Computación Numérica

Primer Parcial A - Mayo 2014

Una máquina almacena números en punto flotante en 11 bits. El primer bit se usa para el signo del número, los cinco siguientes para el exponente sesgado y los últimos cinco bits para la mantisa. Si se sigue un criterio similar al de la norma IEEE 754:

1. Calcular el número $(11011001011)_2$ en base 10.

$$signo$$
 1
 $exponente$ 10110
 $mantisa$ 01011

Cálculo del exponente

El exponente sesgado es $1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 22$. Si el número de bits del exponente es m = 5, entonces el $sesgo = 2^{m-1} - 1 = 15$ y el exponente es

$$22 - sesgo = 22 - 15 = 7$$

Cálculo de la mantisa

Teniendo en cuenta el bit escondido, la mantisa es (1),01011 que en base 10 es

$$1 + 2^{-2} + 2^{-4} + 2^{-5} = 1.34375$$

Número

Como el bit del signo es 1 el número es negativo. Teniendo en cuenta la mantisa y el exponente, el número en base 10 es

$$-(1+2^{-2}+2^{-4}+2^{-5})\times 2^7 = -172$$

2. ¿Cual sería el ϵ de máquina?

Formato decimal

El número 1 teniendo en cuenta el bit escondido sería (1),00000 \times 20. El ϵ de máquina es el número más pequeño que se puede alinear con este 1 en este formato para sumárselo:

$$\begin{cases} 1 = 1,00000 & \times 2^{0} \\ \epsilon = 0,00001 & \times 2^{0} = 1,000000 \times 2^{-5} = 2^{-5} = 0,03125 \end{cases}$$

Formato binario

Como el número es $1,000000 \times 2^{-5}$ la mantisa es (1),00000. El exponente sesgado será $-5 + sesgo = -5 + 15 = 10 = 2^3 + 2^1$ y que en binario será 01010. Por lo tanto.

$$\epsilon = 0\,01010\,000000$$

3. ¿Cual sería el mayor real desnormalizado?

Formato decimal

Si se almacena en forma desnormalizada:

- El exponente en binario es 00000.
- Se asume como exponente el menor posible de los normalizados.
- Se asume que el bit escondido es 0.

El menor exponente posible es, sesgado, 00001 y sin sesgar 1 - sesgo = 1 - 15 = -14. Por lo tanto, el mayor número desnormalizado es $(0),11111 \times 2^{-14}$ que en decimal es

$$(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5}) \times 2^{-14} = 0,0000591278$$

Formato binario

Ya se ha dicho que el exponente se escribe 00000 y como la mantisa es (0),111111, teniendo en cuenta el signo, se escribirá

4. Calcular el número $(11,25)_{10}$ en este formato de 11 bits. Cálculo de la mantisa

Para convertir a binario la parte decimal, multiplicamos por 2, le quitamos la parte entera, que será nuestro dígito binario y repetimos el proceso.

 $\begin{array}{cccc} Decimal: & 0.25 & 0.5 & 0 \\ Entera: & & 0 & 1 \end{array}$

y tomamos los dígitos:

0,01

Para convertir a binario la parte entera 11 dividimos de forma reiterada por 2 y guardamos los restos:

cocientes: 11 5 2 1restos: 1 1 0

empezando por el último cociente, seguimos con los restos en orden inverso:

1011

que almacenamos

$$1011,01 = (1),01101 \times 2^3$$

y no hace falta almacenar el primer uno. Por lo tanto la mantisa se almacena como

01101

Cálculo del exponente

El sesgo = 15. El exponente sesgado será $exponente + sesgo = 3 + 15 = 18 = 2^4 + 2^1$, que en binario es 10010.

