

# Relación I Métodos Numéricos I

Javier Gómez López

2020/2021

**Ejercicio 1.** Comprueba que si  $N \in \mathbb{N}$  y  $1 \leq p < \infty$ , entonces la aplicación  $\|\cdot\|_p : \mathbb{R}^N \rightarrow \mathbb{R}$  definida en cada  $\mathbf{x} \in \mathbb{R}^N$  como

$$\|\mathbf{x}\|_p := \left( \sum_{j=1}^N |x_j|^p \right)^{1/p}$$

es una norma en dicho espacio vectorial.

Tal y como se afirma en la indicación del enunciado, el caso  $p = 1$  es fácil de comprobar (es tratar una consecuencia inmediata de las propiedades del valor absoluto) y para  $p > 1$ , la única de las tres propiedades que definen a una norma que no es trivial en este caso es la desigualdad triangular, por lo que nos centramos en su prueba. Siguiendo la sugerencia del enunciado, seguimos tres pasos. Para ello, notamos por  $p'$  al único real tal que  $\frac{1}{p} + \frac{1}{p'} = 1$

## Paso 1

$$x, y \geq 0 \Rightarrow xy \leq \frac{x^p}{p} + \frac{y^{p'}}{p'}$$

Con el fin de probar esta desigualdad, sea  $y \geq 0$  y consideremos la función  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  definida en cada  $x \geq 0$  por

$$f(x) := \frac{x^p}{p} + \frac{y^{p'}}{p'} - xy$$

y estudiemos su monotonía.

$$f'(x) = 0 \iff x^{p-1} - y = 0 \iff x = y^{\frac{1}{p-1}}$$

Pero  $f''(x) = (p-1)x^{p-2}$ , por lo que  $f$  es convexa, y por tanto,  $f$  alcanza su valor mínimo global en  $x = y^{\frac{1}{p-1}}$ , que es

$$\begin{aligned} f\left(y^{\frac{1}{p-1}}\right) &= \frac{y^{\frac{p}{p-1}}}{p} + \frac{y^{p'}}{p'} - y^{\frac{1}{p-1}}y = \frac{y^{\frac{p}{p-1}}}{p} + \frac{y^{p'}}{p'} - y^{\frac{p}{p-1}} = \\ &= y^{\frac{p}{p-1}} \left( \frac{1}{p} - 1 \right) + \frac{y^{\frac{p}{p-1}}}{p'} = 0 \end{aligned}$$

Así:

$$x \geq 0 \Rightarrow 0 \leq f(x)$$

## Paso 2

Desigualdad de Hölder: dados  $x_1, \dots, x_N, y_1, \dots, y_N \in \mathbb{R}$ ,

$$\sum_{j=1}^N |x_j y_j| \leq \|x\|_p \|y\|_{p'}$$

En efecto: si  $x = 0$  o  $y = 0$ , la desigualdad es obvia. Si  $\|x\|_p = 1 = \|y\|_{p'}$ ,

$$\sum_{j=1}^N |x_j y_j| \leq \sum_{j=1}^N \frac{|x_j|^p}{p} + \sum_{j=1}^N \frac{|y_j|^{p'}}{p'} = \frac{1}{p} + \frac{1}{p'} = 1 = \|x\|_p \|y\|_{p'}$$

Y en general, si  $x \neq 0 \neq y$ , basta aplicar esto último a los escalares  $\frac{x_j}{\|x\|_p}$  e  $\frac{y_j}{\|y\|_{p'}}$ ,  $j = 1, \dots, N$ .

## Paso 3

Desigualdad triangular:

$$\begin{aligned} \|x + y\|_p^p &= \sum_{j=1}^N |x_j + y_j|^p \leq \sum_{j=1}^N |x_j| |x_j + y_j|^{p-1} + \sum_{j=1}^N |y_j| |x_j + y_j|^{p-1} \leq \\ &\leq (\|x\|_p + \|y\|_p) \left( \sum_{j=1}^N |x_j + y_j|^{(p-1)p'} \right)^{\frac{1}{p'}} = (\|x\|_p + \|y\|_p) \|x + y\|_p^{\frac{p}{p'}} \end{aligned}$$

La desigualdad triangular se sigue del principio y el fin de esta cadena de desigualdades y el hecho de ser  $p - \frac{p}{p'} = 1$ .

Ahora probemos las otras dos propiedades que definen una norma:

- $\|x\| \geq 0$  y  $\|x\| = 0 \iff x = 0$ .

$$\begin{aligned} \|x\|_p &= \left( \sum_{j=1}^N |x_j|^p \right)^{1/p} = \sqrt[p]{\sum_{j=1}^N |x_j|^p} \\ \sqrt[p]{\sum_{j=1}^N |x_j|^p} &= 0 \iff \sum_{j=1}^N |x_j|^p = 0 \iff x = 0 \end{aligned}$$

Y podemos afirmar que:

$$\|x\| \geq 0 \quad \|x\| = 0 \implies x = 0$$

- $\lambda \in \mathbb{R} \quad \|\lambda x\| = |\lambda| \|x\|$ .

$$\|\lambda x\|_p = \left( \sum_{j=1}^N |\lambda x_j|^p \right)^{1/p} = \left( |\lambda|^p \sum_{j=1}^N |x_j|^p \right)^{1/p} = |\lambda| \left( \sum_{j=1}^N |x_j|^p \right)^{1/p} = |\lambda| \|x\|$$

Y queda probado.

**Ejercicio 2.** *Símbolos de Landau* Sean  $N \in \mathbb{N}$ ,  $A$  un subconjunto no vacío de  $\mathbb{R}^N$ ,  $f, g : A \rightarrow \mathbb{R}$  y supongamos que  $\mathbf{x}_0 \in \mathbb{R}^N$  es un punto de acumulación de  $A$  y que existe  $\gamma > 0$  de forma que

$$0 < \|\mathbf{x} - \mathbf{x}_0\| < \gamma \mid \mathbf{x} \in A \Rightarrow g(\mathbf{x}) \neq 0.$$

- Se dice que  $f$  es de orden *O grande* (u *O mayúscula*) respecto a  $g$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$  si

$$\text{existen } M > 0 \text{ y } 0 < \delta < \gamma : \quad 0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta \mid \mathbf{x} \in A \Rightarrow \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} \right| < M,$$

hecho que se nota como  $f(\mathbf{x}) = O(g(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$ .

- La función  $f$  es de orden *o pequeña* (u *o minúscula*) respecto a  $g$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$  si

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f(\mathbf{x})}{g(\mathbf{x})} = 0,$$

El papel de  $\mathbf{x}_0$  puede ser reemplazado, cuando  $N = 1$ , por  $+\infty$  (respectivamente  $-\infty$ ) si  $A$  no está acotado superiormente (respectivamente inferiormente).

Comprueba  $(f_1, f_2, g_1, g_2 : A \rightarrow \mathbb{R})$  y  $g_1$  y  $g_2$  satisfacen la misma hipótesis que  $g$  de nulidad en un entorno de  $\mathbf{x}_0$ :

- $f(\mathbf{x}) = o(g(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0 \Rightarrow f(\mathbf{x}) = O(g(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$ .

Suponemos  $f(\mathbf{x}) = o(g(\mathbf{x}))$  cuando  $x \rightarrow \mathbf{x}_0$ .

