Métodos Numéricos

Principios de Matemática Numérica

Diego Passarella

Universidad Nacional de Quilmes

2^{do} Cuatrimestre de 2016

Matrices

Determinantes:

Regla de Laplace:

$$det(A) = \begin{cases} a_{11} & \text{si } n = 1 \\ \sum_{j=1}^{n} \Delta_{ij} a_{ij} & \text{si } n > 1 \, \forall \, i = 1, 2, \cdots, n \end{cases}$$

con $\Delta_{ij} = (-1)^{i+j} det\,(A_{ij})$. Siendo A_{ij} la matriz que resulta de eliminar la i-ésima fila y la j-ésima columna.

Se cumple que:

$det(A) = det(A^T)$, $det(AB) = det(A)det(B)$, $det(A^{-1}) = 1/det(A)$, $det(\alpha A) = \alpha^n det(A)$

Cálculo de inversa:

$$A^{-1} = \frac{1}{det(A)} C$$

Siendo C la matriz cuyos elementos son Δ_{ij}

Producto Escalar y Normas

Producto Escalar

Un producto escalar en un espacio vectorial V definido sobre K, es un mapeo $(\cdot,\cdot)$ $V \times V \to K$ que posee las siguientes propiedades:

1. Es lineal con respecto a los elementos de V

$$(\lambda x + \gamma z, y) = \lambda(x, y) + \gamma(z, y), \quad \forall x, y, z \in V, \quad \forall \lambda, \gamma \in K$$

2. Es definido positivo

$$(x, x) \geq 0, \quad \forall x \neq 0 \in V, \quad (x, x) = 0, \quad \Leftrightarrow x = 0$$

Producto Escalar y Normas (cont.)

Normas

Una norma en un espacio vectorial V definido sobre K, es un mapeo $||\cdot||\ V \to K$ si se cumplen los siguientes axiomas:

1. (i) $||v|| \geq 0 \quad \forall v \in V$ y (ii) $||v|| = 0 \quad \Leftrightarrow v = 0$
2. $||\alpha v|| = |\alpha|\,||v|| \quad \alpha \in K, \forall v \in V$ (propiedad de homogeneidad)
3. $||v + w|| \leq ||v|| + ||w||, \forall v, w \in V$ (desigualdad triangular)

El par $(V, ||\cdot||)$ se llama espacio normado.

El mapeo $|\cdot|$ de V en $\mathbb{R}$ que posee solamente las propiedades 1(i), 2 y 3, es llamado seminorma.

Producto Escalar y Normas (cont.)

Normas vectoriales en $\mathbb{R}^n$

Sea $v \in \mathbb{R}^n$:

- Norma L_1:

$$||v||_1 = \sum_{i=1}^{n} |v_i|$$

- Norma L_2:

$$||v||_2 = \left\{ \sum_{i=1}^{n} |v_i|^2 \right\}^{1/2} = (v, v)^{1/2}$$

Producto Escalar y Normas (cont.)

Normas vectoriales en $\mathbb{R}^n$ (cont.)

Sea $v \in \mathbb{R}^n$:

- Norma L_p:

$$||v||_p = \left\{ \sum_{i=1}^{n} |v_i|^p \right\}^{1/p}$$

- Norma L_∞:

$$||v||_\infty = \max_{i \in \{1,2,\cdots,n\}} |v_i|$$

Producto Escalar y Normas (cont.)

Convergencia

Utilización de normas como herramientas de medida

Convergencia de una sucesión de vectores $\{x^{(k)}\}$ en V a un dado vector x

$$\lim_{k\to\infty} x^{(k)} = x \Leftrightarrow \lim_{k\to\infty} ||x^{(k)} - x|| = 0$$

Producto Escalar y Normas (cont.)

Normas matriciales

- Norma L_1:

$$||A||_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{i,j}|$$

- Norma L_∞:

$$||A||_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{i,j}|$$

Pueden plantearse normas subordinadas (o inducidas) a partir de las normas vectoriales.

$$||A||_p = \sup_{x \neq 0} \frac{||Ax||_p}{||x||_p}$$

Problemas bien planteados

Problemas matemáticamente bien planteados

Sea el siguiente problema matemático: Encontrar x tal que

$$F(d, x) = 0$$

Donde d es un conjunto de datos de los cuales depende la solución y F es la relación funcional entre x y d.

