Solución del examen - 04/02/25

Ejercicio 1.

a) Ver la Definición 1.4.1 y la Observación 1.4.1 de las notas de teórico.

b) Observamos que, para hallar $f(x)$, la primera operación que se debe resolver corresponde a calcular $(1-x)$ en precisión doble. Si $x \in [0, \varepsilon_M/4]$, entonces $(1-x) \in [1-\varepsilon_M/4, 1]$. Ahora, como el mayor número representable menor que 1 en precisión doble es $1-\varepsilon_M/2$, observamos que $\operatorname{dist}(1-x, 1) < \operatorname{dist}(1-x, 1-\varepsilon_M/2)$ para todo $x \in [1-\varepsilon_M/4, 1]$. Por lo tanto, $1-x$ es aproximado a 1 y entonces $f(x) = 0$.

c) Como hicimos en la parte anterior, si $x \approx \varepsilon_M/4$ con $x < \varepsilon_M/4$, se tiene que $\operatorname{dist}(1-x, 1) > \operatorname{dist}(1-x, 1-\varepsilon_M/2)$, y por lo tanto $f(x) = 0$.

Sin embargo, si $x \approx \varepsilon_M/4$ con $x \geq \varepsilon_M/4$, se tiene que $\operatorname{dist}(1-x, 1) \leq \operatorname{dist}(1-x, 1-\varepsilon_M/2)$ y en este caso $1-\varepsilon_M/4$ se aproxima por $1-\varepsilon_M/2$. Esto implica que

$$f(x) \simeq \frac{1-(1-\varepsilon_M/2)}{\varepsilon_M/4} = 2.$$

d) La oscilación de f en este rango se debe a las limitaciones en la precisión de las operaciones de punto flotante, en particular a los errores cometidos al computar $1-x$. Cada vez que x toma un valor tal que $1-x$ puede ser computado de forma exacta en punto flotante ($x = \varepsilon_M/2, \varepsilon_M, 3\varepsilon_M/2, 2\varepsilon_M$), se evalúa $f(x) = 1$ como quisiéramos. Pero cuando x no toma estos valores, $1-x$ es aproximado a un número de punto flotante, lo que hace que $1-(1-x)$ solamente pueda tomar los valores $1, \varepsilon_M/2, \varepsilon_M, 3\varepsilon_M/2, 2\varepsilon_M$ en el intervalo y luego se computa

$$f(x) = \begin{cases} \frac{\varepsilon_M/2}{x}, & \text{para } x \in [\varepsilon_M/4, 3\varepsilon_M/4) \\ \frac{\varepsilon_M}{x}, & \text{para } x \in [3\varepsilon_M/4, 5\varepsilon_M/4) \\ \frac{3\varepsilon_M/2}{x}, & \text{para } x \in [5\varepsilon_M/4, 7\varepsilon_M/4) \\ \dots & \dots \end{cases}.$$

Ejercicio 2.

a) Dado el sistema

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mathbf{x} = \mathbf{b},$$

la iteración asociada al método de Jacobi es,

$$\begin{aligned} \mathbf{x}^{k+1} &= \begin{pmatrix} a_{11}^{-1} & 0 \\ 0 & a_{22}^{-1} \end{pmatrix} \begin{pmatrix} 0 & -a_{12} \\ -a_{21} & 0 \end{pmatrix} \mathbf{x}^k + \begin{pmatrix} a_{11}^{-1} & 0 \\ 0 & a_{22}^{-1} \end{pmatrix} \mathbf{b} \\ &= \begin{pmatrix} 0 & -a_{12}a_{11}^{-1} \\ -a_{21}a_{22}^{-1} & 0 \end{pmatrix} \mathbf{x}^k + \begin{pmatrix} a_{11}^{-1} & 0 \\ 0 & a_{22}^{-1} \end{pmatrix} \mathbf{b} \\ &= Q\mathbf{x}^k + D^{-1}\mathbf{b}. \end{aligned}$$

Ahora, para ver que el método converge, basta ver que el radio espectral de Q cumple $\rho(Q) < 1$. Calculando el polinomio característico de Q, observamos

$$0 = |Q - \lambda I| = \left| \begin{pmatrix} -\lambda & -a_{12}a_{11}^{-1} \\ -a_{21}a_{22}^{-1} & -\lambda \end{pmatrix} \right| = \lambda^2 - a_{11}^{-1}a_{22}^{-1}a_{21}a_{12}.$$

