TEMA 3

Métodos numéricos para resolver Problemas de Valores Iniciales.

Versión 21/4/2022

Índice

1. Introducción

1.1. El problema de Cauchy. Existencia y unicidad.

Sólo una minoría de ecuaciones diferenciales ordinarias puede resolverse mediante los métodos que aparecen en las obras dedicadas a su estudio. Y aún dentro de las resolubles no siempre se podrá calcular explícitamente la solución que pasa por un punto dado, o se podrá evaluar con facilidad esa solución en cualquier punto. De ahí el gran interés de la resolución aproximada de estos problemas por métodos numéricos.

Se recuerdan los resultados principales de existencia y unicidad de solución del *problema de valores iniciales* (PVI) o *problema de Cauchy*

$$\begin{array}{|l} x' = f(t,x) \\ x(t_0) = \mu \end{array} \quad \begin{array}{l} f : D \subseteq \mathbb{R}^2 \to \mathbb{R} \\ (t_0, \mu) \in D \end{array} \tag{1}$$

siendo $x = x(t)$ una función desconocida de t.

Se dice que $x(t)$ es una *solución* de (1) en un intervalo $I \ni t_0$ si es derivable en I y verifica (1) $\forall t \in I$.

Teorema 1 (*existencia y unicidad*)

Si $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ es continua en $D = [a,b] \times \mathbb{R} = \{(t,x) : t \in [a,b],\ x \in \mathbb{R}\}$ y satisface la condición de Lipschitz *respecto de su segunda variable*

$$|f(t,u) - f(t,v)| \leq L|u - v| \quad \forall t \in [a,b] \tag{2}$$

entonces el problema (1) admite una única solución $x(t)$ en $[a,b]$. A L se la denomina constante de Lipschitz *respecto de x.*

La condición de Lipschitz (2) puede sustituirse por que la derivada $\left|\frac{\partial f}{\partial x}\right|$ exista y sea continua y acotada en D. En general diremos que $f \in \mathcal{F}_p(D)$ si f posee derivadas parciales continuas y acotadas en $D = [a,b] \times \mathbb{R}$ hasta orden al menos p. Supondremos siempre que $f \in \mathcal{F}_p(D)$ con $p \geq 1$ de modo que la existencia y unicidad de solución de (1) en el intervalo $[a,b]$ esté asegurada. Por otro lado, este teorema se puede extender fácilmente a sistemas de ecuaciones

diferenciales ordinarias de primer orden sin más que cambiar en la condición (2) los valores absolutos por normas vectoriales y matriciales apropiadas.

Para dominios más reducidos se tiene el

Teorema 2 (*existencia y unicidad*)

Si $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ *es continua en el rectángulo* $R = \{(t,x) : |t - t_0| \leq r_t,\ |x - \mu| \leq r_x\}$ *y satisface (2), entonces el problema (1) admite una única solución* $x(t)$ *en* $[t_0 - h, t_0 + h]$ *donde* $h = \min\left\{r_t, \frac{r_x}{M}\right\}$ *y* $|f(t,x)| \leq M$ *en* R.

2. Métodos de discretización

En muy pocos casos se puede encontrar la solución exacta $x(t)$ de un PVI y, por tanto, es imprescindible establecer métodos aproximados de cálculo, entre los que se encuentran los basados en desarrollos en serie de potencias, en serie de Frobenius, etc, además de una categoría diferente formada por métodos numéricos.

Resolver numéricamente el PVI (1) no significa obtener la expresión analítica de $x(t)$, sino más bien una aproximación de su valor en una serie de puntos del intervalo de trabajo $[a, b]$. Para resolver numéricamente el PVI hay que transformar los elementos continuos, no computables, en elementos discretos, computables. Para ello se emplean diversos *métodos de discretización*.

2.1. Generalidades

Un método numérico de discretización para resolver un PVI consiste básicamente en

- Tomar una partición $\{a = t_0 < t_1 < \cdots < t_N = b\}$ con puntos o *nodos* del intervalo $[a, b]$ donde se requiere evaluar la solución única $x(t)$. Habitualmente es una partición homogénea, es decir

$$h = \frac{b-a}{N}, \quad \begin{cases} t_{n+1} = t_n + h, \ n = 0, 1, \ldots, N-1 \text{ o bien} \\ t_n = a + nh, \ n = 0, 1, \ldots, N. \end{cases} \tag{3}$$

  A h se le llama *paso* o *longitud de paso*.

- Obtener una serie de valores $\{x_n\}_{n=0}^{N}$ asociados a los nodos, que se denomina *solución numérica* del PVI.

El objetivo es que la solución numérica sea una aproximación de la solución exacta: $x_n \approx x(t_n)$, tanto mejor cuanto más densa sea la partición, es decir cuando $N \to \infty$ o equivalentemente $h \to 0$.

Son varias las técnicas (integración y derivación numéricas, desarrollos de Taylor, interpolación, etc) usadas para el cálculo de tales valores. Los métodos se suelen clasificar de varias formas.

En general, un *método numérico de k pasos* para el PVI (1) adopta la forma

$$\begin{cases} x_0, x_1, \ldots, x_{k-1} = \text{ valores iniciales} \\ x_{n+k} = \displaystyle\sum_{j=0}^{k-1} \alpha_j x_{n+j} + h\Phi(x_{n+k}, x_{n+k-1}, \ldots, x_n; t_n, h) \end{cases} \tag{4}$$

donde la función Φ habrá de cumplir $\forall h < h_0$

$$\begin{aligned} & f \equiv 0 \Rightarrow \Phi \equiv 0; \\ & |\Phi(u_0, \ldots, u_k; t, h) - \Phi(v_0, \ldots, v_k; t, h)| \leq M \sum_{j=0}^{k} |u_j - v_j| \end{aligned} \tag{5}$$

es decir, Φ cumple una condición de Lipschitz respecto de las variables representativas de los valores numéricos de la solución $x(t)$.

Definición 1 (*Clasificación de métodos.*)

Cuando $k = 1$ se hablará de un método de un paso *y, en caso contrario, de un* método multipaso*. Por otro lado, en el caso de que en (4) la función Φ no dependa de su primera variable x_{n+k}, el método se clasifica como* método explícito*. En caso contrario, se llamará* método implícito *y requerirá resolver una ecuación en cada paso.*

2.2. Consistencia, convergencia, estabilidad

Se definen los conceptos básicos que permiten establecer las características de comportamiento deseables para un método numérico para PVI. En lo que sigue se considera que $x(t)$ se refiere a la solución exacta de (1).

Definición 2 (*Error global y local.*)

- *Se llama* error de truncatura global *o simplemente* error global *a*

$$e_n = x(t_n) - x_n.$$

- *Se llama* error de truncatura local *o simplemente* error local *a*

$$R_{n+k} = x(t_{n+k}) - \sum_{j=0}^{k-1} \alpha_j x(t_{n+j}) - h\Phi(x(t_{n+k}), x(t_{n+k-1}), \ldots, x(t_n); t_n, h).$$

Para un método de un paso, el error local toma la forma

$$R_{n+1} = x(t_{n+1}) - x(t_n) - h\Phi(x(t_{n+1}), x(t_n); t_n, h)$$

y si además fuera explícito,

$$R_{n+1} = x(t_{n+1}) - x(t_n) - h\Phi(x(t_n); t_n, h).$$

El error global mide simplemente las diferencias entre la solución aproximada y la exacta, y lo deseable es que sea lo más pequeño posible y tienda a cero cuando $N \to \infty$. El error local mide la desviación que introduce el método en cada paso, es decir, el error que produce si partiera de los valores de la solución exacta. El error local podría provocar un efecto acumulador perjudicial. Aunque es de esperar que $R_{n+k} \to 0$ cuando $h \to 0$, no basta para controlar el efecto de acumulación del error local. Para ello se define la propiedad de *consistencia.*

Definición 3 (*Consistencia.*)

Un método (4) se dirá consistente *con (1) si*

$$\lim_{\substack{h\to 0 \\ t=a+nh}} \frac{R_{n+k}}{h} = 0. \tag{6}$$

Definición 4 (*Primer polinomio característico.*)

Dado el método (4), su primer polinomio característico *es*

$$p(\lambda) = \lambda^k - \sum_{j=0}^{k-1} \alpha_j \lambda^j. \tag{7}$$

Teorema 3 (*Caracterización de la consistencia.*)

El método (4) es consistente si y solo si

$$\begin{cases} p(1) = 0, \\ \Phi(x(t_n), \ldots, x(t_n); t_n, 0) = p'(1) f(t_n, x(t_n)). \end{cases}$$

Es como hacer $h=0$ en Φ

Definición 5 (*Convergencia.*)

El método (4) se dice convergente *para (1) si y solo si*

$$\lim_{N\to\infty} \max_{0\le n\le N} |x_n - x(t_n)| = 0 \tag{8}$$

o bien (definición alternativa equivalente)

$$\lim_{\substack{h\to 0 \\ t=a+nh}} x_n = x(t). \tag{9}$$

Un método no convergente se dice divergente.

La propiedad de convergencia es la que se debe exigir a todo método que pretenda ofrecer utilidad práctica. La convergencia de un método implica su consistencia, pero el recíproco no es cierto: el método puede ser consistente, pero muy sensible a otras perturbaciones de Φ u otras causas, como pueden ser los errores de redondeo en la computación o los errores inherentes a la discretización. Se requiere otra definición más para precisar el concepto.

