T) = 0$ y $z_w(0) = 0$. Luego $\alpha \int \langle \hat{y} - y_d, z_w\rangle = \int \langle \hat{\varphi}, w\rangle$, y sustituyendo en el paso 1,

$$\int_0^T \langle \hat{\varphi} + \mu\hat{v},\, w\rangle\, dt = 0 \quad \text{para toda } w \implies \hat{v} = -\frac{1}{\mu}\hat{\varphi}.$$

Paso 3: el sistema, escrito con la matriz de este problema. Como $A^t = \left(\begin{smallmatrix} 1 & a \\ -a & 1 \end{smallmatrix}\right)$,

$$\boxed{\begin{cases} \hat{y}' = \begin{pmatrix} 1 & -a \\ a & 1 \end{pmatrix} \hat{y} + \hat{v}, & t \in (0,T), \\ \hat{y}(0) = (y_{01}, y_{02})^t, & \\ -\hat{\varphi}' = \begin{pmatrix} 1 & a \\ -a & 1 \end{pmatrix} \hat{\varphi} + \alpha(\hat{y} - y_d), & t \in (0,T), \\ \hat{\varphi}(T) = 0, & \\ \hat{v} = -\dfrac{1}{\mu}\hat{\varphi}. & \end{cases}}$$

LA IDEA: QUÉ SISTEMA ES ESTE, Y POR QUÉ EL ADJUNTO ES SU REFLEJO

La matriz se descompone como $A = I + aJ$ con $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ el giro de noventa grados. Como I y J conmutan,

$$e^{At} = e^t \begin{pmatrix} \cos at & -\sin at \\ \sin at & \cos at \end{pmatrix}:$$

un estado libre no nulo **gira con velocidad angular** a **mientras su norma crece con factor** e^t. Si $a \neq 0$ describe una espiral; si $a = 0$, una semirrecta. La traspuesta $A^t = I - aJ$ es la misma espiral girando al revés, y con los mismos autovalores $1 \pm ai$: también se escapa.

La ecuación adjunta, escrita hacia adelante, es $\hat{\varphi}' = -A^t\hat{\varphi} - \alpha(\hat{y} - y_d)$: su parte homogénea decae hacia adelante, pero crece al recorrerla desde T hacia 0. Por ejemplo, una perturbación terminal de norma 1 en esa parte homogénea se amplifica hasta $e^3 \simeq 20.09$ si $T = 3$; el giro no cambia la norma.

El sistema de optimalidad es un problema de contorno porque da $\hat{y}(0)$ y $\hat{\varphi}(T)$, mientras que ambas ecuaciones están acopladas mediante $\hat{v} = -\hat{\varphi}/\mu$. No basta integrar una sola vez con los datos conocidos. Se puede usar un método de disparo, ajustando el dato inicial del adjunto para satisfacer su condición final, pero la amplificación de errores puede hacerlo mal condicionado. Las prácticas de control escalar y control de un sistema lo resuelven directamente como problema de contorno. Cambiar la diagonal de A a -1 invertiría la estabilidad de las partes homogéneas aisladas, pero conservaría el acoplamiento y los datos en extremos distintos.

Un estado de segundo orden (cuaderno de 2016, ejercicio 8)

Enunciado. Se considera el problema de control óptimo

$$\begin{cases} \min J(v) = \dfrac{\alpha}{2}\displaystyle\int_0^T |y - y_d|^2 dt + \dfrac{\mu}{2}\int_0^T |v|^2 dt \\ \text{Sujeto a } v \in L^2(0,T), \text{ con } (y,v) \text{ solución de:} \\ \qquad y'' = ay + v(t),\ t \in (0,T), \qquad y(0) = y_0, \quad y'(0) = y_1. \end{cases} \tag{C.29}$$

Probar que el control óptimo, junto con el estado asociado y una función adicional φ (el estado adjunto asociado), verifica el sistema de optimalidad siguiente:

$$\begin{cases} y'' = ay + v(t), & t \in (0,T), \\ y(0) = y_0, \quad y'(0) = y_1, & \\ \varphi'' = a\varphi + \alpha(y - y_d), & t \in (0,T), \\ \varphi(T) = 0, \quad \varphi'(T) = 0, & \\ v = -\dfrac{1}{\mu}\varphi. & \end{cases}$$

Resolución. Es el sistema de optimalidad de segundo orden con $b = 0$, y aquí se pide la demostración.

Paso 1. Con z_w la solución de $z'' = az + w$, $z(0) = z'(0) = 0$,

$$J'(\hat{v})w = \alpha \int_0^T (\hat{y} - y_d) z_w \, dt + \mu \int_0^T \hat{v}\, w \, dt = 0.$$

Paso 2 (el núcleo): dos integraciones por partes, y por eso no hay cambio de signo. Sea $\hat{\varphi}$ la solución de $\hat{\varphi}'' = a\hat{\varphi} + \alpha(\hat{y} - y_d)$ con $\hat{\varphi}(T) = \hat{\varphi}'(T) = 0$. Despejando $w = z_w'' - a z_w$,

$$\int_0^T \hat{\varphi}\, w \, dt = \int_0^T \hat{\varphi}\, z_w'' \, dt - a \int_0^T \hat{\varphi}\, z_w \, dt,$$

y en la primera integral se integra por partes **dos** veces:

$$\int_0^T \hat{\varphi} z_w'' = \left[\hat{\varphi} z_w'\right]_0^T - \int_0^T \hat{\varphi}' z_w' = \underbrace{\left[\hat{\varphi} z_w'\right]_0^T}_{=0} - \underbrace{\left[\hat{\varphi}' z_w\right]_0^T}_{=0} + \int_0^T \hat{\varphi}'' z_w.$$

Los cuatro términos de frontera se anulan, y no por casualidad: los de $t = T$ porque $\hat{\varphi}(T) = \hat{\varphi}'(T) = 0$ —esa es la razón de que el adjunto necesite *dos* condiciones terminales, una por cada integración por partes— y los de $t = 0$ porque $z_w(0) = z_w'(0) = 0$. Sustituyendo,

$$\int_0^T \hat{\varphi}\, w = \int_0^T (\hat{\varphi}'' - a\hat{\varphi}) z_w = \alpha \int_0^T (\hat{y} - y_d) z_w.$$

Llevado al paso 1, $\int (\hat{\varphi} + \mu\hat{v}) w = 0$ para toda $w \in L^2(0,T)$, y el lema fundamental da $\hat{v} = -\hat{\varphi}/\mu$. ■

Lo que hay que quedarse de aquí. Dos integraciones por partes cambian el signo dos veces, así que el operador $\frac{d^2}{dt^2}$ es su propio adjunto formal y la ecuación de $\hat{\varphi}$ sale con el mismo signo que la de $\hat{y}$. Lo único que se invierte es *dónde* se ponen las condiciones: las del estado en $t = 0$, las del adjunto en $t = T$. El adjunto sigue siendo, como en todo el capítulo 4, un precio que se lee hacia atrás.

Un coste con término final añadido (cuaderno de 2017, ejercicio 3)

Enunciado. Se considera el problema de control óptimo

$$\begin{cases} \text{Minimizar } J(v) = \dfrac{\alpha}{2}\displaystyle\int_0^T |y(t) - y_d|^2 dt + \frac{\beta}{2}|y(T)|^2 + \frac{\mu}{2}\int_0^T |v|^2 dt \\ \text{Sujeto a } v \in L^2(0,T), \end{cases} \tag{C.30}$$

donde, para cada v, la función y es la solución de

$$y' = ay + v(t),\ t \in (0,T), \qquad y(0) = y_0.$$

Aquí $y_d \in C^0([0,T])$, $a, y_0, y_1 \in \mathbb{R}$ y $\alpha, \beta, \mu > 0$. Determinar y demostrar una condición necesaria de óptimo para este problema (sistema de optimalidad).

Resolución. Es la **suma** de los dos problemas básicos del capítulo 4: el de observación distribuida (sección 4.1) y el de observación final (sección 4.2). El término integral aporta el segundo miembro del adjunto y el coste final aporta su condición terminal; la ecuación de estado y la relación $\hat{v} = -\hat{\varphi}/\mu$ se mantienen.

Paso 1. Con z_w solución de $z' = az + w$, $z(0) = 0$,

$$J'(\hat{v})w = \alpha \int_0^T (\hat{y} - y_d) z_w \, dt + \beta\, \hat{y}(T)\, z_w(T) + \mu \int_0^T \hat{v}\, w \, dt = 0.$$

Los tres sumandos vienen de los tres términos del coste, en el mismo orden.

Paso 2 (el núcleo): el adjunto que absorbe los dos primeros. Se busca $\hat{\varphi}$ tal que $\alpha \int (\hat{y} - y_d) z_w + \beta \hat{y}(T) z_w(T) = \int \hat{\varphi} w$. Tómese

$$-\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y} - y_d) \text{ en } (0,T), \qquad \hat{\varphi}(T) = \beta\, \hat{y}(T).$$

Entonces

$$\frac{d}{dt}(\hat{\varphi} z_w) = \big(-a\hat{\varphi} - \alpha(\hat{y} - y_d)\big) z_w + \hat{\varphi}(a z_w + w) = -\alpha(\hat{y} - y_d) z_w + \hat{\varphi} w,$$

e integrando entre 0 y T, con $z_w(0) = 0$,

$$\underbrace{\hat{\varphi}(T)}_{=\beta \hat{y}(T)} z_w(T) = -\alpha \int_0^T (\hat{y} - y_d) z_w + \int_0^T \hat{\varphi} w,$$

que es justo lo que se buscaba. Sustituyendo en el paso 1 queda $\int (\hat{\varphi} + \mu \hat{v}) w = 0$ para toda w, y por el lema fundamental $\hat{v} = -\hat{\varphi}/\mu$.

