Cálculo Numérico - Elementos de Cálculo Numérico

Métodos de un paso - Diferencias finitas

Clase dictada por: Mercedes Pérez Millán

(Dto. de Matemática–FCEyN–Universidad de Buenos Aires, IMAS-CONICET)

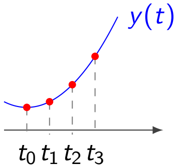
28 de agosto de 2024

Problema de valores iniciales

$$\begin{cases} y'(t) = f(t, y(t)) \\ y(t_0) = y_0 \end{cases}$$

Son *pocos* los casos donde podemos encontrar la solución analíticamente, por lo que vamos a estudiar métodos numéricos para aproximar los valores de la solución.

$$(y_0, y_1, y_2, y_3, \dots) \sim (y(t_0), y(t_1), y(t_2), y(t_3), \dots)$$



Método de Euler

Método:

Sean $h = \frac{T_F - t_0}{N}$, $t_n = t_0 + nh$,

$$\begin{cases} y_0 & = y(t_0) \\ y_{n+1} & = y_n + hf(t_n, y_n) \end{cases}$$

$n = 0, 1, \ldots, N - 1.$

¿Y el error?

Sea

$$\hat{y}_n := y(t_{n-1}) + hf(t_{n-1}, y(t_{n-1})).$$

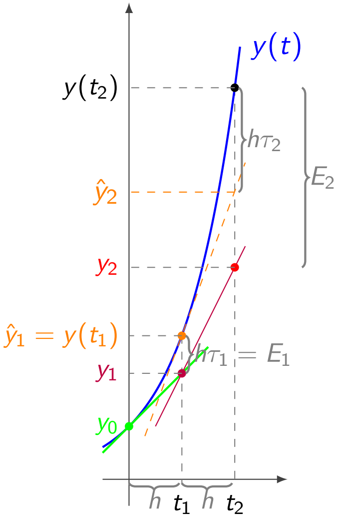
(la recta tangente a y en t_{n-1} evaluada en t_n.)
Definimos

$$h\tau_n := y(t_n) - \hat{y}_n,$$

el *error de truncado local*.
Definimos

$$E_n = y(t_n) - y_n,$$

el *error global*.



Error de truncado local

Recordemos el desarrollo de Taylor de y centrado en t_{n-1}:

$$y(t) = y(t_{n-1}) + y'(t_{n-1})(t - t_{n-1}) + \frac{y''(\theta)}{2}(t - t_{n-1})^2,$$

para θ entre t y t_{n-1}. Como $y'(t_{n-1}) = f(t_{n-1}, y(t_{n-1}))$ y $t_n - t_{n-1} = h$, deducimos:

$$\begin{aligned} h\tau_n &= y(t_n) - \hat{y}_n = y(t_n) - (y(t_{n-1}) + hf(t_{n-1}, y(t_{n-1}))) \\ &= \frac{y''(\theta_n)}{2}h^2, \quad \theta_n \in (t_{n-1}, t_n). \end{aligned}$$

Y podemos acotar

$$|\tau_n| \leq \max_{t\in[t_0,t_F]} |y''(t)| \frac{h}{2}, \qquad \text{para } n = 1, \ldots, N.$$

Método de Runge-Kutta 4

Sean $h = \frac{T_F - t_0}{N}$, $t_n = t_0 + nh$:

$$\begin{cases} y_0 = y(t_0) \\ K_1 = f(t_n, y_n) \\ K_2 = f(t_n + \frac{h}{2}, y_n + \frac{h}{2}K_1) \\ K_3 = f(t_n + \frac{h}{2}, y_n + \frac{h}{2}K_2) \\ K_4 = f(t_n + h, y_n + hK_3) \\ y_{n+1} = y_n + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) \end{cases}$$

para $n = 0, 1, \ldots, N-1$.

El error local de R-K 4 es una $O(h^5)$.

Error global

Decimos que $f : \mathbb{R}^2 \to \mathbb{R}^2$ satisface la *condición de Lipschitz en la segunda variable* si existe $L > 0$ tal que

$|f(t,u) - f(t,v)| \leq L|u - v|$ para toda elección posible de t, u, v.