Las variables x y d pueden ser números reales, vectores o funciones.

Si se conoce:

- F y d, el problema se llama directo
- F y x, el problema se llama inverso
- x y d, el problema es de identificación

Problemas bien planteados (cont.)

Problemas matemáticamente bien planteados (cont.)

Si la solución x es única y varia de forma continua con los datos d, se dice que el problema matemático está bien planteado, o equivalentemente, es estable.

Si x varia de forma discontinua, el problema se define como mal planteado, o inestable. Problemas mal planteados pueden llegar a reformularse y regularizarse.

Ejemplo de problema mal planteado

Encontrar las raíces de:

$$p(x) = x^4 - x^2(2a - 1) + a(a - 1)$$

en el rango $a = [-0{,}5, 1{,}5]$

Problemas bien planteados (cont.)

Problemas matemáticamente bien planteados (cont.)

La continuidad de la solución x, significa que para pequeñas variaciones de d, la solución también tendrá también pequeñas variaciones.

$$F(d + \delta d, x + \delta x) = 0$$

donde

$$\forall \eta > 0, \exists K(\eta, d) : ||\delta d|| < \eta \Rightarrow ||\delta x|| \leq K(\eta, d)||\delta d||$$

Problemas bien planteados (cont.)

Número de condición

$$K(d) = \sup_{\delta d \in D} \frac{||\delta x||/||x||}{||\delta d||/||d||}, \qquad K_{abs}(d) = \sup_{\delta d \in D} \frac{||\delta x||}{||\delta d||}$$

Números de condición relativos y absolutos.

Un problema matemáticamente bien condicionado posee números de condición "bajos".

Estabilidad de los métodos numéricos

Suponiendo un problema $F(d, x) = 0$ bien planteado, un método numérico para aproximar su solución, consiste en una secuencia de problemas aproximantes:

$$F_n(d_n, x_n) = 0, \qquad n \geq 1$$

Se espera que $x_n \to x$, $d_n \to d$ y $F_n \to F$ cuando $n \to \infty$. El método numérico es consistente si el dato d es admisible en el problema original.

$$F_n(d, x) = F_n(d, x) - F(d, x) = 0, \qquad n \to \infty$$

El concepto de número de condicionamiento para un método numérico dependiente de n es equivalente a lo anteriormente visto.

Convergencia de métodos numéricos

Un método numérico es convergente, sí y solo sí

$$\forall \varepsilon > 0 \quad \exists n_0(\varepsilon), \quad \exists \delta(n_0, \varepsilon) > 0 :$$

$$\forall n > n_0(\varepsilon), \quad \forall ||\delta d_n|| < \delta(n_0, \varepsilon) \Rightarrow ||x(d) - x_n(d + \delta d_n)|| \leq \varepsilon$$

Donde d es un dato admisible del problema matemático, $x(d)$ es su solución correspondiente y $x_n(d + \delta d_n)$ es la solución numérica para el dato $d + \delta d_n$.

Convergencia de métodos numéricos (cont.)

Condiciones de convergencia:

- Problema matemático bien planteado.
- Método matemático estable ($||\delta x_n||$ acotado)