Paso 3: el sistema de optimalidad.

$$\boxed{\begin{cases} \hat{y}' = a\hat{y} + \hat{v}, & t \in (0,T), \\ \hat{y}(0) = y_0, & \\ -\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y} - y_d), & t \in (0,T), \\ \hat{\varphi}(T) = \beta\, \hat{y}(T), & \\ \hat{v} = -\dfrac{1}{\mu}\hat{\varphi}. & \end{cases}}$$

Y la condición es también suficiente. J es la suma de tres funciones convexas de v —dos de ellas convexas por ser una convexa compuesta con una afín, la tercera estrictamente convexa por ser la norma al cuadrado—, luego J es estrictamente convexo. Por tanto cualquier punto crítico es el único mínimo. Además, $J(v) \geq \mu\|v\|_2^2/2$ y los argumentos de existencia del capítulo 4 garantizan que ese mínimo existe: la condición necesaria del enunciado es en realidad una caracterización.

LA IDEA: LOS TRES SISTEMAS DE OPTIMALIDAD, EN UNA TABLA

Para el estado escalar $y' = ay + v$, la aportación de cada coste se resume así:

Coste	Ecuación del adjunto	Condición terminal
$\frac{\alpha}{2}\int \lvert y - y_d\rvert^2$	$-\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y} - y_d)$	$\hat{\varphi}(T) = 0$
$\frac{\alpha}{2}\lvert y(T)\rvert^2$	$-\hat{\varphi}' = a\hat{\varphi}$	$\hat{\varphi}(T) = \alpha\hat{y}(T)$
las dos cosas	$-\hat{\varphi}' = a\hat{\varphi} + \alpha(\hat{y} - y_d)$	$\hat{\varphi}(T) = \beta\hat{y}(T)$
$\frac{\alpha}{4}\lvert y(T) - y_d\rvert^4$	$-\hat{\varphi}' = a\hat{\varphi}$	$\hat{\varphi}(T) = \alpha\lvert\hat{y}(T) - y_d\rvert^2(\hat{y}(T) - y_d)$

(La última fila es el examen de junio de 2025.) La regla es la que se anunció: lo que se paga *durante* el trayecto es una fuerza que actúa sobre el adjunto en todo $(0, T)$; lo que se paga *al llegar* es el valor con el que arranca el adjunto en su viaje hacia atrás. En los cuatro casos $\hat{v} = -\hat{\varphi}/\mu$: eso viene del término $\frac{\mu}{2}\int |v|^2$, que es el mismo en los cuatro.

C.9.4 Mecánica lagrangiana: tres sistemas

Los tres ejercicios de este grupo son del Principio de Mínima Acción de la sección 3.3, y los tres se resuelven con el mismo método de cuatro pasos, que conviene tener escrito porque el trabajo está en los tres primeros y no en el último.

LA IDEA: CÓMO SE MONTA UN PROBLEMA DE MECÁNICA LAGRANGIANA, EN CUATRO PASOS

1. **Elegir las coordenadas** que describen la posición usando el menor número posible de números. Si el sistema tiene ligaduras (una varilla rígida, una superficie), esas coordenadas las incorporan. Para ligaduras ideales, cuyas reacciones no realizan trabajo virtual, esas fuerzas no aparecen en las ecuaciones de las coordenadas libres.
2. **Escribir la posición cartesiana** en función de esas coordenadas y derivarla, para obtener la energía cinética $E_c = \frac{m}{2}|x'|^2$.
3. **Escribir la energía potencial** E_p sumando todas las contribuciones (gravedad, muelles) en función de las mismas coordenadas.
4. **Aplicar Euler-Lagrange** a $L = E_c - E_p$: una ecuación por coordenada, $\frac{d}{dt}\left(\partial L/\partial q_i'\right) - \partial L/\partial q_i = 0$.

Y una comprobación que ahorra disgustos: si L **no depende** de una coordenada q_i —se dice entonces que es *cíclica*—, su ecuación se reduce a $\frac{d}{dt}(\partial L/\partial q_i') = 0$, es decir, $\partial L/\partial q_i'$ es una **cantidad conservada**. Eso es el teorema de Noether en su versión más simple, y en el tercer ejercicio de este grupo da la conservación del momento angular sin hacer ninguna cuenta.

Un péndulo sujeto además a un muelle (cuaderno de 2016, ejercicio 9)

Enunciado. Se considera el movimiento de una masa m, debido a la gravedad, situada en el extremo de una varilla delgada de masa despreciable y longitud ℓ que gira alrededor del punto $(0,0)$ (péndulo). La masa se encuentra además sujeta al punto $(\ell,0)$ a través de un muelle de masa despreciable. Sabemos que la energía potencial del muelle es $k\,m\,d^2/2$, siendo d la distancia de la masa al punto $(\ell,0)$. Describiendo la posición a través del ángulo que forma la varilla con la vertical, obtener las ecuaciones del movimiento.

Resolución.

Paso 1: la coordenada. Una sola, el ángulo φ que forma la varilla con la vertical, medido hacia el lado del punto $(\ell,0)$. La varilla es rígida, así que la masa está siempre a distancia ℓ del origen y

$$x_1(t) = \ell \operatorname{sen}\varphi(t), \qquad x_2(t) = -\ell\cos\varphi(t).$$

Paso 2: la energía cinética. Derivando, $x_1' = \ell\cos\varphi\cdot\varphi'$ y $x_2' = \ell\operatorname{sen}\varphi\cdot\varphi'$, de donde $|x'|^2 = \ell^2\varphi'^2$ y

$$E_c = \frac{m}{2}|x'|^2 = \frac{m\ell^2}{2}\,\varphi'^2.$$

Paso 3 (el núcleo): la energía potencial, y la simplificación que la hace manejable. Hay dos contribuciones. La gravitatoria es $mgx_2 = -mg\ell\cos\varphi$. La del muelle es $kmd^2/2$ con

$$d^2 = (\ell - x_1)^2 + x_2^2 = \ell^2\big[(1-\operatorname{sen}\varphi)^2 + \cos^2\varphi\big] = \ell^2\big[1 - 2\operatorname{sen}\varphi + 1\big] = 2\ell^2\big(1-\operatorname{sen}\varphi\big),$$

donde se ha usado $\operatorname{sen}^2 + \cos^2 = 1$. La identidad pitagórica es la que convierte una expresión con cuadrados en una lineal en $\operatorname{sen}\varphi$, y es lo que hace que el problema salga. Por tanto

$$E_p = -mg\ell\cos\varphi + km\ell^2\big(1-\operatorname{sen}\varphi\big).$$

Paso 4: Euler-Lagrange. Con $L = E_c - E_p$,

$$L(\varphi,\varphi') = \frac{m\ell^2}{2}\varphi'^2 + mg\ell\cos\varphi - km\ell^2\big(1-\operatorname{sen}\varphi\big),$$

se tiene $\partial L/\partial\varphi' = m\ell^2\varphi'$, luego $\frac{d}{dt}(\cdot) = m\ell^2\varphi''$; y $\partial L/\partial\varphi = -mg\ell\operatorname{sen}\varphi + km\ell^2\cos\varphi$. La ecuación $\frac{d}{dt}L_{\varphi'} - L_\varphi = 0$ queda

$$m\ell^2\varphi'' + mg\ell\operatorname{sen}\varphi - km\ell^2\cos\varphi = 0,$$

y dividiendo por $m\ell^2$,

$$\boxed{\varphi'' = -\frac{g}{\ell}\operatorname{sen}\varphi + k\cos\varphi.}$$

LA IDEA: QUÉ SE VE EN LA ECUACIÓN SIN RESOLVERLA

Con $k = 0$ es el péndulo de la sección 3.3.2, $\varphi'' = -\frac{g}{\ell}\operatorname{sen}\varphi$. El muelle añade el término $k\cos\varphi$, que es positivo cuando $\varphi \in (-\pi/2, \pi/2)$: tira hacia el lado donde está el anclaje. Los equilibrios ya no son la vertical hacia abajo y hacia arriba, sino las

soluciones de

$$\frac{g}{\ell}\operatorname{sen}\varphi = k\cos\varphi \quad\Longleftrightarrow\quad \tan\varphi = \frac{k\ell}{g},$$

Hay dos posiciones por vuelta. La estable es $\varphi_s = \arctan(k\ell/g)$, inclinada hacia el anclaje; la opuesta, $\varphi_s + \pi$, es inestable. En efecto, la derivada del segundo miembro es negativa en la primera, donde el potencial tiene un mínimo estricto, y positiva en la segunda, donde tiene un máximo estricto. En la rama estable, $\varphi_s \to 0$ cuando $k \to 0$, y $\varphi_s \to \pi/2$ cuando $k \to \infty$: la posición de equilibrio se aproxima al anclaje.
La masa desaparece porque las tres energías contienen el mismo factor m. Esto depende de cómo está definido el muelle: su rigidez física es $K = km$. Si se mantiene K fijo y se cambia la masa, entonces $k = K/m$ cambia y el movimiento sí depende de m. La cancelación vale al mantener fijo el parámetro k del enunciado, además de los datos iniciales.