Supongamos que el método (4) sufre una perturbación δ_{n+k} en la evaluación de Φ de manera que en lugar de x_{n+k} se obtiene una solución perturbada

Convergente ⇒ Consistente
Consistente + Estable ⇒ Convergente

z_{n+k} en la forma

$$z_n = x_n + \delta_n, \quad n = 0, 1, \ldots, k-1$$
$$z_{n+k} = \sum_{j=0}^{k-1} \alpha_j z_{n+j} + h\left(\Phi(z_{n+k}, z_{n+k-1}, \ldots, z_n; t_n, h) + \delta_{n+k}\right), \ n \geq 0.$$

Definición 6 (*Estabilidad o 0-estabilidad.*)

Sean $\{\delta_n\}$ y $\{\delta_n^*\}$ dos perturbaciones del método (4) y $\{z_n\}$ y $\{z_n^*\}$ sus respectivas soluciones perturbadas. Diremos que el método es estable *o* cero-estable *si existen constantes M y h_0 tales que $\forall h \leq h_0$ se cumple*

$$|z_n - z_n^*| \leq M\varepsilon, \ \forall n \quad \text{siempre que} \ |\delta_n - \delta_n^*| \leq \varepsilon \quad \forall n$$

Teorema 4 (*Caracterización de la estabilidad.*)

El método (4) es estable si y solo si el primer polinomio característico tiene todos sus ceros en el disco unidad, y los ceros de módulo 1 son simples.

Teorema 5 (*Caracterización de la convergencia.*)

Un método (4) es convergente si y sólo si es consistente y cero-estable.

Por último, dos o más métodos para un mismo problema pueden compararse respecto de su precisión usando el concepto de *orden*.

Definición 7 (*Orden.*)

Diremos que (4) es de orden *$p \geq 1$ si $\forall f \in \mathcal{F}_p(D)$ se cumple*

$$R_{n+k} = O(h^{p+1}).$$

3. Métodos de un paso

3.1. Generalidades

Lo que sigue es básicamente una particularización de lo anterior para el caso $k = 1$. Un método de un paso para el PVI viene expresado, según el caso, como

Método de un paso explícito

$$\begin{aligned} &x_0 = \mu,\ t_0 = a, \\ &\text{para } n = 0, 1, \ldots, N-1 \\ &\quad t_{n+1} = t_n + h \\ &\quad x_{n+1} = x_n + h\Phi(x_n; t_n, h) \end{aligned} \tag{10}$$

Método de un paso implícito

$$\begin{aligned} &x_0 = \mu,\ t_0 = a, \\ &\text{para } n = 0, 1, \ldots, N-1 \\ &\quad t_{n+1} = t_n + h \\ &\quad \text{resolver } x_{n+1} = x_n + h\Phi(x_{n+1}, x_n; t_n, h) \end{aligned} \tag{11}$$

La condición de Lipschitz (5) para Φ asegura no solamente la existencia y unicidad de solución exacta, sino también la de solución numérica en el caso implícito (11). En efecto, x_{n+1} es la solución de una ecuación (en general no lineal) del tipo $z = g(z)$ donde $g(z) = x_n + h\Phi(z, x_n; t_n, h)$ y, por tanto se cumple

Constante de Lipschitz de Φ

$$|g(u) - g(v)| \leq hM|u - v|$$

siendo M la constante de Lipschitz de Φ en (5). Entonces siempre habrá un h_0 tal que $h_0 M < 1$ y para cualquier $h \leq h_0$, $g(z)$ será contráctil, por lo que la ecuación $z = g(z)$ tendrá solución y el método de iteración funcional asociado (ver Tema 1) convergerá hacia ella.

Los conceptos asociados a los métodos de un paso toman la forma que se comenta seguidamente. Por simplicidad en las notaciones se presenta el caso explícito.

- **Error de truncatura global**. Es el mismo.

- **Error de truncatura local**.

$$R_{n+1} = x(t_{n+1}) - x(t_n) - h\Phi(x(t_n); t_n, h).$$

- **Consistencia**. Es igual.

- **Primer polinomio característico**.

$$p(\lambda) = \lambda - 1.$$

- **Caracterización de la consistencia**.

$$\begin{cases} p(1) = 0 \text{ ya la cumple }, \\ \Phi(x(t_n); t_n, 0) = f(t_n, x(t_n)). \end{cases}$$

- **Convergencia**. La definición es igual.

- **Estabilidad**. La definición es igual.

- **Caracterización de la estabilidad**. El método es estable porque la única raíz de $p(\lambda)$ es $\lambda = 1$ simple.

- **Caracterización de la convergencia**. Al ser estable, la convergencia equivale a la consistencia.

- **Orden**. (10) es de $\boxed{\textit{orden}}$ $p \geq 1$ si $\forall f \in \mathcal{F}_p(D)$ se cumple

$$R_{n+1} = x(t_{n+1}) - x(t_n) - h\Phi(x(t_n); t_n, h) = O(h^{p+1}).$$

Dado que el análisis del orden requiere usar desarrollos de Taylor tanto de la solución $x(t)$ como de la función $f(t, x)$, será de gran ayuda la notación abreviada que facilita el operador $D^m_{\phi,\psi}$ que actúa sobre cualquier función de dos variables $g = g(t, x)$ suficientemente derivable.

$$D^m_{\phi,\psi} g = \left(\phi \frac{\partial \bullet}{\partial t} + \psi \frac{\partial \bullet}{\partial x}\right)^m g = \sum_{j=0}^{m} \binom{m}{j} \phi^{m-j} \psi^j \frac{\partial^m g}{\partial t^{m-j} \partial x^j} \tag{12}$$

En particular para la función f de (1) usaremos

$$F = D_{1,f} f, \; G = D^2_{1,f} f \quad \text{y} \quad H = D^3_{1,f} f.$$

Así, el desarrollo de Taylor de una función cualquiera g de dos variables alrededor del punto (t, x) viene expresado por (donde $g = g(t, x)$

$$\begin{aligned} g(t+h, x+\kappa) &= g + D_{h,\kappa}g + \frac{1}{2!}D^2_{h,\kappa}g + \cdots + \frac{1}{p!}D^p_{h,\kappa}g + (\text{Resto}) \\ &= T_p(g; h, \kappa) + (\text{Resto}) \end{aligned}$$

y, de igual modo, las derivadas sucesivas de la solución $x(t)$ de (1) pueden ponerse como

$$\begin{array}{rcll} x' &=& f & \\ x'' &=& f' &= f_t + f_x x' = f_t + f_x f = F \\ x''' &=& f'' &= F' = F_t + F_x f = G + f_x F \\ x^{iv} &=& f''' &= (f'')_t + (f'')_x f = H + 3F(f_{xx}f + f_{tx}) + f_x(G + f_x F) \\ & \vdots & & \end{array} \tag{13}$$

Ejercicio: desarrollar y comprobar.

3.2. El método de Euler. Variantes

Se trata de un método elemental de un paso, explícito, para (1) en el que la elección $\Phi(x; t, h) = f(t, x)$ es la más simple posible. A partir de (10) se formula así:

Método de Euler
$x_0 = \mu,\ t_0 = a,$
para $n = 0, 1, \ldots, N-1$
$\quad t_{n+1} = t_n + h$
$\quad x_{n+1} = x_n + hf(t_n, x_n)$

(14)

La deducción de éste y otros métodos se puede hacer por diversas vías, entre las que se pueden destacar las siguientes.

1. **Deducción por integración numérica.** Se trata de usar la representación integral de la solución exacta en cada subintervalo $[t_n, t_{n+1}]$, a saber

$$x(t_{n+1}) - x(t_n) = \int_{t_n}^{t_{n+1}} x'(s)\, ds = \int_{t_n}^{t_{n+1}} f(s, x(s))\, ds$$

y usando la fórmula de integración numérica del rectángulo izquierda (ver Tema 2) se tiene

$$\cdots \approx hf(t_n, x(t_n)) \;\Rightarrow\; \boxed{x_{n+1} = x_n + hf(t_n, x_n)}.$$

Rectángulo derecha: $x_{n+1} = x_n + h\, g(t_{n+1}, x_{n+1})$

Trapecio: $x_{n+1} = x_n + \frac{h}{2}\left(g(t_n, x_n) + g(t_{n+1}, x_{n+1})\right)$ } Implícitos

2. **Deducción por derivación numérica.** Usando la fórmula de derivación numérica de diferencia progresiva para $x'(t_n)$ (ver Tema 2)

$$x'(t_n) = f(t_n, x(t_n)) \approx \frac{x(t_{n+1}) - x(t_n)}{h} \;\Rightarrow\; \boxed{x_{n+1} = x_n + hf(t_n, x_n)}.$$

3. **Deducción geométrica.** La aproximación x_{n+1} del valor $x(t_{n+1})$ se obtiene como el valor en $t = t_{n+1}$ de la recta que pasa por (t_n, x_n) y tiene pendiente $f(t_n, x_n)$. Dicha recta es una aproximación a la recta tangente a la curva solución $(t, x(t))$ en el punto $(t_n, x(t_n))$.

