
Capítulo 2

El problema de valor inicial: métodos de Runge-Kutta

2.1. Métodos de Taylor

El método de Euler es tan sólo de primer orden. Eso supone un inconveniente. Si queremos un error muy pequeño nos veremos forzados a tomar h muy pequeño, lo que supondrá tener que dar muchos pasos, con un coste computacional que puede ser inaceptable. La regla del trapecio es de segundo orden, pero a cambio es implícita. ¿Podemos construir métodos “mejores”, de mayor orden que el de Euler, y que a la vez sean explícitos?

Una idea es tomar más términos en el desarrollo de Taylor, lo que da lugar a los llamados *métodos de Taylor*. Como ejemplo construimos el método de Taylor de orden 2, en el caso $d = 1$.

El desarrollo de Taylor produce

$$y(x_{n+1}) = y(x_n) + h_n y'(x_n) + \frac{h_n^2}{2} y''(x_n) + \frac{h_n^3}{3!} y'''(\xi_n).$$

El término de error, de orden h_n^3 , se desprecia. Para calcular $y'(x_n)$ e $y''(x_n)$ usamos la ecuación diferencial,

$$\begin{aligned}y'(x) &= f(x, y(x)), \\y''(x) &= f_x(x, y(x)) + f_y(x, y(x))y'(x) = f_x(x, y(x)) + f_y(x, y(x))f(x, y(x)).\end{aligned}$$

Obtenemos el método

$$y_{n+1} = y_n + h_n f(x_n, y_n) + \frac{h_n^2}{2} (f_x(x_n, y_n) + f_y(x_n, y_n) f(x_n, y_n)).$$

En este caso el error de truncación $R_n = \frac{h_n^3}{3!} y'''(\xi_n)$ satisface que

$$\|R_n\| \leq Ch^3,$$

y el método será de orden 2 (habría que probar que la función de incremento es continua y satisface una condición de Lipschitz con respecto a su segunda variable; esto es cierto si, por ejemplo, f tiene derivadas parciales hasta orden 2 acotadas).

Por el mismo procedimiento podemos generar métodos de Taylor del orden que queramos. El inconveniente es que involucran cada vez más derivadas.

Comparemos con el método de Euler. Los métodos de Taylor tienen mayor orden, luego habrá que dar menos pasos. A cambio, dar un paso es computacionalmente más caro. No sólo hay que evaluar f , sino también sus derivadas, lo que puede ser muy costoso, sobre todo si éstas son difíciles. Así, si no queremos mucha precisión y la función es complicada, Euler es preferible. Si, por el contrario, queremos bastante precisión y la función y sus derivadas no son complicadas, es preferible usar un método de Taylor de orden más alto. En resumen: no hay un método que sea “el mejor”. Cuál sea el más conveniente depende de la ecuación diferencial y de la precisión que queramos conseguir.

2.2. Métodos de Runge-Kutta

Los métodos de Taylor son conceptualmente sencillos. Sin embargo:

- Exigen calcular derivadas de orden superior, lo que puede ser engorroso.
- Exigen evaluar dichas derivadas, lo que puede ser computacionalmente costoso.

Para evitar estas dificultades, los métodos de Runge-Kutta evalúan $f(x, y)$ en más puntos, intentando igualar la precisión de los métodos de Taylor. Concretamente toman

$$\phi(x_n, y_n; h_n) = \sum_{i=1}^s b_i f(\theta_i, \gamma_i)$$

de forma que $\phi(x_n, y_n; h_n)$ sea una media ponderada de las pendientes cerca de (x_n, y_n) . La idea es elegir los parámetros b_i, θ_i, γ_i de manera que la función de incremento coincida lo más posible con la función de incremento de un método de Taylor.

Veamos por ejemplo cómo hacer esto con dos evaluaciones de la función y $d = 1$. Parece razonable tomar $(\theta_1, \gamma_1) = (x_n, y_n)$. Para θ_2 tomamos $\theta_2 = x_n + \alpha h_n$. Una elección sensata para γ_2 consiste en estimar el valor $y(\theta_2)$ avanzando con el método de Euler, es decir

$$\gamma_2 = y_n + \alpha h_n f(x_n, y_n).$$

Queda por tanto

$$y_{n+1} = y_n + h_n (b_1 f(x_n, y_n) + b_2 f(x_n + \alpha h_n, y_n + \alpha h_n f(x_n, y_n))).$$

¿Cómo elegir b_1, b_2 y α ? Desarrollamos por Taylor en potencias de h_n y obtenemos

$$\begin{aligned} y_{n+1} = & y_n + (b_1 + b_2)h_n f(x_n, y_n) \\ & + \alpha b_2 h_n^2 (f_x(x_n, y_n) + f_y(x_n, y_n)f(x_n, y_n)) + O(h_n^3). \end{aligned}$$

Si hacemos

$$b_1 + b_2 = 1, \quad \alpha b_2 = 1/2,$$

entonces la función de incremento ϕ coincide hasta orden uno con la función de incremento ϕ^{T2} del método de Taylor de orden 2,

$$\phi(x_n, y_n; h_n) = \phi^{T2}(x_n, y_n; h_n) + O(h_n^2).$$

Veamos cuál es el orden de consistencia del método:

$$\begin{aligned} R_n &= y(x_{n+1}) - y(x_n) - h_n \phi(x_n, y(x_n); h_n) \\ &= y(x_{n+1}) - y(x_n) - h_n \phi^{T2}(x_n, y(x_n); h_n) + O(h_n^3) \\ &= O(h_n^3). \end{aligned}$$