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f(\mathbf{x})}{g(\mathbf{x})} = 0, \text{ sea } h(\mathbf{x}) = \frac{f(\mathbf{x})}{g(\mathbf{x})} \Rightarrow \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} h(\mathbf{x}) = L = 0$$

Recordemos que la caracterización de la convergencia de una sucesión en  $\mathbb{R}^N$ :

$$\forall \varepsilon \in \mathbb{R}^+ \quad \exists \gamma \in \mathbb{R}^+ : \quad \left. \begin{array}{l} 0 < \|\mathbf{x} - \mathbf{x}_0\| < \gamma \\ \mathbf{x} \in \mathbb{R}^N \end{array} \right\} |h(\mathbf{x}) - L| < \varepsilon$$

De lo que deducimos que:

$$\exists M > 0 \text{ y } 0 < \delta < \gamma : \quad \left. \begin{array}{l} 0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta \\ \mathbf{x} \in A \end{array} \right\} \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} \right| < M \iff f(\mathbf{x}) = O(g(\mathbf{x})) \text{ cuando } \mathbf{x} \rightarrow \mathbf{x}_0$$

- $f(\mathbf{x}) = O(g_1(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$  y  $g_1(\mathbf{x}) = O(g_2(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0 \Rightarrow f(\mathbf{x}) = O(g_2(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$  (Ídem para o).

De la primera condición, obtenemos que:

$$\exists M_1 > 0 \text{ y } 0 < \delta_1 < \gamma_1 : \quad \left. \begin{array}{l} 0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta_1 \\ \mathbf{x} \in A \end{array} \right\} \left| \frac{f(\mathbf{x})}{g_1(\mathbf{x})} \right| < M_1 \Rightarrow |g_1(\mathbf{x})| > \frac{f(\mathbf{x})}{M_1}$$

$$\exists M_2 > 0 \text{ y } 0 < \delta_2 < \gamma_2 : \quad \left. \begin{array}{l} 0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta_2 \\ \mathbf{x} \in A \end{array} \right\} \left| \frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} \right| < M_2$$

Por tanto:

$$\frac{\frac{|f(\mathbf{x})|}{M_1}}{|g_2(\mathbf{x})|} < \left| \frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} \right| < M_2 \Rightarrow \left| \frac{f(\mathbf{x})}{g_2(\mathbf{x})} \right| < M_2 \cdot M_1$$

lo que significa

$$\left. \begin{array}{l} \exists M_3 = M_1 \cdot M_2 > 0 \text{ y } 0 < \delta_3 < \gamma_3 : \\ \mathbf{x} \in A \end{array} \right\} \left| \frac{f(\mathbf{x})}{g_2(\mathbf{x})} \right| < M_3$$

y por tanto,  $f(\mathbf{x}) = O(g_2(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$ .

Veamos ahora el caso para *o minúscula*:

De la condición inicial obtenemos que:

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f(\mathbf{x})}{g_1(\mathbf{x})} = 0 \quad \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} = 0$$

Ambos límites existen y podemos multiplicarlos:

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f(\mathbf{x})}{g_1(\mathbf{x})} \cdot \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} = 0 = \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f(\mathbf{x})}{g_1(\mathbf{x})} \cdot \frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} = \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f(\mathbf{x})}{g_2(\mathbf{x})} = 0 \iff f(\mathbf{x}) = o(g_2(\mathbf{x})) \text{ cuando } \mathbf{x} \rightarrow \mathbf{x}_0$$

- $f_1(\mathbf{x}) = O(g_1(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$  y  $f_2(\mathbf{x}) = O(g_2(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0 \Rightarrow f_1(\mathbf{x})f_2(\mathbf{x}) = O(g_1(\mathbf{x})g_2(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{x}_0$  (Ídem para *o*).

Por hipótesis:

$$\left. \begin{array}{l} \exists M_1 > 0 \text{ y } 0 < \delta_1 < \gamma_1 : \\ \mathbf{x} \in A \end{array} \right\} \left| \frac{f_1(\mathbf{x})}{g_1(\mathbf{x})} \right| < M_1$$

$$\left. \begin{array}{l} \exists M_2 > 0 \text{ y } 0 < \delta_2 < \gamma_2 : \\ \mathbf{x} \in A \end{array} \right\} \left| \frac{f_2(\mathbf{x})}{g_2(\mathbf{x})} \right| < M_2$$

Si multiplicamos ambas desigualdades:

$$\left| \frac{f_1(\mathbf{x})}{g_1(\mathbf{x})} \right| \left| \frac{f_2(\mathbf{x})}{g_2(\mathbf{x})} \right| = \left| \frac{f_1(\mathbf{x})f_2(\mathbf{x})}{g_1(\mathbf{x})g_2(\mathbf{x})} \right| < M_1 \cdot M_2 > 0 \Rightarrow$$

$$\left. \begin{array}{l} \exists M_3 = M_1 \cdot M_2 > 0 \text{ y } 0 < \delta_3 < \gamma_3 : \\ \mathbf{x} \in A \end{array} \right\} \left| \frac{f_1(\mathbf{x})f_2(\mathbf{x})}{g_1(\mathbf{x})g_2(\mathbf{x})} \right| < M_3 \Rightarrow$$

$$f_1(\mathbf{x})f_2(\mathbf{x}) = O(g_1(\mathbf{x})g_2(\mathbf{x})) \text{ cuando } \mathbf{x} \rightarrow \mathbf{x}_0$$

Veamos ahora el caso para *o minúscula*:

Por hipótesis:

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f_1(\mathbf{x})}{g_1(\mathbf{x})} = 0 \quad \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f_2(\mathbf{x})}{g_2(\mathbf{x})} = 0$$

Si multiplicamos y aplicamos propiedades de límites obtenemos que:

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f_1(\mathbf{x})}{g_1(\mathbf{x})} \cdot \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f_2(\mathbf{x})}{g_2(\mathbf{x})} = 0 = \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f_1(\mathbf{x})f_2(\mathbf{x})}{g_1(\mathbf{x})g_2(\mathbf{x})} = 0 \Rightarrow f_1(\mathbf{x})f_2(\mathbf{x}) = o(g_1(\mathbf{x})g_2(\mathbf{x})) \text{ cuando } \mathbf{x} \rightarrow \mathbf{x}_0$$

- Dados  $a, b, c \in \mathbb{R}$ ,  $an^2 + bn + c = O(n^2)$  cuando  $n \rightarrow \infty$ .

Sean  $f(n) = an^2 + bn + c$  y  $g(n) = n^2$  y comprobemos que

$$\left. \begin{array}{l} \exists M > 0 \text{ y } 0 < \delta < \gamma : \\ n \in \mathbb{R} \end{array} \right\} \left| \frac{f(n)}{g(n)} \right| < M$$

\* Cambiamos la notación,  $n \rightarrow \infty$  con errores menores que  $\delta$ .

Calculemos ahora el límite:

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \lim_{n \rightarrow \infty} \frac{an^2 + bn + c}{n^2} = \frac{\infty}{\infty} \Rightarrow \lim_{n \rightarrow \infty} \frac{a + \frac{b}{n} + \frac{c}{n^2}}{1} = a \Rightarrow \lim_{n \rightarrow \infty} \left| \frac{f(n)}{g(n)} \right| = |a|$$

Luego, tomando  $M = |a| + 1$  se cumple que  $f(n) = O(g(n))$  cuando  $n \rightarrow \infty$

- Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  es una función polinómica, entonces  $f(x) = o(e^x)$  cuando  $x \rightarrow \infty$ .

Desarrollemos el límite:

$$\lim_{x \rightarrow \infty} \frac{f(x)}{e^x} = \frac{\infty}{\infty} \Rightarrow \lim_{x \rightarrow \infty} \frac{a_n + \frac{a_{n-1}}{x} + \dots + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}}{e^x} = \frac{a_n}{\infty} = 0$$

y queda probado.

**Ejercicio 3.** Calcula el radio espectral de la matriz

$$\mathbf{A} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 0 \end{bmatrix}$$

¿Qué podemos afirmar sobre la sucesión  $\{\mathbf{A}^n\}_{n \geq 1}$ ?

