Análisis y Modelamiento Numérico I

Tipos de errores. Pérdida de dígitos significativos

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

2023-1

Contenidos

Notación de Landau

Si tenemos varios algoritmos que resuelven un problema es importante comparar la velocidad con la que se acercan a la solución.

Definición:

Sean $\{x_n\}$ y $\{y_n\}$ sucesiones de números reales. Diremos que $\{x_n\}$ es "big Oh" de $\{y_n\}$ y escribimos $\{x_n\} = O(\{y_n\})$ si existen C y n_0 tales que

$$|x_n| \leq C|y_n|, \forall n \geq n_0$$

Ejemplo:

- $\frac{n+1}{n} = O(\frac{1}{n})$
- $\frac{5}{n} + e^{-n} = O(\frac{1}{n})$

"little Oh"

Definición

Diremos que x_n es "little Oh" de y_n y escribimos $x_n = o(y_n)$ si existe una sucesión $\{\epsilon_n\}$ tal que

- $\epsilon_n \geq 0$
- $\epsilon_n \to 0$
- $|x_n| \leq \epsilon_n |y_n|$

Ejemplo:

- $\frac{1}{n \ln n} = o(\frac{1}{n})$
- $\frac{1}{n} = o(\frac{1}{n \ln n})$
- $e^{-n} = o(\frac{1}{n^2})$

"little Oh"

Ejemplo:

Si $x_n = 3n^2 + 5n + 2$ entonces debemos determinar c tal que

$$3 + 5/n + 2/n^2 \leq c, \text{ si } n > N$$

como $n^2 \geq n$ entonces $3 + 5/n + 2/n^2 \leq 3 + 5/n + 2/n = 3 + 7 = 10$ pues $1/n \leq 1$. Asi que $N = 1$ y $c = 10$.

Propiedades

Se cumplen:

- Si x_n es $O(y_n)$ y y_n es $O(z_n)$ entonces x_n es $O(z_n)$.
- Si $f(n)$ es $O(h(n))$ y $g(n)$ es $O(h(n))$ entonces $f(n) + g(n)$ es $O(h(n))$.
- an^k es $O(n^k)$
- n^k es $O(n^{k+j})$ para $j > 0$
- Si $f(n) = cg(n)$ entonces $f(n)$ es $O(g(n))$.
- Si $\log_a(n)$ es $O(\log_b(n))$, para $a, b > 0$.

Propiedades

Ejemplo:

$x_n = 3n^2 + 2n + 5$ es $O(n^2)$ y también $O(n^3)$, $O(n^4)$, etc.

La notación asintótica puede aparece a ambos lados de la ecuación, por ejemplo

$$3n^2 + O(n) = O(n^2)$$

esto se puede interpretar como: para toda $x_n \in O(n)$ existe $y_n \in O(n^2)$ tal que $3n^2 + x_n = y_n$. La ecuación

$$\begin{aligned} 3n^2 + 2n + 4 &= 3n^2 + O(n) \\ &= O(n^2) \end{aligned}$$

quiere decir que: existe $x_n \in O(n)$ tal que $3n^2 + 2n + 4 = 3n^2 + x_n$ y además para toda $y_n \in O(n)$ existe $z_n \in O(n^2)$ tal que $3n^2 + y_n = z_n$.

Dígitos significativos

Tomemos como ejemplo el número real en forma normalizada

$$\hat{x} = 0{,}45632 \times 10^2$$

Los dígitos mas significativos están asociados a las potencias menos negativas en la mantisa. Podemos asumir que $\hat{x}$ aproxima x con 5 cifras significativas si:

$$|x - \hat{x}| \leq \frac{1}{2} 10^{-5}$$

Definición

Diremos que $\hat{x}$ aproxima x con n cifras significativas si:

$$\left| \frac{x - \hat{x}}{x} \right| \leq \frac{10^{-n}}{2}$$

Dígitos significativos

Ejemplo:

- $\hat{x} = 0{,}222$ aproxima $x = 2/9$ con 3 cifras significativas.
- $\hat{x} = 23{,}496$ aproxima $x = 23{,}494$ con 4 cifras significativas.
- $\hat{x} = 0{,}02138$ aproxima $x = 0{,}02144$ con 2 cifras significativas.

Pérdida de dígitos significativos

Consideremos la diferencia de números cercanos, por ejemplo

$$\begin{aligned} x &= 0{,}4568933456 \\ y &= 0{,}4567863423 \\ x - y &= 0{,}0001070033 \end{aligned}$$

si utilizamos aritmética de punto flotante con una mantisa de 5 cifras vemos que

$$\begin{aligned} \mathrm{fl}(x) &= 0{,}45689 \\ \mathrm{fl}(y) &= 0{,}45679 \\ \mathrm{fl}(x) - \mathrm{fl}(y) &= 0{,}00010 = 0{,}10000 \times 10^{-3} \end{aligned}$$

Pérdida de dígitos significativos(cont.)