N	 m	 	

signo	exponente	mantisa	
0	10010	01101	

Primer Parcial B - Mayo 2014

Sea la ecuación

$$h(t) = t^3 - 6t^2 + t + \ln t$$

1. Demostrar que esta función tiene un único extremo relativo en [3,4].

Para que el punto α será un extremo de h es necesario que $h'(\alpha) = 0$. Como

$$h'(x) = 3t^2 - 12t + 1 + \frac{1}{t} = \frac{3t^3 - 12t^2 + t + 1}{t}$$

si tomamos $f(x) = 3t^3 - 12t^2 + t + 1$ entonces

h tiene un extremo en $[3,4] \iff f$ tiene una raíz en [3,4].

O lo que es lo mismo

$$h'(x) = 0$$
 en $[3, 4] \iff f(x) = 0$ en $[3, 4]$.

Las condiciones (suficientes, no necesarias) que ha de cumplir f en [3,4] para que exista una raíz en el intervalo son:

- a) f continua: f es continua porque es la suma de funciones continuas, el polinomio lo es siempre y $\ln x$ es continua en todo su dominio.
- b) f tiene distinto signo en los extremos del intervalo: f(3) = -23 y f(4) = 5
- c) f' > 0 o f' < 0 en (3, 4):

$$f'(t) = 9t^2 - 24t + 1$$

Si calculamos las raíces y factorizamos:

$$9t^2 - 24t + 1 = 9(t - 0.042)(t - 2.62)$$

Como (t-0.042) es siempre positivo en (3,4) y (t-2.62) también $9t^2-24t+1>0$ en (3,4).

Por lo tanto f' > 0 en (3, 4).

2. ¿Se puede calcular por el método de bisección partiendo de dicho intervalo?¿Por qué?

Si porque se cumplen las condiciones necesarias, que son las condiciones a) y b) de la pregunta anterior.

3. Aproximar el extremo haciendo dos iteraciones con el método de bisección.

k	a	c	b	f(a)	f(b)	f(c)	cota de error
0	3		4	-23		5	4-3=1
1	3	(a+b)/2 = (3+4)/2 = 3.5	4	-23	-13.9	5	3.5 - 3 = 0.5
2	3.5	(a+b)/2 = (3.5+4)/2 = 3.75	4	-13.9	-5.8	5	4-3.75=0.25

Y podemos dar como raíz aproximada 3,75 (la raíz verdadera es 3,89236 \in [3,75,4]).

4. Dar una cota del error.

La raíz está en el intervalo [3,75,4] por lo tanto la cota de error es 4-3,75=0,25

5. Aproximar dicho extremo haciendo dos iteraciones con el método de Newton con $x_0 = 4$.

$$t_{k+1} = t_k - \frac{f(t_k)}{f'(t_k)} = t_k - \frac{3t_k^3 - 12t_k^2 + t_k + 1}{9t_k^2 - 24t_k + 1} = \frac{6t_k^3 - 12t_k^2 - 1}{9t_k^2 - 24t_k + 1}$$

k	$t_{k+1} = \frac{6t_k^3 - 12t_k^2 - 1}{9t_k^2 - 24t_k + 1}$
0	4
1	3,89796
2	3,89238

(la raíz verdadera es 3,89236).

Computación Numérica

Segundo Parcial - Mayo 2014

1. Dados los nodos:

x	0	1	2
y	1	0	2

Si escribimos el spline natural que los ajusta como

$$s(x) = \begin{cases} s_1(x) = a(x-1)^3 + b(x-1)^2 + c(x-1) + d & \text{si } x \in [0,1] \\ s_2(x) = e(x-1)^3 + f(x-1)^2 + g(x-1) + h & \text{si } x \in [1,2] \end{cases}$$

plantear las ecuaciones del sistema cuya solución son los coeficientes a, b, c, d, e, f, g y h. Especificar las condiciones aplicadas.

Se tienen que cumplir las siguientes condiciones:

Condiciones de interpolación: la curva ha de pasar por los tres puntos. Por lo tanto:

$$s_1(0) = 1$$
, $s_1(1) = 0$, $s_2(1) = 0$, $s_2(2) = 2$.

