Equazioni differenziali ordinarie

Alvise Sommariva

Università degli Studi di Padova
Dipartimento di Matematica

10 maggio 2021

Problema di Cauchy

Problema. (Problema di Cauchy)

Si determini la funzione y tale che

$$\begin{cases} y'(x) = f(x, y(x)), \ x \geq x_0 \\ y(x_0) = y_0 \end{cases} \tag{1}$$

dove f è a valori in $\mathbb{R}^n$ e definita in un sottoinsieme Ω di $\mathbb{R} \times \mathbb{R}^n$,

con $(x_0, y_0) \in \Omega$.

Nota.

Di seguito supporremo che tale problema abbia una sola soluzione.

Esempi

Esempio (Equazione differenziale ordinaria)

$$\begin{cases} y'(x) = y(x), \ x \geq 0 \\ y(0) = 1 \end{cases} \tag{2}$$

la cui soluzione è $\exp(x)$*, è un problema di Cauchy con* $f(x, y(x)) = y(x)$.

Esempio (Sistema di equazioni differenziali ordinarie)

$$\begin{cases} y_1'(x) = -y_2(x), \ x \geq 0 \\ y_2'(x) = y_1(x), \ x \geq 0 \\ y_1(0) = 1, y_2(0) = 0 \end{cases} \tag{3}$$

la cui soluzione è $y(x) = (y_1(x), y_2(x)) = (\cos(x), \sin(x))$*, è un problema di Cauchy con* $y = (y_1, y_2)$, $x_0 = 0$, $y(x_0) = (1, 0)$ *e*

$$f(x, y(x)) = f(x, y_1(x), y_2(x)) = (-y_2(x), y_1(x)).$$

Teorema di Picard-Lindelöf

Teorema (Teorema di Picard-Lindelöf)

Sia $D = [t_0 - \varepsilon, t_0 + \varepsilon] \times \mathbb{R}^n$. *Dato il problema ai valori iniziali*

$$\begin{cases} y'(t) = f(t, y(t)), \ t \in [t_0 - \varepsilon, t_0 + \varepsilon] \\ y(t_0) = y_0 \end{cases}$$

se

- $f : D \to \mathbb{R}$ *è una funzione lipschitziana in* y *cioè*

$$|f(t, \mathbf{y_1}) - f(t, \mathbf{y_2})| \leq K \|\mathbf{y_1} - \mathbf{y_2}\|, (t, \mathbf{y_1}), (t, \mathbf{y_2}) \in D$$

- f *è continua in* t *allora per qualche* $\varepsilon > 0$,

allora esiste un'unica soluzione y *al problema ai valori iniziali sull'intervallo* $[t_0 - \varepsilon, t_0 + \varepsilon]$.

Teorema di Cauchy in piccolo

Teorema (Cauchy in piccolo, [2, p.7])

Si supponga che

- *D sia un aperto connesso di $\mathbb{R} \times \mathbb{R}^n$;*
- *$f : D \to \mathbb{R}$ sia una funzione continua in D;*
- *(t_0, y_0) un punto interno a D;*
- *la funzione f verifichi la condizione di Lipschitz*

$$\|f(t, \mathbf{y_1}) - f(t, \mathbf{y_2})\| \leq K\|\mathbf{y_1} - \mathbf{y_2}\|, (t, \mathbf{y_1}), (t, \mathbf{y_2}) \in D.$$

per qualche $K \geq 0$.

Allora esiste una unica funzione $\mathbf{y}$ definita su un intervallo $[t_0 - \alpha, t_0 + \alpha]$, per qualche $\alpha > 0$ tale che

$$\begin{cases} \mathbf{y}'(t) = f(t, \mathbf{y}(t)), \ t \in [t_0 - \alpha, t_0 + \alpha] \\ \mathbf{y}(t_0) = \mathbf{y_0} \end{cases} \tag{4}$$

Teorema di Cauchy in grande

Teorema (Cauchy in grande, [5, p.331])

Sia $f : [a, b] \times \mathbb{R}^n \to \mathbb{R}$ *tale che*

- $f(t, \mathbf{y})$, $t \in [a, b]$, $\mathbf{y} \in \mathbb{R}^n$ *continua nella prima variabile,*
- *soddisfi rispetto alla seconda variabile la condizione di Lipschitz*

$$\|f(t, \mathbf{y_1}) - f(t, \mathbf{y_2})\| \leq K\|\mathbf{y_1} - \mathbf{y_2}\|, t \in [a, b], \mathbf{y_1}, \mathbf{y_2} \in \mathbb{R}^n.$$

per qualche norma vettoriale $\|\cdot\|$ *e* $K \geq 0$.

Allora il problema di Cauchy

$$\begin{cases} \mathbf{y}'(t) = f(t, \mathbf{y}(t)), \ t \in [a, b] \\ \mathbf{y}(a) = \mathbf{y_0} \end{cases} \tag{5}$$

ha **una e una sola** *soluzione* $\mathbf{y}$ *in* $[a, b]$, *per un arbitrario* $\mathbf{y_0} \in \mathbb{R}^n$.

Teoremi di Cauchy

Nota.

- *Nel teorema di Cauchy in piccolo si fornisce un teorema di esistenza e unicità locale.*
- *Nel teorema di Cauchy in grande si fornisce un teorema di esistenza e unicità globale.*

Nota. (Soluzioni multiple)

Esistono casi in cui il problema di Cauchy ha soluzioni multiple. Ad esempio, il problema

$$\begin{cases} y'(x) = 2\sqrt{y(x)}, \ x > 0 \\ y(x_0) = 0 \end{cases} \tag{6}$$

ha per soluzioni, per $\alpha \in \mathbb{R}^+$

$$y(x) = \begin{cases} y(x) = (x-\alpha)^2, x \geq \alpha \\ y(x) = 0, \ \textit{altrimenti.} \end{cases} \tag{7}$$

Metodo di Eulero esplicito

Indichiamo con $\mathcal{I}(x, \overline{x})$ il più piccolo intervallo aperto contenente x e $\overline{x}$ (cioè $(x, \overline{x})$ oppure $(\overline{x}, x)$). Assumendo che la soluzione sia sufficientemente regolare, abbiamo dalla formula di Taylor per $x \approx \overline{x}$, $\xi \in \mathcal{I}(x, \overline{x})$ e (1)

$$\begin{aligned} y(x) &= y(\overline{x}) + y'(\overline{x})(x-\overline{x}) + \frac{y''(\xi)(x-\overline{x})^2}{2} \\ &\approx y(\overline{x}) + y'(\overline{x})(x-\overline{x}) = y(\overline{x}) + f(\overline{x}, y(\overline{x}))(x-\overline{x}) \quad (8) \end{aligned}$$

cioè

$$\boxed{y(x) \approx y(\overline{x}) + f(\overline{x}, y(\overline{x}))(x-\overline{x})}$$

Metodo di Eulero esplicito

Metodo (Eulero esplicito, 1768-1770)

Tale metodo consiste nell'approssimare $y(x_{k+1})$ *con* u_{k+1} *dove*

$$\boxed{u_{k+1} = u_k + h \cdot f(x_k, u_k),\ \ u_0 = y(x_0),} \tag{9}$$

ove $h = x_{k+1} - x_k$.

Supponiamo di voler determinare una approssimazione della soluzione nei punti equispaziati $x_{k+1} = x_0 + kh = x_k + h$ con $k > 0$. Da

$$y(x) \approx y(\overline{x}) + f(\overline{x}, y(\overline{x}))(x - \overline{x})$$

ponendo $x = x_{k+1}$, $\overline{x} = x_k$

$$y(x_{k+1}) \approx y(x_k) + h \cdot f(x_k, y(x_k)). \tag{10}$$

Metodo di Eulero implicito

Similmente, supposto $x_{k+1} = x_0 + kh = x_k + h$ con $k > 0$, da

$$y(x) \approx y(\overline{x}) + f(\overline{x}, y(\overline{x}))(x - \overline{x})$$

scelti $x = x_k$, $\overline{x} = x_{k+1}$ abbiamo

$$y(x_k) \approx y(x_{k+1}) + f(x_{k+1}, y(x_{k+1}))(x_k - x_{k+1}), \tag{11}$$

e quindi portato il termine $f(x_{k+1}, y(x_{k+1}))(x_k - x_{k+1})$ nel membro di sinistra

$$y(x_{k+1}) \approx y(x_k) + h \cdot f(x_{k+1}, y(x_{k+1})). \tag{12}$$

Metodo (Metodo di Eulero implicito)

Tale metodo consiste nell'approssimare $y(x_{k+1})$ con u_{k+1} definito da

$$\boxed{u_{k+1} = u_k + h \cdot f(x_{k+1}, u_{k+1}),\ \ u_0 = y(x_0)} \tag{13}$$

Eulero implicito

Nota. (Risoluzione equazioni nonlineari)

Evidentemente ad ogni iterazione del metodo di Eulero implicito

$$u_{k+1} = u_k + h \cdot f(x_{k+1}, u_{k+1}), \;\; u_0 = y(x_0)$$

si richiede di risolvere un'equazione nonlineare nella variabile z del tipo

$$\boxed{z = u_k + h \cdot f(x_{k+1}, z)}$$

la cui soluzione può essere calcolata utilizzando ad esempio

- *col metodo di Newton o*
- *del punto fisso o*
- *delle secanti.*

Osservazione

Nota. (Metodi di Eulero e sistemi di equazioni differenziali)

Il problema di Cauchy è definito da

$$\begin{cases} y'(x) = f(x, y(x)), \ x \geq x_0 \\ y(x_0) = y_0 \end{cases} \tag{14}$$

dove f è a valori in $\mathbb{R}^n$ e definita in un sottoinsieme Ω di $\mathbb{R} \times \mathbb{R}^n$, con

$(x_0, y_0) \in \Omega$.

Nei casi multivariati con $n > 1$, indipendentemente dalla derivazione dovuta alla espansione di Taylor, i metodi

- *Eulero esplicito:* $u_{k+1} = u_k + h \cdot f(x_k, u_k)$, $u_0 = y(x_0)$
- *Eulero implicito:* $u_{k+1} = u_k + h \cdot f(x_{k+1}, u_{k+1})$, $u_0 = y(x_0)$,

possono essere utilizzati, in quanto la sequenza è ben definita.

Linear Multistep Methods

Definizione (Linear Multistep Methods)

I metodi Linear Multistep (abbreviati con LMM) approssimano $y(x_n)$ *con* u_n*, dove* $x_n = x_0 + nh$ *e*

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j})$$

Nota.