Primero calcularemos su radio espectral:

(% i1) A: matrix([1/2,1/2],[1/4,0]);

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & 0 \end{pmatrix} \quad (\text{A})$$

(% i2) float(apply(max,abs(eigenvalues(A)[1])));

$$0.6830127018922193 \quad (\% \text{ o2})$$

(% i3) apply(max,abs(eigenvalues(A)[1]));

$$\frac{\sqrt{3} + 1}{4} \quad (\% \text{ o3})$$

Así:

$$\rho(\mathbf{A}) = 0.683$$

Existe un teorema que dice:

$$\lim_{n \rightarrow \infty} \mathbf{A}^n = 0 \iff \rho(\mathbf{A}) < 1 \quad (1)$$

Por tanto podemos afirmar que la sucesión  $\{\mathbf{A}^n\}_{n \geq 1}$  converge a 0.

**Ejercicio 4.** Encuentra una matriz  $A \in \mathbb{R}^{3 \times 3}$  y una norma matricial  $\|\cdot\|$  en  $\mathbb{R}^{3 \times 3}$  de forma que  $\rho(\mathbf{A}) < 1$  pero  $\|\mathbf{A}\| \geq 1$ .

Tomemos la siguiente matriz:

$$\mathbf{A} = \begin{bmatrix} 1/2 & 1 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Observamos que es una matriz triangular y por tanto sus valores propios son los elementos de su diagonal. Recordemos que

$$\rho(\mathbf{A}) := \max\{|\lambda| : \lambda \in \mathbb{C} \wedge \det(\mathbf{A} - \lambda \mathbf{I}) = 0\}$$

es decir, el radio espectral de una matriz es el máximo de los valores absolutos de los valores propios de dicha matriz. Por tanto, podemos afirmar que

$$\rho(\mathbf{A}) \frac{1}{2} < 1$$

Definamos ahora

$$\|\mathbf{A}\|_{\infty} = \max_{1 \leq i \leq m} \sum_{j=1}^N |a_{ij}| \text{ en } \mathbb{R}^{N \times M}$$

y por tanto

$$\|\mathbf{A}\|_{\infty} = \frac{3}{2} \geq 1$$

**Ejercicio 5.** Demuestra que toda función real definida en un intervalo y de clase  $C^1$  es estable, pero que el recíproco no es cierto. Comprueba además que la función  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  definida en cada  $x \in \mathbb{R}_+$  como

$$f(x) := \sqrt{x}$$

no es estable en cero. ¿Podemos asegurar que toda función real definida en un intervalo que sea estable en todo su dominio es de clase  $C^1$ ? Justifica tu respuesta.

Primero probemos que si  $f : \mathbb{R} \rightarrow \mathbb{R}$  de clase  $C^1 \Rightarrow f$  es estable.

Recordemos que  $y_0 \in \mathbb{R}$ ,  $f$  estable en  $y_0$  si:

$$\exists M, \delta > 0 : \sup_{0 < |y - y_0| < \delta} \frac{|f(y) - f(y_0)|}{|y - y_0|} < M$$

Recordemos también el Teorema del Valor Medio:  $f : [a, b] \rightarrow \mathbb{R}$  y derivable en  $]a, b[ \Rightarrow \exists c \in ]a, b[ :$

$$f(b) - f(a) = f'(c)(b - a)$$

Así, sea  $y_0 \in \mathbb{R}$ , elijo cualquier  $\delta > 0$ . Entonces, sea

$$M := \max\{|f'(x)| : x \in [y_0 - \delta, y_0 + \delta]\}$$

el cual es finito por ser  $f'$  continua. Por tanto,

$$0 < |y - y_0| < \delta \Rightarrow \frac{|f(y) - f(y_0)|}{|y - y_0|} = |f'(c)| \leq M < M + 1$$

Así pues,  $f$  es estable en  $y_0$ , y como este es arbitrario,  $f$  es estable.

Probemos ahora que el recíproco no es cierto, es decir, que  $f : \mathbb{R} \rightarrow \mathbb{R}$  de clase  $C^1$   $\nRightarrow$   $f$  estable.

Contraejemplo:

$$f(x) = |x|, \quad f \in C^1(\mathbb{R})$$

Por la desigualdad triangular:

$$|\alpha + \beta| \leq |\alpha| + |\beta| \Rightarrow ||\alpha| - |\beta|| \leq |\alpha - \beta|$$

Tomando cualquier  $\delta > 0$ , tenemos que:

$$\sup \frac{||y| - |y_0||}{|y - y_0|} \geq \sup \frac{||y| - |y_0||}{||y| - |y_0||} = 1$$

y queda probado.

Probemos ahora que  $f(x) = \sqrt{x}$  no es estable en 0. Sea  $y_0 = 0$ , así:

$$\frac{|\sqrt{y} - \sqrt{y_0}|}{|y - y_0|} = \frac{|\sqrt{y} - 0|}{|y - 0|} = \frac{|\sqrt{y}|}{|y|}$$

Tenemos que

$$\lim_{y \rightarrow 0^+} \frac{|\sqrt{y}|}{|y|} = \frac{0}{0} \stackrel{\text{L'H}}{=} \lim_{y \rightarrow 0^+} \frac{\frac{1}{2\sqrt{y}}}{1} = \frac{1}{0} = \infty$$

Y observamos que no es estable, pues el conjunto  $\left\{ \frac{|f(y) - f(y_0)|}{|y - y_0|} : y \in [y_0 - \delta, y_0 + \delta] \right\}$  no está mayorado.

**Ejercicio 6.** Decide en función del parámetro  $a \in \mathbb{R}$  si el problema:

$$\text{dado } \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^2 : \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

está bien planteado. Para los valores de  $a$  que hagan que dicho problema esté bien planteado, estudia su condicionamiento en  $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1.1 \\ 0.22 \end{bmatrix}$  y el de la matriz de coeficientes, referidos ambos a una conveniente norma que elijas.

Las dos condiciones para que un problema esté bien planteado son:

- Unisolvente  $\Rightarrow \det(A) \neq 0$

$$\det(A) = 1 - a^2 = 0 \iff a = \pm 1$$

- Estabilidad

$$f(x) = Ax \quad g = f^{-1} : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \quad y \rightarrow g(y) = A^{-1}y$$

$$||g(y) - g(\bar{y})||_1 = ||A^{-1}y - A^{-1}\bar{y}||_1 = ||A^{-1}(y - \bar{y})||_1 \leq ||A^{-1}||_1 ||y - \bar{y}||_1$$

Es decir:

$$y \neq \bar{y} \Rightarrow \frac{||g(y) - g(\bar{y})||_\infty}{||y - \bar{y}||_\infty} \leq ||A^{-1}||_1$$

En cuanto al condicionamiento:

$$c(g, y_0) = \frac{\|A^{-1}\|_1 \|y_0\|_1}{\|x_0\|_1}, \text{ donde } Ax_0 = y_0$$

Además, como el problema está asociado a un sistema de ecuaciones lineales, vamos a hallar también el condicionamiento de su matriz de coeficientes (usaremos la norma 1 para facilidad de cálculos),  $A$ .

$$c(A) = \|A\|_1 \|A^{-1}\|_1$$

$$A^{-1} = \frac{1}{1-a^2} \cdot \begin{bmatrix} 1 & -a \\ -a & 1 \end{bmatrix}$$

$$\begin{aligned} c(g, y_0) &= \frac{\|A^{-1}\|_1 \|y_0\|_1}{\|x_0\|_1} = \frac{\|A^{-1}\|_1 \|y_0\|_1}{\|A^{-1} \cdot y_0\|_1} = \frac{\left( \left| \frac{1}{1-a^2} \right| + \left| \frac{-a}{1-a^2} \right| \right) \cdot (|1, 1| + |0.22|)}{\left| \frac{1.1-0.22a}{1-a^2} \right| + \left| \frac{0.22-1.1a}{1-a^2} \right|} = \\ &= \frac{\frac{1+|a|}{|1-a^2|} \cdot 1.32}{\frac{1.1-0.22a+0.22-1.1a}{|1-a^2|}} = \frac{1.32 \cdot (1+|a|)}{0.22 \cdot (|5-a| + |1-5a|)} = 6 \cdot \frac{1+|a|}{|5-a| + |5a-1|} \Rightarrow \text{buen condicionamiento} \end{aligned}$$

Sin embargo:

$$c(A) = (1+|a|) \frac{(1+|a|)}{|1-a^2|} = \frac{1+|a|}{|1-|a||}$$

Y por tanto, cuando  $|a| \rightarrow 1$ ,  $c(A) \rightarrow \infty$ , mal condicionamiento.