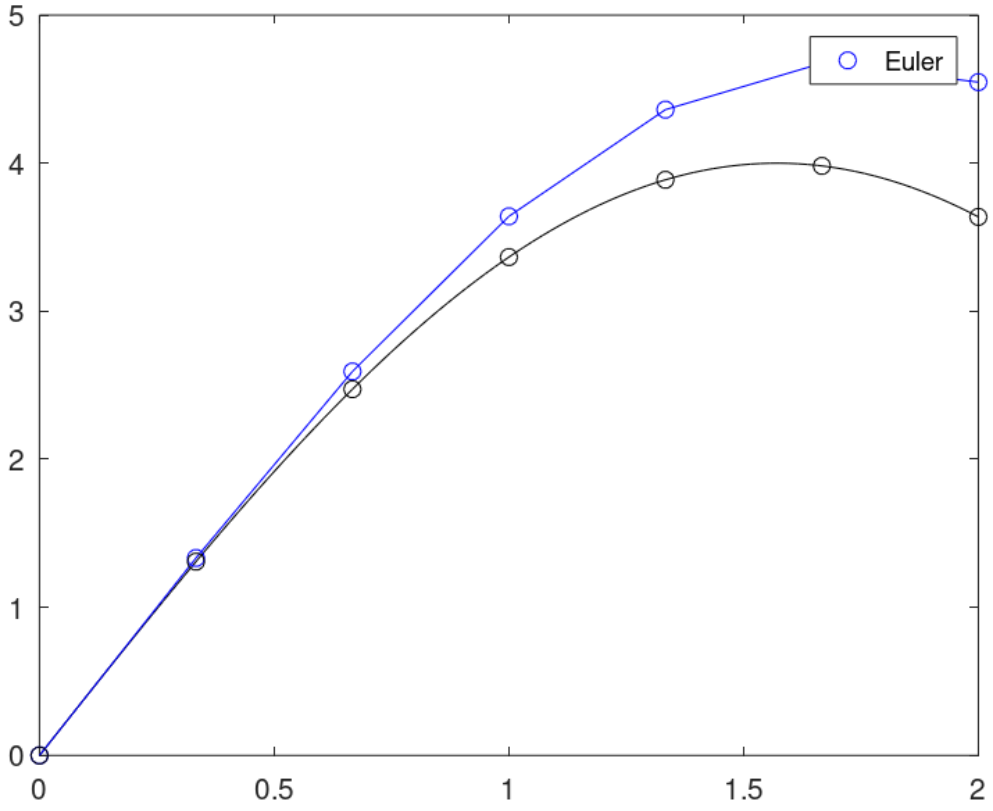

Figura 1: Método de Euler para el PVI $x' = 4\cos t,\ x(0) = 0$.

```
#########################
#### metodo de Euler ####
#########################
clear all;
clc;
more off;
function z = f(t,x) % la ecuacion diferencial
  z = 4*cos(t);
endfunction
function z = x(t) % la solucion
  z = 4*sin(t);
endfunction
a = 0; b = 2; mu = 0; N = 6; h = (b-a)/N;
tn = linspace(a,b,100);
xn = x(tn);
hf = figure(1); % para guardarla en archivo
plot(tn,xn,'k');
hold on; %voy a superponer en la misma grafica
%metodo de Euler
tn = []; xn = []; tt = a; xx = mu;
do
  tn = [tn tt]; xn = [xn xx];
  xx = xx+h*f(tt,xx); tt = tt + h;
until tt>b
plot(tn,xn,'b',tn,xn,'ob;Euler;',tn,x(tn),'ok');
hold off; %he terminado la grafica
print (hf, "T3.1.Euler.png", "-dpng");
close; %cierra la ventana de la figura
```

Cuadro 1: Código Octave empleado para la Figura 1.

3.2.1. Método de Euler implícito.

Si en lugar de la fórmula de integración del rectángulo izquierda se usa la del rectángulo derecha, o en lugar de la fórmula de derivación de diferencia progresiva se usa la regresiva, o en lugar de la tangente en (t_n, x_n) se usa la tangente en (t_{n+1}, x_{n+1}), entonces se obtiene el

Método de Euler implícito
$x_0 = \mu,\ t_0 = a,$
para $n = 0, 1, \ldots, N-1$
$\quad t_{n+1} = t_n + h$
$\quad$ resolver $x_{n+1} = x_n + hf(t_{n+1}, x_{n+1})$

(15)

Es sencillo comprobar (hágase) que ambos métodos (14) y (15) son consistentes con el problema (1), estables, y sus errores locales de truncatura son

$$R_{n+1} = \frac{h^2}{2}x''(t) + O(h^3) \quad \text{y} \quad R_{n+1} = -\frac{h^2}{2}x''(t) + O(h^3)$$

respectivamente, por lo que ambos métodos tienen orden $p = 1$.

3.2.2. Método de Euler mejorado, o del punto medio.

Esta variante del método de Euler se consigue aplicando la fórmula de integración del punto medio, o trazando la tangente en el punto $(t_n + \frac{h}{2})$. El problema en esta fórmula es cómo obtener el valor de $x\left(t_n + \frac{h}{2}\right)$. Runge[1] pensó que se podría aproximar ese valor por el obtenido al aplicar la fórmula de Euler con un paso mitad, de manera que $x\left(t_n + \frac{h}{2}\right) \approx x_n + \frac{h}{2}f(t_n, x_n)$. De esta forma queda el

Método de Euler mejorado (punto medio)
$x_0 = \mu,\ t_0 = a,$
para $n = 0, 1, \ldots, N-1$
$\quad t_{n+1} = t_n + h$
$\quad x_{n+1} = x_n + hf\left(t_n + \frac{h}{2}, x_n + \frac{h}{2}f(t_n, x_n)\right)$

(16)

Este método tiene un error de truncatura local $R_{n+1} = \dfrac{h^3}{6}\left(\dfrac{1}{4}G + f_x F\right) + O(h^4)$, por lo que tiene orden $p = 2$.

[1] Carl Runge (1856-1927), matemático, físico y espectroscopista alemán.

3.2.3. Método de Euler modificado, o de Heun.

Se consigue aplicando la fórmula de integración del trapecio, o la de derivación de diferencia centrada, o la recta secante que pasa por (t_n, x_n) y (t_{n+1}, x_{n+1}). Análogamente al anterior, se presenta aquí el problema de evaluar $x(t_{n+1})$, que se resuelve aproximándolo por Euler. Así queda el

Método de Euler modificado (Heun)
$x_0 = \mu,\ t_0 = a,$
para $n = 0, 1, \ldots, N-1$
$\quad t_{n+1} = t_n + h$
$\quad x_{n+1} = x_n + \frac{h}{2}\left(f(t_n, x_n) + f(t_{n+1}, x_n + hf(t_n, x_n))\right)$

(17)

Su error de truncatura local es $R_{n+1} = \dfrac{h^3}{6}\left(-\dfrac{1}{2}G + f_x F\right) + O(h^4)$, por lo que también tiene orden $p = 2$.

La Figura 2 muestra una comparación de los métodos de Euler, Punto medio y Heun con $N = 6$ para el PVI $x' = 4\cos t,\ x(0) = 0$ en $[0, \pi]$, de solución trivial $x(t) = \sin t$. Se puede apreciar que el método de Euler sufre un grave efecto de acumulación del error de truncatura local, que lo hace desviarse de la curva exacta, aunque a partir de $\frac{\pi}{2}$ los errores locales son de signo contrario y van contrarrestando a los primeros, de forma que en el extremo del intervalo el error global llega a anularse. Los métodos de Punto medio y de Heun, ambos de orden 2, parecen comportarse mejor, aunque hay que tener en cuenta que exigen mayor esfuerzo computacional porque emplean dos evaluaciones de f en casa paso.

3.3. Métodos de Taylor

Si la función f del PVI (1) es suficientemente diferenciable, entonces es sencillo obtener un método de orden $p \geq 1$ sin más que usar el desarrollo de Taylor para $x(t)$

$$x(t+h) = x(t) + hx'(t) + \frac{h^2}{2}x''(t) + \cdots + \frac{h^p}{p!}x^{(p)}(t) + \frac{h^{p+1}}{(p+1)!}x^{(p+1)}(\xi_t) \quad (18)$$

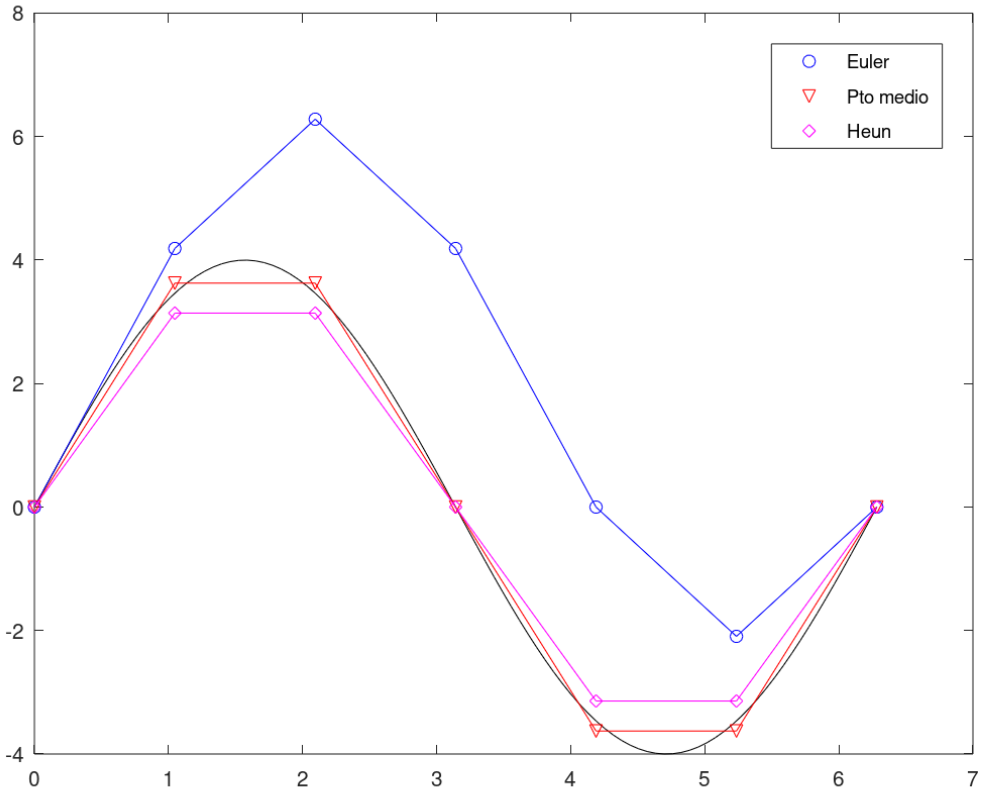

Figura 2: Comparación de los métodos de Euler, punto medio y Heun para el PVI $x' = 4\cos t,\ x(0) = 0$.

```
#########################
#### metodo de Euler ####
#########################
clear all;
clc;
more off;
function z = f(t,x) % la ecuacion diferencial
  z = 4*cos(t);
endfunction
function z = x(t) % la solucion
  z = 4*sin(t);
endfunction
a = 0; b = 2*pi; mu = 0; N = 6; h = (b-a)/N;
tn = linspace(a,b,100);
xn = x(tn);
hf = figure(1); % para guardarla en archivo
plot(tn,xn,'k');
hold on; %voy a superponer en la misma grafica
%metodo de Euler
tn = []; xn = []; tt = a; xx = mu;
do
  tn = [tn tt]; xn = [xn xx];
  xx = xx+h*f(tt,xx); tt = tt + h;
until tt>b
plot(tn,xn,'b',tn,xn,'ob;Euler;');
%metodo del punto medio
tn = []; xn = []; tt = a; xx = mu;
do
  tn = [tn tt]; xn = [xn xx];
  xx = xx+h*f(tt+h/2,xx+h/2*f(tt,xx)); tt = tt + h;
until tt > b;
plot(tn,xn,'r',tn,xn,'vr;Pto medio;');
%metodo de Heun
tn = []; xn = []; tt = a; xx = mu;
do
  tn = [tn tt]; xn = [xn xx];
  xx = xx+(h/2)*(f(tt,xx)+f(tt+h,xx+h*f(tt,xx))); tt = tt + h;
until tt > b;
plot(tn,xn,'m',tn,xn,'dm;Heun;');
hold off; %he terminado la grafica
print (hf, "T3.2.EulerPMH.png", "-dpng");
close; %cierra la ventana de la figura
```

Cuadro 2: Código Octave empleado para la Figura 2.

para deducir el

Método de Taylor
$x_0 = \mu,\ t_0 = a,$
para $n = 0, 1, \ldots, N-1$
$\quad t_{n+1} = t_n + h$
$\quad x_{n+1} = x_n + hx'_n + \frac{h^2}{2}x''_n + \cdots + \frac{h^p}{p!}x_n^{(p)}$

(19)

donde los valores $x_n^{(r)}$ se obtienen evaluando en (t_n, x_n) las sucesivas derivadas de $f(t, x(t))$ que se dan en (13).