Vamos a considerar mayormente:

$$\max_{\substack{t \in [t_0, t_F] \\ y \in [a,b]}} |f_y(t,y)| \leq L,$$

para algún intervalo $[a, b]$ (*ya veremos...*).
Tenemos entonces la siguiente cota para el error global:

$$\boxed{|E_N| = |y(t_F) - y_N| \leq \frac{e^{L(t_F - t_0)} - 1}{L} \tau_{max}.}$$

Error global para el método de Euler

Veamos cómo se obtiene la cota.

$$E_n = y(t_n) - y_n$$

Método de Euler, que usa para aproximar el valor en t_n el valor *aproximado* en t_{n-1}:

$$\begin{aligned}
y_1 &= y_0 + hf(t_0, y_0)\\
y_2 &= y_1 + hf(t_1, y_1)\\
y_3 &= y_2 + hf(t_2, y_2)\\
&\vdots\\
y_N &= y_{N-1} + hf(t_{N-1}, y_{N-1})
\end{aligned}$$

Método que tendría si usara para aproximar el valor en t_n el valor *real* en t_{n-1}:

$$\begin{aligned}
\hat{y}_1 &= y(t_0) + hf(t_0, y(t_0))\\
\hat{y}_2 &= y(t_1) + hf(t_1, y(t_1))\\
\hat{y}_3 &= y(t_2) + hf(t_2, y(t_2))\\
&\vdots\\
\hat{y}_N &= y(t_{N-1}) + hf(t_{N-1}, y(t_{N-1}))
\end{aligned}$$

Para comparar $y(t_n)$ con y_n vamos a intercalar $\hat{y}_n$.

$$
\begin{aligned}
E_1 &= y(t_1) - y_1 = h\tau_1 \text{ en el primer paso } \hat{y}_1 = y_1 \\
E_2 &= y(t_2) - y_2 = y(t_2) - \hat{y}_2 + \hat{y}_2 - y_2 \\
&= h\tau_2 + y(t_1) + hf(t_1, y(t_1)) - y_1 - hf(t_1, y_1) \\
&= h\tau_2 + E_1 + h(f(t_1, y(t_1)) - f(t_1, y_1))
\end{aligned}
$$

Tomando módulo y usando que f es Lipschitz en la segunda variable:

$$
\begin{aligned}
|E_2| &\leq |h\tau_2| + |E_1| + h|f(t_1, y(t_1)) - f(t_1, y_1)| \\
&\leq |h\tau_2| + |E_1| + hL|y(t_1) - y_1| \\
&= |h\tau_2| + |E_1| + hL|E_1| = |h\tau_2| + (1 + hL)|E_1| \\
&= |h\tau_2| + (1 + hL)|h\tau_1|
\end{aligned}
$$

$$
\begin{aligned}
E_3 = y(t_3) - y_3 &= y(t_3) - \hat{y}_3 + \hat{y}_3 - y_3 \\
&= h\tau_3 + y(t_2) + hf(t_2, y(t_2)) - y_2 - hf(t_2, y_2) \\
&= h\tau_3 + E_2 + h(f(t_2, y(t_2)) - f(t_2, y_2))
\end{aligned}
$$

Tomando módulo y usando que f es Lipschitz en la segunda variable:

$$
\begin{aligned}
|E_3| &\leq |h\tau_3| + |E_2| + h|f(t_2, y(t_2)) - f(t_2, y_2)| \\
&\leq |h\tau_3| + |E_2| + hL|y(t_2) - y_2| \\
&= |h\tau_3| + |E_2| + hL|E_2| = |h\tau_3| + (1 + hL)|E_2|
\end{aligned}
$$

Usando la cota para $|E_2|$ nos queda

$$
\begin{aligned}
|E_3| &\leq |h\tau_3| + (1 + hL)(|h\tau_2| + (1 + hL)|h\tau_1|) \\
&= |h\tau_3| + (1 + hL)|h\tau_2| + (1 + hL)^2|h\tau_1|
\end{aligned}
$$

Si $\tau_{MAX} = \max\limits_{1 \leq n \leq N} |\tau_n|$, tenemos

$$\begin{aligned} |E_3| &\leq h\tau_{MAX} + (1+hL)h\tau_{MAX} + (1+hL)^2 h\tau_{MAX} \\ &= (1 + (1+hL) + (1+hL)^2) h\tau_{MAX} \end{aligned}$$