Si osservi che

- *Sono metodi per cui se voglio calcolare* u_{n+1} *necessitano* $u_{n-p}, \ldots, u_n$ *(si chiamano metodi a* $p+1$ *passi). Siccome inizialmente si dispone solo di* u_0*, per calcolare* $u_1, \ldots, u_p$*, necessari ad innescare il metodo, si utilizzano di solito altre strategie (come ad esempio applicare più volte metodi a un passo, cioè in cui* $p = 0$*, come ad esempio Eulero implicito).*
- *Se* $b_{-1} = 0$ *allora il metodo è esplicito, altrimenti è implicito.*

Linear Multistep Methods

Nota.

Osserviamo che il metodo di Eulero esplicito

$$u_{n+1} = u_n + h \cdot f(x_n, u_n)$$

e Eulero implicito

$$u_{n+1} = u_n + h \cdot f(x_{n+1}, u_{n+1})$$

hanno la forma

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j})$$

ponendo rispettivamente

- *Eulero esplicito:* $p = 0,\ a_0 = 1,\ b_{-1} = 0,\ b_0 = 1,$
- *Eulero implicito:* $p = 0,\ a_0 = 1,\ b_{-1} = 1,\ b_0 = 0.$

Metodi basati sull'integrazione numerica ([8, 125])

Si supponga di dover studiare il problema di Cauchy

$$\begin{cases} y'(x) = f(x, y(x)), \ x \geq x_0 \\ y(x_0) = y_0 \end{cases} \tag{15}$$

e di dover approssimare la sua soluzione y nei punti equispaziati $x_k = x_0 + kh$. Osserviamo che

$$y(x_{n+k}) = y(x_{n-j}) + \int_{x_{n-j}}^{x_{n+k}} y'(t)dt \tag{16}$$

$$= y(x_{n-j}) + \int_{x_{n-j}}^{x_{n+k}} f(t, y(t))dt \tag{17}$$

Sostituiamo ora l'integranda con il polinomio di grado al piú p

$$\pi_p(x) = \sum_{i=0}^{p} f(x_{n-i}, y(x_{n-i}))L_i(x)$$

che interpola le coppie $(x_l, f(x_l, y(x_l))$ per $l = n, n-1, \ldots, n-p$, dove

$$L_i(x) = \prod_{k=0, k\neq i}^{p} \frac{x - x_{n-k}}{x_{n-i} - x_{n-k}}.$$

Metodi basati sull'integrazione numerica ([8, 125])

Ricaviamo cosí, da $f(x, y(x)) \approx \pi_p(x) = \sum_{i=0}^{q} f(x_{n-i}, y(x_{n-i}))L_i(x)$, posto

$$\beta_{p,i}^{(k,j)} = \frac{1}{h}\int_{x_{n-j}}^{x_{n+k}} L_i(t)dt = \int_{-j}^{k} \prod_{m=0, m\neq i}^{p} \frac{s+m}{-i+m} ds, i = 0, 1, \ldots, p$$

i metodi di tipo LMM, a più passi

$$\begin{aligned}
y(x_{n+k}) &= y(x_{n-j}) + \int_{x_{n-j}}^{x_{n+k}} f(t, y(t))dt \\
&\approx y(x_{n-j}) + \int_{x_{n-j}}^{x_{n+k}} \sum_{i=0}^{p} f(x_{n-i}, y(x_{n-i}))L_i(t)dt \\
&= y(x_{n-j}) + \sum_{i=0}^{p} f(x_{n-i}, y(x_{n-i})) \int_{x_{n-j}}^{x_{n+k}} L_i(t)dt \\
&= y(x_{n-j}) + h\sum_{i=0}^{p} \beta_{p,i}^{(k,j)} f(x_{n-i}, y(x_{n-i})). \qquad (18)
\end{aligned}$$

Metodi di Adams-Bashforth (1883)

Ponendo $k = 1$, $j = 0$ in

$$y(x_{n+k}) \approx y(x_{n-j}) + h \sum_{i=0}^{p} \beta_{p,i}^{(k,j)} f(x_{n-i}, y(x_{n-i})).$$

otteniamo la famiglia di metodi espliciti di Adams-Bashforth (1883)

$$u_{n+1} = u_n + \sum_{i=0}^{p} \beta_{p,i}^{(1,0)} f(x_{n-i}, u_{n-i}), \quad u_{n-i} \approx y(x_{n-i}).$$

Alcuni esempi sono i metodi a $p+1$ passi:

- Adams-Bashforth (p=0): $u_{n+1} = u_n + hf_n$ (Eulero esplicito);
- Adams-Bashforth (p=1): $u_{n+1} = u_n + \frac{h}{2}(3f_n - f_{n-1})$;
- Adams-Bashforth (p=2): $u_{n+1} = u_n + \frac{h}{12}(23f_n - 16f_{n-1} + 5f_{n-2})$;
- Adams-Bashforth (p=3): $u_{n+1} = u_n + \frac{h}{24}(55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3})$.

dove $f_{n-3} = f(x_{n-3}, u_{n-3})$, $f_{n-2} = f(x_{n-2}, u_{n-2})$, $f_{n-1} = f(x_{n-1}, u_{n-1})$, $f_n = f(x_n, u_n)$.

Metodi di Adams-Moulton (1926)

Ponendo $k = 0$, $j = 1$ in

$$y(x_{n+k}) \approx y(x_{n-j}) + h \sum_{i=0}^{p} \beta_{p,i}^{(k,j)} f(x_{n-i}, y(x_{n-i})).$$

otteniamo la famiglia di metodi impliciti di Adams-Moulton (1926)

$$u_n = u_{n-1} + h \sum_{i=0}^{p} \beta_{p,i}^{(0,1)} f(x_{n-i}, u_{n-i}), \quad u_{n-i} \approx y(x_{n-i}).$$

Sostituendo $n+1$ al posto di n, abbiamo cosí

$$\begin{aligned} u_{n+1} &= u_n + h \sum_{i=0}^{p} \beta_{p,i}^{(0,1)} f(x_{n+1-i}, u_{n+1-i}) \\ &= u_n + h \sum_{j=-1}^{p-1} \beta_{p,j}^{(0,1)} f(x_{n-j}, u_{n-j}). \end{aligned} \tag{19}$$

Metodi di Adams-Moulton (1926)

Alcuni esempi di metodi di Adams-Moulton sono i metodi impliciti

- Adams-Moulton (p=0): $u_{n+1} = u_n + hf_{n+1}$ (Eulero implicito);
- Adams-Moulton (p=1): $u_{n+1} = u_n + \frac{h}{2}(f_{n+1} + f_n)$ (Crank-Nicolson);
- Adams-Moulton (p=2): $u_{n+1} = u_n + \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1})$;
- Adams-Moulton (p=3): $u_{n+1} = u_n + \frac{h}{24}(9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2})$.

dove $f_{n-2} = f(x_{n-2}, u_{n-2})$, $f_{n-1} = f(x_{n-1}, u_{n-1})$, $f_n = f(x_n, u_n)$, $f_{n+1} = f(x_{n+1}, u_{n+1})$.

- Si verifica facilmente che tanto i metodi Adams-Bashforth quanto gli Adams-Moulton sono di tipo Linear Multistep Methods.
- *The Adams-Bashforth methods were designed by John Couch Adams to solve a differential equation modelling capillary action due to Francis Bashforth. Bashforth (1883) published his theory and Adams'numerical method* (Goldstine 1977).
- *The Adams-Moulton methods are solely due to John Couch Adams, like the Adams-Bashforth methods. The name of Forest Ray Moulton became associated with these methods because he realized that they could be used in tandem with the Adams-Bashforth methods as a predictor-corrector pair.*

Metodi di Nyström (1925)

Ponendo $k = 1$, $j = 1$ in

$$y(x_{n+k}) \approx y(x_{n-j}) + h \sum_{i=0}^{p} \beta_{p,i}^{(k,j)} f(x_{n-i}, y(x_{n-i})).$$

otteniamo la famiglia di metodi di Nyström (1925)

$$u_{n+1} = u_{n-1} + h \sum_{i=0}^{p} \beta_{p,i}^{(1,1)} f(x_{n-i}, u_{n-i}), \quad u_{n-i} \approx y(x_{n-i}).$$

Un esempio è il *metodo del punto medio* $u_{n+1} = u_{n-1} + 2hf_n$.

Nota.

- *Un metodo molto noto in questa famiglia è quello di Milne*

$$u_{n+1} = u_{n-1} + (h/3)\,(f_{n-1} + 4f_n + f_{n+1})$$

  ottenuto ponendo in (18) i parametri $k = 0$, $j = 2$, $p = 2$ e sostituendo n con $n + 1$,
- *Esistono altre grandi famiglie di metodi LMM, come i metodi BDF (cf. [6, p.437]), utili in problemi stiff.*

Consistenza

Definizione (Errori di troncamento, [2], p.112)

Si risolva un problema di Cauchy (1) nel compatto $[x_0, b]$ *con il metodo LMM*

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j}), \qquad u_k \approx y(x_k)$$

con y *soluzione di (1),* $x_k = x_0 + kh$, $k = 0, \ldots, N$, $b - h < x_N \leq b$, *e sia*

$$\tau_n(h) := \frac{1}{h} \cdot \left(y(x_{n+1}) - (\sum_{j=0}^{p} a_j y(x_{n-j}) + h \sum_{j=-1}^{p} b_j f(x_{n-j}, y(x_{n-j}))) \right).$$

- *La quantità* $\tau_n(h)$ *chiama errore locale di troncamento del metodo.*
- *La quantità* $\tau(h) = \max_{n=0,\ldots} |\tau_n(h)|$, *assunto che sia stato commesso errore nullo nei passi iniziali, si chiama* **errore globale di troncamento** *del metodo.*
- *Un metodo per cui* $\tau(h)$ *tende a* 0, *quando* $h \to 0$ *si dice* **consistente**.
- *Se* $\tau(h) = O(h^k)$ *allora il metodo ha* **ordine di consistenza** *"k".*

Analisi convergenza: consistenza

La consistenza rende conto di come la soluzione y del problema di Cauchy verifichi lo schema discreto del metodo linear multistep

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j}).$$

Osserviamo inoltre che per metodi espliciti la quantitá detta *residuo*

$$y(x_{n+1}) - (\sum_{j=0}^{p} a_j y(x_{n-j}) + h \sum_{j=-1}^{p} b_j f(x_{n-j}, y(x_{n-j})))$$

non è altro che l'errore dello schema numerico, qualora si supponga $u_{n-j} = y(x_{n-j})$, $j = 0, \ldots, p$.

