1 EQUAZIONI DIFFERENZIALI

1.1 INTRODUZIONE

PROBLEMA ai valori iniziali (Problema di Cauchy): *Data una funzione* $f : [t_0, t_f] \times \mathbb{R}^m \to \mathbb{R}^m$ *ed un vettore (valore iniziale)* $y_0 \in \mathbb{R}^m$, *trovare una funzione* $y(t) : [t_0, t_f] \to \mathbb{R}^m$ *tale che*

$$\begin{aligned} y'(t) &= f(t, y(t)), \quad t_0 \le t \le t_f, \\ y(t_0) &= y_0. \end{aligned} \tag{1}$$

□

DEFINIZIONE. *Diremo soluzione del problema ai valori iniziali ogni funzione* $y \in C^1([t_0, t_f], \mathbb{R}^m)$ *che lo verifica.*

□

DEFINIZIONE. *Introdotta in* $\mathbb{R}^m$ *una norma, diremo che la funzione* $f(t, y)$ *è Lipschitziana nella striscia* $[t_0, t_f] \times \mathbb{R}^m$ *rispetto alla seconda variabile se esiste* $L > 0$ (**costante di Lipschitz classica**) *tale che*

$$\|f(t, y) - f(t, z)\| \le L \|y - z\|, \text{ per ogni } t \in [t_0, t_f] \text{ e per ogni } y, z \in \mathbb{R}^m. \tag{2}$$

□

TEOREMA *Se* $f(t, y)$ *è continua in* t *ed* y *e vale la* (2) *allora il problema ai valori iniziali ammette una e una sola soluzione.*

□

TEOREMA. *Se* $f(t, y)$ *è continua in* t *ed* y *e vale la* (2), *allora considerata la soluzione* $z(t)$ *del problema con dato iniziale* z_0, *cioè*

$$\begin{aligned} z'(t) &= f(t, z(t)), \quad t_0 \le t \le t_f, \\ z(t_0) &= z_0, \end{aligned} \tag{3}$$

si ha che

$$\|y(t) - z(t)\| \le e^{(t-t_0)L} \|y_0 - z_0\|, \ t_0 \le t \le t_f.$$

□

DEFINIZIONE. *Intrododuciamo in* $\mathbb{R}^m$ *un prodotto scalare* $< \cdot, \cdot >$ *e la norma ad esso associata* $\|\cdot\|$. *Un numero reale* M *si dirà una* **costante di Lipschitz destra** *di* $f(t, y)$, *rispetto ad* y, *se*

$$< f(t, y) - f(t, z), y - z > \le M \|y - z\|^2, \tag{4}$$

per ogni $t \in [t_0, t_f]$ e per ogni $y, z \in \mathbb{R}^m$.

□

Si noti che M può essere un numero negativo. Si vede facilmente che se L è una costante di Lipschitz classica (rispetto alla stessa norma), allora esiste una costante di Lipschitz destra M tale che

$$|M| \leq L.$$

TEOREMA. *Se $f(t,y)$ è continua in t ed y e vale la* (4), *allora, considerati i problemi* (1) *e* (3), *si ha*

$$\|y(t) - z(t)\| \leq e^{(t-t_0)M} \|y_0 - z_0\|, \ t_0 \leq t \leq t_f.$$

□

OSSERVAZIONE. Consideriamo il caso di sistemi differenziali lineari, ossia quando f è del tipo

$$f(t, y(t)) = A(t)y(t) + g(t),$$

dove $A(t) \in \mathbb{R}^{m\times m}$ è una matrice che può dipendere da t e $g(t)$ è una funzione data. Considerato il prodotto scalare euclideo, la (4) diventa

$$< A(t)(y-z), y-z > \leq M \|y-z\|^2,$$

per ogni $t \in [t_0, t_f]$ e per ogni $y, z \in \mathbb{R}^m$. Essendo la matrice $A(t)$ reale, per ogni $y, z \in \mathbb{R}^m$ avremo $< A(t)(y-z), y-z > = < A^T(t)(y-z), y-z >$ e dunque M dovrà verificare, per ogni $t \in [t_0, t_f]$ e per ogni $y, z \in \mathbb{R}^m$:

$$< \frac{(A(t) + A^T(t))}{2}(y-z), y-z > \leq M \|y-z\|^2, \tag{5}$$

Poichè, per ogni $t \in [t_0, t_f]$, la matrice $A(t) + A^T(t)$ è simmetrica, indicato con $\lambda_{\max}(\frac{A(t)+A^T(t)}{2})$ il massimo degli autovalori di $\frac{A(t)+A^T(t)}{2}$, per il Teorema di Courant-Fisher la (5) vale quando

$$M \geq \lambda_{\max}(\frac{A(t) + A^T(t)}{2}), \text{ per ogni } t \in [t_0, t_f].$$

□

1.2 METODI AD UN PASSO

Introduciamo in $[t_0, t_f]$ una successione di punti $t_0 < t_1 < ... < t_n < t_{n+1} < ....$ Partendo dal valore iniziale noto $y(t_0) = y_0$, vogliamo approssimare i valori della soluzione nei successivi punti t_1, t_2, etc.

