Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática

Semestre Primvavera 2007
Cálculo Numérico
MA33A-03

Profesor: Gonzalo Hernández.
Auxiliar: Gonzalo Ríos, Constanza Maturana
Fecha: 16 de Agosto

Pauta Control 1

Importante: En el control use aritmética finita de 3 cifras significativas con redondeo

1) **(50%) Representación numérica y errores**

(a) Sea $x \in \mathbb{R}$ tal que su codificación binaria inicial está dada por: $fl(x) = (-1)^{s_x} m \times 2^{-(1)^{s_z} z}$ donde $s_x = 0, s_z = 0, z = 0$. Los bits $(\alpha(k))_{k=1}^{n}$ de la mantisa se calculan con el algoritmo:

$Para\ k = 1 : n$
  $\alpha(k) = 0;$
end
$r(1) = x;$
$Para\ k = 1 : n$
  $Si\ (2^{-k} \leq r(k))$
    $\alpha(k) = 1;$
  end
  $r(k+1) = r(k) - 2^{-k} * \alpha(k);$
end

Suponiendo que para evaluar 2^{-k} se necesita una ops, determine la cantidad de ops como función de n.

Hint: Considere el caso para el cual se hacen mas operaciones.

(b) Considere el polinomio cúbico $q(x) = 3.12x^3 - 2.11x^2 + 4.01x + 10.33$ en su forma standard

i) Calcule $q(1.32)$ evaluando de la izquierda a la derecha utilizando aritmética finita de 4 cifras significativas con redondeo en cada operación que realice.
Determine el error absoluto y relativo.

ii) Calcule $q(1.32)$ evaluando de la derecha a la izquierda utilizando aritmética finita de 4 cifras significativas con redondeo en cada operación que realice.
Determine el error absoluto y relativo.

iii) Repita la parte i) utilizando la forma anidada de $q(x) = ((3.12x - 2.11)x + 4.01)x + 10.33$. Calcule la cantidad de operaciones realizadas por la forma standard y anidada.
Determine el error absoluto y relativo.

iv) Explique en detalle por que los resultados de la parte i) y ii) son diferentes.
Explique en detalle que forma de evaluar $q(x)$ es mejor.

2) **(50%) Sistemas de ecuaciones lineales**

(a) Considere el SEL:

$$\begin{aligned} \epsilon x + by &= c \\ dx + ey &= f \end{aligned} \qquad (*)$$

donde $\epsilon \approx 0$ en comparación con los reales b, c, d, e, f.

i) Resuelva el SEL $(*)$ mediante el método de Gauss exacto. Determine primero y bajo el supuesto $\epsilon \approx 0$ y luego el valor de x.

ii) Suponga ahora que:

$$\begin{aligned} f &= d + e \\ c &= \epsilon + b \end{aligned}$$

Resuelva el sistema $(*)$ pivoteando en forma parcial, es decir intercambiando las ecuaciones:

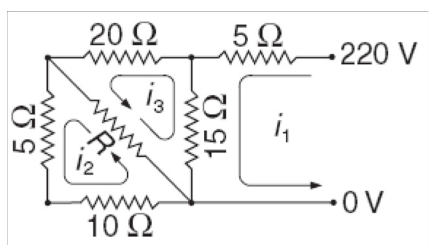
$$\begin{aligned} dx + ey &= f \\ \epsilon x + by &= c \end{aligned} \qquad (+)$$

mediante el método de Gauss exacto.

iii) Por qué la solución de los SEL $(*)$ y $(+)$ son tan diferentes ? Explique en detalle.

iv) Repita los cálculos de la parte ii) asignando valores a los parámetros ϵ, b, c, d, e, f. Resuelva ahora su SEL sin pivotear y con aritmética finita de 3 cifras significativas con redondeo. Qué observa ? Explique los resultados que obtuvo en detalle.