**Ejercicio 7.** Sean  $N \in \mathbb{N}$  y  $\|\cdot\|$  una norma matricial en  $\mathbb{R}^{N \times N}$  inducida por una norma en  $\mathbb{R}^N$ . Prueba que

$$\min\{c(\mathbf{A}) : \mathbf{A} \in \mathbb{R}^{N \times N} \text{ regular}\} = 1,$$

y que si  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N \times N}$  son matrices regulares, entonces

$$c(\mathbf{AB}) \leq c(\mathbf{A})c(\mathbf{B})$$

Sea  $I_N$  la matriz identidad de orden  $N$  y recordemos que todas las normas de un espacio normado son equivalentes:

$$I_N = A^{-1}A \Rightarrow 1 = \|I\| = \|A^{-1}A\| \leq \|A^{-1}\| \cdot \|A\| = c(\mathbf{A})$$

De forma análoga:

$$\begin{aligned} c(\mathbf{AB}) &= \|AB\| \cdot \|(AB)^{-1}\| = \|AB\| \cdot \|B^{-1}A^{-1}\| \leq (\|A\| \cdot \|B\|) \cdot (\|B^{-1}\| \cdot \|A^{-1}\|) = \\ &= (\|A\| \cdot \|A^{-1}\|) \cdot (\|B\| \cdot \|B^{-1}\|) = c(\mathbf{A})c(\mathbf{B}) \end{aligned}$$

Y queda probado que  $c(\mathbf{AB}) \leq c(\mathbf{A})c(\mathbf{B})$ .

**Ejercicio 8.** Estudia el condicionamiento de las siguientes funciones en todos los puntos de su dominio y discute los puntos de mal condicionamiento que eventualmente puedan existir:



- $f(x) = e^x \sin x, \quad (0 < x < \pi).$

$$c(f, y_0) = \left| \frac{f'(y_0)y_0}{f(y_0)} \right| = \frac{e^{y_0}(\sin(y_0) + \cos(y_0))y_0}{e^{y_0} \sin(y_0)} = \frac{(\sin(x) + \cos(x))y_0}{\sin(y_0)} = y_0 \left( 1 + \frac{\cos(y_0)}{\sin(y_0)} \right)$$

Observamos que el condicionamiento en los únicos puntos en los que se dispara es cuando  $\sin(y_0) = 0$ , es decir,  $y_0 = 0$  o  $y_0 = \pi$ . Concluimos que  $f(x)$  tiene un buen condicionamiento en todo su dominio excepto alrededor de  $0, \pi$ .

- $f(x) = 2 - 4 \cos(x), \quad (-\pi/2 < x < \pi/2).$

$$c(f, y_0) = \left| \frac{f'(y_0)y_0}{f(y_0)} \right| = \left| \frac{4 \sin(y_0)y_0}{2 - 4 \cos(y_0)} \right| = \left| y_0 \frac{2 \sin(y_0)}{1 - 2 \cos(y_0)} \right|$$

Observamos que los únicos valores que harían que el valor del condicionamiento se disparase son aquellos que anulan el denominador de la fracción:

$$1 - 2 \cos(y_0) = 0 \iff \cos(y_0) = \frac{1}{2} \iff y_0 = \pm \frac{\pi}{3}$$

Es decir, los puntos de mal condicionamiento de la función son los próximos a  $\pm \frac{\pi}{3}$ . Estudiamos los límites:

$$\lim_{x \rightarrow \frac{\pi}{3}} \frac{2|x| |\sin(x)|}{|1 - 2 \cos(x)|} = \frac{1}{0^+} \cdot \left( 2 \cdot \frac{\pi}{3} \sin\left(\frac{\pi}{3}\right) \right) = +\infty \Rightarrow \text{mal condicionamiento}$$

- $f(x) = \log \sqrt{x}, \quad (x > 0).$

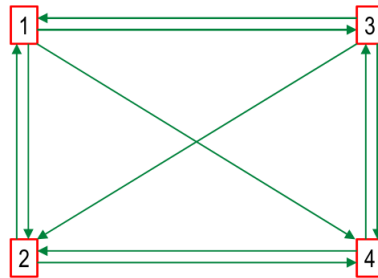
$$c(f, y_0) = \left| \frac{f'(y_0)y_0}{f(y_0)} \right| = \left| \frac{\frac{1}{2y_0}y_0}{\sqrt{y_0}} \right| = \left| \frac{1}{2\sqrt{y_0}} \right|$$

Observamos que el condicionamiento se dispara cuando el denominador tiende a 0, es decir,

$$2\sqrt{y_0} \longrightarrow 0 \iff y_0 \longrightarrow 0$$

Concluimos que la función presenta un mal condicionamiento cuando tiende a 0.

**Ejercicio 9.** Modeliza matemáticamente la relevancia de cada una de las 4 páginas web que aparecen en la figura, de acuerdo con el algoritmo PageRank de Google y teniendo en cuenta la estructura de enlaces mutuos.



Usando la teoría vista en clase, podemos establecer la relevancia de cada página  $x_i$  como la suma de la relevancia de las páginas que establecen un enlace con  $i$  dividido por el número de enlaces que salen de ellas. De esta manera:

$$x_1 = \frac{x_2}{2} + \frac{x_3}{3}, \quad x_2 = \frac{x_1}{3} + \frac{x_3}{3} + \frac{x_4}{2}, \quad x_3 = \frac{x_1}{2} + \frac{x_4}{2}, \quad x_4 = \frac{x_1}{2} + \frac{x_2}{2} + \frac{x_3}{3}$$

De aquí podemos extraer un sistema de ecuaciones:

$$\left. \begin{aligned} -x_1 + \frac{x_2}{2} + \frac{x_3}{3} &= 0 \\ \frac{x_1}{3} - x_2 + \frac{x_3}{3} + \frac{x_4}{2} &= 0 \\ \frac{x_1}{3} - x_3 + \frac{x_4}{2} &= 0 \\ \frac{x_1}{3} + \frac{x_2}{2} + \frac{x_3}{3} - x_4 &= 0 \end{aligned} \right\} \Rightarrow \det \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{3} & 0 \\ \frac{1}{3} & -1 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{3} & 0 & -1 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{3} & -1 \end{pmatrix} = 0$$

Observamos que es un sistema incompatible determinado, y por tanto podemos tomar solo 3 de las ecuaciones que plantea pues una de ellas no aporta información al resto.