Concluimos que $\tau = O(h^2)$, es decir, el método es consistente de orden 2. Por tanto, para ver que el método es convergente de orden 2 basta con comprobar que la función de incremento

$$\phi(x, y; h) = b_1 f(x, y) + b_2 f(x + \alpha h, y + \alpha h f(x, y)),$$

satisface una condición de Lipschitz con respecto a su segunda variable. Esto se prueba fácilmente usando que f es Lipschitz en y . En efecto,

$$\begin{aligned} |\phi(x, y; h) - \phi(x, \hat{y}; h)| &= |b_1| |f(x, y) - f(x, \hat{y})| \\ &\quad + |b_2| |f(x + \alpha h, y + \alpha h f(x, y)) - f(x + \alpha h, \hat{y} + \alpha h f(x, \hat{y}))| \\ &\leq |b_1| L |y - \hat{y}| + |b_2| L |y - \hat{y} + \alpha h (f(x, y) - f(x, \hat{y}))| \\ &\leq (|b_1| + |b_2|) L |y - \hat{y}| + |b_2| L^2 |\alpha| h |y - \hat{y}| \\ &\leq K |y - \hat{y}|, \end{aligned}$$

para todo $h < 1$, donde $K = (|b_1| + |b_2|)L + |b_2|L^2|\alpha|$.

Hemos obtenido $b_2 = \frac{1}{2\alpha}$, $b_1 = 1 - \frac{1}{2\alpha}$ y $\alpha \neq 0$ arbitrario. No tenemos un único método, sino una familia uniparamétrica, con parámetro α .

Ejemplos.

- Tomando $\alpha = 1/2$ obtenemos el llamado método de *Euler modificado*,

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

- Tomando $\alpha = 1$ se tiene el método de *Euler mejorado*, también conocido como método de *Heun de orden 2*,

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

En ambos casos se realizan dos evaluaciones de función, frente a las tres necesarias en el método de Taylor de orden 2 (a saber, $f(x_n, y_n)$, $f_x(x_n, y_n)$ y $f_y(x_n, y_n)$) cuando $d = 1$, siendo la cosa aún peor para problemas de dimensión mayor.

El método de Euler modificado se puede escribir alternativamente como

$$\begin{aligned} k_1 &= f(x_n, y_n), \\ k_2 &= f(x_n + \frac{h_n}{2}, y_n + \frac{h_n}{2}k_1), \\ y_{n+1} &= y_n + h_n k_2. \end{aligned}$$

En cuanto a Euler mejorado, se puede escribir como

$$\begin{aligned} k_1 &= f(x_n, y_n), \\ k_2 &= f(x_n + h_n, y_n + h_n k_1), \\ y_{n+1} &= y_n + h_n (\frac{k_1}{2} + \frac{k_2}{2}). \end{aligned}$$

Siguiendo estos ejemplos, definimos un *método de Runge-Kutta de s etapas* por

$$\begin{cases} k_i = f(x_n + c_i h_n, y_n + h_n \sum_{j=1}^s a_{ij} k_j), & i = 1, 2, \dots, s \\ y_{n+1} = y_n + h_n \sum_{i=1}^s b_i k_i. \end{cases} \quad (2.1)$$

Cada una de las evaluaciones de función k_i es una *etapa*.

El método (2.1) se representa por medio de su *tablero de Butcher*:

c_1	a_{11}	a_{12}	\dots	a_{1s}
c_2	a_{21}	a_{22}	\dots	a_{2s}
\vdots	\vdots	\vdots	\ddots	\vdots
c_s	a_{s1}	a_{s2}	\dots	a_{ss}
	b_1	b_2	\dots	b_s

Si escribimos

$$c = (c_1, c_2, \dots, c_s)^T, \quad b = (b_1, b_2, \dots, b_s)^T, \quad A = (a_{ij}),$$

el tablero se puede resumir como

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

Si $a_{ij} = 0$ para $j \geq i$, $i = 1, 2, \dots, s$, es decir, si la matriz A es triangular inferior estricta, entonces cada uno de los k_i viene dado explícitamente en términos de los anteriormente calculados, k_j , $j = 1, 2, \dots, i - 1$. En este caso se dice que el método es *explícito*. Al escribir su tablero se suelen omitir los ceros sobre y por encima de la diagonal principal.

Ejemplos.

- El método de Euler modificado es explícito. Su tablero es

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

- El método de Euler mejorado también es explícito. Su tablero es

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

Si el método no es explícito, es *implícito*. En general es necesario entonces resolver en cada paso un sistema no lineal para calcular los k_i . Este sistema tiene dimensión ds . Hay una situación intermedia; si $a_{ij} = 0$ para $j > i$, $i = 1, 2, \dots, s$, entonces cada k_i está definido individualmente por

$$k_i = f(x_n + c_i h_n, y_n + h_n \sum_{j=1}^i a_{ij} k_j), \quad i = 1, 2, \dots, s.$$

En vez de resolver en cada paso un sistema no lineal de dimensión ds , tendremos que resolver s sistemas no acoplados de dimensión d . Estos métodos se llaman *semi-implícitos*. Al escribir sus tableros se omiten los ceros por encima de la diagonal principal.