(b) Considere el circuito eléctrico de la figura:



Aplicando las leyes de Kirhoff y Ohm que establecen:

- La suma de todas las corrientes eléctricas en un nodo es igual a cero
- La corriente electrica que atraviesa una resistencia es igual a la diferencia de voltaje dividida por el valor de la resistencia

se obtiene el siguiente SEL:

$$\begin{aligned} 5i_1 + 15(i_1 - i_3) &= 220\,[V] \\ R(i_2 - i_3) + 5i_2 + 10i_2 &= 0 \\ 20i_3 + R(i_3 - i_2) + 15(i_3 - i_1) &= 0 \end{aligned}$$

Mediante el método de Gauss determine el valor de i_1, i_2, i_3 para $R = 5\,[\Omega]$. Puede predecir el efecto de aumentar R en un 1% ?

1) **(50%) Representación numérica y errores: Solución**

(a) $Para\ k = 1 : n$

$\quad \alpha(k) = 0;$

end

n operaciones

$r(1) = x;$

1 operacion

$Para\ k = 1 : n$

$\quad Si\ (2^{-k} \leq r(k))$ -> 2 operaciones por ciclo

$\quad\quad \alpha(k) = 1;$ -> 1 operacion por ciclo

$\quad end$

$\quad r(k+1) = r(k) - 2^{-k} * \alpha(k);\ - >$ 4 operaciones por ciclo

end

7n operaciones

TOTAL: 8n+1 operaciones

(b) $q(x) = 3.12x^3 - 2.11x^2 + 4.01x + 10.33$

i) Calcule $q(1.32)$ evaluando de la izquierda a la derecha
Potencias de x:
$1.32 \times 1.32 = 1.7424 \approx 1.74$
$1.74 \times 1.32 = 2.2968 \approx 2.3$
Multiplicaciones:
$3.12 \times 2.3 = 7.176 \approx 7.18$
$2.11 \times 1.74 = 3.6714 \approx 3.67$
$4.01 \times 1.32 = 5.2932 \approx 5.29$
Sumas:
$7.18 - 3.67 = 3.51$
$3.51 + 5.29 = 8.8$
$8.8 + 10.33 = 19.13 \approx 19.1$
Errores
$q(1.32) \approx 19.12263616$
$E_{abs} = |19.13 - 19.12263616| = 0.00736384$
$E_{rel} = \frac{|19.13 - 19.12263616|}{19.12263616} = 3.850849819 \times 10^{-4}$

ii) Calcule $q(1.32)$ evaluando de la derecha a la izquierda
Las potencias y las multiplicaciones son las mismas.
Sumas:
$10.3 + 5.29 = 15.59 \approx 15.6$
$15.6 - 3.67 = 11.93 \approx 11.9$
$11.9 + 7.18 = 19.08 \approx 19.1$
Errores
$E_{abs} = |19.08 - 19.12263616| = 0.04263616$
$E_{rel} = \frac{|19.08 - 19.12263616|}{19.12263616} = 2.229617278876 \times 10^{-3}$

iii) Repita la parte i) utilizando la forma anidada de $q(x) = ((3.12x - 2.11)x + 4.01)x + 10.33$

$3.12 \times 1.32 = 4.1184 \approx 4.12$
$4.12 - 2.11 = 2.01$
$2.01 \times 1.32 = 2.6532 \approx 2.65$
$2.65 + 4.01 = 6.66$
$6.66 \times 1.32 = 8.7912 \approx 8.79$
$8.79 + 10.33 = 19.12 \approx 19.1$
Errores
$E_{abs} = |19.12 - 19.12263616| = 0.00263616$
$E_{rel} = \frac{|19.12 - 19.12263616|}{19.12263616} = 1.3785547023 \times 10^{-4}$
Numero operaciones forma standard: 9 operaciones
Numero operaciones forma anidada: 6 operaciones

iv) Comparaciones
La diferencia entre $i)$ y $ii)$ se debe por la magnitud de los sumandos, ya que 10.33 es un orden de magnitud mayor que los demás. Si se suman los numeros pequeños primero entre si, se logra tener la misma magnitud al momento de sumar con 10.33. Si se empieza sumando con 10.33, entonces todas las sumas tienen una gran diferencia de magnitud, produciendose un mayor error. Por esta razon el error de $i)$ es menor que el error de $ii)$
La mejor forma de evaluar $q(x)$ es la forma anidada, ya que al hacer menos operaciones se tienen menos errores de redondeo, obteniéndose un resultado más exacto.

