Métodos Numéricos III

Profesor Guillermo González

Transcriptor:
Bartomeu Kane Binimelis

Primavera de 1708

Índice general

Capítulo 1

Introducción

En este capítulo presentamos muy brevemente algunos de los temas que iremos desarrollando a lo largo del curso. Introducimos tanto problemas como métodos para poder resolverlos.

El Problema de Valor Inicial

$$\begin{cases} y' = f(x,y) \\ y(a) = \eta \end{cases}$$

nos lo encontramos frecuentemente. Dicho problema es resoluble analíticamente en contadas ocasiones. Es por tanto esperable que se hayan investigado diversas técnicas numéricas para abordarlo. Nuestro objetivo será obtener una solución numérica aproximada en el intervalo $[a,b]$.

Procederemos a discretizar el problema en el intervalo que queremos la solución:

$$x_n = x_0 + nh \quad \text{donde} \quad x_0 = a$$

Aquí h se conoce como *paso* y $n = 0, 1 \ldots, N$. De esta forma observamos que $N = \frac{b-a}{h}$ y $x_N = b$.

Mediante la aplicación de algún método pasaremos de (x_n, y_n) a (x_{n+1}, y_{n+1}). Habitualmente supondremos $y_n \simeq y(x_n)$, donde $y(x_n)$ es la solución exacta del PVI en el punto x_n. Una vez realizadas todas las iteraciones habremos obtenido un conjunto $\{y_n : \ n = 0, 1 \ldots, N\}$, que es el que nos da la información sobre la función solución del PVI inicial.

Los métodos que estudiaremos se nos presentan en forma de fórmulas recurrentes. Estas fórmulas pueden ser *explícitas*, cuando en ellas no aparece la imagen por alguna función de un punto desconocido o, en caso contrario, *implícitas*, cuando se involucra la imagen por una función de algún punto que todavía tenemos que determinar.

Ejemplo 1.1. Supongamos que tenemos $y' = f(x,y)$. Para aproximar la siguiente imagen en nuestro PVI podemos utilizar el denominado *método de Euler* en su variante explícita o implícita.

- Explícito:

$$\frac{y_{n+1} - y_n}{h} \simeq f(x_n, y_n)$$

$$y_{n+1} = y_n + hf_n, \quad \text{donde} \quad f_n = f(x_n, y_n)$$

- Implícito:

$$\frac{y_{n+1}-y_n}{h} \simeq f(x_{n+1}, y_{n+1})$$
$$y_{n+1} = y_n + hf_{n+1}, \quad \text{donde} \quad f_{n+1} = f(x_{n+1}, y_{n+1})$$

Es corriente que los métodos implícitos presenten una serie de ventajas frente a los explícitos (si no fuera así no habría motivo aparente para utilizarlos, puesto que con uno explícito la resolución del problema es más simple):

- Mayor orden de convergencia.
- Únicos útiles para resolver sistemas *"stiff"*.

Estudiaremos dos tipos de sistemas de EDOs. Son los conocidos como "stiff" y los no "stiff". Los primeros son menos habituales pero nos enfrentan a mayores dificultades a la hora de resolverlos. En este curso solo los introduciremos, y profundizaremos más en el caso no "stiff".

Dependiendo del número de puntos anteriores necesarios al que estamos calculando hablaremos de métodos *multipaso*. En éstos será necesario contar con más de un punto (recordemos que el inmediantamente anterior es (x_n, y_n)) para calcular el siguiente. En un método de k pasos deberemos tener la información de k puntos anteriores al que vamos a aproximar ahora.

$$\left.\begin{matrix}(x_{n-j}, y_{n-j}) \\ j = 0, 1 \dots k-1\end{matrix}\right\} \quad \xrightarrow{\text{Metodo}} \quad (x_{n+1}, y_{n+1})$$

Ejemplo 1.2. Método de dos pasos.

$$\frac{y_{n+2}-y_n}{2h} \simeq f(x_{n+1}, y_{n+1}) \quad \Longrightarrow \quad y_{n+2} = y_n + 2hf_{n+1}$$

En este caso se requieren dos valores iniciales: $(x_0, y_0), (x_1, y_1)$.

Estudiaremos también un tipo distinto de métodos: los que pertenecen a la familia de los *métodos Runge-Kutta*. Estos se caracterizan por evaluar la función $f(x, y)$ en puntos diferentes a los $x_n,\ n = 0, \dots, N$.

Ejemplo 1.3. Método Runge-Kutta (un paso)

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

Ejemplo 1.4. Método de Heun de orden 2 (un paso)

$$\left.\begin{aligned} k_1 &= f(x_n, y_n) = f_n \\ k_2 &= f(x_{n+1}, y_n + hk_1) \\ y_{n+1} &= y_n + h\left(\frac{k_1 + k_2}{2}\right)\end{aligned}\right\}$$

También se conoce como Runge-Kutta de dos niveles explícito.

Capítulo 2

Problemas de Condiciones Iniciales

Consideremos el PVI $\begin{cases} y' = f(x,y) \\ y(a) = \eta \end{cases}$ en $x \in [a,b]$. Queremos obtener una solución aproximada de dicho problema mediante un método numérico.

Los métodos que consideraremos son los denominados de *discretización* o de *variable discreta*. Como hemos descrito brevemente en la introducción, estos métodos se basan en aproximar la función solución en N puntos.

2.1. Método de k pasos

Supongamos que tenemos $y_n, y_{n+1} \dots y_{n+k-1}$. En este caso nos interesará conocer el valor aproximado de y_{n+k}. Para ello utilizaremos un algoritmo de la siguiente forma:

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \cdot \phi_f(y_{n+k} \dots y_n, x_n; h)$$

donde h es el paso y k el número de pasos del método. Habitualmente supondremos que $y_j \simeq y(x_j)$, siendo y_j el valor aproximado de la solución que hemos obtenido en la iteración anterior y $y(x_j)$ el valor exacto de tal solución. Asimismo, notemos que $x_j = x_0 + jh$, de forma que $x_0 = a$ y $x_N = b$.

Los valores α_j son los pesos específicos que cada método en particular asigna a los valores anteriores para obtener y_{n+k}. Ocurre lo mismo con la *función de iteración* ϕ: esta función varía según el método, y como la notación indica, depende de f y de los puntos anteriores que conocemos.

2.1.1. Ejemplos

Acabamos de introducir los métodos de discretización en general. Para comprender mejor cómo funcionan veamos algunos ejemplos de los algoritmos más conocidos y utilizados.

Método de Euler

Es el más simple de todos: consiste en tomar $\phi_f = f(x_n, y_n)$. Observamos que se trata de un método de un solo paso, con las ventajas de complejidad para la implementación que

ello representa, pero con los inconvenientes de bajo orden de convergencia que probablemente conlleve.

Método Runge-Kutta Clásico

Se trata de un algoritmo con orden de convergencia igual a 4[1]. Consiste en aproximar el punto y_{n+1} por:

$$y_{n+1} = y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

con

$$k_1 = f(x_n, y_n) \qquad k_3 = f(x_n + \frac{h}{2}, y_n + \frac{h}{2}k_2)$$
$$k_2 = f(x_n + \frac{h}{2}, y_n + \frac{h}{2}k_1) \qquad k_4 = f(x_n + h, y_n + hk_3)$$

Notemos que aquí la función de iteración es lineal en k_i.

Método Lineal Multipaso

Como su nombre sugiere, la función de iteración es lineal en $f(x_{n+j}, y_{n+j}) = f_{n+j}$. Tomamos $\phi_f = \sum_{j=1}^{k} \beta_j f_{n+j}$. Aplicando la expresión general que hemos presentado al inicio del capítulo tenemos que:

$$\sum_{j=1}^{k} \alpha_j y_{n+j} = h[\sum_{j=1}^{k} \beta_j f_{n+j}]$$

con $\alpha_j, \beta_j \in \mathbb{R}$, $\alpha_k = 1$ y $|\alpha_0| + |\beta_0| \neq 0$. Cabe destacar que dependiendo del valor de β_k nos encontraremos frente a un método explícito ($\beta_k = 0$) o implícito ($\beta_k \neq 0$).

Definición 2.1. Consideremos un método dado por la expresión $\sum_{j=1}^{k} \alpha_j y_{n+j} = h \sum_{j=1}^{k} \beta_j f_{n+j}$, con $\alpha_k = 1$. Entonces, su *polinomio característico* es:

$$\rho(r) = \sum_{j=0}^{k} \alpha_j r^j$$

Esta definición nos permite denotar unívocamente un método por la pareja $\{\rho(r), \phi_f\}$.

2.2. Errores, Convergencia, Consistencia y Orden

Para poder comprender cómo funcionan los métodos numéricos es conveniente analizar las diversas formas de error que estos inducen. Esto motiva las definiciones siguientes.

Definición 2.2. *Función tau del método*

$$\tau(x, h) = h^{-1} \sum_{j=0}^{k} \alpha_j y(x + jh) - \phi_f(y(x + kh) \dots y(x), x; h)$$

Si tomamos $x_n \in [a, b]$, $x_{n+j} = x_n + jh$ entonces en x_n nos queda:

$$\tau(x_n, h) = h^{-1} \sum_{j=0}^{k} \alpha_j y(x_{n+j}) - \phi_f(y(x_{n+k}) \dots y(x_n), x_n; h)$$

[1] En breve definiremos el concepto de *orden de convergencia*.

Definición 2.3. El *error local de discretización (ELD)* es el error que resulta de aplicar el método por primera vez, partiendo de la solución exacta. Veamos que está íntimamente relacionado con la función tau:

$$ELD_{n+1} = y(x_{n+k}) - \tilde{y}_{n+k} = h\tau(x_n, h)$$
$$\tilde{y}_{n+k} = -\sum_{j=0}^{k-1} \alpha_j y(x_{n+j}) + h\phi_f(y(x_{n+k}) \dots y(x_n), x_n; h)$$

En este cálculo hemos supuesto que los valores calculados para puntos anteriores son exactos, como ya hemos indicado.

Definición 2.4. El *error global de discretización (EGD) en* x_{n+k} para una solución numérica $\{y_n : n = 0 \dots N = \frac{b-a}{h}\}$ es $e_{n+k} = y(x_{n+k}) - y_{n+k}$. Notemos que se diferencia de la ELD por estar la solución calculada a partir de los valores anteriormente hallados por el método.

Definición 2.5. El *error global de discretización (EGD) de la solución numèrica* $\{y_n : n = 0 \dots N = \frac{b-a}{h}\}$ se calcula a partir de $\max\limits_{0 \le n \le N} \|y(x_n) - y_n\|$.

Ejemplo 2.6. Análisis del método $y_{n+1} = y_n + hf(x_n, y_n)$.
Consideremos el ELD asociado al método:

$$h\tau(x, h) = y(x_n + h) - y_n - hf(x_n, y(x_n))$$

Aplicando la aproximación de Taylor de orden 2 sobre $y(x_n + h)$ y suponiéndola $\mathcal{C}^\infty[a, b]$ nos queda:

$$h\tau(x_n, h) = \frac{h^2}{2} y''(x_n) + O(h^3)$$
$$y(x_{n+1}) - \tilde{y}_{n+1} = h\tau(x_n, h) = y(x_{n+1}) - y(x_n) - hf(x_n, y(x_n))$$

Respecto al error global, ¿qué podemos decir de $y(x_{n+1}) - y_{n+1}$?

$$\begin{aligned}
e_1 &= y(x_1) - y_1 \overset{(*)}{=} \frac{h^2}{2} y''(x_0) + O(h^3) \\
e_2 &= y(x_2) - y_2 = [y(x_1) + hf(x_1, y(x_1)) + h\tau(x_1, h)] - [y_1 + hf(x_1, y_1)] = \\
&= [\underbrace{y(x_1) - y_1}_{e_1}] + h[\underbrace{f(x_1, y(x_1)) - f(x_1, y_1)}_{e_1 f_y(x_1, y_1) + O(e_1)} + \tau(x_1, h)] = \\
&= e_1 + \frac{h^2}{2} y''(x_1) + O(h^3) \overset{(**)}{=} 2\left[\frac{h^2}{2} y'(x_0)\right] + O(h^3)
\end{aligned}$$

donde hemos aplicado:

$(*)$ $y(x_0) = y(a) = \eta$
$(**)$ aplicando Taylor entorno a x_0

Para obtener el error global en la enésima iteración se aplica inducción: suponemos que en $e_n = n\left[\frac{h^2}{2} y''(x_0)\right] + O(h^3)$ y se deduce que $e_{n+1} = (n+1)\left[\frac{h^2}{2} y''(x_0)\right] + O(h^3)$.
Recordemos que $x_{n+1} = x_0 + (n+1)h \Rightarrow n + 1 = \frac{x_{n+1} - a}{h}$. Por tanto,

$$e_{n+1} = h\left[\frac{x_{n+1} - a}{2} y''(x_0)\right] + O(h^2)$$

Observamos que $\|e_{n+1}\| = O(h)$, comportamiento que también presenta la función tau $\tau(x,h) = O(h)$. Por último, destacar que existe la posibilidad de escribir los sucesivos errores globales como ecuaciones lineales en diferencias, introduciendo la sucesión de funciones φ_n para tal objetivo.

$$\begin{aligned} e_2 - e_1 &= \varphi_1(x_1, h) \\ &\cdots \\ e_{n+1} - e_n &= \varphi_n \end{aligned}$$

Hasta ahora, algunos de los conceptos más importantes que hemos considerado son:

- Los problemas (o PVI), definidos por:

$$\left.\begin{aligned} y' = f(x, y(x)) \\ y(a) = \eta \end{aligned}\right\}$$

  donde f cumple la condición Lipschitz para la existencia y unicidad de soluciones.

- Los métodos de la forma $\{\rho(r), \phi_f\}$, donde $\alpha_k = 1$ y $y_i = \eta_i(h)$ para $i = 0 \div k-1$ son los valores iniciales.

$$\sum_{j=1}^{k} \alpha_j y_{n+j} = h\phi_f(y_{n+k} \dots y_n, x; h)$$

- El error global de discretización asociado al método: $EGD = \max\limits_{0 \le n \le N} \|y(x_n) - y_n\|$.

Definición 2.7. Un método es *convergente* si para todo problema se cumple que $EGD \underset{h\to 0}{\longrightarrow} 0$

Definición 2.8. (Alternativa) Un método es *convergente* si para todo problema y toda solución numérica $\{y_n, n = 0 \div N, N = \frac{b-a}{h}\}$ tal que $\lim\limits_{h\to 0} \eta_i(h) = y(x_0)$ se cumple que $\forall n \ \lim\limits_{h\to 0} y_n = y(x_n)$ con $x = a + nh$ fijado.

La función τ del método induce la definición de otra función íntimamente relacionada con la consistencia del método que estemos analizando:

$$\tau(h) = \max_{x\in[a,b-kh]} \tau(x,h) = \max_{x\in[a,b-kh]} h^{-1} \sum_{j=0}^{k} \alpha_j y(x+jh) - \phi_f(y(x+kh) \dots y(x), x; h)$$

Definición 2.9. Un método es *consistente* con el sistema $y' = f(x, y(x))$ cuando $\lim\limits_{h\to 0} \tau(h) = 0$

Definición 2.10. Un método es (al menos) *de orden* p ($p \in \mathbb{Z}$) si para todo problema tal que $y \in \mathcal{C}^p[a,b]$ se cumple que $\tau(h) = O(h^p)$. Si, además $\tau(h) \neq 0(h^{p+1})$ se dice que el método es de orden p *exactamente* (es de orden p).

Teorema 1. *Sea $\tau(x,h)$ continua $\forall h \ge 0, \forall x \in [a,b]$. Si $\tau \to 0$ cuando $h \to 0$ puntualmente para $x \in [a,b]$ entonces:*

1. *$\tau(x,h)$ converge a cero uniformemente cuando $h \to 0 \quad \forall x \in [a,b]$*
2. *El método es consistente: $\lim\limits_{h\to 0} \tau(h) = 0$*
3. *Si $\tau(x,h) = O(h^p)$ con $x \in [a,b]$ entonces $\tau(h) = O(h^p)$ (el método es de orden p)*

Demostración: *Buscar en la bibliografia.*

Teorema 2. *Consideremos un método lineal multipaso* $\{\rho(r), \phi_f\}$ *tal que* ϕ_f *es continua* $\forall x \in [a,b]$ *y* $\forall h \geq 0$. *Si se cumplen:*

1. $\sum_{j=0}^{k} \alpha_j = \rho(1) = 0$
2. $\phi_f(y(x), \ldots, x; 0) = \rho'(1) f(x, y(x)) \quad \forall x \in [a,b]$

entonces el método es consistente. Si además $y(x)$ *no es la función cero entonces (1) y (2) son condiciones necesarias para la consistencia del método.*

Demostración. Veamos para empezar que si se cumplen (1) y (2) entonces el método es consistente. Consideremos $y \in \mathcal{C}^{\infty}[a,b]$. Sea

$$y(x+jh) \overset{(*)}{=} y(x) + jh\hat{y}'(\theta_j)$$

$(*)$ Por el TVM, $\hat{y}'(\theta_j) = [{}^1y'(\theta_{1j}) \ldots {}^m y'(\theta_{mj})]^T, \quad \theta_{tj} \in [x, x+jh],\ t = 1 \div m^{\text{o}}$

Aplicamos este resultado a la función tau:

$$\tau(x,h) = h^{-1}\left[\sum_{j=0}^{k} \alpha_j\right] y(x) + h^{-1}\left[\sum_{j=0}^{k} j\alpha_j h\hat{y}'(\theta_j)\right] - \phi_f(y(x+kh), \ldots y(x), x; h)$$

Ahora, notemos que el límite cuando $h \to 0$ converge uniformemente (por el teorema anterior). Por tanto,

$$\lim_{h\to 0} \tau(x,h) = \sum_{j=1}^{k} (\alpha_j j) y'(x) - \phi_f(y(x) \ldots y(x), x; h)$$

Por la hipótesis (2) deducimos

$$\lim_{h\to 0} \tau(x,h) = 0 \quad \Rightarrow \quad \lim_{h\to 0} \tau(h) = 0$$

que quiere decir que el método es consistente.

Ahora supongamos que tenemos un método consistente. Demostremos que se cumplen:

- $\rho(1) = 0$
- $\rho'(1) f(x, y(x)) = \phi_f(y(x), \ldots y(x), x; 0), \quad x \in [a,b]$

Tenemos la hipótesis de $\lim_{h\to 0} h\tau(x,h) = 0$. Por otro lado,

$$h\tau(x,h) = \sum_{j=0}^{k} \alpha_j y(x+jh) - \underbrace{h\phi_f(y(x) \ldots y(x), x; h)}_{(*)}$$

Observamos que $(*)$ tiende a cero cuando $h \to 0$, y aprovechando que se trata de una función continua podemos separar la suma en dos límites y obtenemos

$$0 = \lim_{h\to 0} = \lim_{h\to 0} \sum_{j=0}^{k} \alpha_j y(x+jh) = \left(\sum_{j=0}^{k} \alpha_j\right) y(x), \qquad \forall x \in [a,b]$$

Por tanto, si $y(x) \neq 0 \; \forall x \in [a,b]$, se deduce el primero de los resultados buscados.
Por otro lado,

$$\tau(x,h) = \sum_{j=0}^{k} (\alpha_j j)\hat{y}'(\theta_j) - \phi_f(y(x+kh)\dots y(x), x; h)$$

$$\lim_{h\to 0} \tau(x,h) \underset{(**)}{=} \sum_{j=0}^{k} (\alpha_j j)\, y'(x) - \phi_f(y(x)\dots y(x), x; 0)$$

y $(**)$ se cumple por consistencia del método, con lo que queda probado (2).

Teorema 3. *Para todo método $\{\rho(r), \phi_f\}$ con ϕ_f continua la consistencia es condición necesaria para la convergencia del método.*

Demostración. Veremos que la convergencia implica los dos puntos del teorema anterior, que a su vez implican la consistencia:

1. $$\sum_{j=0}^{k} \alpha_j y_{n+j} = 0 = h\phi_f(y_{n+k}, \dots y_n, x; h) \quad \Rightarrow \lim_{h\to 0} \sum_{j=0}^{k} \alpha_j y_{n+j} = 0$$

   Notemos que $\lim\limits_{h\to 0} \sum\limits_{j=0}^{k} \alpha_j y_{n+j} = \left(\sum_{j=0}^{k} \alpha_j\right) y(x_n)$, por la convergencia del método. En este caso, si y es no nula se deduce $\sum\limits_{j=0}^{k} \alpha_j = \rho(1) = 0$.

2. $$\sum_{j=0}^{k} \alpha_j y_{n+j} - \sum_{j=0}^{k} \alpha_j y_n = h\phi_f(y_{n+k}, \dots y_n, x_n; h)$$
   $$\sum_{j=1}^{k} j\alpha_j \left[\frac{y_{n+j} - y_n}{jh}\right] = \phi_f(y_{n+k}, \dots y_n, x_n; h)$$

   Tomando ahora el límite para h tendiendo a cero:

   $$\lim_{\substack{h\to 0 \\ x_n = a+nh}} \phi_f = \phi_f(y(x_n), \dots y(x_n), x_n; 0)$$
   $$\lim_{\substack{h\to 0 \\ x_n = a+nh}} \frac{y_{n+j} - y_n}{jh} \underset{(*)}{=} \lim_{h\to 0} \frac{y(x_n + jh) - y(x_n)}{jh} = y'(x_n) = f(x_n, y_n)$$
   $$(*) x_{n+j} = x_n + jh$$

   y para $x = x_n \in [a,b]$ se deduce la condición (2) de consistencia.

2.3. Ecuaciones Lineales en Diferencias (EDLin)

2.3.1. Introducción a las EDLin

Antes de entrar a enunciar (y demostrar en algunos casos) los resultados relacionados con la Ecuaciones Lineales en Diferencias, primero las definiremos de forma rigurosa para mejorar la comprensión del resto del capítulo. También introduciremos algunos teoremas importantes sobre su resolución.

Definición 2.11. Sea $F : \mathbb{R}^m \mapsto \mathbb{R}^n$ y $k \in \mathbb{N}$. Una *Ecuación en Diferencias* es una relación definida por la ecuación:

$$y_{n+k} = F(n, y_n, y_{n+1}, \ldots, y_{n+k}) \quad \forall n = 0, 1, \ldots$$

Notemos pues que una ED tiene como solución una succesión y_n dónde los y_n cumplen la ecuación y $\{y_0, \ldots, y_{k-1}\}$ son valores iniciales conocidos a priori. Hay dos maneras de encontrar y_n:

- Por una parte, usar F de forma succesiva e ir generando valores de la sucesión solución.
- Por otra, una forma más práctica es encontrar una fórmula cerrada del tipo $y_{n+k} = G(n, y_0, \ldots, y_{k-1})$.

En nuestro caso el objetivo será usar esta segunda vía. Nos centraremos en un tipo específico de ecuación que definimos a continuación:

Definición 2.12. Una *ecuación en diferencias lineal con coeficientes constantes de orden k*, que denotaremos por EDLin, es aquella relación de la forma:

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = \varphi_n, \quad \forall n \in \mathbb{N}, \quad \alpha_k = 1$$

donnde $\alpha_j \in \mathbb{R}$ constantes y y_n, ϕ_n tienen la misma dimensión.

Definición 2.13. Decimos que $\pi(r) = r^k + \alpha_{k-1} r^{k-1} + \ldots + \alpha_0$ es el *polinomio característico* de la EDLin.

Por último, introduciremos un par de teoremas básicos sobre EDLin.