De las 5 cifras decimales solo la más significativa es correcta, las demás son de relleno, no representan cifras correctas.
La pérdida de significancia se evidencia cuando calculamos el error relativo:

$$\left|\frac{x-y-(\mathrm{fl}(x)-\mathrm{fl}(y))}{x-y}\right| \left|\frac{0{,}0001070033-0{,}00010}{0{,}0001070033}\right| \approx 6{,}5\,\%$$

Como regla general se debe evitar la sustracción de cantidades casi iguales, por medio de manipulaciones algebraicas.

Pérdida de dígitos significativos(cont.)

Ejemplo:

La función $f(x) = \sqrt{x^2+1} - 1$ involucra una pérdida de dígitos significativos cuando x es pequeño, una manera de evitarlo es reescribir la función como

$$g(x) = \frac{x^2}{\sqrt{x^2+1}+1}$$

Si operamos con un mantisa de 5 dígitos entonces

$$f(0{,}1) = \sqrt{0{,}01+1} - 1 = 1{,}0050 - 1 = 0{,}50000 \times 10^{-2}$$

por otro lado

$$g(0{,}1) = \frac{0{,}01}{\sqrt{1{,}01}+1} = \frac{0{,}01}{2{,}005} = 0{,}49875 \times 10^{-2}$$

Pérdida de bits significativos

Teorema

Si x e y son números binarios de puntos flotante positivos tal que

$$2^{-q} \leq 1 - \frac{y}{x} \leq 2^{-p}$$

entonces la sustracción $x - y$ perderá de p a q bits significativos.

Pérdida de bits significativos (cont.)

Prueba

Si $x = r \times 2^n$, $y = s \times 2^m$ entonces

$$x - y = r \times 2^n - (s \times 2^{m-n}) \times 2^n = (r - s \times 2^{m-n}) \times 2^n$$

La mantisa de $x - y$ se escribe como

$$d = (r - s \times 2^{m-n}) = r\left(1 - \frac{s \times 2^m}{r \times 2^n}\right) = r\left(1 - \frac{y}{x}\right) < 2^{-p}$$

es decir tiene la siguiente forma $(0{,}000\ldots00 b_{p+1}\ldots b_n)_2$

Pérdida de bits significativos

Al normalizar se rellenan con p ceros como mínimo.
Por otro lado

$$2^{-q-1} \leq r\left(1-\frac{y}{x}\right) = d$$

es decir d tiene la forma $(0{,}000\underbrace{b_k}_{=1} b_{k+1}\ldots b_q b_{q+1}\ldots b_n)_2$, para algún $1 \leq k \leq q+1$ y como máximo se pierden q bits.

Pérdida de bits significativos

Ejemplo:

No siempre es evidente que una aproximación conlleve perdida de cifras significativas, por ejemplo

$$e^{-5} = 1 + \frac{(-5)}{1!} + \frac{(-5)^2}{2!} + \frac{(-5)^3}{3!} + \frac{(-5)^4}{4!} + \dots$$

si realizamos los cálculos con una mantisa de 4 decimales obtenemos

Término	Suma parcial	Término	Suma parcial	Término	Suma parcial
1	-4.0000	9	-1.8240	17	0.0095
2	8.5000	10	0.8670	18	0.0101
3	-12.3300	11	-0.3560	19	0.0100
4	13.7100	12	0.1537	20	0.0100
5	-12.3300	13	-0.0423	21	0.0100
6	9.3700	14	0.0277	22	0.0100
7	-6.1300	15	0.0044	23	0.0100
8	3.5580	16	0.0117	24	0.0100

Pérdida de bits significativos

El valor de e^{-5} con 4 cifras significativas es 0.006738 que es muy distinto del valor obtenido usando 24 términos. La perdida de significación se da por la diferencia de términos de magnitud parecida para formar un numero más pequeño. Para evitar la cancelación de números cercanos podemos calcular e^5 y luego aproximar e^{-5} como $1/e^5$.