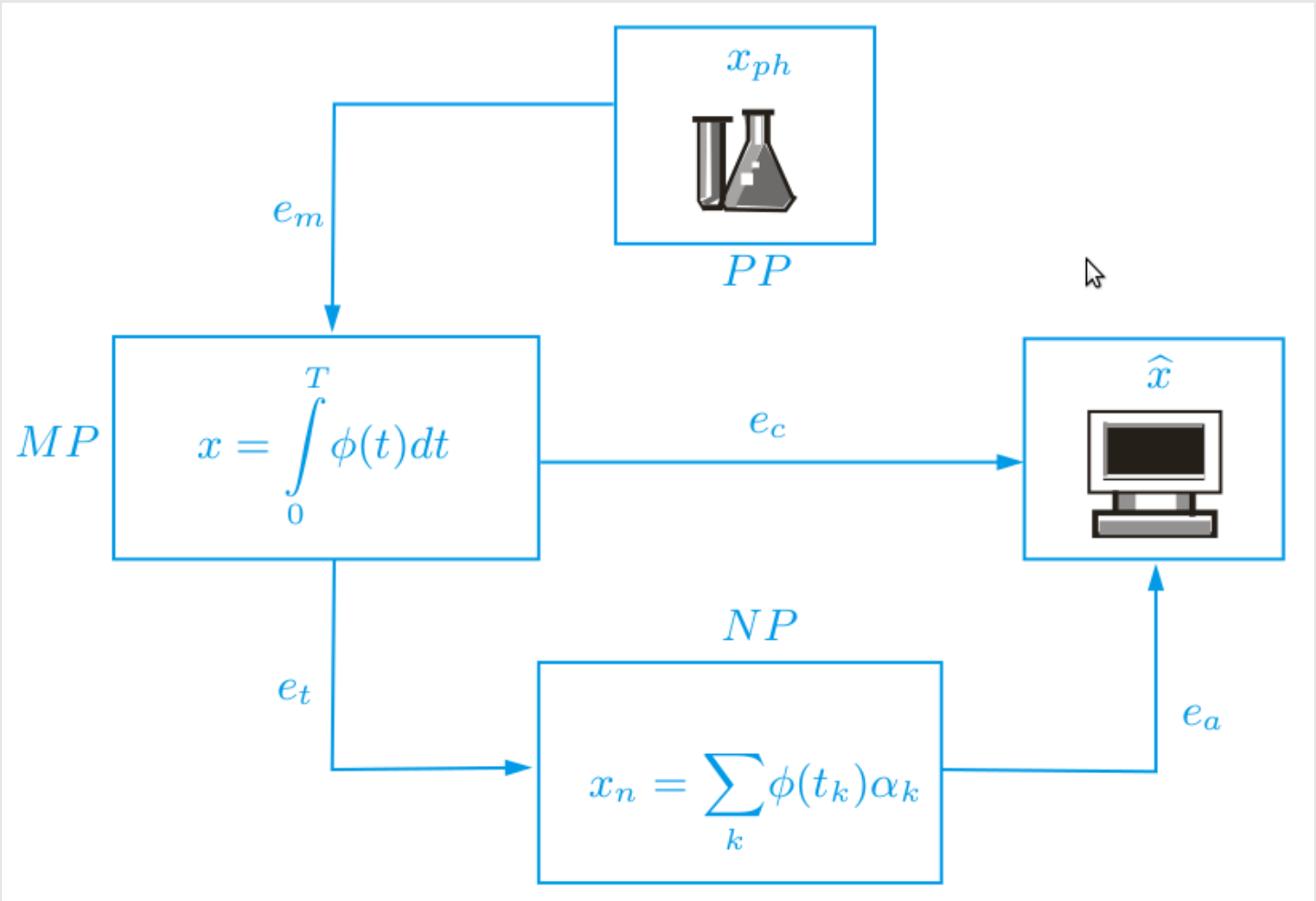
Recordar el concepto de número de condicionamiento, esta vez aplicado a un método numérico.

Recordando el concepto de consistencia:

$$||x(d) - x_n(d)|| \to 0, \qquad n \to \infty$$

Para un método numérico consistente, la condición de estabilidad es equivalente a la de convergencia.

Errores en simulación numérica



Errores

Soluciones

- x_{ph} es la solución física del problema
- x es la solución del modelo matemático
- $\widehat{x}_n$ es la solución numérica del problema matemático

Errores

El error global (e) es la suma del error de modelado ($e_m = x - x_{ph}$) y del error computacional ($e_c = \widehat{x}_n - x$).

$$e = e_m + e_c$$

A su vez, el error computacional contiene errores de truncamiento (e_t) y de redondeo (e_a)

Errores (cont.)

Error de truncamiento (e_t):

- Representación de un número utilizando una menor cantidad de cifras significativas.

$$\pi = 3{,}14159265358979... \simeq 3{,}1415$$

- Este error puede ocurrir en la representación de números u operaciones

$$\int f(x)dx = \sum_{i=1}^{\infty} f(z_i)\Delta x_i \approx \sum_{i=1}^{N} f(z_i)\Delta x_i$$

Errores (cont.)

Error de redondeo (e_a):

- Debido a la representación utilizando un número finito de bits
- Se modifica la última cifra significativa en función de las cifras que no pueden ser representadas

$$\pi = 3{,}14159265358979... \simeq 3{,}1416$$

- Actualmente no presenta muchos problemas inherentes (ver errores históricos). Puede amplificarse debido al algoritmo.

