IMT3410: Métodos para Ecuaciones Diferenciales

Manuel A. Sánchez

Índice general

Prólogo

Apuntes de clase IMT3410 Métodos para Ecuaciones Diferenciales.

Capítulo 1

Métodos para Ecuaciones Diferenciales Ordinarias

1.1. Terminología

- Una **ecuación diferencial de primer orden** es una ecuación para y una variable dependiente y x una variable independiente de la forma

$$\frac{dy}{dx} = y'(x) = f(x, y(x))$$

para $f(x, y)$ una función dada. Una función $y = y(x)$ es llamada **solución** de esta ecuación si este satisface

$$y'(x) = f(x, y(x)) \qquad \forall x.$$

Las soluciones usualmente tienen un parámetro libre así pueden ser únicamente determinadas con un **valor inicial** o condición inicial

$$y(x_0) = y_0.$$

- Una **ecuación diferencial de segundo orden** para $y = y(x)$ es de la forma

$$y'' = f(x, y, y')$$

Usualmente es necesario 2 parámetros para determinar una solución única, los cuales pueden ser únicamente determinados por 2 valores iniciales

$$y(x_0) = y_0, \qquad y'(x_0) = y'_0.$$

Podemos escribir estas como un **sistema de primer orden**, introduciendo las variables of funciones $y_1(x) = y(x)$; $y_2(x) = y'(x)$,

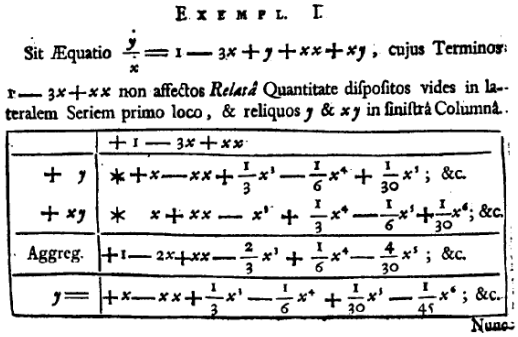
$$\begin{cases} y'_1 = y_2, & y_1(x_0) = y_0 \\ y'_2 = f(x, y_1, y_2), & y_2(x_0) = y'_0. \end{cases}$$

- Podemos extender la noción de **sistema de primer orden** a n ecuaciones, por

$$\begin{cases} y'_1 = f_1(x, y_1, \ldots, y_n), & y_1(x_0) = y_{10} \\ \quad \vdots & \quad \vdots \\ y'_n = f_n(x, y_1, \ldots, y_n), & y_n(x_0) = y_{n0}. \end{cases}$$

Esto en forma vectorial se escribe

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}); \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad \mathbf{f}(x, \mathbf{y}) = \begin{bmatrix} f_1(x, \mathbf{y}) \\ \vdots \\ f_n(x, \mathbf{y}) \end{bmatrix}$$



EXEMPL. I.

Sit Æquatio $\frac{\dot{y}}{\dot{x}} = 1 - 3x + y + xx + xy$, cujus Terminos $1 - 3x + xx$ non affectos *Relatâ* Quantitate disposito vides in lateralem Seriem primo loco, & reliquos y & xy in sinistrâ Columnâ.

	$+1 - 3x + xx$
$+ y$	$\ast + x - xx + \frac{1}{3}x^3 - \frac{1}{6}x^4 + \frac{1}{30}x^5$; &c.
$+ xy$	$\ast \quad xx + xx - x^3 + \frac{1}{3}x^4 - \frac{1}{6}x^5 + \frac{1}{30}x^6$; &c.
Aggreg.	$+1 - 2x + xx - \frac{2}{3}x^3 + \frac{1}{6}x^4 - \frac{4}{30}x^5$; &c.
$y =$	$+x - xx + \frac{1}{3}x^3 - \frac{1}{6}x^4 + \frac{1}{30}x^5 - \frac{1}{45}x^6$; &c.

Nunc

1.1.1. Ejemplos históricos

- **Newton (Differential Calculus 1671):**

$$y' = 1 - 3x + y + x^2 + xy$$

Una de las primeras ecuaciones diferenciales. Se puede resolver usando series infinitas. Solución por series infinitas:

Supongamos que tenemos la ecuación $y' = 1 - 3x + y + x^2 + xy, \quad y(0) = 0.$ Usamos la ecuación y obtenemos la derivada en 0

$$\begin{aligned} &\to \quad y' = 1 - 3 \cdot 0 + 0 + 0^2 + 0 \cdot 0 = 1 \\ &\to \quad y = x \end{aligned}$$

Luego reemplazamos la expresión $y = x$ en la ecuación diferencial y obtenemos

$$\begin{aligned} &\to \quad y' = 1 - 3x + x + x^2 + x \cdot x = 1 - 2x + 2x^2 \\ &\to \quad y = x - x^2 + \frac{2}{3}x^3 \end{aligned}$$

- **Leibniz (1684) y Jacob Bernoulli (1690):** problema de la tangente inversa, se busca una curva $y(x)$ cuya tangente AB es dada,

$$y' = -\frac{y}{\sqrt{a^2 - y^2}}$$

1.1.2. Ejemplo: existencia y unicidad

Considere el **Problema de Valor Inicial** (PVI)

$$\begin{cases} y' = |y|^\alpha, \quad \alpha \in (0,1) \\ y(0) = 0. \end{cases}, \qquad f(x,y) = |y|^\alpha,\ \alpha \in (0,1).$$

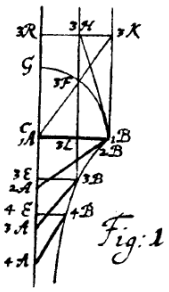


Fig. 2.2. Illustration from Leibniz (1693)

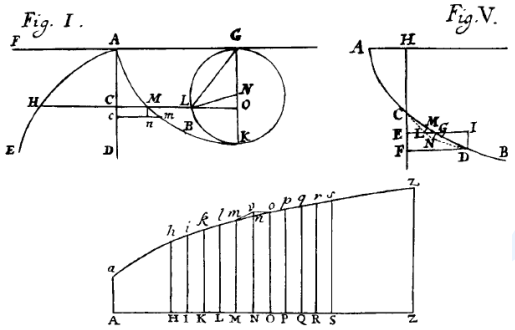


Fig. 2.4. Solutions of the variational problem (Joh. Bernoulli, Jac. Bernoulli, Euler)

Se puede corroborar que para todo real no negativo c

$$y_c(x) = \begin{cases} (1-\alpha)^{\frac{1}{1-\alpha}}\,(x-c)^{\frac{1}{1-\alpha}}, & c \le x < \infty \\ 0, & 0 \le x < c \end{cases}$$

es una solución del problema de valor inicial sobre $[0, \infty)$. Es decir, el problema tiene infinitas soluciones (una por cada valor que pueda tomar c). Observemos que si $\alpha \ge 1$ el PVI tiene solución única.

Bajo que condiciones podemos asegurar la unicidad?

1.1.3. Un Teorema general de existencia

Aproximación de Taylor del problema:

$$y'(x) = f(x,y), \quad y(x_0) = y_0, \quad y(x) = ?$$

Taylor:

$$\begin{aligned} y_1 - y_0 &= (x_1 - x_0) f(x_0, y_0) \\ y_2 - y_1 &= (x_2 - x_1) f(x_1, y_1) \\ &\vdots \\ y_n - y_{n-1} &= (x_n - x_{n-1}) f(x_{n-1}, y_{n-1}) \end{aligned}$$

Si definimos $h_i = x_{i+1} - x_i$, $i = 0, \ldots, n-1$, tenemos los llamados **polígonos de Euler** (solución método de Euler)

$$y_h(x) = y_i + (x - x_i) f(x_i, y_i), \qquad \text{para } x_i \le x \le x_{i+1}$$

Lema 1.1.3.1. *Asuma que $|f|$ es acotado por una constante A sobre el dominio*

$$D = \{(x, y) : \quad x_0 \leq x \leq X, \, |y - y_0| \leq b\}.$$

Si $X - x_0 \leq b/A$, entonces la solución numérica (x_i, y_i) se mantiene en D para cualquier subdivisión y tenemos que

$$|y_h(x) - y_0| \leq A|x - x_0|$$
$$|y_h(x) - (y_0 + (x - x_0)f(x_0, y_0))| \leq \varepsilon|x - x_0|,$$

si

$$|f(x, y) - f(x_0, y_0)| \leq \varepsilon, \quad \textit{sobre } D.$$

Demostración. Tenemos que

$$|y_n - y_0| = |\sum_{i=0}^{n-1}(x_{i+1} - x_i)f(x_i, y_i)| \leq A|\sum_{i=0}^{n-1}(x_{i+1} - x_i)| = A|x_n - x_0|$$

Si $x_i \leq x \leq x_{i+1}$, entonces

$$|y_h(x) - y_0| \leq |y_i - y_0| + |x - x_i||f(x_i, y_i)| \leq A|x_i - x_0| + A|x - x_i| = A|x - x_0|.$$

Demuestre la segunda desigualdad □

Lema 1.1.3.2. *Para una subdivisión fija h sea $y_h(x)$ y $z_h(x)$ los polígonos de Euler correspondientes a los valores iniciales y_0 y z_0, respectivamente. Si*

$$\left|\frac{\partial f}{\partial y}(x, y)\right| \leq L$$

en una región convexa que contiene a los puntos $(x, y_h(x))$ y $(x, z_h(x))$ para todo $x_0 \leq x \leq X$. Entonces,

$$|z_h(x) - y_h(x)| \leq \exp(L(x - x_0))|z_0 - y_0|$$

Demostración. Observe que

$$\left|\frac{\partial f}{\partial y}(x, y)\right| \leq L \implies |f(x, z) - f(x, y)| \leq L|z - y|$$

entonces, como

$$|z_1 - y_1| = |z_0 - y_0 + (x_1 - x_0)\left(f(x_0, z_0) - f(x_0, y_0)\right)|$$

se sigue que

$$|z_1 - y_1| \leq (1 + (x_1 - x_0)L)\,|z_0 - y_0| \leq \exp(L(x_1 - x_0))|z_0 - y_0|$$

el resultado sigue despues de aplicar estos para $i = 2, 3, ...$ □

1.1.4. Teorema polígonos de Euler

Teorema 1.1.4.1. *Sea $f(x, y)$ una función contínua con $|f|$ acotado por A y satisface la condición de Lipschitz en*

$$D = \{(x, y) : \quad x_0 \leq x \leq X,\ |y - y_0| \leq b\}.$$

Si $X - x_0 \leq b/A$, entonces tenemos :

1. *Para $|h| \to 0$, los polígonos de Euler $y_h(x)$ convergen uniformemente a una función contínua $\varphi(x)$.*
2. *La función $\varphi(x)$, es contínua y diferenciable, y es solución del PVI en $x_0 \leq x \leq X$.*
3. *No existe otra solución del PVI en $x_0 \leq x \leq X$.*

Demostración. Sea $\varepsilon > 0$. La función f es uniformemente continua en D (compacto), entonces existe $\delta > 0$ tal que

$$|u_1 - u_2| \leq \delta, \quad |v_1 - v_2| \leq A\delta \implies |f(u_1, v_1) - f(u_2, v_2)| \leq \varepsilon.$$

Suponga que la subdivisión de puntos satisface que $|h| = |x_{i+1} - x_i| \leq \delta$.

Dada la configuración inicial con h, consideramos primero una subdivisión $h(1)$, la cual se obtiene agregando nuevos puntos solo al primer subintervalo. Se sigue que, para la nueva solución $y_{h(1)}(x_1)$, tenemos

$$|y_{h(1)}(x_1) - y_h(x_1)| \leq \varepsilon |x_1 - x_0|$$

Así aplicamos el Lemma y obtenemos

$$|y_{h(1)}(x_1) - y_h(x_1)| \leq \exp(L(x - x_1))(x_1 - x_0)\varepsilon, \quad \text{para } x_1 \leq x \leq X.$$

Ahora repetimos el proceso y subdivimos el intervalo (x_1, x_2), denotamos esta subdivisiónpor $h(2)$. Obtenemos

$$|y_{h(2)}(x) - y_{h(1)}(x) \leq \exp(L(x - x_2))(x_2 - x_1)\varepsilon, \quad \text{para } x_2 \leq x \leq X.$$

Repetimos el proceso hasta el último intervalo y denotamos por $\widehat{h}$ el refinamiento final, obtenemos para $x_i < x \leq x_{i+1}$

$$\begin{aligned} |y_{\widehat{h}}(x) - y_h(x)| &\leq \varepsilon \left(\exp(L(x - x_1))(x_1 - x_0) + \ldots + \exp(L(x - x_i))(x_i - x_{i-1})\right) + \varepsilon(x - x_i) \\ &\leq \varepsilon \int_{x_0}^{x} \exp(L(x - s))ds \end{aligned}$$

$$= \frac{\varepsilon}{L}(\exp(L(x - x_0)) - 1)$$

Si ahora tenemos dos subdivisiones h y $\tilde{h}$ que satisfacen $|h| \leq \delta$ y $|\tilde{h} \leq \delta$. Entonces introducimos una tercera subdivisión $\widehat{h}$, la cual es una refinamiento de h y $\tilde{h}$ y aplicamos el análisis anterior dos veces. Así, obtenemos

$$|y_h(x) - y_{\tilde{h}}(x)| \leq 2\frac{\varepsilon}{L}(\exp(L(x - x_0)) - 1)$$

Tenemos una sucesión de Cauchy uniforme de funciones contínuas. Para $\varepsilon > 0$, esto muestra que los polígonos de Euler convergen a una función contínua $\varphi(x)$.

Sea

$$\varepsilon(\delta) := \sup\{|f(u_1, v_1) - f(u_2, v_2)|;\ |u_1 - u_2| \leq \delta,\ |v_1 - v_2| \leq A\delta,\ (u_i, v_i) \in D\}$$

entonces, para $(x, y_h(x))$ y $x + \delta$ tenemos

$$|y_h(x + \delta) - y_h(x) - \delta f(x, y_h(x))| \leq \varepsilon(\delta)\delta$$

Tomando el límite $|h| \to 0$ obtenemos

$$|\varphi(x + \delta) - \varphi(x) - \delta f(x, \varphi(x))| \leq \varepsilon(\delta)\delta$$

Lo que muestra que φ es diferenciable con $\varphi'(x) = f(x, \varphi(x))$.

Sea $\psi(x)$ una segunda solución del PVI y suponga que la subdivisión h satisface $|h| \leq \delta$. Sea $y_h^{(i)}(x)$ el polígono de Euler con $(x_i, \psi(x_i))$

$$\psi(x) = \psi(x_i) + \int_{x_i}^{x} f(s, \psi(s))ds$$

$$|\psi(x) - y_h^{(i)}| \leq \varepsilon|x - x_i|, \quad x_i \leq x \leq x_{i+1}$$

Usando el segundo Lema deducimos que

$$|\psi(x) - y_h(x)| \leq \frac{\varepsilon}{L}(\exp(L(x - x_0) - 1))$$

tomando límite $|h| \to 0$ y $\varepsilon \to 0$ obtenemos que $\psi = \varphi$, de donde se tiene la unicidad. □

Observación

El Teorema anterior is un resultado de existencia y unicidad local. Sin embargo, si interpretamos el punto final de la solución como un nuevo valor inicial, entonces podemos aplicar el Teorema nuevamente y continuar con la solución. Repitiendo el procedimiento obtenemos:

Teorema 1.1.4.2. *Asuma que U es un conjunto abierto en $\mathbb{R}^2$ y sea f y $\partial f/\partial y$ contínuas en U. Entonces, para cada $(x_0, y_0) \in U$, existe una única solución del PVI la cual puede continuarse hasta la frontera de U.*

1.1.5. Estimación de error

Teorema 1.1.5.1. *Suponga que en una vecidad de la solución se satisface*

$$|f| \leq A, \quad \left|\frac{\partial f}{\partial y}\right| \leq L, \quad \left|\frac{\partial f}{\partial x}\right| \leq M.$$

Entonces tenemos el siguiente estimado de error de los polígonos de Euler

$$|y(x) - y_h(x)| \leq \frac{M + AL}{L} \left(\exp(L(x - x_0) - 1)|h|\right),$$

para un $|h|$ suficientemente pequeno.

Demostración. Para $|u_1 - u_2| \leq |h|$ y $|v_1 - v_2| \leq A|h|$ obtenemos

$$|f(u_1, v_1) - f(u_2, v_2)| \leq (M + AL)|h|$$

de donde se sigue el resultado. □

1.1.6. Teorema de existencia de Peano

Que pasa si no asumimos la condición de Lipschitz en el Teorema de existencia y unicidad?

Por ejemplo, consideremos el problema

$$y' = 4(\text{signo}(y)\sqrt{|y|} + \text{máx}\{0, x - \frac{|y|}{x}\} \cos(\frac{\pi \log(x)}{\log(2)})), \quad y(0) = 0$$

Esta función f es tal que satisface:

$$f(h, 0) = 4(-1)^i h, \qquad \text{para } h = 2^{-i},$$
$$f(x, y) = 4\text{signo}(y)\sqrt{|y|}, \qquad \text{para } |y| \leq x^2.$$

Así hay infinitas soluciones para este valor inicial. Los polígonos de Euler convergen para $h = 2^{-i}$ a $y = 4x^2$ si i es par y a $y = -4x^2$ si i es impar.

Teorema 1.1.6.1. *Sea $f(x,y)$ una función contínua y $|f|$ acotado por A en el dominio*

$$D = \{(x,y): \quad x_0 \le x \le X,\, |y - y_0| \le b\}.$$

Si $X - x_0 \le b/A$, entonces existe una subsucesión de la sucesión de pol'igonos de Euler la cual converge a una solución del PVI.

Demostración.
- Demostración original Peano.
- Reinterpretación por Arzela 1895
- Demostración moderna por Perro 1918, Hahn 1921.
- Ver Theorem 7.6, page 42 en Libro.

□

1.1.7. Teorema de Picard-Lindelof

Teorema 1.1.7.1. *Sea la función $(x,y) \mapsto f(x,y)$. Suponga que:*

- *Continua en $D = \{(x,y) : x_0 \le x \le x_M,\ y_0 - C \le y \le y_0 + C\}$.*
- *$|f(x,y_0)| \le K$, para $x_0 \le x \le x_M$.*
- **(Condición de Lipschitz):** *Existe $L > 0$ tal que*

$$|f(x,u) - f(x,v)| \le L|u - v| \qquad \forall (x,u),(x,v) \in D.$$

- *Se satisface que: $C \ge \frac{K}{L}\left(e^{L(x_M - x_0)} - 1\right)$*

Entonces, existe una única función $y \in C^1([x_0, x_M])$ solución del PVI en $[x_0, x_M]$. Además se tiene que

$$|y(x) - y_0| \le C, \qquad \forall x \in [x_0, x_M].$$

Demostración. Definimos una sucesión de funciones $\{y_n\}$ con

$$\begin{aligned} y_0(x) &= y_0 \\ y_n(x) &= y_0 + \int_{x_0}^{x} f(s, y_{n-1}(s))\, ds, \quad n = 1, 2, \ldots \end{aligned}$$

Como f es continua en D se sigue que cada función $y(x)$ es continua en $[x_0, x_M]$. Además, por la condición de Lipschitz

$$|y_{n+1}(x) - y_n(x)| \le \left| \int_{x_0}^{x} (f(s, y_n(s)) - f(s, y_{n-1}(s)))\, ds \right|$$

$$\leq L \int_{x_0}^{x} |y_n(s) - y_{n-1}(s)|\, ds \tag{1.1}$$

Por otro lado, asuma que para algún valor de n

$$|y_n(x) - y_{n-1}(x)| \leq \frac{K}{L} \frac{(L(x - x_0))^n}{n!}, \quad x_0 \leq x \leq x_M$$

$$|y_k(x) - y_0(x)| \leq \frac{K}{L} \sum_{j=1}^{k} \frac{(L(x - x_0)))^j}{j!}, \quad x_0 \leq x \leq x_M,\ k = 1, \ldots, n.$$

Observe que el caso $n = 1$ se satisface y que la hipótesis de inducción y el cuarto supuesto implican que

$$|y_k(x) - y_0| \leq \frac{K}{L} e^{L(x_M - x_0)} - 1 \leq C, \quad x_0 \leq x \leq x_M,\ k = 1, \ldots, n$$

Así, $(x, y_{n-1}(x)) \in D$ y $(x, y_n(x)) \in D$ para todo $x \in [x_0, x_M]$. Entonces, usando (1.1) y la hipótesis de inducción

$$|y_{n+1} - y_n(x)| \leq L \int_{x_0}^{x} \frac{K}{L} \frac{(L(s - x_0))^n}{n!} ds = \frac{K}{L} \frac{(L(x - x_0))^{n+1}}{(n+1)!}, \quad x \in [x_0, x_M].$$

Además se satisface que

$$\begin{aligned} |y_{n+1} - y_0| &\leq |y_{n+1}(x) - y_n(x)| + |y_n(x) - y_0| \\ &\leq \frac{K}{L} \frac{(L(x - x_0))^{n+1}}{(n+1)!} + \frac{K}{L} \sum_{j=1}^{n} \frac{(L(x - x_0)))^j}{j!} \\ &= \frac{K}{L} \sum_{j=1}^{n+1} \frac{(L(x - x_0)))^j}{j!}, \quad x \in [x_0, x_M], \end{aligned}$$

lo que concluye la inducción. Por lo tanto, como

$$\sum_{j=1}^{\infty} \frac{c^j}{j!} = e^c - 1,\ c \in \mathbb{R},$$

tenemos que para $c = L(x_M - x_0)$:

$$\sum_{j=1}^{\infty} \frac{(L(x_M - x_0))}{j!} = e^{L(x_M - x_0)} - 1.$$

Así, como además

$$|y_n(x) - y_{n-1}(x)| \leq \frac{K}{L} \frac{(L(x - x_0))^n}{n!}$$

lo que implica que la serie

$$\sum_{j=1}^{\infty} |y_j(x) - y_{j-1}(x)|$$

converge uniformemente en $[x_0, x_M]$ y el límite es continuo, el cual llamaremos $y(x)$.

$$\begin{aligned} y(x) &= \lim_{n\to\infty} y_{n+1}(x) = y_0 + \lim_{n\to\infty} \int_{x_0}^{x} f(s, y_n(s))ds \\ &= y_0 + \int_{x_0}^{x} f(s, \lim_{n\to\infty} y_n(s))ds \\ &= y_0 + \int_{x_0}^{x} f(s, y(s))ds. \end{aligned}$$

De aquí y es continua y diferenciable con $y'(x) = f(x, y(x)), \quad y(x_0) = y_0$. Además $(x, y(x)) \in D$.

Para demostrar unicidad, supongamos por contradicción que existen 2 soluciones y, z. Entonces

$$y(x) - z(x) = \int_{x_0}^{x} (f(s, y(s)) - f(s, z(x)))\, ds$$

$$|(y(x) - z(x)| \leq L \int_{x_0}^{x} |y(s) - z(s)|ds.$$

Si $m = \max_{x\in[x_0,x_M]} |y(x) - z(x)|$, entonces

$$|y(x) - z(x)| \leq mL(x - x_0) \leq m\frac{(L(x - x_0))^2}{2!} \quad \dots \quad \leq m\frac{(L(x - x_0))^k}{k!} \xrightarrow{k\to\infty} 0$$

lo que contradice la suposición inicial. Luego, la solución es única.

□

1.2. Aplicaciones del Teorema de Picard

Ejemplo 1

Consideremos la siguiente ecuación diferencial **lineal**, para $p, q \in \mathbb{R}$:

$$y' = p \cdot y + q, \quad \text{esto es} \quad f(x, y) = p \cdot y + q$$

Vemos que la condición de Lipschtz se cumple, pues

$$|f(x, u) - f(x, v)| \leq |p| \cdot |u - v|$$

Además tomando $K = |py_0| + |q|$, tenemos que $|f(x, y_0)| \leq K$.

Entonces, para todo intervalo $[x_0, x_M]$, las condiciones se satisfacen escogiendo C suficientemente grande

$$C \geq \frac{K}{L}(e^{L(x_M - x_0)} - 1)$$

Ejemplo 2

Consideremos el PVI:

$$y' = y^2,\ y(0) = 1,\ x \in [0, x_M], \quad \text{esto es} \quad f(x, y) = y^2$$

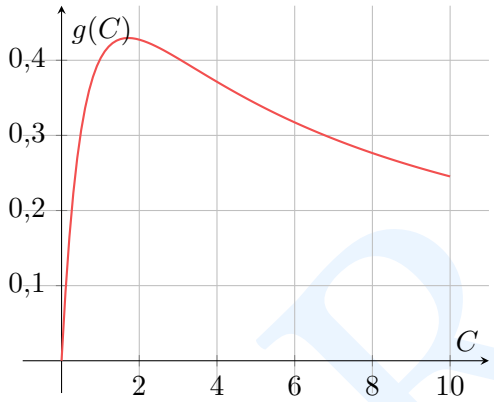
Como $|y - 1| \leq C$ tenemos que

$$|u^2 - v^2| = |u + v||u - v| \leq (|u| + |v|)|u - v| \leq 2(1 + C)|u - v|,$$

es decir, se satisface la condición de Lipschitz. También $|f(x, 1)| = |1^2| \leq 1$.

Para asegurar la unicidad de la solución necesitamos imponer que

$$C \geq \frac{1}{2(1+C)}\left(e^{2(1+C)x_M} - 1\right) \iff x_M \leq \frac{\ln(1 + 2C(1 + C))}{2(1 + C)} =: g(C)$$



La función de la gráfica alcanza su máximo en $C = 1{,}714$ con $x_M \leq 0{,}43$. Así, el Teorema garantiza la existencia y unicidad para $x \in [0, 0{,}43]$. Recuerde que las condiciones son sólo suficientes y no necesarias.

Por otro lado notemos que

$$y' = y^2 \iff \frac{y'}{y} = 1 \iff \int \frac{dy}{y^2} = \int dx \iff -\frac{1}{y} = x + Cte \iff y = \frac{-1}{x + Cte}$$

Reemplazando la condición inicial $y(0) = 1$ obtenemos que $Cte = -1$. Por lo tanto la solución del PVI es

$$y(x) = \frac{1}{1 - x}, \quad 0 \leq x < 1$$

Por otroa lado, el Teorema de Picard nos garantiza que esta solución es única en $[0, 0{,}43]$, sin embargo esta función está bien definida en $[0, 1)$ ¿es única en este intervalo?

Ejercicio: Método de Picard

Podemos usar la sucesión creada en la demostración del Teorema de Picard para construir aproximaciones de la solución.

Consideremos el problema anterior

$$y' = py + q, \;\; y(0) = 1.$$

Encontremos la sucesión $\{y_n\}$:

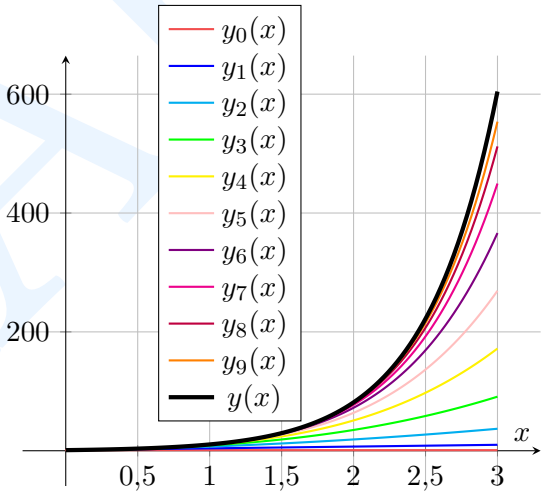
$$\begin{aligned} y_0 &= 1 \\ y_1 &= 1 + \int_0^x (p \cdot 1 + q) ds = 1 + (p+q)x \\ y_2 &= 1 + \int_0^x p(1 + (p+q)x) + q ds = 1 + (p+q)x + \frac{p(p+q)}{2} x^2 \end{aligned}$$

Encuentre la forma general de la sucesión y_n

Resolviendo la ecuación diferencial obtenemos que

$$y(x) = \frac{(p+q)e^{px} - q}{p}$$

Dando valores a $p = 2$ y $q = 1$, uno puede ver como se comportan las soluciones



1.3. Discretización

Aproximaremos numéricamente la solución del PVI.