Consistenza Eulero esplicito

Teorema (Errore locale di troncamento di Eulero esplicito)

Sia $y \in C^2([a,b])$ l'unica soluzione del problema di Cauchy (1) in $[a,b]$ e siano $x_k = a + kh$, $b - h < x_N \leq b$. Allora l'errore locale di troncamento di Eulero esplicito risulta

$$\boxed{\tau_k(h) = \frac{y''(\xi_k)h}{2}}$$

per qualche $\xi_k \in (x_k, x_{k+1})$, $k \in 0, \ldots, N-1$.

Nota.

Se $y \in C^2(a,b)$, allora per $M = \max_{x\in[a,b]} |y''(x)|$

$$\tau(h) = \max_{n=0,\ldots,N} |\tau_k(h)| = \max_{n=0,\ldots,N} \left|\frac{y''(\xi_k)h}{2}\right| \leq \frac{Mh}{2}$$

ovvero il metodo di Eulero esplicito è consistente.

Consistenza Eulero esplicito

Dimostrazione.

Osserviamo che se $y \in C^2([a,b])$, *dalla formula di Taylor*

$$y(x_{k+1}) = y(x_k) + y'(x_k)(x_{k+1} - x_k) + \frac{y''(\xi)(x_{k+1} - x_k)^2}{2}$$

per qualche $\xi \in (x_n, x_{n+1})$ *ovvero da* $h = x_{k+1} - x_k$, $y'(x_k) = f(x_k, y(x_k))$

$$y(x_{k+1}) - y(x_k) - h \cdot f(x_k, y(x_k)) = \frac{y''(\xi)h^2}{2}$$

Di conseguenza,

$$\begin{aligned} \tau_k(h) &= \frac{1}{h}\left(y(x_{k+1}) - y(x_k) - hf(x_k, y(x_k))\right) \\ &= \frac{1}{h}\frac{y''(\xi_k)h^2}{2} = \frac{y''(\xi_k)h}{2}. \end{aligned} \tag{20}$$

△

Consistenza Eulero implicito

Teorema (Errore locale di troncamento di Eulero implicito)

Sia $y \in C^2([a,b])$ l'unica soluzione del problema di Cauchy (1) in $[a,b]$ e siano $x_k = a + kh$, $b - h < x_N \leq b$. Allora l'errore locale di troncamento di Eulero implicito risulta

$$\boxed{\tau_k(h) = -\frac{y''(\xi_k)h}{2}}$$

per qualche $\xi \in (x_k, x_{k+1})$, $k = 0, \ldots, N-1$.

Nota.

Se $y \in C^2(a,b)$, allora per $M = \max_{x\in[a,b]} |y''(x)|$

$$\tau(h) = \max_{n=0,\ldots,N} |\tau_k(h)| = \max_{k=0,\ldots,N-1} \left|\frac{-y''(\xi_k)h}{2}\right| \leq \frac{Mh}{2} \qquad (21)$$

ovvero il metodo di Eulero implicito è consistente.

Consistenza Eulero implicito

Dimostrazione.

Osserviamo che se $y \in C^2([a,b])$, *dalla formula di Taylor*

$$y(x_k) = y(x_{k+1}) + y'(x_{k+1})(x_k - x_{k+1}) + \frac{y''(\xi_k)(x_k - x_{k+1})^2}{2}$$

per qualche $\xi_k \in (x_k, x_{k+1})$ *ovvero da* $h = x_{k+1} - x_k$,
$y'(x_k) = f(x_k, y(x_k))$

$$y(x_{k+1}) - y(x_k) - h \cdot f(x_{k+1}, y(x_{k+1})) = -\frac{y''(\xi_k i)h^2}{2}$$

Di conseguenza,

$$\begin{aligned} \tau_k(h) &= \frac{1}{h}\left(y(x_{k+1}) - y(x_k) - hf(x_k, y(x_k))\right) \\ &= \frac{1}{h} \cdot \frac{-y''(\xi_k)h^2}{2} = -\frac{y''(\xi_k)h}{2}. \end{aligned} \tag{22}$$

△

Consistenza LMM

Teorema ([6, p.437])

Un metodo linear multistep

- *è consistente se e solo se*

$$\sum_{j=0}^{p} a_j = 1, \quad -\sum_{j=0}^{p} j a_j + \sum_{j=-1}^{p} b_j = 1,$$

- *se $y \in C^{q+1}$ per $q \geq 1$, ha ordine di consistenza q se e solo se è consistente e*

$$\sum_{j=0}^{p} (-j)^i a_j + i \sum_{j=-1}^{p} (-j)^{i-1} b_j = 1, i = 2, \ldots, q.$$

Consistenza LMM

Esempio

- *Il metodo di Eulero Esplicito ha parametri* $a_0 = 1$, $b_{-1} = 0$ *e* $b_0 = 1$, $p = 0$ *ed é immediato vedere che vale formula relativa alla consistenza di ordine* 1.
- *Il metodo di Eulero Implicito ha parametri* $a_0 = 1$, $b_{-1} = 1$ *e* $b_0 = 0$, $p = 0$ *ed é immediato vedere che vale formula relativa alla consistenza di ordine* 1.
- *Il metodo di Crank-Nicolson ha parametri* $a_0 = 1$, $b_{-1} = 1/2$, $b_0 = 1/2$, $p = 0$, *ed é immediato vedere che vale formula relativa alla consistenza di ordine* 1. *Inoltre per* $i = 2$ *si ha*

$$\sum_{j=0}^{p}(-j)^i a_j + 2\sum_{j=-1}^{p}(-j)^{i-1} b_j = \ldots = 2b_{-1} = 1$$

e quindi se $y \in C^3([a,b])$ *ha consistenza di ordine 2.*

Stabilità di un metodo LMM

Definizione (Zero stabile, [9, p.332])

Si consideri un metodo numerico LMM a p passi per la soluzione di problemi di Cauchy (1) nell intervallo $[a, b]$. *Siano* $x_k = a + kh_N$ *con* $h_N = (b-a)/N$, $k = 0, \ldots, N$. *Tale metodo si dice zero-stabile se qualsiasi siano le sequenze* $\{u_k^{(h_N)}\}$, $\{v_k^{(h_N)}\}$ *che sono state generate dal metodo con dati iniziali rispettivamente* $u_0^{(h_N)}, \ldots, u_{p-1}^{(h_N)}$, $v_0^{(h_N)}, \ldots, v_{p-1}^{(h_N)}$, *abbiamo*

$$\max_{s=0,\ldots,N} |u_k^{(h_N)} - v_k^{(h_N)}| \leq K \max_{s=0,\ldots,p-1} (|u_s^{(h_N)} - v_s^{(h_N)}|)$$

per $h_N \leq h^*$, *con* K *indipendente da* h_N.

- Questo concetto di stabilità viene detto zero-stabilità, in quanto determinato esclusivamente dallo studio relativo al problema (1) con $f(x, y) = 0$ (cf. [7, p.440]).
- In queste note ci interesseremo alla convergenza e consistenza di metodi LMM, ma esistono definizioni che permettono l'analisi di altri metodi piú generali.

Stabilità di un metodo LMM

In generale risulta tedioso studiare caso per caso la stabilità dei LMM

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j})$$

Risulta più semplice studiare la cosidetta root-condition che è un concetto equivalente.

In particolare questa è verificata se e solo se il polinomio caratteristico

$$\rho(z) = z^{p+1} - \sum_{k=0}^{p} a_k z^k$$

ha tutte le radici nel disco $B(0,1) \subset \mathbb{C}$ e quelle in $\partial B(0,1)$ hanno molteplicità 1.

Stabilità di un metodo LMM

Esempio

Nel caso di

- *Eulero esplicito:* $u_{n+1} = u_n + hf_n$,
- *Eulero implicito:* $u_{n+1} = u_n + hf_{n+1}$,
- *Crank-Nicolson:* $u_{n+1} = u_n + \frac{h}{2}(f_{n+1} + f_n)$,

abbiamo

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j})$$

con $a_0 = 1$, $p = 0$, *da cui*

$$\rho(z) = z^{p+1} - \sum_{k=0}^{p} a_k z^k = z - 1$$

che ha quale unica soluzione $z^* = 1$ *che ovviamente è in* $\partial B(0,1)$ *e ha molteplicità* 1. *Quindi tali metodi sono zero-stabili.*

Convergenza di un metodo LMM

Definizione (Convergenza)

Supponiamo di analizzare il problema di Cauchy nell'intervallo compatto $I = [x_0, x_{fin}]$. *Siano* $x_s = x_0 + sh_N$, *con* $Nh_N = x_{fin} - x_0$,

- $y^{(h_N)} = \{y(x_k)\}_{k=0,\ldots,N}$ *la soluzione esatta di un fissato problema di Cauchy,*
- $u^{(h_N)} = \{u_k\}_{k=0,\ldots,N}$ *dove* u_k *è l'approssimazione di* $y(x_k)$ *fornita nei punti* x_k *da un metodo numerico a* $p+1$ *passi.*

Un metodo Linear Multistep a $p+1$ *passi si dice* **convergente** *se qualora i passi iniziali* $u_0^{(h_N)}, \ldots, u_p^{(h_N)}$ *sono tali che*

$$\eta(h_N) = \max_{0 \le k \le p} |y_k^{(h_N)} - u_k^{(h_N)}| \to 0 \text{ per } h_N \to 0$$

si ha che

$$\|y^{(h_N)} - u^{(h_N)}\|_\infty \to 0 \text{ per } h_N \to 0.$$

Qualora $\eta(h_N) \le C_1 h^q$ *si ha* $\|y^{(h_N)} - u^{(h_N)}\|_\infty \le C_2 h^q$, *con* C_1, C_2 *indipendenti da* h, *allora il metodo si dice convergente con ordine* q.

Analisi convergenza

Importante.

Un teorema dovuto a Lax e Richtmyer (1956), mostra che un metodo è convergente se e solo se consistente e stabile.

Di conseguenza, se vengono verificate

- *le condizioni di consistenza*
- *la root condition,*

un metodo è convergente.

Corollario

I metodi di Eulero esplicito ed Eulero implicito sono convergenti.