Teorema 4. *Toda EDLin de orden k con coeficientes constantes es equivalente a una EDLin vectorial de primer orden de la forma:*

$$z_{n+1} = Bz_n + d_n, \ \forall n = 0, 1, \ldots$$

Siendo $z_n = (y_{n+k-1}, \ldots, y_n)^T$ *y* $d_n = (\phi_n, 0, \ldots, 0)^T$. *Por otra parte:*

$$B = \begin{pmatrix} -\alpha_{k-1} & -\alpha_{k-2} & \ldots & -\alpha_0 \\ 1 & 0 & \ldots & 0 \\ & \ddots & \ddots & \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Además, la matriz B cumple:

1. *$\pi(\mu) = det(B - \mu I)$ es el polinomio característico de la EDLin inicial.*
2. *Si λ es un valor propio de B de multiplicidad algebraica m, entonces la forma canónica de Jordan de B solo tiene un único bloque de Jordan asociado a λ que tiene dimensión $m \times m$.*

Teorema 5. *La solución general de una EDLin es la suma de una solución particular y de la solución general de la EDLin homogénea asociada al problema. Si $\lambda_1, \ldots, \lambda_s$ son las raíces de $\pi(\lambda)$, el polinomio característico de la EDLin, y $k_1, \ldots, k_s$ son sus respectivas multiplicidades, entonces la solución homogénea es combinación lineal de:*

$$\lambda_1^n, n\lambda_1^n, \ldots, n^{k_1-1}\lambda_1^n, \ldots, \lambda_s^n, .., n^{k_s-1}\lambda_s^n$$

2.3.2. Algunos resultados sobre EDLin

El motivo de nuestro interés en las ecuaciones lineales en diferencias es el siguiente: cuando tenemos un método que aplicamos con un paso h y hacemos tender h a cero, las soluciones numéricas y_n que obtenemos se comportan como las soluciones de una EDLin homogénea,

$$\left.\begin{array}{c} \displaystyle\sum_{j=0}^{k} \alpha_j y_{n+j} = 0 \\ \displaystyle\sum_{j=0}^{k} \alpha_j y(x+jh) = h\phi_f + h\tau(x,h) \\ \displaystyle\sum_{j=0}^{k} \alpha_j y_{n+j} = h\phi_f(y_{n+k}, \ldots y_n, x_n; h) \end{array}\right\} \quad \Rightarrow \quad x \equiv x_n$$

Si consideramos $e_{n+j} = y(x+jh) - y_{n+j} \quad \forall n \geq 0,\ j = 0 \div k$ entonces,

$$\sum_{j=0}^{k} \alpha_j e_{n+j} = \varphi_n; \quad \varphi_n = h\left[\phi_f(\text{sol. exacta}) - \phi_f(\text{sol. numérica}) + \tau\right]$$

Algunos resultados sobre normas matriciales

Antes de entrar en el estudio de la estabilidad de las EDLin, introduciremos definiciones y teoremas importantes para la comprensión del resto del capítulo. Son teoremas básicos sobre normas matriciales que usaremos para analizar las condiciones de estabilidad de las EDLin (convenientemente transformadas a un sistema matricial) y que después extenderemos para razonar la estabilidad de los métodos.

Definición 2.14. Decimos que una matriz A es de *clase M* $\iff$ $\forall\lambda$ VAP de A tal que $|\lambda| = \rho(A)$ entonces los bloques de Jordan asociados a λ son de tamaño 1×1.

Teorema 6. *Sea A una matriz. Existe una norma subordinada tal que $||A|| = \rho(A) \iff A$ es de clase M.*

Teorema 7. *Sea A una matriz.*

$$\lim_{k\to\infty} A^k = 0 \iff \rho(A) < 1$$

Además, se cumple que A^k acotada cuando $k \to \infty \iff \rho(A) < 1$ ó A es de clase M y $\rho(A) = 1$.

Estabilidad de EDLin

Consideremos una ecuación en diferencias de primer orden:

$$z_{n+1} = G(z_n, n), \quad z_n \in \mathbb{R}^m, \quad G : \mathbb{R}^m \to \mathbb{R}^m$$

Definición 2.15. Una solución $\{y_n\}_{n\geq 0}$ de la ecuación anterior diremos que es:

- *Estable* si dado $\epsilon > 0$ existe $\delta > 0$ tal que para $\{\hat{y}_n\}$ cualquier otra solución de la ecuación con $\|y_0 - \hat{y}_0\| < \delta$ entonces se tiene que $\|y_n - \hat{y}_n\| < \epsilon$, con $n > 0$.
- *Asintóticamente estable*, si además de ser estable se cumple que $\|y_n - \hat{y}_n\| \underset{n\to+\infty}{\longrightarrow} 0$.

Teorema 8. *(estabilidad de EDLin de orden 1) Consideremos el siguiente sistema de ecuaciones:*

$$z_{n+1} = Bz_n + d, \quad z_n, d \in \mathbb{R}^m, \quad B \in M_m\mathbb{C} \tag{2.1}$$

Entonces una solución $\{y_n\}_{n\geq 1}$ es:

- *Estable* $\iff$ $\rho(B) < 1$ *o* $\rho(B) = 1$ *y* B *de clase M.*
- *Asintóticamente estable* $\iff$ $\rho(B) < 1$.

Demostración. Sea $\{\hat{y}_n\}_{n\geq 0}$ otra solución de la EDLin (2.1).

$$\hat{y}_n - y_n = w_n = Bw_{n-1} = B^2w_{n-2} = \ldots = B^nw_0$$

(a) Primera parte:
$\boxed{\Leftarrow}$ Usando la segunda parte del teorema 7 sabemos que $\exists\sigma \in \mathbb{R}$ tal que $\|B^n\| < \sigma$. Notemos que σ depende de n.

$$\|\hat{y}_n - y_n\| = \|B^nw_0\| \leq \|B^n\| \cdot \|w_0\|$$

Si $\|w_0\| = \|\hat{y}_0 - y_0\| < \delta = \frac{\epsilon}{\sigma}$ entonces $\forall\epsilon > 0$, $\forall n > 0$ puedo encontrar $\delta = \frac{\epsilon}{\sigma}$ tal que si $\|\hat{y}_0 - y_0\| < \delta \quad \Rightarrow \quad \|\hat{y}_n - y_n\| < \epsilon$
$\boxed{\Rightarrow}$ Supongamos que $\|B^n\|$ es no acotada cuando $n \to \infty$. Entonces $\exists x \in \mathbb{R}^n$ tal que $\|B^nx\| \to \infty$. Esto se demuetra tomando $J_B = P^{-1}BP^{-1}$ y distinguiendo entre dos posibles casos para λ_1.

1. Si $|\lambda_1| > 1, \quad \rho(B) > 1$ tomamos $x = P^{-1}e_1$
2. Si $|\lambda_1| = 1$ y $\rho(B) = 1$ tomamos $x = P^{-1}e_2$.,

$$e_i = P^{-1}x, \quad B^n = PJ_B^nP^{-1} \quad \Rightarrow \quad B^nx = PJ_B^ne_i$$

En ambos casos B^nx no es acotado, y tenemos $\|\hat{y}_n - y_n\| = \|w_n\| = \|B^nw_0\|$. Sea $x = \alpha w_n, \quad \alpha \in \mathbb{R}$. Observamos que $\|\hat{y}_n - y_n\| = |\alpha^{-1}| \cdot \|B^nx\|$, y cuando $n \to \infty$ $\|B^nx\|$ no está acotado.
Por tanto, $\|\hat{y}_n - y_n\|$ no está acotado aunque $\|w_0\| = \|\hat{y}_0 - y_0\|$ sí lo esté. En este caso $\{y_n\}_{n\geq 0}$ no es estable.

(b) Segunda parte:
$\boxed{\Leftarrow}$ $\lim\limits_{n\to\infty} B^n = 0 \quad \Rightarrow \quad \lim\limits_{n\to\infty} \|\hat{y}_n - y_n\| = 0$, y por tanto la solución es asintóticamente estable.
$\boxed{\Rightarrow}$ Supongamos que $\rho(B) \geq 1$. Dado que $\exists x \in \mathbb{R}^m$ tal que $\|B^nx\|$ no está acotado entonces $\lim\limits_{n\to\infty} \|\hat{y}^n - y_n\| \neq 0 \quad \Rightarrow \quad \{y_n\}_{n\geq 0}$ no es asintóticamente estable.

Definición 2.16. Dada la ecuación en diferencias de orden k:

$$y_{n+k} + \alpha_{k-1} y_{n+k-1} + \ldots + \alpha_1 y_{n-1} + \alpha_0 y_n = d \tag{2.2}$$

diremos que una solución $\{y_n\}_{n\geq 0}$ es

- *Estable* si dada otra solución $\{\hat{y}_n\}_{n\geq 0}$ se cumpe que $\forall \epsilon > 0\ \exists \delta > 0$ tal que si $|\hat{y}_i - y_i| \leq \delta \quad i = 0, 1, \ldots, k-1$ entonces $|\hat{y}_n - y_n| < \epsilon$ para $n \geq k$.
- *Asintóticamente estable* si además de lo anterior se cumple que $\lim_{n\to\infty} |\hat{y}_n - y_n| = 0$.

Teorema 9. *Sea* (2.3) *la EDLin de primer orden equivalente obtenida de* (2.2) *por el método presentado en el teorema (4).*

$$z_{n+1} = Bz_n + \varphi, \quad z_n = \begin{pmatrix} y_{n+k-1} \\ \vdots \\ y_n \end{pmatrix} \tag{2.3}$$

Las soluciones de (2.2) *son* $\left\{ \begin{matrix} \textit{Estables} \\ \textit{Asintóticamente Estables} \end{matrix} \right\}$ $\iff$ *las soluciones de* (2.3) *son* $\left\{ \begin{matrix} \textit{Estables} \\ \textit{Asintóticamente Estables} \end{matrix} \right\}$

Definición 2.17. Un polinomio $p(x)$ *cumple la condición de la raíz* si:

1. Todas sus raíces son de módulo ≤ 1.
2. Todas sus raíces de módulo 1 son simples.

Teorema 10. *(estabilidad de EDLin de orden k) La EDLin de orden k de la forma* (2.2) *tiene soluciones estables si y solo si el polinomio característico de la EDLin cumple la condición de la raíz.*

Demostración. Dada una EDLin podemos encontrar el sistema $z_{n+1} = Bz_n + d_n$. El polinomio característico asociado a este sistema es:

$$\pi(r) = \sum_{j=0}^{k} \alpha_j r^j = \det(B - rI)$$

Entonces si $\rho(B) \leq 1$ todos los VAPs de B son de módulo $\leq 1 \iff$ todas las raíces de $\pi(r)$ son de módulo ≤ 1.
Por otro lado, si B es de clase M resulta que tiene bloques de Jordan 1×1. Además, B ha sido obtenida a partir de transformaciones del sistema, de forma que tendremos solo un VAP por bloque de Jordan. De esta forma queda demostrado que los VAPs de módulo 1 solo tienen un bloque 1×1 en la forma de Jordan de B, que son raíces simples de $\det(B - rI)$, que es lo mismo que decir que las raíces de módulo 1 de $\pi(r)$ son simples.

Comentario 2.18. Recordemos las dos maneras que tenemos de escribir una EDLin:

$$\sum_{j=0}^{k} \alpha_j y_{n+k} = \phi_n, \quad \alpha_k = 1 \tag{2.4}$$

$$z_{n+1} = Bz_n + d_n, \quad z_n = \begin{pmatrix} y_{n+k-1} \\ \vdots \\ y_n \end{pmatrix}, \quad d_n = \begin{pmatrix} \phi_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.5}$$

Teorema 11. *Sea una EDLin de la forma* (2.5) *tal que* $z_{n+1} = Bz_n$ *tiene una solución nula estable. Entonces existe una norma vectorial tal que*

$$\|z_n\| \leq |\|z_0\| + \sum_{j=0}^{n-1} \|d_j\|$$

Demostración.

$$\begin{aligned} z_n &= Bz_{n-1} + d_{n-1} = B[Bz_{n-2} + d_{n-2}] + d_{n+1} = \\ &= B^2 z_{n-2} + Bd_{n-2} + d_{n-1} = \cdots \end{aligned}$$

Por inducción se puede establecer

$$z_n = B^n z_0 + \sum_{j=0}^{k} B^{n-j-1} d_j$$

Dado que la solución es estable entonces podemos aplicar los resultados ya estudiados:

1. $\rho(B) < 1 \quad \Rightarrow \|B\| \leq \rho(B) + \epsilon < 1$, si tomamos ϵ suficientemente pequeño.

2. $\rho(B) = 1$ y B de clase M. Existe una norma matricial subordinada tal que $\|B\| = \rho(B) = 1$. También existe una norma vectorial $\|\cdot\|$ cuya norma matricial subordinada satisface $\|B\| \leq 1$.

Por las propiedades de la norma matricial subordinada obtenemos

$$\begin{aligned} \|z_n\| \leq \|B\|^n \|z_0\| + \sum_{j=0}^{n-1} \|B\|^{n-j-1} \|d_j\| \leq \\ \|z_0\| + \sum_{j=0}^{n-1} \|d_j\| \end{aligned}$$

Teorema 12. *Sea EDLin de la forma* (2.4) *cuyo polinomio característico cumple la condición de la raíz. Entonces existe una constante* $c \geq 1$ *tal que toda solución* $\{y_n\}_{n\geq 0}$ *de la EDLin* (2.5) *cumple*

$$|y_n| \leq c \left[\max_{0\leq i\leq k-1} |y_i| + \sum_{j=0}^{n-k} |\phi_j| \right], \quad y_n \in \mathbb{R},\ n \geq 0$$

Demostración. De la condición de estabilidad de las soluciones de la EDLin de 1er orden asociada a (2.4) deducimos

$$\|z_n\| \leq \|z_0\| + \sum_{j=0}^{n-1} \|d_j\|, \quad n \geq 1 \tag{2.6}$$

Asimismo, por el teorema de equivalencia de normas $\exists c_1, c_2 \in \mathbb{R} \quad c_2 \geq c_1 > 0$ tales que

$$c_1\|z\|_\infty \leq \|z\| \leq c_2\|z\|_\infty$$

$$\|z_{n-k+1}\| \geq c\|z_{n-k+1}\|_\infty \geq c \cdot \max_{0\leq i\leq k-1} |y_i|, \quad \text{ya que} \quad z_{n-k+1} = \begin{pmatrix} y_n \\ \vdots \\ y_{n-k+1} \end{pmatrix}$$

$$\|z_0\| + \sum_{j=0}^{k} \|d_j\| \leq c_2\|z_0\|_\infty + c_2 \sum_{j=0}^{n-1} |\phi_j|$$

Si cambiamos n por $n-k+1$ en la ecuación anterior, entonces

$$c_1|y_n| \leq \|z_{n-k+1}\| \leq c_2[\max_{0\leq i\leq k-1} |y_i| + \sum_{j=0}^{n-k} |\phi_j|]$$

Tomamos $c = c_2/c_1$ y queda demostrado el teorema.

2.4. Estabilidad y Condiciones de Convergencia

Definición 2.19. Un método dado por $\{\rho(r), \phi_f\}$ decimos que es *cero estable* (sinónimo de estabilidad) si las soluciones de la EDLin homogénea cuyo polinomio característico es $\rho(r)$ son estables.

Comentario 2.20. Un método es cero estable $\iff$ $\rho(r)$ cumple la condición de la raíz.

Condiciones sobre ϕ_f

I) $\phi_f \equiv 0 \quad$ cuando $\quad f \equiv 0$

II) Debe existir C independiente de h tal que $\forall y_{n+i}, y^*_{n+i},\ i = 0, \ldots, k-1$ del dominio ϕ_f y $\forall x_n \in [a, b],\ \forall h \geq 0$

$$|\phi_f(y_{n+k} \ldots y_n, x_n; h) - \phi_f(y^*_{n+k} \ldots y^*_n, x_n; h)| \leq C \max_{0\leq i\leq k-1} |y_{n+i} - y^*_{n+i}| \tag{2.7}$$

Es decir, la función ϕ_f debe ser Lipschitz respecto a las condiciones iniciales.

Teorema 13 (acotación del EGD)**.** *Sea el PVI $y' = f(x, y)$ y un método dado por $\{\rho(r), \phi_f\}$. Sea $y_n(h)$ la solución numérica aproximada en $x_n = a + hn, \quad n = 0, \ldots, N = \frac{b-a}{h}$. Si el método es cero estable entonces existen c_1, c_2 independientes de h de forma que*

$$|y(x_n) - y_n(h)| \leq c_1 r(h) + c_2\tau(h)$$

donde $r(h) = \max_{0\leq i\leq k-1} |y(x_i) - y_i(h)|$ es el máximo de los errores iniciales.

Demostración.

$$\begin{aligned} y_{n+k}(h) + \ldots + \alpha_0 y_n(h) &= h\phi_f(y_{n+k}(h), \ldots, y_n(h), x; h) \\ y(x + kh) + \ldots + \alpha_0 y(x_n) &= h\phi_f(y(x_n + kh), \ldots, y(x_n), x_n; h) + \tau(x_n, h) \\ e_{n+j} &= y(x_n + jh) - y_{n+j}(h) \\ e_{n+k} + \ldots + \alpha_0 e_n = \varphi_n &= h[\phi_f(\text{sol. exacta}) - \phi_f(\text{sol. numérica}) + \tau(x_n, h)] \end{aligned}$$

Usando el teorema 12 para EDLin, es decir, que el polinomio característico de la EDLin del error global es un polinomio de un método cero estable, podemos deducir

$$\begin{aligned}\exists \beta \geq 1: \quad |e_{n+k}| &\leq \beta \left[\max_{0\leq i\leq k-1} |e_i| + \sum_{j=0}^{n} |\varphi_j| \right] \underset{(*)}{\leq} \\ &\leq \beta r(h) + \beta h \sum_{j=0}^{n} \left[C \max_{0\leq i\leq k} |e_{j+i}| + \tau(h) \right] \\ &\leq \beta r(h) + \beta h(n+1) \left[C \max_{0\leq i\leq n+k} |e_i| + \tau(h) \right]\end{aligned}$$

$(*)$ condición II de ϕ_f

Definimos $w_{n+k} = \max_{0\leq i\leq n+k} |e_i|$, y observamos que $\{w_n\}_{n\geq 0}$ es una sucesión creciente. Así, tenemos que

$$\max_{0\leq i\leq n+k} |e_i| = w_{n+k} \leq \beta r(h) + C\beta h(n+1) w_{n+k} + \beta h(n+1)\tau(h),$$

donde C es independiente de h y su cumple que $C\beta h(n+1) > 0$.
Ahora supongamos que para ciertas n y h se verifica

$$1 \geq 1 - \beta C h(n+1) \geq 1/2. \tag{2.8}$$

En tal caso $1 \leq [1 - \beta C h(n+1)]^{-1} \leq 2$. Entonces,

$$w_{n+k} \leq [1 - \beta C h(n+1)]^{-1} [\beta(n+1)h\tau(h)] \tag{2.9}$$

De acuerdo con las consideraciones anteriores escogemos h y n de forma que $h(n+1) \leq (2\beta C)^{-1} \equiv \delta$, que es independiente de h (puesto que β y C lo son). Así tendremos que, puesto que $\beta C h(n+1) \leq 1$, entonces es cierta nuetra suposición de la existencia de n y h para que se cumpla lo anterior.
Por otro lado, y basado en la misma desigualdad $h(n+1) \leq \delta$, también se cumple que $h < \frac{\delta}{n+1}$, así que:

$$\begin{aligned} w_{n+k} &\leq 2\beta r(h) + 2\beta\delta\tau(h) \leq \\ &\leq c_1 r(h) + c_2 \tau(h)\end{aligned}$$

con c_1, c_2 independientes de h. Esta desigualdad es precisamente la que queríamos demostrar.

Comentario 2.21. Notemos que en la demostración anterior debíamos asegurar que $h(n+1) \leq \delta$. Por tanto, la acotación para el EGD que hemos demostrado está probada para el problema $y' = f(x,y),\ y(a) = \eta,\ x \in [a,b]$ únicamente en el subintervalo $[a, a+\delta]$.
Si queremos extender la acotación a $[a,b]$ debemos proceder como sigue. Para empezar, tomamos los k últimos valores de la solución numérica en $[a, a+\delta]$. Seguidamente continuamos calculando la solución numérica, y aplicando el mismo procedimiento de hasta entonces obtenemos en $[a+\delta, a+2\delta]$ la siguiente acotación:

$$\left.\begin{aligned} w_{n+k} &\leq 2\beta[r_2(h) + \delta\tau(h)] \\ r_2(h) &\leq 2\beta[r(h) + \delta\tau(h)] \end{aligned}\right\} \quad \Longrightarrow \quad \begin{aligned} w_{n+k} &\leq 4\beta^2 r(h) + 2\delta(2\beta+1)\tau(h) \\ &\leq c_1' r(h) + c_2' \tau(h)\end{aligned}$$

Observamos que a medida que extendemos la acotación ésta puede volverse menos fina. A pesar de ésto, tiene la forma exigida por el teorema.

2.5. Teoría de Estabilidad Lineal

Consideremos el problema $y' = f(x, y)$. Si tomamos una aproximación de primer orden entonces podemos aproximar el sistema $y' = f(x, y)$ por un sistema $y' = Ay$, con $A \in M_m(\mathbb{R})$ y $y : \mathbb{R} \mapsto \mathbb{R}^m$.

Sea $\lambda \in \mathbb{C}$ un VAP de la matriz A, $Re(\lambda) < 0$ y consideremos el problema $y' = \lambda y$,que es un problema escalar.

Con este procedimiento conseguiremos establecer el comportamiento del método en función del paso que queramos utilizar de una forma bastante simple.

Por tanto, una vez obtenido el problema a analizar ($y' = \lambda y$) y sabiendo que la solución exacta es de la forma $y(x) = e^{\lambda y}$, aplicamos ahora un método lineal multipaso al problema y podremos establecer el comportamiento del mismo:

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

Escribimos la ecuación en diferencias lineal homogénea con $\bar{h} = \lambda h$:

$$\sum_{j=0}^{k} (\alpha_j - \bar{h}\beta_j) y_{n+j} = 0 \tag{2.10}$$

Notemos que para analizar las soluciones de la EDLin es necesario que consideremos el polinomio característico de la misma. De acuerdo con la ecuación (2.10) el polinomio es:

$$\pi(r; \bar{h}) = \rho(r) - \bar{h}\sigma(r) \tag{2.11}$$

Denotaremos como *polinomio de estabilidad absoluta del método* a $\pi(r; \bar{h})$.

Definición 2.22. El método es *absolutamente estable* para $\bar{h} \in \mathbb{C}$ si las soluciones de la EDLin (2.10) son asintóticamente estables.

Proposición 2.1. *Las soluciones de* (2.10) *son asintóticamente estables si y solo si escogido* $\bar{h}$ *entonces:*

$$\lim_{n \to \infty} ||y_n|| \to 0$$

Sea $R_A = \{\bar{h} \in \mathbb{C}/\text{MLM es absolutamente estable}\}$, la región de estabilidad absoluta del método. Podemos establecer una representación de R_A en $\mathbb{R}^2$ de la siguiente manera:

1. Eje $x \ \rightarrow \ Re(\bar{h})$
2. Eje $y \ \rightarrow \ Im(\bar{h})$

Podríamos también definir el conjunto como:

$$R_A = \{\bar{h} \in \mathbb{C} \mid \forall r_t \in \mathbb{C}, \ \pi(r_t, \bar{h}) = 0, \ |r_t| < 1\} \tag{2.12}$$

Ejemplo 2.23. Calcularemos los conjuntos R_A de los métodos:

- Método de Euler: $y_{n+1} = y_n + hf_n$
- Método de Euler implícito: $y_{n+1} = y_n + hf_{n+1}$

Sus polinomios de estabilidad absoluta son:

- $\pi_E(r;\bar{b}) = r - 1 - \bar{h} \to r_t = 1 + \bar{h} \to |r_t| < 1 \iff |1 + \bar{h}| < 1$. Por lo tanto nos define el interior de una circumferencia de radio 1 y centro $(1, 0)$.
- $\pi_{EI}(r;\bar{h}) = r - 1 - \bar{h}r \to r'_t = \frac{1}{1-h} \to |r'_t| < 1 \iff |1 - \bar{h}| > 1$. Nos define el exterior de una circumferencia de radio 1 y centro $(1, 0)$.

