Universidad Nacional de Ingeniería
Facultad de Ciencias
Escuela Profesional de Matemática Ciclo 2021–I

EXAMEN PARCIAL DE ANÁLISIS Y MODELAMIENTO NUMÉRICO I CM4F1 A/B

---

1. Determine si cada proposición es verdadero o falso. Justifique su respuesta.

   (a) (0.4 ptos) Let A be the matrix:

   $$A = \begin{pmatrix} 2 & 1 \\ 2 & -1 \\ 1 & 0 \end{pmatrix}.$$

   The least squares solution of the problem

   $$Ax = b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{is} \quad x = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \end{pmatrix}.$$

   (b) (0.4 ptos) A Householder matrix H is Hermitian or Unitary.

   (c) (0.4 ptos) Let $\rho(A)$ be the spectral radius of matrix A. If $\rho(A) = 0$ then $A = 0$.

   (d) (0.4 ptos) Given two vectors x and y, the conditions on these vectors such that one can find a Householder matrix H satisfying $Hx = y$ are:

   $$\|x\|_\infty = \|y\|_\infty \quad \text{and} \quad \langle x, y\rangle \in \mathbb{R}.$$

   (e) (0.4 ptos) The binary expansion of 9.75 is 1001.11

   (f) (0.4 ptos) Consider the following linear system:

   $$\begin{aligned} x_1 + x_2 + 3x_3 &= 5 \\ x_1 + 3x_2 + x_3 &= 2 \\ 3x_1 + x_2 + x_3 &= -4 \end{aligned}$$

   Re-ordering the equations the Jacobi method will converge to the exact solution.

   (g) (0.4 ptos) In $\mathbb{F}(10, 3, L, U)$ with rounding the sum -0.0321 + 0.000136 is equal to -0.032.

   (h) (0.4 ptos) When a matrix A has their rows linearly independent then there exists its QR descomposition.

   (i) (0.4 ptos) For a general $n \times n$ matrix the Gaussian elimination takes $\mathcal{O}(n^3/3)$ operations.

(j) (0.4 ptos) Consider the system $Ax = b$. If $\rho(A) < 1$ then the Gauss-Seidel method converge to exact solution.

Solución: La clave de respuestas es:

(a) **V**.

(b) **F**.

(c) **F**.

(d) **F**.

(e) **V**.

(f) **V**.

(g) **V**.

(h) **F**.

(i) **V**.

(j) **F**.

2. Decida si la proposición es verdadera o falsa. Justifique adecuadamente su respuesta.

(a) (1 pto) In fixed point arithmetic $a + (b + c) = (a + b) + c$.

(b) (1 pto) Al representar el número $\sqrt{2} \approx 1.4142$ en base 10 en un computador que tiene 2 bits para la parte entera y tres para la parte decimal nos da un error absoluto de 0.0392. (Use truncamiento de ser necesario).

(c) (1 pto) Let A will be $m \times n$ matrix with $m \geq n$. Then the formula for the solution x of a linear system $Ax = b$ that minimizes $\|Ax - b\|_2$ using the QR descomposition of the matrix $A = QR$ is $x = R^{-1}Q^T b$. (Use the fact that the vectors $Q(Rx - Q^T b)$ and $(I - QQ^T)b$ are orthogonal).

(d) (1 pto) Consider the linear system:

$$\begin{aligned} x_1 + \frac{1}{2}x_2 &= \frac{3}{2} \\ \frac{1}{2}x_1 + \frac{1}{3}x_2 &= \frac{5}{6} \end{aligned}$$

Using 2-digit chopping arithmetic and the vector $x = (1.4, 0.35)$ is proposed as an approximate solution. Then the corresponding residual vector is:

$$r = \begin{pmatrix} -0.1 \\ 0.1 \end{pmatrix}.$$

and in the $\|\cdot\|_\infty$, the condition number of the coeffient matrix for the system is 15.

Solución:

(a) Considere un sistema de punto fijo de 1 bit para la parte entera y 4 bits para la parte decimal. Considere también los siguientes números:

$$a = 9.1, \quad b = 0.9999, \quad c = -0.9999$$

luego:

$$a + (b + c) = 9.1000 + 0.0000 = 9.1000$$

pero:

$$(a+b)+c = 10.0999 - 0.9999$$

observe que $a+b$ no puede representarse en el sistema de punto fijo dado. Por tanto la suma $(a+b)+c$ no puede realizarse.