3.3.1. Método de Taylor de orden $p = 1$.

Se trata del método de Euler explícito (14).

3.3.2. Método de Taylor de orden $p = 2$.

Método de Taylor de orden $p = 2$
$x_0 = \mu,\ t_0 = a,$
para $n = 0, 1, \ldots, N-1$
$\quad t_{n+1} = t_n + h$
$\quad x_{n+1} = x_n + hf(t_n, x_n) + \frac{h^2}{2}F(t_n, x_n)$

(20)

3.4. Métodos de Runge-Kutta

A la vista de la complejidad que involucra el desarrollo de las derivadas sucesivas de $x(t)$ para un orden elevado $p \geq 3$, los métodos de Taylor no son muy populares. Por ello se intenta conseguir métodos de orden elevado que eviten la evaluación de derivadas. Es el objetivo de los métodos de Runge-Kutta[2] (en adelante RK), que se basan en varias evaluaciones de $f(t, x)$ en puntos del intervalo $[t_n, t_{n+1}]$ de modo que resulten equivalentes en cierta medida a métodos de Taylor de orden alto.

De forma más precisa, un método de RK de m evaluaciones es como

[2]Martin Wilhelm Kutta (1867-1944), físico y matemático alemán. En 1901 desarrolló el método RK junto con Runge.

sigue.

Método de RK de m evaluaciones

$$
\begin{aligned}
&x_0 = \mu,\ t_0 = a,\\
&\text{para } n = 0, 1, \ldots, N-1\\
&\quad t_{n+1} = t_n + h\\
&\quad x_{n+1} = x_n + h\sum_{j=1}^{m} b_j K_j(t_n, x_n)\\
&\text{donde}\\
&K_i(t,x) = f\left(t + c_i h, x + h\textstyle\sum_{j=1}^{m} a_{ij} K_j(t,x)\right)\ \ i = 1, \ldots, m
\end{aligned}
\tag{21}
$$

Definición 8 (*Arreglo de Butcher.*)

Cualquier método (21) queda determinado por el arreglo de Butcher

$$
\begin{array}{c|ccc}
c_1 & a_{11} & \cdots & a_{1m}\\
c_2 & a_{21} & \cdots & a_{2m}\\
\vdots & \vdots & \ddots & \vdots\\
c_m & a_{m1} & \cdots & a_{mm}\\
\hline
 & b_1 & \cdots & b_m
\end{array}
\tag{22}
$$

Observaciones.

- El método de RK es, en general implícito.
- Si $a_{ij} = 0\ \forall i \leq j$, entonces el método será explícito.
- Si $a_{ij} = 0\ \forall i < j$, entonces el método se dirá *diagonalmente implícito*.
- El método de RK es consistente si y solo si $b_1 + \cdots + b_m = 1$.
- Para simplificar el análisis supondremos $c_i = a_{i1} + a_{i2} + \cdots + a_{im}\ \forall i$.

- Un método de RK explícito tendrá, por tanto,

$$\begin{aligned} K_1 &= f(t,x) \\ K_i &= f\left(t+c_ih, x+h\sum_{j=1}^{i-1} a_{ij}K_j\right) \quad i=2,\ldots,m \end{aligned}$$

A continuación veremos algunos métodos de Runge-Kutta clásicos.

3.4.1. Método de RK explícito de 2 evaluaciones

Teniendo en cuenta las condiciones de consistencia, podemos escribir

Método de RK de 2 evaluaciones explícito
$x_0 = \mu,\ t_0 = a,$
para $n = 0, 1, \ldots, N-1$
$\quad t_{n+1} = t_n + h$
$\quad x_{n+1} = x_n + h\left((1-\alpha)K_1 + \alpha K_2\right)$
donde
$\quad K_1 = f(t_n, x_n)$
$\quad K_2 = f(t_n + \beta h, x_n + h\beta K_1)$

(23)

donde α y β han de ser elegidos adecuadamente, tras un análisis del error de truncatura local para deducir el orden máximo del método. Así

$$R_{n+1}(t) = x(t+h) - x(t) - h\left((1-\alpha)K_1 + \alpha K_2\right)$$

donde $K_1 = f(t, x(t))$, $K_2 = f(t+\beta h, x + h\beta f(t, x(t)))$ que, desarrollando por Taylor alrededor del punto $(t, x(t))$ y abreviando $x = x(t)$, $f = f(t, x(t))$, tenemos

$$K_2 = f(t+\beta h, x+h\beta f) = T_2(f; \beta h, \beta h f) + O(h^3) = f + \beta h F + \frac{\beta^2 h^2}{2} G + O(h^3)$$

y por lo tanto tendremos que

$$\begin{aligned} R_{n+1}(t) &= hx' + \frac{h^2}{2}x'' + \frac{h^3}{6}x''' + O(h^4) \\ &\quad - h\left((1-\alpha)f + \alpha\left(f + \beta h F + \frac{\beta^2 h^2}{2}G + O(h^3)\right)\right) \\ &= \cdots = h^2\left(\frac{1}{2} - \alpha\beta\right)F + \frac{h^3}{6}((1-3\alpha\beta^2)G + f_xF) + O(h^4) \end{aligned}$$

de donde se deduce que el método será de orden 2 si el coeficiente de h^2 se hace cero, es decir, $\alpha\beta = \frac{1}{2}$. Esta igualdad conduce a toda una familia de métodos de RK de orden 2. En particular tendremos

- Para $\alpha = 1,\ \beta = \frac{1}{2}$ se obtiene el método de Euler mejorado o Punto medio (16.
- Para $\alpha = \frac{1}{2},\ \beta = 1$ se obtiene el método de Euler modificado o Heun (17).
- Para $\alpha = \frac{3}{4},\ \beta = \frac{2}{3}$ el método resultante es

$$\begin{array}{l} x_0 = \mu,\ t_0 = a, \\ \text{para } n = 0, 1, \ldots, N-1 \\ \quad t_{n+1} = t_n + h \\ \quad x_{n+1} = x_n + \dfrac{h}{4}\left(f(t_n, x_n) + 3f\left(t_n + \dfrac{2}{3}h, x_n + \dfrac{2}{3}hf(t_n, x_n)\right)\right) \end{array} \tag{24}$$

Además, podemos decir que en cierto sentido la elección de estos parámetros α, β es óptima. ¿En qué sentido?

3.4.2. Método de RK explícito de 4 evaluaciones (Runge-Kutta clásico)

Es un método de orden 4 así:

Método de RK de 4 evaluaciones explícito

$$\begin{array}{l} x_0 = \mu,\ t_0 = a, \\ \text{para } n = 0, 1, \ldots, N-1 \\ \quad t_{n+1} = t_n + h \\ \quad x_{n+1} = x_n + \dfrac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\ \text{donde} \\ \quad K_1 = f(t_n, x_n) \\ \quad K_2 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}K_1) \\ \quad K_3 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}K_2) \\ \quad K_4 = f(t_n + h, x_n + hK_3) \end{array} \tag{25}$$

El orden del método se obtiene mediante desarrollos de Taylor adecuados.

Después de las oportunas simplificaciones se llega a la expresión

$$\begin{aligned} R_{n+1}(t) &= hx' + \frac{h^2}{2}x'' + \frac{h^3}{6}x''' + \frac{h^4}{24}x^{iv} + O(h^5) - h\frac{K_1 + 2K_2 + 2K_3 + K_4}{6} \\ &= \cdots + \frac{h^4}{24}\left(x^{iv} - (H + f_x G + F(f_x^2 + 3f_{tx} + 3ff_{xx}))\right) + O(h^5) \end{aligned}$$

y usando las relaciones (13) se comprueba que se anulan los términos en h, h^2, h^3 y h^4, por lo que el orden del método será al menos 4. Si se desea conocer el término principal de error de truncatura local, es decir el sumando que corresponde a h^5, entonces se debe usar en cada desarrollo un sumando más, lo que complica un poco los cálculos.

3.5. Análisis de errores y convergencia

Un aspecto sumamente importante al aplicar un método para PVI es el de disponer de una estimación del error global a través de una expresión que lo permita. El siguiente teorema la proporciona.

Ejercicio

$x' = \frac{t-x}{2} \quad x(0) = 1$

a) Runge-Kutta 2 evaluaciones con $h = \frac{1}{4} \quad \alpha = \frac{3}{4} \quad \beta = \frac{2}{3} \quad [0,3]$

$x_{n+1} = x_n + \frac{h}{4}\left(g(t_n, x_n) + 3g\left(t_n + \frac{2}{3}h, x_n + \frac{2}{3}hg(t_n, x_n)\right)\right)$

$x_0 = 1$

$x_1 = 1 + \frac{1}{16}\left(-\frac{1}{2} + 3\cdot\left(\frac{1/6 - (1 + \frac{1}{6}\cdot(-\frac{1}{2}))}{2}\right)\right) =$ 0.89843375 ??