Una partícula entre dos muelles (cuaderno de 2016, ejercicio 11)

Enunciado. Se considera una partícula de masa m en el campo gravitatorio terrestre, cerca de la superficie, sujeta a dos muelles de masa despreciable.

a) Supongamos que el movimiento es plano, que los muelles están fijados a los puntos $P_1 = (-1, 0)$, $P_2 = (1, 0)$ y que la energía potencial debida al muelle en P_i es $k\,m\,\ell_i^2/2$, siendo ℓ_i la distancia a P_i y k una constante positiva. Recordando que la energía potencial debida al campo gravitatorio es mgx_2, obtener las ecuaciones del movimiento de la partícula.

b) Suponiendo $m = 0.25$, $k = 8$ y recordando que la aceleración gravitatoria es $g = 9.8$, las ecuaciones anteriores quedan así:

$$\begin{cases} x_1'' = -16x_1, \\ x_2'' = -16x_2 - 9.8. \end{cases} \tag{C.31}$$

Resolver este sistema suponiendo que $x_1(0) = x_2(0) = 0$, $x_1'(0) = a$ y $x_2'(0) = 0$, siendo a un número real. Comprobar que la partícula describe una órbita periódica de periodo $T = \pi/2$ s correspondiente a una elipse de centro a determinar.

Resolución.

a) Las ecuaciones. Aquí las coordenadas son directamente las cartesianas $x = (x_1, x_2)$, porque no hay ninguna ligadura. La energía cinética es $E_c = \frac{m}{2}(x_1'^2 + x_2'^2)$. Para la potencial, con

$$\ell_1^2 = (x_1 + 1)^2 + x_2^2, \qquad \ell_2^2 = (x_1 - 1)^2 + x_2^2,$$

la suma se simplifica sola, y ese es el paso que hay que ver:

$$\ell_1^2 + \ell_2^2 = 2x_1^2 + 2 + 2x_2^2 = 2\big(1 + x_1^2 + x_2^2\big),$$

porque los dobles productos $\pm 2x_1$ se cancelan. Luego

$$E_p = mgx_2 + \frac{km}{2}\big(\ell_1^2 + \ell_2^2\big) = mgx_2 + km\big(1 + x_1^2 + x_2^2\big).$$

Con $L = E_c - E_p$, las dos ecuaciones de Euler-Lagrange son $mx_i'' = \partial L/\partial x_i$, es decir

$$\boxed{x_1'' = -2kx_1, \qquad x_2'' = -g - 2kx_2.}$$

Con $m = 0.25$, $k = 8$, $g = 9.8$ se obtiene $x_1'' = -16x_1$ y $x_2'' = -16x_2 - 9.8$, que es (C.31).

b) La solución explícita. Las dos ecuaciones están **desacopladas** y se resuelven por separado. La primera es un oscilador armónico de frecuencia $\omega = \sqrt{2k} = 4$: $x_1(t) = C_1 \cos 4t + C_2 \operatorname{sen} 4t$, y con $x_1(0) = 0$, $x_1'(0) = a$,

$$x_1(t) = \frac{a}{4} \operatorname{sen}(4t).$$

La segunda es la misma con un término independiente; una solución particular constante es $x_2 \equiv -9.8/16 = -0.6125$, así que $x_2(t) = K_1 \cos 4t + K_2 \operatorname{sen} 4t - 0.6125$, y con $x_2(0) = 0$, $x_2'(0) = 0$ resulta $K_1 = 0.6125$, $K_2 = 0$:

$$x_2(t) = \frac{9.8}{16}\big(\cos(4t) - 1\big).$$

La órbita. Las dos componentes oscilan con la **misma** frecuencia 4, luego el periodo común es $T = 2\pi/4 = \pi/2$ segundos, como decía el enunciado. Y eliminando el tiempo con $\operatorname{sen}^2 + \cos^2 = 1$, para $a \neq 0$,

$$\boxed{\Big(\frac{x_1}{a/4}\Big)^2 + \Big(\frac{x_2 + 9.8/16}{9.8/16}\Big)^2 = 1,}$$

que es una **elipse de centro** $(0, -9.8/16) = (0, -0.6125)$ y semiejes $|a|/4$ en horizontal y $9.8/16$ en vertical. (Si $a = 0$ la elipse degenera en el segmento vertical que une $(0,0)$ con $(0, -2 \cdot 0.6125)$, recorrido de arriba abajo y vuelta.)

LA IDEA: POR QUÉ LA ÓRBITA ES UNA ELIPSE CERRADA Y NO UNA FIGURA ENREDADA

Dos oscilaciones perpendiculares producen en general una figura de Lissajous, que solo se cierra si el cociente de frecuencias es racional y que puede ser complicadísima. Aquí las dos frecuencias son *iguales*, y por eso la órbita es una elipse limpia. ¿Por qué son iguales? Porque los dos muelles, con la energía potencial $km\ell^2/2$, producen un potencial que resulta ser $km(x_1^2 + x_2^2)$ más una constante: es **isótropo**, tira igual en todas las direcciones. La gravedad, que solo actúa en vertical, no cambia la frecuencia: únicamente **desplaza el centro** de $(0,0)$ a $(0, -g/(2k))$, que es la posición en la que los dos muelles equilibran el peso. Es el principio de superposición en estado puro: la gravedad mueve el punto de equilibrio, no la manera de oscilar a su alrededor. Y otra vez la masa desaparece de las ecuaciones, porque el enunciado ha metido un factor m en la energía de los muelles.

Una partícula sobre un cono (cuaderno de 2017, ejercicio 4)

Enunciado. La posición en el instante t de una partícula de masa m puede ser descrita, utilizando coordenadas polares, en función de las variables r y ω como sigue:

$$x(t) = r(t)\cos\big(\omega(t)\big), \qquad y(t) = r(t)\operatorname{sen}\big(\omega(t)\big), \qquad z(t) = a\,r(t),$$

siendo $a > 0$.

a) Teniendo en cuenta que la única fuerza que actúa sobre la partícula es la de la atracción gravitatoria terrestre (que depende sólo de la altura z de la partícula), justifica que la acción (energía cinética menos energía potencial) viene dada por

$$L(r,\omega,r',\omega') = \frac{(1+a^2)m}{2}|r'|^2 + \frac{m}{2}r^2|\omega'|^2 - mgar.$$

b) Suponiendo que en el instante inicial $t = 0$ tenemos $(r(0),\omega(0)) = (r_0,\omega_0)$ y en el instante final $t = T$ tenemos $(r(T),\omega(T)) = (r_T,\omega_T)$, enuncia el problema del Cálculo de Variaciones que se deduce del Principio de Mínima Acción.

c) Obtén el problema de contorno que debe resolver la trayectoria seguida por la partícula (condición necesaria de mínimo). Explica la idea de la demostración.

Resolución.

Lo primero, que el enunciado no dice: la partícula está sobre un cono. Las tres ecuaciones dicen que la distancia al eje vertical es r y la altura es ar, es decir,

$$z = a\sqrt{x^2+y^2},$$

que es un **cono** de vértice en el origen, abierto hacia arriba. Las dos coordenadas (r,ω) describen esa superficie para $r > 0$: la distancia al eje y el ángulo. En el vértice $r = 0$ las coordenadas son singulares. Estudiaremos trayectorias que permanecen en $r > 0$, con $r_0, r_T > 0$; las ecuaciones obtenidas en estas coordenadas no describen por sí solas un paso por el vértice.

La ligadura requiere una reacción normal del cono. Se supone ideal y sin rozamiento: esa reacción no realiza trabajo virtual, por lo que solo se incorpora el potencial gravitatorio. Esta es la interpretación de la frase del enunciado «la única fuerza» compatible con que la partícula permanezca sobre el cono.

a) El lagrangiano. La expresión L es el lagrangiano; la acción es su integral temporal. Derivando las tres coordenadas cartesianas,

$$x' = r'\cos\omega - r\omega'\operatorname{sen}\omega, \qquad y' = r'\operatorname{sen}\omega + r\omega'\cos\omega, \qquad z' = a\,r',$$

y al sumar los cuadrados los dobles productos se cancelan y las identidades pitagóricas hacen el resto:

$$\begin{aligned}|x'|^2 + |y'|^2 + |z'|^2 &= r'^2\big(\cos^2\omega + \operatorname{sen}^2\omega\big) + a^2r'^2 \\ &\quad + r^2\omega'^2\big(\operatorname{sen}^2\omega + \cos^2\omega\big) \\ &= (1+a^2)r'^2 + r^2\omega'^2.\end{aligned}$$

Luego $E_c = \frac{m}{2}\big[(1+a^2)r'^2 + r^2\omega'^2\big]$. Y como la fuerza gravitatoria es $F = (0,0,-mg) = -\nabla V$ con $V(x,y,z) = mgz$, la energía potencial es

$$E_p = mgz = mgar.$$

Restando se obtiene exactamente la L del enunciado.

b) El problema variacional. El Principio de Mínima Acción dice que la trayectoria seguida hace estacionaria la acción entre los extremos dados; no garantiza que toda trayectoria sea un mínimo. El problema de minimización que pide el enunciado es

$$\begin{cases} \text{Minimizar } \displaystyle\int_0^T \Big(\frac{(1+a^2)m}{2} |r'(t)|^2 + \frac{m}{2} r(t)^2 |\omega'(t)|^2 - mg\, a\, r(t) \Big) dt \\ \text{Sujeto a } (r,\omega) \in \big(C^1([0,T])\big)^2, \quad r(t) > 0 \text{ en } [0,T], \\ (r,\omega)(0) = (r_0,\omega_0), \quad (r,\omega)(T) = (r_T,\omega_T). \end{cases}$$