Como $\det(A) > 0$, tenemos que
$$a_{11}a_{22} > a_{21}a_{12},$$
y por lo tanto,
$$\lambda^2 = a_{11}^{-1}a_{22}^{-1}a_{21}a_{12} < 1.$$
Luego $|\lambda| < 1$, y en particular, $\rho(Q) < 1$, por lo que el método converge.

b) Para el sistema
$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \mathbf{x} = \mathbf{b},$$
la iteración de Jacobi es
$$\mathbf{x}^{k+1} = \begin{pmatrix} 0 & -2 \\ -2 & 0 \end{pmatrix} \mathbf{x}^k + \mathbf{b} =: Q\mathbf{x}^k + \mathbf{b}.$$
Entonces, el polinomio característico de Q es
$$0 = \left| \begin{pmatrix} -\lambda & -2 \\ -2 & -\lambda \end{pmatrix} \right| = \lambda^2 - 4 = (\lambda + 2)(\lambda - 2),$$
por lo que $\rho(Q) = 2$ y por lo tanto el método no converge.

c) Para ver si existe alguna relajación que haga que el método converja, debemos examinar el radio espectral de la matriz
$$Q_\omega := \omega Q + (1 - \omega)I = \begin{pmatrix} 1-\omega & -2\omega \\ -2\omega & 1-\omega \end{pmatrix}$$
para $\omega > 0$. Calculando su polinomio característico, obtenemos
$$\begin{aligned} \det(Q_\omega - \lambda I) &= \left| \begin{pmatrix} 1-\omega-\lambda & -2\omega \\ -2\omega & 1-\omega-\lambda \end{pmatrix} \right| \\ &= (1-\omega-\lambda)^2 - 4\omega^2 = \lambda^2 - 2(1-\omega)\lambda - 4\omega^2 + (1-\omega)^2 = 0, \end{aligned}$$
por lo que
$$\lambda = \frac{2(1-\omega) \pm \sqrt{4(1-\omega)^2 + 16\omega^2 - 4(1-\omega)^2}}{2} = (1-\omega) \pm 2\omega.$$
Luego, las raíces son $\lambda_1 = 1 + \omega$ y $\lambda_2 = 1 - 3\omega$. Dado que $\omega > 0$, observamos que $\lambda_1 > 1$ para todo ω, y por lo tanto no existen relajaciones que hagan que el método converja.

Ejercicio 3. a) Observamos que $f(x_0) = 0$, $f(x_1) = 1$ y $f(x_2) = 0$, por lo que para calcular el polinomio interpolante, basta calcular el polinomio de base de Lagrange $L_2^1(x)$. Este polinomio está dado por
$$L_2^1(x) = \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} = -4x(x-1).$$
Entonces, el polinomio interpolante es
$$p_2(x) = -4x(x-1).$$
Usando el teorema de error de interpolación polinomial, obtenemos que para todo $x \in [0,1]$ existe $\gamma_x \in (0,1)$ tal que
$$f(x) - p_2(x) = \frac{f^{(3)}(\gamma_x)}{3!} x(x-1/2)(x-1).$$

Como $f^{(3)}(x) = -\pi^3 \cos(x)$ y acotando $|x(x-1/2)(x-1)| \leq 1$ (porque $x \in [0,1]$; esta cota no es la óptima), tenemos

$$|f(x) - p_2(x)| \leq \frac{\pi^3}{6}.$$

b) Usando que

$$L(x) = f(x_i) + (x - x_i)\frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \quad \forall x \in [x_i, x_{i+1}],$$

vemos que la interpolante lineal a trozos está dada por,

$$L(x) = \begin{cases} 2x & \text{si } x \in [0, 1/2], \\ 2 - 2x & \text{si } x \in [1/2, 1]. \end{cases}$$