$-3/8$

Teorema 6 (*Acotación del error global*)

Dado un método (10) para el PVI (1) con Φ lipschitziana respecto de x con constante de Lipschitz M, entonces

1. *El método es convergente si y sólo si es consistente.*
2. *Si el método es de orden $p \geq 1$, entonces el error global es $O(h^p)$. Más concretamente*

$$|x(t) - x_n| \leq Ch^p \frac{e^{M(b-a)} - 1}{M}.$$

La acotación del teorema anterior no tiene en cuenta las perturbaciones debidas a la presencia de errores de redondeo en la computación, que en la práctica siempre se producen y podrían tener un efecto acumulativo. En lugar de $\{x_n\}_{n=0}^N$, en la realidad se obtiene una sucesión de valores perturbados $\{\tilde{x}_n\}_{n=0}^N$ en la que se introduce un error de redondeo en el cálculo de cada nuevo término:

$$\begin{aligned} \tilde{x}_0 &= \mu + \delta_0 \\ \tilde{x}_{n+1} &= \tilde{x}_n + h\Phi(\tilde{x}_n, t_n, h) + \delta_n, \quad n \geq 0 \end{aligned}$$

Teorema 7 (*Acotación del error global con error de redondeo*)

En las condiciones descritas, si $|\delta_j| \leq \delta \ \forall j$ entonces

$$|x(t) - \tilde{x}_n| \leq \left(Ch^p + \frac{\delta}{h} \right) \frac{e^{M(b-a)} - 1}{M}.$$

Observación. Si bien teóricamente $x_n \to x(t_n)$ cuando $h \to 0$, hay que tener precaución con el tamaño del paso h ya que el término $\frac{\delta}{h}$ puede hacerse grande.

El siguiente teorema nos asegura la convergencia de los métodos clásicos vistos hasta ahora, lo que viene a confirmar su utilidad práctica.

Teorema 8 (*Convergencia de métodos clásicos*)

Sea $f \in \mathcal{F}_r(D)$ con r adecuado. Entonces

- *Los métodos de Euler son convergentes.*
- *El método de Taylor de orden p es convergente.*
- *Los métodos de Runge-Kutta explícitos de orden p son convergentes.*

3.6. Control del tamaño del paso

La siguiente argumentación permite obtener una estimación computable del error de truncatura local.

Supongamos que tenemos un método de orden $p \geq 1$ en el que el término principal del error de truncatura local es Ch^{p+1}, es decir, $R_{n+1} = Ch^{p+1} + O(h^{p+2})$. Por un lado, el cálculo de x_{n+1} desde x_{n-1} en dos pasos de tamaño h acumula un error de truncatura local doble del error local de cada paso; por otro lado, si desde t_{n-1} se aplica el método con un solo paso de tamaño doble $2h$ se obtiene una solución aproximada $x_{n+1}^{(2h)}$ de tal modo que

$$\begin{aligned} x(t_{n+1}) - x_{n+1} &\approx 2R_{n+1} \approx 2Ch^{p+1} \\ x(t_{n+1}) - x_{n+1}^{(2h)} &\approx R_{n+1}^{(2h)} \approx C(2h)^{p+1} = 2^{p+1}Ch^{p+1} \end{aligned}$$

por lo que restando se tiene

$$x_{n+1} - x_{n+1}^{(2h)} \approx (2^p - 1)2Ch^{p+1} \approx (2^p - 1)2R_{n+1}$$

lo que nos permite obtener una estimación del error de truncatura local respecto del paso h como

$$\boxed{R_{n+1} \approx \frac{x_{n+1} - x_{n+1}^{(2h)}}{2(2^p - 1)}.}$$

Así, es posible estimar el error de truncatura local mediante los valores x_{n+1} y $x_{n+1}^{(2h)}$ para poder decidir si conviene aumentar o disminuir el paso a lo largo del proceso de aplicación práctica del método, permitiendo de este modo una implementación adaptativa.

Siempre cae y la gente se la suele dejar

3.7. A-estabilidad o estabilidad numérica

A la hora de elegir un método para resolver numéricamente un PVI no solo hemos de tener en cuenta sus propiedades de convergencia y orden, sino también su comportamiento frente a dos aspectos diferentes pero que conducen a un mismo criterio o concepto.

- ¿Cómo se comporta el método frente a errores de redondeo? Es decir, para un tamaño fijo de paso h ¿cómo se acumulan los errores de redondeo conforme avanza n?

- ¿Qué ocurre si resolvemos numéricamente un PVI cuya solución tiene una parte que tiende a cero (transitoria) y otra que permanece (estacionaria) cuando $t \to \infty$? Un caso muy simple de ello es el PVI $x' = \lambda x$ con $x(0) = \mu$ y $\Re(\lambda) < 0$ cuya solución es $x(t) = \mu e^{\lambda t}$, donde $\Re(\lambda)$ denota la parte real de λ.

Es precisamente este PVI el que se utiliza como patrón o test para definir la estabilidad numérica de un método.

Definición 9 (*A-estabilidad o estabilidad numérica*)

El método (10) o (11) se dirá A-estable *o* numéricamente estable *si al aplicarlo al PVI $x' = \lambda x$ con $x(0) = \mu$ y $\Re(\lambda) < 0$ se cumple*

$$\lim_{n\to\infty} x_n = 0 \quad \forall h > 0.$$

Insiste mucho en que lo va a meter

Ejemplo: Si aplicamos el método de Euler al problema citado tendremos

$$x_{n+1} = x_n + h\lambda x_n = (1 + h\lambda)x_n = \cdots = (1 + h\lambda)^{n+1} x_0$$

luego la solución numérica tenderá a cero (como debería) si, siendo $w = \lambda h$, se cumple

Disco de centro -1 y radio 1

$$|1 + w| < 1$$

Los valores de w que cumplen esta desigualdad están en el círculo de centro -1 y radio 1 en el plano complejo. Sin embargo, puede haber valores de $h > 0$ que

hagan a w estar fuera del círculo. En consecuencia, **el método de Euler no es A-estable**.

Ejemplo: Si aplicamos el método de Euler implícito al mismo problema tendremos

$$x_{n+1} = x_n + h\lambda x_{n+1} \Rightarrow x_{n+1} = \frac{1}{1-h\lambda} x_n = \cdots = \left(\frac{1}{1-h\lambda}\right)^{n+1} x_0$$

luego la solución numérica tenderá a cero (como debería) si se cumple

$$\left|\frac{1}{1-w}\right| < 1$$

condición que es cierta para cualquier valor w con $\Re(w) < 0$ y por lo tanto para cualquier $h > 0$. En consecuencia, **el método de Euler implícito es A-estable**.

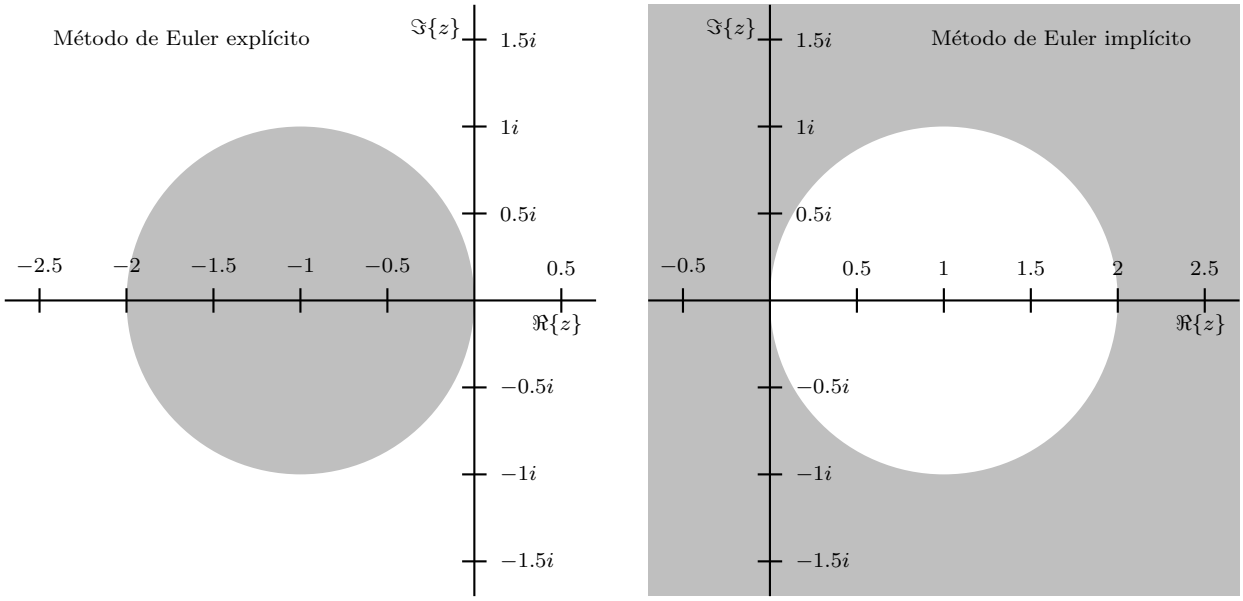

Figura 3: Regiones (sombreadas) de A-estabilidad de métodos de Euler

En general, cuando apliquemos un método de un paso al PVI patrón, la solución numérica adoptará la forma $x_{n+1} = Q(w)x_n$ siendo $Q(w)$ una función polinómica en los casos de métodos explícitos, y racional en los de implícitos. De esta forma, la comprobación de A-estabilidad pasará por analizar la desigualdad $|Q(w)| < 1$ con $w \in \mathbb{C}$.

Por desgracia, la condición de A-estabilidad es con frecuencia demasiado exigente. En la práctica se buscan más bien métodos que cumplan $|Q(w)| < 1$ en

una región de $\mathbb{C}$ lo más amplia posible. Así, se define la *región de A-estabilidad* de un método como el conjunto

$$R_A = \{w \in \mathbb{C} : |Q(w)| < 1\}$$

y se llama *intervalo de A-estabilidad* a $R_A \cap \mathbb{R}_-$.