De esta manera podemos escoger el paso adecuado sin tener que probar directamente el método. Para finalizar, es necesario hacer un análisis para encontrar el margen de h. En el caso de Euler, considerando $\lambda \in \mathbb{R}$ con $\lambda < 0 \Rightarrow \bar{h} \in \mathbb{R}$.

Escogeríamos, por el resultado anterior, $-2 < \bar{h} < 0 \Rightarrow -2 < h\lambda < 0 \Rightarrow \frac{-2}{\lambda} > h > 0$.
En el caso complejo, $\lambda \in \mathbb{C} \Rightarrow 2 > |\lambda\bar{h}| > 0 \Rightarrow \frac{2}{|\lambda|} < |h| > 0$

Hagamos ahora el análisis de los MLM de un paso convergentes:

$$y_{n+1} = y_n + h(\beta f_{n+1} + (1 - \beta)f_n)$$

Calculamos ahora $\rho(r) = r - 1$. Por tanto, $\rho'(1) = \sigma(1)$.
Comprobamos la consistencia del método: $1 = \beta + (1 - \beta) = 1$.

El polinomio de estabilidad absoluta del método es

$$\pi_B(r, \bar{h}) = r - 1 - \bar{h}(\beta r + (1 - \beta)) = (1 - \bar{h}\beta)r + (-1 - (1 - \beta)\bar{h})$$

Calculando las raíces obtenemos:

$$r_t = \frac{1 + (1 - \beta)\bar{h}}{1 - \bar{h}\beta}$$

Para hacer un análisis más simple del método, analizamos la frontera de R_A, que denotaremos por ∂R_A.
∂R_A estará formada por los valores de $\bar{h}$ tales que $\pi(r, \bar{h})$ admite raíces r_t tales que $r_t = e^{i\phi}$, con $\phi \in [0, 2\pi]$.

Sabemos de antes que $\pi(r, \bar{h}) = p(r) - \bar{h}\sigma(r)$. Por lo tanto:

$$\pi(e^{i\phi}, \bar{h}) = 0 \iff \bar{h_f}(\phi) = \frac{p(e^{i\phi})}{\sigma(e^{i\phi})} \tag{2.13}$$

La ecuación (2.13) nos define ∂R_A como una curva uniparamétrica en $\mathbb{R}^2$.

Ejemplo 2.24. MLM de un paso convergente con $\beta = \frac{1}{2}$, alternativamente denominado como la regla del trapecio.

Tenemos $y_{n+1} = y_n + \frac{h}{2}(f_{n+1} + f_n)$, por lo tanto $\sigma(r) = \frac{1}{2}(1 + r)$. Obtenemos como curva $\bar{h_f}$:

$$\bar{h_f}(\phi) = \frac{e^{i\phi} - 1}{\frac{1}{2}(e^{i\phi} + 1)} = i(2\tan(\frac{\phi}{2})) \tag{2.14}$$

Por lo tanto, obtenemos:

$$\partial R_A = \{\bar{h} \in \mathbb{C} / Re(\bar{h}) = 0\} \tag{2.15}$$

Una vez obtenida la frontera, podemos obtener por continuidad los puntos $\bar{h} \in R_A$.

Notemos que, en general, que R_A puede ser el conjunto $\emptyset$, ya que su frontera no siempre es una curva continua ni conexa.

Definición 2.25. El *intervalo de estabilidad absoluta* $I_A \subset R_A$ es $I_A = \{\bar{h} \in R_A / Im(\bar{h}) = 0\}$.

Recordemos que para un MLM se satisfacía:

$$|y_n(h) - y(x_n)| \leq C_1 r(h) + C_2 \tau(h)$$

con C_1, C_2 constantes. Por lo tanto, las condiciones suficientes de convergencia son:

- Condición II sobre ϕ_f
- Consistencia
- Cero Estabilidad

Teorema 14. *Si un método es convergente entonces es cero estable si para $f \equiv 0$ se tiene $\phi_f = 0$ (condición I).*

Demostración. Sea el problema $y' = 0,\ y(a) = \nu$. Para el método dado por $\rho(r) = \sum_{j=0}^{k} \alpha_j r^j$ y ϕ_f cualquiera. Aplicando el método al problema se obtiene una EDLin homogénea:

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = 0 \tag{2.16}$$

Utilizaremos el argumento del contrarecíproco para demostrar el teorema. Supongamos pues que el método no es cero estable. Entonces existe λ una raíz de $\rho(r)$ tal que $|\lambda| > 1$ o bien $|\lambda| = 1$ y raíz múltiple. Toda solución numérica del problema tiene términos de la forma λ^n con $|\lambda| > 1$ o proporcional a n si $|\lambda| = 1$. En general pues, la solución general $\{y_n\}$ va a crecer arbitrariamente cuando $n \to \infty$. Por lo tanto, podemos asegurar que:

$$|y(x) - y_n| \underset{n\to\infty}{\to} \infty \tag{2.17}$$

Ésto se cumplirá aunque los valores inicial usados en (2.16) tiendan a ν. Por lo tanto, tenemos que el método no es convergente y queda demostrado por contrarecíproco.

Observacion 2.26. Es importante notar que para la demostración del teorema anterior es básica la condición de $\phi_f = 0$ ya que en caso contrario no podríamos obtener la fórmula (2.16).

Teorema 15. *Sea un método de la forma $\sum_{j=0}^{k} \alpha_j y_{n+j} = h\phi_f$ tal que cumpla las condiciones (2.4) y (2.7). Entonces, se cumple que:*

$$\textit{El método es convergente} \iff \textit{es consistente y cero-estable.}$$

Además, si el método es de orden p (i.e, $\tau(h) = O(h^p)$), suponiendo que el error de los valores iniciales sea $O(h^q)$ con $q \geq p$ entonces:

$$EDG = O(h^p)$$

Demostración. La demostración del teorema anterior se sigue de todos los teoremas trabajados en este capítulo.

Capítulo 3

Métodos Lineales Multipaso

3.1. Métodos Lineales Multipaso y Teorema de Dahlquist

Como ya se ha visto a lo largo del capítulo anterior, un método lineal de k pasos es

$$\sum_{i=0}^{k} \alpha_i y_{n+i} = h \sum_{i=0}^{k} \beta_i f_{n+i} \quad \text{con } \alpha_k = 1$$

Este método tiene asociados dos polinomios de gran importancia:

$$\rho(r) = \sum_{i=0}^{k} \alpha_i r^i, \quad \sigma(r) = \sum_{i=0}^{k} \beta_i r^i$$

Anteriormente se ha estudiado cuándo un método lineal multipaso es estable y de qué manera se puede conocer su orden. Como el orden de un método está totalmente relacionado con el número de pasos, es necesario preguntarse cuál puede ser el orden más grande posible de un método de este tipo bajo la condición de estabilidad.

En este apartado nos centraremos en la relación entre el paso h y el orden del método, en el caso particular de los método lineales multipaso. Para establecer el orden p hay que imponer que los coeficientes del desarrollo de Taylor de $h\tau(x,h)$ y $h\tau(h)$ se anulen, ya que son éstos los que nos indican el orden de convergencia:

$$C_i = 0, \quad i = 0 \dots p, \tag{3.1}$$

que da un total de $p+1$ condiciones. El teorema que presentaremos a continuación nos da una cota superior del orden a partir del número de pasos. Este teorema será muy importante para poder escoger el paso deseado en nuestro método:

Teorema 16 (de Dahlquist)**.** *Para un MLM de k pasos y orden p, si el método es cero estable entonces*

- *Si k es impar: $p \leq k+1$*
- *Si k es par: $p \leq k+2$*

Definición 3.1. Decimos que un método es *óptimo* cuando es cero estable de k pasos y orden $k+2$.

Teorema 17. *Un MLM de k pasos cero estable y de orden $k+2$ tiene un polinomio característico tal que todas sus raíces son de módulo 1 y simples.*

Ejemplo 3.2. Consideremos el método de Simpson definido por $y_{n+2}-y_n = \frac{h}{3}(f_{n+2}+4f_{n+1}+f_n)$, i.e, $\rho(r) = r^2 - 1$. Se puede comprobar que este método es óptimo.

3.1.1. Estabilidad Lineal

Consideremos la fórmula del polinomio de estabilidad:

$$\pi(r, \bar{h}) = \rho(r) - \bar{h}\sigma(r)$$

Si el método es convergente, se cumple que:

$$\begin{cases} \lim_{h\to 0} \pi(r, \bar{h}) = \rho(r) \\ r_t = 1 \quad \text{es raíz de } \rho(r) \end{cases}$$

Por lo tanto existirá $r_1(\bar{h})$ raíz de $\pi(r, \bar{h})$ tal que $\lim\limits_{h\to 0} r_1(\bar{h}) = 1$.

Teorema 18. *Para un MLM convergente de orden p se cumple:*

$$r_1(\bar{h}) = e^{\bar{h}} + O(\bar{h}^{p+1})$$

en el límite cuando $\bar{h} \to 0$.

Corolario. *Para todo MLM existe ϵ tal que el conjunto de puntos $\{(x, 0) \in \mathbb{R}^2 \quad | \quad x \in (0, \epsilon)\}$ no pertenecen nunca a la region de estabilidad.*

Demostración. (del teorema)

$$h\tau(x; h) = O(h^{p+1}) \text{ con } y' = f(x, y)$$

Para $y' = \lambda y \;\Rightarrow\; h\tau(x, h) = \sum_{j=0}^{k} \alpha_j y(x + jh) - h\sum_{j=0}^{k} \beta_j y'(x + jh)$. Con la condición inicial $y(0) = 1$ se tiene que $y(x) = e^{\lambda x}$. Entonces:

$$\sum_{j=0}^{k} \alpha_j e^{\lambda(x+jh)} - h\sum_{j=0}^{k} \beta_j \lambda e^{\lambda(x+jh)} = O(h^{p+1})$$

$$\underbrace{\sum_{j=0}^{k} \alpha_j (e^{\bar{h}})^j - \bar{h}\beta_j (e^{\bar{h}})^j}_{\pi(e^{\bar{h}}, \bar{h}) = \rho(e^{\bar{h}}) - \bar{h}\sigma(e^{\bar{h}})} = O(\bar{h}^{p+1})$$

Escogemos pues $\bar{h} < \frac{1}{\beta_k} \quad \Rightarrow \quad \bar{h}\beta_k \neq 1$ ya que el grado de π es k. Ahora podemos tomar el límite $h \to 0$, con lo que tenemos que $\pi(e^{\bar{h}}, \bar{h}) = O(h^{p+1})$ implica que $\pi(e^{\bar{h}}, \bar{h}) \to 0$ debido a que $e^{\bar{h}} - r_1(\bar{h}) \to 0$. En tal caso se cumple $r_i(\bar{h}) \neq 1$ para $i = 2, \ldots, k$. Ésto sucede al ser el método convergente, la unidad es raíz simple y $r_1(\bar{h}) \to 1$.

$$e^{\bar{h}} - r_1(\bar{h}) = O(\bar{h}^{p+1}) \quad \Rightarrow \quad r_1(\bar{h}) = e^{\bar{h}} + O(\bar{h}^{p+1})$$

3.2. Métodos Implícitos y Predictor-Corrector

Definición 3.3. Un par *predictor-corrector (P-C) de k pasos* está formado por dos métodos lineales multipaso, uno explícito llamado predictor (P) y otro implícito denominado corrector (C). La expresión general de un P-C es la siguiente:

$$\begin{cases} \sum_{j=0}^{k} \alpha_j^* y_{n+j} = h \sum_{j=0}^{k-1} \beta_j^* f_{n+j} & (P) \\ \sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f_{n+j} & (C) \end{cases} \tag{3.2}$$

donde k es siempre el número de pasos del P, representado por la primera de las igualdades. En la segunda, que representa el C, no se exige la condición $|\alpha_0| + |\beta_0| \neq 0$, pero se sigue asumiendo que $\alpha_k = \alpha_k^* = 1$. A continuación se describen los pasos a seguir y las fórmulas que se utilizan para calcular la solución numérica de un PVI mediante un método P-C como el (3.2) en los modos $P(EC)^\mu E^{1-t}$, con $t \in \{0, 1\}$, donde $t = 0$ indica que se efectua la evaluación final de f, mientras que $t = 1$ indica que no se efectua dicha evaluación final.

Partiendo de los valores iniciales $y_j,\ j = 0 \div k-1$, y para $n \geq 0$:

$$P: \qquad y_{n+k}^{[0]} = h \sum_{j=0}^{k-1} \beta_j^* f_{n+j}^{[\mu-t]} - \sum_{j=0}^{k-1} \alpha_j^* y_{n+j}^{[\mu]}$$

$$(EC)^\mu: \qquad f_{n+k}^{[\nu]} = f(x_{n+k}, y_{n+k}^{[\nu]})$$

$$y_{n+k}^{[\nu+1]} + \sum_{j=0}^{k-1} \alpha_j y_{n+j}^{[\mu]} = h\beta_k f_{n+k}^{[\nu]} + h \sum_{j=0}^{k-1} \beta_j f_{n+j}^{[\nu-k+j]}, \quad \nu = 0 \div \mu - 1$$

$$E^{1-t}: \qquad f_{n+k}^{[\nu]} = f(x_{n+k}, y_{n+k}^{[\mu]}), \quad \text{si } t = 0.$$

Definición 3.4. Un MLM implícito de k pasos se caracteriza por tener el coeficiente $\beta_k \neq 0$.

$$\sum_{i=0}^{k} \alpha_i y_{n+i} = h \sum_{i=0}^{k} \beta_i f_{n+i}$$

Cada paso requiere resolver una ecuación implícita:

$$y_{n+k} = h\beta_k f(x_{n+k}, y_{n+k}) + g_n$$

donde g_n es la solución numérica en x_{n+j} con $j = 0, \ldots, k-1$. Dado $y_{n+k}^{[0]}$, iteramos:

$$y_{n+k}^{[l+1]} = h\beta_k f(x_{n+k}, y_{n+k}^{[l]}) + g_n \quad \forall l = 0, 1, \ldots$$

$$||y_{n+k}^{[l+1]} - y_{n+k}^{[l]}|| \underset{l\to\infty}{\longrightarrow} 0 \qquad \forall\ y_{n+k}^{[0]} \ \text{ si } \ h|\beta_k|L < 1$$

donde L es la constante de Lipschitz de f. Entonces, para garantizar un buen comportamiento, debemos escoger un paso tal que:

$$h < \frac{1}{|\beta_k| L}$$

Si $L \gg 1$ necesitamos $1 \gg h$. Estos casos se dan en problemas mal condicionados (sistemas stiff).

En general, para acabar la iteración, fijamos ϵ para detener la iteraciones con la condición $||y_{n+k}^{[l+1]} - y_{n+k}^{[l]}|| < \epsilon$. Alternativamente, podemos fijar el número de iteracions a $\mu \geq 1$, en cada paso. El método resultante en este caso es un MLM explícito que se denomina Predictor-Corrector cuando $y_{n+k}^{[0]}$ se obtiende de un MLM explícito.

3.3. Criterio de Routh-Hurwitz

Este criterio puede resultar de gran utilidad cuando se analiza la estabilidad absoluta de un método, concepto que introduciremos a lo largo del capítulo.

Definición 3.5. Consideremos un polinomio de grado k, coeficientes reales y variable $z \in \mathbb{C}$

$$P(z) = a_0 z^k + a_1 z^{k-1} + \cdots + a_{k-1} z + a_k,$$

con $a_0 > 0$. La matriz $k \times k$ determinada por

$$H = \begin{bmatrix} a_1 & a_3 & a_5 & \cdots & a_{2k-1} \\ a_0 & a_2 & a_4 & \cdots & a_{2k-2} \\ 0 & a_1 & a_3 & \cdots & a_{2k-3} \\ 0 & a_0 & a_2 & \cdots & a_{2k-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_k \end{bmatrix}$$

tomando $a_j = 0$ cuando $j > k$, se denomina *matriz de Hurwitz* de $P(z)$.

Proposición 3.1. *Las raíces de $P(z)$ tienen parte real estrictamente negativa si y solo si todos los menores principales de la matriz de Hurwitz de $P(z)$ son estrictamente positivos.*

Se entiende por menor principal de orden j de una matriz $k \times k$ el determinante de la submatriz obtenida con los elementos comunes de las j primeras filas y las j primeras columnas de la matriz, siendo $1 \leq j \leq k$. En general, el criterio anterior se traduce en un sistema de k inecuaciones para los coeficientes de $P(z)$.

Definición 3.6. Un polinomio de grado k con coeficientes reales de dice que es un *polinomio de Schur* si todas sus raíces tienen módulo estrictamente inferior a al unidad, es decir, si todas sus raíces se encuentran en el interior de la circunferencia unidad del plano complejo.

Dados $r, z \in \mathbb{C}$ el cambio de variable

$$r = \frac{1+z}{1-z}$$

define una aplicación biyectiva entre $\{r \in \mathbb{C} \mid |r| < 1\}$ y $\{z \in \mathbb{C} \mid \mathtt{Re}(z) < 0\}$, subconjuntos de $\mathbb{C}$. Se deduce a partir de ella el siguiente criterio para que un polinomio sea de Schur:

Proposición 3.2. *Un polinomio $\pi(r)$ con coeficientes reales y de grado k es de Schur si y solo si el polinomio*

$$P(z) = (1-z)^k \pi\left(\frac{1+z}{1-z}\right) = a_0 z^k + a_1 z^{k-1} + \cdots + a_{k-1} z + a_k,$$

con $a_0 > 0$, cumple el criterio de Routh-Hurwitz[1].

[1]Si el polinomio no cumple que $a_0 > 0$, deberemos multiplicarlo por -1 para obtener un polinomio que se ajuste al exigido por el teorema.

3.4. Estimación del ELD para métodos de 1 paso

La idea es estimar el ELD de un método como la diferencia entre el valor que arroja el método y el valor que obtenemos del mismo método al dividir el paso por la mitad.

Consideremos el método $y_{n+1} = y_n + hf_n$. Sabemos que el ELD para éste vale:

$$ELD_{n+1} = h\tau(x_n, h) = \frac{y''(x_n)}{2}h^2 + O(h^3)$$

donde definimos el *término principal* como el primer coeficiento no nulo del desarrollo de Taylor del ELD, que denotaremos por PELD. En nuestro caso, el término principal es $PELD_{n+1} = \frac{y''(x_n)}{2}h^2$. Si tomamos como paso h entonces

$$\left.\begin{matrix} x_n \\ y_n \end{matrix}\right\} \quad \xrightarrow[h]{} \quad \begin{cases} x_{n+1} \\ y_{n+1}(h) \end{cases}$$

Alternativamente, podemos escoger como paso $\frac{h}{2}$, en cuyo caso el esquema nos queda así:

$$\left.\begin{matrix} x_{n+\frac{1}{2}} \\ y_{n+\frac{1}{2}} \end{matrix}\right\} \quad \xrightarrow[\frac{h}{2}]{} \quad \begin{cases} x_{n+1} \\ y_{n+1}\left(\frac{h}{2}\right) \end{cases}$$

En el caso $\frac{h}{2}$ tenemos:

$$\begin{aligned} y_{n+1}\left(\frac{h}{2}\right) &= y_{n+\frac{1}{2}} + \frac{h}{2}f(x_n + \frac{h}{2}, y_{n+\frac{1}{2}}) = y_{n+\frac{1}{2}} + \frac{h}{2}f_{n+\frac{1}{2}} \\ \Rightarrow y_{n+\frac{1}{2}} &= y_n + \frac{h}{2}f_n \\ y_{n+1}(\frac{h}{2}) &= y_n + \frac{h}{2}\left[f_n + f_{n+\frac{1}{2}}\right] \quad \longrightarrow \text{(equivalente al método Runge-Kutta)} \\ \Rightarrow y_{n+1} &= y_n + \frac{h}{2}(k_1 + k_2) \end{aligned}$$

Con respecto al error local (Euler de dos pasos con paso $\frac{h}{2}$)

$$\begin{aligned} ELD_{n+1} = h\tau^{[2]}(x_n, h) &= y(x_n, h) - y(x_n + h) = \\ &= y(x_n) - h[f(x_n, y(x_n)) + \underbrace{f(x_n + \frac{h}{2}, y(x_n) + \frac{h}{2}f(x_n, y(x_n)))}_{(*)}] \end{aligned}$$

(∗) Taylor $\quad f(x_n, y_n) = f_x\frac{h}{2} + f_y\frac{h}{2}f + O(h^2), \quad y''(x) = \frac{d}{dx}f = f_x + f_y f$

De acuerdo con las expresiones anteriores,

$$h\tau^{[2]}(x_n,h) = 2\left[\frac{y''(x_n)}{2}\right]\frac{h}{2}^2 + O(h^3)$$

$$\left.\begin{aligned} h\tau(x_n,h) &= y(x_{n+1}) - \tilde{y}_{n+1}(h) \\ h\tau^{[2]}(x_n,h) &= y(x_{n+1}) - \tilde{y}_{n+1}\left(\frac{h}{2}\right) \end{aligned}\right\} \Longrightarrow$$

$$\Longrightarrow \quad PELD_{n+1} = \frac{y''(x_n)}{2}h^2 = 2\left[\tilde{y}_{n+1}\left(\frac{h}{2}\right) - \tilde{y}_{n+1}(h)\right] + O(h^3)$$

Estimamos el $PELD_{n+1}$ usando $2[y_{n+1}\left(\frac{h}{2}\right) - y_{n+1}(h)]$.
Esquema de control del ELD de la solución numérica con el método de Euler:

$$\frac{2\|y_{n+1}(\frac{h}{2}) - y_{n+1}(h)\|}{\|y_{n+1}(h)\|} < \epsilon_r$$

ϵ_r es la tolerancia relativa al error. Si $\epsilon_r = 10^{-l}$, entonces los l primeros dígitos de la solución numérica son correctos.

Generalización a métodos de 1 paso de orden p

Supongamos que tenemos un método de 1 paso, ahora de orden p. En este caso, el ELD asociado a y_{n+1} con paso h es

$$ELD_{n+1}(h) = G(y(x_n))h^{p+1} + O(h^{p+2}) = y(x_{n+1}) - \tilde{y}_{n+1}(h)$$

Procediendo de manera análoga a la anterior, veamos qué ocurre si tomamos como paso $\frac{h}{2}$, considerando $y_{n+1}\left(\frac{h}{2}\right)$:

$$ELD_{n+1}\left(\frac{h}{2}\right) = 2G(y(x_n))\left(\frac{h}{2}\right)^{p+1} + O(h^{p+2}) = y(x_{n+1}) - \tilde{y}_{n+1}\left(\frac{h}{2}\right)$$

Si restamos las dos ecuaciones anteriores obtenemos:

$$\tilde{y}_{n+1}\left(\frac{h}{2}\right) - \tilde{y}_{n+1}(h) = G(y(x_n))h^{p+1}\left[1 - \frac{1}{2^p}\right] + O(h^{p+2}),$$

con lo que el PELD es

$$PELD_{n+1} = G(y(x_n))h^{p+1} = \left[1 - \frac{1}{2^p}\right]^{-1}\left[\tilde{y}_{n+1}\left(\frac{h}{2}\right) - \tilde{y}_{n+1}(h)\right] + O(h^{p+2})$$

Notamos que $[1 - \frac{1}{2^p}]^{-1} \to 1$ cuando $p \gg 1$. Si se cumple la condición del error escogida continuamos el proceso. En caso contrario redefinimos $h := \frac{h}{2}$ y observamos qué ocurre con ϵ_r.