Condiciones de continuidad: han de coincidir las derivadas primera y segunda en los puntos intermedios:

$$s_1'(1) = s_2'(1), \ s_1''(1) = s_2''(1).$$

Y como hay 8 incógnitas y de momento sólo tenemos 6 condiciones (ecuaciones) imponemos dos más en los extremos. Como el spline es natural las condiciones adicionales son:

$$s_1''(0) = 0, \ s_2''(2) = 0.$$

Calculamos

$$s'(x) = \begin{cases} s'_1(x) = 3a(x-1)^2 + 2b(x-1) + c & \text{si } x \in [0,1] \\ s'_2(x) = 3e(x-1)^2 + 2f(x-1) + g & \text{si } x \in [1,2] \end{cases}$$
$$s''(x) = \begin{cases} s''_1(x) = 6a(x-1) + 2b & \text{si } x \in [0,1] \\ s''_2(x) = 6e(x-1) + 2f & \text{si } x \in [1,2] \end{cases}$$

y las ecuaciones son:

1.
$$s_1(0) = 1 \Longrightarrow -a + b - c + d = 1$$

$$2. \ s_1(1) = 0 \Longrightarrow d = 0$$

3.
$$s_2(1) = 0 \Longrightarrow h = 0$$

4.
$$s_2(2) = 2 \Longrightarrow e + f + g + h = 2$$

5.
$$s'_1(1) = s'_2(1) \Longrightarrow c = g$$

6.
$$s_1''(1) = s_2''(1) \Longrightarrow 2b = 2f$$

7.
$$s_1''(0) = 0 \Longrightarrow -6a + 2b = 0$$

8.
$$s_2''(2) = 0 \Longrightarrow 6e + 2f = 0$$

Y tenemos un sistema lineal de ocho ecuaciones para calcular ocho incógnitas.

2. La ley de Hooke es F(x) = kx donde F es la fuerza aplicada al resorte, k la constante del muelle y x la variación en la longitud del resorte.

x (mm)	8	15	23	30
Fuerza (g)	5.8	11.6	17.4	23.2

Calcular la recta que pasa por el origen y que minimiza los errores cuadráticos. Deducir la fórmula empleada para calcular la recta a partir de los errores cuadráticos. Calcular la constante k de este muelle. La recta que pasa por el origen será de la forma:

$$F(x) = kx$$

Queremos calcular la recta que minimiza la suma de los errores cuadráticos:

$$E(k) = \sum_{k=1}^{4} (F(x_k) - y_k)^2 = \sum_{k=1}^{4} (k \times x_k - y_k)^2.$$

O lo que es lo mismo:

$$E(k) = (F(8) - 5.8)^{2} + (F(15) - 11.6)^{2} + (F(23) - 17.4)^{2} + (F(30) - 23.2)^{2} = (k \times 8 - 5.8)^{2} + (k \times 15 - 11.6)^{2} + (k \times 23 - 17.4)^{2} + (k \times 30 - 23.2)^{2}$$

Para hallar el error mínimo derivamos:

$$E'(k) = \sum_{k=1}^{4} 2(k \times x_k - y_k) x_k = \sum_{k=1}^{4} 2(k \times x_k^2 - y_k x_k) = 0$$
$$k \sum_{k=1}^{4} x_k^2 - \sum_{k=1}^{4} y_k x_k = 0$$

Y por lo tanto:

$$k = \frac{\sum_{k=1}^{4} y_k x_k}{\sum_{k=1}^{4} x_k^2}$$

O lo que es lo mismo:

$$E'(k) = 2(k \times 8 - 5.8) \times 8 + 2(k \times 15 - 11.6) \times 15 + 2(k \times 23 - 17.4) \times 23 + 2(k \times 30 - 23.2) \times 30$$
$$= 2k(8 \times 8 + 15 \times 15 + 23 \times 23 + 30 \times 30) - 2(8 \times 5.8 + 15 \times 11.6 + 23 \times 17.4 + 30 \times 23.2) = 0$$