Con esto, la A-estabilidad se puede redefinir así:

Definición 10 (*A-estabilidad o estabilidad numérica (alternativa)*)
Un método es estable si su región de A-estabilidad contiene al semiplano izquierdo de $\mathbb{C}$.

Es justo lo que hemos visto en el ejemplo

4. Métodos multipaso lineales (MML)

Recordemos de (4) que un método multipaso para el PVI (1) es de la forma

$$\begin{cases} x_0, x_1, \ldots, x_{k-1} = \text{ valores iniciales} \\ x_{n+k} = \displaystyle\sum_{j=0}^{k-1} \alpha_j x_{n+j} + h\Phi(x_{n+k}, x_{n+k-1}, \ldots, x_n; t_n, h) \end{cases} \tag{26}$$

donde la función Φ cumple la condición de Lipschitz (5).

Todas las definiciones y teoremas dados en la Sección 2 son válidos para métodos multipaso. Nos limitaremos aquí a los métodos multipaso en los que la función Φ es lineal.

Definición 11 (*Método multipaso lineal MML.*)

Diremos que un método (26) de k pasos es un ***método multipaso lineal (MML)*** *si se puede escribir en la forma*

$$x_{n+k} = \sum_{j=0}^{k-1} \alpha_j x_{n+j} + h \sum_{j=0}^{k} \beta_j f_{n+j} \tag{27}$$

donde se usa la notación $f_i = f(t_i, x_i)$.

Observaciones.

- Obsérvese que podría ocurrir que $\alpha_0 = \beta_0 = 0$ con lo que el método ya no sería de k pasos. Así, el método (27) es exactamente de k pasos si al menos uno de estos dos coeficientes no es nulo, esto es, $|\alpha_0| + |\beta_0| \neq 0$.

- Si $\beta_k = 0$ entonces el método (27) es explícito; de lo contrario es implícito.

En el caso $\beta_k \neq 0$ (método MML implícito), se plantea la cuestión de la existencia de solución para x_{n+k}.

Proposición 1

Sea $f(t,x)$ lipschitziana con constante de Lipschitz L, y sea el método con $\beta_k \neq 0$. Si $h < \dfrac{1}{|\beta_k|L}$ entonces (27) admite solución.

Gran parte del análisis de los MML se fundamenta en las propiedades que satisfacen los polinomios asociados. Uno de ellos es el *primer polinomio característico* ya definido en (7) pero su definición se reproduce aquí.

Definición 12 (*Polinomios característicos de un MML.*)

Dado el MML (27),

- $p(\lambda) = \lambda^k - \alpha_{k-1}\lambda^{k-1} - \cdots - \alpha_0$ *es su* primer polinomio característico;
- $q(\lambda) = \beta_k\lambda^k + \beta_{k-1}\lambda^{k-1} + \cdots + \beta_0$ *es su* segundo polinomio característico.

Así, un MML (27) es *consistente* con el PVI (1) si y solo si se verifica

1. $p(1) = 0$, es decir, $\displaystyle\sum_{j=0}^{k-1} \alpha_j = 1$,

2. $p'(1) = q(1)$, es decir, $\displaystyle k - \sum_{j=1}^{k-1} j\alpha_j = \sum_{j=0}^{k} \beta_j$.

4.1. Diseño de MML

Tanto para ayudarnos en la obtención de MML como para comprobar su orden, nos será útil el operador lineal asociado

$$\mathcal{L}_k(z(t);h) = z(t+kh) - \sum_{j=0}^{k-1} \alpha_j z(t+jh) - h\sum_{j=0}^{k} \beta_j z'(t+jh) \qquad (28)$$

donde $z \in \mathcal{C}^1[a,b]$. Con esto es fácil observar que el error de truncatura local es $R_{n+k} = \mathcal{L}_k(x(t_n);h)$ y su orden será $p \geq 1$ si y solo si $\mathcal{L}_k(z(t);h) = O(h^{p+1})$.

Proposición 2

Si $z(t)$ es suficientemente derivable, entonces

$$\mathcal{L}_k(z(t);h) = C_0 z(t) + C_1 z'(t)h + \cdots + C_m z^{(m)}(t)h^m + \cdots$$

donde

$$\boxed{\begin{aligned} C_0 &= 1 - \sum_{j=0}^{k-1} \alpha_j; \\ C_1 &= k - \sum_{j=1}^{k-1} j\alpha_j - \sum_{j=0}^{k} \beta_j; \\ C_m &= \frac{k^m}{m!} - \sum_{j=1}^{k-1} \frac{j^m}{m!}\alpha_j - \sum_{j=1}^{k} \frac{j^{m-1}}{(m-1)!}\beta_j, \quad m = 2,3,\ldots \end{aligned}} \tag{29}$$

Para la demostración bastará aplicar desarrollos de Taylor para cada uno de los términos $z(t+jh)$ y $z'(t+jh)$ en el punto t y agrupar por potencias sucesivas de h.

Teorema 9 (*Orden de un MML.*)

El MML (27) es de orden exactamente $p \geq 1$ si y solo si $C_0 = C_1 = \cdots = C_p = 0$ y $C_{p+1} \neq 0$.

La constante C_{p+1} se llama *constante principal de error* y el término $C_{p+1}h^{p+1}x^{(p+1)}(t_n)$ *parte principal del error de truncatura local*.

Teniendo en cuenta estas definiciones y resultados, pueden obtenerse MML de k pasos y orden p resolviendo sistemas de ecuaciones lineales de $p+1$ ecuaciones con $2k$ o $2k+1$ incógnitas α_j y β_j según se desee un MML explícito o implícito, respectivamente. Por tanto, en principio parece factible obtener MML explícitos de k pasos y orden $2k-1$, y MML implícitos de k pasos y orden $2k$. Sin embargo, podrían no ser convergentes. Un resultado debido a Dahlquist limita el orden de los MML convergentes.

Teorema 10 (*Dahlquist*)

El orden máximo de un MML (27) de k pasos cero-estable (o convergente) es $p = k + 1$ si k es impar, o $p = k + 2$ si k es par. Además existen métodos de orden máximo.

¿Qué métodos podrían tener orden máximo según el teorema de Dahlquist y máximo *global* teórico?. Hemos visto que el máximo global es $2k$ (implícitos), así que dependiendo de la paridad de k tendría que ser $k + 1 = 2k$ o $k + 2 = 2k$, lo cual solo deja dos posibilidades: $k = 1$ y $p = 2$ (por ejemplo el método del trapecio, del que veremos un ejemplo más adelante) o $k = 2$ y $p = 4$ (por ejemplo el método de Simpson, del que veremos un ejemplo más adelante). Un método con orden máximo global se califica de *óptimo y maximal*.

4.2. MML basados en cuadraturas

No hay que memorizarlos, nos los daría en el examen

Una de las principales vías de obtención de MML para el PVI (1) es utilizando las fórmulas de integración numérica de tipo interpolatorio clásico vistas en el Tema 2.

Nos pondrá convergencia, estabilidad y error

Supongamos que deseamos diseñar un MML de k pasos. Teniendo en cuenta que la ecuación diferencial $x'(t) = f(t, x(t))$ se puede expresar de forma equivalente como $x(\tau_2) - x(\tau_1) = \int_{\tau_1}^{\tau_2} f(s, x(s))\, ds$, entonces podemos poner

$$x(t_{n+k}) - x(t_{n+k-q}) = \int_{t_{n+k-q}}^{t_{n+k}} f(s, x(s))\, ds$$

con q comprendido entre 1 y k. Este enfoque asegura que el primer polinomio característico cumplirá $p(1) = 0$ ya que $\alpha_{k-q} = 1$ es su único coeficiente no nulo.

Ahora bien, para aproximar la integral se puede aplicar una fórmula de cuadratura basada en un cierto subconjunto de nodos anteriores $\{t_j\} \subseteq \{t_n, \ldots, t_{n+k}\}$ y sus correspondientes $\{f_j\} \subseteq \{f_n, \ldots, f_{n+k}\}$. Por simplificar supongamos un subconjunto de nodos consecutivos desde t_{n+m} hasta t_{n+k-r} (no tiene mucha utilidad un subconjunto de nodos no uniformemente espaciados). Tanto m como r pueden tomar valores enteros desde 0 hasta k, pero para que haya como mínimo un nodo se ha de cumplir $m+r \leq k$. No hay inconveniente en

que algunos de los nodos caigan fuera del intervalo de integración $[t_{n+k-q}, t_{n+k}]$. Por otro lado, para que el método resultante sea realmente de k pasos, debe aparecer x_n en algún lugar por lo que o bien $q = k$ o bien $m = 0$. Aparte de esto, podemos considerar al parámetro q independiente de m y r.

Aplicando a la integral la correspondiente fórmula de cuadratura de tipo interpolatorio clásico en los nodos elegidos, tenemos

$$\int_{t_{n+k-q}}^{t_{n+k}} f(s, x(s))\, ds \approx h \sum_{j=m}^{k-r} \beta_j f_{n+j}$$

lo cual nos permite escribir el *MML basado en cuadratura*

$$\boxed{x_{n+k} = x_{n+k-q} + h \sum_{j=m}^{k-r} \beta_j f_{n+j}.} \tag{30}$$

que será implícito si $r = 0$, explícito en caso contrario.

Dependiendo de los valores de q, m, r surgen algunas conocidas familias de métodos. Se citan seguidamente.

4.2.1. Métodos tipo Adams

Son aquellos con $q = 1$ (y por tanto $m = 0$).

Polinomio: $\lambda^k - \lambda^{k-1}$ ⇓ Es estable

$$x_{n+k} = x_{n+k-1} + h(\beta_0 f_n + \beta_1 f_{n+1} + \cdots + \beta_{k-r} f_{n+k-r}).$$

siendo $\alpha_0 = \cdots = \alpha_{k-2} = 0,\ \alpha_{k-1} = 1,\ \beta_0 \neq 0,\ \beta_{k-r+1} = \cdots = \beta_k = 0$.