2.3. Condiciones de orden

Los métodos de Runge-Kutta son métodos de un paso con función de incremento

$$\phi(x_n, y_n; h_n) = \sum_{i=1}^s b_i k_i.$$

Si hacemos $h_n = 0$, entonces $k_i = f(x_n, y_n)$ para todo $i = 1, 2, \dots, s$. Así,

$$\phi(x, y; 0) = \left(\sum_{i=1}^s b_i \right) f(x, y).$$

Por tanto, un método de Runge-Kutta es consistente si y sólo si

$$\sum_{i=1}^s b_i = 1. \quad (2.2)$$

Por otra parte, puesto que las etapas k_i son evaluaciones de la función f , no es difícil convencerse de que ϕ satisface una condición de Lipschitz con respecto a su segunda variable si f satisface una condición de Lipschitz en y . Así pues, la condición de consistencia (2.2) es suficiente para garantizar la convergencia. Veamos que también es necesaria.

Teorema 2.1. *Un método de Runge-Kutta es convergente si y sólo si es consistente.*

Prueba. Si es convergente, en particular lo es para el problema

$$y'(x) = 1, \quad y(0) = 0,$$

cuya solución es $y(x) = x$. El método, aplicado a este problema con paso fijo h , se reduce a

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

Tomando como valor de arranque $y_0 = 0$ tenemos que la solución numérica viene dada por

$$y_n = nh \sum_{i=1}^s b_i = x_n \sum_{i=1}^s b_i.$$

Por tanto

$$y(x_n) - y_n = x_n \left(1 - \sum_{i=1}^s b_i \right),$$

y la solución numérica converge a la teórica si y sólo si se cumple la condición de consistencia (2.2). \square

¿Qué condiciones sobre los coeficientes garantizan que el método es convergente de orden p ? Tenemos que desarrollar el residuo

$$R_n = y(x_{n+1}) - (y(x_n) + h_n \phi(x_n, y(x_n); h_n))$$

alrededor de x_n en potencias de h_n . Nótese que el residuo coincide con la diferencia entre la solución teórica y la numérica,

$$R_n = y(x_{n+1}) - y_{n+1},$$

suponiendo que en el paso n no se cometía ningún error, $y(x_n) = y_n$ (esta hipótesis se conoce como *hipótesis de localización*).

Así pues, lo que hay que hacer es obtener desarrollos en potencias de h_n para las soluciones teórica y numérica (esta última bajo la hipótesis de localización). Si estos desarrollos coinciden hasta orden p , entonces $R_n = O(h_n^{p+1})$ y tendremos convergencia de orden p .

Para que los desarrollos de las soluciones teórica y numérica sean iguales hasta orden p , es necesario que sus derivadas hasta orden p en $h_n = 0$ coincidan. Eso dará lugar a unas ciertas condiciones sobre los coeficientes del método, las llamadas *condiciones de orden*. Obtenerlas será nuestro siguiente objetivo.

Notación. Los superíndices en mayúsculas denotarán componentes, y variarán por tanto entre 1 y d . Los subíndices en minúsculas denotan etapas; varían entre 1 y s . Finalmente, los subíndices en mayúsculas significan derivación parcial, y también varían entre 1 y d . Ejemplo:

$$f_{KL}^J = \frac{\partial f^J}{\partial y^K \partial y^L}.$$

Si el sistema es autónomo, $y' = f(y)$, podemos obtener más simetría en las fórmulas sustituyendo k_i por g_i tal que $k_i = f(g_i)$. De esta forma el método (2.1) se transforma (para sistemas autónomos) en

$$\begin{cases} g_i = y_n + h_n \sum_{j=1}^s a_{ij} f(g_j), & i = 1, \dots, s, \\ y_{n+1} = y_n + h_n \sum_{j=1}^s b_j f(g_j). \end{cases} \quad (2.3)$$

Así, la única diferencia entre las definiciones de g_i e y_{n+1} es que las constantes a_{ij} en la primera fórmula pasan a ser las constantes b_j en la segunda. Esta semejanza nos permitirá simplificar notablemente algunos cálculos.

Si el sistema no es autónomo, pero el método de Runge-Kutta es *consistente* y satisface la *condición de suma por filas*

$$c_i = \sum_{j=1}^s a_{ij}, \quad i = 1, 2, \dots, s, \quad (2.4)$$

el método también se puede escribir en la forma (2.3). Para ello basta con definir

$$\begin{aligned} \hat{g}_n^1 &= x_n, & \hat{f}^1 &= 1, & \hat{g}_n^1 &= x_n + c_i h_n, \\ \hat{g}_n^{J+1} &= y_n^J, & \hat{f}^{J+1} &= f^J, & \hat{g}_n^{J+1} &= y_n^J + h_n \sum_{j=1}^s a_{ij} k_j^J, \quad J = 1, \dots, d. \end{aligned}$$

En lo que sigue escribiremos siempre los métodos en la forma (2.3). Por ello, para no dejar fuera los sistemas no autónomos, de aquí en adelante nos restringiremos a métodos que cumplan la condición de suma por filas.

Como ejemplo obtendremos las condiciones de orden 3. Las condiciones para tener un orden más alto se obtienen de manera similar. Tenemos que comparar las derivadas hasta orden 3 con respecto a h_n de las soluciones numérica, y_{n+1} , y teórica, $y(x_{n+1}) = y(x_n + h_n)$, evaluadas en $h_n = 0$.