Tomamos la última ecuación, la eliminamos y tomamos  $x_4$  como un parámetro  $\lambda \in \mathbb{R}$ . Por tanto:

$$\begin{aligned} \left( \begin{array}{ccc|c} -1 & 1/2 & 1/3 & 0 \\ 1/3 & -1 & 1/3 & (-1/2)\lambda \\ 1/3 & 0 & -1 & (-1/2)\lambda \end{array} \right) &= \left( \begin{array}{ccc|c} -6 & 3 & 2 & 0 \\ 2 & -6 & 2 & -3\lambda \\ 2 & 0 & -6 & -3\lambda \end{array} \right) = \left( \begin{array}{ccc|c} -6 & 3 & 2 & 0 \\ 0 & -5 & 8/3 & -3\lambda \\ 0 & 1 & -16/3 & -3\lambda \end{array} \right) = \\ \left( \begin{array}{ccc|c} -6 & 3 & 2 & 0 \\ 0 & -5 & 8/3 & -3\lambda \\ 0 & 0 & -24/5 & -18/5\lambda \end{array} \right) &= \left( \begin{array}{ccc|c} -6 & 3 & 2 & 0 \\ 0 & -15 & 8 & -9\lambda \\ 0 & 0 & 4 & -3\lambda \end{array} \right) \Rightarrow \left. \begin{aligned} x_1 &= \frac{3}{4}x_4 \\ x_2 &= x_4 \\ x_3 &= \frac{3}{4}x_4 \end{aligned} \right\} \Rightarrow \left. \begin{aligned} x_1 &= \frac{3}{4}\lambda \\ x_2 &= \lambda \\ x_3 &= \frac{3}{4}\lambda \\ x_4 &= \lambda \end{aligned} \right\} \\ \lambda &\in \mathbb{R} \end{aligned}$$

Para hacernos una idea de la importancia de las páginas, si le damos valores a  $\lambda$  como  $\lambda = 4$ , obtenemos que  $x_1 = x_3 = 3$ , y  $x_2 = x_4 = 4$ , observando así que las páginas con mayor relevancia son la 1 y la 4.

**Ejercicio 10.** Decide razonadamente la validez del siguiente razonamiento: “En la sucesión de Fibonacci, al ser

$$n \in \mathbb{N} \Rightarrow x_{n+2} = x_{n+1} + x_n$$

dividiendo por  $x_{n+1}$  obtenemos

$$\frac{x_{n+2}}{x_{n+1}} = 1 + \frac{1}{\frac{x_{n+1}}{x_n}}$$

luego notando

$$\ell = \lim_{n \rightarrow \infty} \frac{x_{n+1}}{x_n}$$

se tiene de la igualdad anterior

$$\ell = 1 + \frac{1}{\ell} \Rightarrow \ell^2 - \ell - 1 = 0$$

y como  $\ell \geq 0$ , entonces  $\ell = \Phi$ ”

(Indicación: habría que probar antes que la sucesión  $\{x_{n+1}/x_n\}_{n \in \mathbb{N}}$  es convergente. En general, el hecho de que una sucesión verifique una relación de recurrencia no garantiza su convergencia, como ocurre por ejemplo con la sucesión

$$x_1 = 1$$

y

$$n \in \mathbb{N} \Rightarrow x_{n+1} = -x_n + 1).$$

Primero de todo, observemos que en el ejemplo efectivamente tenemos una sucesión no convergente:

$$x_1 = 1, \quad x_2 = 0, \quad x_3 = 1 \dots$$

Pero si aplicamos el álgebra de límites:

$$\ell = -\ell + 1 \Rightarrow \ell = \frac{1}{2} \text{ lo cual no tiene sentido porque la serie no converge}$$

Siguiendo este ejemplo, podemos afirmar que el razonamiento utilizado **no** es correcto, puesto que no se prueba que la sucesión  $\{x_{n+1}/x_n\}_{n \in \mathbb{N}}$  sea convergente y se le aplica el álgebra de límites directamente, pudiendo llegar a un resultado falso como hemos visto anteriormente.

Primero comprobemos si la sucesión es convergente:

$$\frac{x_{n+1}}{x_n} = \frac{\frac{1}{\sqrt{5}}(\Phi^{n+1} - \Psi^{n+1})}{\frac{1}{\sqrt{5}}(\Phi^n - \Psi^n)} = \frac{\Phi^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}}{\Phi^n - \left(\frac{1-\sqrt{5}}{2}\right)^n} = \frac{\Phi - \Phi \left(\frac{1-\sqrt{5}}{2\Phi}\right)^{n+1}}{1 - \left(\frac{1-\sqrt{5}}{2\Phi}\right)^n} = \frac{x_{n+1}}{x_n}$$

Por otro lado:

$$\left| \frac{1 - \sqrt{5}}{2\Phi} \right| \approx 0.382 \dots < 1 \Rightarrow \lim_{n \rightarrow \infty} \left( \frac{1 - \sqrt{5}}{2\Phi} \right)^n = 0$$

Y queda probado que la sucesión converge y ahora sí que podemos aplicar el álgebra de límites:

$$\lim_{n \rightarrow \infty} \frac{x_{n+1}}{x_n} = \lim_{n \rightarrow \infty} \frac{\Phi - \Phi \left(\frac{1-\sqrt{5}}{2\Phi}\right)^{n+1}}{1 - \left(\frac{1-\sqrt{5}}{2\Phi}\right)^n} = \lim_{n \rightarrow \infty} \frac{\Phi}{1} = \Phi$$

Ahora el razonamiento sí que estaría bien realizado.

**Ejercicio 11.** Obtén la representación posicional binaria de los números

$$8.275, \quad -6.6875, \quad \frac{5}{7}.$$

■ 8.275

$$8)_{10} = 2^3)_{10} = 1000)_{10}$$

$$0.275 \cdot 2 = 0.55 \rightarrow 0$$

$$0.55 \cdot 2 = 1.1 \rightarrow 1$$

$$0.1 \cdot 2 = 0.2 \rightarrow 0$$

$$0.2 \cdot 2 = 0.4 \rightarrow 0$$

$$0.4 \cdot 2 = 0.8 \rightarrow 0$$

$$0.8 \cdot 2 = 1.6 \rightarrow 1$$

$$0.6 \cdot 2 = 1.2 \rightarrow 1$$

$$0.2 \cdot 2 = 0.4 \rightarrow 0 \dots$$

$$0.275)_{10} = 0.0100011001100\dots \Rightarrow 0.01001\widehat{1100}$$

Así:

$$8.275)_{10} = 1000.010001\widehat{1100}$$

■ -6.6875

$$-6)_{10} = 2^2 + 2)_{10} = -110)_2$$

$$0.6875 \cdot 2 = 1.375 \rightarrow 1$$

$$0.375 \cdot 2 = 0.75 \rightarrow 0$$

$$0.75 \cdot 2 = 1.5 \rightarrow 1$$

$$0.5 \cdot 2 = 1 \rightarrow 1$$

$$0 \cdot 2 = 0 \rightarrow 0$$

$$-6.6875)_{10} = -110.10110)_2$$

■  $\left(\frac{5}{7}\right)_{10}$

$$\left(\frac{5}{7}\right)_{10} = 0.\widehat{714285})_{10}$$

$$\frac{5}{7} = \frac{1}{2} \left(1 + \frac{3}{7}\right) = \frac{1}{2} + \frac{1}{8} \cdot \frac{12}{7} = \frac{1}{2} + \frac{1}{8} \left(1 + \frac{5}{7}\right) = \frac{1}{2} + \frac{1}{8} + \frac{1}{8} \cdot \left(\frac{5}{7}\right)$$

$$\left(\frac{5}{7}\right)_{10} = 0.\widehat{101})_2$$

**Ejercicio 12.** ¿Es posible encontrar un número real con representación decimal infinita y binaria finita? ¿Por qué? De forma más general, ¿qué debe cumplir una base  $b$  para que toda representación posicional finita en dicha base sea también finita en base 10?

Por la teoría vista en clase, podemos afirmar que existen bases de representación binaria infinita que implican representación decimal finita. En este caso se nos pregunta por el contrarrecíproco  $(A \Rightarrow B) \iff (\bar{B} \Rightarrow \bar{A})$ .