Y por lo tanto:

$$k = \frac{8 \times 5.8 + 15 \times 11.6 + 23 \times 17.4 + 30 \times 23.2}{8 \times 8 + 15 \times 15 + 23 \times 23 + 30 \times 30} = \frac{1316.6}{1718} = 0.766 \text{gr/mm}$$

Y la recta de aproximación es

$$F(x) = 0.766x$$

3. Supongamos que tenemos tres puntos (x_0, y_0) , (x_1, y_1) y (x_2, y_2) de una función f, de forma que $x_1 = x_0 + h$ y $x_2 = x_0 + 2h$ con 0 < h < 1. Construir una fórmula de derivación de tipo interpolatorio que aproxime $f'(x_0)$ o $f'(x_2)$ y que utilice estos tres puntos.

Si usamos un polinomio de interpolación de f de 2^o grado en x_0 , x_1 , x_2 , siendo $y_j = f(x_j)$, el polinomio de interpolación en la forma de Newton es:

$$p(x) = [y_0] + [y_0, y_1](x - x_0) + [y_0, y_1, y_2](x - x_0)(x - x_1)$$

con

$$[y_0] = y_0, \quad [y_0, y_1] = \frac{y_0 - y_1}{x_0 - x_1} = \frac{y_1 - y_0}{h},$$

$$[y_0, y_1, y_2] = \frac{[y_0, y_1] - [y_1, y_2]}{x_0 - x_2} = \frac{1}{2h} \left(\frac{y_0 - y_1}{h} - \frac{y_1 - y_2}{h} \right) = \frac{1}{2h^2} \left(y_0 - 2y_1 + y_2 \right).$$

$$[y_0] = y_0, \quad [y_0, y_1] = \frac{y_1 - y_0}{h}, \quad [y_0, y_1, y_2] = \frac{1}{2h^2} (y_0 - 2y_1 + y_2).$$

Si derivamos p(x)

$$f'(x) \approx p'(x) = [y_0, y_1] + [y_0, y_1, y_2]((x - x_0) + (x - x_1)),$$

y para el caso particular del punto x_0

$$f'(x_0) \approx p'(x_0) = [y_0, y_1] + [y_0, y_1, y_2]((x_0 - x_0) + (x_0 - x_1)),$$

$$f'(x_0) \approx \frac{y_1 - y_0}{h} + \frac{1}{2h^2} (y_0 - 2y_1 + y_2)(-h),$$

$$f'(x_0) \approx \frac{-3y_0 + 4y_1 - y_2}{2h}.$$

Y si aproximamos la derivada en el punto x_2

$$f'(x_2) \approx p'(x3h_2) = [y_0, y_1] + [y_0, y_1, y_2]((x_2 - x_0) + (x_2 - x_1)),$$

$$f'(x_1) \approx \frac{y_1 - y_0}{h} + \frac{1}{2h^2} (y_0 - 2y_1 + y_2)(2h + h),$$

$$f'(x_1) \approx \frac{y_0 - 4y_1 + 3y_2}{2h}.$$

Computación Numérica

Tercer Parcial - Mayo 2014

1. Sea el sistema Ax = b donde

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 5 & -1 \\ 0 & -1 & 3 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ -3 \end{pmatrix}$$

a) Calcular la matriz de iteración de Gauss-Seidel B_{G-S} .

La matriz $B_{G-S} = -(L+D)^{-1}U$ con

$$A = L + D + U = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix} + \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

Por lo que

$$L+D = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 0 & -1 & 3 \end{pmatrix} \quad y \quad U = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

Calculemos por Gauss-Jordan $(L+D)^{-1}$

Por lo tanto

$$(L+D)^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2/5 & 1/5 & 0 \\ -2/15 & 1/15 & 1/3 \end{pmatrix}$$