Empezamos calculando las derivadas de la solución numérica (siempre bajo la hipótesis de localización). Aplicando la regla de Leibniz para la derivación

de un producto se tiene que

$$y_{n+1}^{(q)} \Big|_{h_n=0} = \sum_{j=1}^s b_j q (f(g_j))^{(q-1)} \Big|_{h_n=0}. \quad (2.5)$$

Así, la derivada primera de la solución numérica es

$$y_{n+1}^{(1)} \Big|_{h_n=0} = \sum_{j=1}^s b_j (f(g_j)) \Big|_{h_n=0} = \left(\sum_{j=1}^s b_j \right) f(y(x_n)). \quad (2.6)$$

Nótese que en la última igualdad hemos usado la hipótesis de localización. La derivada segunda viene dada por

$$y_{n+1}^{(2)} \Big|_{h_n=0} = 2 \sum_{j=1}^s b_j (f(g_j))^{(1)} \Big|_{h_n=0}.$$

Pero

$$(f(g_j))^{(1)} = \sum_{K=1}^d f_K(g_j) (g_j^K)^{(1)}.$$

Usando la semejanza de las definiciones de g_i e y_{n+1} junto con (2.6) se tiene que

$$(g_i^J)^{(1)} \Big|_{h_n=0} = \left(\sum_{j=1}^s a_{ij} \right) f^J(y(x_n)), \quad (2.7)$$

por lo que concluimos que

$$y_{n+1}^{(2)} \Big|_{h_n=0} = 2 \sum_{j,k=1}^s b_j a_{jk} \sum_{K=1}^d f_K(y(x_n)) f^K(y(x_n)). \quad (2.8)$$

Para calcular la derivada tercera necesitamos la derivada segunda de $f(g_j)$, que viene dada por

$$(f(g_j))^{(2)} = \sum_{K,L=1}^d f_{KL}(g_j) (g_j^K)^{(1)} (g_j^L)^{(1)} + \sum_{K=1}^d f_K(g_j) (g_j^K)^{(2)}.$$

La derivada segunda de g_j^K se obtiene utilizando de nuevo la semejanza entre g_i e y_{n+1} . Finalmente llegamos a

$$\begin{aligned} y_{n+1}^{(3)} \Big|_{h_n=0} &= 3 \sum_{j,k,l=1}^s b_j a_{jk} a_{jl} \sum_{K,L=1}^d f_{KL}(y(x_n)) f^K(y(x_n)) f^L(y(x_n)) \\ &\quad + 6 \sum_{j,k,l=1}^s b_j a_{jk} a_{kl} \sum_{K,L=1}^d f_K(y(x_n)) f_L^K(y(x_n)) f^L(y(x_n)). \end{aligned} \quad (2.9)$$

En cuanto a la solución teórica, se tiene que

$$\begin{aligned}\left. \frac{d(y(x_{n+1}))}{dh_n} \right|_{h_n=0} &= f(y(x_n + h_n))|_{h_n=0} = f(y(x_n)), \\ \left. \frac{d^2(y(x_{n+1}))}{dh_n^2} \right|_{h_n=0} &= \sum_{K=1}^d f_K(y(x_n))(y^K)^{(1)}(x_n) = \sum_{K=1}^d f_K(y(x_n))f^K(y(x_n)), \\ \left. \frac{d^3(y(x_{n+1}))}{dh_n^3} \right|_{h_n=0} &= \sum_{K,L=1}^d f_{KL}(y(x_n))f^K(y(x_n))(y^L)^{(1)}(x_n) \\ &\quad + \sum_{K,L=1}^d f_K(y(x_n))f_L^K(y(x_n))(y^L)^{(1)}(y(x_n)). \\ &= \sum_{K,L=1}^d f_{KL}(y(x_n))f^K(y(x_n))f^L(y(x_n)) \\ &\quad + \sum_{K,L=1}^d f_K(y(x_n))f_L^K(y(x_n))f^L(y(x_n)).\end{aligned}$$

Comparando las derivadas de la solución numérica y de la teórica obtenemos las *condiciones de orden 3*:

- Orden 1: $\sum_{j=1}^s b_j = 1,$
- Orden 2: $2 \sum_{j,k=1}^s b_j a_{jk} = 1,$
- Orden 3: $3 \sum_{j,k,l=1}^s b_j a_{jk} a_{jl} = 1, \quad 6 \sum_{j,k,l=1}^s b_j a_{jk} a_{kl} = 1.$

Un método de Runge-Kutta tendrá orden de consistencia 3 si cumple estas cuatro condiciones. Se puede demostrar, aunque no lo haremos aquí, que estas condiciones son también necesarias.

Observación. La condición de consistencia coincide con la condición de consistencia de orden 1. Así pues, si un método de Runge-Kutta es convergente, converge con orden al menos uno.

Observación. Algunas de estas condiciones se pueden simplificar usando la

condición de suma por filas, quedando

$$\begin{aligned}
\bullet \text{ Orden 1: } & \sum_{j=1}^s b_j = 1, \\
\bullet \text{ Orden 2: } & 2 \sum_{j=1}^s b_j c_j = 1, \\
\bullet \text{ Orden 3: } & 3 \sum_{j=1}^s b_j c_j^2 = 1, \quad 6 \sum_{j,k=1}^s b_j a_{jk} c_k = 1.
\end{aligned} \tag{2.10}$$

Para tener orden 4 se tienen que cumplir además las siguientes condiciones,

$$\begin{aligned}
4 \sum_{j=1}^s b_j c_j^3 &= 1, & 8 \sum_{j,l=1}^s b_j c_j a_{jl} c_l &= 1, \\
12 \sum_{j,k=1}^s b_j a_{jk} c_k^2 &= 1, & 24 \sum_{j,k,l=1}^s b_j a_{jk} a_{kl} c_l &= 1,
\end{aligned}$$

que se obtienen de manera análoga.