Usaremos inducción:

$$\forall x \in \mathbb{R} \Rightarrow x = (-1)^s \sum_{n=-\infty}^N x_n b^n$$

donde  $N = -1$  en nuestro caso, puesto que solo queremos la parte decimal. Puesto que estamos trabajando en binario, podemos afirmar que  $b = 2$  y que  $x_n = 0$  o  $x_n = 1$ . A partir de ahora trabajaremos con números negativos, entonces nuestras condiciones de inducción se tendrán que cumplir  $\forall -n \in \mathbb{N}$ .

$$x \in \mathbb{R} \quad x = (-1)^s \sum_{i=-1}^{-1} (x_i 2^i) = (-1)^s x_{-1} 2^{-1} = (-1)^s \cdot x_{-1} \cdot \frac{1}{2} \quad \left\{ \begin{array}{l} -1 \cdot 0 \cdot \frac{1}{2} = 0)_{10} \rightarrow \text{finito} \\ -1 \cdot 1 \cdot \frac{1}{2} = -0.5)_{10} \rightarrow \text{finito} \\ 1 \cdot 0 \cdot \frac{1}{2} = 0)_{10} \rightarrow \text{finito} \\ 1 \cdot 1 \cdot \frac{1}{2} = 0.5)_{10} \rightarrow \text{finito} \end{array} \right.$$

Por tanto  $n = -1$  verifica que si la representación posicional es finita, también lo es la decimal.

Suponemos que se verifica para  $n$  verifica la anterior condición y comprobemos para el caso  $n - 1$ :

$$x = (-1)^s \sum_{i=n}^{-1} (x_i 2^i) \text{ tiene representación decimal finita por hipótesis}$$

Sea

$$x' = (-1)^{s'} \sum_{i=n-1}^{-1} (x'_i 2^i) = (-1)^{s'} \left( \sum_{i=n}^{-1} (x'_i 2^i) + (x'_{n-1} 2^{n-1}) \right) = (-1)^{s'} \sum_{i=n}^{-1} (x'_i 2^i) + (-1)^{s'} (x'_{n-1} 2^{n-1})$$

$$(-1)^{s'} (x'_{n-1} 2^{n-1}) \quad \left\{ \begin{array}{l} -1 \cdot 0 \cdot \frac{1}{2^{n-1}} = 0 \rightarrow \text{finito} \\ -1 \cdot 1 \cdot \frac{1}{2^{n-1}} = \frac{1}{2^{n-1}} \rightarrow \text{finito} \\ 1 \cdot 0 \cdot \frac{1}{2^{n-1}} = 0 \rightarrow \text{finito} \\ 1 \cdot 1 \cdot \frac{1}{2^{n-1}} = \frac{1}{2^{n-1}} \rightarrow \text{finito} \end{array} \right.$$

Queda probado que supuesto que  $n$  cumple la condición,  $n - 1$  la cumple también. Así, los números decimales finitos al que se le suma otro con decimales finitos da como resultado otro número con decimales finitos.

Respondiendo a la segunda parte del ejercicio, y utilizando lo demostrado en el ejercicio, podemos concluir que dicha base  $b$  debe cumplir que:

$$\frac{1}{b^n} \quad n \in \mathbb{N}$$

tiene representación posicional finita en base 10.

**Ejercicio 13.** Fijada una base  $b$ , sean  $k \geq 1$ ,  $0 \leq a_{k+1}, a_{k+2} \dots \leq b - 1$  y  $0 \leq a_k < b - 1$ . Demuestra que

$$(0.0 \dots 0a_k a_{k+1} a_{k+2} \dots)_b \leq (0.0 \dots 0a_k + 1)_b.$$

Basta con fijarse en que

$$(0.0 \dots 0a_k a_{k+1} a_{k+2} \dots)_b = \frac{a_k}{b^k} + \sum_{j=k+1}^{\infty} \frac{a_j}{b^j} \leq \frac{a_k}{b^k} + (b-1) \cdot \frac{\frac{1}{b^{k+1}}}{1 - \frac{1}{b}} \leq \frac{a_k}{b^k} + \frac{1}{b^k} = (0.0 \dots 0a_{k+1})_b$$

Y queda probado.

**Ejercicio 14.** Describe todos los números estrictamente positivos del sistema de punto flotante  $\mathbb{F}(2, 3, -1, 2)$ . Calcula su epsilon máquina y su precisión. Obtén además la truncatura y el redondeo en dicho sistema del número real 3.25, comprobando que los correspondientes errores relativos están acotados por el epsilon máquina y la precisión.

Observando el sistema de punto flotante, podemos afirmar que  $b = 2$ ,  $t = 3$  y  $e = -1, 0, 1, 2$ . De esta manera:

$$\begin{aligned} (0.111) \cdot 2^{-1} &= \frac{7}{16} & (0.110) \cdot 2^{-1} &= \frac{3}{8} & (0.101) \cdot 2^{-1} &= \frac{5}{16} & (0.100) \cdot 2^{-1} &= \frac{1}{4} \\ (0.111) \cdot 2^0 &= \frac{7}{8} & (0.110) \cdot 2^0 &= \frac{3}{4} & (0.101) \cdot 2^0 &= \frac{5}{8} & (0.100) \cdot 2^0 &= \frac{1}{2} \\ (0.111) \cdot 2^1 &= \frac{7}{4} & (0.110) \cdot 2^1 &= \frac{3}{2} & (0.101) \cdot 2^1 &= \frac{5}{4} & (0.100) \cdot 2^1 &= 1 \\ (0.111) \cdot 2^2 &= \frac{7}{2} & (0.110) \cdot 2^2 &= 3 & (0.101) \cdot 2^2 &= \frac{5}{2} & (0.100) \cdot 2^2 &= 2 \end{aligned}$$

Calculemos ahora el epsilon máquina:

$$\varepsilon_M = b^{1-t} = 2^{1-3} = 2^{-2} = \frac{1}{4}$$

Precisión:

$$u = \frac{1}{2} \varepsilon_M = \frac{1}{8}$$

Ahora representamos 3.25 en dicho sistema de punto flotante:

$$\left. \begin{aligned} 3)_{10} &= 011)_2 \\ 0.25)_{10} &= \frac{1}{4} = 0.010)_2 \end{aligned} \right\} \Rightarrow 3.25)_{10} = 011.010)_2 = (0.11010) \cdot 2^2$$

Y pasamos a calcular el truncamiento y el redondeo:

■ Truncamiento

$$tr(x) = (0.110) \cdot 2^2 = 3)_{10}$$

Ahora comprobamos que el truncamiento está acotado:

$$\text{error relativo} = \frac{|x - tr(x)|}{|x|} = \frac{|3.25 - 3|}{|3.25|} = \frac{1}{13} < \frac{1}{4} = \varepsilon_M$$

■ Redondeo

$$rd(x) = tr\left(x + (-1)^s \frac{b}{2} \frac{b^e}{b^{t+1}}\right) = tr\left((0.1101) \cdot 2^2 + \frac{2}{2} \cdot \frac{2^2}{2^4}\right) = tr((0.1101) \cdot 2^2 + 2^{-2}) =$$

$$\text{tr}((0.1101) \cdot 2^2 + (0.0001) \cdot 2^2) = \text{tr}((0.1110) \cdot 2^2) = (0.111) \cdot 2^2 = \frac{7}{2} = 3.5)_{10}$$

Ahora comprobamos que el redondeo está acotado:

$$\frac{|x - \text{rd}(x)|}{|x|} = \frac{|3.25 - 3.5|}{|3.25|} = \frac{1}{13} < \frac{1}{8} = u$$

**Ejercicio 15.** Considera un sistema de punto flotante  $\mathbb{F}(b, t, L, U)$  con  $L < U$ . Sean  $L \leq e < e' \leq U$  y  $(0.a_1 \dots a_t) \cdot b^e, (0.\alpha_1 \dots \alpha_t) \cdot b^{e'} \in \mathbb{F}(b, t, L, U)$ . Demuestra que

$$(0.a_1 \dots a_t) \cdot b^e < (0.\alpha_1 \dots \alpha_t) \cdot b^{e'}$$

Deduce la distribución de los puntos del sistema de punto flotante  $\mathbb{F}(b, t, L, U)$ .