Esempio: Convergenza di Eulero esplicito

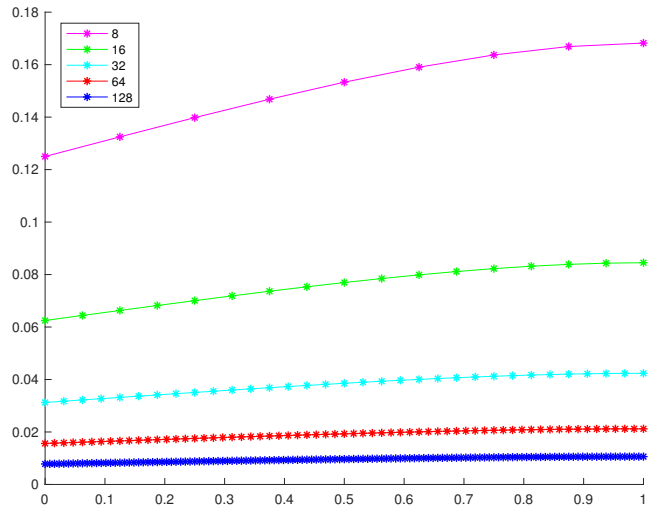

Figura: Si consideri il problema di Cauchy $y'(x) = y(x)$ in $[0, 1]$ con dato iniziale $y(0) = 1$, avente soluzione $y(x) = \exp(x)$ e si applichi il metodo di Eulero Esplicito $u_{n+1} = u_n + h_k f(x_n, u_n)$ per approssimare la soluzione nei punti $x_n = nh_k$, $n = 0, \ldots, N_k$ ove $h_k = 1/N_k$, con dato iniziale $u_0 = 1 + h_k$ per i valori $N_1 = 8$, $N_2 = 16$, $N_3 = 32$, $N_4 = 64$, $N_5 = 128$. Nel grafico in figura sono illustrati gli errori rispetto alla soluzione esatta, e il metodo sembra essere convergente.

Esempio: un metodo consistente ma non stabile

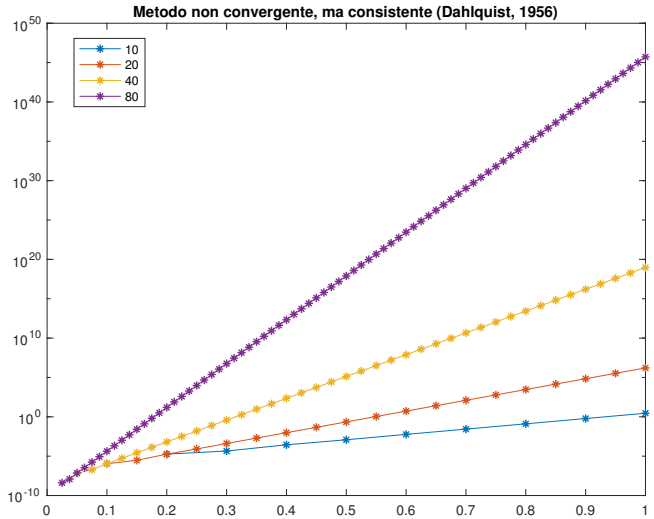

Figura: Si consideri il problema di Cauchy $y'(x) = y(x)$ in $[0, 1]$ con dato iniziale $y(0) = 1$, avente soluzione $y(x) = \exp(x)$ e si applichi il metodo consistente $u_{n+2} = -4u_{n+1} + 5u_n + h_k(4f(x_{n+1}, u_{n+1}) + 2f(x_n, u_n))$ per approssimare la soluzione nei punti $x_n = nh_k$, $n = 0, \ldots, N_k$ ove $h_k = 1/N_k$, con dato iniziale $u_0 = 1$, $u_1 = exp(h_k)$ per i valori $N_k = 2^{k-1} \cdot 10$. Nel grafico in figura sono illustrati gli errori rispetto alla soluzione esatta. Gli esperimenti mostrano che il metodo, suggerito quale esempio da Dahlquist nel 1956 in [4, p.37], non sembra convergere (anzi, al crescere di N_k diverge maggiormente).

Convergenza di un metodo LMM

Teorema (Convergenza LMM (cf.[2, p.116]), facoltativo)

Sia $u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j})$ *e si supponga*

- *si debba studiare il problema di Cauchy (1) in* $[x_0, x_{fin}]$, *che* $f : [x_0, x_{fin}] \times \mathbb{R} \to \mathbb{R}$ *sia continua e L-lipschitziana, rispetto la seconda variabile e che la soluzione* $y \in C^{(1)}([x_0, x_{fin}])$,
- *i dati iniziali siano calcolati in modo che*

$$\eta(h) = \max_{i=0,\ldots,p} |y(x_i) - u_i| \to 0, \quad \text{per } h \to 0;$$

- *il metodo sia consistente;*
- $a_j \geq 0, \quad j = 0, \ldots, p;$
- *il passo di discretizzazione* h *sia tale che* $h \leq h^* = 1/(2c)$, *dove* $c = L \sum_{j=-1}^{p} |b_j|$ *e* L *è la costante di Lipschitz di* f.

Allora il metodo LMM converge ed inoltre esistono C_1, C_2 *positive tali che per ogni* n

$$|y(x_n) - u_n| \leq C_1 \eta(h) + C_2 \tau(h).$$

Se $y \in C^{(m+1)}([x_0, x_{fin}])$, *il metodo ha ordine di consistenza* m *e gli errori iniziali soddisfano* $\eta(h) = O(h^m)$ *allora il metodo è di ordine* m.

Convergenza di un metodo LMM

Nota.

Nelle ipotesi del teorema precedente, per ricavare la convergenza chiedavamo:

- *i dati iniziali siano calcolati in modo che*

$$\eta(h) = \max_{i=0,\ldots,p} |y(x_i) - u_i| \to 0, \quad \text{per } h \to 0;$$

- *il metodo sia consistente;*
- $a_j \geq 0, \quad j = 0, \ldots, p;$
- *il passo di discretizzazione h sia tale che $h \leq 1/(2c)$, dove $c = L \sum_{j=-1}^{p} |b_j|$ e L è la costante di Lipschitz di f.*

1. Se prendiamo un metodo di Adams-Bashforth o Adams-Moulton di quelli esposti, sono tutti consistenti con ordine $p+1$. Inoltre $a_0 = 1, a_1 = 0, \ldots, a_p = 0$. Di conseguenza, essendo gli zeri del polinomio caratteristico $z^{p+1} - 1 = 0$ uguali alle radici $p+1$-esime primitive dell'unitá, tutti i metodi esposti sono convergenti.

2. Se la soluzione è $p+2$ volte derivabile con continuità, il metodo ha ordine di consistenza $p+1$ e gli errori iniziali soddisfano $\eta(h) = O(h^m)$ allora il metodo è di ordine $p+1$.

Convergenza di un metodo LMM

In particolare, nelle ipotesi richieste,

- *il metodo di Eulero esplicito, avendo ordine di consistenza* 1, *è convergente con ordine di convergenza* 1;
- *il metodo di Eulero implicito, avendo ordine di consistenza* 1, *è convergente con ordine di convergenza* 1;
- *il metodo di Crank-Nicolson, avendo ordine di consistenza* 2, *è convergente con ordine di convergenza* 2.

Regione di assoluta stabilità

Nell'ambito delle equazioni differenziali ordinarie, esistono vari criteri di stabilitá. Un classico problema è quello di vedere se un metodo è assolutamente stabile (spesso A-stabile).
Definito il problema di Cauchy

$$\begin{cases} y'(x) = \lambda y(x), \ x \geq 0 \\ y(0) = 1 \end{cases} \tag{23}$$

per un certo $\lambda \in \mathbb{C}$ con $\Re(\lambda) < 0$, visto che l'unica soluzione è $y(x) = \exp(\lambda x)$ si cerca di definire il passo h cosicchè il metodo numerico per la soluzione del problema di Cauchy (1) abbia lo stesso comportamento asintotico di $\exp(\lambda x)$.

La regione di assoluta stabilità è composta dagli $h\lambda$, per cui il metodo numerico con passo h è tale da avere lo stesso comportamento asintotico del problema di Cauchy (23) relativo al parametro λ (con $\Re(\lambda) < 0$).

A-stabilità. Una nota.

Nota. (Dahlquist, 1979)

I didn't like all these strong, perfect, absolute, generalized, super, hyper, complete and so on in mathematical definitions, I wanted something neutral; and having been impressed by David Young's property A, I chose the term A-stable.

Regione di assoluta stabilità

Nota.

Stabilire la stabilità per il problema di Cauchy

$$\begin{cases} y'(x) = f(x, y(x)), \ x \geq 0 \\ y(x_0) = y_0 \end{cases} \tag{24}$$

è in generale complicato. Se invece consideriamo

$$\begin{cases} y'(x) = \lambda y(x) + g(x), \ x \geq 0 \\ y(0) = y_0 \end{cases} \tag{25}$$

la questione è più semplice.
Ricordiamo che la soluzione di (25) è

$$y(x) = y_0 \cdot \exp(\lambda x) + \int_0^x \exp(\lambda(x-t)) g(t) dt$$

(cf. [1, p.369]).

Regione di assoluta stabilità

Sia y la soluzione di

$$\begin{cases} y'(x) = \lambda y(x) + g(x), \ x \geq 0 \\ y(0) = y_0 \end{cases} \tag{26}$$

e y_ϵ quella di

$$\begin{cases} y'(x) = \lambda y(x) + g(x), \ x \geq 0 \\ y(0) = y_0 + \epsilon \end{cases} \tag{27}$$

Posto $z_\epsilon = y_\epsilon - y$, da (26), (27),

$$\begin{cases} z'_\epsilon(x) = \lambda z_\epsilon(x), \ x \geq 0 \\ y(0) = \epsilon \end{cases} \tag{28}$$

la cui soluzione è $z_\epsilon = \epsilon \cdot \exp(\lambda x)$.
Di solito nelle applicazioni ci si interessa al caso $\lambda < 0$ o complesso con parte reale negativa. In questi casi $z_\epsilon \to 0$ per $n \to \infty$ e quindi l'effetto delle perturbazioni si annulla per valori grandi di x.
Si desidera che il metodo numerico goda delle stesse proprietà e la risposta la si ha nuovamente dallo studio della regione di stabilità.

Regione di assoluta stabilità

Nota.