Definición 3.7. Dado $y_{n+1}(h)$ obtenida con un método de orden p, consideramos

$$\tilde{y}_{n+1}(h) = y_{n+1}(h) + \left[1 - \frac{1}{2^p}\right]^{-1}[y_{n+1}\left(\frac{h}{2}\right) - y_{n+1}(h)].$$

Dicho método se conoce como *método de extrapolación (local)* asociado al método original.

3.5. Estimación del ELD con métodos Predictor-Corrector

Supongamos que estamos tratando con un método predictor-corrector de la forma (3.3), para el cual tenemos:

- Predictor: p^* orden, $C^*_{p^*+1}$ constante de error.
- Corrector: p orden, C_{p+1} constante de error.

Cuando se usa el par P-C en el modo $P(EC)^\mu E^{1-t}$ con $t \in \{0,1\}$ en realidad se está utilizando un método explícito. Es por ello que podemos analizar el correspondiente error local de truncación o discretización (ELD) de la forma habitual: examinaremos la diferencia $y(x_{n+k}) - \tilde{y}_{n+k}$, siendo $\tilde{y}_{n+k}$ la aproximación numérica en x_{n+k} obtenida con el método cuando se toman valores exactos,

$$y_{n+j} = y(x_{n+j}), \quad j = 0 \div k-1 \tag{3.3}$$

Para empezar, suponiendo $y(x)$ suficientemente diferenciable, se tiene que

$$h\tau^*(x,h) = C^*_{p^*+1} h^{p^*+1} y^{(p^*+1)}(x) + O(h^{p^*+2}) \tag{3.4}$$

$$h\tau(x,h) = C_{p+1} h^{p+1} y^{(p+1)}(x) + O(h^{p+2}) \tag{3.5}$$

Por otro lado, a partir de la definición de la función τ^* para el P, podemos escribir

$$y(x_{n+k}) + \sum_{j=0}^{k-1} \alpha^*_j y(x_{n+j}) = h \sum_{j=0}^{k-1} \beta^*_j f(x_{n+j}, y(x_{n+j})) + h\tau^*(x_n, h).$$

Si utilizamos (3.3) en la fórmula del P se obtiene

$$\tilde{y}^{[0]}_{n+k} + \sum_{j=0}^{k-1} \alpha^*_j y(x_{n+j}) = h \sum_{j=0}^{k-1} \beta^*_j f(x_{n+j}, y(x_{n+j})).$$

Restando las dos ecuaciones anteriores de deduce:

$$y(x_{n+k}) - y^{[0]}_{n+k} = C^*_{p^*+1} h^{p^*+1} y^{(p^*+1)}(x_n) + O(h^{p^*+2}) \tag{3.6}$$

Repetimos ahora el proceso para C. A partir de la definición de $\tau(x,h)$ se tiene que

$$y(x_{n+k}) + \sum_{j=0}^{k-1} \alpha_j y(x_{n+j}) = h \sum_{j=0}^{k-1} \beta_j f(x_{n+j}, y(x_{n+j})) + h\tau(x_n, h).$$

y utilizando (3.3) en la fórmula del C para cada iteración nos queda

$$\tilde{y}^{[\nu+1]}_{n+k} + \sum_{j=0}^{k-1} \alpha_j y(x_{n+j}) = h\beta_k f(x_{n+k}, \tilde{y}^{[\nu]}_{n+k}) + h \sum_{j=0}^{k-1} \beta_j f(x_{n+j}, y(x_{n+j})).$$

para $\nu = 0 \div \mu - 1$. Finalmente, restando las dos últimas ecuaciones se deduce para $0 \le \nu \le \mu - 1$

$$\begin{aligned} y(x_{n+k}) - \tilde{y}^{[\nu+1]}_{n+k} &= h\beta_k \left[f(x_{n+k}, y(x_{n+k})) - f(x_{n+k}, \tilde{y}^{[\nu]}_{n+k}) \right] + h\tau(x_n, h) = \\ &= h\beta_k \frac{\overline{\partial f}}{\partial y}(x_{n+k}, \xi_\nu)[y(x_{n+k}) - \tilde{y}^{[\nu]}_{n+k}] + C_{p+1} h^{p+1} y^{(p+1)}(x_n) + O(h^{p+2}) \end{aligned} \tag{3.7}$$

con $\nu = 0 \div \mu - 1$. En la segunda igualdad se ha aplicado el teorema del valor medio para funciones vectoriales, con ξ_v representado los distintos valores intermedios para cada componente de la derivada parcial.

En función de los órdenes del P y el C (los valores de p^* y p respectivamente) se obtienen los siguientes resultados:

1. Caso $p^* \geq p$:
   Sustituyendo (3.6) en la fórmula (3.7) para $\nu = 0$ se deduce
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[1]} = C_{p+1} h^{p+1} y^{(p+1)}(x_n) + O(h^{p+2}).$$
   A su vez, esta igualdad puede sustituirse de nuevo en (3.7) para $\nu = 1$ deduciéndose:
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[2]} = C_{p+1} h^{p+1} y^{(p+1)}(x_n) + O(h^{p+2}).$$
   Repitiendo el proceso el número necesario de veces se acaba demostrando que
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[\mu]} = C_{p+1} h^{p+1} y^{(p+1)}(x_n) + O(h^{p+2}).$$
   Es decir, para todo $\mu \geq 1$ el orden del método P-C es igual al orden del C y además el PELD (término principal del ELD) del P-C es idéntico al del C.

2. Caso $p^* = p - 1$:
   Siguiendo el mismo razonamiento que el presentado en el caso anterior, la primera sustitución de (3.6) en (3.7), con $\nu = 0$, da lugar a
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[1]} = \left[\beta_k \overline{\frac{\partial f}{\partial y}} C_p^* y^{(p)}(x_n) + C_{p+1} y^{(p+1)}(x_n) \right] h^{p+1} + O(h^{p+2}).$$
   Sin embargo, si $\mu \geq 2$ las posteriores sustituciones en (3.7), con $\nu \geq 1$, conducen siempre a la igualdad:
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[\mu]} = C_{p+1} h^{p+1} y^{(p+1)}(x_n) + O(h^{p+2}).$$
   Por tanto, para todo $\mu \geq 1$ el orden del método P-C es igual al del C, aunque si $\mu = 1$ entonces el PELD del P-C no coincide con el del C. En cambio, si $\mu \geq 2$ ambos PELD son idénticos.

3. Caso $p^* = p - 2$:
   En este caso, cuando sustituimos (3.6) en (3.7) obtenemos:
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[1]} = \beta_k \overline{\frac{\partial f}{\partial y}} C_{p-1}^* h^p y^{(p-1)}(x_n) + O(h^{p+1}).$$
   Notamos que ahora si $\mu = 1$ el orden del P-C es $p - 1$, inferior en una unidad al orden del C. Si $\mu \geq 2$, sustiyendo la ecuación anterior en (3.7) con $\nu = 1$ se deduce
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[2]} = \left[\left(\beta_k \overline{\frac{\partial f}{\partial y}} \right)^2 C_{p-1}^* y^{(p-1)}(x_n) + C_{p+1} y^{(p+1)}(x_n) \right] h^{p+1} + O(h^{p+2}),$$
   de modo que para $\mu = 2$ el orden del método P-C es igual al del C, pero los PELD de ambos métodos no coinciden. Sucesivas sustituciones en (3.7) conducen a la siguiente fórmula, válida únicamente cuando $\mu \geq 3$
   $$y(x_{n+k}) - \tilde{y}_{n+k}^{[\mu]} = C_{p+1} h^{p+1} y^{(p+1)}(x_n) + O(h^{p+2}).$$
   lo cual muestra que además del mismo orden, el P-C y el C tienen el mismo PELD.

Por tanto, queda claro que el orden y el PELD de un método P-C en el modo $P(EC)^{\mu}E^{1-t}$ dependen de la diferencia de los órdenes del P y del C, así como del valor μ (número de iteraciones) del modo. El siguiente teorema enuncia el resultado general.

Teorema 19. *En las condiciones anteriores, consideremos un método P-C definido por* (3.3) *y utilizado en el modo* $P(EC)^{\mu}E^{1-t}$.

1. *Si* $p^* \geq p$, *o bien si* $p^* < p$ *y* $\mu > p - p^*$, *entonces el método P-C y el corrector tienen el mismo orden y el mismo PELD.*
2. *Si* $p^* < p$ *y* $\mu = p - p^*$, *entonces el método P-C y el corrector tienen el mismo orden pero diferentes PELD.*
3. *Si* $p^* < p$ *y* $\mu < p - p^*$, *entonces el orden del P-C es igual a* $p^* + \mu$ *y siempre es estrictamente inferior al orden del corrector.*

Además, los modos $P(EC)^{\mu}E$ *y* $P(EC)^{\mu}$ *tienen el mismo orden y el mismo PELD.*

Demostración. Empezamos considerando la función $h\tau(x_n, h)$, es decir:

$$h\tau^*(x_n,h) = y(x_{n+k}) - \tilde{y}^{[0]}_{n+k} = \bar{C}_{p^*+1}y^{p^*+1}(x_n)h^{p^*+1} + O(h^{p^*+2})$$
$$h\tau(x_n,h) = y(x_{n+k}) - \tilde{y}^{[\mu]}_{n+k} = C_{p+1}y^{p+1}(x_n)h^{p+1} + O(h^{p+2})$$

donde p^* es el orden del predictor y p el del corrector. Si tomamos la diferencia de las dos expresiones anteriores tenemos:

$$\tilde{y}^{[\mu]}_{n+k} - \tilde{y}^{[0]}_{n+k} = \bar{C}_{p^*+1}y^{p^*+1}(x_n)h^{p^*+1} - C_{p+1}y^{p+1}(x_n)h^{p+1} + O(h^{p+2}) \tag{3.8}$$

Consideremos los posibles casos para valores de p y p^*:

- Si $p^* < p$ entonces

  $$C_{p+1}y^{p+1}(x_n)h^{p+1} = \tilde{y}^{[0]}_{n+k} - \tilde{y}^{[\mu]}_{n+k} + O(h^{p^*+1})$$

  Observamos que en este caso el resultado carece de sentido, así que no lo consideramos.

- Si $p^* > p$ entonces (3.8) como $p^* + 1 \geq p + 2$ de manera que:

  $$C_{p+1}y^{p+1}(x_n)h^{p+1} = \tilde{y}^{[0]}_{n+k} - \tilde{y}^{[\mu]}_{n+k} + O(h^{p+1})$$

  Para el Predictor-Corrector, el resultado final será:

  $$PELD_{n+k} = \tilde{y}^{[0]}_{n+k} - \tilde{y}^{[\mu]}_{n+k}$$

- Si $p = p^*$ entonces

  $$C_{p+1}y^{p+1}(x_n)h^{p+1} = \underbrace{\frac{C_{p+1}}{\bar{C}_{p+1} - C_{p+1}}}_{\text{Constante de Milne}} (\tilde{y}^{[\mu]}_{n+k} - \tilde{y}^{[0]}_{n+k}) + O(h^{p+2})$$

Comentario 3.8 (Modos con extrapolación local)**.** Dado un P-C en modo $P(EC)^{\mu}E^{1-t}$ se construye un nuevo método como solución de:

$$\tilde{y}^{[\mu]}_{n+k} = y^{[\mu]}_{n+k} + \frac{C_{p+1}}{\bar{C}_{p+1} - C_{p+1}}(y^{[\mu]}_{n+k} - y^{[0]}_{n+k}).$$

A partir de $y^{[\mu]}_{n+k}$ en modo $P(EC)^{\mu}E^{1-t}$ se pueden construir nuevos modos P-C del tipo $P(ECL)^{\mu}E^{1-t}$ o del tipo $P(EC)^{\mu}LE^{1-t}$.

3.6. Estabilidad lineal para métodos P-C

Ejemplo 3.9. Consideremos para fijar ideas el método:

$$y_{n+2} - y_{n+1} = \frac{h}{2}(f_{n+2} + f_{n+1}) \to C$$
$$y_{n+2} - y_{n+1} = \frac{h}{2}(3f_{n+1} - f_n) \to P$$

- Modo $P(EC)E$: Si en el ejemplo tomamos $y' = \lambda y$, entonces

$$y_{n+2} - y_{n+1} = \frac{h}{2}(\lambda y_{n+2}^{[0]} + \lambda y_{n+1})$$
$$y_{n+2} - y_{n+1} = \frac{h}{2}(y_{n+1} + \frac{3h}{2}\lambda y_{n+1} - \lambda y_n) + \lambda\frac{h}{2}y_{n+1}$$
$$y_{n+2} - y_{n+1} - \bar{h}(\frac{y_{n+1}}{2} + \frac{3}{4}\bar{h}y_{n+1} - \frac{\bar{h}}{2}y_n) - \frac{\bar{h}}{2}y_{n+1} = 0$$

La forma de los polinomios de estabilidad absoluta de los métodos P-C en este modo es:

$$\pi(r, \bar{h}) = \rho(r) - \bar{h}\sigma(r) + M_\mu(H)(\rho^*(r) - \bar{h}\sigma^*(r))$$
$$\begin{cases} p, \sigma \to C \\ p^*, \sigma^* \to P \end{cases}$$
$$M_\mu(H) = \frac{H^\mu(1-H)}{1-H^\mu}, \quad \text{donde} \begin{cases} H = \bar{h}\beta_k \\ \sigma = \beta_k r^k + \beta_{k-1}r^{k-1} + \cdots + \beta_0 \end{cases}$$

- Modo $P(EC)^\mu$: En este caso, la forma de los polinomios de estabilidad es

$$\pi(r, \bar{h}) = \beta_k r^k(\rho(r) - \bar{h}\sigma(r)) + M_\mu(H)(\rho^*(r)\sigma(r) - \sigma^*(r)\rho(r))$$

Por lo tanto, si P y C son consistentes entonces $\lim_{h\to 0} \pi(r, \bar{h}) = \rho(r)$. En tal caso existe $r_1(\bar{h})$ raíz de $\pi(r, \bar{h})$ tal que $\lim_{h\to 0} r_1(\bar{h}) = 1$. Finalmente, se establece

$$\pi(e^{\bar{h}}, \bar{h}) = O(h^{p+1})$$

donde p es el orden del método P-C. Luego, se deduce (utilizando el mismo razonamiento que para MLM) que$r_1(\bar{h}) = e^{\bar{h}} + O(h^{p+1})$ cuando $h \to 0$. A modo de resumen, para los diferentes modos los polinomios de estabilidad son

$$P(EC)^\mu E: \qquad \pi(r, \bar{h}) = \rho(r) - \bar{h}\sigma(r) + M_\mu(H)[\rho^*(r) - \bar{h}\sigma^*(r)] \tag{3.9}$$
$$P(EC)^\mu: \qquad \pi(r, \bar{h}) = \beta_k r^k[\rho(r) - \bar{h}\sigma(r)] + M_\mu(H)[\rho^*(r)\sigma(r) - \rho(r)\sigma^*(r)] \tag{3.10}$$

donde

$$r_1(\bar{h}) \underset{h\to 0}{\to} 1, \quad M_\mu(H) = \frac{H^\mu(1-H)}{1-H^\mu} = O(h^\mu), \quad H = \beta_k\bar{h}$$

Por otro lado,

$$\pi(e^{\bar{h}}, \bar{h}) = O(\bar{h}^{p+1}) \quad \Longrightarrow \quad r_1(\bar{h}) = e^{\bar{h}} + O(h^{p+1})$$

Si p es el orden del C y p^* es el orden del P con P y C convergentes entonces, para los modos $P(EC)^\mu E$ (se deduce de la ecuación (3.9)) tenemos

$$\pi(e^{\bar{h}}, \bar{h}) = O(h^{p+1}), \quad \text{si } p^* + \mu \geq p,$$

Con respecto a los modos $P(EC)^\mu$, tenemos que para la ecuación (3.10)

$$\left.\begin{array}{r} \rho^* - \bar{h}\sigma^* = O(h^{p^*+1}) \;\Rightarrow\; \rho^*\sigma - \bar{h}\sigma\sigma^* = \sigma O(h^{p^*+1}) = O(h^{p^*+1}) \\ \rho - \bar{h}\sigma = O(h^{p+1}) \;\Rightarrow\; \rho\sigma^* - \bar{h}\sigma\sigma^* = \sigma^* O(h^{p+1}) = O(h^{p+1}) \end{array}\right\} \Rightarrow$$
$$\Rightarrow \quad \rho^*\sigma - \rho\sigma^* = O(h^{p+1})$$

Dado que $h \to 0$ entonces $e^{\bar{h}} \to 1$ y además:

$$\sigma(1) = \rho'(1) \neq 0$$
$$\sigma^*(1) = (\rho^*)'(1) \neq 0$$

Por tanto, teniendo en cuenta que los métodos son convergentes entonces, para los modos $P(EC)^\mu$:

$$\pi(e^{\bar{h}}, \bar{h}) = O(h^{p+1})$$

3.7. Construcción de MLM por Interpolación

Consideremos un método de la forma

$$y_{n+1} - y_n = \int_{x_n}^{x_{n+1}} I(x)dx$$

donde $\{y_n\}$ es la solución numérica en $x_n = a + hn$ e $I(x)$ es el polinomio interpolador de $f(x, y(x))$. Si tomamos el caso:

$$I(x_{n-j}) = f_{n-j} \quad j = 0, 1 \ldots k-1$$

se obtiene

$$y_{n+1} - y_n = h \sum_{j=0}^{k-1} \beta_j^* f_{n+1-k+j}$$

y observamos que se trata de un método expícito de k pasos. Esta familia de métodos se conoce como métodos de Adams-Bashford.
Por otro lado, podemos considerar el caso

$$I(x_{n+1-j}) = f_{n+1-j} \quad j = 0, 1 \ldots k$$

del que se extrae la familia de métodos

$$y_{n+1} - y_n = h \sum_{j=0}^{k} \beta_j f_{n+1-k}$$

Cabe notar que se trata de un conjunto de métodos implícitos de k pasos, conocidos como métodos de Adams-Moulton.

La fórmula de interpolación de Newton nos servirá para explicitar la forma de $I(x)$. A continuación la presentamos.

$$I(x) = P^*_{k-1}(r) = \sum_{i=0}^{k-1} (-1)^i \binom{-r}{i} \nabla^i f_{n+j}$$

donde

$$r = \frac{x - x_n}{h}, \qquad \begin{aligned} \nabla f_n &= f_n - f_{n-1} \\ \nabla^2 f_n &= \nabla(\nabla f_n) = f_n - 2f_{n-1} + f_{n-2} \end{aligned}$$

Definimos ahora un operador denotado por E para facilitar la notación.

$$\begin{aligned} Ef_n &= f_{n+1} \\ E^{-1} f_n &= f_{n-1} \end{aligned}$$

de modo que podemos reescribir ∇ como $\nabla = 1 - E^{-1}$. Modificando la notación anterior por la introducida con E nos queda:

$$\nabla^i = (1 - E^{-1})^i = \sum_{l=0}^{i} \binom{i}{l} E^{-l}$$

$$\nabla^i f_n = \sum_{l=0}^{i} \binom{i}{l} f_{n-l}$$

donde $\binom{-r}{i} = \frac{1}{i!} \prod_{l=0}^{i-1} (-r - l)$. En resumen, los métodos anteriormente descritos quedan determinados así:

- Adams-Bashfort: $I(x) = P^*_{k-1}(r)$ donde $r = \frac{x - x_k}{h}$.
- Adams-Moulton: $I(x) = P_k(r) = \sum_{i=0}^{k} (-1)^i \binom{-r}{i} \nabla^i f_{n+1}$ con $r = \frac{x - x_{n+1}}{h}$.

Notemos el uso intensivo del funcional E en la nueva formulación de los métodos.

3.7.1. Métodos Adams-Bashfort de k pasos

Tenemos para este método la fórmula:

$$y_{n+1} - y_n = \int_{x_n}^{x_{n+1}} P^*_{k-1}(r) \mathrm{d}x = h \int_0^1 P^*_{k-1}(r) \mathrm{d}r$$

$$\int_0^1 P_{k-1}(r) h \mathrm{d}r = h \sum_{j=0}^{k-1} \gamma^*_i \nabla^i f_n, \qquad \gamma^*_i = (-1)^i \int_0^1 \binom{-r}{i} \mathrm{d}r$$

Construimos una función generatriz denotada por $G^*(t)$:

$$G^*(t) = \sum_{i=0}^{\infty} \gamma^*_i t^i = \sum_{i=0}^{\infty} \left[(-1)^i \int_0^1 \binom{-r}{i} \mathrm{d}r \right] t^i = \int_0^1 \left(\sum_{i=0}^{\infty} \binom{-r}{i} (-t)^i \right) \mathrm{d}r =$$

$$= \int_0^1 (1-t)^{-r} \mathrm{d}r = \left. \frac{-(1-t)^{-r}}{\ln(1-t)} \right|_0^1 = \frac{-t}{(1-t)\ln(1-t)}$$

Así, tenemos que el funcional definido satisface la siguiente relación:

$$G^*(t)\frac{\ln(1-t)}{-t} = (1-t)^{-1}$$

Si la desarrollamos en serie de potencias obtenemos:

$$(\gamma_0^* + \gamma_1^* t + \ldots)(1 + \frac{t}{2} + \ldots) = 1 + t + t^2 + \ldots$$

que nos aporta los las ecuaciones que deben satisfacer los coeficientes γ_i^*:

$$\gamma_i^* + \frac{\gamma_{i-1}^*}{2} + \ldots + \frac{\gamma_0^*}{i+1} = 1 \qquad i = 0, 1 \ldots k-1$$

Observamos que puesto que para un método A-B de k pasos únicamente se requieren los valores γ_i^* para $i = 0, 1 \ldots k-1$, entonces las ecuaciones anteriores lo determinan unívocamente.

3.7.2. Métodos Adams-Moulton de k pasos

Tenemos en el planteo del método:

$$y_{n+1} - y_n = \int_{-1}^{0} P_k(r) h \mathrm{d}r = h \sum_{i=0}^{k} \gamma_i \nabla^i f_{n+1}$$
$$\gamma_i = \int_{-1}^{0} (-1)^i \binom{-r}{i} \mathrm{d}r$$

De nuevo construimos una función generatriz, ahora denotada por $G(t)$:

$$G(t) = \sum_{i=0}^{\infty} \gamma_i t^i = \frac{-t}{\ln(1-t)} \quad \Rightarrow \quad G(t)\frac{\ln(1-t)}{-t} = 1$$

En este caso la relación que satisfacen las γ, una vez hecho el correspondiente desarrollo en serie de potencias es:

$$\gamma_0 = 1, \ \frac{\gamma_0}{2} + \gamma_1 = 0 \ \ldots$$
$$\gamma_i + \frac{\gamma_{i-1}}{2} + \ldots + \frac{\gamma_0}{i+1} = 0, \quad i = 0 \div k$$

El método A-M de k pasos requiere γ_i hasta $i = 0 \div k$. Una vez incorporadas y resueltas las ecuaciones de las γ_i obtenemos para el método A-M de k pasos:

$$y_{n+1} - y_n = h \sum_{i=0}^{k} \gamma_i \nabla^i f_{n+1}$$

Capítulo 4

Métodos Runge-Kutta

En este capítulo analizaremos los métodos Runge-Kutta en su formulación general, trataremos las condiciones de orden (según la estrategia de Butcher), y por último trabajaremos sobre el análogo de la estabilidad lineal introducido en el tema anterior.