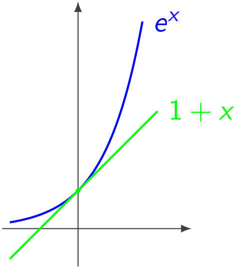
y así...

$$\begin{aligned} |E_N| &\leq (1 + (1+hL) + (1+hL)^2 + \cdots + (1+hL)^{N-1}) h\tau_{MAX} \\ &= \left(\sum_{k=0}^{N-1} (1+hL)^k \right) h\tau_{MAX} \end{aligned}$$

<u>Serie geométrica:</u> Para $R \neq 1$, $\sum\limits_{k=0}^{N-1} R^K = \frac{R^N - 1}{R-1}$ y nos queda:

$$|E_n| \leq \frac{(1+hL)^N - 1}{hL} h\tau_{MAX}.$$

Por último usamos que $1 + x \leq e^x$ para todo $x \in \mathbb{R}$:



y por lo tanto $(1 + x)^N \leq e^{Nx}$ para todo $x \in \mathbb{R}$. De esta forma:

$$|E_n| \leq \frac{(1 + hL)^N - 1}{hL} h\tau_{MAX} \leq \frac{e^{NhL} - 1}{L} \tau_{MAX}$$

Y, finalmente, como $h = \frac{t_F - t_0}{N}$, resulta:

$$|E_n| \leq \frac{e^{L(t_F - t_0)} - 1}{L} \tau_{MAX}.$$

Para el método de Euler, $\tau_{MAX} \leq C_{MAX} \frac{h}{2}$ y nos queda:

$$\boxed{|E_N| = |y(t_F) - y_N| \leq \frac{e^{L(t_F - t_0)} - 1}{L} C_{MAX} \frac{h}{2}.}$$

Orden de convergencia

- Decimos que $\varphi : \mathbb{R} \to \mathbb{R}$ es $\varphi(h) = O(h^k)$ cuando $h \to 0$ si existen $C \geq 0$ y $\delta > 0$ tales que $|\varphi(h)| \leq Ch^k$ cuando $h < \delta$.

  Por ejemplo, si $\varphi(h) = \frac{y^{(k)}(\theta)}{k!}h^k$, puedo tomar $C = \frac{\max_z |y^{(k)}(z)|}{k!}$ y entonces $\varphi(h) = O(h^k)$.

- Decimos que un método converge con orden k si el error global, E_N, es una $O(h^k)$.

- En general, si el error de truncado local de un método de un paso es una $O(h^k)$, el error global es una $O(h^k)$.

¿Cómo podemos estimar el orden de un método?

Idea:

$$|E_N| \sim Ch^k \;\Rightarrow\; \log(|E_N|) \sim \log(C) + k\log(h)$$

Si pudiéramos graficar $\log(|E_N|)$ en función de $\log(h)$, tendríamos que k es la pendiente de "la recta".

Definición:

Un método de un paso se dice consistente si $\lim_{h \to 0} \tau_{MAX} = 0$.

Proposición:

Para un método de un paso con función de incremento φ continua en sus tres variables se tiene que $\lim_{h \to 0} \tau_{MAX} = 0$ si y solo si $\varphi(t, y, 0) = f(t, y)$.

Corolario:

Si un método de un paso con función de incremento φ es consistente y φ es Lipschitz en la segunda variable, entonces el método es convergente (i.e. $y_N \underset{N \to \infty}{\longrightarrow} y(T_F)$).

Método de Euler Implícito

Si desarrollamos Taylor centrado en $t+h$:

$$y(t) = y(t+h) - y'(t+h)h + \cdots = y(t+h) - f(t+h, y(t+h))h + \ldots$$

y deducimos:

$$y(t+h) \sim y(t) + hf(t+h, y(t+h)),$$

surge el método implícito:

Método:

Sean $h = \frac{T_F - t_0}{N}$, $t_n = t_0 + nh$,

$$\begin{cases} y_0 & = y(t_0) \\ y_{n+1} & = y_n + hf(t_{n+1}, y_{n+1}) \quad \text{para } n = 0, 1, \ldots, N-1. \end{cases}$$

Observación: el método implícito requiere resolver una ecuación (en general no lineal) en cada paso.