Consideriamo il generico intervallo $[t_n, t_{n+1}]$. In esso vale l' identità

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

e pertanto

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

Supponendo di conoscere un' approssimazione, che indichiamo con y_n, di $y(t_n)$ potremo allora ottenerne una, diciamo y_{n+1} di $y(t_{n+1})$ applicando una formula di quadratura. Se nel far questo si utilizzano solo informazioni relative all' intervallo $[t_n, t_{n+1}]$ un tale metodo si dirà **ad un passo**.

ESEMPI.

Consideriamo per semplicità il caso scalare ($m = 1$). Poniamo $h = t_{n+1} - t_n$.

1. **Metodo di Eulero esplicito**.

Si usa la formula del rettangolo in t_n:

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) = hf(t_n, y(t_n)) + \sigma_1(t_n, h) \tag{6}$$

dove

$$\sigma_1(t_n, h) = \frac{1}{2}\frac{d}{dt}f(\xi_n, y(\xi_n))h^2 = \frac{1}{2}y"(\xi_n)h^2, \quad t_n < \xi_n < t_{n+1} \ .$$

Trascurando l' errore $\sigma_1(t_n, h)$, si definisce il metodo

$$y_{n+1} = y_n + hf(t_n, y_n) \text{ (metodo di Eulero esplicito).}$$

□

2. **Metodo di Eulero implicito**.

Si usa la formula del rettangolo in t_{n+1}:

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) = hf(t_{n+1}, y(t_{n+1})) + \sigma_2(t_n, h) \tag{7}$$

dove

$$\sigma_2(t_n, h) = -\frac{1}{2}\frac{d}{dt}f(\xi_n, y(\xi_n))h^2 = -\frac{1}{2}y"(\xi_n)h^2, \quad t_n < \xi_n < t_{n+1} \ .$$

Trascurando l' errore $\sigma_2(t_n, h)$, si definisce il metodo

$$y_{n+1} = y_n + hf(t_{n+1}, y_{n+1}) \text{ (metodo di Eulero implicito).}$$

Se vale la (2), si può vedere che, per ogni h sufficientemente picccolo, y_{n+1} esiste ed è unico.

□

2. **Metodo dei trapezi**.

Si usa la formula del trapezio in t_n e t_{n+1}:

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) = \frac{h}{2}(f(t_n, y(t_n)) + f(t_{n+1}, y(t_{n+1})) + \sigma_3(t_n, h) \tag{8}$$

dove

$$\sigma_3(t_n,h) = -\frac{1}{12}\frac{d^2}{dt^2}f(\xi_n,y(\xi_n))h^3 = -\frac{1}{12}y'''(\xi_n)h^3, \quad t_n < \xi_n < t_{n+1}\ .$$

Trascurando l' errore $\sigma_3(t_n,h)$, si definisce il metodo

$$y_{n+1} = y_n + \frac{h}{2}(f(t_n,y_n) + f(t_{n+1},y_{n+1})) \text{ (metodo dei trapezi).}$$

Anche in questo caso, se vale la (2), si può vedere che, per ogni h sufficientemente picccolo, y_{n+1} esiste ed è unico.

Gli errori $\sigma_i(t_n,h)$, $i = 1,2,3$, sono gli **errori di quadratura**.

I tre metodi di cui sopra si estendono anche al caso vettoriale ($m \geq 1$). In questo caso, introdotta una norma vettoriale, per quanto riguarda gli errori di quadratura in (6), (7), (8), se f è sufficientemente regolare in $[t_0,t_f]$ si ha che

$$\|\sigma_1(t_n,h)\| = O(h^2), \|\sigma_2(t_n,h)\| = O(h^2), \|\sigma_3(t_n,h)\| = O(h^3). \tag{9}$$

La scrittura $\|\sigma_i(t_n,h)\| = O(h^k)$ significa che esiste $C > 0$ tale che per ogni $t_n \in [t_0,t_f]$ si ha: $\|\sigma_i(t_n,h)\| \leq Ch^k$.