Los dos extremos están fijados, así que *no* habrá condiciones naturales: las cuatro condiciones del problema de contorno son las cuatro que impone el enunciado. Todo mínimo satisface las ecuaciones de Euler–Lagrange; satisfacerlas es una condición de estacionariedad, no un certificado de mínimo.

c) El problema de contorno. La idea de la demostración es la de siempre, y aquí se resume porque el enunciado la pide explícitamente: se perturba con $h = (h_1, h_2) \in (C^1([0,T]))^2$ que se anule en los dos extremos, se deriva la función $\sigma \mapsto \int_0^T L(\ldots + \sigma h \ldots)$ en $\sigma = 0$. Las variaciones pequeñas son admisibles porque r tiene mínimo positivo en $[0,T]$. El lema de Du Bois–Reymond da que los momentos $(1+a^2)mr'$ y $mr^2\omega'$ son C^1; como $r > 0$, de ahí se obtiene $r, \omega \in C^2$. Ya se puede integrar por partes el término con h'. Los términos de frontera desaparecen porque $h(0) = h(T) = 0$, y el lema fundamental, aplicado a cada componente, da las ecuaciones de Euler–Lagrange

$$\frac{d}{dt} \nabla_{(r',\omega')} L - \nabla_{(r,\omega)} L = 0.$$

Calculando las cuatro derivadas parciales:

$$\frac{\partial L}{\partial r'} = (1+a^2) m\, r', \qquad \frac{\partial L}{\partial \omega'} = m\, r^2 \omega', \qquad \frac{\partial L}{\partial r} = m\, r\, \omega'^2 - mga, \qquad \frac{\partial L}{\partial \omega} = 0,$$

y sustituyendo, tras dividir por m:

$$\boxed{\begin{cases} (1+a^2)\, r'' = r\, \omega'^2 - ga, & t \in (0,T), \\ \dfrac{d}{dt}\big(r^2 \omega'\big) = 0, \quad \text{es decir} \quad r\, \omega'' + 2r'\omega' = 0, & t \in (0,T), \\ (r,\omega)(0) = (r_0,\omega_0), \qquad (r,\omega)(T) = (r_T,\omega_T). \end{cases}}$$

LA IDEA: LA SEGUNDA ECUACIÓN ES LA CONSERVACIÓN DEL MOMENTO ANGULAR, Y NO HA HABIDO QUE CALCULARLA

La acción **no depende de** ω, solo de ω': girar el embudo entero alrededor de su eje no cambia nada. Esa simetría se paga sola: la segunda ecuación de Euler-Lagrange se reduce a $\frac{d}{dt}(\partial L/\partial \omega') = 0$, es decir,

$$m\, r^2 \omega' = \text{constante},$$

que es exactamente el **momento angular** respecto del eje vertical. Es el teorema de Noether: a cada simetría continua del lagrangiano le corresponde una cantidad conservada. En la sección 3.3 aparece la otra cara de la misma moneda, la observación de

Beltrami: cuando L no depende del *tiempo*, lo que se conserva es la energía.
Y con la conservación se entiende el movimiento sin resolver nada. Llamando $\ell := r^2\omega'$ a la constante y sustituyendo $\omega' = \ell/r^2$ en la primera ecuación,

$$(1+a^2)\,r'' = \frac{\ell^2}{r^3} - ga,$$

una ecuación para r sola, con dos fuerzas enfrentadas: la gravedad $-ga$, que tira hacia el vértice, y el término ℓ^2/r^3, la *fuerza centrífuga*, que empuja hacia fuera y crece sin límite cuando $r \to 0$. Para justificar que no se alcanza el vértice se usa además la energía conservada:

$$\frac{E}{m} = \frac{1+a^2}{2}r'^2 + \frac{\ell^2}{2r^2} + gar.$$

Sus sumandos son no negativos, y, si $\ell \neq 0$, el segundo tiende a infinito cuando $r \to 0$. Una energía inicial finita mantiene por tanto r separado de cero. Los dos términos de la ecuación radial se equilibran donde $\ell^2/r^3 = ga$, es decir en $r_* = \left(\ell^2/(ga)\right)^{1/3}$, que es la órbita circular; alrededor de ella la partícula puede subir y bajar mientras gira. Este modelo ideal no incluye el rozamiento ni la rodadura de una moneda sobre un embudo.

C.9.5 Prácticas: tres poblaciones y un problema inverso

Los dos ejercicios de prácticas de estos cuadernos son de 2017. El segundo utiliza el mismo esquema de problema inverso por mínimos cuadrados que `InvSEAIR`, del examen de junio de 2025.

Dos datos que los enunciados no dan. En el primero **no aparece el valor de** μ, y en el segundo **no aparece** T. No se inventan: aquí se dice qué elección se hace para poder ejecutar los programas ($\mu = 1$ en el primero, $T = 100$ en el segundo, que es lo justo para pasar de la última observación), y todos los números que siguen dependen de esa elección.

Un modelo de tres poblaciones (cuaderno de 2017, prácticas)

Enunciado. Se considera el siguiente modelo de tres poblaciones:

$$\begin{cases} u' = u(1 - v - 0.5w), \\ v' = v(1 - u - w), \\ w' = \mu w(-1 + u + v). \end{cases} \tag{C.32}$$

Escribir un programa

```
function Poblaciones(mu, u0, v0, w0, T)
```

que, a partir de los valores de los datos iniciales μ, u_0, v_0, w_0 y T, lleve a cabo las tareas siguientes:

a) Resolver el problema de Cauchy asociado a (C.32) en el intervalo $[0, T]$, tomando como condiciones iniciales $u(0) = u_0$, $v(0) = v_0$, $w(0) = w_0$, y representar las trayectorias juntas

en una misma gráfica. Añadir a la gráfica las anotaciones necesarias para su correcta comprensión. Utilizar los siguientes datos para ejecutar el programa: $u_0 = 1.3$, $v_0 = 1.3$, $w_0 = 1$, $T = 40$.

b) En el caso de los datos anteriores, la componente w de la solución es periódica a partir de un cierto tiempo. Calcular el periodo.

c) En otra gráfica aparte, representar, junto con la tercera componente w ya calculada, la obtenida cuando se cambia el dato inicial w_0 por la mitad de su valor. Añadir las anotaciones adecuadas.

d) Interpretar los resultados numéricos obtenidos en el apartado anterior. (Escribir y entregar junto con el examen de teoría.)

Resolución.

Qué modelo es, antes de teclear. Para $\mu > 0$, los signos describen las interacciones. Las especies u y v crecen solas (el 1 de sus paréntesis) y se estorban mutuamente ($-v$ en la de u, $-u$ en la de v): son **competidoras**. La tercera, w, se extingue sola (el -1) y se beneficia de las otras dos: es un **depredador** que come de las dos, y a cambio les hace daño ($-0.5w$ y $-w$). Obsérvese la asimetría: w le hace la mitad de daño a u que a v, y por tanto v lleva las de perder.

Lo que hay que anticipar del apartado (b). La periodicidad «a partir de un cierto tiempo» del enunciado debe entenderse como una aproximación numérica tras el transitorio. Con $v_0 > 0$, la solución exacta satisface

$$v(t) = v_0 \exp\left(\int_0^t (1 - u(s) - w(s))\, ds \right) > 0$$

en todo instante finito de existencia: el sistema no pierde una ecuación de repente porque la gráfica de v parezca tocar el eje. Con $\mu = 1$ y los datos dados, v se hace muy pequeña. Si se desprecia su contribución, queda