Lo primero de todo, vamos a acotar ambos miembros de la desigualdad para que nos sea más sencilla la demostración:

$$\blacksquare (0.a_1 \dots) \cdot b^e = \lambda$$

$$\lambda \leq (0.b - 1 \dots b - 1) \cdot b^e = \sum_{n=1}^t \frac{b-1}{b^n} \cdot b^e = b^e(b-1) \sum_{n=1}^t \left(\frac{1}{b}\right)^n$$

$$\blacksquare (0.\alpha_1 \dots \alpha_t) \cdot b^{e'} = \beta$$

$$\beta \geq (0.10 \dots 0) \cdot b^{e'} = 1 \cdot b^{-1} \cdot b^{e'} = b^{e'-1}$$

Ahora bastará con que demostremos que se cumple la desigualdad entre ambas acotaciones, es decir,

$$b^e(b-1) \sum_{n=1}^t \left(\frac{1}{b}\right)^n < b^{e'-1}$$

Desarrollemos el primer lado de la desigualdad (usamos un resultado visto en clase sobre series geométricas):

$$\begin{aligned} b^e(b-1) \sum_{n=1}^t \left(\frac{1}{b}\right)^n &= b^e \cdot (b-1) \cdot \frac{\frac{1}{b} - \frac{1}{b^{t+1}}}{1 - \frac{1}{b}} = b^e(b-1) \frac{\frac{b^t-1}{b^{t+1}}}{\frac{b-1}{b}} = b^e \cdot (b-1) \frac{(b^t-1)b}{(b-1)(b^{t+1})} = b^e \cdot \left(1 - \frac{1}{b^t}\right) = \\ &= b^e - b^{e-t} \end{aligned}$$

Y obtenemos que

$$b^e - b^{e-t} < b^{e'-1}$$

Para comprobar que esta desigualdad se cumple, operemos:

$$b^{e+1} - b^{e+1-t} < b^{e'} \iff b^{e+1} \cdot \left(1 - \frac{1}{b^t}\right) < b^{e'} \cdot 1$$

Pero sabemos que

$$b^e < b^{e'} \Rightarrow b^{e+1} \leq b^{e'} \quad 1 - \frac{1}{b^t} < 1$$

Y observamos que la desigualdad se cumple.

Tratemos ahora la segunda parte del ejercicio. Usando el ejercicio anterior (ejercicio 14) y los apuntes de teoría de la asignatura deducimos que podemos organizar los puntos de cualquier sistema de punto flotante  $\mathbb{F}$  por grupos.

Así, todos los puntos de la forma  $(0.a_1 \dots a_t) \cdot b^e$  con  $e$  un exponente fijo, guardan una distancia entre sí proporcional con la que guardan los puntos del siguiente grupo.

De esta manera, sea  $x = (0.a_1 \dots a_t) \cdot b^e$  un punto cualquiera. El punto consecutivo será  $x + (0.0 \dots 1) \cdot b^e$ . De lo que deducimos, que en cada grupo, la distancia que guarda un punto con el siguiente y su consecutivo es de  $b^e \cdot \frac{1}{b^t} = b^{e-t}$ .

**Ejercicio 16.** Sea  $\{x_n\}_{n \geq 0}$  la sucesión de números reales definida para cada  $n \in \mathbb{N} \cup \{0\}$  como

$$x_n := \int_0^1 \frac{x^n}{x+3} dx.$$

Justifica razonadamente que esta sucesión verifica la recurrencia

$$x_0 = \log(4/3)$$

y

$$n \geq 1 \Rightarrow x_n = \frac{1}{n} - 3x_{n-1}$$

(utiliza para ello la identidad

$$\frac{x}{x+3} = 1 - \frac{3}{x+3},$$

multiplica por  $x^{n-1}$  e integra). Para estudiar la propagación del error, parte en la recurrencia anterior de un redondeo de  $x_0$  con 5 cifras significativas ( $t = 5$ ). De forma más general, analiza la propagación del error cuando se parte de  $x_0 + \delta$ , con  $\delta \in \mathbb{R}$ . Expresa  $x_n$  como  $f_n(x_0)$ , para una conveniente función  $f_n$  y halla su condicionamiento en  $x_0$ .

Primero de todo, calculemos  $x_0$ :

$$x_0 = \int_0^1 \frac{1}{x+3} dx = \ln|x+3| \Big|_0^1 = \ln \frac{4}{3} = x_0$$

Ahora calculamos  $x_n$  haciendo uso de la identidad que nos proporciona el ejercicio:

$$\frac{x}{x+3} = 1 - \frac{3}{x+3} \iff \frac{x}{x+3} \cdot x^{n-1} = \left(1 - \frac{3}{x+3}\right) \cdot x^{n-1} \iff \frac{x^n}{x+3} = x^{n-1} - \frac{3x^{n-1}}{x+3}$$

De esta manera:

$$\int_0^1 \frac{x^n}{x+3} dx = \int_0^1 x^{n-1} dx - \int_0^1 \frac{3x^{n-1}}{x+3} dx \Rightarrow x^n = \frac{1}{n} - 3x_{n-1}$$

y queda probado que se cumple la relación de recurrencia.

Ahora estudiaremos su propagación del error usando en  $x_0$  5 cifras significativas.

$$x_0 = \ln\left(\frac{4}{3}\right) \approx 0.28768 = x_0 + \delta \quad \delta \in \mathbb{R}$$



Queremos encontrar ahora una apropiada función  $f_n$  tal que  $x_n = f_n(x_0)$ , para así poder estudiar su condicionamiento en  $x_0$  y poder ver cómo se reproduce el error que acabamos de ver que se genera al redondear:

$$\left. \begin{aligned} x_1 &= f_1(x_0) = 1/1 - 3x_0 \\ x_2 &= f_2(x_0) = \frac{1}{2} - 3x_1 = -\frac{5}{2} + 9x_0 \\ x_3 &= f_3(x_0) = \frac{1}{3} - 3x_2 = \frac{47}{6} - 27x_0 \end{aligned} \right\} \Rightarrow f_n(x_0) = \alpha_n + (-3)^n x_0$$

Ahora se nos pide estudiar el condicionamiento de  $f_n$  en  $x_0$ . La definición del condicionamiento relativo en un punto es:

$$c(g, y_0) := \left| \frac{g'(y_0)y_0}{g(y_0)} \right|$$

para un  $g \in C^1(\mathbb{R})$ . Por ser  $f_n$  una composición de funciones elementales, podemos afirmar que  $f_n \in C^1(\mathbb{R}, \forall n \in \mathbb{N})$ . Obtenemos que:

$$c(f_n, x_0) = \left| \frac{(-3)^n \cdot x_0}{x_n + \alpha_n} \right| \underset{\substack{\geq \\ \downarrow \\ \text{La sucesión es decreciente}}}{\geq} \left| \frac{(-3)^n \cdot x_0}{x_0} \right| = |(-3)^n| = 3^n$$

Mal condicionamiento cuando  $n \rightarrow \infty$ .

Ahora estudiaremos la propagación del error sustituyendo  $x_0$  por  $x_0 + \delta$ .

$$\begin{aligned} f_1(x_0 + \delta) &= \alpha_1 - 3x_0 - 3\delta \\ f_2(x_0 + \delta) &= \alpha_2 + 9x_0 + 9\delta \\ &\vdots \\ f_n(x_0 + \delta) &= \alpha_n \end{aligned}$$

Definimos el error como

$$|f_n(x_0) - f_n(x_0 + \delta)|$$

Así,

$$\begin{aligned} e_1 &= 3\delta \\ e_2 &= 9\delta \\ &\vdots \\ e_n &= 3^n \delta \end{aligned}$$

## Ejercicios extra

**Ejercicio 17.** Sea  $\|\cdot\|_1$  en  $\mathbb{R}^N$  y  $\mathbb{R}^M \rightsquigarrow$  norma  $\|\cdot\|_1$  inducida en  $\mathbb{R}^{M \times N}$ . Comprueba que  $A \in \mathbb{R}^{M \times N} \Rightarrow \|A\|_1 = \max_{j=1, \dots, N} \sum_{i=1}^M |a_{ij}|$ .