Término	Suma parcial	Término	Suma parcial	Término	Suma parcial
1	6.0000	9	143.7000	17	148.4000
2	18.5000	10	146.4000	18	148.4000
3	39.3300	11	147.6000	19	148.4000
4	65.3700	12	148.1000	20	148.4000
5	91.4100	13	148.3000	21	148.4000
6	113.1000	14	148.4000	22	148.4000
7	128.6000	15	148.4000	23	148.4000
8	138.3000	16	148.4000	24	148.4000

$$e^5 \approx 148{,}4000 \implies e^{-5} \approx 0{,}006739$$

Recomendación general:

Si una serie de números se suman para obtener una respuesta S de magnitud menor que los términos de la serie entonces el error total disminuye si los términos se suman en orden creciente a la magnitud.
En efecto, si $S = a_1 + a_2 + \ldots + a_n$ y S_n es la n-ésima suma parcial entonces $S_n = fl(a_n + S_{n-1}) = (a_n + S_{n-1})(1 + \epsilon_n)$, y

$$S - S_n \approx -a_1(\epsilon_2 + \ldots + \epsilon_n) - a_2(\epsilon_2 + \ldots + \epsilon_n) - a_3(\epsilon_3 + \ldots + \epsilon_n) - \ldots - a_n\epsilon_n$$

para minimizar el error lo términos grandes como $\epsilon_2 + \ldots + \epsilon_n$ debe ser multiplicados por factores pequeños, por lo tanto si ordenamos según la magnitud

$$|a_1| \leq |a_2| \leq |a_3| \leq \ldots \leq |a_n|$$

el error total ser a minimizado.

Condicionamiento

Podemos formular u problema como una función $f : X \to Y$ de un conjunto X de datos de entrada a un conjunto Y de soluciones o respuestas. Se espera que f sea no lineal y por lo menos continua.

Decimos que un problema esta *mal condicionado* si pequeños cambios en x originan grandes cambios en $f(x)$ la solución del problema. El *número de condición* es un indicador asociado a un problema especifico, si es grande indica mal condicionamiento.

Número de condición

Número de condición

¿Cuál será el efecto en $f(x)$ si perturbamos ligeramente x en h?. Si f es suave y aproximamos por diferenciales

$$f(x+h) - f(x) \approx hf'(x) \implies \frac{f(x+h) - f(x)}{f(x)} \approx \frac{hf'(x)}{f(x)}$$

el cambio relativo en x es $\frac{(x+h)-x}{x} = \frac{h}{x}$, el cambio relativo en el valor de la función es $\frac{f(x+h)-f(x)}{f(x)}$

$$\frac{f(x+h) - f(x)}{f(x)} \approx \frac{xf'(x)}{f(x)} \frac{h}{x}$$

el factor $\dfrac{xf'(x)}{f(x)}$ es el número de condición para este problema.

Número de condición (cont.)

Denotemos $f(x+h)-f(x)$ por $df(x;h)$. En caso f no sea diferenciable definimos el número de condición de la siguiente manera

Definición

El número de condición absoluto $\hat{\kappa}=\hat{\kappa}(x)$ del problema f en x es $\hat{\kappa}=\lim\limits_{\delta\to 0}\sup\limits_{|h|\leq\delta}\frac{df(x;h)}{h}$

si f es diferenciable $\hat{\kappa}=|f'(x)|$

Definición

El numero de condición relativo $\kappa=\kappa(x)$ del problema f en x es $\kappa=\lim\limits_{\delta\to 0}\sup\limits_{|h|\leq\delta}\left|\frac{\frac{df(x;h)}{f(x)}}{\frac{h}{x}}\right|$

si f es diferenciable $\kappa=\left|\frac{xf'(x)|}{f(x)}\right|$

Número de condición (cont.)

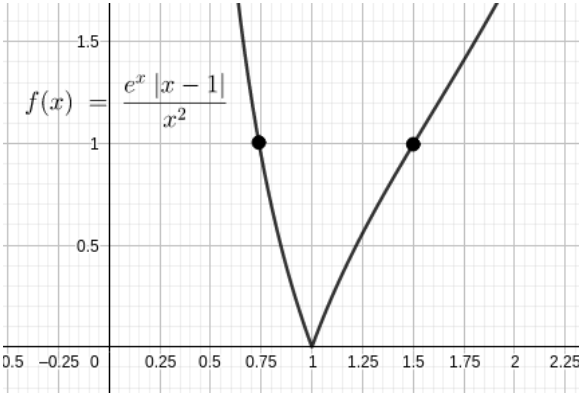
¿Cuales son los números de condición de las siguientes funciones?, ¿Donde son grandes?

- $\ln x$
- $x^{-1}e^{x}$
- $\arccos(x)$

Para $f(x) = x^{-1}e^{x}$ calculamos el número de condición relativo $\kappa = \left|\dfrac{xf'(x)}{f(x)}\right| = |x-1|$, por lo tanto para valores de x, tal que $|x-1| < 1 \implies 0 < x < 2$ el problema estará bien condicionado.

Número de condición (cont.)

Si consideramos $\hat{\kappa} = e^x \left| \frac{x-1}{x^2} \right|$, por lo tanto para valores de $0{,}75 < x < 1{,}52$ el problema estará bien condicionado $\hat{\kappa} < 1$.