$$u' = u\big(1 - 0.5w\big), \qquad w' = \mu w(-1 + u),$$

el modelo de Volterra de la sección 2.2.4, con equilibrio $(u, w) = (1, 2)$ y órbitas positivas cerradas. Este sistema reducido explica las oscilaciones casi periódicas observadas. El programa estima su periodo mediante los dos últimos máximos; esa medida no demuestra periodicidad exacta del sistema de tres poblaciones.

```
function Poblaciones(mu, u0, v0, w0, T)
% Nombre Apellido Apellido
% Poblaciones(mu, u0, v0, w0, T)
% u' = u*(1 - v - 0.5*w)
% v' = v*(1 - u - w)
% w' = mu*w*(-1 + u + v)
% El enunciado no da el valor de mu; aqui se ejecuta con mu = 1:
% Poblaciones(1, 1.3, 1.3, 1, 40)

close all
```

```
opc = odeset('RelTol', 1e-9, 'AbsTol', 1e-11);

% ---------- (a) Resolucion y trayectorias ----------------------------
sol = resolver(mu, [u0; v0; w0], T, opc);
t = linspace(0, T, 4000);
Y = deval(sol, t);

figure(1)
plot(t, Y(1,:), 'b-', t, Y(2,:), 'r-', t, Y(3,:), 'g-', 'LineWidth', 1.5)
xlabel('t'), ylabel('numero de individuos'), grid on
title(['Tres poblaciones, \mu = ', num2str(mu), ...
       ', (u_0,v_0,w_0) = (', num2str(u0), ',', num2str(v0), ',', num2str(w0), ')'])
legend('u(t): competidora 1', 'v(t): competidora 2', 'w(t): depredadora', ...
       'Location', 'northwest')

% ---------- (b) Estimacion del periodo tras el transitorio -----------
w = Y(3,:);
pk = find(w(2:end-1) > w(1:end-2) & w(2:end-1) >= w(3:end)) + 1;  % maximos
if numel(pk) >= 2
   t1 = fminbnd(@(s) -deval(sol, s, 3), t(pk(end-1)-1), t(pk(end-1)+1));
   t2 = fminbnd(@(s) -deval(sol, s, 3), t(pk(end)-1), t(pk(end)+1));
   P = t2 - t1;
   fprintf('Periodo estimado de w (maximos en t = %7.4f y %7.4f): P = %7.4f \n', ...
           t1, t2, P)
   figure(1), hold on
   plot([t1 t2], [deval(sol,t1,3) deval(sol,t2,3)], 'ks', 'MarkerSize', 9)
else
   disp('No hay dos maximos de w en [0,T]: hay que alargar T')
end
fprintf('Valores finales: u(T) = %8.4f, v(T) = %8.2e, w(T) = %8.4f \n', ...
        Y(1,end), Y(2,end), Y(3,end))

% ---------- (c) w con el dato inicial w0 y con w0/2 --------------------
sol2 = resolver(mu, [u0; v0; w0/2], T, opc);
W2 = deval(sol2, t, 3);
figure(2)
plot(t, w, 'g-', 'LineWidth', 1.6), hold on
plot(t, W2, 'm--', 'LineWidth', 1.6)
xlabel('t'), ylabel('w(t)'), grid on
title('Tercera poblacion con dos datos iniciales distintos')
legend(['w_0 = ', num2str(w0)], ['w_0 = ', num2str(w0/2)], 'Location', 'northwest')
shg
end

% ======================================================================
function sol = resolver(mu, y0, T, opc)
F = @(t, y) [ y(1)*(1 - y(2) - 0.5*y(3));
              y(2)*(1 - y(1) - y(3));
              mu*y(3)*(-1 + y(1) + y(2)) ];
sol = ode45(F, [0 T], y0, opc);
end
```

Valores de referencia contrastados con una integración independiente; los últimos dígitos pueden variar con las tolerancias.

SALIDA ESPERADA

```
>> Poblaciones(1, 1.3, 1.3, 1, 40)
Periodo estimado de w (maximos en t = 31.6914 y 38.0209): P = 6.3295
```

```
Valores finales: u(T) = 0.7468, v(T) = 5.36e-35, w(T) = 1.7992
```

Apartado (d): la interpretación. Con $\mu = 1$ y los datos dados:

- **Una competidora queda prácticamente ausente.** $v(40)$ es del orden de 10^{-35}, aunque sigue siendo positiva. El depredador daña a v el doble que a u (coeficientes 1 frente a 0.5), lo que favorece a u.
- **El régimen final observado se aproxima al Volterra entre u y w.** En la parte final de la ventana, w oscila aproximadamente entre 1.46 y 2.65. El valor inicial $w_0 = 1$ pertenece al transitorio y queda por debajo de esos mínimos posteriores.
- **Al cambiar w_0 por $w_0/2$, cambia la amplitud y también el periodo.** El intervalo observado pasa de $[1.46, 2.65]$ a $[1.68, 2.36]$, y el periodo estimado, de 6.3295 a 6.2983. En el sistema reducido de Volterra, el periodo depende de la amplitud; para oscilaciones pequeñas tiende al de la linealización, $2\pi/\sqrt{\mu} = 2\pi$ cuando $\mu = 1$. Las dos órbitas rodean el mismo centro y tienen amplitudes pequeñas, por lo que sus periodos se parecen, pero no son iguales.

LA IDEA: UNA POBLACIÓN PEQUEÑA NO ES UNA POBLACIÓN EXACTAMENTE NULA

La gráfica deja de distinguir v del eje mucho antes de $t = 40$. La fórmula exponencial anterior explica por qué sigue siendo positiva y cuándo puede crecer: $v' > 0$ si $u + w < 1$. En la trayectoria calculada ocurre lo contrario durante la ventana estudiada. La notación exponencial muestra la escala de $v(T)$, pero por sí sola no garantiza sus cifras: valores muy inferiores a `AbsTol` requieren una comprobación adicional. Aquí el orden 10^{-35} se ha contrastado integrando la ecuación de $\log v$, que evita trabajar directamente con una cantidad tan pequeña.

Un problema inverso de epidemias (cuaderno de 2017, prácticas)

Enunciado. Se considera el problema de Cauchy relativo a un modelo de epidemias:

$$\begin{cases} I' = \beta IS - \gamma I, & t \in (0, T) \\ S' = -\beta IS + \gamma I + \alpha S \\ I(0) = I_0, \quad S(0) = S_0 \end{cases} \tag{C.33}$$

donde los datos iniciales (I_0, S_0) son desconocidos. Por el contrario, se conocen los valores de S en ciertos instantes $t_i > 0$:

$$S(t_i) = S_i, \qquad i = 1, \ldots, m. \tag{C.34}$$

Se desea calcular los valores I_0, S_0 de forma que la solución del problema (C.33) verifique (C.34). Se pide:

a) Formular este problema como un problema de mínimos cuadrados (escribir este apartado en papel y entregarlo como parte del examen de teoría).
b) Escribir una M-función

```
function [ysol] = InvEpid(tobs, sobs, Icero, Scero)
```

que calcule una aproximación `ysol` de (I_0, S_0) resolviendo el problema del apartado a), tomando `(Icero,Scero)` como aproximación inicial. El programa debe representar en una sola gráfica: (a) las trayectorias de I y S correspondientes a la solución (I_0, S_0) calculada; (b) los puntos (t_{obs}, S_{obs}) correspondientes a los datos observados; (c) la trayectoria de la población total $I + S$. El programa debe también imprimir, en la ventana de comandos o en la misma gráfica, la aproximación de la solución calculada.

Datos: $\beta = 0.0005$, $\gamma = 0.08$, $\alpha = 0.02$, `tobs` $= [10, 35, 90]$, `sobs` $= [203, 264, 178]$. Sugerencia para aproximación inicial: `Icero` $= 35$, `Scero` $= 180$.

Resolución.

Qué modelo es. No es un SIR: no hay clase de recuperados. El término $+\gamma I$ en la segunda ecuación dice que **el que se cura vuelve a ser susceptible**, así que es un **SIS**, como el de la sección 2.1.2, con una diferencia importante: el término $+\alpha S$ añade **natalidad de susceptibles**, exactamente como en el primer parcial de 2022. Y esa natalidad tiene una consecuencia que conviene tener presente al mirar las gráficas: con datos iniciales positivos, al sumar las dos ecuaciones,

$$N' = (I + S)' = \alpha S > 0,$$

se ve que la población total crece. No hay equilibrio no trivial: si $I' = 0$ con $I > 0$ hace falta $S = \gamma/\beta = 160$, y entonces $S' = \alpha \cdot 160 = 3.2 \neq 0$. El único equilibrio es el origen.

Apartado (a): la formulación de mínimos cuadrados. Para cada par $p = (I_0, S_0) \geq 0$ hay una única solución global hacia delante. El campo polinómico garantiza existencia y unicidad *locales*; para extender la solución, se observa que el cuadrante no negativo es invariante y $N' = \alpha S \leq \alpha N$, de donde $0 \leq I(t), S(t) \leq N(0)e^{\alpha t}$. No hay explosión en tiempo finito. Queda así definida la función $t \mapsto S(t; p)$. Como en general las m condiciones (C.34) no se pueden satisfacer exactamente con solo dos incógnitas, se sustituyen por la mejor aproximación cuadrática:

$$\boxed{\min_{p=(I_0,S_0)\in\mathbb{R}^2_+} \Phi(p), \qquad \Phi(p) := \sum_{i=1}^{m} \left|S(t_i; p) - S_i\right|^2.}$$

Con los datos del enunciado hay tres observaciones y dos parámetros. Eso no garantiza por sí solo un residuo positivo: las tres observaciones podrían proceder de una misma trayectoria exacta. El ajuste es de mínimos cuadrados no lineales; para evaluar Φ hay que integrar el sistema de EDO.