Sabemos que  $\|A\|_1 \geq \|Ax\|_1$  para  $\|x\|_1 = 1$ . Esto se deduce de aplicar la definición  $\|A\| := \sup_{x \in \mathbb{R}^N, \|x\|=1} \|Ax\|$ .

Sabiendo como está definida  $\|\cdot\|_1$  en  $\mathbb{R}^N$  y que  $\|x\|_1 = 1$ , el valor absoluto de las coordenadas del vector  $x$  han de sumar entre todas 1. Por ejemplo, escogiendo  $x = [1, 0, \dots, 0]^T$  tenemos que

$$\|A\|_1 \geq \left\| A \cdot \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\|_1 = \sum_{i=1}^N |a_{i1}|$$

Este proceso podemos repetirlo de manera análoga con vectores  $v_2 = [0, 1, \dots, 0]^T$ ,  $v_N = [0, 0, \dots, 1]^T$ , de manera que el 1 ocupase todas las posiciones de  $i = 1, \dots, N$ .

De manera reiterada podemos comprobar que  $\|A\|_1 \geq \max_{j=1, \dots, N} \sum_{i=1}^N |a_{ij}|$  y tendríamos probada la primera parte de la igualdad.

Pasemos ahora a demostrar que  $\|A\|_1 \leq \max_{j=1, \dots, N} \sum_{i=1}^N |a_{ij}|$ . Volvemos a considerar un vector tal que su norma 1 sea 1, pero esta vez no lo limitamos de ninguna manera. Así,

$$\begin{aligned} \|Ax\|_1 &= \left\| \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \right\|_1 = \left\| \begin{bmatrix} \sum_{j=1}^N a_{1j} \cdot x_j \\ \vdots \\ \sum_{j=1}^N a_{Nj} \cdot x_j \end{bmatrix} \right\|_1 = \sum_{i=1}^N \left| \sum_{j=1}^N a_{ij} x_j \right| \leq \\ &\leq \sum_{i=1}^N \sum_{j=1}^N |a_{ij}| |x_j| = \sum_{j=1}^N \left[ |x_j| \sum_{i=1}^N |a_{ij}| \right] \leq \sum_{j=1}^N \left[ |x_j| \cdot \left( \max_{t=1, \dots, N} \sum_{i=1}^M |a_{it}| \right) \right] = \max_{t=1, \dots, N} \sum_{i=1}^M |a_{it}| \cdot \sum_{j=1}^N |x_j| = \\ &= \max_{t=1, \dots, M} \sum_{i=1}^M |a_{it}| \cdot \|x\|_1 = \max_{j=1, \dots, N} \sum_{i=1}^M |a_{ij}| \end{aligned}$$

Y concluimos que

$$\|Ax\| \leq \max_{j=1, \dots, N} \sum_{i=1}^M |a_{ij}| = \delta$$

Usando la definición,  $\|A\| := \sup_{x \in \mathbb{R}^N, \|x\|=1} \|Ax\|$ . Acabamos de probar que  $\delta \geq \|Ax\|_1$ , por tanto  $\delta$  es un mayorante de  $\|A\|_1$  y usando el principio del supremo (cualquier mayorante de un conjunto es mayor o igual que el supremo de dicho conjunto)

$$\|A\|_1 := \sup_{x \in \mathbb{R}^N, \|x\|=1} \|Ax\|_1 \leq \delta = \max_{j=1, \dots, N} \sum_{i=1}^M |a_{ij}|$$

Y usando la doble desigualdad queda probada la igualdad.

**Ejercicio 18.** Demuestra que los términos de la sucesión de Fibonacci  $\{x_n\}_{n \in \mathbb{N}}$  vienen dados por

$$x_n = \frac{1}{\sqrt{5}}(\Phi^n - \Psi^n) \quad n \in \mathbb{N} \quad \Phi = \frac{1 + \sqrt{5}}{2} \quad \Psi = \frac{1 - \sqrt{5}}{2}$$

Para esta demostración, utilizaremos la primera sugerencia de las transparencias del tema 1 de la asignatura y procederemos usando el principio de inducción.

Sea  $A = \{n \in \mathbb{N} : \Phi^{n+1} = \Phi^n + \Phi^{n-1}\}$  y probemos que es inductivo:

■ Comprobemos  $n = 1$ :

$$\Phi^2 = \Phi^1 + \Phi^0 \Rightarrow \left( \frac{1 + \sqrt{5}}{2} \right)^2 = \frac{1 + \sqrt{5}}{2} + 1 \Rightarrow \frac{3 + 2\sqrt{5}}{2} = \frac{3 + 2\sqrt{5}}{2}$$

Vemos que se cumple para  $n = 1$ .

- Ahora supongamos que se cumple para  $n$  y comprobemos para  $n + 1$ :

$$\Phi^{n+2} = \Phi^{n+1} + \Phi^n \Rightarrow \Phi^{n+1} \cdot \Phi = \Phi \cdot (\Phi^n + \Phi^{n-1}) \Rightarrow \Phi^{n+1} = \Phi^n + \Phi^{n-1}$$

Usando la hipótesis de inducción, vemos que se cumple  $\forall n \in \mathbb{N}$  y podemos afirmar que el conjunto  $A$  es inductivo.

De manera análoga procedemos para  $\Psi$ .

Sea  $B = \{n \in \mathbb{N} : \Psi^{n+1} = \Psi^n + \Psi^{n-1}\}$  y probemos que es inductivo:

- Comprobemos  $n = 1$ :

$$\Psi^2 = \Psi^1 + \Psi^0 \Rightarrow \left(\frac{1 - \sqrt{5}}{2}\right)^2 = \frac{1 - \sqrt{5}}{2} + 1 \Rightarrow \frac{3 - 2\sqrt{5}}{2} = \frac{3 - 2\sqrt{5}}{2}$$

Vemos que se cumple para  $n = 1$ .

- Ahora supongamos que se cumple para  $n$  y comprobemos para  $n + 1$ :

$$\Psi^{n+2} = \Psi^{n+1} + \Psi^n \Rightarrow \Psi^{n+1} \cdot \Psi = \Psi \cdot (\Psi^n + \Psi^{n-1}) \Rightarrow \Psi^{n+1} = \Psi^n + \Psi^{n-1}$$

Usando la hipótesis de inducción, vemos que se cumple  $\forall n \in \mathbb{N}$  y podemos afirmar que el conjunto  $B$  es inductivo.

Ahora recordemos como se define la sucesión de Fibonacci de manera recurrente:

$$x_{n+2} = x_{n+1} + x_n$$

Empleando la misma técnica, definamos el conjunto  $C = \{n \in \mathbb{N} : x_n = \frac{1}{\sqrt{5}}(\Phi^n - \Psi^n)\}$  y  $x_n$  el  $n$ -ésimo número de Fibonacci} y probemos que es inductivo:

- Comprobemos  $n = 1$ :

$$x_1 = \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} - \frac{1 - \sqrt{5}}{2} \right) = \frac{1}{\sqrt{5}} \cdot \frac{2\sqrt{5}}{2} = 1$$

Vemos que se cumple para  $n = 1$ .

- Ahora supongamos que se cumple para  $n$  y comprobemos para  $n + 1$ :

$$\begin{aligned} x_{n+2} &= \frac{1}{\sqrt{5}}(\Phi^{n+2} - \Psi^{n+2}) = \frac{1}{\sqrt{5}}(\Phi^{n+1} + \Phi^n - \Psi^{n+1} - \Psi^n) = \\ &= \frac{1}{\sqrt{5}}(\Phi^{n+1} - \Psi^{n+1}) + \frac{1}{\sqrt{5}}(\Phi^n - \Psi^n) = x_{n+1} + x_n \end{aligned}$$

Y queda demostrada la forma explícita del  $n$ -ésimo número de Fibonacci.