Ricordiamo a tal proposito che dalla formula di Eulero, se $z = a + ib = \Re(z) + i\Im(z)$ *allora*

$$\exp(z) = \exp(a) \cdot (\cos(b) + i \sin(b)).$$

Quindi, visto che per qualsiasi b *si ha che* $|cos(b) + i \sin(b)| = 1$, *se* $\Re(\lambda) < 0$

$$\begin{aligned} |\exp(\lambda x)| &= |\exp(\Re(\lambda x))(\cos(\Im(\lambda x)) + i \sin(\Im(\lambda x)))| \\ &= |\exp(\Re(\lambda x))||\cos(\Im(\lambda x)) + i \sin(\Im(\lambda x))| \\ &= |\exp(\Re(\lambda x))| \to 0, \ \textit{per } x \to \infty. \end{aligned}$$

Problema stiff

Definizione (Problema stiff (Curtiss-Hirschfelder, 1952))

Un problema di Cauchy

$$\begin{cases} y'(x) = \lambda y(x), \ x \geq 0 \\ y(0) = 1. \end{cases} \tag{29}$$

si dice stiff, quando $\Re(\lambda) \ll 0$ *(cioè* $\Re(\lambda)$ *è molto grande in valore assoluto).*

Regione di assoluta stabilità: Eulero esplicito

Nel caso del metodo di Eulero esplicito,

$$u_n = u_{n-1} + hf(x_{n-1}, u_{n-1}) = (1 + h\lambda)u_{n-1}, \;\; u_0 = 1.$$

Visto che

$$u_n = (1+h\lambda)u_{n-1} = (1+h\lambda)^2 u_{n-2} = \ldots = (1+h\lambda)^n u_0 = (1+h\lambda)^n,$$

necessariamente

$$\boxed{u_n = (1 + h\lambda)^n}$$

e quindi

$$u_n \to 0 \text{ se e solo se } |1 + h\lambda| < 1$$

il che significa che la regione di assoluta stabilitá consiste del disco aperto del piano complesso di centro -1 e raggio 1.

Regione di assoluta stabilità: Eulero esplicito

Di conseguenza,

- fissato λ, il metodo risulta stabile se e solo se si sceglie un passo sufficientemente piccolo;
- se $\lambda \in \mathbb{R}^-$, allora bisognerá scegliere un passo minore di $h^* = 2/|\lambda|$.
- nel caso di problemi stiff, cioè con $\Re(\lambda) \ll 0$, si deve scegliere un passo h molto piccolo affinchè la soluzione di Eulero esplicito tenda a 0.

Regione di assoluta stabilità: Eulero esplicito

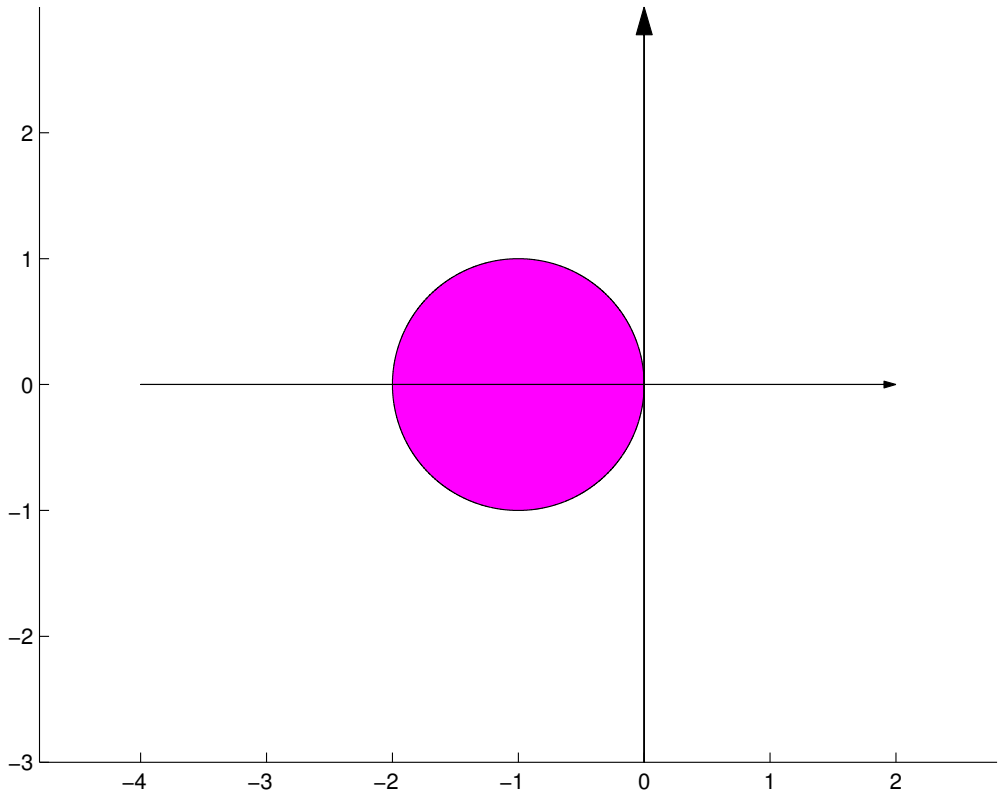

Figura: In magenta: regione di assoluta stabilità di Eulero esplicito: disco unitario centrato in $(-1, 0)$.

Regione di assoluta stabilità: Eulero implicito

Nel caso del metodo di Eulero implicito,

$$u_n = u_{n-1} + hf(x_n, u_n) = u_{n-1} + h\lambda u_n, \;\; u_0 = 1.$$

Portando a primo membro $h\lambda u_n$, dividendo i membri per $1 - h\lambda$

$$u_n = \frac{u_{n-1}}{1 - h\lambda},$$

da cui posto $C = \frac{1}{1-h\lambda}$,

$$u_n = Cu_{n-1} = C^2 u_{n-2} = \ldots = C^n u_0 = C^n = \left(\frac{1}{1 - h\lambda}\right)^n$$

ovvero

$$\boxed{u_n = C^n, \quad C = \frac{1}{1 - h\lambda}}$$

e quindi

$$u_n \to 0 \text{ se e solo se } \frac{1}{|1 - h\lambda|} < 1.$$

Regione di assoluta stabilità: Eulero implicito

Essendo $\Re(\lambda) < 0$, si vede facilmente che $|\frac{1}{1-h\lambda}| < 1$ per qualsiasi h. Infatti, $h\lambda = a + ib$ con $a < 0$, ciò è vero se e solo se

$$\frac{1}{|1-h\lambda|} < 1 \Leftrightarrow 1 < |1-h\lambda| = \sqrt{(1-a)^2+b^2}$$

ovvero, elevando ai quadrati entrambi i membri, se e solo se

$$1 < (1-a)^2 + b^2$$

ovviamente verificata in quanto $a < 0$, $a, b \in \mathbb{R}$. La regione di ass. stabilità è tutto il semipiano negativo $\Re(h\lambda) < 0$.

Nota.

Questa proprietà suggerisce di applicare Eulero implicito invece di Eulero esplicito per risolvere numericamente un problema di Cauchy di tipo stiff.

Regione di assoluta stabilità: Eulero implicito

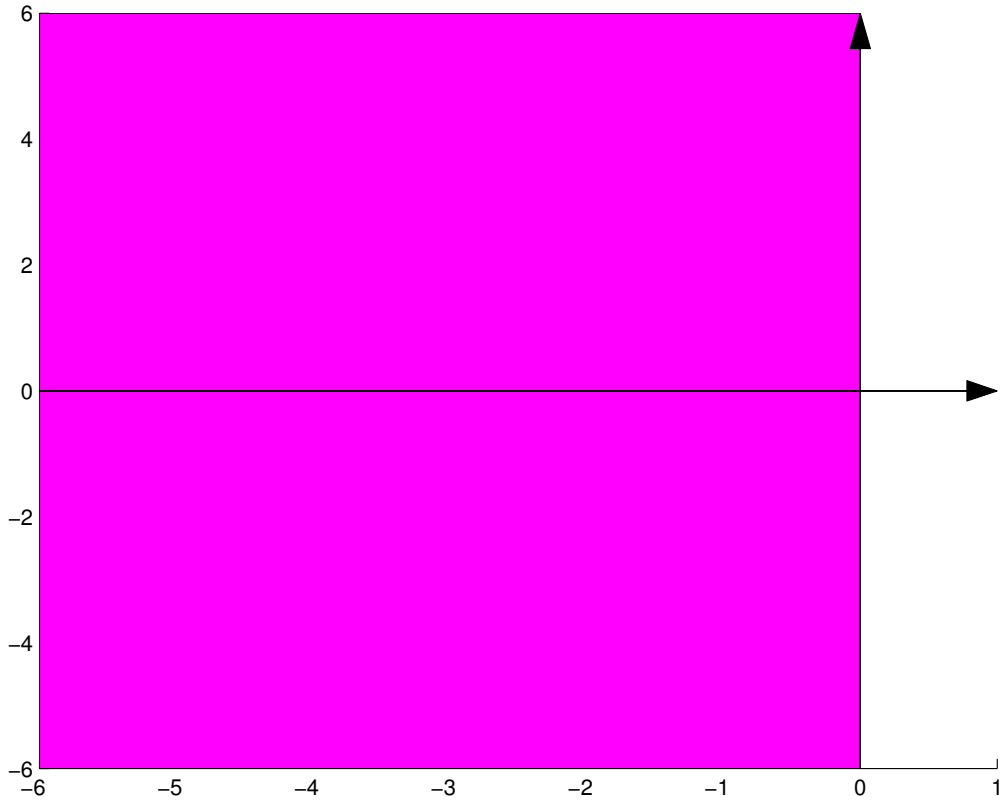

Figura: In magenta: regione di assoluta stabilità di Eulero implicito, semipiano negativo $\Re(h\lambda) < 0$.

