Máster Universitario en Computación Paralela y Distribuida
Algoritmos Matriciales Paralelos en Ingeniería

Tema 1.

Computación con precisión finita. Representación en coma flotante. Normas vectoriales y matriciales.

Bibliografía:

“Matrix Computations”. G.Golub & C. Van Loan. Baltimore ; London : Johns Hopkins University Press, 1996 (u otra edición del libro)

Lecturas recomendadas:

“Matrix Computations”. G.Golub & C. Van Loan.

Capítulo 1. Puntos 1.1, 1.2 y 1.3

Capítulo 2. Puntos 2.1, 2.2, 2.3, 2.4 y 2.7

Computación en precisión finita

- La operaciones aritméticas realizadas en un computador suelen estar afectadas por errores de redondeo
- Solo se dispone de una cantidad limitada de memoria para almacenar números
- Los sistemas de representación en coma flotante solo permiten representar un subconjunto finito de elementos del conjunto de los números reales
- Existen números consecutivos y números máximo y mínimo representables
- Los algoritmos van a trabajar con datos aproximados
- Esto va a dar lugar a que tengan que ser analizados y diseñados con cuidado para evitar errores.

Sistemas de representación en coma flotante

- Se caracterizan por cuatro números enteros:

Base : β. Precisión : t. Rango de exponentes : $[L,U]$

$F \equiv$ Conjunto de números en coma flotante

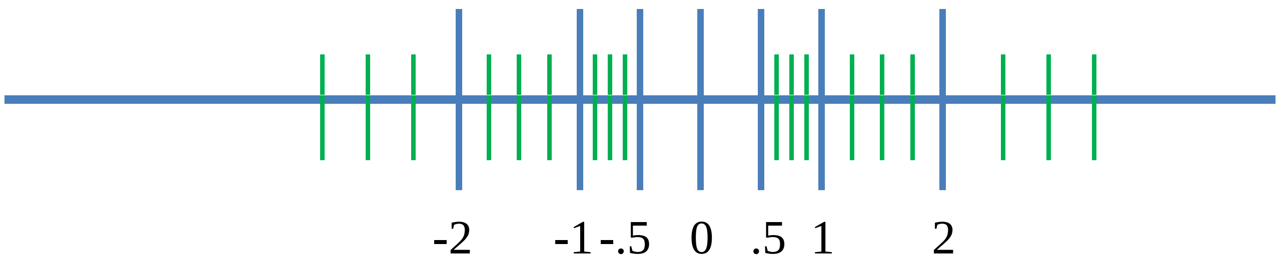
$$F = \{f \in \Re : f = \pm .d_1 d_2 \ldots d_t \times \beta^e, \text{con } 0 \le d_i < \beta,\ d_1 \ne 0,\ L \le e \le U\}$$

- Números extremos representables

Para $f \in F, f \ne 0,$ se tiene

$$m \le |f| \le M, \text{con } m = \beta^{L-1} \text{ y } M = \beta^U (1 - \beta^{-t});$$

- Ejemplo: $\beta = 2;\ t = 3;\ L = 0;\ U = 2.$



Modelo aritmético de las operaciones en coma flotante

* Sea $G = \{x \in \Re : m \le |x| \le M\} \cup \{0\}$

  Se define el operador en coma flotante $fl{:}G \to F$

$$fl(x) = \begin{cases} c \in F \text{ más próximo a } x, \text{ si se usa aritmética redondeada} \\ c \in F \text{ más próximo a } x \text{ que verifique } |c| \le |x|, \text{ si se usa aritmética truncada} \end{cases}$$

* Redondeo unidad

$$\mathrm{u} = \begin{cases} \dfrac{1}{2}\beta^{1-t}, \text{ si se usa aritmética redondeada} \\ \beta^{1-t} \text{ si se usa aritmética truncada} \end{cases}$$