4.1. Formulación del método

Definición 4.1. Un *método Runge-Kutta (RK)* es un método de un paso dado por la ecuación:

$$y_{n+1} - y_n = h \sum_{i=1}^{s} b_i k_i$$

donde

$$k_i = f(x_n + c_i h, y_i), \qquad y_i = y_n + h \sum_{j=1}^{s} a_{ij} k_j$$

A estas definiciones se les impone una condición aparentemente arbitraria aunque estrechamente relacionada con el orden del método, conocida como *condición suma-fila*:

$$\sum_{j=1}^{s} a_{ij} = c_i, \quad i = 1 \div s$$

Definición 4.2. Diremos que s es el *número de niveles del método.*

Es habitual denotar los métodos RK mediante notación matricial. Para ello introducimos:

Definición 4.3. La *matrix de Butcher* es

$$\left(\begin{array}{c|c} c & A \\ \hline 0 & b^T \end{array}\right)$$

donde:

$$A = (a_{ij}) \qquad c = [c_1 \dots c_s]^T \qquad b = [b_1 \dots b_s]^T$$

En el caso en que A sea triangular inferior estricta (con $a_{ii} = 0,\ i = 1 \div s$), entonces el método RK es explícito. Entonces tendrá la estructura:

$$\begin{aligned} k_1 &= f(x_n, y_n) = f_n \\ k_2 &= f(x_n + c_2 h, y_n + h c_2 k_1) \\ k_3 &= f(x_n + c_3 h, y_n + h(c_3 - a_{32}) k_1 + h a_{32} k_2) \\ &\vdots \end{aligned}$$

Ejemplo 4.4. De acuerdo con las definiciones anteriores, el método RK de un nivel es

$$y_{n+1} - y_n = h b_1 f_n$$

que solo converge en el caso $b_1 = 1$, definiendo así, como habrá apreciado el hábil lector, el ya conocido método de Euler.

4.2. Convergencia de métodos RK

Veamos ahora que el método cumple:

- Condición (2.7) sobre ϕ_f. Como tenemos:

$$\phi_f = \sum_{i=1}^{s} b_i f(x_n + c_i h, y_i)$$

  Si f cumple la condición de Lipschitz, entonces ϕ_f cumple la condición (2.7).

- Cero estabilidad. Por ser $\rho(r) = r - 1$, los métodos siempre son cero estables.

- Segunda condición de consistencia (*Teorema 2*):

$$\phi_f(y(x), x, 0) \stackrel{?}{=} \rho'(1) f(x, y(x)) = f(x, y(x))$$

  En nuestro método tenemos:

$$\phi_f(y(x), x, 0) = \sum_{j=1}^{s} b_i f(x, y(x)).$$

  Por tanto, la condición de consistencia para los método RK es $\sum_{i=1}^{s} b_i = 1$, que nos da también la convergencia junto con las otras condiciones verificadas.

4.2.1. Error local de discretización

Consideremos el ELD asociado a métodos RK. De apartados anteriores conocemos que éste viene dado por:

$$\begin{aligned} h\tau(x, h) &= y(x + h) - y(x) - h\phi_f(y(x), x; h) \\ h\tau(x, h) &= y'(x)h + O(h^2) - h\phi_f(y(x), x; 0) + O(h^2) \end{aligned}$$

donde $y'(x) = f(x, y(x))$ y $h\phi_f(y(x), x; 0) = h \sum_i b_i f(x, y(x))$. Si el RK es convergente, al ser consistente en general podremos afirmar que $\tau(x, h) = O(h)$, tiene orden 1 al menos.

4.2.2. Métodos implícitos y semi-implícito

Aquí introduciremos un tipo particular de métodos implícitos.

Definición 4.5. Si A es matriz triangular inferior no estricta[1], el método RK se denomina *semi-implícito.*

Definición 4.6. Decimos que el método es *implícito* cuando no es ni semi-implícito ni explícito. $k_1, \ldots, k_s$ se obtienen resolviendo un sistema de s ecuaciones acopladas e implícitas.

4.2.3. Región de estabilidad absoluta

Proposición 4.1. *La formula general de la función de estabilidad absoluta de los métodos R-K es:*

$$y_{n+1} = R(\hat{h})y_n$$
$$\mathcal{R}_A = \{\hat{h} \in \mathbb{C}| \quad |R(\hat{h})| < 1\}$$

Analizemos ahora la forma general de $R(\widehat{h})$:

$$y_{n+1} = y_n + h\sum_{i=1}^{s} b_i k_i$$

En el caso $\quad y' = \lambda y, \quad \Rightarrow y_{n+1} = y_n + \widehat{h}\sum_{i=1}^{s} b_i Y_i$, donde

$$Y_i = y_n + h\sum_{j=1}^{s} a_{ij}(\lambda Y_j) = y_n + \widehat{h}\sum_{j=1}^{s} a_{ij} Y_j$$
$$Y = [Y_1, \ldots, Y_s]^T \quad e = [1, \ldots, 1]^T \in \mathbb{R}^s$$

Podemos reformular pues las condiciones anteriores de forma vectorial con el uso de el Y, la matriz A, y el vector e:

$$Y = y_n e + \widehat{h}AY$$
$$(I_s - \widehat{h}A)Y = y_n e$$
$$Y = [(I_s - \widehat{h}A)^{-1}e]y_n$$

Condición básica para que ésto suceda es evidentemente que $det(I_s - \widehat{h}A) \neq 0$. Si el método Rk es explícito, se comprueba que $det(I_s - \widehat{h}A) = 1 \quad \forall \widehat{h} \in \mathbb{C}$.

$$Y_{n+1} = Y_n + \widehat{h}b^T(I_s - \widehat{h}A)^{-1}ey_n = (1 + \widehat{h}b^T(I_s - \widehat{h}A)^{-1}e)y_n$$

Además $R(\widehat{h})$ es una función polinómica de grado $\leq s$.

[1]Definimos matriz triangular inferior no estricta como matriz triangular inferior cuya diagonal no es idénticamente cero

Ejercicio. Calculemos ahora los método de orden s de s niveles.

En el caso $s = 2$: $\widehat{h}b^T(I_s - \widehat{h}A)^{-1} =$ RK de dos niveles explícto. $R(\widehat{h}) = 1 + \widehat{h} + \frac{\widehat{h}^2}{2}$.

En el caso $s = 3$: $R(\widehat{h}) = R(\widehat{h}) = 1 + \widehat{h} + \frac{\widehat{h}^2}{2} + \frac{\widehat{h}^3}{6}$.

Veamos ahora por qué R coincide con la forma del desarrollo de la exponencial hasta orden p. El ELD $y(x_{n+1}) - y_{n+1} = O(h^{p+1})$, habiendo obtenido $y_n = y(x_n)$.

$$y(x_{n+1}) = y_n + (\lambda h)y_n + \ldots + \frac{(\lambda h)^p}{p!}y_n + O(h^{p+1})$$

$$y_{n+1} = (1 + \widehat{h} + \frac{\widehat{h}^2}{2} + \ldots + \frac{\widehat{h}^p}{p!})y_n + O(h^{p+1})$$

$$y_{n+1} = R(\widehat{h})y_n \rightarrow \text{para cualquier RK aplicado a} \quad y' = \lambda y$$

Con las condiciones anteriores y la definición de $R(\widehat{h})$, se tiene que $R(\widehat{h})$ no depende de y_n sino de la h y del método:

En el caso $p = s \Rightarrow R(h) = 1 + \widehat{h} + \ldots + \frac{\widehat{h}^p}{p!}$.

Ejercicio. Obtener los intervalos de estabilidad absoluta para RK explícito de orden igual al del numero de niveles y ver como se obtiene la región de estabilidad de un RK.

4.3. Introducción al estudio del orden de los métodos RK

Para estudiar el orden hay que deteminar el desarrollo de Taylor de la función $\tau(x, h)$, que en el caso de los métodos Runge-Kutta (RK) adopta la siguiente expresión:

$$h\tau(x,h) = y(x,h) - y(x) - h\phi_f(y(x), x, h) = y(x+h) - y(x) - h\sum_{i=1}^{s} b_i k_i,$$

siendo

$$k_i = f(x + c_i h, y_i), \qquad y_i = y_n + h\sum_{j=1}^{s} a_{ij}k_j, \qquad c_i = \sum_{j=1}^{s} a_{ij}, \quad i = 1 \div s$$

para un RK de s niveles y un PVI dado por un sistema $y' = f(x, y)$.

Recordemos que si consideramos métodos lineales, las derivadas de la función ϕ_f se pueden expresar en términos de las derivadas de $y(x)$. Ésto no es posible para los métodos RK, en los que las derivadas de ϕ_f se obtienen a partir de las derivadas de $f(x, y)$. Por ello nos vemos forzados a expresar también las derivadas de $y(x)$ en términos de las derivadas de f, complicando considerablemente el estudio de las condiciones de orden.

4.3.1. Orden de un RK explícito de 3 niveles para un PVI escalar

Consideremos el caso general de un PVI escalar, cuya EDO es $y' = f(x, y)$, con $y(x) \in \mathbb{R}$, y un método RK explícito de 3 niveles:

$$\begin{aligned} y_{n+1} &= y_n + h(b_1k_1 + b_2k_2 + b_3k_3) \\ k_1 &= f(x_n, y_n) \\ k_2 &= f(x_n + hc_2, y_n + hc_2k_1) \\ k_3 &= f(x_n + hc_3, y_n + h(c_3 - a_{32})k_1 + ha_{32}k_2) \end{aligned}$$

Desarrollamos en serie de Taylor hasta el término h^4 de:

$$h\tau(x,h) = y(x,h) - y(x) - h(b_1k_1 + b_2k_2 + b_3k_3)$$

Para las derivadas de f se adopta la siguiente notación (todas se evaluan en el mismo punto $(x, y(x))$ y por ello no se especifica):

$$f := f(x,y), \qquad f_x := \frac{\partial f}{\partial x}, \qquad f_{xy} := \frac{\partial^2 f}{\partial x \partial y}, \qquad f_{xx} := \frac{\partial^2 f}{\partial x^2}, \qquad \text{etc.}$$

El desarrollo de Taylor de $y(x+h)$ (en torno a x) hasta $O(h^4)$ es:

$$y(x+h) = y(x) + hy^{(1)}(x) + \frac{1}{2}h^2y^{(2)}(x) + \frac{1}{6}h^3y^{(3)}(x) + O(h^4)$$

donde $y^{(i)}$ es la derivada de orden i respecto a la variable x. Usando la regla de la cadena se obtinenen las siguientes expresiones:

$$\begin{aligned} y^{(1)}(x) &= y' = f \\ y^{(2)}(x) &= f_x + f_y y' = f_x + f_y f \equiv F \\ y^{(3)}(x) &= f_{xx} + 2ff_{xy} + f^2 f_{yy} + f_y(f_x + ff_y) \equiv G + f_y F \end{aligned}$$

Por otro lado, el desarrollo de $\phi_f = b_1k_1 + b_2k_2 + b_3k_3$ hasta orden h^3 requiere hacer el correspondiente desarrollo de k_2 y k_3 entorno a x (observad que $k_1 = f$ no requiere desarrollo). Si reunimos los obtenidos para k_2 y k_3 y el calculado para $y(x+h)$ nos queda:

$$\begin{aligned} h\tau(x,h) = h(1 - b_1 - b_2 - b_3)f + h^2\left(\frac{1}{2} - b_2c_2 - b_3c_3\right)F \\ + \left[\left(\frac{1}{6} - b_3c_2a_{32}\right)\right]f_yF + \left(\frac{1}{6} - \frac{b_2c_2^2 + b_3c_3^2}{2}G\right] + O(h^4) \end{aligned}$$

Se deducen los siguientes resultados sobre el orden de un RK explícito de 3 niveles:

$$\begin{aligned} b_1 + b_2 + b_3 = 1 \quad &\Rightarrow \quad \text{orden 1 al menos,} \\ b_2c_2 + b_3c_3 = \frac{1}{2} \quad &\text{y la anterior} \quad \Rightarrow \quad \text{orden 2 al menos,} \\ b_2c_2^2 + b_3c_3^2 = \frac{1}{3} \quad & \\ b_3c_2a_{32} = \frac{1}{6} \quad &\text{y las anteriores} \quad \Rightarrow \quad \text{orden 3 al menos.} \end{aligned}$$

De hecho, examinando el término de orden h^4 se puede comprobar que si se cumplen las condiciones anteriores, el orden es exactamente 3 y no puede ser mayor.

4.3.2. Algunos resultados generales

- Para un método RK explícito de orden p y s niveles se puede demostrar que $p \leq s$. Solo existen métodos RK explícitos tales que $p = s$ cuando $s \leq 4$.
- Para construir un RK explícito de orden 5 se requieren 6 niveles como mínimo. Si es de orden 7 se requieren 9 niveles y si es de orden 8 se requieren 11 niveles como mínimo.

- El máximo orden que puede alcanzarse con un método RK de s niveles es igual a $2s$ cuando el método es implícito.
- Existen métdodos RK que se comportan con orden diferente según se apliquen a un problema escalar (autónomo o no) o a un sistema de EDOs. Considerar los siguientes enunciados aplicados a un método RK cualquiera:

  Enunciado A: El método tiene orden p para $y' = f(y),\ y(x) \in \mathbb{R}^m$, $m > 1$ (sistema de EDOs, caso general).

  Enunciado B: El método tiene orden p para $y' = f(y),\ y(x) \in \mathbb{R}$ (problema escalar, caso general).

  Enunciado C: El método tiene orden p para $y' = f(y),\ y(x) \in \mathbb{R}$ (problema escalar autónomo).

  Se conocen los siguientes resultados:

  $$\begin{aligned}
  &\text{si } 1 \le p \le 3, \quad A \Leftrightarrow B \Leftrightarrow C,\\
  &\text{si } p = 4, \quad A \Leftrightarrow B \Rightarrow C, \quad \text{pero} \quad C \not\Rightarrow B,\\
  &\text{si } p \ge 5, \quad A \Rightarrow B \Rightarrow C, \quad \text{pero} \quad C \not\Rightarrow B \quad \text{y} \quad B \not\Rightarrow A.
  \end{aligned}$$

4.4. Herramientas para el estudio del orden de los métodos RK

Para estudiar a fondo el orden de los métodos RK en su caso más general (es decir, aplicados a sistemas de ecuaciones diferenciales), tenemos que introducir nuevas herramientas que nos ayudarán a comprender el problema. Estas herramientas son:

- la M-ésima derivada de Frechet,
- los diferenciales elementales, y
- los árboles con raíz.

Es en esta sección donde las introduciremos y estudiaremos con cierto nivel de detalle.

Sin pérdida de generalidad, un sistema de EDOs siempre puede suponerse autónomo. Así pues, consideramos el caso general de un PVI

$$y' = f(y), \quad y(a) = \eta, \quad y, f \in \mathbb{R}^m,\ m > 1$$

Las primeras derivadas de $y(x)$ en el caso escalar autónomo son:

$$y' = f, \qquad y^{(2)} = f_y f, \qquad y^{(3)} = f_{yy} f^2 + f_y^2 f.$$

Para ver que estas expresiones pueden generalizarse de forma inmediata al caso de un sistema de dimensión $m > 1$ es suficiente considerar las mismas derivadas cuando $m = 2$, es decir, cuando

$$y = [{}^1y, {}^2y]^T, \quad f = [{}^1f, {}^2f]^T$$

Introduciendo la notación ${}^{\text{componente}}f_{\text{derivada}}$:

$${}^i f_j := \frac{\partial({}^i f)}{\partial({}^j y)}, \quad {}^i f_{jk} := \frac{\partial^2({}^i f)}{\partial({}^j y)({}^k y)}, \quad \text{etc}$$

a partir de diferenciar $y' = f$ componente a componente se obtiene:

$$y^{(2)} = \begin{bmatrix} {}^1y^{(2)} \\ {}^2y^{(2)} \end{bmatrix} = \begin{pmatrix} {}^1f_1 & {}^1f_2 \\ {}^2f_1 & {}^2f_2 \end{pmatrix} f = (Df)f$$

y tras cierta manipulación se puede escribir

$$y^{(3)} = \begin{bmatrix} {}^1y^{(3)} \\ {}^2y^{(3)} \end{bmatrix} = \begin{bmatrix} f^T D^2({}^1f) f \\ f^T D^2({}^2f) f \end{bmatrix} + (Df)^2 f$$

siendo $D^2({}^if)$ la matriz de derivadas parciales de segundo orden de la función ${}^if$.

4.4.1. Caso $y' = f(y)$

En este caso consideramos el problema $y' = f(y)$ con $y : \mathbb{R} \to \mathbb{R}$. Buscamos los valores de F y G.

$$F_x = f_y f G_x = f_{yy} f^2$$

y observamos que no se cumple $f_y F_x \neq G$. En el caso $y' = f(x, y(x))$ cuando calculamos $y^{(4)}(x)$ nos aparecen dos términos:

$$(f_{xx} + f f_{yy})(f_x + f f_y) \quad \text{(condición (1) de orden)} \tag{4.1}$$

$$f_y(f_{xx} + 2f f_{xy} + f^2 f_{yy}) \quad \text{(condició (2) de orden)} \tag{4.2}$$

Notemos que si f no depende de x, entonces los términos (4.1) y (4.2) se convierten en:

$$(4.1) \quad f_y f_{yy} f^2$$

$$(4.2) \quad f_y f_{yy} f^2$$

Notemos que los dos términos quedan exactamente iguales. Por ese motivo, las dos condiciones colapsan en una sola condición que denominaremos ((4.1) + (4.2)).

4.4.2. La M-ésima derivada de Frechet

Sean $z,\ f(z) \in \mathbb{R}^m$. La derivada M-ésima de Frechet, que se denota $f^{(M)}(z)$, es un operador lineal cuyo dominio es el producto cartesiano $\mathbb{R}^m \times \cdots \times \mathbb{R}^m$ (M veces) y que viene dado por la expresión siguiente:

$$f^{(M)}(z)(K_1 \dots K_M) = \sum_{i=1}^{m} \left[\sum_{j_1=1}^{m} \cdots \sum_{j_M=1}^{m} {}^if_{j_1 \dots j_M} {}^{j_1}K_1 \dots {}^{j_M}K_M \right] \mathbf{e_i}$$

donde

- $\mathbf{e_i},\ i = 1 \div m$ son los vectores de la base canónica de $\mathbb{R}^m$. De esta forma, la M-ésima derivada de Frechet evaluada en el argumento z constituye un vector de m componentes.
- La expresión de la M-ésima derivada de Frechet involucra todas las posibles derivadas parciales de orden M de la función f respecto a las componentes de su argumento z:

$${}^if_{j_1 \cdots j_M} = \frac{\partial^M({}^if(z))}{\partial({}^{j_1}z) \cdots \partial({}^{j_M}z)}$$

- $K_i,\ i = 1 \dots M$ son los operandos de la derivada de Frechet. Hay M operandos si la derivada es de orden M. Cada uno de ellos representa una función vectorial de m componentes, $K_i = [{}^iK_i \dots {}^mK_i]^T$. En particular, los operandos pueden ser a su vez derivadas de Frechet.

4.4.3. Derivadas de Frechet de primer y segundo orden

A modo de ejemplo, consideremos el caso bidimensional, $m = 2$. La derivada de Frechet de orden 1 se obtiene tomando $M = 1$ en la expresión general

$$f^{(1)}(z)(K_1) = \sum_{i=1}^{2} \left[\sum_{j_1=1}^{2} {}^{i}f_{j_i}{}^{j_1}K_1 \right] \mathbf{e_i} = \begin{bmatrix} {}^{1}f_1{}^{1}K_1 + {}^{1}f_2{}^{2}K_1 \\ {}^{2}f_1{}^{1}K_1 + {}^{2}f_2{}^{2}K_1 \end{bmatrix}$$

Reemplazando el argumento z por $y = y(x)$ y el operando K_1 por $f(y)$, siendo (para el caso $m = 2$) $y, f(y) \in \mathbb{R}^2$ la expresión anterior adopta la forma:

$$f^{(1)}(y)(f(y)) = \begin{bmatrix} {}^{1}f_1{}^{1}f + {}^{1}f_2{}^{2}f \\ {}^{2}f_1{}^{1}f + {}^{2}f_2{}^{2}f \end{bmatrix} = D(f)f = y^{(2)}$$

Para simplificar la notación se suprime el argumento y, de forma que la derivada segunda de y se puede expresar como una derivada de Frechet de primer orden:

$$y^{(2)} = f^{(1)}(f).$$

También se puede expresar $y^{(3)}$ utilizando derivadas de Frechet hasta orden 2 como máximo. Si de nuevo consideramos $m = 2$ y adoptamos $M = 2$ en la fórmula general, se obtiene:

$$f^{(2)}(z)(K_1, K_2) = \begin{bmatrix} {}^{1}f_{11}{}^{1}K_1{}^{1}K_2 + {}^{1}f_{12}{}^{1}K_1{}^{2}K_2 + {}^{1}f_{21}{}^{2}K_1{}^{1}K_2 + {}^{1}f_{22}{}^{2}K_1{}^{2}K_2 \\ {}^{2}f_{11}{}^{1}K_1{}^{1}K_2 + {}^{2}f_{12}{}^{1}K_1{}^{2}K_2 + {}^{2}f_{21}{}^{2}K_1{}^{1}K_2 + {}^{2}f_{22}{}^{2}K_1{}^{2}K_2 \end{bmatrix}$$

Reemplazamos z por y, tomando $K_1 = K_2 = f$ y usando ${}^{i}f_{12} = {}^{i}f_{21}$ se obtiene:

$$f^{(2)}(f, f) := f^{(2)}(y)(f(y), f(y)) = \begin{bmatrix} {}^{1}f_{11}({}^{1}f_1)^2 + 2{}^{1}f_{12}({}^{1}f)({}^{2}f) + {}^{1}f_{22}({}^{2}f)^2 \\ {}^{2}f_{11}({}^{1}f_1)^2 + 2{}^{2}f_{12}({}^{1}f)({}^{2}f) + {}^{2}f_{22}({}^{2}f)^2 \end{bmatrix} = \begin{bmatrix} f^T D^2({}^{1}f)f \\ f^T D^2({}^{2}f)f \end{bmatrix}$$

que es el primer término de la derivada tercera de y que se obtuvo al principio de la sección 2. El segundo término de $y^{(3)}$ se puede comprobar que es también una derivada de Frechet, en este caso de orden 1. Concretamente, si se toma $K_1 = f^{(1)}(f)$ como operando de la derivada de Frechet de orden 1, se deduce:

$$f^{(1)}(f^{(1)}(f)) = (Df)^2 f$$

Podemos escribir entonces para la derivada tercera:

$$y^{(3)} = f^{(2)}(f, f) + f^{(1)}(f^{(1)}(f)),$$

que es posible generalizar para cualquier dimensión m, y en particular, para el caso escalar $m = 1$ se recupera la expresión habitual de la derivada tercera.

Comentario 4.7 (sobre la derivación de una derivada de Frechet)**.** Al derivar respecto a x, $y'(x) = f(x)$, una derivada de Frechet de orden M con argumento $y = y(x)$ cuyos M operandos son derivadas de Frechet de orden menor a M, se obtiene una combinación lineal de derivadas de Frechet de orden $\leq M + 1$ con argumentos y y operandos f o derivadas de Frechet de orden $\leq M$.

$$\frac{\mathrm{d}}{\mathrm{d}x}[{}^{i}f_{j_1 \ldots j_M}, {}^{j_1}K_1 \ldots {}^{j_M}K_M]$$
$$\frac{\mathrm{d}}{\mathrm{d}x}{}^{i}f_{j_1 \ldots j_M} = f^{i}f_{j_1 \ldots j_M, j_{M+1}}$$

4.4.4. Diferenciales elementales

Los resultados de la sección anterior se generalizan del siguiente modo: $y^{(p)}$ se puede escribir (en cualquier dimensión) como una combinación lineal de derivadas de Frechet de orden menor o igual a $p-1$. Cada uno de los término de dicha combinación lineal es lo que se conoce como un *diferencial elemental* de f.