Para este grupo de métodos las constantes C_i se reformularían de manera algo más simplificada:

$$\boxed{\begin{aligned} C_0 &= 0; \\ C_1 &= 1 - \sum_{j=0}^{k-r} \beta_j; \\ C_m &= \frac{k^m - (k-1)^m}{m!} - \sum_{j=1}^{k-r} \frac{j^{m-1}}{(m-1)!} \beta_j, \quad m = 2, 3, \ldots \end{aligned}} \tag{31}$$

4.2.2. Métodos de Adams-Bashforth (AB)

Son métodos tipo Adams explícitos con exactitud máxima: $q = 1$, $m = 0$, $r = 1$

$$x_{n+k} = x_{n+k-1} + h(\beta_0 f_n + \beta_1 f_{n+1} + \cdots + \beta_{k-1} f_{n+k-1}).$$

Ejemplos.

- Para $k = 1$ el método **AB1** es el método de Euler $x_{n+1} = x_n + h f_n$.

  Deducción mediante anulación de constantes C_i: el método sería en general

  $$x_{n+1} = x_n + h\beta_0 f_n \quad (\alpha_0 = 1,\ \beta_1 = 0)$$

  y anulando las constantes correspondientes:

  $$\begin{aligned} C_0 &= 0; \\ C_1 &= 1 - \sum_{j=0}^{k-1} \beta_j = 1 - \beta_0 = 0 \Rightarrow \boxed{\beta_0 = 1}. \end{aligned}$$

  Deducción mediante integración: de la regla de Sarrus

  $$\int_{t_n}^{t_{n+1}} f(t, x(t))\, dt = \int_{t_n}^{t_{n+1}} x'(t)\, dt = x(t_n + 1) - x(t_n)$$

  se obtiene

  $$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} f(t, x(t))\, dt$$

  y aplicando la fórmula del rectángulo izquierda se obtiene

  $$x(t_{n+1}) = x(t_n) + h f(t_n, x(t_n)) + \frac{h^2}{2} x''(\xi_n)$$

  de donde surge el método de Euler

  $$x_{n+1} = x_n + h f_n.$$

  ¿Qué pasaría si se usa otra fórmula de cuadratura, u otros nodos?

Con rectángulo derecha saldría el método de Euler implícito

$$x_{n+1} = x_n + hf_{n+1}.$$

Con punto medio tendríamos

$$x(t_{n+1}) \approx x(t_n) + hf\left(t_{n+\frac{1}{2}}, x\left(t_{n+\frac{1}{2}}\right)\right)$$

de donde, aproximando $x\left(t_{n+\frac{1}{2}}\right)$ por $x_n + \frac{h}{2}f_n$ saldría el método de Euler mejorado

$$x_{n+1} = x_n + hf\left(t_n + \frac{h}{2}, x_n + \frac{h}{2}f_n\right).$$

- Para $k = 2$ el método **AB2** es $x_{n+2} = x_{n+1} + \frac{h}{2}(3f_{n+1} - f_n)$.

Deducción mediante anulación de constantes C_i: el método sería en general

$$x_{n+2} = x_{n+1} + h(\beta_0 f_n + \beta_1 f_{n+1}) \quad (\alpha_0 = 0,\ \alpha_1 = 1,\ \beta_2 = 0)$$

y anulando las constantes correspondientes:

$$\begin{aligned}
C_0 &= 0;\\
C_1 &= 1 - \sum_{j=0}^{k-1} \beta_j \;=\; 1 - \beta_0 - \beta_1 \;=\; 0 \;\Rightarrow\; \beta_0 + \beta_1 = 1;\\
C_2 &= \frac{k^2 - (k-1)^2}{2!} - \sum_{j=1}^{k-1} j\beta_j \;=\; \frac{3}{2} - \beta_1 \;=\; 0\\
&\Rightarrow\; \boxed{\beta_1 = \frac{3}{2}} \;\Rightarrow\; \boxed{\beta_0 = -\frac{1}{2}}
\end{aligned}$$

Deducción mediante integración: de la regla de Sarrus

$$\int_{t_{n+1}}^{t_{n+2}} f(t, x(t))\, dt = \int_{t_{n+1}}^{t_{n+2}} x'(t)\, dt = x(t_n + 2) - x(t_{n+1})$$

se obtiene

$$x(t_{n+2}) = x(t_{n+1}) + \int_{t_{n+1}}^{t_{n+2}} f(t, x(t))\, dt.$$

No se puede aplicar la fórmula del trapecio porque saldría implícito. En otras palabras, para que sea explícito no se puede usar el nodo t_{n+2} en

la fórmula de cuadratura. Si aplicásemos rectángulo izquierda saldría trivialmente un método de Euler desplazado

$$x_{n+2} = x_{n+1} + hf_{n+1}.$$

sin ningún interés, y además de 1 paso porque no aparece t_n ni x_n. En definitiva, tiene que usarse el nodo t_n pero no el nodo t_{n+2} en la fórmula de cuadratura. Una fórmula de tipo interpolatorio que use t_n como único nodo sería

$$\int_{t_{n+1}}^{t_{n+2}} f(t, x(t))\, dt \approx hf(t_n, x(t_n))$$

que es una especie de rectángulo 'más a la izquierda', exacta en $\mathbb{P}_0$. Con esta fórmula se tendría el método

$$x_{n+2} = x_{n+1} + hf_n$$

que viene a ser como Euler, pero de muy escaso interés. Es preferible usar dos nodos, t_n y t_{n+1} para la fórmula:

$$\int_{t_{n+1}}^{t_{n+2}} h(t)\, dt \approx \alpha_0 h(t_n) + \alpha_1 h(t_{n+1})$$

que para que sea de tipo interpolatorio ha de ser

$$\alpha_0 = \frac{3}{2}h,\ \alpha_1 = -\frac{1}{2}h,$$

lo que da lugar al método **AB2**

$$x_{n+2} = x_{n+1} + \frac{h}{2}(3f_{n+1} - f_n).$$

- Para $k = 3$ el método **AB3** es $x_{n+3} = x_{n+2} + \frac{h}{12}(5f_n - 16f_{n+1} + 23f_{n+2})$.

  Deducción mediante anulación de constantes C_i: el método sería en general

  $$x_{n+3} = x_{n+2} + h(\beta_0 f_n + \beta_1 f_{n+1} + \beta_2 f_{n+2}) \quad (\alpha_0 = \alpha_1 = 0,\ \alpha_2 = 1,\ \beta_3 = 0)$$

y anulando las constantes correspondientes:

$$\begin{aligned}
C_0 &= 0;\\
C_1 &= 1-\sum_{j=0}^{k-1}\beta_j = 1-\beta_0-\beta_1-\beta_2 = 0 \Rightarrow \beta_0+\beta_1+\beta_2=1;\\
C_2 &= \frac{k^2-(k-1)^2}{2!}-\sum_{j=1}^{k-1} j\beta_j = \frac{5}{2}-\beta_1-2\beta_1 = 0\\
&\Rightarrow \beta_1+2\beta_2=\frac{5}{2}\\
C_3 &= \frac{k^3-(k-1)^3}{3!}-\sum_{j=1}^{k-1}\frac{j^2}{2!}\beta_j = \frac{19}{6}-\frac{1}{2}\beta_1-\frac{4}{2}\beta_2 = 0\\
&\Rightarrow \beta_1+4\beta_2=\frac{19}{3}\\
&\Rightarrow \boxed{\beta_0=\frac{5}{12},\ \beta_1=-\frac{16}{12},\ \beta_2=\frac{23}{12}}
\end{aligned}$$

Deducción mediante integración: de

$$x(t_{n+3}) = x(t_{n+2}) + \int_{t_{n+2}}^{t_{n+3}} f(t,x(t))\,dt,$$

usando una fórmula que emplee los nodos t_n, t_{n+1} y t_{n+2}

$$\int_{t_{n+2}}^{t_{n+3}} h(t)\,dt \approx \alpha_0 h(t_n)+\alpha_1 h(t_{n+1})+\alpha_2 h(t_{n+2})$$

de tipo interpolatorio en $\mathbb{P}_2$ resultaría

$$\alpha_0=\frac{5}{12},\ \alpha_1=-\frac{16}{12},\ \alpha_2=\frac{23}{12}$$

con lo que se tendría el método **AB3**.

- Para $k=4$ se tiene el método **AB4**

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

de obtención algo más laboriosa. En general sería

$$x_{n+4} = x_{n+3}+h(\beta_0 f_n+\beta_1 f_{n+1}+\beta_2 f_{n+2}+\beta_3 f_{n+3})$$

donde $(\alpha_0=\alpha_1=\alpha_2=0,\ \alpha_3=1,\ \beta_4=0)$

4.2.3. Métodos de Adams-Moulton (AM)

Son métodos tipo Adams implícitos con exactitud máxima: $q = 1$, $m = 0$, $r = 0$. El modelo general sería

$$x_{n+k} = x_{n+k-1} + h(\beta_0 f_n + \beta_1 f_{n+1} + \cdots + \beta_k f_{n+k}).$$

Para $k = 1$ se tiene $x_{n+1} = x_n + \frac{h}{2}(f_n + f_{n+1})$ (m. del trapecio).

Para $k = 2$ se tiene $x_{n+2} = x_{n+1} + \frac{h}{12}(-f_n + 8f_{n+1} + 5f_{n+2})$.

Para $k = 3$ se tiene $x_{n+3} = x_{n+2} + \frac{h}{24}(f_n - 5f_{n+1} + 19f_{n+2} + 9f_{n+3})$.

Para $k = 4$ se tiene $x_{n+4} = x_{n+3} + \frac{h}{720}(-19f_n + 106f_{n+1} - 264f_{n+2} + 646f_{n+3} + 251f_{n+4})$.

En general, si comparamos un método **AB** de k pasos con uno **AM** de $k - 1$ pasos se tiene:

- Ambos realizan k evaluaciones de la función f.
- Ambos tienen en el error de truncatura local el factor h^{k+1} y la derivada de orden $k + 1$ de x en un punto intermedio.
- Tanto los coeficientes β_j como la constante del error de truncatura local son más pequeños en **AM** que en **AB**. Esto hace que los métodos implícitos sean más estables que los explícitos.
- La semilla para resolver la ecuación en un método **AM** se suele obtener con un método explícito. Es decir, el método explícito predice el valor inicial y el implícito lo corrige reiteradamente, actuando como método de iteración funcional. Esta combinación se conoce como *predictor-corrector*. En la práctica conviene mejor hacer una sola corrección y usar un paso h pequeño, en lugar de varias correcciones.