Suponga que el PVI satisface las condiciones del Teorema de Picard en $[x_0, x_M]$. Dividamos el intervalo usando punto de malla uniforme de la siguiente forma

$$h = \frac{x_M - x_0}{N}, \; N \in \mathbb{N}, \quad x_n = x_0 + h \cdot n, \quad \text{para } n \in \{0, 1, \ldots, N\}$$

Aproximamos $y(x_n) \sim y_n$

1.3.1. Métodos de paso simple

Consideramos aproximaciones del tipo: $\begin{cases} y_{n+1} &= y_n + h\Phi(x_n, y_n; h) \\ y_0 &= y(x_0) \end{cases}$

Un ejemplo de este tipo de métodos es el de **método de Euler explícito**.

$$\left.\begin{array}{rcll} y' & = & f(x,y) & , x \in [x_0, x_M] \\ y(x_0) & = & y_0 & \end{array}\right\} \qquad \begin{array}{c} y_{n+1} = y_n + hf(x_n, y_n) \\ y_0 = y(x_0) \end{array}$$

Por la expansión de Taylor tenemos

$$\begin{aligned} y(x_{n+1}) = y(x_n + h) &= y(x_n) + hy'(x_n) + O(h^2) \\ &= y(x_n) + hf(x_n, y(x_n)) + O(h^2). \end{aligned}$$

Error del método de paso simple

Definición 1.3.1.1. *Definimos el **error global** del método de paso simple* $y_{n+1} = y_n + h\Phi(x_n, y(x_n); h)$ *por*

$$e_n := y(x_n) - y_n.$$

*Definimos el **error de truncación** por*

$$T_n := \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h).$$

Teorema 1.3.1.2. *Considere el método de paso simple y asuma que* Φ *es continua y satisface la condición de Lipschitz respecto a la segunda variable, es decir, existe* $L_\Phi > 0$ *tal que para* $0 < h \leq h_0$

$$|\Phi(x, u; h) - \Phi(x, v; h)| \leq L_\Phi |u - v|, \quad \forall (x, u), (x, v) \in D$$

donde $D = \{(x, y) : x_0 \leq x \leq x_M, |y - y_0| < C\}$*. Entonces, asumiendo que* $|y_n - y_0| < C$ *para todo* n *se sigue que*

$$|e_n| \leq \frac{T}{L_\Phi} \left(e^{L_\Phi (x_n - x_0)} - 1 \right), n \in \{0, 1, \ldots, N\},$$

donde $T = \max\limits_{0 \leq n \leq N-1} |T_n|$.

Demostración. Tenemos que para $n \in \{0, 1, \ldots, N-1\}$

- $y(x_{n+1}) = y(x_n) + h\Phi(x_n, y(x_n); h) + hT_n$
- $y_{n+1} = y_n + h\Phi(x_n, y_n; h)$

Luego

$$\begin{aligned} e_{n+1} &= y(x_{n+1}) - y_{n+1} \\ &= e_n + h(\Phi(x_n, y(x_n); h) - \Phi(x_n, y_n; h)) + hT_n \\ \Rightarrow |e_{n+1}| &\leq |e_n| + hL_\Phi|e_n| + hT_n \quad \text{(Condición de Lipschitz).} \end{aligned}$$

Afirmamos que $|e_n| \leq \frac{T}{L_\Phi}((1 + hL_\Phi)^n - 1)$. Lo demostraremos por inducción sobre n. Para $n = 1$, tenemos que

$$|e_1| \leq |e_0| + hL_\Phi|e_0| + hT_0 = hT_0 \leq \frac{T}{L_\Phi}\left((1 + hL_\Phi)^1 - 1\right)$$

Luego, asumiendo que $|e_n| \leq \frac{T}{L_\Phi}((1 + hL_\Phi)^n - 1)$ vemos que

$$\begin{aligned} |e_{n+1}| &\leq |e_n| + hL_\Phi|e_n| + hT_n \\ &\leq \frac{T}{L_\Phi}((1 + hL_\Phi)^n - 1) + hL_\Phi\frac{T}{L_\Phi}((1 + hL_\Phi)^n - 1) + hT_n \\ &\leq \frac{T}{L_\Phi}((1 + hL_\Phi)^n - 1)(1 + hL_\Phi) + \frac{hT}{L_\Phi}L_\Phi \\ &= \frac{T}{L_\Phi}\left((1 + hL_\Phi)^{n+1} - 1 - hL_\Phi + hL_\Phi\right) \\ &= \frac{T}{L_\Phi}((1 + hL_\Phi)^{n+1} - 1) \end{aligned}$$

Y con esto, concluimos la inducción. Como $1 + hL_\Phi \leq e^{hL_\Phi}$, se sigue que

$$|e_n| \leq \frac{T}{L_\Phi}\left(e^{L_\Phi(x_n - x_0)} - 1\right).$$

□

1.3.2. Cota de error para el método de Euler

Recordemos que para el método de Euler explícito $\Phi(x_n, y_n; h) = f(x_n, y_n)$. Entonces, el error de truncación está dado por

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - f(x_n, y(x_n)) = \frac{y(x_{n+1}) - y(x_n)}{h} - y'(x_n)$$

Si asumimos que $y \in C^2([x_0, x_M])$ tenemos que

$$y(x_{n+1}) = y(x_n) + hy'(x_n) + \frac{h^2}{2}y''(\xi_n),\ \xi_n \in (x_0, x_M) \Rightarrow T_n = \frac{h}{2}y''(\xi_n)$$

Tomando $M_2 = \max\limits_{\xi\in[x_0,x_M]} |y''(\xi)|$, entonces $|T_n| < T = \frac{1}{2}hM_2$ y por lo tanto

$$|e_n| \leq \frac{M_2 h}{2}\left(\frac{e^{L(x_M-x_0)}-1}{L}\right)$$

Ejemplo

Consideremos el PVI

$$\begin{cases} y' = \tan^{-1}(y) & t \in (0,T), \\ y(0) = y_0 \end{cases}$$

Encuentre una cota para el error global e_n de la aproximación de Euler explícito.

Solución Queremos encontrar una cota superior para el error global e_n. Notemos que $f(x,y) = \tan^{-1}(y)$ es Lipschitz

$$|f(x,u) - f(x,v)| = \left|\frac{\partial f}{\partial y}(x,\xi)(u-v)\right| = \frac{1}{1+\xi^2}|u-v| \leq |u-v|$$

Además

$$y'' = \frac{d}{dx}(\tan^{-1}(y)) = \frac{1}{1+y^2}\cdot y' = \frac{\tan^{-1}(y)}{1+y^2}$$

$$\Rightarrow |y''(x)| \leq \frac{\pi}{2} := M_2$$

y así tenemos que $|e_n| \leq \frac{\pi}{4}h(e^{x_n}-1)$.

Pregunta

Dada una tolerancia $\varepsilon > 0$, podemos encontrar $h > 0$ que asegura que el error $e_n < \varepsilon$, para $n = 0, ..., N$.

Ejemplo

Considere el siguiente PVI

$$\begin{cases} y' = y^2 - \dfrac{x^4 - 6x^3 + 12x^2 - 14x + 9}{(1+x)^2}, \\ y(0) = 2. \end{cases}$$

1.3.3. Consistencia

Definición 1.3.3.1. *El método numérico $y_{n+1} = y_n + h\Phi(x_n, y_n; h)$ es* ***consistente*** *(con la EDO $y' = f(x, y)$) si el error de truncación es tal que para todo $\varepsilon > 0$ existe $h(\varepsilon) > 0$ para el cual*

$$|T_n| < \varepsilon \text{ cuando } 0 < h < h(\varepsilon)$$

y los puntos $(x_n, y(x_n)), (x_{n+1}, y(x_{n+1})) \in D$. Para los métodos de paso simple esto significa que es ***consistente*** *si y solo si*

$$\Phi(x, y; 0) = f(x, y)$$

Teorema 1.3.3.2. *Suponga que el PVI satisface las condiciones del Teorema de Picard y que su aproximación*

$$y_{n+1} = y_n + h\Phi(x_n, y_n; h)$$

cuando $h < h_0$ pertenece a D. Asuma que Φ es continua en $D \times [0, h_0]$ y satisface

- $\Phi(x, y; 0) = f(x, y)$
- $|\Phi(x, u; h) - \Phi(x, v; h)| \leq L_\Phi |u - v|$

Entonces, la sucesión $\{y_n\}$ converge a la solución del PVI como $x_n \to x$ cuando $h \to 0$.

Como resultado del Teorema el método entonces se dice **convergente**.

Demostración. Tenemos, de la definición de error de truncación

$$\begin{aligned}
T_n &= \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h) \\
&= \left(\frac{y(x_{n+1}) - y(x_n)}{h} - f(x_n, y(x_n)) \right) + (\Phi(x_n, y(x_n); 0) - \Phi(x_n, y(x_n); h)) \\
&= (y'(\xi_n) - y'(x_n)) + (\Phi(x_n, y(x_n); 0) - \Phi(x_n, y(x_n); h)) \\
&= \underbrace{\varepsilon/2}_{\text{para } h < h_1(\varepsilon)} + \underbrace{\varepsilon/2}_{\text{para } h < h_2(\varepsilon)}.
\end{aligned}$$

Así $|T_n| < \varepsilon$ para $h < \min\{h_1(\varepsilon), h_2(\varepsilon)\}$.

Entonces, aplicando la continuidad Lipschitz de Φ,

$$\begin{aligned}
|y(x) - y_n| &\leq |y(x) - y(x_n)| + |y(x_n) - y_n| \\
&\leq |y(x) - y(x_n)| + T \left(\frac{e^{L\Phi(x_M - x_0)} - 1}{L_\Phi} \right)
\end{aligned}$$

$$\leq |y(x) - y(x_n)| + \varepsilon \left(\frac{e^{L\Phi(x_M - x_0)} - 1}{L_\Phi} \right)$$

□

1.3.4. Precisión

Definición 1.3.4.1. *El método numérico $y_{n+1} = y_n + h\Phi(x_n, y_n; h)$ se dice que tiene **orden de precisión** p, si p es el entero positivo más grande tal que para toda curva solución $(x, y(x)) \in D$ suficientemente suave del PVI, existen constantes K y h_0 tales que*

$$|T_n| \leq Kh^p, \text{ para } 0 < h < h_0$$

Vimos en uno de los ejemplos anteriores, que para el método de Euler, si exigíamos a la solución $y \in C^2([x_0, x_M])$ entonces $|T_n| \leq Kh$, donde

$$K = \frac{1}{2} \max_{\xi \in [x_0, x_M]} |y''(\xi)|$$

Así, el **orden del método de Euler** es $p = 1$.

1.4. Métodos de la Regla del Trapecio y Runge-Kutta

Preguntas

- ¿Es el método de Euler explícito convergente?
- ¿Cual es el orden de precisión del método de Euler?

El método de Euler es convergente, ya que este es consistente y tiene cota de Lipschitz.

Observamos que

$$|T_n| = |hy''(\xi_n)| \leq Kh, \quad K \text{ es independiente de } h$$

1.4.1. Método de la Regla del Trapecio

La regla trapezoidal es un método de paso simple definido por la iteración

$$y_{n+1} = y_n + \frac{h}{2}(f(x_n, y_n) + f(x_{n+1}, y_{n+1}))$$

Observe que el método se deriva de

$$y(x_{n+1}) - y(x_n) = \int_{x_n}^{x_{n+1}} y'(x)\,dx = \int_{x_n}^{x_{n+1}} f(x, y(x))\,dx$$

Proposición 1.4.1.1. *El orden de la regla trapezoidal es $p = 2$.*

Demostración. Error de truncación:

$$\begin{aligned} T_n &= \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h) \\ &= \frac{y(x_{n+1}) - y(x_n)}{h} - \frac{1}{2}(f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))) \end{aligned}$$

Por Taylor

$$\begin{aligned} y(x_{n+1}) &= y(x_n) + hy'(x_n) + \frac{h^2}{2}y''(x_n) + O(h^3) \\ y'(x_n) &= f(x_n, y(x_n)) \\ f(x_{n+1}, y(x_{n+1})) &= y'(x_{n+1}) = y'(x_n) + hy''(x_n) + O(h^2) \end{aligned}$$

Entonces

$$\begin{aligned} T_n &= (y'(x_n) + \frac{h}{2}y''(x_n) + O(h^2)) - \frac{1}{2}(y'(x_n) + y'(x_n) + hy''(x_n) + O(h^2)) \\ &= O(h^2). \end{aligned}$$

Más específicamente:

$$|T_n| \leq \frac{1}{12} \max_{x \in [x_0, x_M]} |y'''(x)| h^2.$$

□

Proposición 1.4.1.2. *La Regla trapezoidal es convergente.*

Encontremos la constante de Lipschitz L_Φ. Primero, observemos que:

$$h\ \Phi(x_n, y_n; h) = \frac{h}{2}(f(x_n, y_n) + f(x_{n+1}, y_{n+1})) = \frac{h}{2}(f(x_n, y_n) + f(x_{n+1}, y_n + h\ \Phi(x_n, y_n; h)))$$

Así

$$\begin{aligned} &|\Phi(x_n, u; h) - \Phi(x_n, v; h)| \\ &= \frac{1}{2}|f(x_n, u) + f(x_n + h, u + h\ \Phi(x_n, u; h)) - f(x_n, v) - f(x_n + h, v + h\ \Phi(x_n, v; h))| \\ &\leq \frac{1}{2}|f(x_n, u) - f(x_n, v)| + \frac{1}{2}|f(x_n + h, u + h\ \Phi(x_n, u; h)) - f(x_n + h, v + h\ \Phi(x_n, v; h))| \\ &\leq \frac{1}{2}L_f|u - v| + \frac{1}{2}L_f|u + h\ \Phi(x_n, u; h) - v - h\ \Phi(x_n, v; h)| \end{aligned}$$

Demostración.

$$|\Phi(x_n,u;h)-\Phi(x_n,v;h)| \leq \frac{1}{2}L_f|u-v| + \frac{1}{2}L_f|u-v| + \frac{1}{2}L_f h|\Phi(x_n,u;h)-\Phi(x_n,v;h)|$$

De donde tenemos que:

$$|\Phi(x_n,u;h)-\Phi(x_n,v;h)| \leq \frac{L_f}{1-hL_f/2}|u-v|$$

Por lo tanto, si $1-hL_f/2>0$, podemos tomar $L_\Phi \leq \frac{L_f}{1-hL_f/2}$.

Consistencia

$$\Phi(x,y;0) = \frac{1}{2}(f(x,y)+f(x,y)) = f(x,y)$$

Además

$$|e_n| \leq \frac{T}{L_\Phi}(e^{L_\Phi(x_n-x_0)}-1) \underset{h\to 0}{\longrightarrow} 0$$

□

1.4.2. Implementación de la regla trapezoidal

La regla trapezoidal:

$$y_{n+1} = y_n + \frac{h}{2}\left(f(x_n,y_n)+f(x_{n+1},y_{n+1})\right)$$

Dado la iteración y_n, como resolvemos para y_{n+1}?

Este método corresponde a un método **implícito**, necesitamos resolver para y_{n+1}.

Método de Newton-Raphson (encontrar aproximaciones de raíces de una función real):

$$\textbf{Resolver para z:} \quad F(z) = y_n + \frac{h}{2}\left(f(x_n,y_n)+f(x_{n+1},z)\right) - z = 0$$

$$z^{(k+1)} = z^{(k)} - \frac{F(z^{(k)})}{F'(z^{(k)})}, \quad \text{con } z^{(0)} = y_n + hf(x_n,y_n)$$

Necesitamos calcular $F'(z^{(k)})$, es decir $\dfrac{\partial f}{\partial y}(z^{(k)})$.

1.5. Problema no lineal, caso vectorial.

Resolver el problema no lineal $F(\mathbf{z}) = \mathbf{0}$, para $F : \mathbb{R}^n \to \mathbb{R}^n$.

- Newton-Raphson: $\mathbf{z}^{(k+1)} = z^{(k)} - \left(\nabla F(z^{(k)})\right)^{-1} F(\mathbf{z}^{(\mathbf{k})})$
- Métodos de quasi-Newton: $J_k \approx \nabla F(\mathbf{z}^{(k)})$. Broyden 1965.

$$\begin{aligned}
\Delta \mathbf{z}^{(k)} &= \mathbf{z}^{(k)} - \mathbf{z}^{(k-1)} \\
\Delta F_k &= F(\mathbf{z}^{(k)}) - F(\mathbf{z}^{(k-1)}) \\
J_k &= J_{k-1} + \frac{\Delta F_k - J_{k-1}\Delta \mathbf{z}_k}{\|\Delta \mathbf{z}_k\|^2} \Delta \mathbf{z}_k^\top \\
J_k^{-1} &= J_{k-1}^{-1} + \frac{\Delta z_k - J_{k-1}^{-1}\Delta F_k}{\|\Delta F_k\|^2} \Delta F_k^\top \\
z^{(k+1)} &= z^{(k)} - J_k^{-1} F(z^{(k)})
\end{aligned}$$

- BFGS.

$$\begin{aligned}
J_k &= J_{k-1} + \frac{\Delta F_k \Delta F_k^\top}{\Delta F_k^\top \Delta z_k} - \frac{J_{k-1}\Delta z_k (J_{k-1}\Delta z_k)^\top}{\Delta z_k^\top J_{k-1} \Delta z_k} \\
J_k^{-1} &= \left(I - \frac{\Delta z_k \Delta F_k^\top}{\Delta F_k^\top \Delta z_k}\right) J_{k-1}^{-1} \left(I - \frac{\Delta F_k \Delta z_k^\top}{\Delta z_k^\top \Delta F_k}\right) + \frac{\Delta z_k \Delta z_k^\top}{\Delta F_k^\top \Delta z_k}
\end{aligned}$$

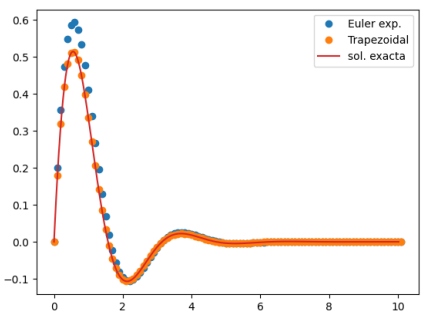
```
from scipy.optimize import fmin\_bfgs
```

1.5.1. Ejemplo 1. Cuando todo funciona

Consideremos el siguiente PVI: $y' = -y + 2\exp(-t)\cos(2t), \quad y(0) = 0.$

Comparamos los dos métodos vistos hasta ahora

- $y^{n+1} = y^n + h(-y^n + 2\exp(-t_n)\cos(2t_n)$
- $(1 + \frac{h}{2})y^{n+1} = (1 - \frac{h}{2})y^n + \frac{h}{2}\left(2\exp(-t_n)\cos(2t_n) + 2\exp(-t_{n+1})\cos(2t_{n+1})\right)$



1.5.2. Ejemplo 2.

Consideremos el siguiente PVI

$$y' = \ln(3)\left(y - \lfloor y \rfloor - \frac{3}{2}\right), \quad y(0) = 0$$

solución exacta: $y(x) = -\lfloor x \rfloor + \frac{1}{2}(1 - 3^{x - \lfloor x \rfloor}), \quad x \geq 0$

1.5.3. Ejercicios

1. Derive la regla del punto medio implícito

$$y_{n+1} = y_n + hf(x_n + \frac{1}{2}h, \frac{1}{2}(y_n + y_{n+1}))$$

   Pruebe que es de orden 2 y convergente.

2. La forma general, método theta

$$y_{n+1} = y_n + (\Delta t)(\theta f(t_n, y_n) + (1 - \theta) f(t_{n+1}, y_{n+1})), \quad \theta \in [0, 1]$$

   a) $\theta = 1 \longrightarrow$ Euler explícito

   b) $\theta = \frac{1}{2} \longrightarrow$ Regla trapezoidal

   c) $\theta = 0 \longrightarrow$ Euler implícito

   es convergente y de orden 1 si $\theta \neq \frac{1}{2}$.

1.6. Métodos de Runge-Kutta

Carl Runge, Martin Wilhelm Kutta $\sim$ 1900.

Aproximación de orden más alto. Nuestro objetivo es introducir métodos de paso simple con orden precisión mas alto. Para ilustrar su derivación, observemos lo siguiente

$$\begin{aligned} y(x_{n+1}) &= y(x_n) + \int_{x_n}^{x_{n+1}} f(\tau, y(\tau)) d\tau \\ &= y(x_n) + h \int_0^1 f(x_n + \tau h, y(x_n + \tau h)) d\tau \end{aligned}$$

Regla de Cuadratura $y(x_{n+1}) = y(x_n) + h \sum_{i=1}^{s} b_i f(x_n + c_i h, y(x_n + c_i h))$.

En la expresión de arriba aparecen valores desconocidos. La idea de los método de Runge-Kutta es aproximarlos por valores $\xi_j \approx y(x_n + c_j h)$.

Así, escribimos la estructura de la iteración de paso simple con

$$\Phi(x_n, y_n; h) = \sum_{i=1}^{s} b_i k_i; \qquad k_i = f(x_n + c_i h, \xi_i)$$

Aproximación de orden más alto. Consideremos como primer ejemplo el siguiente esquema

$$\textbf{(RK2)} \qquad y_{n+1} = y_n + h(ak_1 + bk_2), \quad \begin{cases} k_1 &= f(x_n, y_n) \\ k_2 &= f(x_n + \alpha h, y_n + \beta h k_1) \end{cases}$$

Esto corresponde a un método de paso simple con:

$$\Phi(x_n, y_n; h) = a\, f(x_n, y_n) + b\, f(x_n + \alpha h, y_n + \beta h f(x_n, y_n))$$

Observemos que el método de Euler corresponde a $a = 1, b = 0$.

Consistencia de RK2:

$$\Phi(x, y; 0) = f(x, y) \longrightarrow \boxed{\text{a + b = 1}}$$

Orden de RK2:

- $y'(x_n) = f(x_n, y(x_n)) = f$
- $y''(x_n) = \dfrac{\partial f}{\partial x} + \dfrac{\partial f}{\partial y} y' = \dfrac{\partial f}{\partial x} + \dfrac{\partial f}{\partial y} f$
- $y'''(x_n) = \dfrac{\partial^2 f}{\partial x^2} + \dfrac{\partial^2 f}{\partial x \partial y} f + \left(\dfrac{\partial^2 f}{\partial x \partial y} + \dfrac{\partial^2 f}{\partial y^2} f \right) f + \dfrac{\partial f}{\partial y} \left(\dfrac{\partial f}{\partial x} + \dfrac{\partial f}{\partial y} f \right)$

y escribimos

$$\begin{aligned} \Phi(x_n, y(x_n); h) = af + b(f + \alpha h \frac{\partial f}{\partial x} + \beta h \frac{\partial f}{\partial y} f + \frac{1}{2}(\alpha h)^2 \frac{\partial^2 f}{\partial x^2} \\ + \alpha\beta h^2 \frac{\partial^2 f}{\partial x \partial y} + \frac{1}{2}(\beta h)^2 \frac{\partial^2 f}{\partial y^2} + O(h^3)) \end{aligned}$$

Entonces

$$\begin{aligned} T_n &= \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h) \\ &= f + \frac{h}{2}\left(\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right) + \frac{h^2}{6}\left(\frac{\partial^2 f}{\partial x^2} + 2 \frac{\partial^2 f}{\partial x \partial y} f + \frac{\partial^2 f}{\partial y^2} f^2 + \frac{\partial f}{\partial y}\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial x} f \right) \right) \\ &- (af + b(f + \alpha h \frac{\partial f}{\partial x} + \beta h f \frac{\partial f}{\partial y} + \frac{1}{2}(\alpha h)^2 \frac{\partial^2 f}{\partial x^2} + \alpha\beta h^2 f \frac{\partial^2 f}{\partial x \partial y} \end{aligned}$$

$$+\frac{1}{2}(\beta h)^2 f^2 \frac{\partial^2 f}{\partial y^2})) + O(h^3).$$

Eliminar $f \longrightarrow 1 - a - b = 0$

Eliminar $h \longrightarrow \frac{1}{2}\left(\frac{\partial f}{\partial x} + f\frac{\partial f}{\partial y}\right) - b\alpha\frac{\partial f}{\partial x} - b\beta\frac{\partial f}{\partial y}f = 0$

Por lo tanto $b\alpha = b\beta = \frac{1}{2}$.

El método es de orden $p = 2$ si

$$\beta = \alpha, \quad a = 1 - \frac{1}{2\alpha}, \quad b = \frac{1}{2\alpha}; \quad \alpha \neq 0$$

Pregunta: ¿Puede ser el método de orden $p = 3$?

Ejemplos: Orden $p = 2$

- Euler modificado $\quad (\alpha = \frac{1}{2})$

$$y_{n+1} = y_n + hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hf(x_n, y_n))$$

- Euler mejorado $\quad (\alpha = 1)$

$$y_{n+1} = y_n + \frac{1}{2}h(f(x_n, y_n) + f(x_n + h, y_n + hf(x_n, y_n)))$$

1.6.1. Esquema de Runge-Kutta explícitos

Regla de Cuadratura:

$$y_{n+1} = y_n + h\sum_{j=1}^{s} b_j f(x_n + c_j h, y(x_n + c_j h)) = y_n + h\sum_{j=1}^{s} b_j f(x_n + c_j h, \xi_i))$$

No sabemos los valores de $y(x_n + c_j h) \approx \xi_j$. Fórmulas **explícitas**

Aproximaciones:

$$\begin{aligned}
\xi_1 &= y_n \\
\xi_2 &= y_n + h\, a_{21}\, f(x_n, \xi_1) \\
\xi_3 &= y_n + h\, a_{31}\, f(x_n, \xi_1) + h\, a_{32}\, f(x_n + c_2 h, \xi_2) \\
&\vdots \\
\xi_s &= y_n + h\sum_{i=1}^{s-1} a_{s,i} f(x_n + c_i h, \xi_i)
\end{aligned}$$

para $i = 1, 2, ..., s$, s es el número de pasos del método.

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i$$
$$k_i = f(x_n + c_i h, \xi_i)$$
$$\xi_i = y_n + h \sum_{j=1}^{i-1} a_{ij} f(x_n + c_j h, \xi_j)$$

para $i = 1, ..., s$.

Diagrama de Butcher

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

Diagramas de Butchet ERK

$$\sum_{i=1}^{v} a_{ji} = c_j,\ j = 1, \ldots, v, \quad \text{orden}\, p > 1$$

Ejemplos:

$$\begin{array}{c|cc} 0 & & \\ 1/2 & 1/2 & \\ \hline & 0 & 1 \end{array} \qquad \begin{array}{c|cc} 0 & & \\ 2/3 & 2/3 & \\ \hline & 1/4 & 3/4 \end{array} \qquad \begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & 1/2 & 1/2 \end{array}$$

$$\begin{array}{c|ccc} 0 & & & \\ 1/2 & 1/2 & & \\ 1 & -1 & 2 & \\ \hline & 1/6 & 2/3 & 1/6 \end{array} \qquad \begin{array}{c|ccc} 0 & & & \\ 2/3 & 2/3 & & \\ 2/3 & 0 & 2/3 & \\ \hline & 1/4 & 3/8 & 3/8 \end{array} \qquad \begin{array}{c|cccc} 0 & & & & \\ 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ 1 & 0 & 0 & 1 & \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array}$$

Ejercicio

A partir de los diagramas de Butcher anterior escriba los métodos de Runge-Kutta asociados.