Los diferenciales elementales son por tanto las unidades básicas para representar las derivadas de cualquier orden de y, y finalmente, poder estudiar el orden de los métodos RK.

Definición 4.8. Los *diferenciales elementales* de la función $f \in \mathbb{R}^m$ son las funciones $F_s : \mathbb{R}^m \to \mathbb{R}^m$ determinadas recursivamente del modo siguiente:

(I) f es el único diferencial elemental de orden 1.

(II) Si $F_s,\ s = 1 \div M$, son diferenciales elementales de orden r_s, respectivamente, entonces la derivada de Frechet de orden M

$$f^{(M)}(F_1 \dots F_M)$$

es un diferencial elemental de orden $1 + \sum_{s=1}^{M} r_s$. No es necesario que los F_s sean todos distintos, ni tampoco sus órdenes. La notación que adoptaremos para los diferenciales elementales es:

$$\{F_1 F_2 \dots F_M\} := f^{(M)}(F_1 \dots F_M)$$

donde se sobreentiende que las derivadas involucradas son las de $f(y)$ respecto a las componentes de $y \in \mathbb{R}^m$.

De acuerdo con la definición anterior, para construir los diferenciales elementales de orden $p > 1$, se tienen que utilizar derivadas de Frechet de orden $M < p$ y los M operandos han de ser diferenciales elementales de orden estrictamente inferior a p pero cuyos órdenes sumen $p-1$. De la construcción de los diferenciales elementales hasta orden 4 obtenemos los siguientes resultados:

Orden 1. Existe un único diferencial elemental que por definición es f.

Orden 2. Solo se puede tomar la derivada de Frechet de primer orden, $M = 1$, de un diferencial elemental de orden 1. Por tanto, existe un único diferencial de orden 2, que es $f^{(1)}(f) = \{f\}$.

Orden 3 Hay dos opciones: o bien tomar $M = 2$ y dos difereciales elementales de orden 1 (que por tanto deben ser f), o bien tomar $M = 1$ y un diferencial elemental de orden 2 (que deberá ser $\{f\}$.
Con la primera opción obtenemos $\{f, f\} = f^{(2)}(f, f)$, que denotamos como $\{f^2\}$.
Con la segunda opción se obtiene $\{\{f\}\} = f^{(1)}(f^{(1)}(f))$, que denotamos como $\{{}_2f\}_2$.

Orden 4. Existen 4 diferenciales elementales:

$M = 3$**.** El único operando posible es f y se obtiene un único diferencial elemental, $\{f^3\}$.

$M = 2$**.** En este caso solo pueden usarse los operandos f y $\{f\}$, y el único diferencial elemental que se obtiene es $\{f\{f\}\} = \{\{f\}f\}$.

$M = 1$. Cada uno de los diferenciales elementales de orden 3 se puede usar como argumento, de modo que hay dos posibles resultados: $\{{}_2f^2\}_2$ y $\{{}_3f\}_3$.

De acuerdo con los resultados anteriores se comprueba que podemos escribir las primeras derivadas de y tal como indica la siguiente tabla:

i	$y^{(i)}$
1	f
2	$\{f\}$
3	$\{f^2\} + \{{}_2f\}_2$
4	$\{f^3\} + 3\{f\{f\}\} + \{{}_2f^2\}_3 + \{{}_3f\}_3$

Observacion 4.9. f diferencial elemental de orden $1 \Rightarrow \quad f^{(1)}(f)$ diferencial elemental de orden $2 \Rightarrow f^{(1)}(f^{(1)}(f))$ diferencial elemental de orden 3.

Comentario 4.10. Para el caso de orden 4 y $M = 1$, entonces:

$$f^{(1)}(f^{(2)}(f,f)) \;\rightarrow\; \{\{f,f\}\} \;\rightarrow\; \{{}_2f^2\}_2$$
$$\{{}_3f\}_3 \equiv f^{(1)}(f^{(1)}(f^{(1)}(f)))$$

En el caso escalar $y' = f(y)$ tenemos un término de $y^{(1)}(x)$, $f_y f_y f_y f$.

En resumen, las derivadas de y se construyen utilizando los diferenciales elementales, pero además es necesario conocer:

1. Los coeficientes numéricos de los diferenciales elementales cuya combinación lineal permite calcular con exactitud la expresión de las derivadas de y.

2. El número total de diferenciales elementales que forman la combinación lineal que se identifica con la derivada de $y^{(p)}$.

Estas dos cuestiones se pueden resolver utilizando resultados de combinatoria y teoría de grafos que se describen en la siguiente sección.

4.4.5. Árboles con raíz

El objetivo de esta sección es introducir las herramientas necesarias para representar los diferenciales elementales y las derivadas $y^{(p)}$, ya que nos permitirá formular las condiciones de orden de los métodos RK. Comenzaremos con algunas nociones generales de teoría de grafos.

Definición 4.11.

1. Sea V un conjunto finito. El par (V, A) se denomina *grafo* cuando $A \subset \mathcal{P}_2(V)$, siendo $\mathcal{P}_2(V)$ el conjunto de subconjuntos de cardinal 2 de V. Cuando $A \subset V \times V$ entonces el par (V, A) se denomina *grafo dirigido* o *digrafo*. En ambos casos los elementos de V se denominan *vértices* y los de A *aristas*.

2. Si (V, A) es un digrafo, su *grafo subyacente* es el grafo "sin dirigir". Formalmente, será el par $(\bar{V}, \bar{A})$ tal que $\bar{V} = V$ y [2]

$$\forall\, u, v \in V, \quad \{u, v\} \in \bar{A} \;\Leftrightarrow\; (u, v) \in A \text{ o } (v, u) \in A.$$

[2]Notemos que en el caso de grafos dirigidos, la arista la escribiremos como una pareja *ordenada* de vértices, es decir, (a,b). En cambio, en los grafos no dirigidos, la arista es una pareja *no ordenada*. Por tanto, la escribiremos como {a, b}.

3. Sea un grafo (V, A). Una sucesión de vértices diferentes $v_1, \ldots, v_k$ con $k \geq 2$, tal que $\{v_i, v_{i+1}\} \in A$ para $1 \leq i \leq k-1$ se conoce como un *camino* entre v_1 y v_k. Se dice que el grafo (V, A) es *conexo* si entre cada par de vértices distintos de V existe un camino.

Definición 4.12. Un digrafo (V, E) es un *árbol con raíz* cuando cumple las siguientes condiciones:

(1) Su grafo subyacente es conexo.

(2) Los pares ordenados de E tienen componentes distintas: $(a, b) \in E \ \rightarrow \ a \neq b$.

(3) Existe un único elemento de V, denominado *raíz*, que nunca aparece en la segunda componente de los pares ordenados de E.

(4) Excepto la raíz, todos los elementos de V aparecen exactamente una vez entre las segundas componentes de los pares ordenados de E.

Definición 4.13.

- Un grafo (V, A) es *acíclico* si cumple las condiciones 3 y 4 anteriores. No posee caminos cerrados o ciclos.
- El número de vértices de un árbol con raíz (y de un grafo en general) se denomina *orden*. Dado el par (V, E) el *orden del árbol con raíz* es el cardinal del conjunto de vértices de V.

Comentario 4.14 (Representación gráfica)**.** Los vértices se representan como puntos y las aristas como segmentos de recta uniendo los vértices que las definen. Dada una arista (a, b) el vértice a se representa por debajo del b: las aristas apuntan hacia arriba. La raíz es el vértice que figura en la parte inferior del diagrama.

Definición 4.15. Si (V, E) y (V', E') son dos árboles con raíz, se dice que son *isomorfos* cuando existe una aplicación biyectiva $\phi : V \rightarrow V'$ tal queda

$$\forall\, x, y \in V, \quad (x, y) \in E \ \Leftrightarrow \ (\phi(x), \phi(y)) \in E'$$

Todos los árboles con raíz isomorfos se representan con el mismo diagrama sin etiquetar los vértices. En realidad, cuando no se usan etiquetas para los vértices el diagrama es una representación de toda una clase de equivalencia de árboles con raíz definida por la relación de isomorfía.

Comentario 4.16. A menudo hablaremos de "árbol" para referirnos a un árbol con raíz (a pesar de que en la teoría de grafos son dos objetos distintos). Además, consideraremos iguales todos los árboles isomorfos entre sí.

Es posible construir los distintos árboles de un orden fijado de forma recursiva utilizando la siguiente notación:

- Para orden 1, hay un único árbol con raíz de un vértice (árbol trivial), que denotamos τ.
- Dado un árbol t, consideramos todos los vértices x de t tales que (v, x) es una arista de t, siendo v la raíz de t. En tal caso se dice que x es un *hijo* de la raíz. Si suprimimos v del conjunto de vértices de t, y todas las aristas que tengan como segunda componente a

vértices hijos de la raíz, lo que queda es un conjunto de ároles $t_1, t_2 \dots t_m$ (no necesariamente distintos), que no están conectados entre sí por ningún camino, y cuyas raíces son los vértices hijos de v en el árbol t. Utilizaremos la notación

$$t = [t_1, t_2 \dots t_m]$$

para representar el árbol t. Notemos que el orden de los t_i en la representación anterior es irrelevante, de modo que si hay árboles t_i repetidos, se agrupan y se escribe un exponente para indicar el número de repeticiones. Por ejemplo:

$$[t_1 t_3 t_2 t_1 t_2 t_2] := [t_1^2 t_2^3 t_3]$$

Mediante la notación anterior se obtienen los siguientes árboles con raíz *distintos* de cualquier orden:

Orden 1. τ.

Orden 2. $[\tau]$.

Orden 3. $[\tau^2]$, $[[\tau]] := [{}_2\tau]_2$.

Orden 4. $[\tau^3]$, $[\tau[\tau]]$, $[[\tau^2]] := [{}_2\tau^2]_2$, $[{}_3\tau]_3$.

Notemos que el número total de árboles de un orden dado coincide con el número de diferenciales elementales del mismo orden que se obtuvo en la sección anterior. Observemos también que si $r(t_i)$ son los órdenes de los árboles t_i, $i = 1 \dots m$, entonces el orden de $t = [t_1 \dots t_m]$ es

$$r(t) = 1 + \sum_{i=1}^{m} r(t_i),$$

lo mismo que sucede con la construcción de los diferenciales elementales que se describió en la sección anterior. El siguiente teorema generaliza estas dos observaciones:

Teorema 20. *Para $p \geq 1$, el número de árboles con raíz distintos de orden p coincide con el número de diferenciales elementales de orden p de f. Existe una aplicación biyectiva F entre el conjunto de árboles distintos T y el conjunto de los distintos diferenciales elementales de f. Esta aplicación se define del siguiente modo:*

1. *$F(\tau) = f$. Al árbol trivial τ le corresponde el único diferencial elemental de orden 1, f.*
2. *Si a los árboles con raíz t_s de orden $r(t_s)$, $s = 1 \div M$, les corresponden los diferenciales elementales $F(t_s)$ de orden r_s, $s = 1 \dots M$, entonces al árbol con raíz $[t_1 \dots t_M]$ de orden $1 + \sum_{s=1}^{M} r(t_s)$ le corresponde el diferencial elemental de orden $1 + \sum_{s=1}^{M} r_s$.*

$$F([t_1 \dots t_M]) = \{F(t_1) \dots F(t_M)\},$$

Ejemplo 4.17. Consideramos los árboles $\tau, [\tau], [\tau], [\tau^2]$ y construimos el árbol:

$$t = [\tau[\tau]^2[\tau^2]]$$

El orden, por la proposición enunciada anteriormente, es $1+(3+2+2+1) = 9$. La representació del mismo árbol en notación de derivada de Frechet es $\{f, \{f^2\}, \{f, f\}\}$. En el caso escalar:

$$F(t) = f_{yyyy} f (f_y f)^2 f_{yy} f$$

que es uno de los términos de $f^{9)}$ (ya que tiene orden 9). En este caso, $F(t)$ es uno de los términos de la derivada de orden 9, los otros se construirian a base de hacer todos los casos posibles de árboles de orden 9.

4.4.6. Funciones definidas sobre árboles con raíz

Volviendo a las dos cuestiones que quedaron planteadas al final de la sección 4.3, el teorema anterior ha establecido que el número de diferenciales elementales de orden n coincide con el número de árboles con raíz distintos de orden n. El siguiente teorema resuelve cómo contar ambas cantidades:

Teorema 21. *Sean a_n el número de árboles con raíz distintos de orden n. Es decir, el número de diferentes clases de equivalencia bajo la relación de isomorfía de árboles con raíz de n vértices. Entonces, las cantidades $a_1, a_2 \dots$ satisfacen la siguiente identidad término a término:*

$$a_1 + a_2 x + a_3 x^2 + \cdots = (1-x)^{-a_1}(1-x^2)^{-a_2}(1-x^3)^{-a_3}\cdots .$$

Por identidad término a término se entiende que tras sustituir cada factor del miembro derecho de la igualdad por su serie de Taylor, el coeficiente que resulte para la potencia x^k debe coincidir con a_{k+1}.
Cabe recalcar que el coeficiente de la potencia en el miembro derecho de la igualdad solo depende de $a_1 \dots a_k$. Por tanto, la identidad anterior establece una relación recurrente para a_{k+1}, el cual, para $k > 0$, se obtiene de los valores anteriores a_i, $i = 1 \div k$.

Aplicando el teorema anterior se obtienen los valores ya conocidos $a_1 = 1$, $a_2 = 1$, $a_3 = 2$, $a_4 = 4$. El número de diferenciales elementales crece significativamente con el orden, por ejemplo, $a_5 = 9$, y $a_8 = 115$. Esto da una idea de lo complicado que puede llegar a ser "diseñar" métodos RK de orden elevado.

Por si fuera poco, todavía necesitamos saber los coeficientes numéricos de las combinaciones de difereciales elementales que nos permiten obtener $y^{(p)}$. Para resolver esta cuestión se introducen las siguientes funciones definidas sobre un árbol con raíz y que pueden calcularse de forma recursiva.

Definición 4.18. La *densidad* de un árbol t, $\gamma(t)$, es el producto del orden de todos los subárboles de t:

$$\prod_{v \in V} r_v = \gamma(t)$$

Donde entendemos que el subárbol v del árbol t es un la pareja $(\hat{V}, \hat{E})$ tal que:

- $\hat{V} \subset V, \qquad \hat{E} \subset E.$
- $(\hat{V}, \hat{E})$ es un árbol con raíz.
- $v \in \hat{V} \subset V$ es la raíz de $(\hat{V}, \hat{E})$.

El orden del subárbol v lo hemos denotado por r_v.

Definición 4.19. Consideremos $t = (V, E)$. Diremos que $\phi : V \to V$ es un *automorfismo de t* si cumple

$$\forall v, v' \in V, \quad (v, v') \in E \quad \Leftrightarrow \quad [\phi(v), \phi(v')] \in E$$

Definición 4.20. Decimos que el *grupo de simetrías de t* es:

$$A(t) = \{\phi : V \to V \mid \phi \quad \text{automorfismo de } t\}$$

El orden de $A(t)$, $\sigma(t)$, se conoce como *simetría de t*.

Definición 4.21. Sea un árbol con raíz $t = [t_1^{m_1} t_2^{m_2} \dots t_k^{m_k}]$ donde $t_1, t_2 \dots t_k$ son árboles con raíz distintos 2 a 2. Entonces el *orden* $r(t)$, la *simetría* $\sigma(t)$, y la *densidad* $\gamma(t)$ del árbol t se definen de forma recursiva de acuerdo con las siguientes relaciones:

$$r(t) = 1 + \sum_{i=1}^{k} m_i r(t_i),$$

$$\sigma(t) = \prod_{i=1}^{k} m_i! \left(\sigma(t_i)\right)^{m_i},$$

$$\gamma(t) = r(t) \prod_{i=1}^{k} \left(\gamma(t_i)\right)^{m_i}$$

Y para el caso del árbol trivial $t = \tau$,

$$r(\tau) = \sigma(\tau) = \gamma(\tau) = 1.$$

Finalmente, se define la cantidad $\alpha(t)$ para un árbol t cualquiera como

$$\alpha(t) = \frac{r(t)!}{\sigma(t)\gamma(t)}$$

Comentario 4.22. Observemos que $\alpha(t)$ es el número de formas que tenemos de etiquetar los vértices de t con $\{1 \dots r(t)\}$ de forma que:

1. A cada vértice le asigno una y solo una etiqueta.
2. Etiquetados equivalentes por automorfía cuentan como uno solo.
3. Si (i, j) es una arista tras etiquetar, entonces $i < j$.

Estas definiciones nos permiten enunciar el siguiente teorema:

Teorema 22. *Sea $y = y(x)$, $y' = f(y)$, y $f : \mathcal{R} \to \mathcal{R}$. Entonces,*

$$y^{(q)} = \sum_{r(t)=q} \alpha(t) F(t)$$

donde $F(t)$ es la aplicación biyectiva entre árboles con raíz y diferenciales elementales que se introdujo anteriormente.

La siguiente table resume toda la información que se ha ido acumulando hasta ahora y que permite obtenter las derivadas de $y(x)$ hasta orden 4.

Con estos resultados se puede obtener el desarrollo en serie de Taylor de $y(x+h)$. No obstante, para obtener las condiciones de orden de un método RK se necesita el desarrollo de la función $\tau(x, h)$ al completo. Por tanto, hay que obtener el desarrollo de Taylor de la parte restante de $\tau(x, h)$ que no se ha considerado hasta ahora. Esta parte es:

$$\tilde{y}(h) := y(x) + h \sum_{i=1}^{k} b_i k_i, \quad k_i = f\left(y(x) + h \sum_{j=1}^{s} a_{ij} k_j\right), \quad \sum_{j=1}^{s} a_{ij} = c_i, \quad i = 1 \dots s$$

Orden	Arbol (t)	$F(t)$	$r(t)$	$\sigma(t)$	$\gamma(t)$	$\alpha(t)$
1	τ	f	1	1	1	1
2	$[\tau]$	$\{f\}$	2	1	2	1
3	$[\tau^2]$	$\{f^2\}$	3	2	3	1
3	$[{}_2\tau]_2$	$\{{}_2f\}_2$	3	1	6	1
4	$[\tau^3]$	$\{f^3\}$	4	6	4	1
4	$[\tau[\tau]]$	$\{f\{f\}\}$	4	1	8	3
4	$[{}_2\tau^2]_2$	$\{{}_2f^2\}_2$	4	2	12	1
4	$[{}_2\tau]_2$	$\{{}_3f\}_3$	4	1	24	1

Ahora bien, el desarrollo de Taylor de $\tilde{y}(x)$ se obtiene a partir del de los diferentes k_i, que coinciden con la función f evaluada en los distintos puntos. Puesto que $y' = f$ en realidad este desarrollo se puede expresar también en términos de los diferenciales elementales de f. Por tanto, el siguiente paso es determinar los coeficientes numéricos de la combinación de diferenciales elementales que permiten obtener las derivadas de $\tilde{y}(h)$. Estos coeficientes en principio no tienen por qué coincidir con los que ya se han obtenido al desarrollar directamente $y(x)$. Comenzamos por expresar formalmente el desarrollo de Taylor de $\tilde{y}(h)$ entorno a x, (o equivalentemente entorno a $h = 0$) del siguiente modo:

$$\tilde{y}(h) = \tilde{y}(0) + h\frac{\mathrm{d}}{\mathrm{d}h}\tilde{y}(h)\bigg|_{h=0} + \frac{1}{2}h^2\frac{\mathrm{d}^2}{\mathrm{d}h^2}\tilde{y}(h)\bigg|_{h=0} + \cdots$$

A continuación, adoptando la notación

$$a_{s+1,i} := b_i, \quad i = 1 \dots s$$

se introducen las funciones $\psi_i, i = 1 \dots s+1$ cuyo dominio es el conjunto T de los distintos (bajo isomorfismos) árboles con raíz, y cuyo recorrido es el conjunto de los reales. Es decir, son aplicaciones $\psi_i : T \to \mathbb{R}$ definidas de forma recursiva de acuerdo con las siguientes reglas:

$$\psi_i(\tau) = \sum_{j=1}^{s} a_{ij} = c_i,$$

$$\text{si } t_1, \dots, t_M \in T, \quad \psi([t_1 \dots t_M]) = \sum_{j=1}^{s} a_{ij}\psi_j(t_1)\cdots\psi_j(t_M).$$

Si denotamos $\psi(t) := \psi_{s+1}(t),\ \forall t \in T$, podemos enunciar el siguiente resultado:

Teorema 23. *Para un método RK de s niveles cuyas ecuaciones se han descrito anteriormente y que se aplica al caso general de un sistema $y' = f(y)$ con $y, f \in \mathbb{R}^m$, la expresión de las derivadas de orden $q \geq 1$ de $\tilde{y}(h)$ viene dada por*

$$\frac{\mathrm{d}^q}{\mathrm{d}h^q}\tilde{y}(h)\bigg|_{h=0} = \sum_{r(t)=q} \alpha(t)\gamma(t)\psi(t)F(t), \quad t \in T,$$

donde $F(t)$ son diferenciales elementales de $f(y)$ que están evaluados en $y(x)$.

4.5. Orden de los métodos Runge-Kutta

4.5.1. Expresión de métodos RK

$$h\tau(x,h) = y(x+h) - \tilde{y}(h) = \sum_{q=1}^{p} \frac{h^q}{q!} \left[y^{(q)}(x) - \frac{\mathrm{d}^q}{\mathrm{d}h^q}\tilde{y}(h)\Big|_{h=0} \right] + O(h^{p+1})$$

Para un método de orden p se cumple:

$$h\tau(x,h) = O(h^{(p+1)}) = \frac{h^{p+1}}{(p+1)!} \left[\sum_{r(t)=p+1} \alpha(t)[1-\gamma(t)\psi(t)]F(t) \right] + O(h^{p+2})$$

donde el PELD es:

$$PELD_{n+1} = \frac{h^{p+1}}{(p+1)!} \left[\sum_{r(t)=p+1} \alpha(t)[1-\gamma(t)\psi(t)]F(t) \right]$$

y $F(t)$ son las derivadas de Frechet de $f(y,x(t))$. La expresión anterior la hemos obtenido a partir de:

$$y^{(p+1)}(x) = \sum_{r(t)=p+1} \alpha(t)F(t); \quad \frac{\mathrm{d}^{p+1}}{\mathrm{d}q^{p+1}}\tilde{y}(h)\Big|_{h=0} = \sum_{r(t)=p+1} \alpha(t)\gamma(t)\psi(t)F(t)$$

y F es una biyección entre derivadas de Frechet y diferenciales elementales.

Ejemplo 4.23. Determinar el PELD para RK explícito de 2 niveles y orden máximo.

$$\begin{array}{c|cc} & 0 & 0 \\ c_2 & c_2 & 0 \\ \hline & b_1 & b_2 \end{array}$$

donde se cumplen las condiciones:

$$b_1 + b_2 = 1$$
$$b_2 c_2 = \frac{1}{2}$$

Los dos árboles que consideramos son: $t_1 = [[\tau]],\ t_2 = [\tau^2]$.

Para ellos se tiene:

$$F(t_1) = \{\{f\}\} = f^{(1)}(f^{(1)}(f)) \ \rightarrow \ f_y^2 f \text{ (dimension 1)}$$
$$F(t_2) = \{f^2\} = f^{(2)}(f,f) \ \rightarrow \ f_{yy} f^2 \text{ (dimension 1)}$$

Además, $\gamma(t_1) = 6,\ \gamma(t_2) = 3$ y $\alpha(t_1) = \alpha(t_2) = 1$.