Los métodos de Adams **AM** y **AB** son MML (30)

$$x_{n+k} = x_{n+k-q} + h \sum_{j=m}^{k-r} \beta_j f_{n+j}.$$

con $q = 1,\ m = 0$, ya sean explícitos ($r \geq 1$) o implícitos ($r = 0$). Veamos algunos MML con $q > 1$.

4.2.4. Métodos de Milne-Simpson generalizados

Son implícitos con $q = 2$, $m = 0$, $r = 0$

$$x_{n+k} = x_{n+k-2} + h(\beta_0 f_n + \beta_1 f_{n+1} + \cdots + \beta_k f_{n+k}),$$

es decir, $\alpha_j = 0\ \forall j \neq k-2,\ \alpha_{k-2} = 1,\ \beta_0 \neq 0,\ \beta_k \neq 0$.

4.2.5. Métodos Nyström

Son explícitos con $q = 2$, $m = 0$, $r \geq 1$

$$x_{n+k} = x_{n+k-2} + h(\beta_0 f_n + \beta_1 f_{n+1} + \cdots + \beta_{k-r} f_{n+k-r}).$$

4.2.6. Métodos tipo Newton-Cotes

Son métodos con $q = k$, es decir, del tipo

$$x_{n+k} = x_n + h(\beta_0 f_n + \beta_1 f_{n+1} + \cdots + \beta_k f_{n+k}).$$

Distinguiremos dos clases:

- Si $m = r = 1$ son métodos explícitos o abiertos:

$$x_{n+k} = x_n + h(\beta_1 f_{n+1} + \cdots + \beta_{k-1} f_{n+k-1})$$

- Si $m = r = 0$ son métodos implícitos o cerrados:

$$x_{n+k} = x_n + h(\beta_0 f_n + \cdots + \beta_k f_{n+k}),\ \beta_0 \neq 0,\ \beta_k \neq 0.$$

Ejemplos.

- Método del trapecio: $x_{n+1} = x_n + \dfrac{h}{2}(f_n + f_{n+1})$. Cerrado, implícito, de $k = 1$ paso. Si se sustituyera f_{n+1} por una aproximación $f(t_{n+1}, x_n + hf_n)$ sería el método de Heun, explícito.

- Método de Simpson: $x_{n+2} = x_n + \dfrac{h}{3}(f_n + 4f_{n+1} + f_{n+2})$.

- Fórmula abierta de 4 pasos: $x_{n+3} = x_{n-1} + \dfrac{4h}{3}(2f_n - f_{n+1} + 2f_{n+2})$.

4.3. Métodos predictor-corrector

Por lo general, los MML implícitos son sustancialmente más exactos que los explícitos de los mismos pasos y orden, por lo que vale la pena usarlos a pesar de las dificultades que conllevan por tener que resolver una ecuación en cada paso.

Recordemos que los teoremas 4 (caracterización de la estabilidad) y 5 (caracterización de la convergencia) siguen siendo válidos para MML.

Para resolver la ecuación de un MML implícito

$$x_{n+k} = \sum_{j=0}^{k-1} \alpha_j x_{n+j} + h \sum_{j=0}^{k-1} \beta_j f_{n+j} + h\beta_k f(t_{n+k}, x_{n+k}), \tag{32}$$

lo ideal sería usar un método de iteración funcional (véase el Tema 1) a partir de su misma formulación:

$$x_{n+k}^{(v)} = \sum_{j=0}^{k-1} \alpha_j x_{n+j} + h \sum_{j=0}^{k-1} \beta_j f_{n+j} + h\beta_k f(t_{n+k}, x_{n+k}^{(v-1)}) \tag{33}$$

partiendo de una semilla adecuada $x_{n+k}^{(0)}$. Surge así la cuestión de la convergencia de la sucesión de iteraciones $\{x_{n+k}^{(v)}\}_{v\geq 0}$. Sea K la constante de Lipschitz de f respecto de su segunda variable. Entonces, restando (32) y (33) se tiene

$$x_{n+k} - x_{n+k}^{(v)} = h\beta_k \left(f(t_{n+k}, x_{n+k}) - f(t_{n+k}, x_{n+k}^{(v-1)}) \right).$$

Así, podemos escribir

$$|x_{n+k} - x_{n+k}^{(v)}| \leq h|\beta_k|K|x_{n+k} - x_{n+k}^{(v-1)}|$$

y, por inducción,

$$|x_{n+k} - x_{n+k}^{(v)}| \leq (h|\beta_k|K)^v |x_{n+k} - x_{n+k}^{(0)}|$$

con lo que, finalmente, podríamos asegurar la convergencia del método iterativo sin más que tomar

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

es decir, con el tamaño de paso suficientemente pequeño.

Sin embargo, la velocidad de convergencia del método iterativo no pasa de ser simplemente lineal. Tomar un tamaño de paso h muy pequeño para que la constante asintótica del error sea pequeña y así aumentar la velocidad no suele compensar el esfuerzo computacional necesario. En la práctica es mucho más habitual emplear algún método (explícito, naturalmente) para elegir una semilla $x_{n+k}^{(0)}$ con la máxima precisión (predicción), para así realizar un bajo número de iteraciones (corrección). Bajo estas premisas surgen los *métodos predictor-corrector*.

Un MML predictor-corrector de k pasos tiene la forma general

$$\begin{array}{lrcl} \text{Predictor} & P: & x_{n+k}^{(0)} & = \displaystyle\sum_{j=0}^{k-1} \alpha_j^* x_{n+j} + h \sum_{j=0}^{k-1} \beta_j^* f_{n+j} \\ \text{Corrector} & C^m: & x_{n+k}^{(v)} & = \displaystyle\sum_{j=0}^{k-1} \alpha_j x_{n+j} + h \sum_{j=0}^{k-1} \beta_j f_{n+j} + h\beta_k f(t_{n+k}, x_{n+k}^{(v-1)}) \\ & & & v = 1, \ldots, m \end{array}$$

donde m indica el número (habitualmente fijo) de correcciones.

Ejemplo: Método de Adams-Bashforth-Moulton de orden 5. Se obtiene combinando un MML predictor AB de 5 pasos con un corrector AM de 4 pasos, aplicando una sola corrección.

$$P: x_{n+5}^{(0)} = x_{n+4} + \frac{h}{720}(1901 f_{n+4} - 2774 f_{n+3} + 2616 f_{n+2} - 1274 f_{n+1} + 251 f_n)$$

$$C^1: x_{n+5} = x_{n+4} + \frac{h}{720}(251 f(t_{n+5}, x_{n+5}^{(0)}) + 646 f_{n+4} - 264 f_{n+3} + 106 f_{n+2} - 19 f_{n+1})$$

4.3.1. Orden de un método predictor-corrector

Terminamos dando un resultado que relaciona el orden de un MML predictor-corrector con los órdenes de los métodos que lo componen.

Proposición 3

Sea un método PC^m. Sea p^* el orden del predictor P, p el del corrector C y m el número de correcciones. Entonces

1. *Si $p^* + m > p$, entonces el método PC^m tiene orden p y su constante principal de error de truncatura local es la misma de C.*
2. *Si $p^* + m = p$, entonces el método PC^m tiene orden p pero su constante principal de error de truncatura local es distinta de la de C.*
3. *Si $p^* + m < p$, entonces el método PC^m tiene orden $p^* + m$.*

En suma, el orden de PC^m es mín$\{p^*+m, p\}$. Por tanto, para economizar esfuerzos (porque a más alto orden mayor esfuerzo tanto de diseño como de computación), lo mejor es equilibrar para que sea $p^* + m = p$ y por tanto el mínimo se alcance en los dos términos. Siendo $m = 1$ la costumbre más extendida, entonces $p^* + 1 = p$.

5. Sistemas de ecuaciones diferenciales y ecuaciones de orden superior

La mayor parte de la teoría expuesta es extensible a sistemas de ecuaciones diferenciales. Si sustituimos el PVI (1) por

$$\boxed{\begin{array}{l} X' = F(t, X) \\ X(t_0) = \boldsymbol{\mu} \end{array}} \tag{34}$$

donde $X = X(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{pmatrix}$, $F(t, X) = \begin{pmatrix} f_1(t, x_1, \ldots, x_m) \\ \vdots \\ f_m(t, x_1, \ldots, x_m) \end{pmatrix}$ y $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}$.

Este problema puede resolverse mediante los métodos de discretización anteriores, y los conceptos y resultados relativos pasan a utilizar criterios de normas en lugar de valores absolutos. Cada método se escribiría en forma vectorial en lugar de en forma escalar. Por ejemplo, el método de Euler (14) aplicado al problema (34) se escribe como

$$\begin{array}{l} X_0 = \boldsymbol{\mu},\ t_0 = a \\ t_{n+1} = t_n + h,\ h = \frac{b-a}{N} \\ X_{n+1} = X_n + hF(t_n, X_n) \quad n = 0, \ldots, N-1 \end{array} \tag{35}$$

Por último, una ecuación diferencial de orden superior puede transformarse en un sistema de ecuaciones de primer orden. Como ilustración considérese el PVI de segundo orden

$$\boxed{\begin{array}{l} x'' = f(t, x, x') \\ x(t_0) = \mu_1,\ x'(t_0) = \mu_2 \end{array}} \tag{36}$$

Haciendo el cambio $x_1 = x(t)$, $x_2 = x'(t)$, tendremos que el problema (36) equivale al problema

$$\boxed{\begin{array}{l} X' = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = F(t, X) = \begin{pmatrix} x_2 \\ f(t, x_1, x_2) \end{pmatrix} \\ X(t_0) = \begin{pmatrix} x_1(t_0) \\ x_2(t_0) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \end{array}} \tag{37}$$

y de este modo la solución numérica de (36) estaría formada por las primeras componentes de la solución numérica de (37).