Observación. Las sumas que aparecen en las condiciones de orden se pueden interpretar como productos de matrices, lo que puede ser cómodo a la hora de hacer cálculos. Por ejemplo, la condición $6 \sum_{j,k=1}^s b_j a_{jk} c_k = 1$ se puede reescribir como $6 b^T A c = 1$.

Ejemplo. El método de Heun de orden 3, de tablero

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

cumple la condición de suma por filas y las cuatro condiciones (2.10). Es por tanto de orden 3.

2.4. Métodos explícitos. Orden obtenible

Hemos obtenido las condiciones para que un método de Runge-Kutta tenga un orden dado. Ahora dirigimos nuestra atención al problema de encontrar

soluciones a dichas condiciones de orden. Nos preguntamos qué orden se puede lograr. Nuestro primer resultado da una cota superior para el orden que se puede conseguir con un método *explícito* de s etapas.

Teorema 2.2. *Un método de Runge-Kutta explícito de s etapas no puede tener orden mayor que s .*

Prueba. Aplicamos el método al problema escalar

$$y'(x) = y(x), \quad y(0) = 1,$$

cuya solución es $y(x) = e^x$. Así pues, las derivadas de la solución teórica vienen dadas por

$$\left. \frac{d^p(y(x_n + h_n))}{dh_n^p} \right|_{h_n=0} = e^{x_n+h_n} \Big|_{h_n=0} = e^{x_n} = y(x_n).$$

En cuanto a la solución numérica, aplicando la regla de Leibniz para la derivada de un producto obtenemos que

$$y_{n+1}^{(p)} \Big|_{h_n=0} = p \sum_{j_1=1}^s b_{j_1} k_{j_1}^{(p-1)} \Big|_{h_n=0}.$$

Por otra parte,

$$\begin{aligned} k_{j_1}^{(p-1)} \Big|_{h_n=0} &= (p-1) \sum_{j_2=1}^s a_{j_1 j_2} k_{j_2}^{(p-2)} \Big|_{h_n=0} \\ &= (p-1)(p-2) \sum_{j_2, j_3=1}^s a_{j_1 j_2} a_{j_2 j_3} k_{j_3}^{(p-3)} \Big|_{h_n=0} \\ &= \dots \\ &= (p-1)! \sum_{j_2, j_3, \dots, j_p=1}^s a_{j_1 j_2} a_{j_2 j_3} \dots a_{j_{p-1} j_p} k_{j_p}^{(0)} \Big|_{h_n=0}. \end{aligned}$$

Por consiguiente, asumiendo la hipótesis de localización,

$$y_{n+1}^{(p)} \Big|_{h_n=0} = p! \sum_{j_1, \dots, j_p=1}^s b_{j_1} a_{j_1 j_2} a_{j_2 j_3} \dots a_{j_{p-1} j_p} y(x_n).$$

Concluimos que la condición

$$p! \sum_{j_1, \dots, j_p=1}^s b_{j_1} a_{j_1 j_2} a_{j_2 j_3} \dots a_{j_{p-1} j_p} = 1 \quad (2.11)$$

es necesaria para que el método sea de orden p .

Si el método es explícito, $a_{ij} = 0$ para $j \geq i$. Así, a menos que exista una sucesión j_1, j_2, \dots, j_p de enteros $1, 2, \dots, s$ tal que $j_1 > j_2 > j_3 > \dots > j_p$, se tendrá que $\sum_{j_1, \dots, j_p=1}^s b_{j_1} a_{j_1 j_2} a_{j_2 j_3} \dots a_{j_{p-1} j_p} = 0$, y la condición (2.11) no se podrá cumplir. Dicha sucesión sólo puede existir si $p \leq s$. \square

¿Se puede conseguir orden $p = s$? Para $s = 1, 2, 3, 4$ se puede ver que sí, y es relativamente fácil encontrar familias de soluciones a las correspondientes condiciones de orden.

Ejemplo. Sea $s = 1$. La condición de orden 1 es en este caso $b_1 = 1$. El único método explícito de una etapa y orden 1 es por tanto el método de Euler.

Ejemplo. Sea $s = 2$. Las condiciones de orden 1 y 2 son en este caso $b_1 + b_2 = 1$ y $b_1 c_1 + b_2 c_2 = 1$ respectivamente (estamos suponiendo que se cumple la condición de suma por filas). Si el método es explícito entonces $c_1 = 0$, y la condición de orden 2 queda que $b_2 c_2 = 1/2$. Hay pues una familia uniparamétrica de métodos explícitos de 2 etapas y orden 2,

$$c_2 = \lambda, \quad b_1 = 1 - \frac{1}{2\lambda}, \quad b_2 = \frac{1}{2\lambda}, \quad \lambda \in \mathbb{R}, \lambda \neq 0.$$

¿Qué pasa con $s = 5$? Para que un método sea de orden 5 se tienen que satisfacer 17 condiciones de orden. Si $s = 5$ tenemos sólo 15 parámetros libres. Sin embargo, el sistema es no lineal, así que podría haber solución. Pero no es así.