Por otro lado, para los métodos sabemos que:

$$\psi(t_1) = b^T A c \quad \psi(t_2) = b^T c$$

Si imponemos que sea de dos niveles y explícito, entonces:

$$\psi(t_1) = 0, \quad \psi(t_2) = b_2 c_2^2$$

Por tanto el PELD $= (1 - 3b_2c_2^2) f^{(2)}(f,f)$

Teorema 24. *Si un método RK es de orden p, entonces* $b^T c^{q-1} = \frac{1}{q},\ q = 1 \div p.$

Demostración. Consideremos el árbol con raíz de orden q: $[\tau^{q-1}]$, que existe $\forall q \geq 1$. Para $q = 1$ tenemos el árbol trivial.

$$\psi_i([\tau^{q-1}]) = \sum_j a_{ij}[\psi_j(\tau)]^{q-1}$$
$$\psi_i([\tau^{q-1}]) = \sum_j a_{ij}c_j^{q-1}$$
$$\psi([\tau^{q-1}] = \sum_j b_j c_j^{q-1}$$

La condición de orden q será:

$$\psi([\tau^{p+1}]) = \frac{1}{\gamma([\tau^{q-1}])} \quad \Leftrightarrow \quad \sum_{j=1}^{s} b_j c_j^{q-1} = \frac{1}{q}$$

Teorema 25. *Un RK explícito de s niveles tiene orden* $\leq s$.

Demostración. Sea un RK de orden p y $t = [_{p-1}\tau]_{p-1}$.

$$\begin{aligned}
\gamma(t) &= p \cdot (p-1) \cdots 2 \cdot 1 = p! \\
\psi(t) &= \sum_{i=1}^{s} b_i \psi_i([_{p-2}\tau]_{p-2}) = \\
&= \sum_{i=1}^{s} b_i \sum_{j=1}^{s} a_{ij_1} \psi_{j_1}([_{p-3}\tau]_{p-3}) = \sum_{i,j_1,j_2} b_i a_{ij_1} a_{j_1 j_2} \psi_{j_2}([_{p-4}\tau]_{p-4}) = \\
&= \sum_{i,j_1,j_2 \ldots j_{p-2}} b_i a_{ij_1} a_{j_1 j_2} \cdots a_{j_{p-3} j_{p-2}} \psi_{j_{p-2}}(\tau)
\end{aligned}$$

donde $\psi_{j_{p-2}} = c_{j_{p-2}}$.

Además, sabemos que para un método RK explícito $a_{ij} = 0$ si $j \geq i$. Se cumple también $\psi(t) = 0$ excepto si existe una secuencia de enteros $i, j_1, j_2 \ldots j_{p-2}$ tal que $i > j_1 > j_2 > \cdots > j_{p-2} > 1$ $(*)$.

Utilizando $(*)$ nos queda $j_{p-2} > 1$ porque si $j_{p-1} = 1 \Rightarrow c_{j_{p-2}} = c_1 = 0$.

De $(*)$ se deduce que $i \geq p$, y como $i \leq s$ entonces $p \leq s$. Por tanto $\psi(t) \neq 0$ si $p \leq s$. Si $p > s \Rightarrow \psi(t) = 0 \neq \frac{1}{\gamma(t)}$, con lo que llegamos a una contradicción.

Teorema 26. *Un método RK explícito de s niveles tienen orden* $p < s$ *si* $s > 4$.

4.5.2. Condición suma-fila

Al principio del capítulo hemos visto que habitualmente exigíamos la condición suma-fila, es decir, que para un método de la família RK de s niveles se cumpliese que:

$$\sum_{j=1}^{s} a_{ij} = c_i, \qquad i = 1 \div s$$

En esta sección profundizaremos más en la justificación de este hecho: estudiaremos dos ejemplos, uno que cumple la condición suma-fila, y otro que no la cumple.

Llegaremos a la conclusión de que en general, la condición suma-fila permite alcanzar orden mayor que 1 para el PCI $y' = f(x, y(x))$.

Ejemplo 4.24 (Cumpliendo la condición suma-fila)**.** Sea el método definido por:

$$\begin{cases} k_1 = f(x_n, y_n + \frac{h}{2}k_1) \\ k_2 = f(x_n + h, y_n + \frac{h}{2}k_1) \end{cases}$$

De lo que se deriva que:

$$\begin{aligned} \tau(x,h) &= \frac{y(x+h)-y(x)}{h} - \frac{1}{2}(\hat{k_1} + \hat{k_2}) \\ \hat{k_1} &= f + \frac{h}{2}f_y\hat{k_1} + \frac{1}{2}f_{yy}\frac{\hat{k_1}^2}{2}h^2 + O(h^3) = f + \frac{h}{2}f_yf + \frac{h^2}{4}(f_y^2 f + \frac{1}{2}f_{yy}f) + O(h^3) \\ \hat{k_2} &= f + f_x h + f_y\frac{h}{2}\hat{k_1} + O(h^2) = f + f_x h + f_y f\frac{h}{2} + O(h^2) \end{aligned}$$

Por lo que, nos queda:

$$\frac{1}{2}(\hat{k_1} + \hat{k_2}) = f + f_x\frac{h}{2} + f_yf\frac{h}{2} + O(h^2) = y(x) + \frac{h}{2}y^{(2)}(x) + O(h^2)$$

Es decir, tenemos $\tau(x,h) = O(h^2)$. Desarrollando un orden más, obtenemos la expresión de $\tau(x,h)$:

$$\tau(x,h) = \frac{h^2}{6}y^{(3)}(x) - \frac{h^2}{4}(f_y^2 f + \frac{1}{2}f_{yy}f) - \frac{h^2}{4}(f_{xx} + f_{xx}f) + O(h^3)$$

Por lo que, sabiendo que la tercera derivada tiene la expresión:

$$y^{(3)}(x) = f_y(f_x + f_y f) + (f_{xx} + 2f_{xy}f + f_{yy}f^2)$$

Podemos concluir que $\tau(x,y)$ tiene un coeficiente de orden 3 no nulo, por lo que es orden 2 exacto.

Para generalizar este ejemplo, podemos considerar las definiciones ya conocidas del método RK:

$$\begin{cases} \hat{k_i} = f(x + c_i h, y(x) + h\sum_{j=1}^{s} a_{ij}\hat{k_j}) \\ \hat{k_i} = f(x, y(x)) + f_x(c_i h) + f_y(h\sum_{j=1}^{s} a_{ij}f) + O(h^3) \\ \hat{k_i} = f + h(c_i f_x + (\sum_j a_{ij})f_y f) + O(h^2) \end{cases}$$

Ejemplo 4.25 (Sin cumplir la condición suma-fila)**.** Consideremos un método RK de un nivel, en general, sin cumplir la condición suma fila.

$$\left.\begin{cases} k_1 = f(x + ch, y + \hat{c}hk_1) \\ y_{n+1} = y_n + hbk_1 \end{cases}\right\} \Rightarrow \tau(x,h) = (y'(x) - bf) + h(f_x(\frac{1}{2} - b\hat{c}) + f_y f(\frac{1}{2} - bc)) + O(h^2)$$

Notemos que para alcanzar orden mayor que 1 se tienen que cumplir las dos condiciones siguientes:

$$\frac{1}{2} - b\hat{c} = 0, \qquad \frac{1}{2} - bc = 0. \tag{4.3}$$

Por ser $b = 1$ por motivos de convergencia, enconces la condición a cumplir es que $c \neq \hat{c}$ para que el orden sea mayor que 1. La conclusión del ejercicio es que la condición suma fila es necesaria para obtener ordenes mayor que 1 con los método RK, y a su vez facilita la formulación de las condiciones de orden.

4.5.3. Formulación de las condiciones de orden de los métodos RK

Reuniendo los resultados anteriores podemos regresar a la expresión de la función $\tau(x,h)$ de un método RK:

$$h\tau(x,h) = y(x+h) - y(x) - h\Phi_f(y(x),x,h) = y(x+h) - \tilde{y}(h)$$

Para un método que sea de orden p al menos debe cumplirse

$$\tau(x,h) = O(h^p)$$

o lo que es equivalente,

$$y(x+h) - \tilde{y}(h) = O(h^{p+1}).$$

Por un lado se tiene

$$y(x+h) = y(x) + hy'(x) + \frac{h^2}{2}y^{(2)}(x) + \dots$$

siendo

$$y^{(q)} = \sum_{r(t)=q} \alpha(t)F(t), \quad t \in T$$

con $F(t)$ evaluado en $y(x)$. Por otro lado se tiene:

$$\tilde{y}(h) = \tilde{y}(0) + h\frac{\mathrm{d}}{\mathrm{d}h}\tilde{y}(h)\bigg|_{h=0} + \frac{1}{2}h^2\frac{\mathrm{d}^2}{\mathrm{d}h^2}\tilde{y}(h)\bigg|_{h=0} + \dots$$

siendo $\tilde{y}(0) = y(x)$ y

$$\frac{\mathrm{d}^q}{\mathrm{d}h^q}\tilde{y}(h)\bigg|_{h=0} = \sum_{r(t)=q} \alpha(t)\gamma(t)\psi(t)F(t), \quad t \in T,$$

con $F(t)$ evaluado en $y(x)$. Así pues, es inmediato enunciar el siguiente resultado.

Teorema 27 (Condiciones de orden de los métodos RK)**.** *Un método RK tiene orden p al menos si:*

Para todo árbol con raíz $t \in T$ tal que $r(t) \leq p$ se verifica la igualdad $\psi(t) = 1/\gamma(t)$.

Si además existe $t \in T$ de orden $p+1$ tal que $\psi(t) \neq 1/\gamma(t)$, entonces el método RK es de orden p exactamente.

Por último, cabe destacar algunos comentarios:

- En la forma matricial, los coeficientes del método RK se agrupan de la forma convencional para construir la correspondiente matriz de Butcher. Las potencias del vector c representan

  $$c^m = [c_1^m \dots c_s^m]^T, \quad m \geq 0, c^0 = [1 \dots 1]^T.$$

  La matriz D es de dimensión s y se define como $D = \mathrm{diag}\{c_1 \dots c_s\}$. Las potencias de las matrices D y A se calculan mediante el producto matricial habitual.

- Para que un método RK sea de orden 4 al menos deben verificarse un total de 8 condiciones. Como ya se comentó, el número de árboles con raíz va creciendo con el orden. Puesto que hay 9 árboles de orden 5, el total de condiciones que debe cumplir un RK para alcanzar orden 5 al menos es igual a $8 + 9 = 17$.

4.6. Métodos RK imbricados

Un método RK imbricado es un conjunto de dos métodos RK que comparten gran parte de los coeficientes. Los dos métodos tendrán un orden distinto. Uno de ellos (el de menor orden) servirá para obtener la solución numérica, y el otro (el de orden más grande), lo utilizaremos para aproximar el ELD cometido:

$$\begin{cases} y_n = y_n(\text{método 1}) \\ ELD \approx y_n(\text{método 1}) - y_n(\text{método 2}) \end{cases}$$

- **Método 1** (explícito o semiimplícito) de orden p.

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

de s niveles. Sirve para obtener soluciones numéricas.

- **Método 2** (explícito o semiimplícito) de orden $p+1$ y $\hat{s}$ niveles con $\hat{s} \geq s$

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

donde las s primeras componentes de $\hat{c}$ forman c y las s primeras filas y columnas de $\hat{A}$ forman A.

Es decir, las $k_i,\ i = 1 \div s$ de los métodos 1 y 2 son iguales.

$$y_{n+1}^{(1)} = y_n^{(1)} + \sum_{i=1}^{s} b_i k_i$$

$$y_{n+1}^{(2)} = y_n^{(2)} + \sum_{i=1}^{\hat{s}} \hat{b}_i k_i$$

Si restamos ambas expresiones obtenemos:

$$y_{n+1}^{(2)} - y_{n+1}^{(1)} = h \sum_{i=1}^{\hat{s}} (\hat{b}_i - b_i) k_i, \quad \text{si } y_n^{(1)} = y_n^{(2)}$$

que nos da $b_i = 0$ si $i > s$. Por otro lado,

$$y(x_{n+1}) - \tilde{y}_{n+1}^{(1)} = h\tau^{(1)}(x,h) = \Psi^{(1)}(y(x_n))h^{p+1} + O(h^{p+2})$$
$$y(x_{n+1}) - \tilde{y}_{n+1}^{(2)} = h\tau^{(2)}(x,h) = \Psi^{(2)}(y(x_n))h^{p+2} + O(h^{p+2})$$

donde $\tilde{y}_{n+1}^{(1)}, \tilde{y}_{n+1}^{(2)}$ son las soluciones numéricas obtenidas con los métodos 1 y 2, suponiendo que $y_n^{(1)} = y_n^{(2)} = y(x_n)$.

En particular, se cumple que

$$y_{n+1}^{(2)} - y_{n+1}^{(1)} = \Psi^{(1)}(y(x_n))h^{p+1} + O(h^{p+2})$$

$$PELD_{n+1} = \Psi^{(1)}(y(x_n))h^{p+1} \approx y_{n+1}^{(2)} - y_{n+1}^{(1)} = h \sum_{i=s+1}^{\hat{s}} (\hat{b}_i - b_i) k_i$$

Comentario 4.26. Un método RK imbricado $(p, p+1)$ nos proporciona la solución numérica mediante el método 1, y el $PELD$ lo obtenemos gracias a la diferencia de soluciones de los métodos 1 y 2. Por ello puede resultar interesante denotar los coeficientes de ambos métodos mediante la siguiente matriz de Butcher adaptada:

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

4.7. Problemas

Problema 47. De acuerdo con el enunciado, consideremos el siguiente predictor-corrector:

$$\begin{aligned} P: &\qquad y_{n+1}^{[0]} = y_n + hf_n \\ C: &\qquad y_{n+1} = y_n + \frac{h}{2}\left(f_n + f_{n+1}\right) \end{aligned}$$

Así, para la primera iteración

$$y_{n+1}^{[1]} = y_n + \frac{h}{2}\left[f_n + f(x_n + h, y_{n+1}^{[0]})\right]$$

donde

$$\begin{aligned} f_n &= k_1 \\ y_{n+1}^{[0]} &= y_n + hk_1 \end{aligned}$$

Observamos que el orden del predictor es 1, y el del corrector es 2. El orden del método predictor-corrector se corresponde con el del corrector si $\mu \geq 2 - 1 = 1$. Al aplicar el método en modo $PECE$ obtenemos un método de tipo Runge-Kutta:

$$\begin{aligned} y_{n+1} &= y_n + \frac{h}{2}\left(k_1 + k_2\right) \\ k_1 &= f_n \\ k_2 &= f(x_n + h, y_n + hk_1) \end{aligned}$$

La matriz de Butcher para este caso es

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

que es un método RK de orden 2.

Supongamos que aplicamos el método en el modo $P(EC)^2E$. En este caso

$$y_{n+1}^{[2]} = y_n + \frac{h}{2}\left[f_n + f(x_n + h, y_{n+1}^{[1]})\right] = y_n + \frac{h}{2}\left(k_1 + k_3\right)$$

Los valores para las k_i del RK correspondiente son:

$$\begin{aligned} k_1 &= f_n \\ k_2 &= f(x_n + h, y_n + hk_1) \\ k_3 &= f(x_n + h, y_n + \frac{h}{2}k_1 + \frac{h}{2}k_2) \end{aligned}$$

La matriz de Butcher para $\mu = 2$ es

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

Discutimos el orden según la matriz de Butcher:

$$\begin{aligned} \frac{1}{2} + \frac{1}{2} = 1 \quad &\Rightarrow \quad \text{orden} \geq 1 \\ 0 \cdot 1 + \frac{1}{2} \cdot 1 = \frac{1}{2} \quad &\Rightarrow \quad \text{orden} \geq 2 \\ 0 \cdot 1^2 + \frac{1}{2} \cdot 1^2 = \frac{1}{2} \neq \frac{1}{3} \quad &\Rightarrow \quad \text{orden exactamente 2} \end{aligned}$$

Para un RK implícito de ≤ 3 niveles:

$$b_1 + b_2 + b_3 = 1 = b^T \cdot c^0$$

donde

$$\begin{aligned} c^l &= [c_1^l \dots c_s^l]^T, \quad b^T \cdot c = \frac{1}{2} \\ c^0 &= [1 \dots 1]^T, \quad b^T \cdot c = \frac{1}{3} \end{aligned}$$

Intuimos que para términos en h^q aparece una condición de orden

$$b^T c^{q-1} = \frac{1}{q}$$

Así es. Una condición necesaria para que un RK alcance orden p es

$$b^T c^{q-1}, \quad q = 1 \div p$$

No es una condición suficiente.

Problema 50. Consideremos un método con matriz de Butcher:

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

No podemos aplicar las condiciones de orden. Deberemos obtener el desarrollo de Taylor de $\tau(x, h)$. Por otro lado, tampoco cumple la condición suma-fila.

A partir de la matriz de Butcher sabemos que el método es

$$y_{n+1} = y_n + \frac{h}{2}k_1 + \frac{h}{2}k_2$$

donde

$$k_1 = f(x_n, y_n + \frac{1}{2}k_1) \quad \text{(ídem ejercicio 48)}$$
$$k_2 = f(x_n + h, y_n + \frac{h}{2}k_1) \quad \text{(caso particular } k_2 \text{ para un método explícito)}$$

Problema 52. Se trata de un método de orden 3 en el caso general. Si tomamos $y' = f(y)$ en dimensión 1 entonces tiene orden 4:

$$t = [\tau[\tau]]$$
$$t' = [[\tau^2]]$$

Observamos que el orden es 3 para el caso general puesto que no se cumplen las dos condiciones de orden 4:

$$bAc^2 = \frac{1}{12}$$
$$bDAc = \frac{1}{8}$$

Hay dos términos en potencia de h^3 del desarrollo de Taylor de $\tau(x,h)$ que no se anulan:

$$t \quad \rightarrow \quad f^{(2)}(f, f^{(1)}(f)) = f_{yy}f(f_y f) = f_{yy}f_y f^2$$
$$t' \quad \rightarrow \quad f^{(1)}(f^{(2)}(f,f)) = f_y(f_{yy}f \cdot f) = f_{yy}f_y f^2$$

y notamos que ambos términos coinciden.

Podemos reescribir las condiciones de orden 4 del modo siguiente:

$$1 - 12bAc^2 = 0$$
$$1 - 8bDAc = 0$$

de modo que

$$\alpha(t')[1 - 12bAc^2] = 0$$
$$\alpha(t)[1 - 8bDAc] = 0$$

Con respecto a la función τ:

$$\tau(x,h) = \frac{y(x+h) - y(x)}{h} - \sum_{i=1}^{s} b_i \tilde{k}_i$$

Para este método la $\tau(x,h)$ posee en h^3 esta expresión:

$$\tau(x,h) = \alpha(t')[1 - 12bAc^2]f^{(1)}(f^{(2)}(f,f))\frac{h^3}{4!} + \alpha(t)[1 - 8bDAc]f^{(2)}(f, f^{(1)(f)})\frac{h^3}{4!} + O(h^4)$$

donde la expresión anterior se corresponde con $PELD_{n+1}/h$ del método.

- $\alpha(t)$ es el número de formas de etiquetar los vértices de t con $\{1, 2 \dots r(t)\}$
- $r(t)$ es el orden de t bajo las siguientes condiciones:
  1. Cada vértice (i, j) cumple $i < j$.
  2. Las aristas (i, j) cumplen $i < j$.
  3. Etiquetados automorfos cuentan como uno solo.

Por tanto, tenemos los siguientes árboles:

- t' $\quad (1-2, 2-3, 2-4), \quad \alpha('t) = 1$
- t $\quad (1-2, 1-3, 3-4), \quad (1-3, 1-2, 2-4), \quad (1-4, 1-2, 2-3), \quad \alpha(t) = 3$

$$\alpha(t')[1 - 12bAc^2] + \alpha(t)[1 - 8bDAc] = 0,$$
$$\alpha(t') = 1, \quad bAc^2 = 0, \quad \alpha(t) = 3, \quad bDAc = 1/6$$

Y por tanto $\tau(x, h) = O(h^4)$ para el caso escalar con un sistema autónomo.

Problema. (problema 4 Extra_7Juliol2010.pdf) El método de orden 3 es explícito y su matriz de Butcher es:

$$\left(\begin{array}{c|ccc} 0 & & & \\ 1/2 & 1/2 & & \\ 1/2 & 0 & 1/2 & \\ \hline & b_1 & b_2 & b_3 \end{array}\right)$$

por lo que se trata de un método de 3 niveles.

$$\begin{cases} b_1 + b_2 + b_3 = 1, & \text{orden} \geq 1 \\ \dfrac{b_2}{2} + \dfrac{b_3}{2} = \dfrac{1}{2}, & \text{orden} \geq 2 \end{cases}$$

Por otro lado, las condiciones de orden 3 son:

$$\frac{b_2}{4} + \frac{b_3}{4} = \frac{1}{3}$$

que es incompatible con las dos anteriores, y por tanto no puede construirse el RK (3,4) imbricado.

Problema. Consideremos el siguiente caso:

$$y' = f(y), \quad f : \mathbb{R}^m \to \mathbb{R}^m$$

Para la derivada tercera:

$$y^{(3)} = \{f^2\} + \{{}_2f\}_2$$

Se obtienen los siguientes árboles de tercer orden:

- $[_2\tau]_2$, (1-2,2-3), $\gamma = 6$

$$\alpha([_2\tau]_2) = \frac{3!}{6 \cdot 1} = 1$$

- $[\tau^2]$, (1-3,3-1), $\gamma = 2$

$$\alpha([\tau^2]) = \frac{3!}{3 \cdot 2} = 1$$

Introducimos las funciones ψ_i y las condiciones de orden que se obtienen a partir de éstas:

- Orden 1:

$$\psi_i(\tau) = \sum_{j=1}^{s} a_{ij} = c_i, \quad i = 1 \div s$$

$$\psi(\tau) = \sum_{i=1}^{s} b_i$$

$$\psi(\tau) = \frac{1}{\gamma(\tau)} \quad \Leftrightarrow \quad \sum_i b_i = 1$$

- Orden 2: $[\tau]$

$$\psi_i([\tau]) = \sum_j a_{ij}\psi_j(\tau) \overset{*}{=} \sum_j a_{ij}c_j$$

$$(*) \quad \text{orden 1}$$

$$\psi([\tau]) = \sum_j b_j\psi_j(\tau) = \sum_j b_j c_j$$

$$\psi([\tau]) = \frac{1}{\gamma([\tau])} \quad \Leftrightarrow \quad \sum_j b_j c_j = \frac{1}{2}$$

- Orden 3:

  1. $[\tau^2]$

$$\psi_i([\tau^2]) = \sum_j a_{ij}(\psi_j(\tau))^2 = \sum_j a_{ij}c_j^2$$

$$\psi([\tau^2]) = \sum_j b_j c_j^2$$

$$\gamma([\tau^2]) = 3 \Rightarrow \sum_j b_j c_j^2 = \frac{1}{3}$$

  2. $[_2\tau]_2$

$$\psi_i([_2\tau]_2) = \sum_j a_{ij}\psi_j([\tau]) = \sum_j a_{ij} \sum_k a_{ij}c_k$$

$$\psi([_2\tau]_2) = \sum_j b_j \sum_k a_{jk}c_k$$

$$\gamma([_2\tau]_2) = 6 \Rightarrow \sum_j b_j \sum_k a_{jk}c_k = \frac{1}{6}$$

Problema. En este problema primero hemos comprobado que el método tiene orden 3. Una vez hecho el análisis sobre un problema autónomo y no autónomo se descubre que el primero tiene un comportamiento de orden 4. Ésto sucede debido a que:

$$y_{n+1} = y_n + h(\frac{k_1}{6} + \frac{2}{3}k_2 + \frac{k_4}{6})$$
$$\text{Aplicado al PCI} y_1 = y_0 + h(\ldots)$$
$$k_1 = y_0 = 1$$
$$k_2 = y_0 + \frac{h}{2}k_1 = 1 + \frac{h}{2}$$
$$k_3 = 1 + \frac{h}{2}k_1 - \frac{3h}{2}k_2 = 1 - h - \frac{3}{4}h^2$$
$$k_4 = y_0 + \frac{4}{3}hk_2 - \frac{1}{3}hk_3 = 1 + h + h^2 + \frac{h^3}{4}$$

Por lo que la solución numérica en el primer paso es:

$$y_1 = 1 + h + \frac{h^2}{2} + \frac{h^3}{6} + \frac{h^4}{24}$$

Como la solución exacta es $y(x) = e^x$, tenemos que:

$$y(h) - y_1 = O(h^5)$$

Por lo que el error general de discretización es que orden 4. Para el caso $y' = xy$ vamos a obtener $y(h) - y_1 = O(h^4)$, por lo que el EGD es de orden 3, un orden inferior como queríamos ver. No entramos en cálculos detallados.