1.6.2. Métodos de Runge-Kutta implícitos

$$\xi_j = y_n + h \sum_{i=1}^{s} a_{ji} f(x_n + c_i h, \xi_i)\,, j = 1, \ldots, v$$
$$y_{n+1} = y_n + h \sum_{j=1}^{s} b_j f(x_n + c_j h, \xi_j)$$

Ejemplo:

$$\xi_1 = y_n + \frac{h}{4}(f(x_n, \xi_1) - f(x_n + \frac{2}{3}h, \xi_2))$$
$$\xi_2 = y_n + \frac{h}{12}(5f(x_n, \xi_1) + 5f(x_n + \frac{2}{3}h, \xi_2))$$
$$y_{n+1} = y_n + \frac{h}{4}(f(x_n, \xi_1) + 3f(x_n + \frac{2}{3}h, \xi_2))$$

Diagrama de Butcher

0	1/4	−1/4
2/3	1/4	5/12
	1/4	3/4

1.7. Métodos de pasos múltiples lineales

Método de Euler de primer orden $\longrightarrow$ **Métodos de Runge-Kutta** de orden más alto

Paso simple: $y_{n+1} = \text{función}(y_n)$

Alternativamente, podemos obtener métodos de orden más alto usando múltiples pasos anteriores para calcular una nueva aproximación.

Ejemplo. Si tenemos los puntos x_{n-1}, x_n, x_{n+1},

$$\begin{aligned} y(x_{n+1}) &= y(x_{n-1}) + \int_{x_{n-1}}^{x_{n+1}} y'(x)\,dx \\ &= y(x_{n-1}) + \int_{x_{n-1}}^{x_{n+1}} f(x, y(x))\,dx \\ &\approx y(x_{n-1}) + \frac{(x_{n+1} - x_{n-1})}{6}\left(f(x_{n-1}, y(x_{n-1})) + 4f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))\right) \end{aligned}$$

Para llegar a ordenes mas altos se pueden ocupar métodos como Runge-Kutta, donde se requieren mas evaluaciones de la función f para avanzar un paso, o se pueden ocupar métodos de pasos múltiples donde se ocupan mas puntos para poder integrar.

Definición 1.7.0.1. *Dada una sucesión de puntos x_n con paso h, consideramos el método de k-pasos lineal general*

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f(x_{n+j}, y_{n+j})$$

con coeficientes $\{\alpha\}_{j=0}^{k}$ y $\{\beta\}_{j=0}^{k}$ son constantes reales, tales que $\alpha_k \neq 0$ y α_0 u β_0 no son ambas 0

- $\beta_{n+k} = 0 \rightarrow$ *método explícito*
- $\beta_{n+k} \neq 0 \rightarrow$ *método implícito*

Ejemplos de métodos de pasos múltiples

1. Adams-Bashforth, orden 4 explicito

$$y_{n+4} = y_{n+3} + \frac{1}{24}h(55f_{n+3} - 59f_{n+2} + 37f_{n+1} - 9f_n)$$

2. Adams-Moulton, orden 3 implícito

$$y_{n+3} = y_{n+2} + \frac{1}{24}h(9f_{n+3} + 19f_{n+2} - 5f_{n+1} + 1f_n)$$

Usamos la notación $f_{n+j} = f(x_{n+j}, y_{n+j})$.

Pregunta: ¿Como derivamos estas fórmulas ? $\rightarrow$ Ayudantía.

Pregunta: ¿Como calculamos los primeros k-pasos del método ?

No hay un método definido pero se puede ocupar Runge-Kutta del orden requerido para aproximas los primeros k-pasos

1.7.1. Cero- estabilidad

Definición 1.7.1.1. *Un método de k-pasos lineal se dice* ***cero-estable*** *si existe una constante K tal que para toda $\{y_n\}$ y $\{z_n\}$ que han sido generadas por la misma formula pero por valores iniciales distintos $y_0, ..., y_{k-1}$ y $z_0, ..., z_{k-1}$ respectivamente, tenemos*

$$|y_n - z_n| \leq K \max_{j\in\{0,...,k-1\}} |y_j - z_j|, \quad \text{para } x_n \leq x_M, \ y\, h \to 0$$

Raíces del polinomio característico.

Lema 1.7.1.1. *Considere la relación de secuencia lineal (asociada al problema homogéneo)*

$$\alpha_k y_{n+k} + ... + \alpha_0 y_n = 0, \quad n = 0, 1, 2, ..., N, \tag{1.2}$$

con $\alpha_k \neq 0, \alpha_0 \neq 0, \alpha_j \in \mathbb{R}$ y defina el ***polinomio característico*** $\rho(z) = \sum_{j=0}^{k} \alpha_j z^j$.

Sean $\{z_r\}_{r=1}^{l}, l \leq k$, las ***raíces distintas*** *del polinomio $\rho(z)$ y sea m_r la multiplicidad de z_r, con $\sum_{r=1}^{l} m_r = k$. Si una sucesión $\{y_n\} \subseteq \mathbb{C}$ satisface (1.2), entonces*

$$y_n = \sum_{r=1}^{l} p_r(n) z_r^n, \quad \forall n \geq 0$$

donde p_r es un polinomio, con variable n, de grado $(m_r - 1)$, $1 \leq r \leq l$.

Demostración. Consideremos el caso cuando todas las raíces $z_1, z_2, ..., z_k$ son simples. Como $\alpha_0 \neq 0$ entonces $z_i \neq 0$, $i = 1, ..., k$. Como

$$\rho(z_r) = 0, \implies y_n = (z_r^n) \text{ satisface (1.2)}$$

Para mostrar que una solución de (1.2) es una combinación lineal de $z_1^n, ..., z_k^n$, debemos probar que estas son l.i. Suponemos entonces que existen $C_1, ..., C_k$ tales que

$$C_1 z_1^n + C_2 z_2^n + ... + C_k z_k^n = 0, \quad n = 0, 1, 2$$

Implica que

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_k \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{k-1} & z_2^{k-1} & \dots & z_k^{k-1} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Matriz de Vandermonde

$$D = \begin{vmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_k \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{k-1} & z_2^{k-1} & \dots & z_k^{k-1} \end{vmatrix} = \prod_{r<s} (z_s - z_r) \neq 0 \implies C_1 = C_2 = \ldots = C_k = 0$$

Ahora, si y_n es solución de (1.2), entonces existen únicas $C_1, ..., C_k$ tales

$$y_m = C_1 z_1^m + C_2 z_2^m + ... + C_k z_k^m, \quad m = 0, 1, ..., k-1.$$

Esto implica que al sustituir en (1.2) para $n = 0$

$$\begin{aligned} 0 =& \alpha_k y_k + \alpha_{k-1}(C_1 z_1^{k-1} + ... + C_k z_k^{k-1}) + ... + \alpha_0(C_1 + ... + C_k) \\ =& \alpha_k y_k + C_1(\rho(z_1) - \alpha_k z_1^k) + ... + C_k(\rho(z_k) - \alpha_k z_k^k) \\ =& \alpha_k \left(y_k - (C_1 z_1^k + ... + C_k z_k^k) \right) \\ \implies & y_k = (C_1 z_1^k + ... + C_k z_k^k). \end{aligned}$$

□

Condición de Raíces

Teorema 1.7.1.2. *Un método de paso múltiple es cero-estable para cualquier problema de valor inicial donde f satisface las condiciones del Teorema de Picard, si y solo si, todas las raíces de $\rho(z)$ del método están dentro del disco unitario cerrado en el plano complejo, con toda raíz en el disco unitario es simple.*

Esta condición, se conoce como la condición de raíces.

Demostración. ($\Longrightarrow$) Consideramos el método de k-pasos lineal, aplicado al problema $y' = 0$, esto es:

$$\alpha_k y_{n+k} + ... + \alpha_0 y_n = 0$$

Por el Lema anterior, toda solución de esta ecuación tiene la forma

$$y_n = \sum_{r=1}^{l} p_r(n) z_r^n,$$

donde z_r es una ráiz, de multiplicidad $m_r \geq 1$, del primer polinomio característico ρ, y el polinomio p_r es de grado $m_r - 1$.

Si $|z_r| > 1$, entonces existen valores iniciales $y_0, y_1, ..., y_{k-1}$ para los cuales la solución correspondiente crece como $|z_r|^n$.

Si $|z_r| = 1$, y la multiplicidad $m_r > 1$, la solución crece como $n^{m_r - 1}$.

En cualquiera de los casos, las soluciones crecen cuando $n \to \infty$ lo que implica que el método no es cero estable.

($\Longrightarrow$) Ver página 353, W. Gautschi, Numerical Analysis: an Introduction.

Considerando los valores iniciales $y_0, y_1, ..., y_{k-1}$ □

1.7.2. Ejemplos, métodos cero estable

1. Método de Euler explícito

$$y_{n+1} = y_n + hf_n$$

$$y_{n+1} - y_n = hf_n \to \alpha_0 = -1; \alpha_1 = 1$$

$$\rho(z) = z - 1 \xrightarrow{\text{Raíz}} z_1 = 1$$

2. Método de Euler implícito

$$y_{n+1} = y_n + hf_{n+1}$$

3. Método de Regla trapezoidal

$$y_{n+1} = y_n + \frac{h}{2}(f_{n+1} + f_n)$$

4. Método de Adams-Bashforth

$$y_{n+4} = y_{n+3} + \frac{h}{24}(55f_{n+3} - 59f_{n+2} + 37f_{n+1} - 9f_n)$$

$$\alpha_0 = 0, \alpha_1 = 0, \alpha_2 = 0, \alpha_3 = -1, \alpha_4 = 1$$

$$\rho(z) = z^4 - z^{3-} = z^3(z - 1)$$

5. Método de Adams-Moulton

6. Método de Simpson

7. Determine si el siguiente método es cero estable

$$11y_{n+3} + 27y_{n+2} - 27y_{n+1} - 11y_n = 3h(f_{n+3} + 9f_{n+2} + 9f_{n+1} + f_n)$$

8. Determine si el siguiente método es cero estable

$$y_{n+3} + y_{n+2} - y_{n+1} - y_n = 2h(f_{n+2} + f_{n+1})$$

1.7.3. Consistencia

Definición 1.7.3.1. *El **error de truncación** de un método de k-pasos lineal se define por*

$$T_n = \left(\sum_{j=0}^{k}(\alpha_j y(x_{n+j}) - h\beta_j f(x_{n+j}, y(x_{n+j})))\right) / \left(h\sum_{j=0}^{k}\beta_j\right)$$

Observe que este corresponde al residual que se obtiene al evaluar la solució del PVI en el método.

Definición 1.7.3.2. *El metodo de k-pasos lineal se dice **consistente** con el PVI si el error de truncación es tal que*

$$\forall \epsilon > 0, \exists h(\epsilon) : \ |T_n| < \epsilon, \ \ \textit{para} \ \ 0 < h < h(\epsilon).$$

Segundo polinomio característico

Definición 1.7.3.3. *Se define el segundo polinomio característico asociado al método de k-pasos lineal por:*

$$\sigma(z) = \sum_{j=0}^{k}\beta_j z^j$$

Observe que

$$\sigma(1) = \sum_{j=0}^{k}\beta_j$$

entonces

$$T_n = \frac{1}{h\sigma(1)}\sum_{j=0}^{k}(\alpha_j y(x_{n+j}) - h\beta_j f(x_{n+j}, y(x_{n+j})))$$

1.7.4. Condiciones para la consistencia

A continuación estudiamos condiciones para la consistencia. Usamos la expansión de Taylor y la ecuación

$$y(x_{n+j}) = y(x_n) + (jh)y'(x_n) + \frac{(jh)^2}{2!}y''(x_n) + \frac{(jh)^3}{3!}y'''(x_n) + \ldots$$
$$f(x_{n+j}, y(x_{n+j})) = y'(x_{n+j})$$

Entonces, tenemos

$$\begin{aligned} T_n &= \frac{1}{h\sigma(1)}\sum_{j=0}^{k}(\alpha_j y(x_{n+j}) - h\beta_j f(x_{n+j}, y(x_{n+j}))) \\ &= \frac{1}{h\sigma(1)}(C_0 y(x_n) + C_1 h y'(x_n) + C_2 h^2 y''(x_n) + \ldots + C_q h^q y^{(q)}(x_n) + \ldots) \end{aligned}$$

Donde

$$\begin{aligned} C_0 &= \sum_{j=0}^{k}\alpha_j \\ C_1 &= \sum_{j=0}^{k}(j\alpha_j - \beta_j) \\ C_2 &= \sum_{j=0}^{k}\left(\frac{j^2}{2!}\alpha_j - j\beta_j\right) \\ &\vdots \\ C_q &= \sum_{j=0}^{k}\left(\frac{j^q}{q!}\alpha_j - \frac{j^{q-1}}{(q-1)!}\beta_j\right) \end{aligned}$$

Verificar

Por lo tanto, para que el método sea consistente necesitamos que:

$$C_0 = 0 \quad \text{y} \quad C_1 = 0$$

Observe

- $C_0 = \sum_{j=0}^{k}\alpha_j = \rho(1)$
- $C_1 = \sum_{j=0}^{k}(j\alpha_j - \beta_j) = \rho'(1) - \sigma(1)$

Consistente $\begin{cases} \rho(1) &= 0 \\ \rho'(1) &= \sigma(1) \end{cases}$

Orden

Definición 1.7.4.1. *El método de paso múltiple lineal se dice de orden p, si p es el entero positivo mas grande tal que, para cualquier solución suficientemente suave en D del PVI, existen constantes K y h_0 tales que*

$$|T_n| \leq Kh^p \text{ para } 0 < h \leq h_0$$

para cualquiera de los $k+1$ puntos $(x_n, y(x_n)), ..., (x_{n+k}, y(x_{n+k}))$

Proposición 1.7.4.2. *El método es de orden p si y solos si*

$$C_0 = C_1 = C_2 = ... = C_p = 0, \text{ y } C_{p+1} \neq 0$$

Ejercicio

Determine $b \in \mathbb{R}$ para que el método

$$y_{n+3} + (2b-3)(y_{n+2} - y_{n+1}) - y_n = hb(f_{n+2} + f_{n+1})$$

sea

- Cero-estable
- de orden 4

Además deduzca el orden mayor del método cero-estable.

Ejercicio: Método de 2-pasos explicito mas preciso

Determine los parámetros tales que el método de 2 pasos tenga el orden de precisión mas alto posible

$$\alpha_2 y_{n+2} + \alpha_1 y_{n+1} + \alpha_0 y_n = h(\beta_1 f_{n+1} + \beta_0 f_n)$$

1.7.5. Equivalencia de Dahlquist

Valores iniciales consistentes: $y_j = \eta_j = \eta_j(h)$ con $y_j \to y_0$ cuando $h \to 0$.

Teorema 1.7.5.1. *Para un método de k-pasos lineal que es **consistente** con el PVI, donde f se asume que satisface la condición de Lipschitz, y con valores iniciales consistentes, la **cero-estabilidad** es necesaria y suficiente para **convergencia**. Además si la solución y tiene una derivada continua de orden $p+1$ y error de truncación $\mathcal{O}(h^p)$, entonces el error global del método*

$$e_n = y(x_n) - y_n$$

es de orden $\mathcal{O}(h^p)$

1.7.6. Barrera de Dahlquist

Teorema 1.7.6.1. *El orden de precisión de un método cero-estable de k-pasos no puede exceder:*

- $k+1$ *si* k *es impar*
- $k+2$ *si* k *es par*

Ejemplos

- El Teorema de Barrera de Dahlquist indica que cuando $k=1$, el orden de precisión de un método cero estable no puede ser mayor que 2. La regla del trapecio es de orden 2 y es cero estable.
- Muestre que el siguiente método de 2 pasos es cero-estable y que su orden es 4

$$y_{n+2} = y_n + \frac{h}{3}(f_{n+2} + 4f_{n+1} + f_n)$$

- Determine el orden y la cero-estabilidad del siguiente método de 3-pasos

$$11y_{n+3} + 27y_{n+2} - 27y_{n+2} - 11y_n = 3h(f_{n+3} + 9f_{n+2} + 9f_{n+1} + f_n)$$

  Es de orden 6, no puede ser cero estable

1.8. Sistemas rígidos o stiff

1.8.1. Problema rígido escalar

En esta clase se introducirán lo que se conoce como sistemas rígidos. Considere el ejemplo escalar modelo

$$\begin{cases} y' = \lambda y, \\ y(0) = y_0, \end{cases}$$

cuya solución está dada trivialmente por $y(x) = y_0 e^{\lambda x}$.

Observe que la solución exacta es **exponencialmente decreciente** si $\lambda < 0$,

¿podemos garantizar que todo esquema numérico sea también decreciente?

Consideremos el esquema de Euler explícito:

$$\begin{cases} y' = \lambda y, \\ y(0) = y_0, \end{cases} \qquad \Longrightarrow \qquad y_{n+1} = y_n + h\lambda y_n$$

Luego

$$y_{n+1} = (1 + h\lambda)y_n = (1 + h\lambda)^n y_0.$$

Para garatinzar que la solución numérica sea decreciente necesitamos que:

$$|1 + h\lambda| < 1 \implies 0 < h|\lambda| < 2$$

En caso contrario, la aproximación numérica oscilará con magnitud creciente.

Ver ejemplo: $y' = -20y, y(0) = 1.$ Comparar con Euler implicito.

1.8.2. Sistemas rígidos

Sea la matriz $A \in \mathbb{R}^{m\times m}$ y consideremos el sistema lineal de

$$\begin{cases} \mathbf{y}' = A\mathbf{y}, \\ \mathbf{y}(0) = \mathbf{y}_0. \end{cases}$$

Asumimos, por simplicidad, que todos los valores propios tienen multiplicidad algebraica simple. Sea $MAM^{-1} = \Lambda$ la diagonalización de A. Si hacemos el cambio de variable $\mathbf{z} = M\mathbf{y}$, entonces

$$\begin{cases} \mathbf{z}' = \Lambda\mathbf{z}, \\ \mathbf{z}(0) = M\mathbf{y}_0. \end{cases}$$

De esta manera, nos quedan m EDOs **desacopladas** con solución

$$\mathbf{z} = (z_j), \qquad z_j(x) = z_j(0)e^{\lambda_j x}, \quad j = 1, 2 \dots m,$$

Hacemos el análisis de comportamiento de la solución. Si los valores propios λ_j son reales y negativos, entonces obsrvamos que

$$\lim_{x\to\infty} \|\mathbf{z}(x)\| \to 0 \quad \text{implica} \quad \lim_{x\to\infty} \|\mathbf{y}(x)\| \to 0$$

Al aplicar el esquema de Euler explícito, obtenemos que:

$$\mathbf{y}^{n+1} = \mathbf{y}^n + hA\mathbf{y}^n = (I + hA)\mathbf{y}^n = M(I + h\Lambda)M^{-1}\mathbf{y}^n$$

$$\mathbf{z}^{n+1} = (I + h\Lambda)\mathbf{z}^n \qquad \implies \qquad z_j^{n+1} = z_j^n + h\lambda_j z_j^n$$

necesitaremos $h|\lambda_j| < 2$, $j = 1, 2 \dots m$, para asegurar el decaimiento de la aproximación numérica.

Ejemplo: que tan malo puede ser?

Considere el PVI

$$\begin{cases} \mathbf{y}' = A\mathbf{y}, \\ \mathbf{y}(0) = \mathbf{y}_0. \end{cases}, \quad \text{con} \quad A = \begin{bmatrix} -8003 & 1999 \\ 23988 & -6004 \end{bmatrix}, \quad \mathbf{y}_0 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

Los valores propios de A son $\lambda_1 = -7$ y $\lambda_2 = -14000$, por lo cual la solución exacta al problema está dada por

$$\mathbf{y}(x) = \begin{bmatrix} e^{-7x} \\ 4e^{-14000x} \end{bmatrix}$$

Sistema lineal rígido

De manera general, un sistema de ecuaciones diferenciales ordinarias lineal **rígido** se caracteriza por tener todos los valores propios de A con parte real negativa, a la vez que la razón entre el valor propio más grande y más pequeño en magnitud es grande también. Otros conceptos asociados a rigidez:

- métodos numéricos son numéricamente inestables.
- ecuación tiene solución con rápidas variaciones.

Ejemplo sistema lineal stiff

Considere el sistema: $\mathbf{y}' = A\mathbf{y}, \quad \text{con} \quad \Lambda = \begin{bmatrix} -100 & 1 \\ 0 & -1/10 \end{bmatrix}$

Observe que: $A = V\Lambda V^{-1}, \quad \text{con } V = \begin{bmatrix} 1 & 1 \\ 0 & 999/10 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} -100 & 0 \\ 0 & -1/10 \end{bmatrix}$

Así las soluciones se escribe como

$$\mathbf{y}(t) = \exp(tA)\mathbf{y}_0 = V\exp(t\Lambda)V^{-1}\mathbf{y}_0 = \mathbf{x}_1\exp(-100t) + \mathbf{x}_2\exp(-t/10) \approx \mathbf{x}_2\exp(-t/10)$$

Las iteraciones del método de **Euler explícito** son: $\mathbf{y}^n = (I + hA)\mathbf{y}^0 = V(I + h\Lambda)V^{-1}\mathbf{y}_0$
Calculamos

$$(I + h\Lambda)^n = \begin{bmatrix} (1 - 100h)^n & 0 \\ 0 & (1 - h/10)^n \end{bmatrix}$$

Así

$$\mathbf{y}^n = \mathbf{x}_1(1 - 100h)^n + \mathbf{x}_2(1 - h/10)^n$$

Observamos que, si $h > 1/50$, entonces $1 - 100h < -1$, y así la iteración de Euler crece geométricamente en maginitud.

Suponga que elegimos una condición inicial idética al vector propio correspondiente al valor propio $-1/10$, esto es: $\mathbf{y}_0 = \begin{bmatrix} 1 \\ 999/10 \end{bmatrix}$

Entonces, en aritmética exacta $\mathbf{x}_1 = 0$, $\mathbf{x}_2 = \mathbf{y}_0$. Así la iteración de Euler es

$$\mathbf{y}^n = (1 - h/10)^n \mathbf{y}_0.$$

Así, esperaríamos que todos los cálculos no dieran bien

Regla trapezoidal

Si ahora analizamos las iteraciones de la regla trapezoidal, tenemos:

$$(I - \frac{h}{2}A)\mathbf{y}_1 = (I + \frac{h}{2}A)\mathbf{y}_0$$
$$(I - \frac{h}{2}A)\mathbf{y}_2 = (I + \frac{h}{2}A)\mathbf{y}_1 = (I + \frac{h}{2}A)(I - \frac{h}{2}A)^{-1}(I + \frac{h}{2}A)\mathbf{y}_0$$

en general

$$\mathbf{y}_n = \left((I - \frac{h}{2}A)^{-1}(I + \frac{h}{2}A)\right)^n \mathbf{y}_0$$

Así

$$\mathbf{y}^n = \mathbf{x}_1 \left(\frac{1 - 50h}{1 + 50h}\right)^n + \mathbf{x}_2 \left(\frac{1 - h/20}{1 + h/20}\right)^n$$

Por lo tanto:

$$\left|\frac{1 - 50h}{1 + 50h}\right|, \left|\frac{1 - h/20}{1 + h/20}\right| < 1 \quad \Longrightarrow \quad \lim_{n\to\infty} \mathbf{y}^n = 0$$

1.8.3. Problemas no lineales

Considere una EDO modelo de la forma $y'(x) = f(x, y)$, donde la función f es no-lineal. Ahora, tomando la linearización en torno a algún punto x_n, la EDO queda como

$$y'(x) = y'(x_n) + \frac{\partial f}{\partial x}(x_n, y(x_n))(x - x_n) + J(x_n)(y(x) - y(x_n)) + \dots \tag{1.3}$$

donde J corresponde a la matriz Jacobiana definida como $J(x_n) = \nabla_y f(x_n, y(x_n))$. El problema rígido se encuentra justamente en $J(x_n)$, donde uno esperaría que los valores propios tengan parte real negativa y la razón del mayor valor propio y el menor es grande.

Ejemplo:

Considere el problema de la reacción de Robertson

$$\begin{cases} x' = & -0{,}04x + 10^4 yz \\ y' = & 0{,}04x - 10^4 yz - 3 \cdot 10^7 y^2 \\ z' = & 3 \cdot 10^7 y^2 \end{cases}$$

Observación

- Que los valores propios de $J(x_n)$ tengan parte real negativa y la razón entre la parte real mayor y menor sea grande es considerado rígido o stiff.
- La matriz Jacobiana con valores propios con parte real negativa pero la solcuión

$$\lim_{x\to\infty} \|y(x)\|$$

crece exponencialmente

1.8.4. Estabilidad para métodos de paso múltiple

Recordamos la forma general de un método de pasos múltiples

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f(x_{n+j}, y_{n+j})$$

Aplicando este esquema al problema rígido escalar lineal ($y' = \lambda y$), se obtiene

$$\sum_{j=0}^{k} (\alpha_j - \lambda h \beta_j) y_{n+j} = 0.$$

Recordando las definiciones del primer y segundo polinomio característico

$$\rho(z) = \sum_{j=0}^{k} \alpha_j z^j \quad \text{y} \quad \sigma(z) = \sum_{j=0}^{k} \beta_j z^j$$

Definimos el **polinomio de estabilidad** como: $\pi(z_j \lambda h) = \sum_{j=0}^{k} (\alpha_j - \lambda h \beta_j) z^j = \rho(z) - \lambda h \sigma(z)$

A continuación se definen distintos tipos de estabilidad para métodos de pasos múltiples, cuyo comportamiento queda completamente caracterizado por los polinomios característicos o de estabilidad.

Definición 1.8.4.1. *Un método de paso múltiple se dice **absolutamente estable** para un valor dado de λh si cada raíz $z_r = z_r(\lambda h)$ del polinomio de estabilidad $\pi(\cdot; \lambda h)$ satisface $|z_r(\lambda h)| < 1$*

Definición 1.8.4.2. *La **región de estabilidad absoluta** de un método de paso múltiple es el conjunto de todos los puntos λh en el plano complejo para el cual el método es absolutamente estable.*

Definición 1.8.4.3 (A-estable)**.** *Un método de paso múltiple se dice **A-estable** si su región de estabilidad absoluta contiene al plano negativo complejo.*

Segunda barrera de Dahlquist

Teorema 1.8.4.4. *1. Ningún método de paso múltiple lineal explícito es A-estable*

2. *Ningún método de paso múltiple lineal y A-estable puede ser de orden mayor que 2*
3. *El método de paso múltiple lineal, A-estable y de orden 2 con la menor constante de error es la regla trapezoidal.*

Típicamente la condición de A-estabilidad se relaja para el diseño de esquemas numéricos, pero el objetivo sigue siendo obtener una región de estabilidad tan grande como sea posible.

Región de estabilidad: ejemplos

Ejemplos Se presentan varios ejemplos de región de estabilidad $\mathcal{D}$ para distintos esquemas.