У

$$B_{G-S} = -(L+D)^{-1}U = -\begin{pmatrix} 1 & 0 & 0 \\ -2/5 & 1/5 & 0 \\ -2/15 & 1/15 & 1/3 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -2 & 0 \\ 0 & 4/5 & 1/5 \\ 0 & 4/15 & 1/15 \end{pmatrix}$$

b) Calcular la norma infinito de la matriz de iteración de Gauss-Seidel B_{G-S} . Partiendo de este resultado ¿qué podemos decir de la convergencia del método de Gauss-Seidel?

$$B_{G-S} = \begin{pmatrix} 0 & -2 & 0 \\ 0 & 4/5 & 1/5 \\ 0 & 4/15 & 1/15 \end{pmatrix} \begin{pmatrix} 0+2+0=2 \\ 0+4/5+1/5=1 \\ 0+4/15+1/15=1/3 \end{pmatrix}$$

$$||B_{G-S}||_{\infty} = \text{Max}(2, 1, 1/3) = 2$$

Como $||B_{G-S}||_{\infty} < 1$ es condición suficiente, no necesaria, de convergencia, no podemos asegurar ni que converge ni que no converge.

c) Calcular la norma uno de B_{G-S} . Partiendo de este resultado ¿qué podemos decir de la convergencia del método de Gauss-Seidel?

$$B_{G-S}^{T} = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 4/5 & 4/15 \\ 0 & 1/5 & 1/15 \end{pmatrix} \begin{pmatrix} 0+0+0=0 \\ 0+4/5+1/5=46/15 \\ 0+1/5+1/15=4/15 \end{pmatrix}$$

$$||B_{G-S}||_1 = \text{Max}(0, 46/15, 4/15) = 46/15$$

Como $||B_{G-S}||_1 < 1$ es condición suficiente, no necesaria, de convergencia, no podemos asegurar ni que converge ni que no converge.

d) Calcular los autovalores de B_{G-S} . Partiendo de este resultado ¿qué podemos decir de la convergencia del método de Gauss-Seidel?

$$\begin{vmatrix} 0 - \lambda & -2 & 0 \\ 0 & 4/5 - \lambda & 1/5 \\ 0 & 4/15 & 1/15 - \lambda \end{vmatrix} = 0$$

$$-\lambda \left(\frac{4}{5} - \lambda\right) \left(\frac{1}{15} - \lambda\right) - \frac{4}{15} \frac{1}{5} \left(-\lambda\right) = \frac{13}{15} \lambda^2 - \lambda^3 = (\lambda) \left(\lambda\right) \left(\frac{13}{15} - \lambda\right) = 0$$

Y los autovalores son

$$\lambda_{1,2} = 0, \quad \lambda_3 = \frac{13}{15}$$

La condición necesaria y suficiente para que el método de Gauss-Seidel converja es que todos los autovalores, en valor absoluto, sean menores que 1. Como esta condición se cumple, el método converge.

e)Realizar una iteración con Gauss-Seidel. Utilizar $\mathbf{x}^{(0)}=\mathbf{0}.$

El sistema es

$$\begin{array}{ccccc} x & +2y & = & 1 \\ 2x & +5y & -z & = & 3 \\ & -y & +3z & = & -3 \end{array}$$

En la primera ecuación despejamos la primera incógnita, x, en la segunda, la segunda incógnita, y, y finalmente, z.

$$x = 1 - 2y$$

 $y = (3 - 2x + z)/5$
 $z = (-3 + y)/3$

Realizamos una iteración

$$\begin{array}{lll} x^{(1)} & = & 1-2y^{(0)}=1-2(0)=1 \\ y^{(1)} & = & \left(3-2x^{(1)}+z^{(0)}\right)/5 = \left(3-2(1)+(0)\right)/5 & = 0,2 \\ z^{(1)} & = & \left(-3+y^{(1)}\right)/3 = \left(-3+0,2\right)/3 = -2,8 \end{array}$$

2. Utilizando el método del gradiente, aproximar el mínimo de la función $f(x,y) = x^4 + y^4$. Utilizar el punto (1,1) como punto inicial y realizar una iteración.