(b) El computador usa 2 dígitos para la parte entera y 3 bits para la parte fraccionaria. Luego, convertimos $\sqrt{2} \approx 1.4142$ a base 2:

$$\begin{array}{rcl} 1.4142 & = & 1 + 0.4142 \\ 2(0.4142) & = & 0.8284 \Rightarrow d_1 = 0 \\ 2(0.8284) & = & 1.6568 \Rightarrow d_2 = 1 \\ 2(0.6568) & = & 1.3136 \Rightarrow d_3 = 1 \\ 2(0.3136) & = & 0.6272 \Rightarrow d_4 = 0 \\ 2(0.6272) & = & 1.2544 \Rightarrow d_5 = 1 \end{array}$$

luego: $1.4142 = 1.01101_{(2)}$ cuya representación en el computador dado es:

$$1.011$$

que al ser expresado en base 10 resulta:

$$1.011 = 1 + 0 + \frac{1}{4} + \frac{1}{8} = \frac{11}{8} = 1.375.$$

Por tanto, el error absoluto cometido es:

$$EA = 1.4142 - 1.375 = 0.0392,$$

es decir, la proposición es verdadera.

(c) Para resolver el sistema $Ax = b$ debemos resolver:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2.$$

Observe:

$$\begin{array}{rcl} Ax - b & = & (QR)x - b \\ & = & (QR)x - (QQ^T + I - QQ^T)b \\ & = & Q(Rx - Q^T b) - (I - QQ^T)b \end{array}$$

observe que los vectores $Q(Rx - Q^T b), (I - QQ^T)b$ son ortogonales, por tanto se cumple el teorema de Pitágoras:

$$\begin{array}{rcl} \|Ax - b\|_2^2 & = & \|Q(Rx - Q^T b) - (I - QQ^T)b\|_2^2 \\ & = & \|Q(Rx - Q^T b)\|_2^2 + \|(I - QQ^T)b\|_2^2 \\ & = & \|(Rx - Q^T b)\|_2^2 + \|(I - QQ^T)b\|_2^2 \end{array}$$

donde la última igualdad se da debido a que Q es una matriz ortogonal. Luego, la suma de cuadrados es mínimo cuando:

$$\|(Rx - Q^T b)\|_2^2 = 0 \Rightarrow Rx = Q^T b.$$

Así la proposición es verdadera.

(d) La matriz asociada al sistema es:

$$A = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}.$$

Dado $x = (1.4, 0.35)$, el vector residual correspondiente es:

$$r = b - Ax = \begin{pmatrix} \frac{3}{2} \\ \frac{5}{6} \end{pmatrix} - \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1.4 \\ 0.35 \end{pmatrix}$$

Nos dicen usar aritmética con truncamiento a dos dígitos, por tanto resulta:

$$r = \begin{pmatrix} 1.5 \\ 0.83 \end{pmatrix} - \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.33 \end{pmatrix} \begin{pmatrix} 1.4 \\ 0.35 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.83 \end{pmatrix} - \begin{pmatrix} 1.4 + 0.175 \\ 0.7 + 0.1155 \end{pmatrix}$$

con aritmética truncando al segundo dígito se obtiene:

$$r = \begin{pmatrix} 1.5 \\ 0.83 \end{pmatrix} - \begin{pmatrix} 1.4 + 0.17 \\ 0.7 + 0.11 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.83 \end{pmatrix} - \begin{pmatrix} 1.5 \\ 0.81 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.02 \end{pmatrix}$$

que truncando a dos dígitos se tiene:

$$r = \begin{pmatrix} 0 \\ 0.02 \end{pmatrix}$$

Observe que:

$$A^{-1} = \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix}$$

luego:

$$K_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty = \frac{3}{2}(18) = 27.$$

Por tanto, la proposición es falsa.