Teorema 2.3. *Para $p \geq 5$ no existe ningún método de Runge-Kutta explícito de orden p con $s = p$ etapas.*

Este es un primer ejemplo de *barrera de Butcher*. Hay más barreras. Por ejemplo, se tiene el siguiente teorema.

Teorema 2.4. (i) Para $p \geq 7$ no existe ningún método de Runge-Kutta explícito de orden p con $s = p + 1$ etapas.

(ii) Para $p \geq 8$ no hay ningún método de Runge-Kutta explícito con $s = p + 2$ etapas.

En la siguiente tabla recogemos el estado del arte:

Orden	1	2	3	4	5	6	7	8	9	10
s mínimo	1	2	3	4	6	7	9	11	$12 \leq s \leq 17$	$13 \leq s \leq 17$

2.5. A-estabilidad de métodos de Runge-Kutta

Para estudiar el dominio de estabilidad lineal \mathcal{D} tenemos que aplicar el método de Runge-Kutta (2.1) con paso fijo h al problema escalar

$$y' = \lambda y, \quad x \geq 0, \quad y(0) = 1. \quad (2.12)$$

Para ello conviene reescribir (2.1) en la forma alternativa

$$\begin{cases} Y_i = y_n + h \sum_{j=1}^s a_{ij} f(x_n + c_j h, Y_j), \\ y_{n+1} = y_n + h \sum_{i=1}^s b_i f(x_n + c_i h, Y_i). \end{cases}$$

Obtenemos que

$$\begin{cases} Y_i = y_n + \lambda h \sum_{j=1}^s a_{ij} Y_j, \\ y_{n+1} = y_n + \lambda h \sum_{i=1}^s b_i Y_i. \end{cases} \quad (2.13)$$

Sean $Y, e \in \mathbb{R}^s$ dados por

$$Y = (Y_1, \dots, Y_s)^T, \quad e = (1, 1, \dots, 1)^T.$$

Entonces podemos escribir (2.13) en la forma

$$\begin{cases} Y = e y_n + \lambda h A Y, \\ y_{n+1} = y_n + \lambda h b^T Y. \end{cases}$$

La primera de estas ecuaciones da

$$Y = (I - \lambda h A)^{-1} e y_n,$$

que introducido en la segunda produce

$$y_{n+1} = (1 + \lambda h b^T (I - \lambda h A)^{-1} e) y_n, \quad n = 0, 1, \dots,$$

es decir,

$$y_{n+1} = R(\lambda h) y_n,$$

donde

$$R(z) = 1 + z b^T (I - z A)^{-1} e, \quad z \in \mathbb{C}. \quad (2.14)$$

La función $R(z)$ recibe el nombre de *función de estabilidad del método*, o también *función de amplificación*.

Obsérvese que $y_n = (R(\lambda h))^n$, así que de la definición del dominio de estabilidad lineal \mathcal{D} se sigue inmediatamente que

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

Por tanto un método será A -estable si y sólo si $|R(z)| < 1$ en \mathbb{C}^- .

Notación. \mathbb{P}_α es el conjunto de todos los polinomios de grado menor o igual que α . $\mathbb{P}_{\alpha/\beta}$ es el conjunto de todas las funciones racionales \hat{p}/\hat{q} , con $\hat{p} \in \mathbb{P}_\alpha$ y $\hat{q} \in \mathbb{P}_\beta$.

Un sencillo análisis permite probar el siguiente lema.

Lema 2.5. *Para todo método de Runge-Kutta existe $R \in \mathbb{P}_{s/s}$ tal que la solución por dicho método del problema (2.12) viene dada por*

$$y_n = (R(\lambda h))^n.$$

Es más, si el método de Runge-Kutta es explícito, entonces $R \in \mathbb{P}_s$.

Prueba. Tenemos que probar que la función R dada por (2.14) pertenece a $\mathbb{P}_{s/s}$. Usamos que

$$(I - zA)^{-1} = \frac{(\text{adj}(I - zA))^T}{\det(I - zA)}.$$

Cada elemento de $I - zA$ es lineal en z . Por tanto, como cada elemento de $\text{adj}(I - zA)$ es el determinante de una submatriz $(s - 1) \times (s - 1)$, está en \mathbb{P}_{s-1} . Así pues,

$$b^T(\text{adj}(I - zA))^T e \in \mathbb{P}_{s-1}.$$

Como $\det(I - zA) \in \mathbb{P}_s$, concluimos que $R \in \mathbb{P}_{s/s}$.

Si el método es explícito, entonces A es estrictamente triangular inferior y $I - zA$ es una matriz triangular inferior con unos a lo largo de la diagonal. Por tanto $\det(I - zA) = 1$, y R es un polinomio. \square

Corolario 2.6. *Ningún método de Runge-Kutta explícito puede ser A -estable.*

Prueba. La función de estabilidad de un método de Runge-Kutta explícito es un polinomio. Además $R(0) = 1$. Por otra parte, ningún polinomio, salvo la función constante $R(z) = c$, $|c| < 1$, está estrictamente acotado por uno en \mathbb{C}^- . \square

En el caso de métodos explícitos podemos decir aún más sobre la función de amplificación.