1. Euler explícito:
$$\pi(z;\lambda h) = z - 1 - \lambda h$$
$$z_1 = 1 + \lambda h$$
$$\mathcal{D}_E := \{z \in \mathbb{C} : |1+z| < 1\}$$
2. Regla trapezoidal: (A-estable)
$$\pi(z;\lambda h) = z\left(1 - \frac{\lambda h}{2}\right) - \left(1 + \frac{\lambda h}{2}\right)$$
$$z_1 = \frac{1 + \lambda h/2}{1 - \lambda h/2}$$
$$\mathcal{D}_T := \left\{z \in \mathbb{C} : \left|\frac{1+z/2}{1-z/2}\right| < 1\right\}$$
3. Adam-Bashfort de 2 pasos:
$$\pi(z;\lambda h) = z^2 - z\left(1 - 3\frac{h\lambda}{2}\right) + \frac{h\lambda}{2}$$
$$z_{1,2} = \frac{1}{2}\left(\left(1 - 3\frac{h\lambda}{2}\right) \pm \sqrt{\left(1 - 3h\lambda + \frac{9}{4}(h\lambda)^2\right) - 2h\lambda}\right)$$
4. Adam-Moulton de 2 pasos:
$$\pi(z;\lambda h) = z^2\left(1 - 5\frac{\lambda h}{12}\right) - z\left(1 - 2\frac{h\lambda}{3}\right) + \frac{h\lambda}{12}$$

1.8.5. Estabilidad de métodos de Runge-Kutta

Para métodos de Runge-Kutta la estabilidad absoluta se define de la misma forma que para métodos de paso múltiple

$$y' = \lambda y,\ y(0) = 0 \qquad \lambda \in \mathbb{C},\ \mathrm{Re}(\lambda) < 0$$

Queremos que $y_n \to 0$ a medida que $n \to \infty$ para un valor $h\lambda$ fijo. El conjunto de $z = \lambda h$ en el plano complejo donde el método es absolutamente estable es la región de estabilidad. Los métodos explícitos tienen una pequeña región de estabilidad. Otra alternativa son los métodos de Runge-Kutta implícitos (IRK).

1.9. Fórmulas de diferenciación regresiva

Una alternativa para obtener métodos altamente estables son las fórmulas de diferenciación regresiva (BDF), escritas de manera general como

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h\beta_k f_{n+k}$$

En el caso $k = 1$ se recupera el esquema de Euler implícito. De manera general, los esquema BDF-k, para $k = 1 \ldots 5$ corresponden a

k	α_5	α_4	α_3	α_2	α_1	α_0	β_k
1					1	-1	1
2				3	-4	1	2
3			11	-18	9	-2	6
4		25	-48	36	-16	3	12
5	137	-300	300	-200	75	-12	60

1.10. A-estabilidad de métodos de Runge-Kutta

Consideremos el siguiente problema de valor inicial, donde $f(x, y) = \lambda y$, esto es,

$$y' = \lambda y, \qquad y(0) = 1.$$

Sea

$$\xi_j = y_n + h\lambda \sum_{i=1}^{s} a_{ij}\xi_i \qquad ,\ j = 1, 2, ..., s$$

Si definimos $\xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_s \end{bmatrix}$, $\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^s$ y $A = (a_{ij})$, entonces

$$\xi = \mathbf{1}y_n + h\lambda A\xi \qquad \Longrightarrow \qquad \xi = (I - h\lambda A)^{-1}\mathbf{1}y_n$$

Asuma que $(I - h\lambda A)$ es no singular (si no, no se podrá resolver!). Entonces

$$y_{n+1} = y_n + h\lambda \sum_{j=1}^{s} b_j \xi_j = \left(1 + h\lambda b^T (I - h\lambda A)^{-1}\mathbf{1}\right) y_n$$

1.10.1. Aproximaciones racionales y RK

Definición 1.10.1.1. *Definimos el conjunto de funciones racionales como*

$$\mathbb{P}_{\alpha/\beta} := \{p/q \text{ funciones racionales, con } p \in \mathbb{P}_\alpha,\ q \in \mathbb{P}_\beta\}$$

Lema 1.10.1.1. *Para cada método de Runge-Kutta de s-etapas existe $r \in \mathbb{P}_{s/s}$ tal que*

$$y_n = (r(h\lambda))^n \qquad , n = 0, 1, ...$$

Además, si el método es explícito, entonces $r \in \mathbb{P}_s$ (i.e., r no es racional, sino que un polinomio de grado s).

Demostración. Anteriormente teniamos: $y_{n+1} = y_n + h\lambda \sum_{j=1}^{s} b_j \xi_j = \left(1 + h\lambda b^T (I - h\lambda A)^{-1}\mathbf{1}\right) y_n$

Entonces $r : \mathbb{C} \to \mathbb{C}$:

$$r(z) := 1 + zb^T (I - zA)^{-1}\mathbf{1} \quad , z \in \mathbb{C}$$

satisface $y_n = (r(h\lambda))^n$. Debemos demostrar que $r \in \mathbb{P}_{s/s}$. Recuerde que:

- $(I - zA)^{-1} = \dfrac{\operatorname{adj}(I - zA)}{\det(I - zA)}$
- $b^T \operatorname{adj}(I - zA)1 \in \mathbb{P}_{s-1}$ y que $\det(I - zA) \in \mathbb{P}_s$

Por lo tanto $r(z) \in \mathbb{P}_{s/s}$.

Por último si A es estrictamente triangular inferior, entonces $I - zA$ es triangular inferior con 1s en la diagonal, y así $\det(I - zA) = 1$. □

Lema 1.10.1.2. *Suponga que una aplicación de un método numérico a la ecuación $y' = \lambda y$ produce una sucesión*

$$y_n = (r(h\lambda))^n \qquad , n = 0, 1, ...$$

donde r es una función arbitraria. Entonces la región de estabilidad del método está dado por:

$$\mathcal{D} = \{z \in \mathbb{C} \ : \ |r(z)| < 1\}$$

La demostración de este lema es trivial por la definición de A-estabilidad.

Corolario 1.10.1.2. *Ningún ERK (método de Runge-Kutta explícito) es A-estable.*

Demostración Sabemos que para ERK, $r \in \mathbb{P}_s$ y además que $r(0) = 1$. Sin embargo, los polinomios no puedes ser acotados por 1, excepto el polinomio constante $r(z) = c$, $c \in (-1, 1)$. Por lo tanto, no puede ser A-estable.

Ejemplos

Considere la siguiente forma general para el método de Runge-Kutta:

$$y_{n+1} = y_n + h\sum_{j=1}^{s} b_j f(x_n + c_j h, \xi_j), \qquad \xi_j = y_n + h\sum_{j=1}^{s} a_{ji} f(x_n + c_i h, \xi_i)$$

1) Euler explícito

$$y_{n+1} = y_n + hf(x_n, y_n) \qquad \Rightarrow \qquad s = 1 \qquad \begin{array}{c|c} c_1 & a_{11} \\ \hline & b_1 \end{array} = \begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

Como el método tiene un solo término de f, entonces $s = 1$. Igualando la forma de Euler explícito vemos que $f(x_n, y_n) = b_1 f(x_n + c_1 h, \xi_1)$, por lo que se debe cumplir que $b_1 = 1$, $x_n + c_1 h = x_n \;\Rightarrow\; c_1 = 0$ y que $\xi_1 = y_n$. Para esto último, se debe tener $a_{11} = 0$. De ahí los valores del diagrama de arriba.

Aquí la función racional r sería

$$r(z) = 1 + zb^T(I - zA)^{-1}\mathbf{1} = 1 + z$$

2) Regla trapezoidal $\quad y_{n+1} = y_n + \dfrac{h}{2}\left(f_n + f_{n+1}\right), \quad \xi_1 = y_n, \quad \xi_2 = y_{n+1}$

Entonces: $\quad s = 2, \quad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$, $\quad$ y $\quad r(z) = 1 + zb^T(I - zA)^{-1} = \dfrac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z}$

En efecto, $\quad (I - zA)^{-1} = \begin{bmatrix} 1 & 0 \\ -z/2 & 1 - z/2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ \frac{z/2}{1-z/2} & \frac{1}{1-z/2} \end{bmatrix}$

$$\begin{aligned} r(z) &= 1 + z\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ \frac{z/2}{1-z/2} & \frac{1}{1-z/2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 + \frac{z}{2}\left(1 + \frac{z}{2}\left(\frac{1}{1 - z/2}\right) + \frac{1}{1 - z/2}\right) \\ &= 1 + \frac{z}{2}\left(1 + \left(1 + \frac{z}{2}\right)\left(\frac{1}{1 - z/2}\right)\right) = \left(1 + \frac{z}{2}\right) + \left(1 + \frac{z}{2}\right)\left(\frac{z/2}{1 - z/2}\right) \\ &= \left(1 + \frac{z}{2}\right)\left(1 + \frac{z/2}{1 - z/2}\right) = \frac{1 + z/2}{1 - z/2}. \end{aligned}$$

3) Runge-Kutta implícito.

$$\begin{array}{c|cc} 0 & 1/4 & -1/4 \\ 2/3 & 1/4 & 5/12 \\ \hline & 1/4 & 3/4 \end{array} \qquad \begin{aligned} x(1-z/4)+y(z/4) &= 1 \\ -xz/4+y(1-5z/12) &= 0 \end{aligned}$$

$$(I-zA)^{-1} = \begin{pmatrix} 1-z/4 & z/4 \\ -z/4 & 1-5z/12 \end{pmatrix}^{-1} = \frac{1}{2(z^2-4z+6}\begin{pmatrix} 12-5z & -3z \\ 3z & 12-3z \end{pmatrix}$$

$$\begin{aligned} r(z) &= 1 + z\begin{pmatrix}1/4\\3/4\end{pmatrix}^T \frac{1}{2(z^2-4z+6}\begin{pmatrix} 12-5z & -3z \\ 3z & 12-3z \end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} \\ &= 1 + \frac{z/8}{z^2-4z+6}\begin{pmatrix}1\\3\end{pmatrix}^T\begin{pmatrix}12-8z\\12\end{pmatrix} = 1 - \frac{z/2}{z^2-4z+6}(3-2z+9) \\ &= \frac{z^2-4z+6+6z-z^2}{z^2-4z+6} = \frac{2z+6}{z^2-4z+6} = \frac{1+z/3}{1-2z/3+z^2/6} \end{aligned}$$

A-estabilidad: $z = \rho e^{i\theta} = \rho(\cos\theta + i\sin\theta)$, entonces $r(z) = \frac{1+\rho e^{i\theta}/3}{1-2/3(\rho e^{i\theta})+\rho^2 e^{2i\theta}/6}$

$$|r(z)| < 1 \quad \to \quad |1+\rho e^{i\theta}/3| < |1-2/3(\rho e^{i\theta}) + \rho^2 e^{2i\theta}/6|$$

$$\begin{aligned} |1+\rho e^{i\theta}/3|^2 &= (1+\frac{\rho}{3}\cos\theta)^2 + (\frac{\rho}{3}\sin\theta)^2 \\ &= 1 + \frac{2}{3}\rho\cos\theta + \frac{\rho^2}{9}\cos^2\theta + \frac{\rho^2}{9}\sin^2\theta = 1 + \frac{2}{3}\rho\cos\theta + \frac{\rho^2}{9} \end{aligned}$$

$$\begin{aligned} |1-2/3(\rho e^{i\theta}) + \rho^2 e^{2i\theta}/6| &= \left(1 - \frac{2}{3}\rho\cos(\theta) + \frac{\rho^2}{6}\cos(2\theta)\right)^2 + \left(-\frac{2}{3}\rho\sin(\theta) + \frac{\rho^2}{6}\sin(2\theta)\right)^2 \\ &= 1 + 2\left(-\frac{2}{3}\rho\cos(\theta) + \frac{\rho^2}{6}\cos(2\theta)\right) + \left(-\frac{2}{3}\rho\cos(\theta) + \frac{\rho^2}{6}\cos(2\theta)\right)^2 \\ &\quad + \frac{4}{9}\rho^2\sin^2(\theta) - 2\cdot\frac{2}{3}\rho\sin(\theta)\cdot\frac{\rho^2}{6}\sin(2\theta) + \frac{\rho^4}{36}\sin^2(2\theta) \end{aligned}$$

$$\begin{aligned} |1-2/3(\rho e^{i\theta}) + \rho^2 e^{2i\theta}/6| &= 1 - \frac{4}{3}\rho\cos(\theta) + \frac{\rho^2}{3}\cos(2\theta) + \frac{4}{9}\rho^2\cos^2(\theta) - \frac{2}{9}\rho^3\cos(\theta)\cos(2\theta) \\ &\quad + \frac{\rho^4}{36}\cos^2(2\theta) + \frac{4}{9}\rho^2\sin^2(\theta) - \frac{2}{9}\rho^3\sin(\theta)\sin(2\theta) + \frac{\rho^4}{36}\sin^2(2\theta) \\ &= 1 + \frac{4}{9}\rho^2 + \frac{\rho^4}{36} - \frac{4}{3}\rho\cos(\theta) + \frac{\rho^2}{3}\cos(2\theta) - \frac{2}{9}\rho^3(\cos(\theta)\cos(2\theta) + \sin(\theta)\sin(2\theta)) \end{aligned}$$

$$
\begin{aligned}
&= 1 + \frac{4}{9}\rho^2 + \frac{\rho^4}{36} - \frac{4}{3}\rho\cos(\theta) + \frac{\rho^2}{3}\cos(2\theta) - \frac{2}{9}\rho^3\cos(\theta)(\cos(2\theta) + 2\sin^2(\theta)) \\
&= 1 + \frac{4}{9}\rho^2 + \frac{\rho^4}{36} - \frac{4}{3}\rho\cos(\theta) + \frac{\rho^2}{3}\cos(2\theta) - \frac{2}{9}\rho^3\cos(\theta)(\cos(2\theta) + 1 - \cos(2\theta)) \\
&= 1 + \frac{4}{9}\rho^2 + \frac{1}{36}\rho^4 - \frac{4}{3}\rho\cos(\theta) + \frac{1}{3}\rho^2\cos(2\theta) - \frac{2}{9}\rho^3\cos(\theta)
\end{aligned}
$$

Por lo tanto $|r(z)| = |r(\rho e^{i\theta})| < 1$ si

$$
\begin{aligned}
1 + \frac{2}{3}\rho\cos(\theta) + \frac{\rho^2}{9} &< 1 + \frac{4}{9}\rho^2 + \frac{1}{36}\rho^4 - \frac{4}{3}\rho\cos(\theta) + \frac{1}{3}\rho^2\cos(2\theta) - \frac{2}{9}\rho^3\cos(\theta) \\
\cos(\theta)\left(\frac{2}{3}\rho + \frac{4}{3}\rho + \frac{2}{9}\rho^3\right) &< \frac{\rho^2}{3}(1 + \cos(2\theta)) + \frac{1}{36}\rho^4 \\
\underbrace{\cos(\theta)}_{<0}\left(1 + \frac{\rho^2}{9}\right)2\rho &< \frac{\rho^2}{3}\underbrace{(1 + \cos(2\theta))}_{>0} + \frac{1}{36}\rho^4 \quad \checkmark\checkmark
\end{aligned}
$$

Se tiene A-estabilidad si $|\theta + \pi| < \pi/2$, y en este rango se cumple que $\cos(\theta) < 0$, por lo que podemos afirmar que el método es A-estable.

Ejercicio

4) Analice la A-estabilidad del siguiente método de Runge-Kutta implícito

$1/3$	$5/12$	$-1/12$
1	$3/4$	$1/4$
	$3/4$	$1/4$

Nota: Observamos que no es necesario verificar que para todo $z \in \mathbb{C}^-$ una función r racional dada origina a un método A-estable. (Decimos esta última propiedad que r es **A-aceptable**).

Lema 1.10.1.3. *Sea r una función racional que no es constante. Entonces, $|r(z)| < 1$ para todo $z \in \mathbb{C}^-$ si y solo si todos los polos de r tienen parte real positiva y $|r(ix)| \leq 1$ para todo $x \in \mathbb{R}$.*

Ejemplo:

Considere $r(z) = \dfrac{1 + z/3}{1 - 2z/3 + z^2/6}$. Veamos que satisface las condiciones del Lema.

Calculamos los polos:

$$
(1 - \frac{2z}{3} + \frac{z^2}{6}) = 0 \implies z = \frac{2/3 \pm \sqrt{4/9 - 4/6}}{2} = \frac{2/3 \pm 1/3 \cdot i\sqrt{2}}{2/6} = 2 \pm i\sqrt{2} \quad \checkmark
$$

$$
\begin{aligned}
|r(ix)| \leq 1 &\iff |1+ix/3| \leq |1-2ix/3-x^2/6| \\
&\iff 1+x^2/9 \leq (1-x^2/6)^2+(2x/3)^2 \\
&\iff 1+x^2/9 \leq 1-x^2/3+x^4/36+4x^2/9 \\
&\iff 1+x^2/9 \leq 1+x^2/9+x^4/36 \quad \checkmark
\end{aligned}
$$

1.11. Métodos IRK (Runge-Kutta implícito) de Gauss-Legendre

$$
\xi_j = y_n + h\sum_{i=1}^{s} a_{ji} f(x_n + c_i h, \xi_i), \quad y_{n+1} = y_n + h\sum_{j=1}^{s} b_j f(t_n + c_j h, \xi_j)
$$

Observe que la condición $\displaystyle\sum_{i=1}^{s} a_{ji} = c_j, \quad j = 1, 2, ..., s,$

es necesaria para que el método converja con orden mayor o igual que 1.

Ejemplo IRK de dos fases

0	$1/4$	$-1/4$
$2/3$	$1/4$	$5/12$
	$1/4$	$3/4$

$$
\begin{aligned}
\xi_1 &= y_n + h\left(\frac{1}{4}f(t_n, \xi_1) + -\frac{1}{4}f(t_n + \frac{2}{3}h, \xi_2)\right) \\
\xi_2 &= y_n + h\left(\frac{1}{4}f(t_n, \xi_1) + \frac{5}{12}f(t_n + \frac{2}{3}h, \xi_2)\right) \\
y_{n+1} &= y_n + h\left(\frac{1}{4}f(t_n, \xi_1) + \frac{3}{4}f(t_n + \frac{2h}{3}, \xi_2)\right)
\end{aligned}
$$

Es convergente de orden 3.

Este tipo de métodos rienen s fases y son de **orden** $2s$. Considere el PVI

$$
y' = f(x, y) \quad \text{, para } x \geq x_0 \quad ; \quad y(x_0) = y_0
$$

Teniendo (x_n, y_n) queremos (x_{n+1}, y_{n+1}), con $x_{n+1} = x_n + h$.

Buscamos un polinomio $u \in \mathbb{P}_s$ tal que

$$
\begin{aligned}
u(x_n) &= y_n \\
u'(x_n + c_j h) &= f(x_n + c_j h, u(x_n + c_j h)) \quad , \ j = 1, 2, ..., s
\end{aligned}
$$

Un método de **colocación** consiste en encontrar un u y hacer

$$
y_{n+1} = u(x_{n+1})
$$

1.11.1. Colocación

Lema 1.11.1.1. *Sean*

$$q(x) := \prod_{j=1}^{s}(x - c_j), \qquad q_\ell(x) = \frac{q(x)}{x - c_\ell} \qquad ; \ \ell = 1, ..., s$$

y sean

$$a_{ji} := \int_0^{c_j} \frac{q_i(\tau)}{q_i(c_i)} d\tau, \qquad b_j := \int_0^1 \frac{q_j(\tau)}{q_j(c_j)} d\tau \qquad ; \ i, j = 1, ..., s$$

El método de colocación es idéntico al método IRK.

Demostración. Sea el polinomio de interpolación de **Lagrange** $r(x) := \sum_{\ell=1}^{s} \frac{q_\ell((x - x_n)/h)}{q_\ell(c_\ell)} w_\ell.$

Este satisface $r(x_n + c_\ell h) = w_\ell, \ \ell = 1, 2, ..., s.$

Si escogemos $w_\ell = u'(x_n + c_\ell h), \ \ell = 1, 2, ..., s$, entonces r y u', dos polinomios de $\mathbb{P}_{s-1}$, coinciden en s-puntos, por lo tanto $r = u'$, es decir

$$u'(x) = \sum_{\ell=1}^{s} \frac{q_\ell((x - x_n)/h)}{q_\ell(c_\ell)} f(x_n + c_\ell h, u(x_n + c_\ell h))$$

Integramos u:

$$\begin{aligned} u(x) &= y_n + \int_{x_n}^{x} \sum_{l=\ell}^{s} f(x_n + c_\ell h, u(x_n + c_\ell h)) \frac{q_\ell((\tau - x_n)/h)}{q_\ell(c_\ell)} d\tau \\ &= y_n + h \sum_{\ell=1}^{s} f(x_n + c_\ell h, u(x_n + c_\ell h)) \int_0^{(x - x_n)/h} \frac{q_\ell(\tau)}{q_\ell(c_\ell)} d\tau \end{aligned}$$

Sea $\xi_j := u(x_n + c_j h)$, para $j = 1, 2, ..., s$. Si reemplazamos $x = x_n + c_j h$ obtenemos que:

$$\xi_j = y_n + h \sum_{l=1}^{s} f(x_n + c_l h, u(x_n + c_l h)) a_{ji}$$

Finalmente, al reemplazar en $x = x_{n+1}$ obtenemos que

$$y_{n+1} = y_n + \sum_{j=1}^{s} b_j f(x_n + c_j h, \xi_j)$$

□

Todos los RK son de colocación?

Remark 1.11.1.1. *No todo método de Runge-Kutta se origina a partir de un método de colocación.*

Por ejemplo, para $s = 2$ y $c_1 = 0, c_2 = 2/3$. Esto implica que

$$q(t) = t(t - 2/3), \quad q_1(t) = t - 2/3, \quad q_2(t) = t$$

De donde el diagrama de Butcher del método tiene que ser

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 2/3 & 1/3 & 1/3 \\ \hline & 1/4 & 3/4 \end{array}$$

Por lo tanto el método del comienzo de esta sección no puede ser de colocación.

1.11.2. Orden método de colocación

Teorema 1.11.2.1. *Suponga que*

$$\int_0^1 q(\tau)\tau^j d\tau = 0, \qquad j = 0, 1, ..., m - 1$$

para algún $m \in \{0, ..., s\}$. Entonces el método de colocación es de orden $s + m$.

Corolario 1.11.2.2. *Sean c_1, c_2, ..., c_s los ceros del polinomio $\tilde{P}_s \in \mathbb{P}_s$ que son ortogonal con respecto a la función peso $w(x) \equiv 1$. Entonces el método de colocación es de orden $2s$.*

Polinomios de Gauss-Legendre en $[0, 1]$:

$$\tilde{P}_s(x) = \frac{(s!)^2}{(2s)!} \sum_{k=0}^{s} (-1)^{s-k} \binom{s}{k} \binom{s+k}{k} x^k$$

1.11.3. Ejemplos: Métodos de Gauss-Legendre

- $s = 1$, $\tilde{P}_1(x) = x - \frac{1}{2}$. Así $c_1 = \frac{1}{2}$, y $\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array}$ **Regla del punto medio implícito.**
- $s = 2$ $\tilde{P}_2(x) = x^2 - x + \frac{1}{6}$ así $c_1 = \frac{1}{2} - \frac{\sqrt{3}}{6}, c_2 = \frac{1}{2} + \frac{\sqrt{3}}{6}$

2 stages
4° orden

$$\begin{array}{c|cc} 1/2 - \sqrt{3}/6 & 1/4 & 1/4 - \sqrt{3}/6 \\ 1/2 + \sqrt{3}/6 & 1/4 + \sqrt{3}/6 & 1/4 \\ \hline & 1/2 & 1/2 \end{array}$$

- $s = 3$

$1/2 - \sqrt{15}/10$	$5/36$	$2/9 - \sqrt{15}/15$	$5/36 - \sqrt{15}30$
$1/2$	$5/36 + \sqrt{15}/24$	$2/9$	$5/26 - \sqrt{15}24$
$1/2 + \sqrt{15}/10$	$5/36 + \sqrt{15}30$	$2/ + \sqrt{15}/15$	$5/36$
	$5/18$	$4/9$	$5/18$

Lema 1.11.3.1 (Lema auxiliar para probar oden)**.** *Suponga que la sucesión $\{y_n\}$ la cual es generada aplicando un método de orden p a la ecuación lineal*

$$y' = \lambda y, \qquad y(0) = 1$$

con paso constante y con $y_n = r(\lambda h)^n$. Entonces

$$r(z) = \exp(z) + \mathcal{O}(z^{p+1}), \quad z \to 0$$

Demostración. Como $y_{n+1} = r(\lambda h)y_n$, y la solución exacta, comenzando en $y(t_n) = y_n$, es $\exp(\lambda h)y_n$, entonces

$$r(z) = \exp(z) + \mathcal{O}(z^{p+1})$$

por definición de orden. Así, decimos que r es de orden p. □

Teorema 1.11.3.1. *Dados los enteros $\alpha, \beta \geq 0$, existe una única función $\hat{r}_{\alpha/\beta}$ tal que:*

$$\hat{r}_{\alpha\beta} = \frac{\hat{p}_{\alpha\beta}}{\hat{q}_{\alpha\beta}}, \qquad \hat{q}_{\alpha/\beta}(0) = 1,$$

y $\hat{r}_{\alpha/\beta}$ es de orden $\alpha + \beta$. Las fórmulas de los polinomios del numerador y denominador son

$$\hat{p}_{\alpha/\beta}(z) = \sum_{k=0}^{\alpha} \binom{\alpha}{k} \frac{(\alpha + \beta - k)!}{(\alpha + \beta)!} z^k$$

$$\hat{q}_{\alpha/\beta}(z) = \sum_{k=0}^{\beta} \frac{(\alpha + \beta - k)!}{(\alpha + \beta)!} (-z)^k = \hat{p}_{\alpha/\beta}(-z)$$

Además, $\hat{r}_{\alpha/\beta}$ es el único elemento de $\mathbb{P}_{\alpha/\beta}$ de orden $\alpha+\beta$ y ninguna función en $\mathbb{P}_{\alpha/\beta}$ puede exceder este orden.

Observación: aproximaciónes de Padé

Las funciones $\hat{r}_{\alpha/\beta}$ son llamadas aproximaciones de Padé de la exponencial. Muchas de las funciones r que hemos visto son de este tipo:

$$\hat{r}_{1/0}(z) = 1 + z, \qquad \hat{r}_{1/1}(Z) = \frac{1 + z/2}{1 - z/2}, \qquad \hat{r}_{1/2}(z) = \frac{1 + z/3}{1 - 2z/2 + z^2/6}.$$

Definición 1.11.3.2. *Decimos que una función r es llamada **A-aceptable** si esta origina un método A-estable*

Las aproximaciones de Padé pueden ser clasificadas de acuerdo a esta definición. De inmediato observaos que la condición $\alpha \leq \beta$ es necesaria. Pero observe que esta condición no es suficiente. Por ejemplo, $\hat{r}_{0/3}$ no es A-aceptable.

Teorema de Wanner -Hairer - Norsett

Teorema 1.11.3.3. *La aproximación de Padé $\hat{r}_{\alpha/\beta}$ es A-estable si y sólo si $\alpha \leq \beta \leq \alpha + 2$.*

Corolario 1.11.3.4. *Los métodos IRK de Gauss-Legendre son A-estables para cada $s \geq 1$.*

Demostración.

Un método de Gauss-Legendre de s-fases es de orden $2s$. Así, la función r asociado pertenece a $\mathbb{P}_{s/s}$ y luego aproxima a la exponencial a orden $2s$. Así, este

1.12. Integración numérica geométrica

Este contenido se ve abordado en el libro: **A first course in the numerical analysis of differential equations**, *Iserles Capítulo 5.*

Otra referencia importante es el libro: **Geometric numerical integration**, *Hairer, Lubich, Wanner.*

En muchos problemas físicos es de crucial importancia mantener ciertas propiedades de la solución exacta del PVI. En ciertas ocasiones, se pueden perder al buscar la solución más precisa con el menor costo computacional.