Problema. Veamos ahora que los métodos RK no explícitos tienen zonas de estabilidad no acotadas, contrariamente a los explícitos. Teníamos:

$$R(\bar{h}) = 1 + \bar{h}b^T(I_s - \bar{h}A)^{-1}e$$
$$Adj(I_s - \bar{h}A) \rightarrow \quad \text{sus coef son funciones polinomicas de} \quad \bar{h} \quad \text{de grado} \quad s-1.$$

Si el método no es explícito $det(I_s - \bar{h}A) \neq 1$ el det es un polinomio en $\bar{h}$ de grado $\leq s$. $\Rightarrow$ es una función racional en $\bar{h}$.

$$R(\bar{h}) = \frac{p(\bar{h})}{q(\bar{h})}$$

Dónde p, q son polinomios de grado $\leq s$.

Problema 45. Comprobamos las condiciones de orden y obtenemos:

$$\frac{1}{10} + \frac{1}{2} + \frac{2}{5} = 1 \Rightarrow \quad \text{orden} \quad \geq 1$$
$$b^T c = \frac{1}{2} \Rightarrow \quad \text{orden} \quad \geq 2$$

Veamos ahora si cumple las condiciones de orden 3:

$$\alpha(t_1) = 1 = \alpha(t_2) \tag{4.4}$$

$$\gamma(t_1) = 6 \quad \gamma(t_2) = 3 \tag{4.5}$$

$$b^T Ac = \frac{1}{6}\text{cond de orden 3} \quad \Longleftrightarrow \; b_3 a_{32} c_2 = \frac{1}{6} \tag{4.6}$$

$$\Psi(t_2) = \sum_{i=1}^{s} b_i c_i^2 \tag{4.7}$$

$$b^T c^2 = \frac{1}{3} \quad \text{Condición de orden 3} \tag{4.8}$$

Cumple la condición de orden ≥ 3 y 3 niveles y explícito. *Rightarrow* Orden 3 exacto. Veamos ahora el término principal del error:

$$PELD = \frac{h^4}{4!} \sum_{r(t)=4} \alpha(t)(1 - \gamma(t)\Psi(t))F(t)$$

F biyección arboles con raíz y diferenciales elementales. Ahora haciendo cálculos tenemos ya que:

$$PELD = \frac{h^4}{4!} f^{(1}(f^{(1)}(f^{(1)}(f)))) + \dots$$

Falta pues calcular la contribución de los otros tres árboles: Para t_2, collita propia:

$$\alpha(t_2) = 2$$
$$\gamma(t_2) = 4!$$
$$\Psi(t_2)$$

Capítulo 5

Introducción a los Métodos para Sistemas Rígidos

En este capítulo presentaremos un tipo concreto de problema, caracterizado por las dificultades con añadidas con las que nos encontramos a la hora de resolverlo. El grueso del capítulo se encuentra en el fichero de Atenea *stiffness.pdf*.

5.0.1. Determinación de métodos A-estables

Para resolver sistemas stiff utilizaremos los siguientes métodos:

- Métodos implícitos.
- De un paso.

Es necesario que sean A_0-estables, aunque también es deseable que sean A-estables.

- Utilizaremos métodos de un paso (RK o lineales).
- Los aplicaremos al problema $y = \lambda y$, cuya solución numérica cumple $y_{n+1} = R(\hat{h})y_n$
- Para un método RK: $R(\hat{h}) = 1 + \hat{h}b^T(I_s - \hat{h}A)^{-1}e$, que es de s niveles.
- Para un método Lineal: $R(\hat{h}) = \frac{1+\beta_0\hat{h}}{1-\beta_1\hat{h}}$, con $y_{n+1} - y_n = h(\beta_1 f_{n+1} + \beta_0 f_n)$.
- Son métodos implícitos, $R(\hat{h})$ es una función racional.

$$y' = \lambda y \quad \Rightarrow \quad y(x) = Ke^{(}\lambda x)$$
$$y(x_{n+1}) = y(x_n + h) = e^{(}\lambda h) \cdot y(x_n)$$

Si el método es de orden p entonces:

$$y(x_{n+1}) - \tilde{y}_{n+1} = [\exp(\hat{h}) - R(\hat{h})]y(x_n) = O(h^{p+1})$$
$$R(\hat{h}) = \exp(\hat{h}) + O(h^{p+1})$$

Definición 5.1. Sea $R_T^S(q)$ función racional, donde S es el grado del polinomio del numerador y T es el grado del polinimio del denominador. Se dice que $R_T^S(q),\ q \in \mathbb{C}$ es una aproximación racional de orden p de la exponencial si y solo si

$$R_T^S(q) = \exp(q) + O(q^{p+1})$$

Teorema 28. *Un método de un paso de orden p posee una función de estabilidad absoluta que es una aproximación racional de orden p de* $\exp(\hat{h}),\ \hat{h} = \lambda h$.

Sea

$$R_T^S(q) = \frac{\sum_{i=0}^{S} a_i q^i}{\sum_{i=0}^{T} b_i q^i}$$

Teorema 29. R_T^S *es una aproximación racional de orden p de* $\exp(q)$ *si y solo si:*

$$1 + a_1 q + \ldots + a_S q^S - (1 + b_1 q + \ldots + b_T q^T)(1 + q + \frac{q^2}{2} + \ldots) = O(q^{p+1})$$

Teorema 30. *El máximo orden de una aproximación racional* R_T^S *es* $S + T$. *La aproximación racional* R_T^S *de máximo orden posible es única y se denomina aproximación de Padé de* e^q. *Se denota* $\hat{R}_T^S$.

Definición 5.2. Una aproximación racional R_T^S se denomina:

- A-*aceptable* si $\forall\, q \in \mathbb{C},\ \mathtt{Re}(q) < 0 \ \Rightarrow\ |R_T^S(q)| < 1$.
- A_0-*aceptable* si $\forall\, q \in \mathbb{R},\ q < 0 \ \Rightarrow\ |R_T^S(q)| < 1$.
- L-*aceptable* si es A-aceptable y $|R_T^S| \to 0$ si $\mathtt{Re}(q) \to -\infty$.

Comentario 5.3. Si R_T^S es una función de estabilidad absoluta de un método de un paso entonces:

- A-aceptable $\Leftrightarrow$ método A-estable.
- A_0-aceptable $\Leftrightarrow$ método A_0-estable.
- L-aceptable $\Leftrightarrow$ método L-estable.

Resultados sobre aceptabilidad de funciones racionales

1. La aproximación de Padé $\hat{R}_T^S$ es A-aceptable $\Leftrightarrow$ $T - 2 \leq S \leq T$
2. Si $S = T - 2$ o bien $S = T - 1$ entonces $\hat{R}_T^S$ es L-aceptable.
3. $R_T^S(q)$ es A-aceptable y $T > S$ $\Rightarrow$ $R_T^S(q)$ es L-aceptable.
4. $R_T^S(q)$ con $S > T$ no puede ser A-aceptable.
5. Si $\hat{R}_T^S$ es la aproximación de Padé y $T \geq S$ entonces $\hat{R}_T^S(q)$ es A_0-aceptable.

Capítulo 6

Problemas de Condiciones de Frontera

6.1. Caso general: PCF

Sea $y' = f(x, y) \in \mathbb{R}^m$. Se busca la solución $y(x)$ tal que $r\Big(y(a), y(b)\Big) = 0$, donde se tiene que $r : \mathbb{R}^m \times \mathbb{R}^m \mapsto \mathbb{R}^m$ es una función que se anulará si nuestra solución cumple con las condiciones de frontera.

Es decir, si exigimos que $y_{exacta}(a) = y_1,\ y_{exacta}(b) = y_2$, entonces nuestra función r podría ser:

$$r(y(a), y(b)) = (y(a) - y_1)^2 + (y(b) - y_2)^2.$$

Hay que tener en cuenta que el Problema de Condiciones Frontera:

- Puede no tener solución, o
- puede admitir más de una solución.

Condiciones de Frontera Separables

Son casos en los que tenemos: $r_1(y(a)) = 0$, y $r_2(y(b)) = 0$, donde $r_1,\ r_2 : \mathbb{R}^m \mapsto \mathbb{R}^m$.

Ejemplo 6.1 (Condiciones de Frontera Separables)**.**

$$r_1(y(a)) = y(a) - \alpha, \qquad r_2(y(b)) = y(b) - \beta$$

6.1.1. Problemas de valores singulares ("eigenvalue problems")

Sea $y' = f(x, y; \lambda)$, con $r\Big(y(a), y(b); \lambda\Big) = 0$, donde $r : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R} \mapsto \mathbb{R}^{m+1}$.

El problema consiste en determinar los valores de λ (valores singulares) para los cuales existe una solución $y(x)$. Definimos para $y(x) \in \mathbb{R}^m,\ {}^{m+1}y(x) = \lambda,\ {}^{m+1}y'(x) = 0$ los siguientes vectores:

$$\begin{aligned} \vec{y} &= \Big[{}^1y, \ldots, {}^m y, {}^{m+1}y\Big]^t \\ \vec{f} &= \Big[f, 0\Big]^t \\ \vec{y} &= \vec{f}(x, \vec{y}) \end{aligned}$$

6.1.2. Problema de la frontera libre

El problema en este caso consiste en determinar $\lambda \in \mathbb{R}$ tal que la solución $y(x)$ de $y'(x) = f(x,y)$ cumpla que $r\Big(y(a), y(\lambda)\Big) = 0$.

Ejemplo 6.2.

$$x = a + t(\lambda - a) = a + tz_{m+1}, \qquad 0 \le t \le 1 \Rightarrow \frac{dz_{m+1}}{dt} = \dot{z_{m+1}} = 0$$

$$\tilde{z}(t) = y(a + tz_{m+1})$$

Definimos el vector $\vec{z}(t) = [\tilde{z}, z_{m+1}(t)]^t$, con lo que tenemos que:

$$\frac{d\tilde{z}(t)}{dt} = \dot{\tilde{z}} = z_{m+1}y'(a + tz_{m+1}) = z_{m+1}f(x,y)$$

$$\dot{z} = \vec{f}(t,z); \qquad \vec{f} = [z_{m+1}f(x,y),\ 0]^t$$

Por tanto, $r\Big(y(a), y(\lambda)\Big) = 0 \iff \tilde{r}(\tilde{z}) = 0$, donde se tiene que cumplir que $\tilde{z}(1) = 0$.

6.2. Método del tiro simple

Ejemplo 6.3. Tenemos la ecuacion:

$$yy'' = -1 - y'^2, \qquad y(0) = 1, \quad y(1) = 2$$

Para transformarlo en un sistema de primer orden, operamos de la siguiente manera:

$$y'' = \frac{1 - y'^2}{y} \Rightarrow \begin{pmatrix} y \\ y' \end{pmatrix} = z \in \mathbb{R}^2$$

De donde nos queda el sistema en z:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}' = z' = f(z) = \begin{pmatrix} y \\ y' \end{pmatrix} = \begin{pmatrix} z_2 \\ \frac{1 - z_2^2}{z_1} \end{pmatrix}$$

Tenemos la condicion inicial $y(0) = 1$, y también nos exigen que $y(1) = 2$. Para resolverlo, buscaremos la condición que debe cumplir $y'(0)$ para que la segunda restricción sea cierta (o, por lo menos, obtener una aproximación suficientemente buena).

En nuestro caso, la solución que obtenemos es $y'(0) = 5$, es decir,

$$z(0) = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

Para obtener este valor en la práctica (el valor que debe cumplir $y'(0)$), podemos utilizar el *método del tiro simple*. Este consiste en:

- Plantear una ecuación no lineal cuya raiz nos dará la solución: Sea $y(1,s)$ el valor de $y(1)$ dada la condición $y'(0) = s$. Entonces, en nuestro caso particular la ecuación no lineal a minimizar será $F(s) = y(1,s) - 2 \approx 0$. Habrá que dar una tolerancia ε.

- Resolver la ecuación no lineal: hay que minimizar una función, teniendo en cuenta que cada punto en que evaluemos la función $F(1, s)$ se obtendrá resolviendo una ecuación diferencial concreta (es decir, probablemente aplicando uno de los métodos estudiados anteriormente). Se puede aplicar, por ejemplo, el método de la secante para aproximar el cero de F.

Si tenemos un PCF de la forma: $y' = f(x, y),\ r\Big(y(a), y(b)\Big) = 0$, y queremos encontrar una solución de la ecuación, podemos utilizar el método del tiro simple si f cumple la condición de Lipschitz (ya que esto asegura que existirá una solución $y(x; s)$).

Sea $s \in \mathbb{R}^m$. Consideramos el PCI:

$$y' = f(x, y), \quad y(a) = s.$$

Entonces, obtenemos una solución numérica aproximada en x_n que es $y(x_n; s)$ para $n = 0, 1, \dots N,\ \ x_0 = a,\ \ x_N = n$.

Nuestro objetivo es buscar una solución numérica tal que:

$$r\Big(y(a), y(b)\Big) = r\Big(s, y(x_N; s)\Big) = 0$$

Esto es equivalente a hallar la solución $s \in \mathbb{R}^m$ de una ecuación $F(s) = r\Big(s, y(x_n; s)\Big) = 0$, donde $F(s)$ en general es no lineal y no se puede calcular analíticamente debido a que depende de $y(x_n; s)$.

Aplicando el método de Newton a $F(s) = 0$ (método para encontrar ceros de funciones), se tiene:

$$J_F(s^{[0]})\widetilde{\Delta}s^{[0]} = -F(s^{[\nu]})$$

Donde J_F es la matriz jacobiana de F respecto a s y $\nu = 0, 1, \dots$. Se tiene que:

$$\widetilde{\Delta}s^{[\nu]} = s^{[\nu+1]} - s^{[\nu]}$$

Cada iteración del método requiere:

1. Calcular $F(s^{[\nu]})$. Por tanto, también necesitamos calcular $y(x_N, s^{[\nu]})$, y, por tanto, necesitamos la solución numérica aproximada de $y' = f(x, y),\ \ y(a) = s^{[\nu]}$.
2. Calcular $J_F(s^{[\nu]})$. Este es el paso más costoso.
3. Resolver el sistema lineal $J_F(S^{[0]})\widetilde{\Delta}s^{[0]} = -F(s^{[\nu]})$.

Observemos que en el caso $m = 1$, tenemos que $J_F(s) = \frac{dF(s)}{ds} \approx \frac{\Delta F(s)}{\Delta s}$, con lo que obtener la jacobiana no es demasiado costoso. Esto no es así en general para dimensiones superiores.

6.2.1. Limitaciones del método de tiro simple

Una precisión elevada en la solución de $F(s) = 0$ no implica que $\|y(x_N; s) - y(b; s)\| = EGD_N$ sea lo suficientemente pequeña.

Teorema 31. *Sea $y' = f(x, y),\ \ y(a) = s$. Sean s_1, s_2 dos condiciones iniciales posibles para las cuales f es Lipschitz. Entonces,*

$$\|y(x_N; s_1) - y(x_N; s_2)\| \leq \|s_1 - s_2\| e^{L|x_n - a|}$$

Donde L es l constante de Lipschitz de f.

Supongamos ahora que aplicamos el método del tiro simple a un problema en el que el intervalo de integración es muy grande, y con una función "suave" (esto es, con un buen comportamiento). Por ejemplo, tomemos $b-a=100,\ L\approx 1$.

Sea $\bar{s}$ tal que $y(x;\bar{s})$ es la solución exacta del PCF. Entonces,

$$EGD_n = \|y(x_N;\bar{s}) - y(x_N;s)\| \leq \|\bar{s}-s\|e^{L|x_n-a|}$$

Si al resolver $F(s)=0$ la solución se obtiene para $s=\bar{s}+\varepsilon$, entonces:

$$\|\bar{s}-s\| = \varepsilon,\ \ L\approx 1,\ \ |x_n-a|\approx 100 \Rightarrow EGD_n \leq \varepsilon\cdot e^{100} \approx \varepsilon\cdot 10^3$$

Con lo que obtenemos una cota del error excesivamente grande.

Para evitar este problema, se subdivide el intervalo en trozos, y se aplica el método del tiro libre en cada uno de ellos. Esta solución es conocida como el *método del tiro múltiple*.

6.3. Método del tiro múltiple o tiro paralelo

Tenemos el problema:

$$\begin{cases} y' = f(x,y) \\ r\Big(y(a),y(b)\Big) = 0 \quad [a,b]\subseteq\mathbb{R} \end{cases}$$

Consideremos una partición: $a=x_1<x_2<\ldots<x_N=b$ Ahora, dados $s_1,\ s_2,\ \ldots,\ s_N$, calculamos la solución numérica de:

$$\begin{cases} y'=f(x,y) & y(x_1)=y(a)=s_1 & [x_1,x_2] \\ y'=f(x,y) & y(x_2)=s_2 & [x_2,x_3] \\ \quad\vdots & & \\ y'=f(x,y) & y(x_{N-1})=s_{N-1} & [x_{N-1},x_N] \end{cases}$$

En general, consideramos $N-1$ PCI:

$$y'=f(x,y)\quad y(x_k)=s_k\quad [x_k,x_{k+1}],\qquad k=1,\ldots,N-1$$

La idea consiste en ir ajustando los valores $s_1,\ s_2,\ \ldots,\ s_N$, de tal manera que se minimice una cierta función de error $F(s_1,\ldots,s_N)$. Esta función será mínima cuando los $N-1$ PCI cumplan la restricción en la frontera, es decir, cuando

$$y'=f(x,y)\quad y(x_k)=s_k\quad [x_k,x_{k+1}]\Rightarrow s_{k+1}=y(x_{k+1}),\qquad k=1,\ldots,N-1$$

Notemos que en este momento, la solución "encajará", es decir, será contínua, y se cumplira el PCF requerido.

Para Y,f,r con recorrido $\mathbb{R}^n$, $s_i\in\mathbb{R}^m\quad i=1,\ldots,N$, la expresión vectorial del problema será:

$$F(s)=\begin{bmatrix} F_1(s_1,s_2) \\ F_2(s_2,s_3) \\ \vdots \\ F_{N-1}(s_{N-1},s_N) \\ F_N(s_1,s_N) \end{bmatrix} = 0$$

Las componentes se definen como:

$$F_i(s_i, s_{i+1}) = y(x_{i+1}; \{x_i, s_i\}) - s_{i+1} \quad i = i, \dots, N-1$$
$$F_N(s_1, s_N) = r(s_1, s_N) = r\Big(y(a), y(b)\Big)$$

F_i son funciones con recorrido en $\mathbb{R}^m$. $F : \mathbb{R}^{m\times N} \mapsto \mathbb{R}^{m\times N}$

$$s = [s_1^T, \dots, s_N^T]^T \in \mathbb{R}^{mN}$$

Utilizando el método de Newton para resolver $F(s) = 0$. Obtendremos la solución iterando la aproximación con la siguiente fórmula:

$$J_F(s^{[\nu]})\hat{\Delta}s = -\begin{bmatrix} F_1^{[\nu]} \\ \vdots \\ F_N^{[\nu]}(s_1, s_N) \end{bmatrix} \quad \nu = 0, 1 \dots$$
$$\hat{\Delta}s = \begin{bmatrix} \hat{\Delta}s_1 \\ \vdots \\ \hat{\Delta}s_N \end{bmatrix} = s^{[\nu+1]} - s^{[\nu]}$$

Para $k = 1, \dots, N-1$ se tiene que:

$$\frac{\partial F_k}{\partial s_k} = \frac{\partial y(x_{k+1}; \{x_k, s_k\})}{\partial s_k} = \frac{y(x_{k+1}; \{x_k, s_k + \Delta s_k\}) - y(x_{k+1}; \{x_k, s_k\})}{\Delta s_k} = G_k \in \mathcal{M}_{m\times m}$$
$$\frac{\partial F_k}{\partial s_{k+1}} = -I_m \quad k = 1, \dots, N-1$$
$$\frac{\partial F_N}{\partial S_N} = \frac{\partial r(s_1, s_N)}{\partial s_N} = B \in \mathcal{M}_{m\times m}$$
$$\frac{\partial F_N}{\partial s_1} = \frac{\partial r(s_1, s_N)}{\partial s_1} = A \in \mathcal{M}_{m\times m}$$

Tenemos, que la matriz $J_f(s)$ acabará siendo de la forma:

$$J_F(s) = \begin{bmatrix} G_1 & -I_m & 0 & \dots & 0 \\ 0 & G_2 & -I_m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & G_{N-1} & -I_m \\ A & 0 & \dots & 0 & B \end{bmatrix}$$

$$s^{[\nu]} = J_F(s)\begin{bmatrix} \hat{\Delta}s_1 \\ \vdots \\ \hat{\Delta}s_N \end{bmatrix} = -\begin{bmatrix} F_1 \\ \vdots \\ F_N \end{bmatrix} \iff \begin{cases} G_1\hat{\Delta}s_1 - \hat{\Delta}s_2 = -F_1 \\ G_2\hat{\Delta}s_2 - \hat{\Delta}s_3 = -F_2 \\ \vdots \\ G_{N-1}\hat{\Delta}s_{N-1} - \hat{\Delta}s_N = -F_{N-1} \\ A\hat{\Delta}s_1 + B\hat{\Delta}s_N = -F_N \end{cases}$$

De donde finalmente obtenemos las ecuaciones a resolver:

$$\begin{cases} \hat{\Delta}s_2 = G_1\hat{\Delta}s_1 + F_1 \\ \hat{\Delta}s_3 = G_2\hat{\Delta}s_2 + F_2 \\ \vdots \\ \hat{\Delta}s_N = G_{N-1}\hat{\Delta}s_{N-1} + F_{N-1} \\ A\hat{\Delta}s_1 + B\hat{\Delta}s_N = -F_N \end{cases}$$

Problema. Tenemos el problema:

$$y'' = f(x,y) = q(x)y - g(x)$$
$$y(a) = \alpha \quad y(b) = \beta$$

El método es el siguiente:

$$y_{n+1} - 2y_n + y_{n-1} = h^2 f_n$$
$$x_n = a + hn \quad n = 0, 1 \dots N+1$$
$$y_{n+1} - 2y_n + y_{n-1} = h^2(q(x_n)y_n - g(x_n)) \quad n = 1 \dots N$$

A partir de aquí, las ecuaciones que aparecen son:

$$y_2 - (2 + h^2 q(x_1))y_1 + y_0 = -g(x_1)h^2$$
$$y_3 - (2 + h^2 q(x_2))y_2 + y_0 = -g(x_2)h^2$$
$$\vdots$$

Esto se puede resumir en la matriz A tal que $Av = w$:

$$A = \begin{pmatrix} 2+q_1h^2 & -1 & 0 & \dots & 0 \\ -1 & 2+q_2h^2 & -1 & \dots & 0 \\ 0 & -1 & & \ddots & \\ \vdots & & \ddots & & -1 \\ 0 & & & -1 & 2+q_Nh^2 \end{pmatrix}$$