2) **(50%) Sistemas de ecuaciones lineales: Solución**

(a) Pivoteo parcial en SEL

i) $\begin{bmatrix} \varepsilon & b & | & c \\ d & e & | & f \end{bmatrix} \Rightarrow f_2 - \frac{d}{\varepsilon} f_1 \to f_2 \Rightarrow \begin{bmatrix} \varepsilon & b & | & c \\ 0 & e - \frac{d}{\varepsilon} b & | & f - \frac{d}{\varepsilon} c \end{bmatrix}$

$\Rightarrow \quad y = \frac{f - \frac{d}{\varepsilon} c}{e - \frac{d}{\varepsilon} b} = \frac{\varepsilon f - dc}{e\varepsilon - db} \quad \text{como} \quad \varepsilon \approx 0 \Rightarrow y = \frac{c}{b}$

Luego, reemplazando en la primera ecuación:

$\Rightarrow \varepsilon x + by = c \Rightarrow x = \frac{c - by}{\varepsilon} \approx \frac{c - b\frac{c}{b}}{\varepsilon} = 0$

Por lo tanto, la solución del SEL $(*)$ es:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{c}{b} \end{pmatrix}$$

Notar que esta solución es válida $\forall b, c, d, e, f, \varepsilon$ que cumplan que $\varepsilon \approx 0$ en comparación con b, c, d, e, f.

ii) En particular, escojamos:

$$\begin{aligned} f &= d + e \\ c &= \varepsilon + b \end{aligned}$$

y resolvamos el SEL $(+)$, i.e. permutando las ecuaciones:

$\begin{bmatrix} d & e & | & f \\ \varepsilon & b & | & c \end{bmatrix} = \begin{bmatrix} d & e & | & d+e \\ \varepsilon & b & | & \varepsilon + b \end{bmatrix}$

$\Rightarrow f_2 - \frac{\varepsilon}{d} f_1 \Rightarrow \begin{bmatrix} d & e & | & d+e \\ 0 & b - \frac{\varepsilon}{d} e & | & (\varepsilon + b) - \frac{\varepsilon}{d}(d+e) \end{bmatrix} = \begin{bmatrix} d & e & | & d+e \\ 0 & b - \frac{\varepsilon}{d} e & | & b - \frac{\varepsilon}{d} e \end{bmatrix}$

$\Rightarrow y = \frac{b-\frac{\varepsilon}{d}e}{b-\frac{\varepsilon}{d}e} = 1$

Reemplazando en la primera acuación:

$\Rightarrow dx + ey = d + e \Rightarrow x = \frac{d+e-e}{d} = 1$

Pro lo tanto, la solución del SEL (+) es :

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

iii) Al calcular la solución del SEL ($*$) tomamos como pivote un número muy pequeño, lo que causa mucha disparidad en las magnitudes c y $f - \frac{d}{\varepsilon}c$ introdujendo un error de redondeo que se propaga al despejar y.

Este error no se considera en el SEL (+) ya que tomamos como pivote el mayor número de la columna.

iv) Consideremos el sistema:

$$\begin{aligned} 0.003x + 59.1y &= 59.103 \\ 5.29x + 6.1y &= 11.39 \end{aligned}$$

Al redondear a 3 cifras cignificativas:

$$\begin{aligned} 3 \times 10^{-3}x + 59.1y &= 59.1 \\ 5.29x + 6.1y &= 11.4 \end{aligned}$$

Para repetir los calculos realizados en la parte (ii) antes de empezar a pivotear debemos permutar las ecuaciones, obteniendo la siguiente matriz ampliada:

$$\begin{bmatrix} 5.29 & 6.1 & | & 11.4 \\ 3 \times 10^{-3} & 59.1 & | & 59.1 \end{bmatrix}$$

Nos damos cuenta que la solución del sistema es :

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Si permutamos las filas y comenzamos a pivotear como en la parte (i) :