Ejemplo:

$$\begin{aligned} y_1' &= y_2 y_3 \sin t - y_1 y_2 y_3 \\ y_2' &= -y_1 y_3 \sin t + \frac{1}{20} y_1 y_3 \\ y_3' &= y_1^2 y_2 - \frac{1}{20} y_1 y_2 \end{aligned}$$

Observe que

$$y_1 y_1' + y_2 y_2' + y_3 y_3' = 0$$

Por lo tanto:

$$\frac{1}{2} \frac{d}{dt} \left(y_1^2 + y_2^2 + y_3^2 \right) = 0$$

1.12.1. Sistemas Hamiltonianos

Algunos campos científicos donde son altamente utilizados:

- mecánica
- dinámica molecular

- mecánica de fluidos
- mecánica cuántica
- procesamiento de imágenes
- mecánica celestial
- ingeniería nuclear

Los sistemas Hamiltonianos son una forma de representar la energía que tiene un sistema, siendo la más común la energía mecánica expresada en forma una suma de energía cinética y potencial:

$$H = E_m = T + U$$

Ecuaciones de Hamilton: Sean $p(t), q(t) \in \mathbb{R}^d$ solución

$$\dot{p}_i := \frac{dp_i}{dt} = -\frac{\partial H(p,q)}{\partial q_i}$$
$$\dot{q}_i := \frac{dq_i}{dt} = \frac{\partial H(p,q)}{\partial p_i}$$

para $i = 1, 2, ..., d$, donde $H : \mathbb{R}^d \times \mathbb{R}^d$ es el **Hamiltoniano** (energía).

Lema 1.12.1.1. *El Hamiltoniano $H(p(t), q(t))$ se mantiene constante a lo largo de la trayectoria de la solución.*

Demostración.

$$\frac{\partial H}{\partial t}(p(t), q(t)) = \sum_{i=1}^{d} \left(\frac{\partial H}{\partial q_i}\dot{q}_i + \frac{\partial H}{\partial p_i}\dot{p}_i \right) = \sum_{i=1}^{d} \left(\frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q_i} \right) = 0$$

Decimos que el Hamiltoniano es invariante.

Ejemplo.

Consideremos la familia de ecuaciones diferenciales de orden 2

$$\ddot{y} + a(y) = 0$$

Si definimos las variables $q = y$, $p = \dot{y}$ entonces reescribimos la ecuación como

$$\dot{q} = p$$
$$\dot{p} = -a(q)$$

Hamiltoniano: $$H(p,q) = \frac{1}{2}p^2 - \int_0^q a(x)dx$$

Pregunta: La función H es un invariante para la ODE ¿Podemos tener métodos numéricos que lo preservan?

1.12.2. Intuición Geométrica: Método simpléctico

Los sistemas Hamiltonianos tienen otra característica, aún más importante, su flujo es **simpléctico**. Consideremos la ecuación

$$y' = f(y), \quad \text{mapeo de flujo} \quad \varphi_t(y_0) : \ y_0 \to y(t)$$

Esta definición se extiende a conjuntos medibles $\Omega \subset \mathbb{R}^d$:

$$\varphi_t(\Omega) = \{y(t) : y(0) \in \Omega\}$$

Ilustración. Consideremos el la ecuación $y'' + \sin y = 0$, y el conjunto $\Omega := \{y(t) : y(0) \in \Omega\}$

¡Área constante!

Esto es una *manifestación* de la geometría simpléctica; el mapeo de flujo para sistemas Hamiltonianos $d = 1$ preserva área.

1.12.3. Simplecticidad

Definición 1.12.3.1. *Una función $\varphi : \Omega \to \mathbb{R}^{2d}$, $\Omega \subset \mathbb{R}^d$ se dice simpléctica si*

$$\frac{\partial \varphi}{\partial y}(y)^T J \frac{\partial \varphi}{\partial y}(y) = J$$

donde el operador $J \in \mathbb{R}^{2d \times 2d}$ es el operador antisimétrico canónico $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$.

Teorema 1.12.3.2 (Teorema de Poincaré). *Si $H \in C^2$, entonces el mapeo de flujo φ_t del sistema Hamiltoniano es simpléctico.*

Demostración. Reescribimos el sistema Hamiltoniano: $y' = J^{-1}\nabla H(y)$, donde $y = \begin{bmatrix} p \\ q \end{bmatrix}$.

Denotamos el Jacobiano de $\varphi_t(y)$ por $\frac{\partial \varphi_t}{\partial y}$. Observamos que

$$\frac{\partial}{\partial t}\varphi_t(y) = J^{-1}\nabla H(\varphi_t(y)) \quad \Longrightarrow \quad \frac{\partial}{\partial t}\frac{\partial \varphi_t}{\partial y} = J^{-1}\nabla^2 H(\varphi_t(y))\frac{\partial \varphi_t}{\partial y}$$

Por lo tanto:

$$\frac{\partial}{\partial t}\left(\frac{\partial \varphi_t^T}{\partial y} J \frac{\partial \varphi_t}{\partial y}\right) = \left(\frac{\partial}{\partial}\frac{\partial \varphi_t}{\partial y}\right)^T J \frac{\partial \varphi_t}{\partial y} + \frac{\partial \varphi_t^T}{\partial y} J \left(\frac{\partial}{\partial t}\frac{\partial \varphi_t}{\partial y}\right)$$

$$
\begin{aligned}
&= \frac{\partial \varphi_t}{\partial y}(\nabla^2 H(\varphi_t(y)))^T J^{-T} J \frac{\partial \varphi_t}{\partial y} + \left(\frac{\partial \varphi_t}{\partial y}\right)^T J J^{-1} \nabla^2 H(\varphi_t) \frac{\partial \varphi_t}{\partial y} \\
&= -\frac{\partial \varphi_t}{\partial y}^T \nabla^2 H(\varphi_t) \frac{\partial \varphi_t}{\partial y} + \frac{\partial \varphi_t}{\partial y} \nabla^2 H(\varphi_t) \frac{\partial \varphi_t}{\partial y} \\
&= 0
\end{aligned}
$$

Entonces

$$
\frac{\partial \varphi_t}{\partial y}^T J \frac{\partial \varphi_t}{\partial \varphi} = \frac{\partial \varphi_0}{\partial y}^T J \frac{\partial \varphi_0}{\partial \varphi} = J
$$

■

□

Observación: Se puede demostrar que si φ_t es un mapeo simpléctico entonces este es el flujo de algún sistema Hamiltoniano.

1.12.4. Métodos numéricos simplécticos

Definición 1.12.4.1. *Un método numérico de paso simple*

$$
y^{n+1} = \Phi_h(y^n)
$$

se dice ***simpléctico*** *si la función Φ_h es un mapeo simpléctico.*

Observación: Si discretizamos un sistema Hamiltoniano por un método numérico simpléctico entonces este es una solución exacta de algún sistema Hamiltoniano.

$$
\ddot{y} + \sin y = 0 \quad \Longrightarrow \quad \begin{bmatrix} p^{n+1} \\ q^{n+1} \end{bmatrix} = \Phi_h\left(\begin{bmatrix} p^n \\ q^n \end{bmatrix}\right)
$$

Con $h = 0{,}1$ y $t = 10$

Runge-Kutta simplécticos

Teorema 1.12.4.2. *Dado un método de Runge-Kutta con coeficientes*

$$
A = (a_{ij});\ b = (b_j);\ c = (c_j)
$$

definimos $M = (m_{ij})$ por

$$
m_{ij} = b_i a_{ij} + b_j a_{ji} - b_i b_j,\ i, j = 1, ..., v
$$

Entonces si $M = 0$ el método de RK es ***simpléctico****.*

$$\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array} \qquad \begin{array}{c|ccc} 0 & & & \\ 2/3 & 2/3 & & \\ 2/3 & 0 & 2/3 & \\ \hline & 1/4 & 3/8 & 3/8 \end{array}$$

Demostración. Método de RK a $y' = J\nabla H(y)$

$$\xi_k = f(t_n + c_k h, y_n + h\sum_{l=1}^{s} a_{kl}\xi_l),\ k = 1, ..., s, \quad y_{n+1} = y_n + h\sum_{k=1}^{s} b_k\xi_k$$

Sea $\varphi_n = \frac{\partial y_n}{\partial y_0}$, simplecticidad significa que: $\varphi_{n+1}^T J\varphi_{n+1} = \varphi_n^T J\varphi_n,\ n = 0, 1, ...$

Sea $x_k = \frac{\partial \xi_k}{\partial y_0}$, $G_k = \nabla^2 H(t_n + c_k h, y_n + h\sum_{l=1}^{s} a_{kl}\xi_l)$ no singulares.

Como $\varphi_{n+1} = \varphi_n + h\sum_{k=1}^{s} b_k x_k$, se sigue que

$$\begin{aligned}\varphi_{n+1}^T J\varphi_{n+1} &= (\varphi_n + h\sum_{k=1}^{s} b_k x_k)^T J(\varphi_n + h\sum_{k=1}^{s} b_k x_k) \\ &= \varphi_n^T J\varphi_n + h\sum_{k=1}^{s} b_k x_k^T J\varphi_n + \sum_{l=1}^{s} b_l\varphi_n^T Jx_l + h^2\sum_{k=1}^{s}\sum_{l=1}^{s} b_k b_l x_k^T Jx_l\end{aligned}$$

Observe que $x_K = J^{-1}G_k(\varphi_n + h\sum_{l=1}^{v} a_{kl}x_l)$. Así $\varphi_n = G_k^{-1}Jx_k - h\sum_{l=1}^{s} a_{kl}x_l$

Por lo tanto:

$$\begin{aligned}\sum_{k=1}^{s} b_k x_k^T J\varphi_n &= \sum_{k=1}^{n} b_k x_k^T JG_k^{-1}Jx_k - h\sum_{k=1}^{s}\sum_{l=1}^{s} b_k a_{kl} x_k^T Jx_l \\ \sum_{l=1}^{s} b_l\varphi_n^T Jx_l &= \sum_{l=1}^{s} b_l x_l^T J^T G_l^{-1}Jx_l - h\sum_{k=1}^{s}\sum_{l=1}^{s} b_l a_{lk} x_k^T Jx_l\end{aligned}$$

Como $J^T = -J$ obtenemos que

$$\varphi_{n+1}^T J\varphi_n = \varphi_n^T J\varphi_n - h^2\sum_{k=1}^{s}\sum_{l=1}^{s}(b_k a_{kl} + b_l a_{lk} - b_k b_l)x_k^T Jx_l = \varphi_n^T J\varphi_n$$

□

1.12.5. Sistemas Hamiltonianos separables

Consideramos $H(p,q) = T(p) - V(q)$

$$\dot{p} = -\frac{\partial H}{\partial q} = -\frac{\partial V}{\partial q}$$
$$\dot{q} = -\frac{\partial H}{\partial p} = \frac{\partial T}{\partial p}$$

Es posible discretizar estas ecuaciones usando dos métodos de RK, uno aplicado a la ecuación de $\dot{p}$ y el otro a la de $\dot{q}$.

Ejemplos.

1. Euler simpléctico orden 1:

$$p^{n+1} = p^n - h\frac{\partial H}{\partial q}(t^n, p^{n+1}, q^n)$$
$$q^{n+1} = q^n + h\frac{\partial H}{\partial p}(t^n, p^{n+1}, q^n)$$

$$p^{n+1} = p^n - h\frac{\partial H}{\partial q}(t^n, p^n, q^{n+1})$$
$$q^{n+1} = q^n - h\frac{\partial H}{\partial p}(t^n, p^n, q^{n+1})$$

   Demostrar que estos métodos son simplécticos.

2. Stormer-Verlet orden 2

$$p^{n+1/2} = p^n - \frac{h}{2}\frac{\partial}{\partial q}H(p^{n+1/2}, q^n)$$
$$q^{n+1} = q^n - \frac{h}{2}\left(\frac{\partial}{\partial q}H(p^{n+1/2}, q^n) + \frac{\partial}{\partial q}H(p^{n+1/2}, q^{n+1})\right)$$
$$p^{n+1} = p^{n+1/2} - \frac{h}{2}\frac{\partial}{\partial q}H(p^{n+1}, q^{n+1/2})$$

3. Stormer-Verlet orden 2

$$q^{n+1/2} = q^n + \frac{h}{2}\frac{\partial}{\partial p}H(p^n, q^{n+1/2})$$
$$p^{n+1} = p^n - \frac{h}{2}\left(\frac{\partial}{\partial q}H(p^n, q^{n+1/2}) + \frac{\partial}{\partial q}H(p^{n+1}, q^{n+1/2})\right)$$
$$q^{n+1} = q^n + \frac{h}{2}\frac{\partial}{\partial p}H(p^{n+1}, q^{n+1/2})$$

Ejercicios

Demuestre que los siguientes métodos son simplécticos:

$$\begin{array}{c|cc} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ \hline & 1/2 & 1/2 \end{array}$$

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

Ver articulo reciente: Cockburn, B., Du, S., & Sánchez, M. A. (2023). **Combining finite element space-discretizations with symplectic time-marching schemes for linear Hamiltonian systems.** Frontiers in Applied Mathematics and Statistics, 9, 1165371. Link.

1.13. Método de elementos finitos para ecuaciones diferenciales ordinarias

Estas notas están basadas en el influyente artículo:

Hulme, B. L. (1972). **One-step piecewise polynomial Galerkin methods for initial value problems**. Mathematics of Computation, 26(118), 415-426. Link.

1.13.1. Problema modelo

El problema a resolver planteado es el siguiente

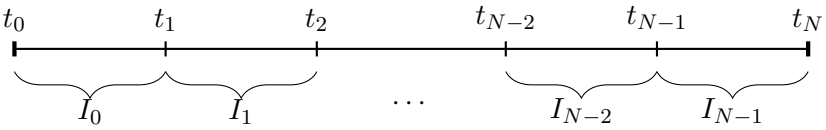
$$\begin{cases} u'(t) = f(t, u(t)), & t_0 \leq t \leq t_N \\ u(t_0) = u_0 \end{cases}$$

donde asumimos que $f(t, x) \in C^{2n}([t_0, t_N] \times \mathbb{R})$ con constante Lipschitz L y $u \in C^{2n+1}([t_0, t_N])$, para $n \geq 1$.

1.13.2. Método de Galerkin e implementación

Consideraremos una **triangulación**[1] uniforme dada por los nodos $t_i = t_0 + ih$, $0 \leq i \leq N$ y los elementos dados por los intervalos $I_i = [t_i, t_{i+1}]$, $0 \leq i \leq N - 1$.

[1] Este concepto lo extenderemos en a dimensiones mas altas mas adelante.



Definimos entonces nuestro **subespacio de dimensión finita** V como

$$V := \{v \in C([t_0, t_N]) : v|_{I_i} \in \mathbb{P}_k(I_i), 0 \leq i \leq N-1\}$$

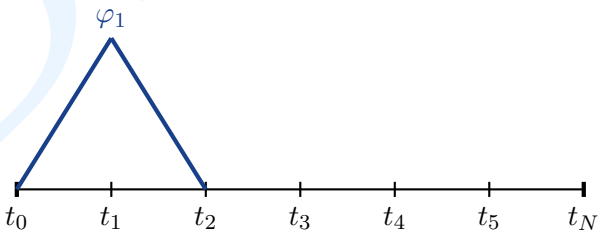
Buscamos aproximaciones $u_h(t) \in V$, por lo que expandiendo en términos de la base,

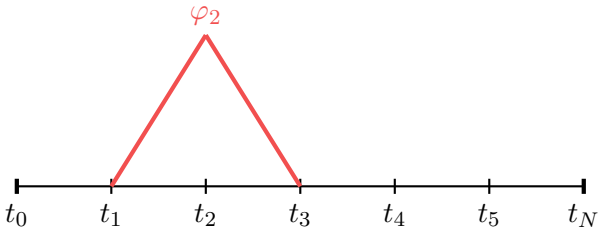
$$u_h(t) = \sum_{j=0}^{n} b_j^{(i)} \varphi_j^{(i)}(t), \quad t_i \leq t \leq t_{i+1},\ 0 \leq i \leq N-1 \tag{1.4}$$

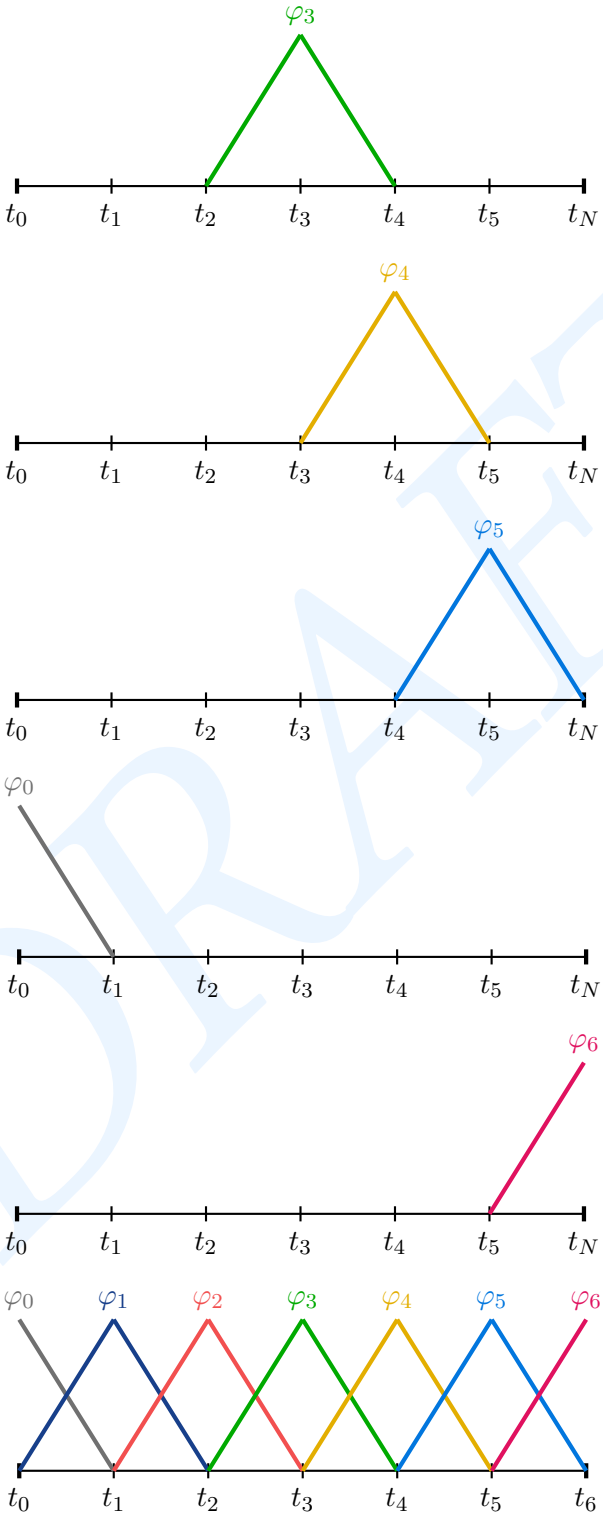
Las funciones base, caso lineal

$$V := \{v \in C([t_0, t_N]) : v|_{I_i} \in \mathbb{P}_1(I_i), 0 \leq i \leq N-1\}, \qquad \dim(V) = ?$$

$$\varphi_i(t) = \begin{cases} \frac{t-t_{i-1}}{t_i-t_{i-1}}, & \text{si } t \in I_{i-1} \\ \frac{t_{i+1}-t}{t_{i+1}-t_i}, & \text{si } t \in I_i, \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{t-t_{i-1}}{h}, & \text{si } t \in I_{i-1} \\ \frac{t_{i+1}-t}{h}, & \text{si } t \in I_i, \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \varphi_1^{i-1}, & \text{si } t \in I_{i-1} \\ \varphi_0^{i}, & \text{si } t \in I_i, \\ 0 & \text{en otro caso} \end{cases}$$






φ_3
t_0 t_1 t_2 t_3 t_4 t_5 t_N
φ_4
t_0 t_1 t_2 t_3 t_4 t_5 t_N
φ_5
t_0 t_1 t_2 t_3 t_4 t_5 t_N
φ_0
t_0 t_1 t_2 t_3 t_4 t_5 t_N
φ_6
t_0 t_1 t_2 t_3 t_4 t_5 t_N
φ_0 φ_1 φ_2 φ_3 φ_4 φ_5 φ_6
t_0 t_1 t_2 t_3 t_4 t_5 t_6

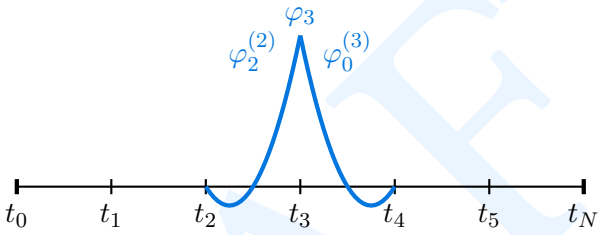
Las funciones base

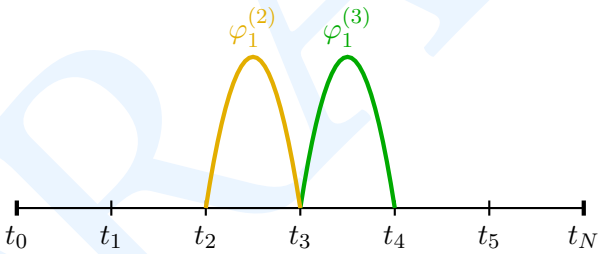
Proposición 1.13.2.1. *El conjunto de funciones gorro o "hat"* $\{\varphi_i\}_{i=0}^{N-1}$ *es una base del subespacio* V.

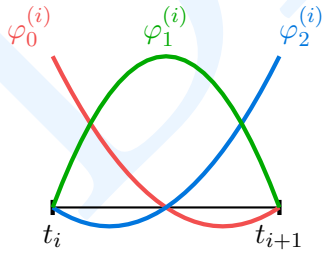
Funciones base de orden 2

Base de Lagrange

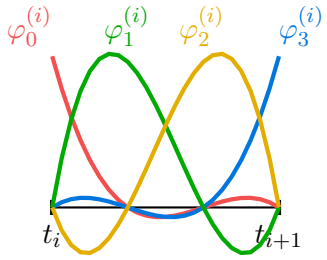
Funciones base de nodos, donde impone continuidad:







Order $n = 2$.



Orden $n = 3$.

Las funciones base

Las funciones $\varphi_j^{(i)}(t)$ son polinomios de grado n en cada intervalo I_i.

Para definir la **formulación débil** del problema, imponemos la continuidad de la solución en

cada nodo y utilizando funciones test $\varphi_k^{(i)}$, con $0 \leq k \leq n$ y $0 \leq i \leq N-1$,

$$\begin{cases} u_h|_{t_i^+} = u_h|_{t_i^-}, & 1 \leq i \leq N \\ u_h|_{t_0} = u_0, & \\ \displaystyle\int_{I_i} \frac{d}{dt}\big(u_h(t)\big)\varphi_k^{(i)}(t)dt = \int_{I_i} f(t, u_h(t))\varphi_k^{(i)}(t)dt, & 1 \leq k \leq n. \end{cases} \tag{1.5}$$

1.13.3. Formulación e implementación

Ahora, imponemos (1.4) reemplanzando en (1.5).

$$\sum_{j=0}^{n} b_j^{(i)} \int_{I_i} \frac{d}{dt}\big(\varphi_j^{(i)}(t)\big)\varphi_k^{(i)}(t)dt = \int_{I_i} f\left(t, \sum_{j=0}^{n} b_j^{(i)}\varphi_j^{(i)}(t)\right)\varphi_k^{(i)}(t)dt \tag{1.6}$$

Implementación.

Para la implementación, asumimos que el lado izquierdo se calcula en forma exacta, mientras que el lado derecho se calcula utilizando una cuadratura de Gauss-Legendre de n-puntos dada por la fórmula

$$\int_{I_i} v(t)dt = h\sum_{k=1}^{n} w_k v(x_{i,k}) + \mathcal{O}(h^{2n+1}), \quad \text{dado} \quad x_{i,k} = t_i + \theta_k h, \quad 1 \leq k \leq n$$

donde (w_k, θ_k) son pesos en los nodos en $[0,1]$.

De esta manera, para cada intervalo I_i, $0 \leq i \leq N-1$ tenemos un sistema de $n+1$ ecuaciones no-lineales

$$Ab^{(i)} = c^{(i)}(b^{(i)}), \quad 0 \leq i \leq N-1$$

donde $b^{(i)} = [b_1^{(i)}, b_2^{(i)}, \ldots, b_{n+1}^{(i)}]^T$, $[A]_{k,j} = A_{k,j}$ y $[c^{(i)}]_k = c_k^{(i)}$.

Las entradas de la matriz A se definen por

$$A_{k,j} = \begin{cases} \varphi_j^{(i)}(t_i), & k = 0, 0 \leq j \leq n \\ \displaystyle\int_{I_i} \frac{d}{dt}\varphi_j^{(i)}(t)\varphi_k^{i}(t)dt, & 1 \leq k \leq n,\ 1 \leq j \leq n+1 \end{cases}$$

y el vector de lado derecho se define por

$$c_k^{(i)}(b^{(i)}) = \begin{cases} u_h|_{t_i^-} = \displaystyle\sum_{j=0}^{n} b_j^{(i-1)}\varphi_j^{(i-1)}, & k = 0, \\ \displaystyle h\sum_{m=1}^{n} w_m f\left(x_{i,m}, \sum_{j=0}^{n} b_j^{(i)}\varphi_j^{(i)}(x_{i,m})\right)\varphi_k^{(i)}(x_{i,m}), & 1 \leq k \leq n \end{cases}$$

Asumimos que A es no-singular. (Muestre que esto se satisface si $\left\{\varphi_k^{(i)}\right\}_{k=1}^{n}$ genera $\mathbb{P}_{n-1}$)

1.13.4. Existencia y unicidad

Para demostrar existencia y unicidad, usaremos el teorema de punto fijo de Banach. Sea $b^\star$ la solución del sistema.

$$\begin{aligned}\|b - b^\star\|_\infty &= \|A^{-1}c^{(i)}(b) - A^{-1}c^{(i)}(b^\star)\|_\infty \\ &\leq \|A^{-1}\|_\infty \|c^{(i)}(b) - c^{(i)}(b^\star)\|_\infty \\ &\leq \|A^{-1}\|_\infty h Q_1 L \|b - b^\star\|_\infty\end{aligned}$$

donde Q_1 es una constante que no depende de la solución dada por

$$Q_1 := \max_{1\leq k\leq n} \sum_{m=1}^{n} w_m |\varphi_k^{(i)}(x_{i,m})| \sum_{j=0}^{n} |\varphi_j^{(i)}(x_{i,m})|$$

Fijando $h < 1/Q_i L\|A^{-1}\|_\infty$ encontramos la contracción.

Teorema 1.13.4.1 (Teorema del punto fijo de Banach)**.** *Sea (X, d) un espacio métrico completo y $T : X \to X$ una aplicación contractiva, es decir, existe una constante $0 \leq c < 1$ tal que para todos $x, y \in X$,*

$$d(T(x), T(y)) \leq c \cdot d(x, y).$$

Entonces, existe un único punto $x^* \in X$ tal que $T(x^*) = x^*$. Además, para cualquier $x_0 \in X$, la sucesión definida por $x_{n+1} = T(x_n)$ converge a x^*.