Para la cota de error, usamos nuevamente el teorema de error de interpolación, con

$$|f(x) - 2x| \leq \frac{\pi^2}{2}|x(x-1/2)| \leq \frac{\pi^2}{2}\frac{1}{2}\frac{1}{2} = \frac{\pi^2}{8} \quad \text{en } [0, 1/2],$$

y

$$|f(x) - (2 - 2x)| \leq \frac{\pi^2}{2}\frac{1}{2}\frac{1}{2} = \frac{\pi^2}{8} \quad \text{en } [1/2, 1].$$

Por lo tanto, el máximo del error de interpolación está acotado por $\pi^2/8$.

c) Cuando n es grande, para la interpolante lineal a trozos vamos a tener un error del orden de $1/n^2$ porque la distancia entre nodos consecutivos es igual a 1/n. En cambio, para la interpolante de grado alto la fórmula del error de interpolación indica

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\gamma_x)}{(n+1)!}\,(x - x_0)(x - x_1)\ldots(x - x_n).$$

Razonando como antes, podemos acotar el polinomio nodal,

$$|(x - x_0)(x - x_1)\ldots(x - x_n)| \leq 1,$$

y la derivada de orden $n+1$,

$$|f^{(n+1)}(\gamma_x)| \leq \pi^{n+1}.$$

Por lo tanto, el error de interpolación se puede acotar como

$$|f(x) - p_n(x)| \leq \frac{\pi^{n+1}}{(n+1)!},$$

que decae a 0 mucho más rápido. Para esta función, la interpolante de grado alto da lugar a un menor error de interpolación.

Ejercicio 4. a) El método de Runge-Kutta descrito es explícito, dado que para el cálculo de K_1 y K_2, se requiere únicamente t_k, y_k y h.

b) Observamos que

$$\begin{aligned} K_1 &= f(t_0, y_0) = f(0,0) = \cos(0) + 0 = 1, \\ K_2 &= f\left(t_0 + \frac{2}{3}, y_0 + \frac{2}{3}K_1\right) = f\left(\frac{2}{3}, \frac{2}{3}\right) = \cos\left(\frac{2\pi}{3}\right) + \frac{2}{3} = -\frac{1}{2} + \frac{2}{3} = \frac{1}{6}, \\ y_1 &= y_0 + \frac{1}{4}K_1 + \frac{3}{4}K_2 = 0 + \frac{1}{4} + \frac{1}{8} = \frac{3}{8}. \end{aligned}$$

c) Debemos ver que el error local de truncamiento es $\ell_{k+1} = O(h^3)$. Consideramos el problema

$$\begin{cases} u(t) = f(t, u(t)), \\ u(t_k) = y_k, \end{cases}$$

y haciendo un desarrollo de Taylor para u obtenemos

$$u(t_{k+1}) = y_k + hf(t_k, y_k) + \frac{h^2}{2} u''(t_k) + O(h^3).$$

Si denotamos

$$f := f(t_k, y_k), \quad f_t := \frac{\partial f}{\partial t}(t_k, y_k), \quad f_y := \frac{\partial f}{\partial y}(t_k, y_k),$$

observamos que

$$u''(t_k) = \frac{d}{dt} f(t, u(t))\Big|_{t=t_k} = f_t + f_y u'(t_k) = (f_t + f_y f)\Big|_{(t_k, u(t_k))}.$$

Por lo tanto,

$$u(t_{k+1}) = y_k + hf + \frac{h^2}{2} f_t + \frac{h^2}{2} f_y f + O(h^3).$$

Vemos que $K_1 = f$, y haciendo un desarrollo de Taylor para K_2,

$$K_2 = f\left(t_k + \frac{2h}{3}, y_k + \frac{2h}{3} K_1\right) = f + \frac{2h}{3} f_t + \frac{2h}{3} f_y f + O(h^2).$$

Luego,

$$\begin{aligned} y_{k+1} &= y_k + \frac{h}{4} K_1 + \frac{3h}{4} K_2. \\ &= y_k + \frac{h}{4} f + \frac{3h}{4} (f + \frac{2h}{3} f_t + \frac{2h}{3} f_y f + O(h^2)) \\ &= y_k + hf + \frac{h^2}{2} f_t + \frac{h^2}{2} f_y f + O(h^3), \end{aligned}$$

por lo que, juntando las expresiones para $u(t_{k+1})$ e y_{k+1}, obtenemos

$$\ell_{k+1} = y_{k+1} - u(t_{k+1}) = O(h^3).$$