Errores (cont.)

Errores computacionales:

- Introducción de errores de truncamiento (e_t) y redondeo (e_a) en las operaciones de un proceso numérico
- En general no se pueden evitar y deben mantenerse acotados. Un algoritmo que amplifique e_t y e_a será inestable.

Error del modelo (e_m):

- Surge de la incapacidad del modelo matemático de representar todas las funcionalidades e interrelaciones de la realidad física.

Errores (cont.)

Representaciones del error cometido:

Sea $\widehat{x}_n$ una solución computacional que aproxima a x, se define:

- Error absoluto:

$$e_c^{abs} = |\widehat{x}_n - x|$$

- Error relativo: (si $|x| \neq 0$)

$$e_c^{rel} = \frac{|\widehat{x}_n - x|}{|x|}$$

Errores - Convergencia

- En general, los procesos de resolución numérica dependen de un parámetro de discretización (h).
- Si cuando $h \to 0$, el proceso numérico devuelve una solución que se aproxima a la del modelo matemático, se dice que el proceso es convergente.
- Si e_c se puede acotar de la forma:

$$e_c \leq C.h^p$$

con C independiente de h y $p > 0$, se dice que el proceso es convergente de orden p.

Números en b_{10} y b_2

Representación de un real en una base dada

$$x = \pm \left(b_n c^n + b_{n-1} c^{n-1} + \cdots + b_1 c^1 + b_0 c^0 + b_{-1} c^{-1} + b_{-2} c^{-2} + \cdots \right)_c$$

con $n \geq 0$ un dado entero y los coeficientes $b_i = 0, 1, \cdots, c-1$.

Números en b_{10} y b_2

Representación de un real en una base dada

$$x = \pm \left(b_n c^n + b_{n-1} c^{n-1} + \cdots + b_1 c^1 + b_0 c^0 + b_{-1} c^{-1} + b_{-2} c^{-2} + \cdots \right)_c$$

con $n \geq 0$ un dado entero y los coeficientes $b_i = 0, 1, \cdots, c-1$.

Ejemplo:

$$(134{,}82)_{10} = 1 \times 10^2 + 3 \times 10^1 + 4 \times 10^0 + 8 \times 10^{-1} + 2 \times 10^{-2}$$

Números en b_{10} y b_2

Representación de un entero en b_{10} y b_2

0	216	
1		
2		
3		
4		
5		
6		
7		

$$(216)_{10} = (11011000)_2$$

$$(216)_{10} = 1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 0 \times 2^0$$

$$(216)_{10} = 128 + 64 + 0 + 16 + 8 + 0 + 0 + 0$$

Números en b_{10} y b_2

Representación de una fracción de b_{10} en b_2

-1	0.625	
-2		
-3		
-4		
-5		
$\vdots$	$\vdots$	$\vdots$

$$(0{,}625)_{10} = (,10100\cdots)_2$$

$$(0{,}625)_{10} = 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} + 0 \times 2^{-4} + 0 \times \cdots$$

$$(0{,}625)_{10} = 1/2 + 0 + 1/8 + 0 + 0 + \cdots$$

Números de Coma Flotante

En una computadora no se suelen representar los números como binarios con punto decimal. Se utilizan números con coma flotante, cuya forma es:

$$valor = signo \times base^{exponente} \times fraccion$$

La base está predeterminada (2), mientras que la información del signo, exponente y fracción están codificadas de acuerdo al formato del número de coma flotante específico.

Notar que con una cantidad finita de bits no se pueden representar todos los números, solo una cantidad discreta (y no equiespaciada) de ellos.

Números de Coma Flotante

Norma ANSI/IEEE 754-1985: "Standard para Artimética binaria de punto flotante"

Distribución de bits (binary digits, 0 ó 1):

Con esta codificación, los valores representables son:

$$v = (-1)^s \times 2^{e-bias} \times (1.f)_{10}$$

Números de Coma Flotante - ANSI/IEEE 754-1985

Formato:

Parámetro	formato	
	Single	Double
bits	32	64
ne	8	11
p-1	23	52
bias	127	1023

Fracción y exponente:

$$(f)_{10} = (0.b_1 b_2 \cdots b_{p-1})_2$$

ne	e_{min}	e_{max}
8	0	255
11	0	2047

Números de Coma Flotante - ANSI/IEEE 754-1985

Con la fórmula:

$$valor = signo \times base^{exponente} \times fraccion$$

se tienen que representar todos los números racionales posibles de la codificación y tres valores adicionales: `NaN`, `+Inf` y `-Inf`. Estos valores tienen exponentes reservados (al igual que el cero).