Teorema 2.7. *Si un método de Runge-Kutta explícito tiene orden de consistencia p , entonces*

$$R(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^p}{p!} + O(z^{p+1}).$$

Prueba. La solución exacta del (2.12) es $y(x) = e^{\lambda x}$. Por ser el método consistente de orden p tenemos que

$$y(h) - y_1 = O(h^{p+1}),$$

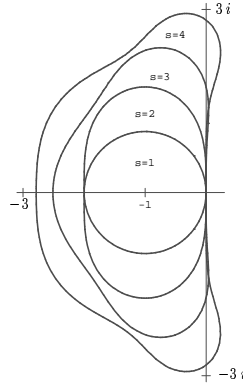


Figura 2.1: Dominio de estabilidad lineal de los métodos de Runge-Kutta explícitos de orden $p = s$.

así que

$$y_1 = e^z + O(h^{p+1}) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^p}{p!} + O(z^{p+1}),$$

donde $z = \lambda h$. Usando que $y_1 = R(z)y_0$ concluimos inmediatamente el resultado. \square

Como consecuencia tenemos que todos los métodos de Runge-Kutta explícitos con $p = s$ tienen como función de estabilidad

$$R(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^s}{s!}.$$

En la figura 2.1 mostramos los correspondientes dominios de estabilidad lineal.

¿Qué podemos decir de los métodos implícitos? Veamos con un ejemplo que pueden ser A -estables.

Ejemplo. Sea el método de dos etapas y orden 3 dado por el tablero

0	$\frac{1}{4}$	$-\frac{1}{4}$
$\frac{2}{3}$	$\frac{1}{4}$	$\frac{5}{12}$
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$

Su función de amplificación es

$$R(z) = \frac{1 + \frac{1}{3}z}{1 - \frac{2}{3}z + \frac{1}{6}z^2}.$$

Veamos si es A -estable. Representamos $z \in \mathbb{C}^-$ en coordenadas polares, $z = \rho e^{i\theta}$, con $\rho > 0$, $|\theta - \pi| < \pi/2$, y nos preguntamos si $|R(\rho e^{i\theta})| < 1$. Esto es equivalente a

$$\left|1 + \frac{1}{3}\rho e^{i\theta}\right|^2 < \left|1 - \frac{2}{3}\rho e^{i\theta} + \frac{1}{6}\rho^2 e^{2i\theta}\right|^2,$$

y por tanto a

$$1 + \frac{2}{3}\rho \cos \theta + \frac{1}{9}\rho^2 < 1 - \frac{4}{3}\rho \cos \theta + \rho^2 \left(\frac{1}{3}\cos 2\theta + \frac{4}{9}\right) - \frac{2}{9}\rho^3 \cos \theta + \frac{1}{36}\rho^4.$$

Reordenando términos, la condición $\rho e^{i\theta} \in \mathcal{D}$ pasa a ser

$$2\rho \left(1 + \frac{1}{9}\rho^2\right) \cos \theta < \frac{1}{3}\rho^2 (1 + \cos 2\theta) + \frac{1}{36}\rho^4 = \frac{2}{3}\rho^2 \cos^2 \theta + \frac{1}{36}\rho^4,$$

que se satisface para todos los $z \in \mathbb{C}^-$, puesto que $\cos \theta < 0$ para todo z de este tipo. El método es por tanto A -estable.

Esto explica por qué son interesantes los métodos implícitos. Son más caros, pero, además de permitir romper las barreras de Butcher, algunos de ellos tienen regiones de estabilidad grandes, y pueden ser convenientes a la hora de tratar problemas *stiff*.

Sin embargo, no todos los métodos implícitos son A -estables.

Ejemplo. Sea el método de dos etapas y orden 1 definido por el tablero

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

Su función de amplificación es

$$R(z) = \frac{1 + \frac{2}{3}z + \frac{1}{6}z^2}{1 - \frac{1}{3}z}.$$

Así pues, $\lim_{|z| \rightarrow \infty} |R(z)| = \infty$, por lo que el método no es A -estable.

En realidad no hace falta comprobar todos los $z \in \mathbb{C}^-$ para ver si una función racional dada tiene su origen en un método A -estable (una $R(z)$ tal se dice A -aceptable). Ése es el contenido del siguiente lema, cuya prueba es una aplicación del principio del máximo.

Lema 2.8. *Sea $R(z)$ una función racional no constante. Entonces $|R(z)| < 1$ para todo $z \in \mathbb{C}^-$ si y sólo si todos los polos de R tienen parte real positiva y $|R(it)| \leq 1$ para todo $t \in \mathbb{R}$.*

Prueba. Si $|R(z)| < 1$ para todo $z \in \mathbb{C}^-$, por continuidad $|R(z)| \leq 1$ para todo $z \in \overline{\mathbb{C}^-}$. En particular R no puede tener polos en el semiplano cerrado izquierdo, y $|R(it)| \leq 1$ para todo $t \in \mathbb{R}$.

Para probar el recíproco observamos que si los polos están a la derecha del eje imaginario, entonces la función racional R es analítica en el cerrado $\overline{\mathbb{C}^-}$. Por tanto, al no ser R constante, $|R|$ no puede tener un máximo en el interior, y alcanza su máximo en la frontera. Parte de la frontera está en el infinito. Sin embargo, por ser R una función racional existe el límite (finito o infinito) $\lim_{|z| \rightarrow \infty} |R(z)|$, y es igual a $\lim_{t \rightarrow \infty} |R(it)|$. Por tanto, si $|R(it)| \leq 1$, $t \in \mathbb{R}$, entonces $|R(z)| < 1$ para todo $z \in \mathbb{C}^-$. \square

Ejemplo. Consideramos el método de Gauss-Legendre de dos etapas (orden 4),

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

Un sencillo cálculo muestra que

$$R(z) = \frac{1 + \frac{1}{2}z + \frac{1}{12}z^2}{1 - \frac{1}{2}z + \frac{1}{12}z^2}.$$

Los polos de esta función, $3 \pm i\sqrt{3}$, están en el semiplano abierto derecho, y $|R(it)| = 1$, $t \in \mathbb{R}$, por lo que el método es A -estable.