1.3 CONVERGENZA DEI METODI AD UN PASSO

In generale un metodo ad un passo si può esprimere nella forma

$$y_{n+1} = y_n + h\Phi(t_n,y_n;h,f). \tag{10}$$

dove Φ è un' opportuna funzione detta **Funzione Incrementale** del metodo. Per brevità, data f, poniamo $\Phi(t_n,y_n;h) = \Phi(t_n,y_n;h,f)$. Introdotta un norma in $\mathbb{R}^m$, assumiamo fin d' ora che la funzione $\Phi(t,y;h)$ sia continua in t e Lipschitziana (in senso classico) rispetto ad y, cioè che esista $Q > 0$ tale che, per ogni h sufficientemente piccolo,

$$\|\Phi(t,y;h) - \Phi(t,z;h)\| \leq Q\|y-z\|\text{, per ogni } t \in [t_0,t_f] \text{ e per ogni } y,z \in \mathbb{R}^m. \tag{11}$$

Si può vedere che i tre metodi di cui al paragrafo precedente si possono rappresentare nella forma (10) e che se vale la (2) allora vale anche la (11).

Consideriamo un metodo ad un passo

$$y_{n+1} = y_n + h\Phi(t_n,y_n;h) \tag{12}$$

e supponiamo valga la (11).

DEFINIZIONE. *Fissato* $h > 0$, *dato* $t \in [t_0,t_f]$ (*tale che* $t+h \in [t_0,t_f]$) *definiamo* **errore locale di troncamento:**

$$\sigma(t,h) = y(t+h) - y(t) - h\Phi(t,y(t);h)$$

e definiamo **errore locale di discretizzazione:**

$$d(t,h) = \frac{\sigma(t,h)}{h}.$$

Poniamo inoltre

$$d(h) = \max_{t\in[t_0,t_f]} \|d(t,h)\| . \tag{13}$$

□

DEFINIZIONE. *Un metodo* (12) *si dice* **consistente** *in* $[t_0,t_f]$ *(con il problema) se*

$$\lim_{h\to 0.} d(h) = 0.$$

Se

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

diremo che il metodo ha **ordine di consistenza** p.

□

DEFINIZIONE. *Un metodo* (12) *si dice convergente in* $[t_0,t_f]$ *se per ogni fissato* $t = t_0 + Nh \in [t_0,t_f]$ (N intero ≥ 0) *si ha*

$$\lim_{h\to 0.} \|y_N - y(t)\| = 0, \ (N \to \infty \text{ , essendo } Nh = t - t_0),$$

uniformemente su $[t_0,t_f]$. *Se esiste* $C > 0$ *tale che, per ogni* h *sufficientemente piccolo e per ogni* $t = t_0 + Nh \in [t_0,t_f]$,

$$\|y_N - y(t)\| \leq Ch^p$$

allora diremo che il metodo ha **ordine di convergenza** p.

□

TEOREMA. S*upponiamo che per ogni* $h \leq h_0$ *valga la* (11) *e che il metodo definito dalla* (12) *sia consistente di ordine* p *in* $[t_0,t_f]$, *allora esso è convergente in* $[t_0,t_f]$ *con ordine di convergenza* p.

Dim. Sia $h \leq h_0$ e $t = t_0 + Nh \in [t_0,t_f]$. Poniamo $t_n = t_0 + nh$, $e_n = y(t_n) - y_n$. Sottraendo membro a membro la (12) dall' identità

$$y(t_{n+1}) = y(t_n) + h\Phi(t_n, y(t_n); h) + \sigma(t_n, h),$$

si ottiene, per $0 \leq n \leq N-1$,

$$e_{n+1} = e_n + h(\Phi(t_n, y(t_n); h) - \Phi(t_n, y_n; h)) + \sigma(t,h).$$

Da questa, per la (11) e la (13), otteniamo

$$\|e_{n+1}\| \leq \|e_n\| + hQ\|e_n\| + hd(h). \tag{14}$$

Utilizzeremo ora il seguente:

LEMMA. *Sia* $\{s_n\}$, $n = 0, 1,...$, *una successione di numeri non negativi che soddisfano la relazione*

$$s_{n+1} \leq (1 + hQ)s_n + D,$$

dove $D \geq 0$. *Allora per ogni* $k \leq N$ *si ha*

$$s_k \leq (1 + hQ)^N s_0 + D\frac{(1 + hQ)^N - 1}{hQ}.$$

□

Applicando questo risultato alla (14) (posto $s_n = \|e_n\|$), essendo$\|e_0\| = 0$, si ottiene

$$\begin{aligned}\|e_N\| &\leq (1 + hQ)^N \|e_0\| + hd(h)\frac{(1 + hQ)^N - 1}{hQ} \\ &= (\frac{(1 + hQ)^N - 1}{Q})d(h) \\ &\leq (\frac{e^{NhQ} - 1}{Q})d(h) = (\frac{e^{Q(t-t_0)} - 1}{Q})d(h)\end{aligned}$$

da cui si ha la tesi.