Regione di assoluta stabilità: Crank-Nicolson

Nel caso del metodo di Crank-Nicolson, da $f(x,y) = \lambda y$

$$\begin{aligned} u_n &= u_{n-1} + (h/2)(f(x_n,u_n) + f(x_{n-1},u_{n-1})) \\ &= u_{n-1} + (h/2)\lambda(u_n + u_{n-1}),\ u_0 = 1 \end{aligned} \tag{30}$$

Quindi,

$$(1 - \frac{h\lambda}{2})u_n = u_{n-1} + \frac{h\lambda}{2}u_{n-1}$$

cioè

$$\frac{2 - h\lambda}{2} \cdot u_n = \frac{2 + h\lambda}{2} \cdot u_{n-1}$$

e di conseguenza

$$u_n = \frac{2 + h\lambda}{2 - h\lambda}u_{n-1}$$

Regione di assoluta stabilità: Crank-Nicolson

Posto $C = \frac{2+h\lambda}{2-h\lambda}$, essendo $u_n = Cu_{n-1}$, otteniamo come in precedenza $u_n = Cu_{n-1} = C^2 u_{n-2} = \ldots = C^n u_0 = C^n$ e quindi

$$\boxed{u_n = \left(\frac{2+h\lambda}{2-h\lambda}\right)^n}$$

da cui

$$u_n \to 0 \text{ se e solo se } \frac{|2+h\lambda|}{|2-h\lambda|} < 1.$$

Da $\Re(\lambda) < 0$, $|\frac{2+h\lambda}{2-h\lambda}| < 1$ per $h > 0$. Infatti, se $h\lambda = a + ib$, $a < 0$

$$\frac{|2+h\lambda|}{|2-h\lambda|} = \frac{|2+(a+ib)|}{|2-(a+ib)|} = \frac{|(2+a)+ib|}{|(2-a)+ib|} = \frac{\sqrt{(2+a)^2+b^2}}{\sqrt{(2-a)^2+b^2}} < 1$$

in quanto $(2+a)^2 < (2-a)^2$, qualora $a < 0$.

Ne segue che la regione di assoluta stabilità è tutto il semipiano negativo $\Re(h\lambda) < 0$.

Regione di assoluta stabilità di Crank-Nicolson

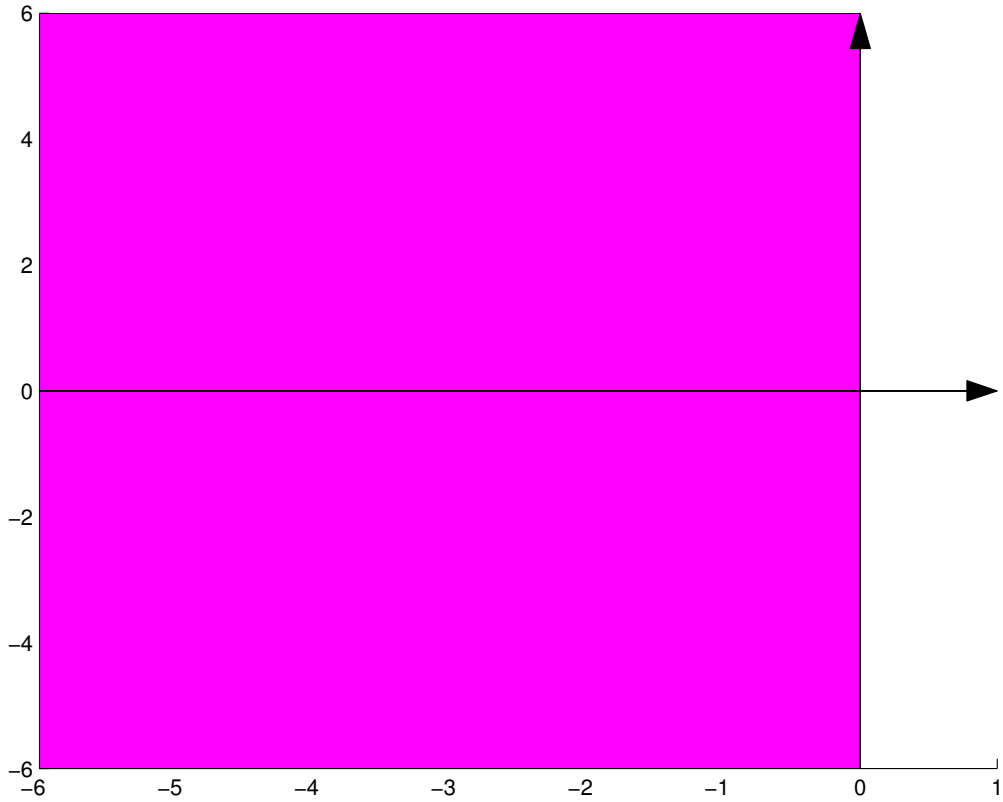

Figura: In magenta: regione di assoluta stabilità di Crank-Nicolson, semipiano negativo $\Re(h\lambda) < 0$.

A-stabilitá e Barriere LMM

Definizione

Un metodo numerico è detto A-stabile se la sua regione di stabilità assoluta contiene tutto il semipiano negativo.

Valgono le seguenti barriere di stabilità (Dahlquist, 1963)

Teorema

Nessun metodo LMM esplicito è A-stabile.

Teorema

Nessun metodo LMM implicito di ordine maggiore di 2 è A-stabile.

Runge-Kutta

Sia $h = (x_{\text{fin}} - x_0)/N_h$, $x_n = x_0 + nh$, $n = 0, \ldots, N_h$, e supponiamo che y sufficientemente regolare sia la soluzione del problema di Cauchy (1) nell'intervallo $[x_0, x_{\text{fin}}]$, ed f continua e L-lipschitziana rispetto la seconda variabile.
Consideriamo un metodo numerico, che diremo di Runge-Kutta, che definisca una sequenza u_n t.c.

$$y(x_{n+1}) \approx u_{n+1} = u_n + hF(x_n, u_n; h)$$

con

$$F(x, y; h) = \gamma_1 f(x, y) + \gamma_2 f(x + \alpha h, y + \beta h f(x, y)).$$

Determiniamo i parametri γ_1, γ_2, α, β così da ottenere un metodo del second'ordine cioè

$$\tau(h) = \max_n(|\tau_n(h)|) = O(h^2)$$

dove

$$\tau_n(h) = \frac{y(x_n) - \overline{u}_n}{h}, \ \overline{u}_{n+1} = y_n + hF(x_n, y_n; h).$$

Runge-Kutta

Per ottenere questo risultato usiamo la formula di Taylor bivariata, ovvero se $g : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ è di classe C^2, Ω aperto contenente $(\overline{x}_0, \overline{y}_0)$, $(\overline{x}_0 + \overline{h}, \overline{y}_0 + \overline{k})$,

$$g(\overline{x}_0 + \overline{h}, \overline{y}_0 + \overline{k}) = g(\overline{x}_0, \overline{y}_0) + g'_x(\overline{x}_0, \overline{y}_0)\overline{h} + g'_y(\overline{x}_0, \overline{y}_0)\overline{k} + O(\overline{h}^2 + \overline{k}^2).$$

Denotate con f'_x, f'_y le derivate parziali rispetto al primo e secondo argomento di f, abbiamo per $\overline{x}_0 = x_n$, $\overline{y}_0 = y_n$, $\overline{h} = \alpha h$, $\overline{k} = \beta h f(x_n, y_n)$,

$$f(x_n + \alpha h, y_n + \beta h f(x_n, y_n)) = f(x_n, y_n) + \alpha h f'_x(x_n, y_n) + \beta h f'_y(x_n, y_n) f(x_n, y_n) + O(h^2)$$

e quindi

$$\begin{aligned} F(x_n, y_n; h) &:= \gamma_1 f(x_n, y_n) + \gamma_2 f(x_n + \alpha h, y_n + \beta h f(x_n, y_n)) \\ &= \gamma_1 f(x_n, y_n) + \gamma_2 (f(x_n, y_n) + \alpha h f'_x(x_n, y_n) + \beta h f'_y(x_n, y_n) f(x_n, y_n)) \\ &+ O(h^2). \end{aligned} \tag{31}$$

da cui

$$\begin{aligned} \overline{u}_{n+1} &= y_n + hF(x_n, y_n; h) = y_n + h(\gamma_1 f(x_n, y_n) + \gamma_2 (f(x_n, y_n) \\ &+ \alpha h f'_x(x_n, y_n) + \beta h f'_y(x_n, y_n) f(x_n, y_n))) + O(h^3) \\ &= y_n + h(\gamma_1 + \gamma_2) f(x_n, y_n) + h^2 \gamma_2 (\alpha f'_x(x_n, y_n) + \beta f'_y(x_n, y_n) f(x_n, y_n))) \\ &+ O(h^3) \end{aligned}$$

Runge-Kutta

Facilmente, dalla formula di Taylor, se y è sufficientemente regolare,

$$\begin{aligned} y_{n+1} &= y_n + hy^{(1)}(x_n) + (h^2/2)y^{(2)}(x_n) + O(h^3) \\ &= y_n + hf(x_n, y_n) + (h^2/2)(f'_x(x_n, y_n) + f'_y(x_n, y_n)f(x_n, y_n)) + O(h^3) \end{aligned}$$

Quindi affinchè $y_{n+1} - \overline{u}_{n+1} = O(h^3)$, per confronto con

$$\begin{aligned} \overline{u}_{n+1} &= y_n + h(\gamma_1 + \gamma_2)f(x_n, y_n) + h^2\gamma_2(\alpha f'_x(x_n, y_n) + \beta f'_y(x_n, y_n)f(x_n, y_n))) \\ &+ O(h^3) \end{aligned}$$

basta richiedere

$$\begin{aligned} &\gamma_1 + \gamma_2 = 1 \\ &\gamma_2\alpha = 1/2, \\ &\gamma_2\beta = 1/2. \end{aligned}$$

Runge-Kutta

Vediamo alcuni metodi in cui

$$\boxed{\gamma_1 + \gamma_2 = 1, \;\; \gamma_2\alpha = 1/2, \;\; \gamma_2\beta = 1/2.}$$

Metodo di Heun ($\alpha = 1$), (scoperto nel 1900):

$$\boxed{u_{n+1} = u_n + \frac{h}{2}(f(x_n, u_n) + f(x_n + h, u_n + hf(x_n, u_n))).}$$

Metodo di Eulero modificato ($\alpha = 1/2$):

$$\boxed{u_{n+1} = u_n + hf(x_n + \frac{h}{2}, u_n + \frac{h}{2}f(x_n, u_n)).}$$

Runge-Kutta

La scelta particolare di un maggior numero di vincoli permette, con qualche fatica, di calcolare un metodo di ordine 4 (scoperto da Kutta nel 1901). Posto

$$u_{n+1} = u_n + h \cdot \frac{f_1 + 2f_2 + 2f_3 + f_4}{6}$$

ove

$$\begin{aligned} f_1 &= f(x_n, u_n) \\ f_2 &= f(x_n + \frac{h}{2}, u_n + \frac{h \cdot f_1}{2}) \\ f_3 &= f(x_n + \frac{h}{2}, u_n + \frac{h}{2} f_2) \\ f_4 &= f(x_n + h, u_n + h \cdot f_3) \end{aligned} \tag{32}$$

si ottiene un metodo del quart'ordine.