$\begin{bmatrix} 3 \times 10^{-3} & 59.1 & | & 59.1 \\ 5.3 & 6.1 & | & 11.4 \end{bmatrix} \Rightarrow f_2 - \frac{5.3}{3\times10^{-3}} f_1 \to f_2$

$\Rightarrow \begin{bmatrix} 3 \times 10^{-3} & 59.1 & | & 59.1 \\ 0 & 6.1 - \frac{5.3}{3\times10^{-3}}59.1 & | & 11.4 - \frac{5.3}{3\times10^{-3}}59.1 \end{bmatrix} = \begin{bmatrix} 3 \times 10^{-3} & 59.1 & | & 59.1 \\ 0 & 1.04 \times 10^{5} & | & 1.04 \times 10^{5} \end{bmatrix}$

$\Rightarrow \; y = 1$

Luego, de la primera ecuación:

$\Rightarrow 3 \times 10^{-3}x + 59.1y = 59.1 \Rightarrow x = \frac{59.1-59.1}{3\times10^{-3}} = 0$

Por lo tanto, la solución de SEL es:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(b) Consideremos el siguiente SEL:

$$\begin{aligned} 5i_1 + 15(i_1 - i_3) &= 220\,[V] \\ R(i_2 - i_3) + 5i_2 + 10i_2 &= 0 \\ 20i_3 + R(i_3 - i_2) + 15(i_3 - i_1) &= 0 \end{aligned}$$

Escrito matricialmente:

$$\begin{bmatrix} 20 & 0 & -15 \\ 0 & R+15 & -R \\ -15 & -R & R+35 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} = \begin{bmatrix} 220 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 20 & 0 & -15 & | & 220 \\ 0 & R+15 & -R & | & 0 \\ -15 & -R & R+35 & | & 0 \end{bmatrix} \Rightarrow f_3 + \frac{3}{4} f_1 \to f_3 \Rightarrow \begin{bmatrix} 20 & 0 & -15 & | & 220 \\ 0 & R+15 & -R & | & 0 \\ 0 & -R & R+23.7 & | & 165 \end{bmatrix}$$

$$\Rightarrow f_3 + \frac{R}{(R+15)} f_2 \to f_3 \Rightarrow \begin{bmatrix} 20 & 0 & -15 & | & 220 \\ 0 & R+15 & -R & | & 0 \\ 0 & 0 & \frac{38.7R+356}{R+15} & | & 165 \end{bmatrix}$$

$$\Rightarrow \begin{matrix} i_3 = \frac{165}{38.7R+356}(R+15) \\ i_2 = \frac{1}{20}\frac{165R}{38.7R+356}(R+15) \\ i_1 = (220 + 15\frac{165}{38.7R+356}(R+15))\frac{1}{20} \end{matrix} \qquad (1)$$

Reemplazando con $R = 5$:

$$\begin{matrix} i_3 = 6 \\ i_2 = 1.5 \\ i_1 = 15.5 \end{matrix}$$

Para saber como varia la solucion si aumentamos R a 5.05 tenemos dos opciones:

Calcular el condicionamiento de la matriz

Reemplazar $R = 5.05$ en (1). Claramente el primer camino es mas engorroso, pues debemos calcular la inversa de la matriz

Llamemos M a la matriz del sistema, entonces:

$$M^{-1} = \begin{bmatrix} 0.0705 & 0.0068 & 0.0273 \\ 0.0068 & 0.0523 & 0.0091 \\ 0.0273 & 0.0091 & 0.0364 \end{bmatrix}$$

$\Rightarrow \|M\|_\infty = 60 \quad \text{y} \quad \left\|M^{-1}\right\|_\infty = 0.105$

$\Rightarrow Cond(M) = 60 \cdot 0.105 = 6.3$

Como el condicionamiento de la matriz no es muy lejano a 1 la solución no debería variar mucho con un incremento de 1% en R.

Ahora, reemplacemos $R = 5.05$ en (1) $\Rightarrow$

$$\begin{matrix} i_3 = 0.476 \\ i_2 = 1.5 \\ i_1 = 11.4 \end{matrix}$$

Los cambios en la solución del SEL no son significativos debido al condicionamiento de la matriz.