El programa parte de datos iniciales no negativos y asigna coste infinito a cualquier candidato con una coordenada negativa, para respetar la restricción del modelo. `fminsearch` realiza una búsqueda local sin restricciones propias: este rechazo limita las evaluaciones, pero su resultado no certifica un mínimo global. Se comprueba también su indicador de terminación.

```
function [ysol] = InvEpid(tobs, sobs, Icero, Scero)
% Nombre Apellido Apellido
% [ysol] = InvEpid(tobs, sobs, Icero, Scero)
% Problema inverso para un modelo SIS con natalidad:
```

```
% I' = beta*I*S - gamma*I
% S' = -beta*I*S + gamma*I + alpha*S
% Se buscan los datos iniciales (I0,S0) que mejor ajustan S(tobs) = sobs.
% Ejecucion pedida:
% tobs = [10 35 90]; sobs = [203 264 178];
% ysol = InvEpid(tobs, sobs, 35, 180)

beta = 0.0005; gamma = 0.08; alpha = 0.02;
T = max(tobs) + 10;  % el enunciado no da T; basta con pasar del ultimo dato
tobs = tobs(:)'; sobs = sobs(:)';

% ---------- (a) El funcional de minimos cuadrados ---------------------
if Icero < 0 || Scero < 0
   error('Los datos iniciales de partida deben ser no negativos.')
end
Phi = @(p) coste(p, tobs, sobs, T, beta, gamma, alpha);

opciones = optimset('Display', 'iter', 'TolX', 1e-10, 'TolFun', 1e-12);
[ysol, ~, estado] = fminsearch(Phi, [Icero, Scero], opciones);
if estado <= 0
   warning('La busqueda no ha confirmado convergencia; revisar el resultado.')
end

fprintf('\n Datos iniciales calculados: I0 = %10.4f, S0 = %10.4f \n', ysol(1), ysol(2))
fprintf(' Suma de cuadrados de partida: %10.4f \n', Phi([Icero, Scero]))
fprintf(' Suma de cuadrados al final: %10.4f \n', Phi(ysol))

% ---------- (b) La grafica ----------------------------------------------
close all
sol = resolver(ysol, T, beta, gamma, alpha);
t = linspace(0, T, 1000);
Y = deval(sol, t);

figure(1)
plot(t, Y(1,:), 'r-', 'LineWidth', 1.6), hold on
plot(t, Y(2,:), 'b-', 'LineWidth', 1.6)
plot(t, Y(1,:) + Y(2,:), 'k--', 'LineWidth', 1.4)
plot(tobs, sobs, 'bo', 'MarkerSize', 9, 'MarkerFaceColor', 'b')
xlabel('t (dias)'), ylabel('numero de individuos'), grid on
title(['Problema inverso: I_0 = ', num2str(ysol(1)), ', S_0 = ', num2str(ysol(2))])
legend('I(t): infectados', 'S(t): susceptibles', 'I(t)+S(t): poblacion total', ...
       'valores observados de S', 'Location', 'northwest')
shg
end

% ======================================================================
function valor = coste(p, tobs, sobs, T, beta, gamma, alpha)
if any(p < 0)
   valor = Inf;  % no integrar poblaciones iniciales negativas
   return
end
sol = resolver(p, T, beta, gamma, alpha);
valor = sum((deval(sol, tobs, 2) - sobs).^2);
end

function sol = resolver(p, T, beta, gamma, alpha)
% Resuelve el sistema con datos iniciales I(0) = p(1), S(0) = p(2).
```

```
F = @(t, y) [ beta*y(1)*y(2) - gamma*y(1);
             -beta*y(1)*y(2) + gamma*y(1) + alpha*y(2) ]; % y = [I; S]
sol = ode45(F, [0 T], [p(1); p(2)], odeset('RelTol', 1e-9, 'AbsTol', 1e-11));
end
```

Valores de referencia verificados con dos métodos numéricos independientes; no son una transcripción de una ejecución de MATLAB.

SALIDA ESPERADA

```
>> tobs = [10 35 90];
>> sobs = [203 264 178];
>> ysol = InvEpid(tobs, sobs, 35, 180)
 (fminsearch escribe aqui sus iteraciones)

 Datos iniciales calculados: I0 = 30.3446, S0 = 170.0468
 Suma de cuadrados de partida: 127.7100
 Suma de cuadrados al final: 0.3233
ysol =
   30.3446 170.0468
```

Lo que dicen esos números. La suma de cuadrados baja de 127.7 a 0.3233. La raíz del error cuadrático medio por observación, en unidades de población, es

$$\sqrt{\Phi/3} \simeq 0.3283.$$

No significa que cada residuo mida 0.33: son aproximadamente $(0.0473, -0.0013, 0.5666)$, sobre observaciones del orden de 200. El candidato calculado da un ajuste cercano, aunque no exacto. La coincidencia de dos métodos numéricos respalda estas cifras, pero no demuestra que no pueda existir otro ajuste mejor. Un residuo pequeño y no nulo tampoco es, por sí solo, un fallo del programa.

LA IDEA: POR QUÉ LOS DATOS OBSERVADOS SUBEN Y LUEGO BAJAN, Y EL MODELO TAMBIÉN

Los tres valores de `sobs` son 203, 264 y 178: S sube y luego baja. Con el modelo eso se lee sin resolverlo, pero hay que leer la ecuación entera. La segunda es

$$S' = S(\alpha - \beta I) + \gamma I,$$

y el sumando $+\gamma I$ importa: el que se cura **vuelve** a la clase de susceptibles, así que para que S baje no basta con que el contagio supere a la natalidad. Despejando,

$$S' < 0 \iff I(\beta S - \gamma) > \alpha S \iff I > \frac{\alpha S}{\beta S - \gamma} \quad (\text{si } S > \gamma/\beta),$$

un umbral que *depende de* S y que solo baja hasta $\alpha/\beta = 40$ cuando S es enorme. Con los valores ajustados, S sigue creciendo hasta $t = 36.58$, donde $S = 264.34$ y I **vale ya** 101.34: el 40 se cruza mucho antes, en $t = 15.52$, y allí todavía es $S' = +3.20$.
La otra mitad es más simple: I crece mientras $S > \gamma/\beta = 160$, y con los valores ajustados $S_0 \simeq 170.05 > 160$, así que los infectados empiezan a crecer. Además, en $S = 160$ se tiene $S' = 3.2 > 0$: esa recta no puede cruzarse hacia abajo. Por ello I sigue creciendo. En $[0, 100]$ la susceptible nunca baja de 160 —su mínimo es el propio $S_0 = 170.05$—, de modo que I crece durante toda la ventana. Los $I(100) = 452.05$ que se ven al final del dibujo **no son un máximo**, son el borde de la ventana: con $T = 200$

> se llega a 802.93 y sigue subiendo. Es lo coherente con lo dicho arriba —la población total crece sin parar y el único equilibrio es el origen—, y conviene tenerlo presente para no leer el último punto de una gráfica como si fuera un límite.

Qué se practica con estos ejercicios

- **Los de poblaciones.** Adimensionalizar, contar equilibrios de una ecuación trascendente por la forma de una función, y leer un plano de fases por sus regiones de signo. Los dos ejercicios de Gompertz permiten comparar una cuota fija (2016) con una tasa de extracción proporcional (2017), y entender por qué cambian los equilibrios.
- **Los de cálculo de variaciones.** Deducir las condiciones de optimalidad desde la perturbación, y **sobre todo**, saber cuántas condiciones de frontera hay y de dónde sale cada una: unas están impuestas y otras se deducen de la libertad de las variaciones.
- **Los de control óptimo.** Construir el estado adjunto y saber por qué es el que es. Los tres usan el mismo mecanismo de integración por partes, con distintos términos de frontera.
- **Los de mecánica.** Montar el lagrangiano: elegir las coordenadas, escribir las dos energías y aplicar Euler-Lagrange. Y reconocer una coordenada cíclica cuando la hay.
- **Los de prácticas.** Formular un problema inverso como mínimos cuadrados, programarlo sabiendo que cada evaluación es una integración, y **interpretar** los números, que es lo que se pregunta expresamente en los dos.

Errores típicos

- **En Gompertz, tomar $y = 0$ como punto crítico.** El dominio es $(0, +\infty)$: el logaritmo no está definido en cero, y el campo ni siquiera es lipschitziano al acercarse.
- **Confundir las dos extracciones.** Con cuota fija $-h$ puede haber extinción; con extracción proporcional $-hy$, todo h finito deja un equilibrio positivo en este modelo. Sin embargo, ambas políticas tienen el mismo rendimiento máximo en equilibrio, $\lambda K/e$.
- **Poner condiciones de frontera de más o de menos.** En estos problemas de primer orden, las variaciones se anulan en el extremo fijado; en el libre dan una condición natural.
- **Olvidar la derivada del coste final en la condición natural.** Con $G(x(1)) = |x(1)|^2$, la condición es $f_{x'}(1) + 2x(1) = 0$, no $f_{x'}(1) = 0$.
- **Equivocar el signo en Euler-Lagrange.** Es $\frac{d}{dt} f_{x'} - f_x = 0$; con $f = x'^2 + \operatorname{sen} x$ eso da $2x'' = \cos x$.
- **Escribir la ecuación adjunta de un sistema sin trasponer la matriz.** Es $-\hat{\varphi}' = A^t \hat{\varphi} + \ldots$, y con $A = \left(\begin{smallmatrix} 1 & -a \\ a & 1 \end{smallmatrix}\right)$ la traspuesta *no* coincide con A.
- **Poner $-\hat{\varphi}''$ en el problema de segundo orden.** Dos integraciones por partes no cambian el signo; lo que cambia es dónde van las condiciones.
- **Olvidar una de las dos condiciones terminales del adjunto** en el problema de segundo orden. Hacen falta las dos, $\hat{\varphi}(T) = \hat{\varphi}'(T) = 0$, una por cada integración por partes.
- **En mecánica, cambiar la ley del muelle al construir el potencial.** Los enunciados dan una energía proporcional al cuadrado de la distancia al anclaje: esa es la expresión que se debe usar antes de simplificar y derivar.
- **En el modelo de tres poblaciones, decir que el periodo estimado no depende del dato inicial.** Se parece mucho, pero no es igual: 6.3295 frente a 6.2983.

- **Decidir el residuo por el número de datos.** Tener más observaciones que parámetros no garantiza un residuo positivo ni nulo: depende de si esos datos son compatibles con el modelo y de si el ajuste encuentra el mínimo.