Parte facoltativa

Gli argomenti che seguono sono facoltativi. Servono a dare una immagine più completa sul tema dell'analisi numerica per la risoluzione di problemi di Cauchy.

Convergenza Eulero esplicito, caso Lipschitziano

Di seguito intendiamo mostrare, sotto opportune ipotesi e senza usare i teoremi precedenti, la **convergenza del metodo di Eulero esplicito**.

Consideriamo il metodo di Eulero esplicito, con $x_n = x_0 + nh$ per un prefissato passo $h = (x_{\text{fin}} - x_0)/N$. Sia $y \in C^2([x_0, x_{\text{fin}}])$ soluzione di (1), e

$$\overline{u}_n = y(x_{n-1}) + h \cdot f(x_{n-1}, y(x_{n-1})),$$

$$u_n = u_{n-1} + h \cdot f(x_{n-1}, u_{n-1}).$$

Osserviamo che

- la prima sequenza $\overline{u}^{(h)} = \{\overline{u}_n\}$, ha il generico $\overline{u}_n$ ottenuto applicando il metodo di Eulero esplicito partendo dal valore assunto dalla soluzione y in ogni x_{n-1},
- la seconda sequenza $u^{(h)} = \{u_n\}$, ha il generico u_n ottenuto applicando il metodo di Eulero esplicito partendo dal valore assunto u_{n-1}, che a priori non coincide con $y(x_{n-1})$.

Convergenza Eulero esplicito, caso Lipschitziano

Sketch della dimostrazione: Basta mostrare che

1. $|e_n| \leq (1 + (1 + hL) + \ldots + (1 + hL)^{n-1})h\tau(h) + (1 + hL)^n|e_0|$;
2. $(1 + hL)^n \leq \exp(L(x_{fin} - x_0))$;
3. $1 + (1 + hL) + \ldots + (1 + hL)^{n-1} \leq \frac{\exp(L(x_{fin} - x_0))}{hL}$.

Infatti ne consegue che

$$|e_n| \leq \frac{\exp(L(x_{fin} - x_0))}{L}\tau(h) + \exp(L(x_{fin} - x_0))|e_0|.$$

Da $y \in C^2([x_0, x_{\text{fin}}])$, posto $M := \max_{x\in[x_0,x_{fin}]} |y''(x)| \in (0, +\infty)$, per quanto visto sulla consistenza di Eulero Esplicito in (21)

$$\tau(h) = \max_n |\tau_n(h)| \leq Mh/2.$$

Di conseguenza $\max_n(|e_n|) \to 0$ per $h \to 0$, perchè il secondo membro è indipendente da n, infinitesimo per $h \to 0$ visto che $\tau(h) \leq Mh^2/2$ e $|e_0| \to 0$ per $h \to 0$.

Convergenza Eulero esplicito, caso Lipschitziano

Punto 1. Se f è L-Lipschitziana (rispetto al secondo argomento) allora

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2|.$$

Mostriamo per prima cosa che da $e_n := y(x_n) - u_n$,

$$\boxed{|\overline{u}_n - u_n| \leq (1 + hL)|e_{n-1}|.}$$

Infatti da

$$\overline{u}_n = y(x_{n-1}) + h \cdot f(x_{n-1}, y(x_{n-1})), \quad u_n = u_{n-1} + h \cdot f(x_{n-1}, u_{n-1})$$

ricaviamo dalla L-lipschitzianità

$$\begin{aligned}
|\overline{u}_n - u_n| &= |(y(x_{n-1}) + h \cdot f(x_{n-1}, y(x_{n-1}))) - (u_{n-1} + h \cdot f(x_{n-1}, u_{n-1}))| \\
&= |(y(x_{n-1}) - u_{n-1}) + h \cdot (f(x_{n-1}, y(x_{n-1})) - f(x_{n-1}, u_{n-1}))| \\
&\leq |y(x_{n-1}) - u_{n-1}| + h|f(x_{n-1}, y(x_{n-1})) - f(x_{n-1}, u_{n-1}))| \\
&\leq |y(x_{n-1}) - u_{n-1}| + hL|y(x_{n-1}) - u_{n-1}| \\
&= (1 + hL)|y(x_{n-1}) - u_{n-1}| = (1 + hL)|e_{n-1}|. \qquad (33)
\end{aligned}$$

Convergenza Eulero esplicito, caso Lipschitziano

Da

- $e_n := y(x_n) - u_n = (y(x_n) - \overline{u}_n) + (\overline{u}_n - u_n)$,
- $|y(x_n) - \overline{u}_n| = h|\tau_n(h)| \leq h\tau(h)$,
- $|\overline{u}_n - u_n| \leq (1 + hL)|e_{n-1}|$

si deduce

$$|e_n| = |(y(x_n) - \overline{u}_n) + (\overline{u}_n - u_n)| \leq |y(x_n) - \overline{u}_n| + |\overline{u}_n - u_n| \leq h\tau(h) + (1 + hL)|e_{n-1}|$$

e quindi

$$\begin{aligned} |e_n| &\leq h\tau(h) + (1 + hL)|e_{n-1}| \\ &\leq h\tau(h) + (1 + hL)(h\tau(h) + (1 + hL)|e_{n-2}|) \\ &= (1 + (1 + hL))h\tau(h) + (1 + hL)^2|e_{n-2}| \leq \dots \\ &\leq (1 + (1 + hL) + \dots + (1 + hL)^{n-1})h\tau(h) + (1 + hL)^n|e_0| \end{aligned}$$

e cioè

$$\boxed{|e_n| \leq (1 + (1 + hL) + \dots + (1 + hL)^{n-1})h\tau(h) + (1 + hL)^n|e_0|.}$$

Convergenza Eulero esplicito, caso Lipschitziano

Punto 2. Notiamo che per $\gamma > 0$

$$1 + \gamma = \sum_{k=0}^{1} \frac{\gamma^k}{k!} \leq \sum_{k=0}^{\infty} \frac{\gamma^k}{k!} = \exp(\gamma) \tag{34}$$

e quindi $(1+\gamma)^n \leq (\exp(\gamma))^n = \exp(n\gamma)$ implica che per $\gamma = hL$, essendo $nh = x_n - x_0$,

$$\boxed{(1+hL)^n \leq \exp(nhL) = \exp(L(x_n - x_0)) \leq \exp(L(x_{fin} - x_0)).}$$

Punto 3. Ricordando che

$$1 + s + \ldots + s^k = \frac{(1 - s^{k+1})}{(1-s)}$$

posto $s = 1 + hL$ deduciamo che

$$1 + (1+hL) + \ldots + (1+hL)^{n-1} = \frac{1 - (1+hL)^n}{1 - (1+hL)} = \frac{(1+hL)^n - 1}{hL}.$$

Convergenza Eulero esplicito, caso Lipschitziano

Da

- $1 + (1 + hL) + \ldots + (1 + hL)^{n-1} = \frac{1-(1+hL)^n}{1-(1+hL)} = \frac{(1+hL)^n-1}{hL}$
- $(1 + hL)^n \leq \exp(L(x_{fin} - x_0))$,

concludiamo che

$$\begin{aligned} 1 + (1 + hL) + \ldots + (1 + hL)^{n-1} &= \frac{(1 + hL)^n - 1}{hL} \leq \frac{\exp(L(x_{fin} - x_0)) - 1}{hL} \\ &\leq \frac{\exp(L(x_{fin} - x_0))}{hL}. \end{aligned} \tag{35}$$

Quindi, per quanto detto in precedenza, il metodo di Eulero esplicito è convergente. △

Convergenza Eulero implicito, caso Lipschitziano

Sia $h = (x_{\text{fin}} - x_0)/N_h$, $x_n = x_0 + nh$, $n = 0, \ldots, N_h$, e supponiamo che y sia la soluzione del problema di Cauchy (1) nell'intervallo $[x_0, x_{\text{fin}}]$, ed f continua e L-lipschitziana rispetto la seconda variabile, con $L > 0$.

Consideriamo il metodo di Eulero implicito

$$u_n = u_{n-1} + hf(t_n, u_n),\ u_0 = y_0$$

con u_n che si desidera approssimi $y_n := y(x_n)$. Di seguito richiederemo $y \in C^2([x_0, x_{\text{fin}}])$.
Dalla definizione di errore locale di troncamento, posto

$$\bar{u}_n = y_{n-1} + hf(t_n, y_n)$$

abbiamo

$$y_n - \bar{u}_n = h\tau_n(h)$$

con $\tau_n(h) = -hy''(\xi_n)/2$, $\xi_n \in [x_0, x_{\text{fin}}]$. Di conseguenza, visto che il teorema di Weierstrass esiste M finito tale che $M = \max_{x\in[x_0,x_{\text{fin}}]} |y''(x)|$,
necessariamente

$$\tau(h) := \max_{n=0,\ldots,N_h} |\tau_n(h)| \leq Mh/2.$$

Convergenza Eulero implicito, caso Lipschitziano

Per prima cosa notiamo che per la disuguaglianza triangolare, posto $e_n := y_n - u_n$

$$\begin{aligned} |e_n| &= |y_n - u_n| = |(y_n - \overline{u}_n) + (\overline{u}_n - u_n)| \leq |y_n - \overline{u}_n| + |\overline{u}_n - u_n| \\ &\leq h|\tau_n(h)| + |\overline{u}_n - u_n| \leq h\tau(h) + |\overline{u}_n - u_n|. \end{aligned} \tag{36}$$

Dalla lipschitzianità (rispetto la seconda variabile)

$$\begin{aligned} |\overline{u}_n - u_n| &= |y_{n-1} + hf(t_n, y_n) - u_{n-1} - hf(t_n, u_n)| \\ &= |y_{n-1} - u_{n-1} + hf(t_n, y_n) - hf(t_n, u_n)| \\ &\leq |y_{n-1} - u_{n-1}| + h|f(t_n, y_n) - f(t_n, u_n)| \\ &\leq |y_{n-1} - u_{n-1}| + hL|y_n - u_n| = |e_{n-1}| + hL|e_n|. \end{aligned} \tag{37}$$