3. (4 ptos) Transform the matrix A to the tridiagonal form using Given's rotation.

$$A = \begin{pmatrix} 5 & 4 & 3 \\ 4 & 6 & 1 \\ 3 & 1 & 7 \end{pmatrix}.$$

Solución:
Necesitamos que los elementos siguientes sean nulos: a_{13} y a_{31}. Tenemos varias opciones para ello, por ejemplo, considerar las siguientes matrices de Givens:

$$G(1,3), G(3,1), G(2,3), G(3,2)$$

Si elegimos $G(1,3)$, es decir:

$$G(3,1) = \begin{pmatrix} c & 0 & s \\ 0 & 1 & 0 \\ -s & 0 & c \end{pmatrix}$$

luego:

$$G(3,1)A = \begin{pmatrix} 3s+5c & s+4c & 7s+3c \\ 4 & 6 & 1 \\ 3c-5s & c-4s & 7c-3s \end{pmatrix}$$

para que a_{13} sea nulo, se debe cumplir:

$$3c - 5s = 0, \quad \text{y como} \quad c^2 + s^2 = 1 \Rightarrow \left(\frac{5}{3}s\right)^2 + s^2 = 1 \Rightarrow s = \frac{3}{\sqrt{34}} \Rightarrow c = \frac{5}{\sqrt{34}},$$

así se obtiene:

$$G(3,1)A = \begin{pmatrix} \sqrt{34} & \dfrac{23}{\sqrt{34}} & \dfrac{36}{\sqrt{34}} \\ 4 & 6 & 1 \\ 0 & -\dfrac{7}{\sqrt{34}} & \dfrac{26}{\sqrt{34}} \end{pmatrix}$$

Desde que $G(3,1) = G(1,3)^T$ entonces el producto $G(3,1)AG(1,3)$ no resulta tridiagonal. Nos queda la alternativa de la matriz $G(2,3)$:

$$G(2,3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & -s \\ 0 & s & c \end{pmatrix}$$

luego:

$$G(2,3)A = \begin{pmatrix} 5 & 4 & 3 \\ 4c-3s & 6c-s & c-7s \\ 4s+3c & 6s+c & s+7c \end{pmatrix}$$

para que a_{31} sea nulo, se debe cumplir:

$$4s + 3c = 0, \quad \text{y como} \quad c^2 + s^2 = 1 \Rightarrow \left(-\frac{4}{3}s\right)^2 + s^2 = 1 \Rightarrow s = \frac{3}{5} \Rightarrow c = -\frac{4}{5},$$

por tanto:

$$G(2,3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\dfrac{4}{5} & -\dfrac{3}{5} \\ 0 & \dfrac{3}{5} & -\dfrac{4}{5} \end{pmatrix}$$

así se obtiene:

$$G(2,3)AG(2,3)^T = \begin{pmatrix} 5 & -5 & 0 \\ -5 & \dfrac{183}{25} & \dfrac{19}{25} \\ 0 & \dfrac{19}{25} & \dfrac{142}{25} \end{pmatrix}$$

que es una matriz tridiagonal.

4. (4 ptos) Sea A una matriz $m \times n$ de rango completo y b un punto de $\mathbb{R}^m$. Para $\lambda \geq \|b\|_2$ los conjuntos

$$K_\lambda = \{x \in \mathbb{R}^n : \|Ax - b\|_2 \leq \lambda\}$$

son cerrados y acotados.

(a) Pruebe que si $\lambda = 2\|b\|_2$ entonces

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_2 = \min_{x\in K_\lambda} \|Ax - b\|_2$$

(b) Pruebe que si B, $n \times m$ es tal que $BA = I$ y x es la solución de mínimos cuadrados de $Ax = b$ entonces

$$\|x\|_2 \leq 2\|b\|_2\|B\|_2$$

Solución:

(a) Como A es de rango completo entonces existen Q ortogonal y R triangular superior tal que $QA = R$.

Si $R = \left[\frac{R_1}{\text{O}}\right]$ y $Q = \left[\frac{Q_1}{Q_2}\right]$ entonces

$$Rx - Qb = \left[\frac{R_1x - Q_1b}{Q_2b}\right]$$

Como $\|Ax - b\|_2 = \|Rx - Qb\|_2$ entonces

$$\|Ax - b\|_2^2 = \|R_1x - Q_1b\|_2^2 + \|Q_2b\|_2^2$$

entonces $\|Ax - b\|_2$ alcanza valor mínimo $\|Q_2b\|_2$ en $x = R_1^{-1}Q_1b$.