APÉNDICE D

Añadidos y omisiones

Este apéndice reúne los principales desarrollos incorporados por el libro y materiales del curso que no se han convertido en prácticas completas. Complementa las indicaciones de procedencia del cuerpo del texto. Que una explicación se desarrolle aquí no determina por sí solo si forma parte de la evaluación: las indicaciones sobre esta se recogen en el apéndice correspondiente.

Precisión en la ecuación de estado del capítulo 4. Las definiciones de solución de las páginas 88, 97 y 102 de *ApuntesMOM.pdf* exigen continuidad y la ecuación diferencial en casi todo punto. Se añade continuidad absoluta (o, de forma equivalente, la ecuación integral), necesaria para la prueba de unicidad: una función continua con derivada cero casi en todas partes puede no ser constante, como muestra la función de Cantor. Véase J. K. Hunter, *Notes on Partial Differential Equations*, sección 3.A.2, teorema 3.60 y ejemplo 3.61.

Existencia del óptimo no lineal. El teorema 4.13 de la página 102 de *ApuntesMOM.pdf* afirma existencia bajo regularidad C^1 y derivadas acotadas. Esas condiciones garantizan que cada estado exista, pero no bastan para la existencia del óptimo: falta una condición sobre la dependencia en el control. Aquí se añade la dependencia afín para el teorema de existencia y se prueba el paso al límite. La condición necesaria posterior conserva el g general, suponiendo que exista un óptimo. Sobre el papel de la convexidad en los teoremas generales, véase D. Liberzon, *Calculus of Variations and Optimal Control Theory*, sección 4.5.

Iteración no lineal para μ grande. Los ejercicios 4.5 y 4.6 (página 104 de *ApuntesMOM.pdf*) se desarrollan con la hipótesis adicional C^2 y derivadas segundas acotadas. La prueba trabaja en una bola fija que contiene a todos los óptimos y que la iteración conserva; estas condiciones justifican la aplicación del teorema del punto fijo.

Hamiltoniano y restricciones. En las páginas 106–107 de *ApuntesMOM.pdf*, se corrigen la inclusión del coste final en el hamiltoniano y la escritura del dato final: aquí $H = G + g \cdot \varphi$ y

$\varphi(T) = \nabla\Psi(y(T))$. Se distingue la minimización puntual para controles con cotas en cada instante de la desigualdad integral para un convexo general de L^2. La estacionariedad o la desigualdad variacional no equivalen a minimizar H sin una hipótesis adicional, como su convexidad en el control.

Variables biológicas en los ejemplos de control. En (4.40), Y representa células T CD4$^+$ no infectadas, como en el modelo de Joshi citado por los apuntes; no anticuerpos, que es la descripción de la página 107. En (4.44), la ecuación $x_3' = v_1$ hace de x_3 un acumulador de tratamiento, no una densidad de anticuerpos (página 108). Se explicitan además el dominio positivo del logaritmo y las condiciones usadas para conservar estados no negativos en el modelo del VIH.

Lo que el libro añade y no está en los apuntes

- La sesión cero de MATLAB entera, basada en el material de preparación del profesor y en el libro previo del autor. La sección «Resolver ecuaciones diferenciales: `ode45`», sigue el guion de EDO de ese material y resuelve sus once ejercicios, que allí van propuestos. Las funciones `fsistem.m` y `odeparam.m` también proceden de ese material. Lo añadido es la resolución comentada de los ejercicios, la comparación con la solución exacta en los casos en que se sabe (Bernoulli, Malthus, logística), la lectura de la fase inicial del modelo de epidemia, y la explicación de la relación entre los extremos de las componentes en los ejemplos de Volterra, a partir de su adimensionalización.
- En el capítulo 1: el teorema de las cinco caracterizaciones de la irreducibilidad con su demostración; el índice de imprimitividad y el criterio del máximo común divisor; la idea de la demostración de Frobenius (Wielandt); la demostración completa del teorema de Leslie y del de convergencia; la proposición de velocidad de convergencia de PageRank; los ejercicios 1.1 (Vandermonde) y 3.1 (máximos y medias) del repaso de MATLAB; y, como escalones de la exposición, la matriz de Leslie 2×2 resuelta entera a mano, la nota de Collatz-Wielandt dibujada en ese caso, la lectura de $\varphi(\lambda) = 1$ con el criterio «$\hat{\lambda} > 1$ si y solo si $\varphi(1) > 1$», una iteración de PageRank a mano con un nodo sin salida y las ecuaciones normales de la regresión, con un caso de tres puntos.
- En el capítulo 2: la prueba directa de la observación tras el teorema de las EDO autónomas; el ejercicio del carbono 14 y el café; la linealización en el modelo general de competición; las cajas de historia y de conexión; la resolución completa del caso degenerado $\alpha = \beta = \lambda = 1$ mediante la integral primera u/v (véase la lista de erratas); y, en cada modelo, la discusión de las hipótesis una a una y de lo que se pierde al simplificar, que en los apuntes va implícita. La subsección de simbiosis desarrolla la inversión del tiempo que el material original solo menciona, añade el modelo con autolimitación en la segunda especie y compara sus regiones de crecimiento y extinción. Y, como escalones de la exposición: la línea de fases de $y' = y^2 - 1$, con la explosión en tiempo finito; el tamaño final del SIR con números; el pico también con mortalidad; la adimensionalización paso a paso, empezando por la logística; y la tabla de las cuatro ramas de la competición simplificada.
- En el capítulo 3: la demostración del lema de Du Bois-Reymond y el paso que lo aplica a Euler-Lagrange sin suponer la derivabilidad del momento; la obtención de la cicloide y la justificación de su minimalidad por convexidad con $z = \sqrt{u}$; la resolución de las geodésicas, de la superficie de revolución y de los dos extremos libres; la estabilidad de

los equilibrios bajos del péndulo por Liapunov; los períodos del péndulo no lineal; la explicación de por qué la parametrización por el ángulo cubre también el tramo ascendente de la braquistócrona; y la lectura física de cada término de cada funcional; el control del impacto y de los invariantes en la práctica de dos partículas; y la cota que relaciona el residuo con el error en la viga lineal; y, como escalones, cómo se comprueba la convexidad con la hessiana, el ejemplo $x'^2 - x^2$ (sin mínimo si $T > \pi$) y el método lagrangiano resumido en cuatro pasos.

- En el capítulo 4: el ejercicio resuelto de observación final; la demostración de la contractividad para μ grande; la demostración de que un convexo cerrado es débilmente cerrado; la lectura de Euler-Lagrange como caso particular del control; la discusión de qué modela cada pieza del funcional de coste; el término b en el teorema de la EDO de segundo orden, que en los apuntes se enuncia sin él aunque el fichero `ControlOrden2.m` sí lo programe para b constante; la forma $(\varphi' + b\varphi)'$ para coeficientes variables continuos; la deducción de la función de conmutación de los dos problemas bilineales, con la positividad del estado y el dato final del adjunto como piezas del argumento; y la prueba de existencia del problema no lineal bajo una hipótesis adicional de dependencia afín respecto del control. Y, como escalones: de dónde sale la convexidad del coste, con un término final no convexo que tiene dos óptimos; el sistema de optimalidad resuelto en cerrado en el caso escalar más simple; la existencia con observación final para cualquier término final continuo y acotado inferiormente; el carrito $y'' = v$ como sistema y como EDO de segundo orden; un problema con cota en el control comparado con su recorte; el horizonte de inversión del problema bilineal con números; y el problema inverso de una tasa de crecimiento resuelto a mano.
- En todos los capítulos, los resúmenes «Lo que hay que llevarse» detrás de las demostraciones principales, «Qué se ha ganado» al final de algunas secciones y «Dónde se falla» al final de los ejercicios principales.
- Los programas desarrollados en estos apuntes para ejercicios de las hojas: `UmbralTecho.m` (tema 2), `VigaElasticaF.m`, `VigaElasticaNH.m` y `Autofun.m`, que cubre también `Autofun2.m` y `Autofun3.m` (tema 3), y `InvRicatti.m`, `InvRicatti2.m`, `InvLotka.m` y `ControlSDO.m` (tema 4).
- En el apéndice A: el lema de Gronwall con demostración, las ideas de demostración de las proposiciones sobre órbitas, y las remisiones al libro de EDO del autor, capítulo y sección, en cada resultado.
- Los apéndices de orientación, añadidos y omisiones, y el simulacro de examen. Este se distingue de las pruebas reales conservadas en el archivo.

Lo del profesor que no está en el libro

Ficheros que el profesor reparte y el libro no ha convertido en práctica, porque no están en las hojas de ejercicios: `Lorentz.m` (el atractor de Lorenz), `quasipar.m` (un péndulo forzado con solución cuasiperiódica), `TumorQuimio.m` y `TumorRadio.m` (dos modelos biomédicos con dosis fijas), `SSolarCompleto2.m` (una animación del sistema solar) y `CompeticionBis.m` (comparación de órbitas, adaptada para la figura de la práctica 13).

Demostraciones no incluidas. El libro desarrolla argumentos de competición y control que los apuntes solo esbozan, pero no demuestra todos sus resultados generales. En el apéndice A se enuncian resultados de EDO (Picard, prolongación, matriz fundamental, LaSalle y Poincaré-Bendixson) y se remite a sus demostraciones en el libro de EDO del autor.

Bibliografía: la lista de veintitrés referencias de los apuntes no se reproduce.