* Errores debidos al uso del operador en coma flotante

$$fl(x) = x(1+\varepsilon),\ con\ \ |\varepsilon| \le u$$

Sistemas de representación en coma flotante en el Matlab

$$\text{Doble precisión}: \beta = 2;\ t = 52;\ L = -1023;\ U = 1024.$$

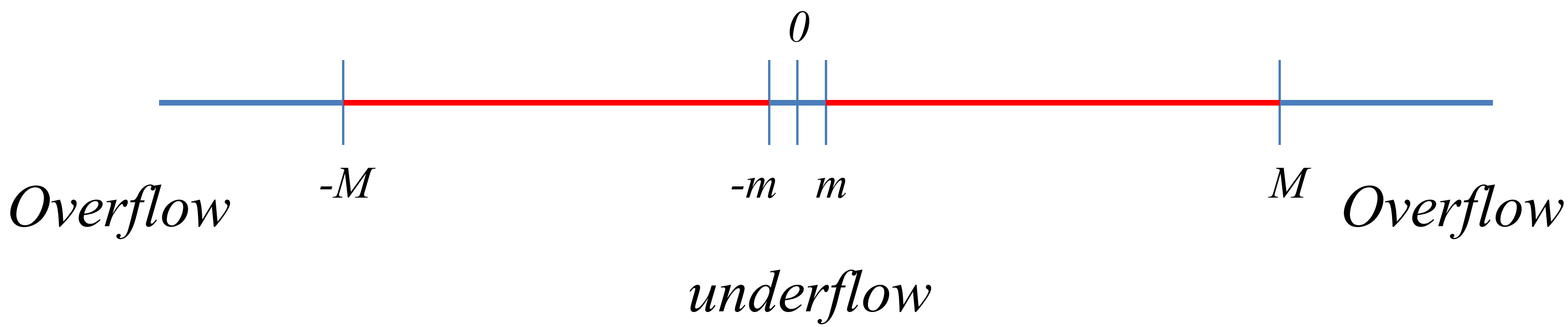
$$\text{Simple precisión}: \beta = 2;\ t = 23;\ L = -127;\ U = 128.$$

- Help/Math Constants

	Simple precisión	**Doble precisión**
Redondeo unidad: **eps**	1.1920929e-007	2.220446049250313e-016
Real máximo: **realmax**	3.4028235e+038	1.797693134862316e+308
|Real mínimo|: **realmin**	1.1754944e-038	2.225073858507201e-308
Infinito: **Inf**	1/0 (Overflow)	1/0 (Overflow)
Not-a-Number: **NaN**	0/0 , Inf-Inf (Indeterminación)	0/0, Inf-Inf (Indeterminación)

Operaciones que producen inestabilidad

Representaciones fuera de rango: *Overflow* y *underflow*



Overflow: puede presentarse al dividir números grandes por números pequeños

Cancelaciones catastróficas: puede presentarse al obtener números pequeños a partir de operaciones con números grandes, especialmente restas

Normas Vectoriales

$$f : \Re^n \to \Re \text{ es una norma vectorial si :}$$

$$1.\ f(x) \geq 0 \text{ y } f(x) = 0 \leftrightarrow x = 0, \forall x \in \Re^n$$

$$2.\ f(x + y) \leq f(x) + f(y), \forall x, y \in \Re^n$$

$$3.\ f(\alpha x) = |\alpha| f(x), \forall \alpha \in \Re, \forall x \in \Re^n$$

Ejemplos: p-normas vectoriales

$$x \in \Re^n$$

$$\|x\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}$$

$$\|x\|_1 = (|x_1| + |x_2| + \ldots + |x_n|)$$

$$\|x\|_2 = \sqrt{x_1^{\ 2} + x_2^{\ 2} + \ldots + x_n^{\ 2}} = \sqrt{x^T x}$$

$$\|x\|_\infty = \max_{1 \leq i \leq n}(|x_i|)$$

Algunas propiedades de las normas vectoriales

- Desigualdad de Holder: $\left|x^T y\right| \le \|x\|_p \|y\|_q, \text{ si } \dfrac{1}{p}+\dfrac{1}{q}=1$

- Desigualdad de Cauchy-Schwartz: $\left|x^T y\right| \le \|x\|_2 \|y\|_2$

- Equivalencia de todas las normas vectoriales en $\Re^n$

Si $\|\bullet\|_\alpha$ y $\|\bullet\|_\beta$ son normas en $\Re^n$, existen constantes positivas c_1 y c_2 tales que $c_1\|x\|_\alpha \le \|x\|_\alpha \le c_2\|x\|_\alpha \ , \forall x \in \Re^n$