1.13.5. Galerkin como método de colocación

Observamos que la solución aproximada $u_h(t)$ satisface la EDO en los nodos de cuadratura de cada intervalo, es decir $u_i = u_h|_{t_i}$. Tenemos que

$$b_j^{(i)} = (A^{-1})_{j,0} u_i + \sum_{m=1}^{n} \gamma_{j,m} f(x_{i,m}, u_h(x_{i,m})), \quad 0 \leq j \leq n \tag{1.7}$$

donde

$$\gamma_{j,m} = h w_m \sum_{k=1}^{n} A_{j,k}^{-1} \varphi_k^{(i)}(x_{i,m})$$

Por lo tanto, reemplazando (1.7) en (1.4), derivando y evaluando para algún nodo $x_{i,k}$ se tiene que

$$u_h'(x_{i,k}) = \alpha_k u_i + \sum_{m=1}^{n} \beta_{m,k} f(x_{i,m}, u_h(x_{i,m})), \quad 1 \leq k \leq n \tag{1.8}$$

donde los coeficientes están dados por

$$\alpha_k = \sum_{j=0}^{n} (A^{-1})_{j,0} \frac{d}{dt}\varphi_j^{(i)}(x_{i,k}) \quad \text{y} \quad \beta_{m,k} = \sum_{j=0}^{n} \gamma_{j,m} \frac{d}{dt}\varphi_j^{(i)}(x_{i,k}),$$

lo cual corresponde a la forma general de un método de colocación.

Proposición 1.13.5.1. *Muestre que:*

- *Si f es independiente de u y $f \in \mathbb{P}_{n-1}$, entonces $u \in \mathbb{P}^n$ y $u_h \equiv u$.*
- *Sea $q(t) \in \mathcal{P}^n$ con $q(t_i) = 1$, $q'(x_{i,k}) = 0$ dado $1 \leq k \leq n$. Si $f = q'$ se tiene entonces que $u = q = u_h$ en I_i y que $\alpha_k = 0$.*
- *Sea $q_r(t) \in \mathcal{P}^n$ con $q_r(t_i) = 0$, $q_r'(x_{i,k}) = \delta_{r,k}$, dado $1 \leq r \leq n$. Si $f = q_r'$ y $u(t_i) = 0$ entonces $u = q_r = u_h$ y los coeficientes $\beta_{r,k} = \delta_{r,k}$.*

1.13.6. Método de Galerkin como IRK

Ejercicio.

También es posible reescribir el método de Galerkin como un IRK, donde debemos buscar pesos w_m y funciones $g_m(t_i, u_i; h) = f(x_{i,m}, u_h(x_{i,m}))$ tales que

$$u_{i+1} = u_i + h\Phi(t_i, u_i; h) \quad \text{donde} \quad \Phi(t_i, u_i; h) = \sum_{m=1}^{n} w_m g_m(t_i, u_i; h)$$

1.13.7. Análisis de error y estabilidad

El método propuesto interpola a f en $x_{i,k}$ mediante la aproximación $u_h'(t)$. Esto se puede representar como

$$u_h'(t) = \sum_{k=1}^{n} \ell_k f(x_{i,k}, u_h(x_{i,k})) \quad \text{donde} \quad \ell_k(t) = \prod_{j=1, j\neq k}^{n} \frac{t - x_{i,j}}{x_{i,k} - x_{i,j}}, \quad 1 \leq k \leq n \tag{1.9}$$

Integrando sobre algún intervalo $t \in I_i$,

$$u_h(t) = u_i + \sum_{k=1}^{n} f(x_{i,k}, u_h(x_{i,k})) \int_{t_i}^{t} \ell_k(s) ds \tag{1.10}$$

y de la forma de Runge-Kutta se puede escribir en extenso

$$f(x_{i,m}, u_h(x_{i,m})) = f\left(t_i + \theta_m h, u_i + \sum_{k=1}^{n} g_k(t_i, y_i; h) \int_{t_i}^{t_i + \theta_m h} \ell_k(s) ds\right) \tag{1.11}$$

Usando lo anterior, es posible demostrar lo siguiente

Teorema 1.13.7.1 (Cotas de error). *Asuma que $f \in C^{2n}([t_0, t_N] \times \mathbb{R})$ así $u \in C^{2n+1}([t_0, t_N])$ y denote por L la constante Lipschitz para f. Considere el método de Galerkin con funciones continuas y polinomiales a trozos de grado $n \geq 1$ y con cuadratura de Gauss-Legendre de n-puntos. Entonces existen h_0, M que para $0 < h < h_0$*

$$|u(t_i) - u_h(t_i)| \leq Mh^{2n}, \quad 0 \leq i \leq N$$

Además existen constante E_j, $0 \leq j \leq n$ tales que

$$\max_{t_0 \leq y \leq t_N} |u(t) - u_h(t)| \leq E_0 h^{n+1}$$

y

$$\max_{t_i \leq y \leq t_{i+1}} |u^{(j)}(t) - u_h^{(j)}(t)| \leq E_j h^{n-j+1}, \quad | \leq j \leq n, \ 0 \leq i \leq N-1$$

Para el análisis de la **estabilidad** del método, podemos reescribir el problema rígido $u' = \lambda u$ como $u_{i+1} = P_{nn}(\lambda h) u_i$ donde $P_{nn}(\lambda h)$ es la n-ésima diagonal de la aproximación de Padé de $\exp(\lambda h)$.

1.13.8. Problema propuestos:

Resuelva los siguientes problemas de valores iniciales usando elementos finitos.

1. $$\begin{cases} u' = -2tu^2, & 0 \leq t \leq 1 \\ u(0) = 1 \end{cases} ; \qquad u(t) = \frac{1}{1+t^2}$$

2. $$\begin{cases} u' = -u, & 0 \leq t \leq 100 \\ u(0) = 1 \end{cases} ; \qquad u(t) = e^{-t}$$

3. $$\begin{cases} u_1' = u_1^2 u_2, & u_1(0) = 1 \\ u_2' = -1/u_1, & u_2(0) = 1 \end{cases} ; \qquad \begin{matrix} u_1(t) = e^t \\ u_2(t) = e^{-t} \end{matrix} \quad 0 \leq t \leq 1$$

1.14. Método de Galerkin Discontinuo para Ecuaciones Diferenciales Ordinarias

Las notas de esta clase están basadas en el artículo, origen del los métods DG:

Lesaint, P., & Raviart, P. A. (1974). **On a finite element method for solving the neutron transport equation**. Publications des séminaires de mathématiques et informatique de Rennes, (S4), 1-40. Link.

Considere el problema de valores iniciales dado por:

$$\begin{aligned} u'(t) &= f(t, u(t)), \quad t \geq 0, \\ u(t_0) &= u_0. \end{aligned}$$

Sea $t_i = t_0 + ih$ con $0 \leq i \leq N$ e $I_i = [t_i, t_{i+1}]$ con $0 \leq i \leq N-1$. Aproximaremos u en cada subintervalo por $u_h \in \mathbb{P}_n$, que satisface:

$$\text{(DG)} \quad \begin{cases} \left(u_h(t_i^+) - u_h(t_i^-)\right) v(t_i) + \displaystyle\int_{I_i} [u_h'(t) - f(t, u_h(t))] v(t)\, dt = 0, \quad \forall v \in \mathbb{P}_n \\ u_h(t_0^-) = u_0. \end{cases}$$

Es importante enfatizar de que la función u_h es generalmente **discontinua** en los puntos de la malla t_i. En lo que sigue, (DG) corresponderá al método de Galerkin Discontinuo.

Implementación: Para implementar el método de Galerkin discontinuo, consideremos la fórmula de cuadratura:

$$\int_{I_i} \varphi(t)\, dt = h \sum_{j=1}^{n+1} w_j \varphi(x_{i,j}) + O(h^{p+1}),$$

donde $x_{i,j} = x_i + \theta_j h$ con $1 \leq j \leq n+1$ y $\theta_1 = 0$. Además, $(w_j, x_{i,j})$ corresponden a los pesos y nodos de cuadratura en $[0, 1]$.

De esta manera, para todo $v \in \mathbb{P}_n$ se sigue que:

$$[u_h(t_i^+) - u_h(t_i^-)] v(t_i) + h \sum_{j=1}^{n+1} b_j [u_h'(x_{i,j}) - f(x_{i,j}, u_h(x_{i,j}))] v(x_{i,j}) = 0.$$

1.14.1. DG como IRK

A continuación, visualizaremos el método de Galerkin Discontinuo para EDO's como un método de Runge - Kutta Implícito (IRK). Definamos:

$$\begin{cases} u_i = u_h(t_i^-), \\ u_{i,1} = u_h(t_i^+) = u_h(x_{i,1}), \\ u_{i,j} = u_h(x_{i,j}, \quad 2 \leq j \leq n+1. \end{cases}$$

De la Interpolación de Lagrange, sabemos:

$$\ell_j(t) = \prod_{j=2, j \neq i}^{n+1} \frac{t - \theta_j}{\theta_i - \theta_j}, \quad 2 \leq j \leq n+1.$$

El siguiente resultado será clave para vincular el método de Galerkin Discontinuo como método de Runge Kutta Implícito.

Lema 1.14.1.1. *El Método de Galerkin Discontinuo (DG) es equivalente al siguiente método de Runge Kutta Implícito (IRK):*

$$\textit{(IRK)} \quad \begin{cases} u_{i,j} = u_i + h \displaystyle\sum_{k=1}^{n+1} a_{jk} f(x_{ik}, u_{ik}), \quad 1 \leq j \leq n+1, \\ u_{i+1} = u_i + h \displaystyle\sum_{k=1}^{n+1} b_k f(x_{ik}, u_{ik}), \quad (b_k = w_k). \end{cases}$$

en donde

$$a_{j1} = b_1, \quad 1 \leq j \leq n+1, \quad a_{jk} = \int_0^{\theta_j} [\ell_k(x)\,dx - b_1 \ell_k(\theta_1)],$$

con $1 \leq j \leq n+1$ *y* $2 \leq k \leq n+1$.

Demostración. Sea $\{v_j\}_{1 \leq j \leq n+1}$ una base de $\mathbb{P}_n$ definida por:

$$v_j(x_{ik}) = \delta_{jk},$$

con $1 \leq j, k \leq n+1$. Reescribimos (DG) como sigue:

$$\begin{cases} u_h(t_i^+) - u_h(t_i^-) + hb_1(u_h'(x_{i1}) - f(x_{i1}, u_h(x_{i,1}))) = 0 \\ u_h'(x_{ij}) - f(x_{ij}, u_h(x_{ij})) = 0, \quad 2 \leq j \leq n+1. \end{cases}$$

En I_i, $u_h' \in \mathbb{P}_{n-1}$. Entonces u_h' se reescribe como:

$$u_h'(t) = \sum_{k=2}^{n+1} \ell_k \left(\frac{t - t_i}{h} \right) f(x_{ik}, u_h(x_{ik})).$$

Al evaluar en $t = t_i = x_{i,1}$, obtenemos:

$$u_{i,1} - u_i + hb_1 \left(\sum_{k=2}^{n+1} \ell_k(\theta_1) f(x_{ik}, u_{ik}) - f(x_{i1}, u_{i1}) \right) = 0.$$

Por otro lado, para $2 \leq j \leq n+1$:

$$u_h(x_{ij}) = u_h(x_{i1}) + \int_{x_{i1}}^{x_{ij}} u_h'(x)\,dx.$$

Reemplazando, se consigue:

$$u_{ij} = u_i + h(b_1 f(x_{i1}, u_{i1})) + \sum_{k=2}^{n+1} \left(\int_0^{\theta_j} \ell_k(x)\,dx - b_1 \ell_k(\theta_1) f(x_{ik}, u_{ik}) \right).$$

Similarmente, al evaluar en $x = t_{i+1}$ se tiene:

$$u_h(t_{i+1}^-) = u_h(x_{i1}) + \int_{I_1} u'(x)\,dx.$$

Luego:

$$u_{i+1} = u_i + h \left[b_1 f(x_{i1}, u_{i1}) + \sum_{k=2}^{n+1} \left(\int_0^1 \ell_k(x)\,dx - b_1 \ell_k(\theta_1) \right) f(x_{ik}, u_{ik}) \right].$$

Observe que:

$$\int_0^1 \ell_k(x)\,dx = \sum_{j=1}^{n+1} W_J \ell_k(\theta_j) = w_1\ell_k(\theta_1) + w_k = b_1\ell_k(\theta_1 + b_k.$$

Esto nos permite concluir que:

$$u_{i+1} = u_i + h\sum_{k=1}^{n+1} b_k f(x_{ik}, u_{ik}).$$

□

1.14.2. Orden del método DG

Teorema 1.14.2.1. *El Método de Galerkin Discontinuo (DG) es un método de paso simple de orden p.*

Demostración. Basándonos en lo propuesto por Butcher (*Implicit Runge - Kutta process, 1964*) y Crouzeix (*PhD Thesis, 1974*), las condiciones necesarias y suficientes para que un método Runge - Kutta sea de orden p son las siguientes:

$$\text{(IRK)} \quad \begin{cases} \displaystyle\sum_{k=1}^{n+1} b_k\theta_j^l = \frac{1}{l+1}, & 0 \le l \le p-1, \\ \displaystyle\sum_{k=1}^{n+1} a_{jk}\theta_k^l = \frac{\theta_j^{l+1}}{l+1}, & 0 \le l \le n-1,\ 0 \le j \le n+1 \\ \displaystyle\sum_{k=1}^{n+1} b_j a_{jk}\theta_j^l = \frac{b_k(1-\theta_k^{l+1})}{l+1}, & n+l \le p-1,\ 1 \le k \le n+1 \end{cases}$$

□

1.14.3. A - estabilidad del Método de Galerkin Discontinuo

Antes de analizar la A - estabilidad del Método de Galerkin Discontinuo (DG), debemos introducir el siguiente resultado preliminar:

Teorema 1.14.3.1. *El Método de Galerkin Discontinuo (DG) aplicado a la ecuación $u' = \lambda u$ permite obtener $u_{i+1} = R(\lambda h)u_i$, $R(z) = \dfrac{P(z)}{Q(z)}$ con $P \in \mathbb{P}_n$ y $Q \in \mathbb{P}_{n+1}$.*

Demostración. Recordemos que para $1 \leq j \leq n+1$:

$$u_{ij} = u_i + \lambda h \sum_{k=1}^{n+1} a_{jk} u_{ik}, \quad u_{i+1} = u_i + \lambda h \sum_{k=1}^{n+1} b_k u_{ik},$$
$$(I - \lambda h A) u_i = u_{i1},$$

en donde $u_i = \begin{bmatrix} u_{i1} & u_{i2} & \dots & u_{i(n+1)} \end{bmatrix}$. Como $a_{j1} = b_1$ para $1 \leq j \leq n+1$, por la Regla de Cramer se sigue que para $1 \leq j \leq n+1$:

$$u_{ij} = \frac{P_j(\lambda h)}{Q(\lambda h)} u_i,$$

en donde $P_1 \in \mathbb{P}_n$ con coeficiente $b_1^{-1} \det(A)$, $P_j \in \mathbb{P}_{n-1}$ con $2 \leq j \leq n+1$ y $Q \in \mathbb{P}_{n+1}$. Por lo tanto:

$$u_{i+1} = \frac{P(\lambda h)}{Q(\lambda h)} u_i, \quad P(z) = Q(z) - z \sum_{k=1}^{n+1} P_k(z).$$

Note que el coeficiente de z^{n+1} en $P(z)$ es nulo. □

Para la A - estabilidad del Método de Galerkin Discontinuo (DG), considere el siguiente resultado:

Teorema 1.14.3.2. *El Método de Galerkin Discontinuo (DG) es A - estable de orden* $2n+1$.

Demostración. Considere el método de Galerkin Discontinuo (DG) con la regla de cuadratura de Gauss - Radau:

$$\int_0^1 \varphi(x)\, dx = w_1 \varphi(0) + \sum_{j=2}^{n+1} w_j \varphi(x_j),$$

para todo $\varphi \in \mathbb{P}_{2n+1}$. Por **??**, sabemos que el método es de orden $p = 2n+1$. De esta manera, el operador es una aproximación racional de la función exponencial:

$$u' = \lambda u \Rightarrow u(t) = e^{\lambda t} \Rightarrow u_{i+1} = R(\lambda h) u_i.$$

De esta manera:

$$R(z) = e^z + O(z^{2n+2}).$$

Ya sabemos que $R(z) = \dfrac{P(z)}{Q(z)}$ con $P \in \mathbb{P}_n$ y $Q \in \mathbb{P}_{n+1}$. Entonces $R(z)$ corresponde a una aproximación de Padé subdiagonal.

Según Axelsson (*A class of A - stable method, 1969*), la aproximación de Padé subdiagonal satisface:

$$\text{Fuertemente A - estable} \quad \begin{cases} |R(z)| < 1, & \text{para} \quad \text{Re}(z) < 0, \\ |R(z)| \to 0, & \text{cuando} \quad \text{Re}(z) \to \infty. \end{cases}$$

Esto nos permite concluir que el método de Galerkin Discontinuo (DG) más Gauss - Radau es A - estable y de orden $2n+1$. □

1.15. Métodos Predictor - Corrector

Motivación:

Considere el problema de valores iniciales dado por:

$$\begin{aligned} y'(t) &= f(t, y(t)), \quad t \geq 0, \\ y(t_0) &= y_0. \end{aligned}$$

Sea $t_i = t_0 + ih$ con $0 \leq i \leq N$ e $I_i = [t_i, t_{i+1}]$ con $0 \leq i \leq N-1$.

Considere un **método de k - pasos lineal** para aproximar la solución del PVI

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f_{n+j}, \quad f_{n+j} = f(t_{n+j}, y_{n+j}).$$

Si $\alpha_k, \beta_k \neq 0$, entonces el método es implícito. Entonces, en cada paso tenemos que resolver para y_{n+k}:

$$\alpha_k y_{n+k} - h\beta_k f(t_{n+k}, y_{n+k}) = \sum_{j=0}^{k-1} (h\beta_j f_{n+j} - \alpha_j y_{n+j}).$$

Para resolver esta ecuación no lineal podemos usar una iteración de punto fijo para y_{n+k}. Así obtenemos una iteración

$$\alpha_k y_{n+k}^{(s+1)} = h\beta_k f(t_{n+k}, y_{n+k}^{(s)}) + h \sum_{j=0}^{k-1} (\beta_j f_{n+j} - \alpha_j y_{n+j}), \quad s = 1, 2, \ldots.$$

Cuando podemos asegurar que esta iteración converge?

Si asumimos que $h < \dfrac{|\alpha_k|}{L|\beta_k|}$, entonces la ecuación tiene solución única.

Desventajas: debemos obtener demasiadas evaluaciones.

Cómo podemos reducir el número de evaluaciones?

Podemos reducir el número de evaluaciones escogiendo $y_{n+k}^{(0)}$ de forma precisa. Esto se puede lograr escogiendo esta primera aproximaciín mediante un método explícito al que llamamos **predictor**. Al método implícito, lo llamamos **corrector**.

1.15.1. Métodos predictor-corrector

$$\textbf{Predictor}: \quad \rho^*(z) = \sum_{j=0}^{k} \alpha_j^* z^j, \quad \alpha_k^* = 1, \quad \sigma^*(z) = \sum_{j=0}^{k-1} \beta_j^* z^j.$$

$$\textbf{Corrector}: \quad \rho(z) = \sum_{j=0}^{k} \alpha_j z^j, \quad \alpha_k = 1, \quad \sigma(z) = \sum_{j=0}^{k-1} \beta_j z^j.$$

en donde:

- $m \in \mathbb{N}$ corresponde al número de veces que se puede aplicar el corrector.
- P indica la aplicación del predictor.
- C indica una aplicación del corrector.
- E indica una evaluación de f.

1. $P(EC)^m E$

$$\begin{aligned}
y_{n+k}^{(0)} + \sum_{j=0}^{k-1} \alpha_j^* y_{n+j}^m &= h \sum_{j=0}^{k-1} \beta_j^* f_{n+j}^m \\
f_{n+k}^{(s)} &= f(t_{n+k}, y_{n+k}^{(s)}) \\
y_{n+k}^{(s+1)} + \sum_{j=0}^{k-1} \alpha_j y_{n+j}^m &= h\beta_k f_{n+k}^{(s)} + h \sum_{j=0}^{k-1} \beta_j f_{n+j}^m, \quad s = 0, 1, .., m = 1 \\
f_{n+k}^{(m)} &= f(t_{n+k}, y_{n+k}^{(m)})
\end{aligned}$$

para $n = 0, 1, 2, ..$

1.16. Problemas de valores de frontera

Tenemos una ecuación diferencial ordinaria donde buscamos la solución $y = y(x)$ en un intervalo $I = (a, b)$

$$F(x, y, y, ..., y^{(n-1)}, y^{(n)}) = 0$$

sujeta a condiciones de frontera o de contorno

$$R_j(y) = R_j((y(a), y'(a), ..., y^{(n-1)}(a)), (y(b), y'(b), ..., y^{(n-1)}(b)))$$

para $j = 1, ..., n$.

Para el caso no lineal general es dificil obtener resultados de existencia y unicidad.

Ilustración

Deflexión de una viga uniforme.

Para deflexiones pequeñas, el modelo matemático es: $E\,I\,y^{(4)}(x) = F(x)$, donde :

- E: constante de modulo de Young, depende del material. $E_{\text{acero}} = 200 \times 10^9[N/m^2]$, $E_{\text{concreto}} = 20 \times 10^9[N/m^2]$.
- I: momento de inercia de la sección transversal de la viga. Por ejemplo, $I = \pi a^4/4$ para una viga cilíndrica con radio a. $F(x)$: denota la fuerza actuando verticalmente sobre la viga y hacia abajo en el punto x. Si $F(x) = w$, peso de la viga , entonces resolvemos y obtenemos

  $$y(x) = \frac{w}{24E\,I}x^4 + Ax^3 + Bx^2 + Cx + D$$

  Las constantes A, B, C, D se determinan por como a viga esta sujeta en los entremos.

Soporte	Condiciones de frontera
Apoyada	$y = y'' = 0$
Empotrada	$y = y' = 0$
Libre	$y' = y''' = 0$

1.16.1. Problema de valores de frontera lineal

Analizamos el caso del operador diferencial lineal

$$(Ly)(x) := \sum_{i=0}^{n} f_i(x)y^{(i)}(x) = g(x), \quad x \in I,\ f_n \neq 0.$$

y con condiciones de frontera

$$R_j(y) = \sum_{k=0}^{n-1} \Big(\alpha_{j,k+1}y^{(k)}(a) + \beta_{j,k+1}y^{(k)}(b)\Big) = \gamma_j, \quad j = 1, 2..., n.$$

En la ecuación diferencial asumimos que $f_i \in C([a, b])$, para $i = 0, 1, ..., n$. Además, decimos que si:

- $g = 0$, la ecuación se dice homogénea
- $\gamma_j = 0$, para $j = 1, 2, .., n$, las condiciones de frontera de dicen homogéneas
- $g = 0, \gamma_j = 0$, para $j = 1, 2, ..., n$, entonces el problema se dice homogéneo.

Problema de valores de frontera de segundo orden lineal

Ilustramos las definiciones anteriores con el problema de frontera de segundo orden y lineal.

$$
\begin{aligned}
(Ly)(x) &= f_2(x)y''(x) + f_1(x)y'(x) + f_0(x)y(x) = g(x) \\
R_1(y) &= \alpha_{11}y(a) + \beta_{11}y(b) + \alpha_{12}y'(a) + \beta_{12}y'(b) = \gamma_1, \\
R_2(y) &= \alpha_{21}y(a) + \beta_{21}y(b) + \alpha_{22}y'(a) + \beta_{22}y'(b) = \gamma_2.
\end{aligned}
$$

$$
A \begin{bmatrix} y(a) \\ y'(a) \\ y(b) \\ y'(b) \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \beta_{11} & \beta_{12} \\ \alpha_{21} & \alpha_{22} & \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y(a) \\ y'(a) \\ y(b) \\ y'(b) \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}
$$

Observación: condiciones homogéneas

Problemas de valores de frontera lineales pueden ser reducidos a un problema con condiciones de frontera **homogéneas**, esto es, $\gamma_j = 0$, $j = 1, ..., n$.

Ejercicio. Muestre un procedimiento para reducir el problema a uno de condiciones de frontera homogéneas.

Ejemplos

1.
$$
\begin{aligned}
y'' - y &= 0 \\
y(0) &= 0 \\
y(b) &= \beta
\end{aligned}
$$

Solución:

$$y(x) = \frac{\beta \sinh(x)}{\sinh(b)},\ 0 \le x \le b$$

Aquí, $\sinh(x)$ es la solución que satisface las condiciones impuestas. *No hay intervalos críticos en este caso.*

2.
$$
\begin{aligned}
y'' + y &= 0 \\
y(0) &= 0 \\
y(b) &= \beta
\end{aligned}
$$

Solución:

$$y(x) = \frac{\beta \sin(x)}{\sin(b)},\ 0 \le x \le b,\ b \neq n\pi,\ n \in \mathbb{N}$$

- Si $b = n\pi$, hay puntos críticos donde la solución puede no ser única.
- Si $\beta = 0$ y $b = n\pi$, hay **infinitas soluciones** de la forma $y(x) = c\sin(x)$.

Este es un ejemplo de cómo la estructura de la solución cambia debido a la naturaleza oscilatoria.

1.16.2. Un problema de valores de frontera lineal de dos puntos

Considere el problema de valores de frontera de dos puntos de segundo orden escalar

$$y'' = f(x, y, y'), \quad a \leq x \leq b$$

con condiciones de frontera lineal

$$\begin{aligned} a_0 y(a) - a_1 y'(a) &= \alpha \\ b_0 y(b) + b_1 y'(b) &= \beta, \end{aligned}$$

dond asumimos que al menos a_0 o a_1 son no cero y lo mismo para b_0 y b_1.

Además asumimos que f es una función continua en $[a, b] \times \mathbb{R} \times \mathbb{R}$ y unifmormente Lipschitz continua para el segundo y tercer argumento, esto es

$$\begin{aligned} |f(x, \tilde{u}_1, u_2) - f(x, u_1, u_2)| &\leq L_1 |\tilde{u}_1 - u_1|, \\ |f(x, u_1, \tilde{u}_2) - f(x, u_1, u_2)| &\leq L_2 |\tilde{u}_2 - u_2| \end{aligned}$$

para todo $x \in [a, b]$ y $u_1, \tilde{u}_1, u_2, \tilde{u_2} \in \mathbb{R}$.

Problema de valores iniciales asociados

Asociamos al problema de valores de frontera anterior el siguiente IVP

$$u'' = f(x, u, u'), \quad a \leq x \leq b, \qquad \text{sujeto a: } \begin{cases} a_0 u(a) - a_1 u'(a) &= \alpha \\ c_0 u(a) - c_1 u'(a) &= s \end{cases}$$

donde debemos asumir que $a_1 c_0 - a_0 c_1 \neq 0$. Escogemos c_0, c_1 tales que: $a_1 c_0 - a_0 c_1 = 1$.

Así las condiciones iniciales queda: $\begin{cases} u(a) &= a_1 s - c_1 \alpha \\ u'(a) &= a_0 s - c_0 \alpha \end{cases}$

Denotando a la solución del IVP por $u(x; s)$, notamos que esta resuelve el BVP si

$$\phi(s) := b_0 u(b; s) + b_1 u'(b; s) - \beta = 0$$

Teorema 1.16.2.1. *El problema de valores de frontera de dos puntos tiene tantas soluciones distintas como la función $\phi(s)$ tiene distintos ceros.*

Demostración. Ejercicio.