Por ejemplo, en el primer paso tenemos

$$y_1 = y_0 + hf(t_1, y_1).$$

Si bien y_0, h y t_1 los conocemos, en general no podemos despejar y_1 y hay que resolver la ecuación

$$y_1 - y_0 - hf(t_1, y_1) = 0$$

usando, por ejemplo, el método de Newton-Raphson.

Ejercicio

Considerar el problema de valores iniciales:

$$\begin{cases} y'(t) &= & t\cos^2(y(t)), \\ y(0) &= & 5. \end{cases}$$

- En “papel”:
  1. Escribir la iteración del método de Euler correspondiente a este problema.
  2. Hallar h para que el error al estimar $y(1)$ con el método de Euler sea menor que 10^{-3}.
- Con Python: Escribir un programa que implemente el método de Euler explícito para este PVI y graficar la solución obtenida para distintos valores de h.

Resolución:

1) La iteración del método de Euler con paso $h = \frac{T_F - t_0}{N} = \frac{1-0}{N} = \frac{1}{N}$, $t_n = t_0 + nh = 0 + nh = nh$ es

$$\begin{cases} y_0 & = 5 \\ y_{n+1} & = y_n + hf(t_n, y_n) = y_n + ht_n \cos^2(y_n) \end{cases}$$

$n = 0, 1, \ldots, N-1.$

Resolución:

2) Buscamos h para que $|E_N| < 10^{-3}$. Hallemos C_{MAX} y L para este problema. En este caso, $f(t,y) = t\cos^2(y)$.

$$\begin{aligned} y''(t) &= f_t(t,y(t)) + f_y(t,y(t))f(t,y(t)) \\ &= \cos^2(y(t)) + t.\,2\cos(y(t))(-\operatorname{sen}(y(t)))f(t,y(t)) \\ &= \underline{\cos^2(y(t))} - 2t\cos(y(t))\operatorname{sen}(y(t))t\underline{\cos^2(y(t))} \\ &= \cos^2(y(t))[1 - 2t^2\cos(y(t))\operatorname{sen}(y(t))]. \end{aligned}$$

De esta forma:

$$\begin{aligned} |y''(t)| &\leq |\cos(y(t))|^2[1 + |2t^2||\cos(y(t))||\operatorname{sen}(y(t))|] \\ &\underset{t\in[0,1]}{\leq} 1^2.\,[1 + 2.\,1.\,1] = 3 =: C_{MAX} \end{aligned}$$

Y $|f_y(t,y)| = |t.\,2\cos(y)(-\operatorname{sen}(y))| \leq |2t| \underset{t\in[0,1]}{\leq} 2 =: L$

Resolución:

Sabemos que

$$\begin{aligned} |E_N| &\leq \frac{e^{L(t_F - t_0)} - 1}{L} C_{MAX} \frac{h}{2} \\ &= \frac{e^{2(1-0)} - 1}{2} \cdot \frac{3h}{2} = \frac{3h}{4}(e^2 - 1) \leq 6h \end{aligned}$$

Ahora busquemos un valor de h para que este error sea menor que 10^{-3}:

$$6h < 10^{-3} \Leftrightarrow h < \frac{1}{6000} \sim 0{,}0001666\ldots.$$

Por lo tanto, cualquier $h < 0{,}0001666$ sirve para tal propósito.

Diferencias finitas

Partimos de un problema de valores de contorno:

$$\begin{cases} u''(x) = p(x)u'(x) + q(x)u(x) + f(x) \\ u(a) = \alpha, \ u(b) = \beta \end{cases}$$

Buscamos $(u_0, \ldots, u_N)$ de forma tal que $u_j \sim u(x_j)$ para $1 \leq j \leq N-1$.

Teorema:

Dada la ecuación de segundo orden $u''(x) = p(x)u'(x) + q(x)u(x)+f(x)$ con $a \leq x \leq b$ y valores en la frontera $u(x) = \alpha$, $u(b) = \beta$. Si p, q, f son continuas en $[a, b]$ y $q > 0$, entonces el problema tiene única solución.

Sea $n \in \mathbb{N}$, tomamos un paso $h = (b - a)/N$ y consideramos los puntos $x_j = a + jh$ para $j = 0, \ldots, N$ (notar que $x_0 = a, x_N = b$). Definimos $u_0 = \alpha$, $u_N = \beta$ y aproximamos las derivadas de u.