2.6. Pares encajados

Nos gustaría tener una forma “barata” de estimar el error local. Esto permitirá al método elegir el paso de manera que dicho error sea menor que una tolerancia dada. Seguiremos una idea debida a Merson.

Sea un método de Runge-Kutta de orden p de tablero

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

La solución numérica por este método es

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

Si tuviéramos otro método de Runge-Kutta con las mismas etapas,

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

pero de distinto orden, $\hat{p} > p$, podríamos calcular la solución

$$\hat{y}_{n+1} = y_n + h_n \sum_{i=1}^s \hat{b}_i k_i$$

sin coste adicional, pues las etapas k_i coinciden. Esta solución más precisa se puede utilizar para estimar el error cometido con el primer método. A un par de métodos con estas características se le conoce como *par encajado*.

Seamos un poco más precisos. Sabemos que los errores locales satisfacen

$$\begin{aligned} u(x_{n+1}) - y_{n+1} &= C\psi(y_n)h_n^{p+1} + O(h_n^{p+2}), \\ u(x_{n+1}) - \hat{y}_{n+1} &= O(h_n^{\hat{p}+1}). \end{aligned}$$

Eliminando $u(x_{n+1})$ y usando que $p + 2 \leq \hat{p} + 1$, se tiene que

$$C\psi(y_n)h_n^{p+1} = \hat{y}_{n+1} - y_{n+1} + O(h^{p+2}).$$

Esto justifica la estimación

$$u(x_{n+1}) - y_{n+1} \approx \hat{y}_{n+1} - y_{n+1} = h_n \sum_{i=1}^s (\hat{b}_i - b_i) k_i. \quad (2.15)$$

Así, podemos tomar

$$\text{ERROR} = \|\hat{y}_{n+1} - y_{n+1}\|_{\infty}$$

y se sigue como en el ejemplo que vimos en el capítulo anterior.

Veamos dos ejemplos debidos a Fehlberg. El primero es el RKF2(3), de tablero

0	0		
1	1		
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	
y_{n+1}	$\frac{1}{2}$	$\frac{1}{2}$	0
\hat{y}_{n+1}	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{4}{6}$.

El método de orden 2 que se usa para avanzar es en realidad de dos etapas, mientras que el de orden 3 que se usa para estimar es de tres etapas. El segundo par encajado que queremos presentar es el RKF2(3)B,

0	0			
$\frac{1}{4}$	$\frac{1}{4}$			
$\frac{27}{40}$	$-\frac{189}{800}$	$\frac{729}{800}$		
1	$\frac{214}{891}$	$\frac{1}{33}$	$\frac{650}{891}$	
y_{n+1}	$\frac{214}{891}$	$\frac{1}{33}$	$\frac{650}{891}$	0
\hat{y}_{n+1}	$\frac{533}{2106}$	0	$\frac{800}{1053}$	$-\frac{1}{78}$.

El método de avance es de 3 etapas y orden 2, y el de estimación es de 4 etapas y orden 3.

Notación. “Nombre $p(q)$ ” quiere decir que el orden de y_{n+1} (el método que se utiliza para avanzar) es p , y el orden de \hat{y}_{n+1} (el método que se usa para estimar) es q .

Comparando los métodos RKF2(3) y RKF2(3)B podríamos pensar que el primero es mucho mejor, por emplear una etapa menos para conseguir el mismo orden. Sin embargo, el método RKF2(3)B tiene la propiedad FSAL (First Same As Last),

$$a_{si} = b_i \quad i = 1, \dots, s.$$

Por consiguiente, la última etapa del paso n viene dada por

$$k_s = f(x_n + c_s h_n, y_n + h_n \sum_{i=1}^{s-1} a_{si} k_i) = f(x_n + h_n, y_n + h_n \sum_{i=1}^{s-1} b_i k_i),$$

donde hemos usado las condiciones de suma por filas y de consistencia para probar que $c_s = \sum_{j=1}^s a_{sj} = \sum_{j=1}^s b_j = 1$. Por otra parte, por ser el método explícito, la primera etapa del paso $n + 1$ viene dada por

$$k_1 = f(x_{n+1}, y_{n+1}) = f(x_n + h_n, y_n + h_n \sum_{i=1}^{s-1} b_i k_i).$$

Es decir, la primera etapa de cada paso coincide con la última del paso anterior. Por tanto la primera etapa sólo hay que calcularla en el primer paso, y desde el punto de vista computacional es como si el método tuviera una etapa menos.

Aunque la estimación (2.15) da el error de truncación para el método del par encajado de orden más bajo, si h es suficientemente pequeño el error para el método de orden alto será aún menor; así que, ¿por qué no hacer extrapolación local, avanzando con el método de orden más alto? Esto es lo que hacen los métodos de *Dormand y Prince* DOPRI5(4) y DOPRI8(7), que son algunos de los mejores métodos existentes hasta el momento.

¿Por qué no se hace siempre extrapolación local? Porque a veces supone desventajas. Por ejemplo, si hiciéramos extrapolación local en el método de RKF2(3)B, no podríamos aprovechar la propiedad FSAL, con el coste computacional extra que ello supone.