Realizaremos una iteración por el método del gradiente usando la fórmula

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - h\nabla f_{(x_0, y_0)} \tag{1}$$

y buscaremos h para que

$$q(h) = f(x_0 - hf_x(1, 1), y_0 - hf_y(1, 1))$$

tenga un valor mínimo.

Tenemos que

$$\nabla f = (f_x, f_y) = (4x^3, 4y^3)$$

у

$$\nabla f_{(x_0,y_0)} = \nabla f_{(1,1)} = (4,4)$$
.

$$g(h) = f(x_0 - hf_x(1, 1), y_0 - hf_y(1, 1)) = f(1 - 4h, 1 - 4h) = (1 - 4h)^4 + (1 - 4h)^4 = 2(1 - 4h)^4$$

Optimicemos esta función:

$$g'(h) = 8(1 - 4h)^3 (-4) = 0 \Longrightarrow 1 - 4h = 0 \Longrightarrow h = \frac{1}{4}$$

y usando este valor en la ecuación (1)

$$\left(\begin{array}{c} x_1 \\ y_1 \end{array}\right) = \left(\begin{array}{c} 1 \\ 1 \end{array}\right) - \frac{1}{4} \left(\begin{array}{c} 4 \\ 4 \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right)$$

Hemos llegado al mínimo porque $f(x_0, y_0) = f(1, 1) = 2$ y $f(x_1, y_1) = f(0, 0) = 0$ que es el menor valor posible puesto que f es siempre positivo.

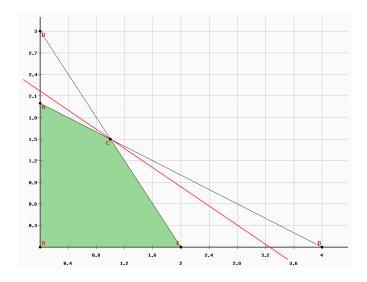


Figura 1: Solución gráfica del problema 3.

3. Maximizar la función f(x,y) = 2x + 3y con las restricciones

$$x + 2y \le 4$$
, $3x + 2y \le 6$, $x \ge 0$, $y \ge 0$

a) Geométricamente.

La región de posibles soluciones es la región verde. Una curva de nivel de f es la línea roja. La función crece hacia la derecha. El punto donde la función es máxima es B(1,1,5) y en él la función vale f(1,1,5)=6,5.

b) Mediante el algoritmo simplex.

Reescribimos las condiciones

$$F - 2x - 3y = 0$$
$$x + 2y + r = 4$$
$$3x + 2y + s = 6$$

y entonces la tabla es:

En la primera fila el valor negativo con mayor valor absoluto es -3 y, de los valores de esta columna debajo de -3 en la segunda fila el Valor/y=4/2=2 y en la tercera 6/2=3. Elejimos la segunda fila como ecuación pivotal porque de los dos valores 2 es menor que 3. Por lo tanto el pivote está en la segunda columna, segunda fila. Realizamos las operaciones por fila

y la matriz se trasforma en

En la primera fila queda el valor negativo -1/2 y, de los valores de esta columna en la segunda fila tenemos el Valor/y=2/(1/2)=4 y en la tercera 2/2=1. Elejimos la tercera fila como ecuación pivotal porque de los dos valores 2 es menor que 3. Por lo tanto el pivote está en la segunda columna, segunda fila. Realizamos las operaciones por fila

\overline{F}	x	y	r	s	Valor
1	0	0	5/2	1/4	13/2
0	0	1	3/4	-1/4	3/2
0	1	0	-1/2	1/2	1

Ya no hay valores negativos en la primera fila y hemos acabado. En la columna de F el valor a la altura del 1 es 13/2 = 6.5, en la columna de y el valor en la fila del 1 es 3/2 = 1.5 y en la columna de x, 1. Resumiendo, el máximo está en (x,y) = (1,1.5) y el valor del máximo es f(1,1.5) = 6.5.