Ejemplos: Si $x \in \Re^n$:

$$\|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2$$

$$\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$$

$$\|x\|_\infty \le \|x\|_1 \le n\|x\|_\infty$$

Normas Matriciales

$$f : \Re^{m\times n} \to \Re \text{ es una norma matricial si :}$$

$$1.\ f(x) \geq 0 \text{ y } f(x) = 0 \leftrightarrow x = 0, \forall x \in \Re^{m\times n}$$

$$2.\ f(x+y) \leq f(x) + f(y), \forall x, y \in \Re^{m\times n}$$

$$3.\ f(\alpha x) = |\alpha| f(x), \forall \alpha \in \Re, \forall x \in \Re^{m\times n}$$

Ejemplos de Normas Matriciales

$A \in \Re^{mxn}, x \in \Re^{n}$

Norma de Frobenius

$$\|A\|_F = (\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}{}^2)^{1/2} = (tr(A^T A))^{1/2} \quad (* \quad tr(A) = \sum_{i=1}^{\min(m,n)} a_{ii} \quad *)$$

p - norma matricial

$$\|A\|_p = \sup_{x\in\Re^n} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\|x\|_p=1} \|Ax\|_p$$

Errores absoluto y relativo al aproximar vectores y matrices

Si $x \in \Re^n$ se aproxima por $x^* \in \Re^n$ $\quad E_{ab}(x) = \left\| x - x^* \right\| \quad E_r(x) = \dfrac{\left\| x - x^* \right\|}{\left\| x \right\|}$

Si $A \in \Re^{m\times n}$ se aproxima por $A^* \in \Re^{m\times n}$ $\quad E_{ab}(A) = \left\| A - A^* \right\| \quad E_r(x) = \dfrac{\left\| A - A^* \right\|}{\left\| A \right\|}$

Distancia entre vectores y entre matrices

Si $x, y \in \Re^n$

$$d(x, y) = \left\| x - y \right\|$$

Si $A, B \in \Re^{m\times n}$

$$d(A, B) = \left\| A - B \right\|$$

Propiedades de las p-normas

1. $\left\| A.B \right\|_p \leq \left\| A \right\|_p . \left\| B \right\|_p$
2. $\left\| Ax \right\|_p \leq \left\| A \right\|_p . \left\| x \right\|_p$

Algunas propiedades de las normas matriciales

Si $A \in \Re^{m\times n}$:

i) $\|A\|_2 \le \|A\|_F \le \sqrt{n}\|A\|_2$

ii) $\max_{i,j}|a_{ij}| \le \|A\|_2 \le \sqrt{n}\max_{i,j}|a_{ij}|$

iii) $\|A\|_1 = \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|$

iv) $\|A\|_\infty = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|$

v) $\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty$

vi) $\frac{1}{\sqrt{m}}\|A\|_1 \le \|A\|_2 \le \sqrt{n}\|A\|_1$

vii) $\|A\|_2 \le \sqrt{\|A\|_1\|A\|_\infty}$

Todas las normas matriciales en $\Re^{m\times n}$ son equivalentes

Límite de una sucesión de matrices, $\left\{A^{(k)}\right\}$:

$$\left\{A^{(k)}\right\} \quad \text{converge si} \quad \lim_{k\to\infty}\left\|A^{(k)} - A\right\| = 0$$

Ejercicios propuestos

- 1. Probar que las transformaciones ortogonales conservan la 2-norma vectorial:

$$\text{Si } v \in \Re^n, \text{ y } Q \in \Re^{n\times n} \text{ es ortogonal } (Q^T Q = I),$$

$$\text{se verifica que:} \qquad \|v\|_2 = \|Qv\|_2$$

- 2. Probar que las transformaciones ortogonales conservan la norma de Frobenius y la 2-norma matricial:

$$\text{Sea } A \in \Re^{m\times n}. \text{ Si } P \in \Re^{m\times m} \text{ y } Q \in \Re^{n\times n} \text{ son matrices ortogonales,}$$

$$\text{se verifica que:} \qquad i)\ \ \|A\|_F = \|PAQ\|_F \qquad\qquad ii)\ \ \|A\|_2 = \|PAQ\|_2$$