Erratas detectadas en los apuntes y en los ficheros

Señaladas en su sitio, con la cuenta que las corrige:

- En 2.2.1, la fórmula de $S(t)$ lleva μ por β.
- En el teorema 2.7, las cotas exponenciales son estrictas solo para $t > 0$; en $t = 0$ hay igualdad. En el análisis de competición simplificada, $\Lambda_{++}(0)$ es una gráfica creciente (aunque ambas poblaciones decrezcan con el tiempo); los límites hacia atrás deben referirse al extremo del intervalo maximal, que puede ser finito. Tampoco se cumple en general $u(t) > v(t)$ en toda la rama $\Lambda_+(C)$.
- En 2.2.5, la figura escribe αv por αu.
- En 2.2.5, $\alpha\beta = 1$ no basta para que haya infinitos equilibrios: se requiere $\alpha = \beta = 1$, pues solo entonces coinciden las dos rectas de nulidad.
- En 2.2.5, caso $\alpha = \beta = 1$: los apuntes afirman que la solución converge a $(1, 0)$ o a $(0, 1)$ según sea $u_0 > v_0$ o $v_0 > u_0$, y concluyen que también aquí una población se extingue. Con $\alpha = \beta = \lambda = 1$ y $u_0, v_0 > 0$, las dos ecuaciones comparten el factor $(1 - u - v)$: u/v es una integral primera y cada semirrecta desde el origen es invariante. La solución converge a $(u_0, v_0)/(u_0 + v_0)$, un punto interior del segmento $\bar{u} + \bar{v} = 1$; ninguna población se extingue.
- En 3.2, la braquistócrona usa un eje vertical hacia abajo: la energía potencial debe ser $-mgu$, no $+mgu$. La condición $u \in C^1([0, \alpha])$ excluye la tangente vertical inicial de la cicloide; se permite esa singularidad y se toma una integral impropia. La identidad de Beltrami se aplica en el interior, donde $u > 0$.
- En 3.3, el principio de Hamilton exige acción estacionaria, no necesariamente mínima. La función $\mathcal{L}$ es la lagrangiana; la acción es su integral. El muelle muestra en el cuerpo del libro una trayectoria estacionaria que no es un mínimo.
- En 3.3.2, la afirmación de que todas las órbitas son periódicas debe excluir las separatrices: se acercan al equilibrio invertido en tiempo infinito. En el nivel crítico, una órbita no recorre todos los ángulos aunque el conjunto de nivel sí se proyecte sobre todo el eje.
- En 3.3.3, la energía cinética de rotación de la varilla es $\frac{1}{3}\rho\ell^3|\varphi'|^2$ y no $\frac{2}{3}$.
- En la hoja del tema 2, ejercicio 8, el dato inicial del enunciado ($2 \cdot 10^3$) y el de la figura ($2 \cdot 10^5$) no coinciden; el libro resuelve los dos. Además, si $0 < N_0 < N_1$, la prolongación negativa explota en tiempo finito: el límite escrito como $t \to +\infty$ debe referirse al extremo finito del intervalo maximal. La población solo se interpreta hasta su primer cero.
- En la hoja del tema 2, ejercicio 11, la figura de la solución usa $P_0 = N_0 = 2.2$, mientras que el enunciado pide $P_0 = 2$ y $N_0 = 0.5$ para el caso sin insecticida. El libro representa los datos del enunciado.
- En `DosParticulas.m`, el último nodo de altura positiva no es el instante de contacto con el suelo: la versión del libro lo localiza mediante un evento. En `VigaElastica.m`, el eje vertical fijo recorta algunas llamadas de prueba; el libro permite que se ajuste a los datos.
- En las prácticas 16 y 17, «amortiguamiento» designa fuerzas proporcionales al desplazamiento. Los funcionales escritos describen restitución elástica, no amortiguación viscosa; se aclara esa diferencia en el texto y en las figuras.
- En la hoja del tema 3, ejercicio 18, y en `VibForzadas.m`, el coeficiente resonante es $A/(2\omega)$ y no $A/2$.

- En la práctica 19, las potencias r^4 y r^2 de la ecuación corresponden a un radio $\sqrt{\varepsilon}\, r(x)$: la sección debe escribirse con $\varepsilon r(x)^2$, no con $\varepsilon r(x)$. Se recupera además la distinción entre el desplazamiento físico $w = u/\varepsilon$ y la incógnita escalada u, y se explicita la densidad en k.
- En `Bratu.m`, los intervalos separados por 1.2 fallan cerca del crítico, y el extremo 5 no alcanza la raíz grande para ciertos λ pequeños. El libro calcula el máximo, amplía el intervalo y comprueba que las dos ramas son distintas.
- En 4.5, la última línea de (4.30) mezcla dos igualdades.
- En 4.7, la expresión (4.38) tiene un `dfrac` sin compilar y el dato final del adjunto debe ser $\nabla\Psi(\hat{y}(T))$, no $\partial H/\partial y$.
- En 4.8.2, los datos iniciales dicen $1 \leq i \leq 4$ para un sistema de tres componentes.
- En `ControlEDOr.m`, el signo de la proyección en el sistema de contorno no coincide con (4.35).
- `ControlOrden2.m` integra el sistema de optimalidad desde $t = 0$ con $\varphi(0) = \varphi'(0) = 0$ en vez de resolver el problema de contorno.
- En A.19 (página 119), la cota de estabilidad exponencial debe incluir $|y_0 - \bar{y}|$, como en el libro de EDO, §6.3.
- La lista de comportamientos tras A.41 (páginas 123–124) omite conexiones homoclínicas y heteroclínicas. El teorema para un compacto sin puntos críticos se conserva.

APÉNDICE E

Cómo usar este libro

El libro combina modelos, explicaciones intuitivas, demostraciones y prácticas. Este apéndice explica cómo orientarse entre esos materiales.

Las referencias a «el profesor» identifican los materiales de la asignatura de Modelización Matemática de la Universidad de Sevilla. Las indicaciones específicas de su evaluación están reunidas en el apéndice B.

Para quién es y desde dónde parte

Se suponen conocimientos de Álgebra Lineal, Análisis y Ecuaciones Diferenciales Ordinarias. El apéndice A recuerda los resultados de EDO utilizados y da referencias para ampliar. No hace falta disponer de los otros libros de la colección para seguir el desarrollo principal.

La sesión cero introduce MATLAB. Las prácticas ofrecen también versiones en Python para estudiar los métodos y comparar resultados fuera del contexto del examen. Conviene ejecutar los programas con los datos indicados y comprobar lo que representa cada salida antes de modificar el modelo.

La estructura

Los cuatro capítulos siguen los temas y el orden de los apuntes de la asignatura. Cada capítulo reúne al final las prácticas correspondientes. El apéndice A contiene resultados auxiliares de EDO; los siguientes reúnen orientación para el examen, exámenes resueltos, añadidos y omisiones, y esta guía.

Los desarrollos adicionales y las correcciones se explican junto al contenido afectado y se recogen en el apéndice D.

El código visual

Los bloques ayudan a distinguir la función de cada pasaje.

- **Ciruela, «La pregunta».** Plantea el problema que se va a estudiar. Conviene intentar una respuesta antes de continuar.
- **Azul.** Desarrolla una demostración.
- **Ocre, «La idea».** Explica la intuición y cómo se relaciona con el argumento formal.
- **Verde, «Antes de leer» y «Repaso intercalado».** Propone preguntas para recuperar lo aprendido y relacionar temas.
- **Gris, «De dónde viene», «Conexión» y «El camino que no se tomó».** Ofrece historia, conexiones y alternativas de modelado.
- **Cursiva entre comillas, sin filete.** Conserva intervenciones del autor en sus notas anteriores; las aclaraciones aparecen después.
- **Listados de código.** Azul para MATLAB, verde para Python y gris para salidas de referencia. Hay programas completos y fragmentos que se usan dentro del ejemplo correspondiente.

Intercalados en el texto hay avisos breves en cursiva verde que empiezan por EXPLÍCATE. Son el punto exacto donde conviene parar, cerrar el libro y decir en voz alta la respuesta antes de seguir.

Fuera de los bloques, unos pocos rótulos en negrita se repiten con la misma función en todo el libro. «El caso más pequeño, a mano» o «Con números» resuelve un ejemplo concreto, con los valores puestos, antes o después del resultado general. «Lo que hay que llevarse», detrás de una demostración, la resume en tres o cuatro frases con las que se puede reconstruir. «Qué se ha ganado» cierra una sección diciendo qué pregunta ha quedado contestada. Y «Dónde se falla», al final de un ejercicio, reúne los errores que se cometen al resolverlo y el síntoma que delata cada uno.

Los ejercicios

Los ejercicios aparecen junto al contenido que utilizan. Cuando hay pistas escalonadas, conviene intentar primero el problema y consultar solo la ayuda necesaria antes de leer la resolución. Después puede repetirse sin mirar, explicando qué resultado permite cada paso y por qué se cumplen sus hipótesis.

Las prácticas incluyen salidas de referencia. Para comparar cifras, deben coincidir el modelo, los parámetros, el dato inicial y el intervalo; el método numérico y sus tolerancias también influyen en el resultado.

Qué hacer si algo no cuadra

Si un paso no se entiende, conviene identificar qué hipótesis o resultado lo justifica y probarlo en un ejemplo sencillo. Una discrepancia se anota con la sección, los datos utilizados y la cuenta o salida que permite reproducirla.