Teorema 1.16.2.2. *Asuma que*

1. *$f(x, u_1, u_2)$ es una función contínua en $[a, b] \times \mathbb{R} \times \mathbb{R}$.*
2. *Asuma que las derivadas parciales f_{u_1} y f_{u_2} son contínuas y satisfacen*

$$0 < f_{u_1}(x, u_1, u_2) \leq L_1, \quad |f_{u_2}(x, u_1, u_2)| \leq L_2, \quad \text{en } [a, b] \times \mathbb{R} \times \mathbb{R}.$$

3. $a_0 a_1 \geq 0$, $b_0 b_1 \geq 0$, $|a_0| + |b_0| > 0$.

Entonces, el problema de valores de frontera tiene una única solución.

La tercera condición la reescribimos como

$$a_0 \geq 0, a_1 \geq 0;\ b_0 \geq 0, b_1 \geq 0;\ a_0 + b_0 > 0.$$

y además a_0 y a_1 no pueden ser cero al mismo tiempo. Lo mismo para b_0 y b_1.

Demostración. Demostraremos que

$$\phi'(s) \geq x > 0, \quad \text{para todo } s \in \mathbb{R}. \qquad (\star)$$

Así la función $\phi(s)$ es monótona creciente de $-\infty$ a ∞ y así es cero en exactamente un valor de s.

Tenemos que

$$\phi'(s) = b_0 \frac{\partial}{\partial s} u(b; s) + b_1 \frac{\partial}{\partial s} u'(b; s).$$

Denotando $v(x) = v(x; s) = \frac{\partial}{\partial s} u(x; s)$, podemos escibir

$$\phi'(s) = b_0 v(b) + b_1 v'(b)$$

Además, sabemos que $u(x; s)$ satisface el problema valores iniciales:

$$\begin{aligned} u''(x; s) &= f(x, u(x; s), u'(x; s)), \quad a \leq x \leq b, \\ u(a; s) &= a_1 s - c_1 \alpha, \\ u'(a; s) &= a_0 s - c_0 \alpha. \end{aligned}$$

De donde, derivando con respecto a s, se sigue que:

$$\begin{aligned} v''(x) &= f_{u_1}(x.u(x; s), u'(x; s))v(x) + f_{u_2}(x, u(x; s), u'(x; s))v'(x), \quad a \leq x \leq b, \\ v(a) &= a_1, \\ v'(a) &= a_0. \end{aligned}$$

Este problema lo reescribimos de forma mas simple

$$\begin{aligned} v''(x) &= q(x)v(x) + p(x)v'(x), \quad a \leq x \leq b, \\ v(a) &= a_1, \\ v'(a) &= a_0. \end{aligned}$$

donde $|p(x)| \leq L_2$, y $0 < q(x) \leq L_1$ sobre $[a, b]$.

Vamos a demostrar que la solución v satisface

$$v(x) > a_1 + a_0 \frac{1 - \exp(-L_2(x - a))}{L_2}, \quad v'(x) > a_0 \exp(-L_2(x - a)), \quad a \leq x \leq b.$$

Esto implica el resultado (⋆). En efecto, como a_0 y a_1 no son cero al mismo tiempo, al menos uno es positivo, se sique que $v(b) > 0$. Si $b_0 > 0$, entonces, como además $b_1 \geq 0$ y $v'(b) > 0$, $\phi'(s) \geq c > 0$. si $b_0 = 0$, entonces $b_1 > 0$ y $\phi'(s) = b_1 v'(b) > 0$.

Resta por demostrar las cotas de v y v'.

Primero, observemos que $v(x) > 0$ en alguna vecindad de a, dado que $(v(a) > 0)$ o $(v(a) = 0$ y $v'(a) > 0)$.

Si $v(x)$ no fuese positivo para todo $x \in [a, b]$, entonces tendrímos $x_0 \in [a, b] : v(x_0) = 0$. Así v tendría un máximo local en algún $a < x_1 < x_0$. Entonces:

$$v(x_1) > 0, \quad v'(x_1) = 0, \quad v''(x_1) < 0.$$

Pero esto contradice la ecuación diferencial, ya que

$$v''(x_1) = q(x_1)v(x_1) + p(x_1)v'(x_1) > 0.$$

Por lo tanto, $v(x) > 0$ para $a \leq x \leq b$.

Usando nuevamente que q es positivo, tenemos que

$$v''(x) - p(x)v'(x) > 0 \implies \frac{d}{dx}\left(\exp(-\int_a^x p(t)dt)\right) > 0$$

de donde

$$v'(x) > a_0 \exp(\int_a^x p(t)dt) > a_0 \exp(-L_2(x-a))$$

Por último, integrando la desigualdad, obtenemos

$$v(x) - v(a) = v(x) - a_1 > a_0 \frac{1 - \exp(-L_2(x-a))L_2}{.}$$

□

Problema de Sturm-Liouville

Corolario 1.16.2.3. *Sea la ecuación diferencial linear de orden 2*

$$Ly := -y'' + p(x)y' + q(x)y = r(x), \quad a \leq x \leq b,$$

con condiciones de frontera

$$a_0 y(a) - a_1 y'(a) = \alpha, \quad b_0 y(b) + b_1 y'(b) = \beta$$

Entonces, si p, q, r con continuas en $[a, b]$ y si $a_,a_,b_0, b_1$ la condicion (3) del teorema anterior, el problema tiene una única solución.

1.17. Método de disparo

La solución del BVP lleva a la solución de una ecuación no lineal para ϕ desde el IVP. Nos referimos a resolver el IVP como el *disparo*, que *apunta* a la segunda condición de frontera, o *blanco*. Un mecanismo que reajusta el disparo basado en cuanto se erro el blanco es derivado por el método de Newton

$$s^{(\nu+1)} = s^{(\nu)} - \frac{\phi(s^{(n)})}{\phi'(s^{(n)})}, \quad \nu = 0, 1, 2, ...$$

Si $s^{\nu} \to s_{\infty}$, entonces, $y(x) = u(x; s_{\infty})$ es solución del BVP.

Cómo calculamos $\phi(s)$ y $\phi'(s)$?

Defina: $y_1(x) = u(x; s), \quad y_2(x) = u'(x; s), \quad y_3(x) = v(x), \quad y_4(x) = v'(x)$

resuelve el IVP

$$\begin{aligned} y_1' &= y_2 & y_1(a) &= a_1 s - c_1 \alpha \\ y_2' &= f(x, y_1, y_2) & y_2(a) &= a_0 s - c_0 \alpha \\ y_3' &= y_4 & y_3(a) &= a_1 \\ y_4' &= f_{u_1}(x, y_1, y_2) y_3 + f_{u_2}(x, y_1, y_2) y_4 & y_4(a) &= a_0 \end{aligned}$$

donde c_0, c_1 se escogen satisfaciendo $c_0 a_1 - c_1 a_0 = 1$, y luego se calcula

$$\phi(s) = b_0 y_1(b) + b_1 y_2(b) - \beta, \qquad \phi'(s) = b_0 y_3(b) + b_1 y_4(b)$$

Cada paso de Newton requiere la solución de IVP con $s = s^{(\nu)}$.

Ejemplo

Escriba el procedimiento por el método del disparo para resolver el problema

$$y'' = -\exp(-y), \quad 0 \le x \le 1, \quad \text{con } y(0) = y(1) = 0.$$

Solución: Primero mostremos que el problema tiene única solución. Para esto consideremos la función

$$f(y) = \begin{cases} -\exp(-y), & \text{si } y \ge 0 \\ \exp(y) - 2, & \text{si } y \le 0 \end{cases} \implies f_y(y) = \begin{cases} \exp(-y) & y \ge 0, \\ \exp(y) & y \le 0. \end{cases}$$

Entonces, $0 < f_y(y) \le 1$ para todo $y \in \mathbb{R}$. Las otras hipótesis del Teorema se satisfacen y por lo tanto existe una única soluci'on del problema

$$y'' = f(y), \quad 0 \le x \le 1, \quad \text{con } y(0) = y(1) = 0.$$

Luego el sistema de primer orden nos queda:

$$\begin{aligned}
\frac{dy_1}{dx} &= y_2, & y_1(0) &= 0,\\
\frac{dy_2}{dx} &= -\exp(-y_1), & y_2(0) &= s,\\
\frac{dy_3}{dx} &= y_4, & y_3(0) &= 0,\\
\frac{dy_4}{dx} &= \exp(-y_1)y_3, & y_4(0) &= 1,
\end{aligned}$$

y $\phi(s) = y_1(1)$, $\phi'(s) = y_3(1)$

1.18. Operador adjunto

Definición 1.18.0.1. *Dado un operador diferencial lineal L, de orden n, de la forma*

$$(Ly)(x) := \sum_{j=0}^{n} f_i(x) y^{(i)}(x)$$

definimos el operador adjunto $L^\star$ por

$$(L^* y)(x) := \sum_{j=0}^{n} (-1)^j \frac{d^j}{dx^j} \left(f_j(x) y(x)\right)$$

El operador L se dice autoadjunto si

$$Ly = L^* y, \quad y \in C^n(a, b)$$

Para el operador lineal de segundo orden, la condición de autoadjunto queda

$$f_2(x)y''(x) + f_1(x)y'(x) + f_0(x)y(x) = \frac{d^2}{dx^2}\left(f_2(x)y(x)\right) - \frac{d}{dx}\left(f_1(x)y(x)\right) + f_0(x)y_0(x)$$

de donde

$$2(f_1(x) - f_2'(x))y'(x) - (f_2''(x) - f_1'(x))y(x) = 0, \quad y \in C^2(a, b)$$

y por lo tanto, el operador L es autoadjunto si y solo si

$$f_1(x) = f_2'(x).$$

Ecuación diferencial linear de segundo orden autoadjunta

$$Ly = \frac{d}{dx}\left(f_2(x)\frac{dy}{dx}\right) + f_0(x)y = g$$

$$Ly = -\frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) + q(x)y = g$$

Teorema 1.18.0.2. *Toda ecuación diferencial lineal de segundo orden*

$$f_2(x)y''(x) + f_1(x)y'(x) + f_0(x)y(x) = g(x)$$

con $f_2(x) \neq 0$ *para todo* $x \in (a, b)$ *puede ser transformada a una ecuación autoadjunta de segundo orden.*

Demostración.

Multiplicar por : $-p(x) = \exp\left(\int_{x_0}^{x} \frac{f_1(t)}{f_2(t)} dt\right)$ para $x_0, x \in (a, b)$

1.19. Métodos de diferencias finitas

Un método numérico clásico para resolver problemas de valores de frontera es el método de diferencias finitas, el cual reemplaza de forma directa los operadores diferenciales, o derivadas, por expresiones de **diferencias finitas** e impone una version discreta del problema en algunos puntos o grilla del dominio.Esta discretización da origen a un sistema lineal o no lineal de ecuaciones con solución la aproximación en los puntos de la grilla.

Vamos a considerar esquemas de una grilla o malla uniforme de un intervalo finito (a, b), esto es:

$$\{x_n\}_{n=0}^{N+1} : \qquad a = x_0 < x_1 < \ldots < x_N < x_{N+1}, \quad x_n = a + nh;\ h = \frac{b-a}{N+1}$$

Definimos **funciones de grilla** $u \in \Gamma_h[a, b]$ como $u : \{x_n\} \mapsto \mathbb{R}^{N+1}$.

1.19.1. Ecuaciones lineales de segundo orden

Consideraremos el problema de Sturm-Liouville

$$L(y) = r(x), \quad a \leq x \leq b, \qquad L(y) := -y'' + p(x)y' + q(x)y$$

con condiciones de frontera simples

$$y(a) = \alpha,\ y(b) = \beta.$$

Asumimos que existen constantes positivas $\bar{p}, \underline{q}$ y $\bar{q}$ tales que

$$|p(x)| \leq \bar{p}, \quad 0 < \underline{q} \leq q(x) \leq \bar{q}, \quad a \leq x \leq b.$$

Operador de diferencias finitas: Definimos para una función de grilla $u \in \Gamma_h[a, b]$

$$(L_h u)_n = -\frac{u_{n+1} - 2u_n + u_{n-1}}{h^2} + p(x_n)\frac{u_{n+1} - u_{n-1}}{2h} + q(x_n)u_n, \quad \text{para } n = 1, 2, ..., N.$$

Definición 1.19.1.1. *Para toda función suficientemente suave v definida sobre $[a,b]$ y operador diferencial L y operador de diferencias finitas L_h asociado a la grilla $\{x_n\}$ definimos el* ***error de truncación*** *T_h por*

$$(T_h v)_n = (L_h v)_n - (Lv)(x_n), \quad n = 1, 2, ..., N$$

Tenemos que, para el operador asociado al problema de Sturm Liouville tenemos, para $v \in C^4([a,b])$

$$(T_h v)_n = -\frac{h^2}{12}\left(v^{(4)}(\xi_1) - 2p(x_n)v^{(3)}(\xi_2)\right), \quad \xi_1, \xi_2 \in [x_n - h, x_n + h],$$

o si $v \in C^6([a,b])$

$$(T_h v)_n = -\frac{h^2}{12}\left(v^{(4)}(x_n) - 2p(x_n)v^{(3)}(x_n)\right) + \mathcal{O}(h^4), \quad h \to 0.$$

Definición 1.19.1.2. *Decimos que un operador de diferencias L_h es* ***estable*** *si existe una constante M, independiente de h, tal que para h suficientemente pequeño, tenemos para toda función de grilla $v = \{v_n\}$*

$$\|v\|_\infty \leq M\left(\text{máx}\{|v_0|, |v_{N+1}|\} + \|L_h(v)\|_\infty\right)$$

Donde

$$\|v\|_\infty = \max_{0 \leq n \leq N+1} |v_n|, \quad \|L_h(v)\|_\infty = \max_{1 \leq n \leq N} |(L_h v)_n|$$

Es el operador de nuestro operador de diferencias estable?

Teorema 1.19.1.3. *Si $h\bar{p} \leq 2$, entonces el operador L_h (aprox. S-L) es estable. En efecto, la constante de estabilidad $M = \text{máx}\{1, 1/\underline{q}\}$*

Ver ejemplo numérico de cuando las condiciones de estabilidad no se satisface

Demostración. Reescribimos el operador de diferencias por

$$\frac{h^2}{2}(L_h v)_n = a_n v_{n-1} + b_n v_n + c_n v_{n+1}, \quad n = 1, 2, ..., N$$

donde

$$a_n = -\frac{1}{2}\left(1 + \frac{1}{2}hp(x_n)\right), \quad b_n = 1 + \frac{1}{2}h^2 q(x_n), \quad c_n = -\frac{1}{2}\left(1 - \frac{1}{2}hp(x_n)\right).$$

Por el supuesto del teorema, tenemos que $\frac{1}{2}h|p(x_n)| \leq \frac{1}{2}h\bar{p} \leq 1$, y así $a_n \leq 0$, $c_n \leq 0$ y

$$|a_n| + |c_n| = \frac{1}{2}\left(1 + \frac{1}{2}hp(x_n)\right) + \frac{1}{2}\left(1 - \frac{1}{2}hp(x_n)\right) = 1$$

Además, tenemos que $b_n \geq 1 + \frac{1}{2}h^2\underline{q}$. Observemos que

$$b_n v_n = -a_n v_{n-1} - c_n v_{n+1} + \frac{1}{2}h^2 (L_h v)_n,$$

Entonces, se sigue que

$$\left(1 + \frac{1}{2}h^2\underline{q}\right)|v_n| \leq \|v\|_\infty + \frac{1}{2}h^2\|L_h v\|_\infty, \quad n = 1, 2, ..., N$$

Observe que, si $\|v\|_\infty = \text{máx}\{|v_0|, |v_{N+1}|\}$ (esto es la norma se alcanza en uno de estos dos valores), entonces $M = 1$.

Por otro lado, si v alcanza el máximo en el interior, esto es, $\|v\|_\infty = |v_{n_0}|$, $1 \leq n_0 \leq N$, se sique

$$\left(1 + \frac{1}{2}h^2\underline{q}\right)|v_{n_0}| \leq |v_{n_0}| + \frac{1}{2}h^2\|L_h v\|_\infty, \quad n = 1, 2, ..., N$$

lo que implica que

$$\|v\|_\infty = |v_{n_0}| \leq \frac{1}{\underline{q}}\|L_h v\|_\infty.$$

□

Una aproximación por diferencias finitas del problema de Sturm Liouville es la función de grilla $u = \{u_n\}$ solución del problema:

$$(L_h u)_n = r(x_n), \quad n = 1, ..., N, \quad u_0 = \alpha,\ u_{N+1} = \beta$$

Sistema lineal asociado: (demuestre que la matriz es estrictamente diagonal dominante)

$$\begin{bmatrix} b_1 & c_1 & 0 & \dots & 0 \\ a_2 & b_2 & c_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & \dots & 0 & a_N & b_N \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N-1} \\ u_N \end{bmatrix} = \frac{1}{2}h^2 \begin{bmatrix} r(x_1) \\ r(x_2) \\ \vdots \\ r(x_{N-1}) \\ r(x_N) \end{bmatrix} - \begin{bmatrix} a_1\alpha \\ 0 \\ \vdots \\ 0 \\ c_N\beta \end{bmatrix}$$

Teorema 1.19.1.4. *Asuma que L_h es estable. Entonces, el problema de diferencias finitas tiene una única solución o equivalentemente la matriz de diferencias finitas es no singular.*

Demostración. Observe que el problema homogéneo asociado, esto es, $r(x) = 0$ y $\alpha = \beta = 0$ pueden sólo tener la solución trivial dado que $L_h u = 0$ y $u_0 = u_{N+1} = 0$, lo que implica, por la condición de estabilidad que $u_n = 0, n = 0, 1, ..., N + 1$.

Teorema 1.19.1.5. *Si $h\bar{p} \leq 2$, entonces*

$$\|u - y\|_\infty \leq M\|T_h y\|_\infty, \quad M = \text{máx}\{1, 1/\underline{q}\}$$

donde $u = \{u_n\}$ es la solución de diferencias finitas, $y = \{y_n = y(x_n)\}$ es la función de grilla inducida por la solución exacta $y(x)$ del problema y T_h el error de truncación. Si $y \in C^4([a,b])$, entonces

$$\|u - y\|_\infty \leq \frac{1}{12} h^2 M \left(\|y^{(4)}\|_\infty + 2\bar{p}\|y^{(3)}\|_\infty \right)$$

Demostración. Sea $v_n := u_n - y(x_n)$. De

$$\begin{aligned} L_h u_n &= r(x_n), \qquad u_0 = \alpha, \qquad u_{N+1} = \beta \\ Ly(x_n) &= r(x_n), \quad y(x_0) = \alpha, \quad y(x_{N+1}) = \beta \end{aligned}$$

obtenemos que

$$(L_h v)_n = (L_h u)_n - (L_h y)_n = r(x_n) - [(Ly)(x_n) + (L_h y)_n - (Ly)(x_n)] = r(x_n) - r(x_n) - (T_h y)_n = -(T_h y)_n$$

de modo que

$$\|L_h v\|_\infty = \|T_h y\|_\infty.$$

Sabemos que L_h es estable con la constante de estabilidad M. Dado que $v_0 = v_{N+1} = 0$, se sigue que $\|v\|_\infty \leq M\|L_h v\|_\infty = M\|T_h y\|_\infty$, lo que demuestra la primera afirmación. La segunda afirmación se sigue directamente de la estimación del error de truncación. □

Otra forma de escribir el teorema anterior es la siguiente.

Teorema 1.19.1.6. *Sean $p, q \in C^2([a,b])$, $y \in C^6([a,b])$, y $h\bar{p} \leq 2$. Entonces,*

$$u_n - y(x_n) = h^2 e(x_n) + \mathcal{O}(h^4), \quad n = 0, 1, ..., N+1,$$

donde $e(x)$ es la solución de

$$Le = \theta(x), \quad a \leq x \leq b; \quad e(a) = 0,\ e(b) = 0,$$

con $\theta(x) = \frac{1}{12}(y^{(4)}(x) - 2p(x)y^{(3)}(x)$.

Demostración. Observe que $\theta(x) \in C^2([a,b])$, $\theta(x) = \frac{1}{12}(y^{(4)}(x) - 2p(x)y^{(3)}(x)$, y así $e(x) \in C^4([a,b])$.

Sea $\tilde{v}_n := \frac{1}{h^2}(u_n - y(x_n))$. Demostraremos que $\tilde{v}_n = e(x_n) + \mathcal{O}(h^2)$.

Desde la demostración del Teorema anterior, tenemos

$$L_h \tilde{v}_n = -\frac{1}{h^2} T_h y_n.$$

Desde la estimación de T_h tenemos para $v = y$,

$$(L_h \tilde{v})_n = \theta(x_n) + O(h^2).$$

Además,

$$(L_h e)_n = (Le)(x_n) + (L_h e)_n - (Le)(x_n) = \theta(x_n) + (T_h e)_n.$$

Dado que $e \in C^4[a,b]$, tenemos que $T_h e_n = O(h^2)$. De donde

$$L_h v_n = O(h^2), \quad \text{donde } v_n = \tilde{v}_n - e(x_n).$$

Dado que $v_0 = v_{N+1} = 0$ y L_h es estable, se sigue que $|v_n| \leq M\|L_h v\|_1 = O(h^2)$. □

Observación

Una aplicación del teorema anterior es el método de corrección de diferencias debido a L. Fox. Una corrección de diferencias es cualquier cantidad E_n tal que

$$E_n = e(x_n) + O(h^2), \quad n = 1, 2, \ldots, N.$$

Entonces se sigue del teorema que

$$u_n - h^2 E_n = y(x_n) + O(h^4),$$

es decir, $\hat{u}_n = u_n - h^2 E_n$ es una aproximación mejorada con orden de exactitud $O(h^4)$. La idea de Fox es construir una corrección de diferencias E_n aplicando el método de diferencias básico al problema de valor en la frontera $Le = \theta(x)$ en el que $\theta(x_n)$ es reemplazado por una aproximación de diferencias adecuada Θ_n:

$$(L_h E)_n = \Theta_n, \quad n = 1, 2, \ldots, N; \quad E_0 = 0, \quad E_{N+1} = 0.$$

Haciendo $v_n = E_n - e(x_n)$, encontramos

$$(L_h v)_n = (L_h E)_n - (L_h e)_n = \Theta_n - \theta(x_n) + O(h^2).$$

Dado que $v_0 = v_{N+1} = 0$, la estabilidad entonces implica

$$|v_n| = |E_n - e(x_n)| \leq M\|\Theta - \theta\|_1 + O(h^2),$$

por lo que para que (7.106) se cumpla, todo lo que necesitamos es asegurarnos de que

$$\Theta_n - \theta(x_n) = O(h^2), \quad n = 1, 2, \ldots, N.$$

Esto se puede lograr reemplazando las derivadas en la definición de $\theta(x)$ por aproximaciones de diferencias adecuadas.

1.20. Una segunda versión de estabilidad

En una dimensión tenemos que la **ecuación de Poisson en una dimensión** toma la siguiente forma

$$\begin{cases} -u''(x) = r(x), x \in (0,1) \\ u(0) = \alpha \\ u(1) = \beta \end{cases}$$

Por lo cual, usando diferencias finitas vemos que

$$\begin{cases} -\dfrac{1}{h^2}\left(u_{n-1}-2u_n+u_{n+1}\right)=r\left(x_n\right), \quad 1\leq n\leq N, \\ u_0=\alpha \\ u_n=\beta \end{cases}$$

El cual, puede ser reescrito de la forma

$$A_{1\mathrm{d}}\mathbf{u}=\mathbf{b}$$

donde $\mathbf{u}=[u_1,u_2,...,u_N]^\top$, y

$$A_{1\mathrm{d}}=\frac{1}{h^2}\begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \quad \text{y} \quad \mathbf{b}=\begin{bmatrix} r\left(x_1\right) \\ r\left(x_2\right) \\ \vdots \\ r\left(x_{n-2}\right) \\ r\left(x_{n-1}\right) \end{bmatrix}-\frac{1}{h^2}\begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{bmatrix}$$

Para este caso, tenemos que el **error de truncación** viene dado por la siguiente expresión

$$\left(T_h u\right)_n=-\frac{1}{h^2}\left(u\left(x_{n-1}\right)-2u\left(x_n\right)+u\left(x_{n+1}\right)\right)-r\left(x_n\right), \quad 1\leq n\leq N$$

y usando series de Taylor vemos que

$$\left(T_h u\right)_n=-\frac{1}{12}h^2u^{(4)}\left(x_n\right)+\mathcal{O}\left(h^4\right)=\mathcal{O}\left(h^2\right), \quad \text{cuando } h\to 0$$

Error Global: Observemos que si $v_n=u_n-y\left(x_n\right)$, entonces

$$-\frac{1}{h^2}\left(v_{n-1}-2v_n+v_{n+1}\right)=-(T_h y)_n, \quad 1\leq n\leq N, \qquad (A_{1\mathrm{d}}\mathbf{v}=-T_h\mathbf{y})$$

y $v_0=v_{N+1}=0$. Desde aquí podemos interpretar estas ecuaciones como una discretización de la ecuación

$$e''(x)=-\tau(x), \quad x\in[a,b], \quad e(0)=e(1)=0.$$

Como la función $\tau(x)\approx\dfrac{1}{12}h^2u^{(4)}(x)$ entonces al integrar dos veces en la ecuación diferencial obtenemos

$$e(x)\approx-\frac{1}{12}h^2u''+\frac{1}{12}h^2\left(u''(0)+x\left(u''(1)-u''(0)\right)\right)\sim\mathcal{O}\left(h^2\right).$$

Esto indica que si **resolvemos** las ecuaciones de diferencias entonces tenemos una buena aproximación de la solución de la ecuación de la ecuación diferencial.

Denotemos por $A^h = A_{1\mathrm{d}}$; $\mathbf{v}^h = \mathbf{v}$; $T^h = T_h\mathbf{y}$, el súper índice, denota una dependencia de h. De esta forma,

$$A^h\mathbf{v}^h = -T^h \quad \Longrightarrow \quad \mathbf{v}^h = -\left(A^h\right)^{-1} T^h \quad \Longrightarrow \quad \left\|\mathbf{v}^h\right\| \leq \left\|\left(A^h\right)^{-1}\right\| \left\|T^h\right\|$$

Entonces para $\left\|\mathbf{v}^h\right\| \sim \mathcal{O}\left(h^2\right)$ necesitamos $\left\|\left(A^h\right)^{-1}\right\| \leq C$.

Definición 1.20.0.1. *Suponga que un método de diferencias finitas para un problema de valores de frontera lineal tiene una forma matricial $A^h\mathbf{u}^h = \mathbf{b}^h$, para h tamaño de malla. Decimos que el método es estable si $\left(A^h\right)^{-1}$ existe para todo h suficientemente pequeño y existe una constante C, independiente de h, tal que*

$$\left\|\left(A^h\right)^{-1}\right\| \leqslant C \quad \forall h < h_0.$$

Definición 1.20.0.2. *Decimos que el método de diferencias finitas es* ***consistente*** *con la ecuación diferencial y las condiciones de frontera si*

$$\left\|T^h\right\| \longrightarrow 0 \text{ cuando } h \longrightarrow 0.$$

Decimos que el método de diferencias finitas es ***convergente*** *si*

$$\left\|\mathbf{v}^h\right\| \longrightarrow 0 \text{ cuando } h \longrightarrow 0.$$

Teorema 1.20.0.3 (Teorema fundamental de métodos de diferencias finitas)**.** *Si el método de diferencias finitas es* ***consistente*** *y* ***estable*** *entonces es* ***convergente****.*

1.20.1. Estabilidad en la norma 2.

Definicion de norma 2:

$$\|\mathbf{x}\|_2 = \left(\sum_{n=1}^{N} |x_n|^2\right)^{1/2}, \qquad \|A\|_2 = \max_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$

Para matrices cuadradas la norma 2 corresponde a

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^*A)} = \sigma_{\max}(A), \quad \text{simétricas } \|A\|_2 = \rho(A) := \max_{1\leq n\leq N} |\lambda_n|$$

y para la matriz inversa (simétrica) tenemos la identidad

$$\left\|A^{-1}\right\|_2 = \rho\left(A^{-1}\right) = \max_{1\leq n\leq N} \left|\lambda_n^{-1}\right| = \left(\min_{1\leq n\leq N} |\lambda_n|\right)^{-1}.$$

Entonces, ara demostrar estabilidad necesitamos mostrar que los valores propios de A están acotados por abajo cuando $h \to 0$.