Falta ver si $x \in K_\lambda$, en efecto como $\|Qb\|_2^2 = \|Q_1b\|_2^2 + \|Q_2b\|_2^2$ tenemos que

$$\|Ax - b\|_2 = \|Q_2b\|_2 \leq \|Qb\|_2 = \|b\|_2 \leq \lambda$$

(b) Como $x = Bb + B(Ax - b)$ entonces

$$\|x\|_2 \leq \|Bb\|_2 + \|B(Ax-b)\|_2 \leq \|B\|_2\|b\|_2 + \|B\|_2\|Ax-b\|_2 \leq \|B\|_2\|b\|_2 + \|B\|_2\|b\|_2$$

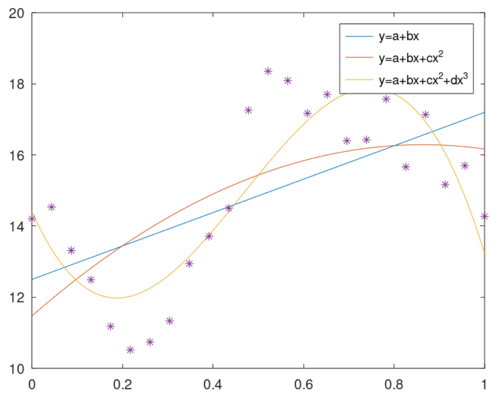
5. (4 ptos) Los datos del archivo adjunto corresponden a valores de temperatura tomados cada hora desde la medianoche hasta las 11 p.m. Realice un ajuste de mínimos cuadrados considerando los siguiente modelos:

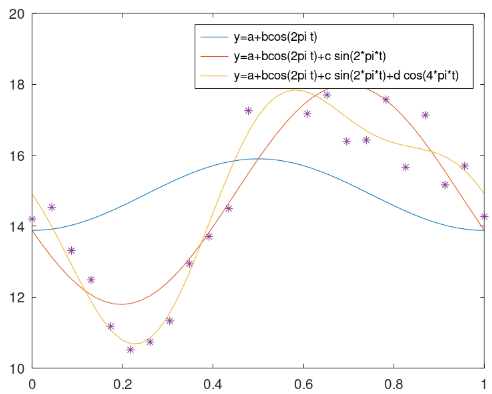
- $y = a + bt$.
- $y = a + bt + ct^2$.
- $y = a + bt + ct^2 + dt^3$.
- $y = a + b\cos(\pi\frac{t}{12})$.
- $y = a + b\cos(\pi\frac{t}{12}) + c\sin(\pi\frac{t}{12})$.
- $y = a + b\cos(\pi\frac{t}{12}) + c\sin(\pi\frac{t}{12}) + d\cos(\pi\frac{t}{6})$.

Genere las gráficas del modelo y los datos observados. ¿Cual considera que es el modelo que mejor ajusta los datos y por qué?. Adjunte los modelos encontrados, las gráficas y el código utilizado.

Solución:

- $y = 12.5005 + 4.6970t$. $r = 9.4050$
- $y = 11.4711 + 11.1540t - 6.4570t^2$. $r = 9.0510$
- $y = 14.400 - 28.242t + 94.156t^2 - 67.075t^3$. $r = 5.7741$
- $y = 14.8911 - 1.0096\cos(\pi\dfrac{t}{12}).r = 11.123$
- $y = 14.8911 - 1.0096\cos(\pi\dfrac{t}{12}) - 2.9264\sin(\pi\dfrac{t}{12})$. $r = 5.0232$
- $y = 14.8464 + -1.0990\cos(\pi\dfrac{t}{12}) - 2.9264\sin(\pi\dfrac{t}{12}) + 1.1629\cos(\pi\dfrac{t}{6})$. $r = 2.9126$





El mejor modelo es el que tiene menor valor de r.

Los profesores[1]
Lima, 2 de Junio del 2021.

[1] Hecho en LaTeX