Comentarios:

- En formato *double*, el mínimo valor representable es el (∼)2.2251e-308, mientras que el mayor es el (∼)1.7977e+308.
- Los números representables poseen 15 cifras significativas como máximo (*double*).

Números de Coma Flotante

Posibles problemas:

1. Asociatividad:

$$(A + B) - C \neq (A - C) + B \neq A + (B - C)$$

2. Pérdida de dígitos significativos:

$$\frac{(1 + x) - 1}{x} \neq 1, \text{ cuando } x \to 0$$

3. Errores de redondeo:
   - Números irracionales
   - Números racionales con más de 15 dígitos significativos
   - otros (0.1 no puede ser representado exactamente!!!)

Los errores de redondeo son poco relevantes con la precisión actual alcanzada (ver errores históricos). Los errores tipo 1 ó 2 pueden ser más graves.

Errores y Estabilidad

La repetición de pequeños errores de redondeo pueden dar lugar a resultados catastróficos. Ver ejemplos en:

`http://www5.in.tum.de/~huckle/bugse.html`

Generalmente se pierde precisión por redondeo cuando:

- Se restan números muy próximos
- Se divide por un número muy pequeño

Es muy difícil evitar esta clase de errores, y más aún, identificar cuando y en qué parte del algoritmo ocurren.

Propagación de errores en operaciones

Modelo de propagación de errores

Se dispone de dos números de máquina x e y, cada uno con su error asociado (ε).
El error de multiplicación de ambos números resulta:

$$x(1+\varepsilon_x)\cdot y(1+\varepsilon_y) = x\cdot y(1+\varepsilon_y+\varepsilon_x+\varepsilon_y\varepsilon_x) \simeq x\cdot y(1+\varepsilon_y+\varepsilon_x)$$

Por lo que su error asociado es:

$$\varepsilon_{x\cdot y} = \varepsilon_y + \varepsilon_x$$

Propagación de errores en operaciones

Modelo de propagación de errores (cont.)

El error de división de ambos números resulta:

$$\frac{x(1+\varepsilon_x)}{y(1+\varepsilon_y)} = \frac{x}{y}(1+\varepsilon_x)(1-\varepsilon_y+\varepsilon_y^2-+\cdots) \simeq \frac{x}{y}(1+\varepsilon_x-\varepsilon_y)$$

Por lo que su error asociado es:

$$\varepsilon_{x/y} = \varepsilon_x - \varepsilon_y$$

Propagación de errores en operaciones

Modelo de propagación de errores (cont.)

El error de adición/sustracción de dos números resulta:

$$x(1+\varepsilon_x)+y(1+\varepsilon_y)=x+y+x\varepsilon_x+y\varepsilon_y=\cdots$$

$$\cdots(x+y)\left(1+\frac{x}{x+y}\varepsilon_x+\frac{y}{x+y}\varepsilon_y\right)$$

Por lo que su error asociado es:

$$\varepsilon_{x\pm y}=\frac{x}{x+y}\varepsilon_x+\frac{y}{x+y}\varepsilon_y$$

Discutir qué ocurre cuando $x\cdot y>0$, qué cuando $x\cdot y<0$ y cuando $|x|\simeq|y|$

Estabilidad

Ejemplos de inestabilidad:

- $((1+x)-1)/x$ ó $\sqrt{x^2+1}-1$, cuando $x \to 0$
- $(x-1)^7$ en $x \to 1$ versus su versión expandida.
  A los polinomios siempre conviene evaluarlos en su forma anidada:

$$c_n x^n + c_{n-1} x^{n-1} + \cdots + c_2 x^2 + c_1 x + c_0 \qquad \text{FAIL}$$