Solución de una ecuación no lineal

Consideremos el problema de resolver la ecuación $f(x) = 0$. Aqui el dato de entrada es f y el dato de salida es un numero r: $f(r) = 0$. Si perturbamos f por $f + \epsilon g$ y el dato de salida r por $r + h$, entonces

$$f(r + h) + \epsilon g(r + h) = 0$$

Si f y g son suaves entonces aproximamos por diferenciales

$$f(r) + hf'(r) + \epsilon \left(g(r) + hg'(r)\right) \approx 0$$

luego

$$h = -\frac{\epsilon g(r)}{f'(r) + \epsilon g'(r)}$$

Solución de una ecuación no lineal (cont.)

Ejemplo:

Si $f(x) = (x-1)(x-2)\ldots(x-20) = x^{20} + a_{19}x^{19} + \ldots + 20!$, y $g(x) = x^{20}$, es decir perturbamos el coeficiente de x^{20} tenemos que

$$h \approx -\frac{\epsilon g(20)}{f'(20) + \epsilon g'(20)} = -\frac{\epsilon 20^{20}}{19! + \epsilon 20^{20}}$$

para $\epsilon = 10^{-5}$, $h \approx -1$. Conclusión cambiar el coeficiente de x^{20} en 10^{-5} origina que una de las raíces cambie de $r = 20$ a $r = 19$.

Multiplicación de una matriz por un vector

Para el condicionamiento de la multiplicación de una matriz por un vector considere $A \in \mathbb{R}^{m\times n}$ y el problema de calcular Ax, por lo tanto

$$\begin{aligned}\kappa &= \sup_h \frac{\frac{\|A(x+h)-Ax\|}{\|Ax\|}}{\frac{\|h\|}{\|x\|}} \\ &= \frac{\|x\|}{\|Ax\|}\sup_h \frac{\|Ah\|}{\|h\|} \\ &= \frac{\|x\|}{\|Ax\|}\|A\|\end{aligned}$$

si A es cuadrada e inversible entonces $\|\|x\| = \|A^{-1}Ax\| \leq \|A^{-1}\|\|Ax\|$, luego $\kappa \leq \|A^{-1}\|\|A\|$

Multiplicación de una matriz por un vector (cont.)

Definición

El numero de condición de una matriz A inversible es

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

si $\kappa(A)$ es pequeño A se dirá bien condicionada, si es muy grande A se dirá mal condicionada. Si A es singular escribimos $\kappa(A) = \infty$.

Condicionamiento de un sistema de ecuaciones

Dada la ecuación $Ax = b$ analizamos el problema de obtener la solución del sistema perturbando la matriz del sistema A con δA .

$$(A + \delta A)(x + h) = b \implies h = -A^{-1}(\delta A)x$$

Esto implica que $\|h\| = \|A^{-1}\|\|\delta A\|\|x\|$

$$\frac{\|h\|}{\|x\|} \Big/ \frac{\|\delta A\|}{\|A\|} \leq \|A^{-1}\|\|A\| = \kappa(A)$$

luego el número de condición del problema respecto a la perturbación de A es

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

eset

Condicionamiento de un sistema de ecuaciones (cont.)

Ejemplo:

Si f es de la forma

$$f(x) = \alpha x^{12} + \beta x^{13}$$

tal que $f(0{,}1) = 6{,}06 \times 10^{-13}$ y $f(0{,}9) = 0{,}03577$. Determine α y β, y analice la sensibilidad de estos parámetros ante ligeros cambios en los valores de f en los dos puntos indicados.
Podemos plantear el problema de encontrar $y = (\alpha, \beta)^T$ dada la ecuación

$$Ay = \begin{bmatrix} (0{,}1)^{12} & (0{,}1)^{13} \\ (0{,}9)^{12} & (0{,}9)^{13} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

como $A^{-1} = |A|^{-1} \begin{bmatrix} (0{,}9)^{13} & -(0{,}1)^{13} \\ -(0{,}9)^{12} & (0{,}1)^{12} \end{bmatrix}$ entonces $\kappa(A) = 0{,}9^{12}|A|^{-1}0{,}9^{12} \approx 9^{12}$. En conclusión el problema esta mal condicionado.

Referencias

- Numerical Analysis: Mathematics of Scientific Computing, Third Edition David Kincaid: University of Texas at Austin, Austin, TX, Ward Cheney.
- Numerical Methods Using Matlab, 4th Edition John H. Mathews, California State University, Fullerton, Kurtis K. Fink, Northwest Missouri State University
- Numerical Lineal Algebra. Lloyd N. Trefethen and David Bau, III xii+361 pages. SIAM, 1997
- Elementary Numerical Analysis, 3rd Edition Kendall Atkinson, Weimin Han