Ejercicio: Para $h = \dfrac{1}{N+1}$ los N valores propios de $A_{1\mathrm{d}}$ son

$$\lambda_n = \frac{2}{h^2}(\cos(n\pi \cdot h) - 1), n = 1, \ldots, N$$

y los vectores propios $\mathbf{u}^n$ asociados a λ_n son

$$u_n^j = \sin(n\pi j h).$$

Así, se satisface que $A^h \mathbf{u}^n = \lambda_n \mathbf{u}^n$, $n = 1, ..., N$

El menor valor propio en magnitud de A^h es

$$\begin{aligned}\lambda_1 &= \frac{\pi}{h^2}(\cos(\pi h) - 1) \\ &= \frac{2}{h^2}\left(-\frac{1}{2!}\pi^2 h^2 + \frac{1}{4!}\pi^4 h^4 + O\left(h^6\right)\right) \\ &= -\pi^2 + O\left(h^2\right)\end{aligned}$$

el cual esta acotado lejos de cero cuando $h \to 0$. Por lo tanto el método es estable y

$$\left\|\mathbf{v}^h\right\|_2 \sim \frac{1}{\pi^2}\left\|T^h\right\|_2.$$

Además, observe que el vector propio $\mathbf{u}^n$ es cercano a la **función propia** del operador diferencial de la ecuación. Para el sistema,

$$\text{Problema de autovalores del Laplaciano} \quad \begin{cases} y'' = \mu y \\ y(0) = 0 \\ y(1) = 0 \end{cases}$$

Entonces las funciones propias y valores propios son

$$y^n(x) = \sin(n\pi x); \quad \mu_n = -n^2\pi^2, \quad n \in \mathbb{Z}.$$

Los valores propios de A^h, λ_n son aproximaciones de $\mathbf{u}^n$, pero tenemos

$$\lambda_n = \frac{2}{h^2}\left(-\frac{1}{2}n^2\pi^2 h^2 + \frac{1}{4!}n^4\pi^4 h^4 + \ldots\right) = \mu_n + \mathcal{O}(h^2).$$

1.20.2. Ecuaciones de segundo orden no lineales

Una extensión natural no lineal del problema lineal de segundo orden es

$$Ky = 0, \quad y(a) = \alpha, \quad y(b) = \beta,$$

donde el operador de segundo orden K es no lineal: $Ky \equiv -y'' + f(x, y, y')$, y $f(x, y, z)$ es una función de clase C^1 definida en $[a, b] \times \mathbb{R} \times \mathbb{R}$ (no lineal en y y/o z).

Analogamente al caso lineal, hacemos la suposición

$$|f_z| \leq \bar{p}, \quad 0 < \underline{q} \leq f_y \leq \bar{q} \quad \text{en} \quad [a,b] \times \mathbb{R} \times \mathbb{R}.$$

Entonces, por el Teorema existencia y unicidad de problemas de segundo orden, este problema tiene una solución única.

Usamos nuevamente la aproximación de diferencias más simple K_h a K,

$$(K_h u)_n \equiv \frac{u_{n+1} - 2u_n + u_{n-1}}{h^2} + f\left(x_n, u_n, \frac{u_{n+1} - u_{n-1}}{2h}\right)$$

Definimos el **error de truncación**, para cualquier función suave v en $[a,b]$, por

$$(T_h v)_n \equiv (K_h v)_n - (Kv)(x_n), \quad n = 1, 2, \ldots, N.$$

Si $v \in C^4[a,b]$, entonces por el teorema de Taylor, aplicado en $x = x_n$, $y = v(x_n)$, $z = v'(x_n)$,

$$\begin{aligned}
(T_h v)_n &= -\frac{v(x_n + h) - 2v(x_n) + v(x_n - h)}{h^2} + v''(x_n) \\
&\quad + f\left(x_n, v(x_n), \frac{v(x_n + h) - v(x_n - h)}{2h}\right) - f(x_n, v(x_n), v'(x_n)) \\
&= -\frac{h^2}{12} v^{(4)}(\xi_1) + f_z(x_n, v(x_n), z_n)\left(\frac{v(x_n + h) - v(x_n - h)}{2h} - v'(x_n)\right) \\
&= -\frac{h^2}{12} v^{(4)}(\xi_1) + f_z(x_n, v(x_n), z_n)\frac{h^2}{6} v^{(3)}(\xi_2),
\end{aligned}$$

donde $\xi_i \in [x_n - h, x_n + h]$, $i = 1, 2$, y $z_n \in [v'(x_n), (2h)^{-1}(v(x_n + h) - v(x_n - h))]$. Así,

$$(T_h v)_n = -\frac{h^2}{12}[v^{(4)}(\xi_1) - 2f_z(x_n, v(x_n), z_n) v^{(3)}(\xi_2)].$$

Dado que K_h es no lineal, la definición de estabilidad necesita ser ligeramente modificada.

Definición 1.20.2.1. *Decimos que el operador de diferencias K_h estable si para h suficientemente pequeño, y para cualquier dos funciones de malla $v = \{v_n\}$, $w = \{w_n\}$, existe una constante M tal que*

$$\|v - w\|_\infty \leq M \max\left(\|v_0 - w_0\|, \|v_{N+1} - w_{N+1}\|\right) + \|K_h v - K_h w\|_\infty), \quad v, w \in \Gamma_h[a,b].$$

Observe que si K_h es lineal, esto se reduce a la definición anterior, ya que $v - w$, al igual que v, es una función de malla arbitraria.

Teorema 1.20.2.2. *Si $h\bar{p} \leq 2$, entonces K_h es estable. De hecho, la desigualdad de estabilidad se cumple con $M = \max(1, 1/q)$.*

Demostración. Ejercicio.

El método de diferencias finitas ahora toma la siguiente forma:

$$(K_h u)_n = 0, \quad n = 1, 2, \ldots, N; \quad u_0 = \alpha, \quad u_{N+1} = \beta.$$

Este es un sistema de N ecuaciones **no lineales** en las N incógnitas $u_1, u_2, \ldots, u_N$.

Ejercicio: Muestre que el error del método satisface que:

$$\|u - y\|_\infty \leq M \|T_h y\|_\infty,$$

y que

$$\|u - y\|_\infty \leq \frac{1}{12} h^2 M \left(\|y^{(4)}\|_\infty - 2\bar{p} \|y^{(3)}\|_\infty \right)$$

Cuando tiene una solución única?

Para mostrar que el método tiene una solución única, escribimos el sistema en forma de punto fijo y aplicamos el principio de mapeo de contracción.

Introducimos un parámetro de relajación $\omega \neq -1$ y escribimos

$$u = \mathbf{g}(u), \quad \mathbf{g}(u) = u - \frac{1}{1+\omega} \frac{1}{2} h^2 K_h u; \quad u_0 = \alpha, \quad u_{N+1} = \beta.$$

Aquí $\mathbf{g} : R^{N+2} \to R^{N+2}$, con $\mathbf{g} = [g_0, g_1, ..., g_N, g_{N+1}]^\top$, definiendo $g_0(u) = \alpha$, $g_{N+1}(u) = \beta$.

Queremos mostrar que g es un mapeo de contracción en R^{N+2} si h satisface la condición $h\bar{p} \leq 2$ y ω se elige adecuadamente. Esto probará la existencia y unicidad de la solución.

Dadas dos funciones de malla $v = \{v_n\}$, $w = \{w_n\}$, podemos escribir

$$g_n(v) - g_n(w) = \frac{1}{1+\omega} [a_n(v_{n-1} - w_{n-1}) + (1 + \omega - b_n)(v_n - w_n) + c_n(v_{n+1} - w_{n+1})], \quad 1 \leq n \leq N;$$

y $g_0(v) - g_0(w) = 0$, $g_{N+1}(v) - g_{N+1}(w) = 0$.

Aqui

$$a_n = \frac{1}{2} \left(1 + \frac{1}{2} h f_z(z_n, y_n, z_n) \right), \quad b_n = 1 + \frac{1}{2} h^2 f_y(x_n, \bar{y}_n, \bar{z}_n), \quad c_n = \frac{1}{2} \left(1 - \frac{1}{2} h f_z(x_n, y_n, z_n) \right)$$

con $\bar{y}_n, \bar{z}_n$ son valores intermedios apropiados. Como $h\bar{p} \leq 2$, tenemos que

$$a_n \geq 0, \quad c_n \geq 0, \quad a_n + c_n = 1.$$

Si asumimos que: $\omega \geq \frac{1}{2} h^2 \bar{q}$, entonces tenemos

$$1 + \omega - b_n \geq 1 + \omega - \left(1 + \frac{1}{2} h^2 \bar{q} \right) = \omega - \frac{1}{2} h^2 \bar{q} \geq 0.$$

Además, como

$$0 \leq 1+\omega-b_n \leq 1+\omega-\left(1+\frac{1}{2}h^2\underline{q}\right) = \omega-\frac{1}{2}h^2\underline{q}$$

tenemos que

$$|g_n(v)-g_n(w)| \leq \frac{1}{1+\omega}\left(a_n+\omega-\frac{1}{2}h^2\underline{q}+c_n\right)\|v-w\|_\infty = \frac{1}{1+\omega}\left(1+\omega-\frac{1}{2}h^2\underline{q}\right)\|v-w\|_\infty.$$

Esto es

$$\|\mathbf{g}(v)-\mathbf{g}(w)\|_\infty \leq \gamma(\omega)\|v-w\|_\infty$$

donde $\gamma(\omega) = 1-\dfrac{h^2\underline{q}/2}{1+\omega} < 1$, lo que muestra que $\mathbf{g}$ es una contracción.

Ejercicios

- Escriba el método de punto fijo (o de iteración sucesiva) para el problema de segundo orden no lineal con diferencias finitas.
- Escriba el método de Newton para el problema de segundo orden no lineal con diferencias finitas.

Ejercicio: La ecuación del péndulo de masa y largo unitario.

$$\theta''(x) = -\sin(\theta(x)), \quad x \in [0,T], \quad \theta(0)=\alpha, \theta(T)=\beta$$

1.21. Métodos Variacionales

1.21.1. Ecuaciones lineales de segundo orden

Consideraremos el problema de Sturm-Liouville

$$L(y) = r(x), \quad a \leq x \leq b, \qquad y(a)=\alpha,\ y(b)=\beta.$$

donde L es el operador autoadjunto

$$L(y) := -\frac{d}{dx}\left(p(x)\frac{dy}{dx}\right)+q(x)y, \quad a \leq x \leq b$$

Asumimos que $p \in C^1([a,b])$ y que q, r son continuas en $[a,b]$ y que existen constantes positivas $\underline{p}$ y $\underline{q}$ tales que

$$p(x) \geq \underline{p} > 0, \quad q(x) \geq \underline{q} > 0, \quad a \leq x \leq b.$$

Bajo estos supuestos, el problema tiene una única solución.

Problema homogéneo

Transformamos el problema a uno con condiciones de frontera homogénea

$$L(y) = r(x), \quad a \leq x \leq b, \qquad y(a) = 0 \; y(b) = 0.$$

Denotando el espacio $C_0^2[a,b] := \{u \in C^2([a,b]) : \; u(a) = u(b) = 0\}$, podemos reescribir el problema de la siguiente forma:

$$Ly = r, \quad y \in C_0^2([a,b]).$$

Observe que $L : C^2([a,b]) \to C([a,b])$ es un operador lineal. Es conveniente definir un espacio mas grande que $C_0^2([a,b])$, así definimos

$$V_0 = \{v \in C([a,b]) : \; v' \text{ es continua por tramos y acotado sobre } [a,b], \; v(a) = v(b) = 0\}.$$

Producto interno

Sobre V_0 denifinimos el producto interno (usual)

$$(u,v) := \int_a^b u(x)v(x)dx, \quad u, v \in V_0$$

Teorema 1.21.1.1. *El operador L es simétrico sobre $C_0^2([a,b])$ relativo al producto interno, esto es:*

$$(Lu, v) = (u, Lv), \quad u, v \in C_0^2([a,b]).$$

Demostración. Usando integración por partes, obtenemos

$$\begin{aligned}(Lu, v) &= \int_a^b \left(-\frac{d}{dx}\big(p(x)\frac{du}{dx}\big) + q(x)u(x) \right) v(x)\, dx \\ &= -(pu')\Big|_a^b + \int_a^b \big(p(x)u'(x)v'(x) + q(x)u(x)v(x)\big)\, dx \\ &= \int_a^b \big(p(x)u'(x)v'(x) + q(x)u(x)v(x)\big)\, dx\end{aligned}$$

El último término es simétrico en u y v, esto es también igual a (Lv, u), lo que prueba el teorema. □

Forma variacional

Observe que la última integral en la demostración es definida no sólo para funciones de $C_0^2([a,b])$, pero también en V_0. Así, tenemos una producto interno alternativo

$$[u,v] := \int_a^b \big(p(x)u'(x)v'(x) + q(x)u(x)v(x)\big)\, dx, \quad u, v \in V_0.$$

La demostración entonces muestra que:

$$(Lu, v) = [u, v], \quad \text{si } u \in C_0^2([a, b]), \quad v \in V_0.$$

En particular, si $u = y$ es la solución del problema de valores de frontera, entonces

$$[y, v] = (r, v), \quad \forall v \in V_0,$$

esta es la **forma variacional**, o forma débil, del problema.

Teorema 1.21.1.2. *Existen constantes positivas $\underline{c}$ y $\bar{c}$ tales que*

$$\underline{c}\|u\|_\infty^2 \leq [u, u] \leq \bar{c}\|u'\|_\infty^2, \quad \forall u \in V_0.$$

En efecto,

$$\underline{c} = \frac{p}{b-a}, \quad \bar{c} = (b-a)\|p\|_\infty + (b-a)^3\|q\|_\infty.$$

Demostración. Para todo $u \in V_0$, como $u(a) = 0$, tenemos que

$$u(x) = \int_a^x u'(t)dt, \quad x \in [a, b].$$

Entonces, por la desigualdad de Schwarz

$$u^2(x) \leq \int_a^x 1dt \int_a^x (u'(t))^2 dt \leq (b-a) \int_a^b (u'(t))^2 dt, \quad x \in [a, b],$$

y por lo tanto

$$\|u\|_\infty^2 \leq (b-a) \int_a^b (u'(t))^2 dt \leq (b-a)^2 \|u'\|_\infty^2.$$

Bajo los supuestos de las funciones p y q, obtenemos

$$[u, u] = \int_a^b \left(p(x)(u'(x))^2 + q(x)u^2\right) dx \geq \underline{p} \int_a^b (u'(x))^2 dx \geq \frac{p}{b-a}\|u\|_\infty^2$$

Lo último demuestra la primera desigualdad del teorema. La segunda desigualdad se obtiene observando que

$$[u, u] \leq (b-a)\|p\|_\infty \|u'\|_\infty^2 + (b-a)\|q\|_\infty \|u\|\infty^2 \leq \bar{c}\|u'\|_\infty^2.$$

□

1.21.2. Unicidad de la solución

Observamos que el teorema anterior demuestra la unicidad de las soluciones del problema homogéneo. En efecto, si

$$Ly = r, \quad Ly^\star = r, \quad y, y^\star \in C_0^2([a, b]),$$

entonces $L(y - y^\star) = 0$. Esto implica que

$$0 = (L(y - y^\star), y - y^\star) = [y - y^\star, y - y^\star] \geq \underline{c}\|y - y^\star\|_\infty^2$$

de donde $y = y^\star$.

1.21.3. El problema del valor extremo

Definimos el funcional cuadrático

$$F(u) := [u,u] - 2(r,u), \quad u \in V_0.$$

Teorema 1.21.3.1. *Sea y la solución del problema: $Ly = r,\ y \in C_0^2([a,b])$. Entonces,*

$$F(u) > F(y), \quad \forall u \in V_0,\ u \neq y.$$

Demostración. Tenemos la solución del problema variacional, $(r,u) = [y,u]$, así

$$F(u) = [u,u] - 2(r,u) = [u,u] - 2[y,u] + [y,y] - [y,y] = [y-u, y-u] - [y,y] > -[y,y].$$

Por otro lado, como $[y,y] = (Ly,y) = (r,y)$, tenemos

$$F(y) = [y,y] - 2(r,y) = (r,y) - 2(r,y) = -(r,y) = -[y,y]$$

lo que demuestra el teorema. □

Propiedad del valor extremo

El teorema anterior nos permite escribir la siguiente propiedad del valor extremo de la solución

$$F(y) = \min_{u \in V_0} F(u).$$

Además, satisface la siguiente identidad

$$[y-u, y-u] = F(u) + [y,y], \quad u \in V_0.$$

Aproximación de la solución del problema extremo

Sea $S \subset V_0$ un **subespacio de dimensión finita** de V_0 y con dimensión $\dim(S) = n$. Sea $u_1, u_2, \ldots, u_n$ una base de S, así

$$u \in S, \text{ si y sólo sí } u = \sum_{i=1}^{n} \xi_i u_i, \quad \xi_i \in \mathbb{R}.$$

Aproximamos la solución y del problema del minimización por $u_S \in S$, el cual satisface que

$$F(u_S) = \min_{u \in S} F(u), \quad u_S \approx y.$$

Cómo estudiamos la calidad de esta aproximación?

1.21.4. El método

Para toda función $u \in S$, tenemos

$$F(u) = \left[\sum_{i=1}^{n} \xi_i u_i, \sum_{j=1}^{n} \xi_j u_j\right] - 2\left(r, \sum_{i=1}^{n} \xi_i u_i\right) = \sum_{i,j=1}^{n} [u_i, u_j]\xi_i\xi_j - 2\sum_{i=1}^{n}(r, u_i)\xi_i$$

Definamos entonces la matriz y los vectores

$$U = \begin{bmatrix} [u_1, u_1] & [u_1, u_2] & \cdots & [u_1, u_n] \\ [u_2, u_1] & [u_2, u_2] & \ddots & \vdots \\ \vdots & \ddots & \ddots & [u_{n-1}, u_n] \\ [u_n, u_1] & \cdots & [u_n, u_{n-1}] & [u_n, u_n] \end{bmatrix}, \quad \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}, \quad \rho = \begin{bmatrix} (r, u_1) \\ (r, u_2) \\ \vdots \\ (r, u_n) \end{bmatrix}$$

U : matriz de rigidez, $\quad \rho$: vector de carga

En términos de esta definición, el funcional puede escribirse como

$$F(u) = \xi^\top U\xi - 2\rho^\top \xi, \quad \xi \in \mathbb{R}^n.$$

Observe que, la matriz U es simétrica y definida positiva ($\xi^\top U\xi = [u, u] > 0$, a menos que $u = 0$). Nuestro problema aproximado nos queda

$$\phi(\xi) = \text{mín}, \quad \phi(\xi) := \xi^\top U\xi - 2\rho^\top \xi, \quad \xi \in \mathbb{R}^n$$

este es un problema de minimización cuadrática sin restricciones en $\mathbb{R}^n$. Como la matrix U es simétrica y definida positiva el problema tiene una solución única $\hat{\xi}$ dada por la solución del sistema lineal:

$$U\xi = \rho$$

Lema 1.21.4.1. *Si $\hat{\xi}$ es solución del sistema lineal, entonces se verifica que:*

$$\phi(\xi) > \phi(\hat{\xi}), \quad \forall \xi \in \mathbb{R}^n,\ \xi \neq \hat{\xi}.$$

Demostración. Tenemos que:

$$\begin{aligned} \phi(\xi) &= \xi^\top U\xi - 2\rho^\top \xi \\ &= \xi^\top U\xi - 2\hat{\xi}^\top U\xi \\ &= \xi^\top U\xi - 2\hat{\xi}^\top U\xi + \hat{\xi}^\top U\hat{\xi} - \hat{\xi}^\top U\hat{\xi} \\ &= (\xi - \hat{\xi})^\top U(\xi - \hat{\xi}) + \phi(\hat{\xi}) \end{aligned}$$

donde en el último paso usamos que:

$$-\hat{\xi}^\top U\xi = -\hat{\xi}^\top \rho = \hat{\xi}^\top \rho - 2\rho^\top \hat{\xi} = \hat{\xi}^\top U\hat{\xi} - 2\rho^\top \hat{\xi} = \phi(\hat{\xi})$$

□

Propiedad de mejor aproximación

El método en la práctica, escoge funciones base del espacio S que proporcionen alguna ventaja. Por ejemplo que tengan soporte pequeño y que resulte en que la matrix U tenga alguna estrucura particular.

Teorema 1.21.4.1. *Si $u_S \in S$ es la aproximación de y la solución del problema de valor extremo, entonces, esta satisface*

$$[y - u_S, y - u_S] = \min_{u \in S} [y - u, y - u]$$

Demostración. Ejercicio.

Estimación del error

Teorema 1.21.4.2. *Se satisface que*

$$\|y - u_S\|_\infty \leq \sqrt{\frac{\bar{c}}{\underline{c}}} \|y' - u'\|_\infty, \quad \forall u \in S.$$

En particular, tenemos que

$$\|y - u_S\|_\infty \leq \sqrt{\frac{\bar{c}}{\underline{c}}} \inf_{u \in S} \|y' - u'\|_\infty$$

Demostración. Se sigue que:

$$\underline{c} \|y - u_S\|_\infty^2 \leq [y - u_S, y - u_S] \leq [y - u, y - u] \leq \bar{c} \|y' - u'\|_\infty^2.$$

Ejemplo, spline cúbica por tramos

Sea Γ_h una subdivisión del intervalo $[a, b]$, esto es, $a = x_1 < x_2 < \ldots < x_n = b$, y sea el subespacio

$$S = \{s \in C^2([a, b]),\ s\big|_{[x, x_{i+1}]} \in \mathbb{P}_3 : s(a) = s(b) = 0\} \subset V_0$$

Verifique que la dimensión de este espacio es n.

Dada la solución y, existe una única $s \in S$ (el interpolante spline cubico completo de y) tal que:

$$s(x_i) = y(x_i), \quad i = 1, 2, \ldots, n, \quad s'(a) = y'(a),\ s'(b) = y'(b).$$

Desde la teoría de spline cúbicas tenemos

$$\|s' - y'\|_\infty \leq \frac{1}{24} \max_i |x_i - x_{i+1}|^3 \|y^4\|_\infty, \quad \text{si } y \in C^4([a, b]).$$

Luego se sigue que

$$\|y - u_S\|_\infty \leq \frac{1}{24} \sqrt{\frac{\bar{c}}{\underline{c}}} \max_i |x_i - x_{i+1}|^3 \|y^4\|_\infty.$$

1.22. Problemas singularmente perturbados

1.22.1. Perturbaciones singulares

Consderamos el model estado estacionario de advección-difusión. Este se deriva del problema dependiente del tiempo:

$$\frac{\partial}{\partial t}u + a\frac{\partial}{\partial x}u = \kappa\frac{\partial^2}{\partial x^2}u + \psi$$

este problema modela la temperatura $u(x,t)$ de un fluído en una tubería con velocidad constante a, donde el fluído tiene una constante de difusión de calor κ y ψ es un término fuente de calor. Condiciones de borde $u(0,t) = \alpha(t)$, $(u(1,t) = \beta(t)$.

En el caso que α, β, ψ sean independiente de t, podemos esperar una solución de estado estacionario. De donde obtenemos el problema

$$au'(x) = \kappa u''(x) + \psi(x),\ x \in [0,1], \quad u(0) = \alpha, u(1) = \beta$$

Reescribimos el problema para el número de Péclet $\epsilon = \kappa/a$ por

$$\epsilon u''(x) - u'(x) = f(x),\ x \in [0,1], \quad u(0) = \alpha, u(1) = \beta$$

Solución de la ecuación:

$$u(x) = \alpha + x + (\beta - \alpha - 1)\left(\frac{\exp(x/\epsilon) - 1}{\exp(1/\epsilon) - 1}\right)$$

Observe que $\epsilon \to 0$, la solución tiene a una función discontinua qee salta al valor β cerca de 1. Esta región de transición rápida se conoce como capa límite (**boundary layer**). El problema anterior con $0 < \epsilon \ll 1$ se conoce como **ecuación singularmente perturbada**.

Problemas singularmente perturbados causan dificultades numérica debido a que la solución cambia rápidamente sobre una región muy pequena. Recordemos que el error en la aproximación de u'' es proporcional a $h^2u^{(4)}$. Si h no es lo suficientemente pequeno entonces el error de truncación sera muy grande en la capa límite.

1.22.2. Capas interiores

Consideremos como ejemplo el problema no linear de valores de frontera

$$\epsilon u'' + u(u' - 1) = 0, \quad x \in [a,b], \quad u(a) = \alpha,\ u(b) = \beta.$$

para valores de ϵ pequenos este es un problema singularmente perturbado porque ϵ multiplica la derivada de mayor orden. Para $\epsilon = 0$ la ecuación queda de primer orden y tenemos que solo una condicón de frontera en general.

$$u(u' - 1) = 0 \quad \Longrightarrow \begin{cases} u(x) = x + \alpha - a, & \text{si } u(a) = \alpha \\ u(x) = x + \beta - b, & \text{si } u(b) = \beta \end{cases}$$

Bibliografía

[1] Evans, G., Blackledge, J., Yardley, P. Analytic Methods for Partial Differential Equations. 1999.

[2] Ern, A. and Guermond, J.-C. Theory and Practice of Finite Elements. Springer Science & Business Media, 2013.

[3] Strauss, W. A. Partial Differential Equations: An Introduction. John Wiley & Sons, 2007.

[4] Atkinson, K. and Han, W. Theoretical Numerical Analysis. Vol. 39. Berlin: Springer, 2005.

[5] Iserles, A.A First Course in the Numerical Analysis of Differential Equations. No.44. Cambridge University Press, 2009.

[6] Braess, D. Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics. Cambridge University Press, 2007.

[7] Brenner, S.C. and Scott,L.R. The Mathematical Theory of Finite Element Methods. Springer, 3o edición, 2008.

[8] Evans, L.C. Partial differential equations. Vol. 19. American Mathematical Soc., 2010.

[9] Evans, G., Blackledge, J., Yardley, P. Numerical Methods for Partial Differential Equations.2000.

[10] Gautschi, W. Numerical Analysis: An Introduction, Springer Science & Business Media, 2nd edition, 2011

[11] Gustafsson, B. High Order Difference Methods for Time Dependent PDEs. Vol.38. Springer Science & Business Media, 2007.

[12] Hairer, E., Lubich, C., and Wanner, G. Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations. Vol. 31. Springer Science & Business Media, 2006.

[13] Logg, A., Mardal, K.A and Wells, G., Automated Solution of Differential Equations by the Finite Element Method, Springer, 2012.

[14] Sauter, S. A. and Schwab, C. Boundary Element Methods. Springer, Berlin, Heidelberg, 2010.183-287.

[15] Strang, G. Introduction to Applied Mathematics. 1986.

[16] Strikwerda, J.C. Finite Difference Schemes and Partial Differential Equations. Vol.88. SIAM, 2004.

[17] Suli