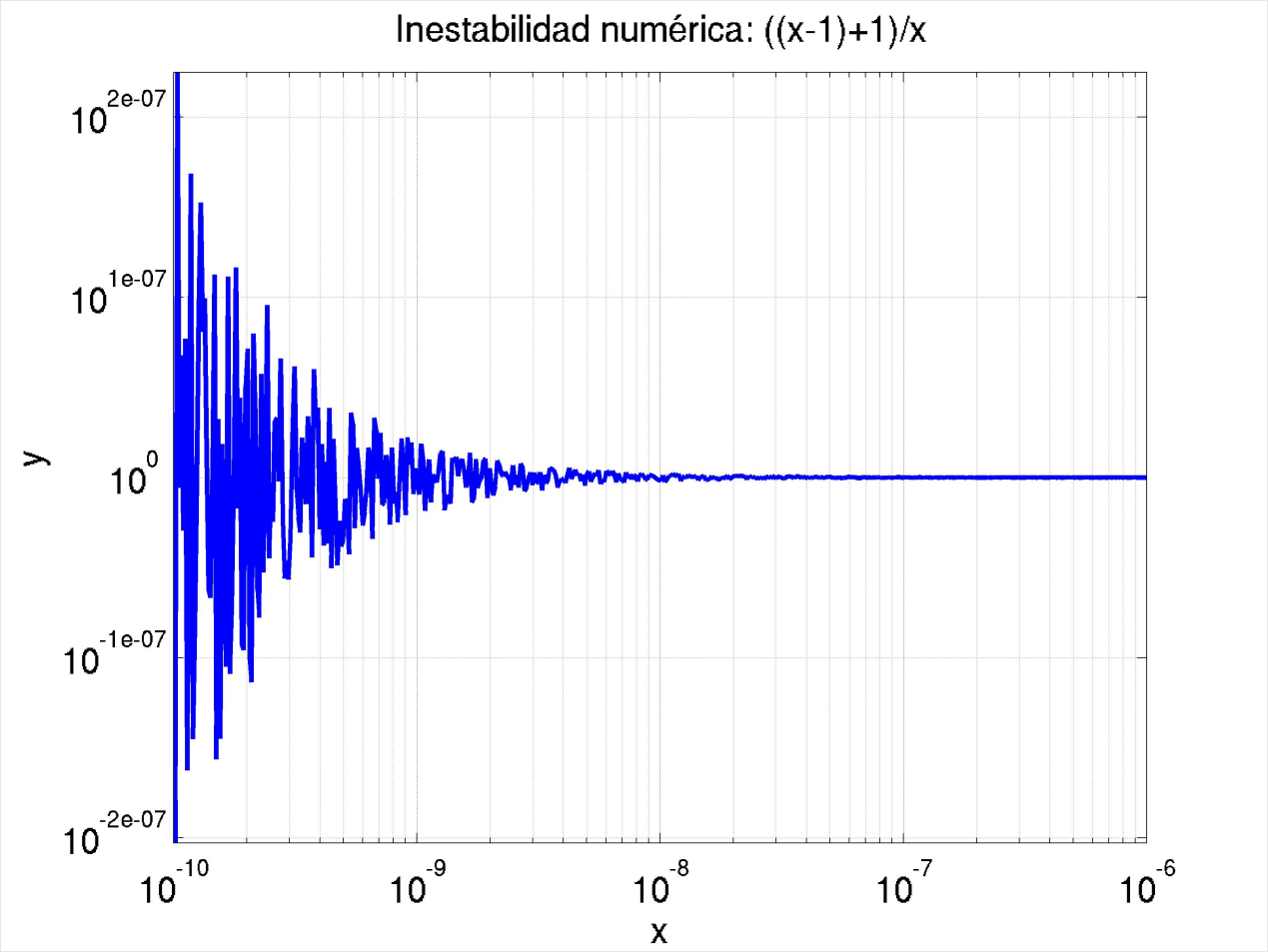
$$((\cdots(c_n x + c_{n-1})\, x + \cdots + c_2)\, x + c_1)\, x + c_0 \qquad \text{OK}$$