Discretizaciones usuales de u'

- Forward: $u'(x_j) \sim \frac{u(x_{j+1})-u(x_j)}{h}$
- Backward: $u'(x_j) \sim \frac{u(x_j)-u(x_{j-1})}{h}$
- Centrada: $u'(x_j) \sim \frac{u(x_{j+1})-u(x_{j-1})}{2h}$

Discretizaciones usuales de u''

$$
\begin{array}{rrcl}
 & u(x+h) & = & u(x) + hu'(x) + \frac{h^2}{2}u''(x) + \frac{h^3}{6}u'''(x) + \frac{h^4}{24}u^{(iv)}(\xi_1) \\
+ & & & \\
 & \underline{u(x-h)} & = & \underline{u(x) - hu'(x) + \frac{h^2}{2}u''(x) - \frac{h^3}{6}u'''(x) + \frac{h^4}{24}u^{(iv)}(\xi_2)} \\
 & {\scriptstyle u(x+h)+u(x-h)} & = & 2u(x) + h^2u''(x) + \frac{h^4}{24}(u^{(iv)}(\xi_1) + u^{(iv)}(\xi_2))
\end{array}
$$

$$
\Rightarrow u''(x) = \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} - \underbrace{\frac{h^4}{24}(u^{(iv)}(\xi_1) + u^{(iv)}(\xi_2))}_{O(h^2)}
$$

Llamamos $p_j = p(x_j)$, $q_j = q(x_j)$, $f_j = f(x_j)$ y usando diferencias centradas para u' proponemos:

$$\frac{u_{j+1} - 2u_j + u_{j-1}}{h^2} = p_j \frac{u_{j+1} - u_{j-1}}{2h} + q_j u_j + f_j$$

Multiplicando por h^2 y reagrupando,

$$u_{j+1} - 2u_j + u_{j-1} - \frac{h}{2} p_j [u_{j+1} - u_{j-1}] - h^2 q_j u_j = h^2 f_j$$

$$\left(1 + \frac{h}{2} p_j\right) u_{j-1} - (2 + h^2 q_j) u_j + \left(1 - \frac{h}{2} p_j\right) u_{j+1} = h^2 f_j$$

$j{=}1$ $\quad -(2 + h^2 q_1) u_1 + \left(1 - \tfrac{h}{2} p_1\right) u_2 = h^2 f_1 - \left(1 + \tfrac{h}{2} p_1\right) \alpha$

$j{=}2{:}N{-}2$ $\quad (1 + \tfrac{h}{2} p_j) u_{j-1} - (2 + h^2 q_j) u_j + (1 - \tfrac{h}{2} p_j) u_{j+1} = h^2 f_j$

$j{=}N{-}1$ $\quad (1 + \tfrac{h}{2} p_{N-1}) u_{N-2} - (2 + h^2 q_{N-1}) u_{N-1} = h^2 f_{N-1} - (1 - \tfrac{h}{2} p_{N-1}) \beta$

En forma matricial:

$$\underbrace{\begin{pmatrix} -(2+h^2q_1) & 1-\frac{h}{2}p_2 & & & \\ 1+\frac{h}{2}p_2 & -(2+h^2q_2) & (1-\frac{h}{2}p_2) & & \\ & & \ddots & & \\ & 1+\frac{h}{2}p_{N-2} & -(2+h^2q_{N-2}) & 1-\frac{h}{2}p_{N-2} \\ & & (1+\frac{h}{2}p_{N-1}) & -(2+h^2q_{N-1}) \end{pmatrix}}_{A_h} \underbrace{\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N-1} \end{pmatrix}}_{u_h} = \underbrace{\begin{pmatrix} h^2f_1 - (1+\frac{h}{2}p_1)\alpha \\ \cdots \\ h^2f_j \\ \cdots \\ h^2f_{N-1} - (1-\frac{h}{2}p_{N-1})\beta \end{pmatrix}}_{b_h}$$

En definitiva,

$$A_h u_h = b_h$$

Teorema 1:

Si $q > 0$ y $h < 2/M$ con $M = \max\{|p(x)| \ : \ a < x < b\}$, entonces la matriz es inversible.

Teorema 2:

Sea $A = \{a_{ij}\}_{i,j=1,\cdots,N}$ tal que $|a_{jj}| > \sum_{k\neq j} |a_{jk}|$. Entonces todos los autovalores son distintos de cero.