Quindi, per $h < h^* = 1/L$ abbiamo $0 < 1 - hL < 1$ e quindi posto $\gamma := 1/(1 - hL) > 1$, da (36), portando $hL|e_n|$ a primo membro e dividendo ambo i membri per $1 - hL$, ricaviamo con facili conti

$$|e_n| \leq h\tau(h) + |e_{n-1}| + hL|e_n| \Leftrightarrow |e_n| \leq \gamma h\tau(h) + \gamma|e_{n-1}|.$$

Convergenza Eulero implicito, caso Lipschitziano

Di conseguenza, da $|e_n| \leq \gamma h\tau(h) + \gamma|e_{n-1}|$ si ha

$$\begin{aligned} |e_n| &\leq \gamma h\tau(h) + \gamma|e_{n-1}| \\ &\leq \gamma h\tau(h) + \gamma(\gamma h\tau(h) + \gamma|e_{n-2}|) = \gamma h\tau(h)(1+\gamma) + \gamma^2|e_{n-2}| \\ &\leq \ldots \leq \gamma h\tau(h)\sum_{k=0}^{n-1}\gamma^k + \gamma^n|e_0|. \end{aligned} \tag{38}$$

Osserviamo che da $(1+x)^k \leq \exp(kx)$, $\gamma - 1 = hL/(1-hL) > 0$, $nh = x_n - x_0$,

$$\begin{aligned} \gamma^n &= (1+(\gamma-1))^n \leq \exp(n(\gamma-1)) = \exp(nhL/(1-hL)) \\ &= \exp((x_n - x_0)L/(1-hL)) \leq \exp((x_{\text{fin}} - x_0)L/(1-hL)), \end{aligned} \tag{39}$$

per cui, visto che da $0 < \gamma - 1$

$$\begin{aligned} \gamma\sum_{k=0}^{n-1}\gamma^k &= \gamma\cdot\frac{\gamma^n - 1}{\gamma - 1} < \gamma\cdot\frac{\gamma^n}{\gamma-1} \leq \frac{1}{1-hL}\cdot\frac{1-hL}{hL}\cdot\exp\left(\frac{(x_{\text{fin}}-x_0)L}{1-hL}\right) \\ &= \frac{1}{hL}\cdot\exp\left(\frac{(x_{\text{fin}}-x_0)L}{1-hL}\right). \end{aligned} \tag{40}$$

Convergenza Eulero implicito, caso Lipschitziano

Quindi da

- $|e_n| \leq \gamma h \tau(h) \sum_{k=0}^{n-1} \gamma^k + \gamma^n |e_0|$,
- $\gamma \sum_{k=0}^{n-1} \gamma^k < \frac{1}{hL} \cdot \exp\left(\frac{(x_{\text{fin}} - x_0)L}{1-hL}\right)$,
- $\gamma^n \leq \exp((x_{\text{fin}} - x_0)L/(1-hL))$,

ricaviamo per $n = 0, \ldots, N_h$

$$\begin{aligned} |e_n| &\leq \gamma h \tau(h) \sum_{k=0}^{n-1} \gamma^k + \gamma^n |e_0| \\ &< h\tau(h) \frac{1}{hL} \cdot \exp\left(\frac{(x_{\text{fin}} - x_0)L}{1-hL}\right) + \exp\left(\frac{(x_{\text{fin}} - x_0)L}{1-hL}\right) |e_0| \\ &\leq \left(\frac{\tau(h)}{L} + |e_0|\right) \exp\left(\frac{(x_{\text{fin}} - x_0)L}{1-hL}\right). \end{aligned} \tag{41}$$

Visto che per $h \to 0$, abbiamo che $\tau(h) \to 0$ e $e_0 \to 0$, e che l'ultimo membro di (41) non dipende da n, deduciamo per il teorema del confronto che

$$\lim_{h \to 0} \max_{n=0,\ldots,N_h} |e_n| = 0,$$

ossia il metodo di Eulero implicito è convergente.

Facoltativo: Analisi convergenza Eulero esplicito, caso dissipativo

Se f non è L-Lipschitziana ma è L-dissipativa, cioè

$$-L \leq \frac{\partial f}{\partial y}(\xi) \leq 0, \quad \xi \in \Omega := (x_0, x_{\mathbf{fin}}) \times \mathbb{R}$$

abbiamo da

- $\overline{u}_n = y(x_{n-1}) + h \cdot f(x_{n-1}, y(x_{n-1}))$,
- $u_n = u_{n-1} + h \cdot f(x_{n-1}, u_{n-1})$,
- $f(x_{n-1}, u_{n-1}) = f(x_{n-1}, y(x_{n-1})) + \frac{\partial f}{\partial y}(\xi)(u_{n-1} - y(x_{n-1}))$,

che per qualche $\xi \in (x_{n-1}, x_n) \times \mathbb{R} \subseteq \Omega$

$$\begin{aligned} u_n - \overline{u}_n &= (u_{n-1} + h \cdot f(x_{n-1}, u_{n-1})) - (y(x_{n-1}) + h \cdot f(x_{n-1}, y(x_{n-1}))) \\ &= (u_{n-1} - y(x_{n-1})) + h \cdot (f(x_{n-1}, u_{n-1}) - f(x_{n-1}, y(x_{n-1}))) \\ &= (u_{n-1} - y(x_{n-1})) + h \frac{\partial f}{\partial y}(\xi)(u_{n-1} - y(x_{n-1})) \\ &= (1 + h \frac{\partial f}{\partial y}(\xi)) \cdot (u_{n-1} - y(x_{n-1})). \end{aligned}$$

Facoltativo: Analisi convergenza Eulero esplicito, caso dissipativo

Da

$$u_n - \overline{u}_n = (1 + h\frac{\partial f}{\partial y}(\xi)) \cdot (u_{n-1} - y(x_{n-1})). \tag{42}$$

abbiamo che se $0 < h \leq 2/L$ (non restrittivo per mostrare la conv. visto che si studia il comportamento per $h \to 0$) allora da $-L \leq \frac{\partial f}{\partial y} \leq 0$,

$$-1 = 1 - 2L/L \leq 1 - hL \leq 1 + h\frac{\partial f}{\partial y}(\xi) \leq 1$$

e quindi $|1 + h\frac{\partial f}{\partial y}(\xi)| \leq 1$ da cui

$$\begin{aligned} |u_n - \overline{u}_n| &= |1 + h\frac{\partial f}{\partial y}(\xi)| \cdot |u_{n-1} - y(x_{n-1})| \\ &\leq |u_{n-1} - y(x_{n-1})| = |e_{n-1}|. \end{aligned} \tag{43}$$

Facoltativo: Analisi convergenza Eulero esplicito, caso dissipativo

Da

- $e_n := y(x_n) - u_n$,
- $|y(x_n) - \overline{u}_n| \leq h\tau(h)$,
- $|\overline{u}_n - u_n| \leq |e_{n-1}|$,
- $|e_n| \leq h\tau(h) + |\overline{u}_n - u_n| \leq h\tau(h) + |e_{n-1}|$

ricaviamo, essendo $h\tau(h) > 0$, $e_0 = 0$, $n \leq N$, $Nh = x_{\text{fin}} - x_0$

$$\begin{aligned}
|\overline{u}_n - u_n| &\leq |e_{n-1}| \leq h\tau(h) + |e_{n-1}| \\
&\leq h\tau(h) + (h\tau(h) + |e_{n-2}|) \\
&= 2h\tau(h) + |e_{n-2}| \\
&\leq 2h\tau(h) + (h\tau(h) + |e_{n-3}|) \\
&= 3h\tau(h) + |e_{n-3}| \leq \dots \\
&\leq nh\tau(h) + |e_0| = nh\tau(h) \\
&\leq Nh\tau(h) = (x_{\text{fin}} - x_0)\tau(h) \leq \frac{(x_{\text{fin}} - x_0)Mh}{2}.
\end{aligned}$$

Predictor-corrector

In generale un metodo LMM implicito ha proprietà di stabilità migliori rispetto ad un LMM esplicito e questo ne suggerisce l'utilizzo.

Purtroppo ad ogni iterazione bisogna risolvere una equazione nonlineare.

Un approccio comunemente utilizzato è quello del predictor-corrector. Consiste nell'utilizzare il metodo di punto fisso con una certa scelta del punto iniziale.

Nel caso di un metodo di Adams implicito (detto corrector), si utilizza ad esempio un metodo di Adams esplicito (detto predictor).

Esempio Predictor-corrector

Consideriamo i seguenti metodi di Adams (porremo $f_i = f(x_i, u_i)$).

$$\text{Predictor: } u_{i+1} = u_i + \frac{h}{2}(3f_i - f_{i-1})$$

$$\text{Corrector: } u_{i+1} = u_i + \frac{h}{2}(f_{i+1} + f_i).$$

Posto $f_i^{(\eta)} = f(x_i, u_i^{(\eta)})$ abbiamo

$$\begin{aligned}
\text{P:} \quad & u_{i+1}^{(0)} = u_i^{(1)} + \frac{h}{2}(3f_i^{(1)} - f_{i-1}^{(1)}) \\
\text{E:} \quad & f_{i+1}^{(0)} = f(x_{i+1}, u_{i+1}^{(0)}) \\
\text{C:} \quad & u_{i+1}^{(1)} = u_i^{(1)} + \frac{h}{2}(f_{i+1}^{(0)} + f_i^{(1)}) \\
\text{E:} \quad & f_{i+1}^{(1)} = f(x_{i+1}, u_{i+1}^{(1)})
\end{aligned}$$

Bibliografia

K. Atkinson e W. Han, *Elementary Numerical Analysis*, Wiley, (2004).

K.E. Atkinson, W. Han, D.E. Steward *Numerical Solution of Ordinary Differential Equations*, Wiley, (2009).

V. Comincioli, *Analisi Numerica, metodi modelli applicazioni*, Mc Graw-Hill, 1990.

G. Dahlquist, *Convergence and stability in the numerical integration of ordinary differential equations*, Math. Scand. 4 (1956), pp.33-53.

W. Gautschi, *Numerical Analysis*, Birkhaüser, second edition, 2012.

A. Quarteroni, F. Saleri, *Introduzione al Calcolo Scientifico*, Springer, (2002).

A. Quarteroni, R. Sacco, F. Saleri, *Matematica Numerica*, Springer, (1998).

J. Stoer, R.Bulirsch *Introduzione all'analisi numerica*, Zanichelli, (1975).

E. Süli, D. Mayers *An Introduction to Numerical Analysis*, Cambridge University Press, (2003).