□

Consideriamo i metodi di Eulero esplicito, implicito e dei trapezi. Si verifica che per questi tre metodi $\|\sigma(t, h)\|$ tende a zero con h (uniformemnte rispetto a $t \in [t_0, t_f]$), rispettivamente come $\|\sigma_1(t, h)\|$, $\|\sigma_2(t, h)\|$ e $\|\sigma_3(t, h)\|$. Pertanto, se la f è sufficientemente regolare, vista la (9), avremo:

1. per Eulero esplicito $d(h) = O(h)$, $p = 1$;
2. per Eulero implicito $d(h) = O(h)$, $p = 1$;
3. per i trapezi $d(h) = O(h^2)$, $p = 2$.

1.4 METODI RUNGE-KUTTA

I metodi Runge-Kutta sono metodi ad un passo costruiti nel seguente modo.

Consideriamo il generico intervallo $[t_n, t_{n+1}]$ e la relazione

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

Vogliamo utilizzare formule di quadratura che possono coinvolgere altri punti in $[t_n, t_{n+1}]$. Siano dunque $c_i \leq 1$, per $i = 1, 2, ..., s$, coefficienti opportuni. Consideriamo i punti $t_n + c_i h$, dove $h = t_{n+1} - t_n$ una formula di quadratura del tipo

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) = h\sum_{i=1}^{s} b_i f(t_n + c_i h, y(t_n + c_i h)) + \overline{\sigma}(t_n, h).$$

Avremo pertanto

$$y(t_{n+1}) = y(t_n) + h\sum_{i=1}^{s} b_i f(t_n + c_i h, y(t_n + c_i h)) + \overline{\sigma}(t_n, h). \qquad (15)$$

Poichè i valori $y(t_n + c_i h)$ non sono noti, cerchiamo di approssimarli applicando delle formule di quadratura (sui nodi $t_n + c_i h$) alle identità

$$y(t_n + c_i h) = y(t_n) + \int_{t_n}^{t_n + c_i h} f(t, y(t))dt, \text{ per } i = 1, 2, ...s.$$

Vale a dire

$$y(t_n + c_i h) = y(t_n) + h\sum_{j=1}^{s} a_{ij} f(t_n + c_j h, y(t_n + c_j h)) + \overline{\sigma}_i(t_n, h). \qquad (16)$$

Nelle (15) e (16) $\overline{\sigma}$ e $\overline{\sigma}_i$ indicano gli **errori di quadratura**. Affinchè le formule di quadratura utilizzate siano esatte almeno per le costanti si richiede che siano verificate le condizioni:

$$\sum_{i=1}^{s} b_i = 1,$$

$$c_i = \sum_{j=1}^{s} a_{ij}, \text{ per ogni } i = 1, 2, ..., s.$$

Trascurando gli errori di quadratura, dalle (15) e (16) otteniamo il metodo **Runge-Kutta ad s livelli**:

$$y_{n+1} = y_n + h\sum_{i=1}^{s} b_i f(t_n + c_i h, Y_i)$$

dove gli Y_i si ottengono risolvendo il sistema

$$Y_i = y_n + h\sum_{j=1}^{s} a_{ij} f(t_n + c_j h, Y_j), \text{per } i = 1, 2, ..., s.$$

Una formulazione equivalente è la seguente:

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

dove i K_i, per $i = 1, 2, ..., s$, risolvono il sistema

$$K_i = f(t_n + c_i h, y_n + h\sum_{j=1}^{s} a_{ij} K_j), \text{per } i = 1, 2, ..., s.$$

Si può vedere che questo è un metodo ad un passo al quale si applica la teoria del paragrafo precedente.

Un metodo Runge-Kutta è univocamente identificato dai coefficienti c_i, b_i, a_{ij} che possono essere raccolti in una tabella detta **Tabella di Butcher**. Precisamente dati $c = [c_1, ..., c_s]^T$, $b^T = [b_1, ..., b_s]$, $A = [a_{ij}]$, per $i, j = 1, ..., s$, la tabella è così formata:

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

In particolare i coefficienti a_{ij} consentono di distinguere il tipo di metodo nel modo seguente:

1. se $a_{ij} = 0$ per ogni $j \geq i$ il metodo è ESPLICITO (la matrice A è triangolare inferiore in senso stretto);

2. se $a_{ij} = 0$ per ogni $j \geq i + 1$, il metodo si dirà SEMI IMPLICITO (la matrice A è triangolare inferiore);

3. in ogni altro caso il metodo è IMPLICITO.

Si osservi che i metodi di Eulero esplicito ed implicito e quello dei trapezi sono metodi Runge-Kutta.

Diamo qui di seguito una tabella degli ordini massimi ottenibili $p(s)$ per un metodo Runge-Kutta ad s livelli.

Metodi espliciti: $p(1) = 1, p(2) = 2, p(3) = 3, p(4) = 4, p(5) = 4, p(6) = 5, p(7) = 5, p(8) = 6$.

Metodi impliciti: $p(s) = 2s$.