- la sucesión $\{x_n\}_{n=2}^{\infty} \to \pi$ con $x_2 = 2$ y

$$x_{n+1} = 2^{n-1/2} \sqrt{1 - \sqrt{1 - 4^{1-n} x_n^2}}$$

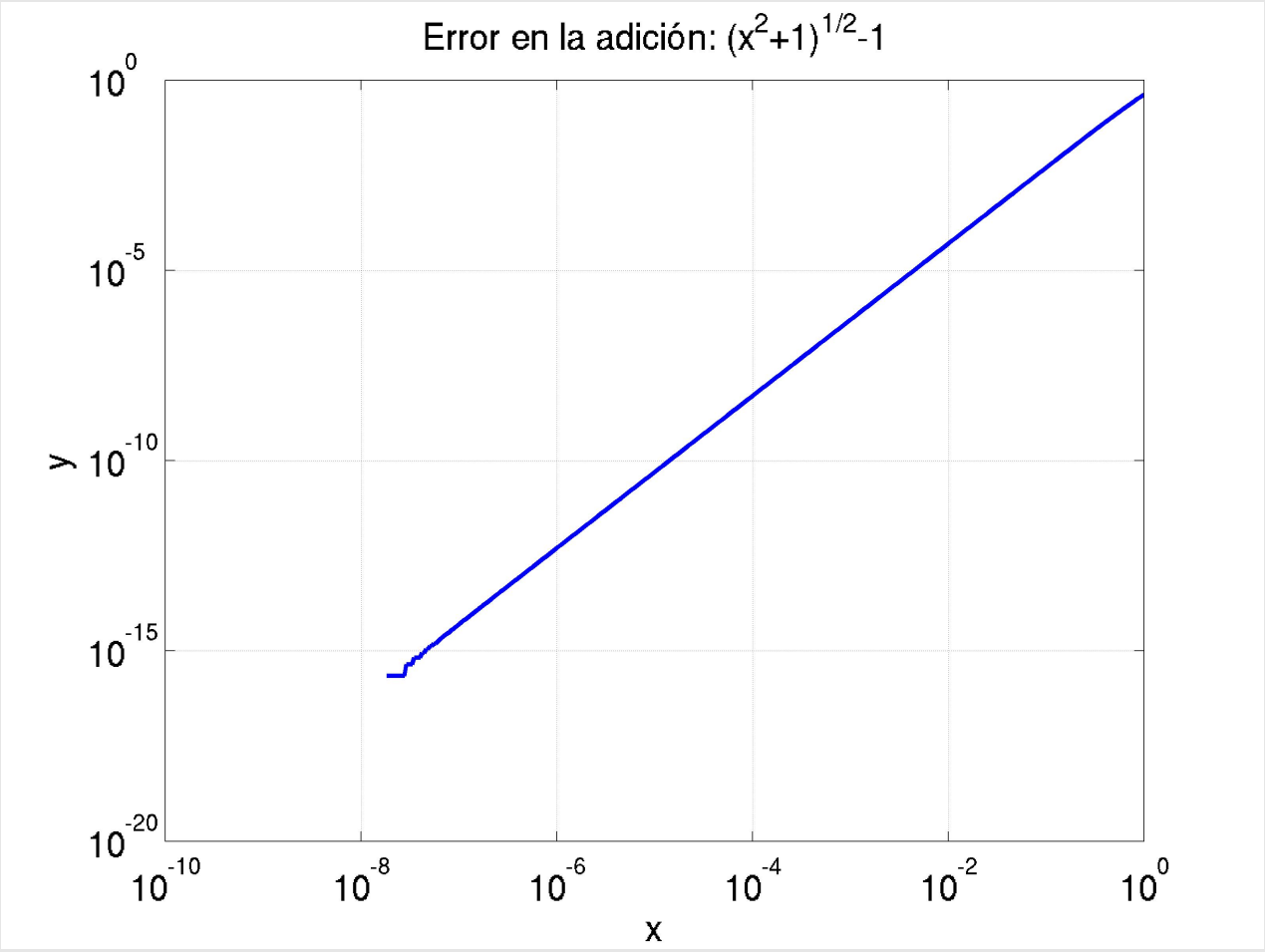
Estabilidad

Ejemplo de inestabilidad: $((1+x)-1)/x$



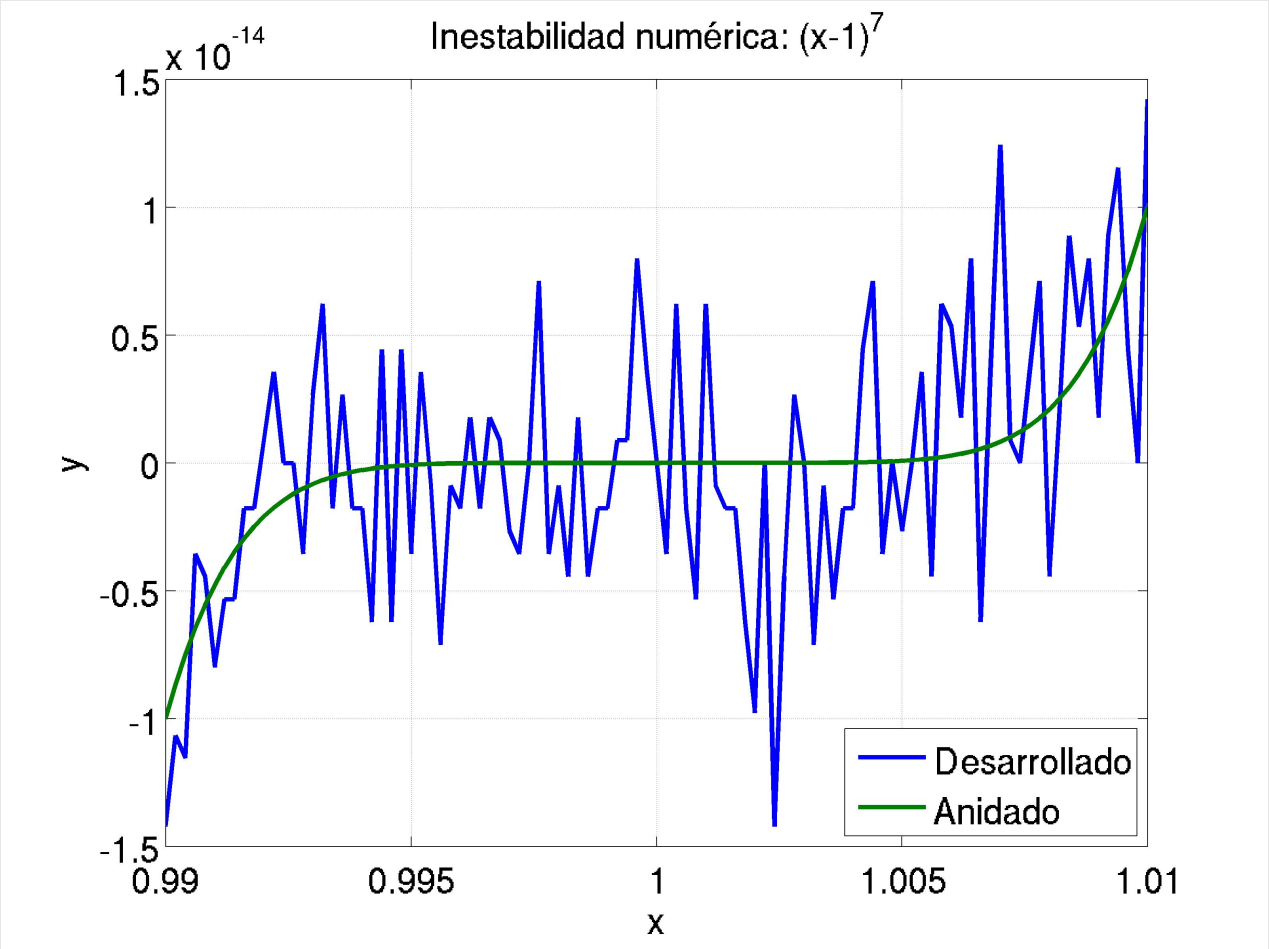
Estabilidad

Ejemplo de falla en la suma: $\sqrt{x^2+1}-1$



Estabilidad

Evaluación de polinomios: $(x-1)^7$



Ejercicios

Utilizando la Norma IEEE 754:

1. ¿Cuáles son los exponentes reservados para definir el 0, $\pm$`Inf`, `NaN`?
2. ¿Qué son los números denormalizados?

Graficar en Octave/Matlab y sacar conclusiones del comportamiento de:

1. La diferencia de la sucesión $x_{n+1} = 2^{n-1/2}\sqrt{1-\sqrt{1-4^{1-n}x_n^2}}$ con $x_2 = 2$ con respecto a π a medida de n crece.