

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Triennale in Matematica

**INTEGRATORI ESPONENZIALI DEL
PRIMO ORDINE PER PROBLEMI
DIFFERENZIALI SEMI - LINEARI**

Tesi di Laurea in Analisi Numerica

Relatore:
Chiar.ma Prof.ssa
VALERIA SIMONCINI

Presentata da:
SIMONE PANDINI

Sessione Unica
Anno Accademico 2021/2022

Introduzione

Gli integratori esponenziali costituiscono un'ampia classe di metodi numerici per la risoluzione di equazioni differenziali ordinarie, in particolare vengono applicati a Problemi di Cauchy ai dati iniziali. I primi esempi di integratori esponenziali risalgono a ottanta anni fa e vennero introdotti per la prima volta da Certaine (*The solution of ordinary differential equations with large time constants*, 1960) e Pope (*An exponential method of numerical integration of ordinary differential equations*, 1963), ma non hanno avuto un ruolo importante nelle applicazioni per molto tempo, in quanto fanno uso esplicito dell'esponenziale di matrice o di funzioni di matrici di grandi dimensioni: un'efficiente implementazione risultava pertanto difficile. Originariamente, tali metodi vennero sviluppati per problemi stiff, ossia problemi la cui risoluzione con alcuni metodi numerici è numericamente instabile, a meno che il passo di discretizzazione sia estremamente piccolo. Successivamente, gli integratori esponenziali vennero applicati a equazioni differenziali alle derivate parziali: in particolare, i primi studi consistevano nell'analisi della stabilità e della convergenza. L'applicazione di questi nuovi metodi a problemi di notevole rilevanza diede inizio a un rigoroso approfondimento teorico e alla costruzione di nuove classi di integratori esponenziali, ad esempio con ordine di convergenza maggiore. La parte metodologica di questa tesi si basa in buona parte sull'articolo di rassegna di Hochbruck e Ostermann, in cui vengono analizzati tali metodi numerici e descritte strategie risolutive avanzate per il trattamento numerico di problemi iperbolici e parabolici (*Exponential integrators*, 2010): studieremo i primi modelli di integratori esponenziali che sfruttano la linearizzazione del problema e fanno uso della matrice Jacobiana. In generale, tale matrice non ha una struttura particolare e varia continuamente nel tempo, pertanto le tecniche standard per la valutazione dell'esponenziale non sono molto efficienti dal punto

di vista computazionale: studi successivi più avanzati sfruttano, ad esempio, i sottospazi di Krylov e l'interpolazione polinomiale sui nodi di Leja, che risultano più interessanti in quanto utilizzano solamente prodotti matrice - vettore e non richiedono un'eccessiva quantità di dati da memorizzare.

Nel primo capitolo studieremo l'esponenziale di matrice e vedremo alcune sue proprietà e un metodo numerico per approssimarlo.

Nel secondo capitolo introdurremo alcuni metodi numerici a un passo per la risoluzione di Problemi di Cauchy.

Nel terzo capitolo studieremo il Metodo di Rosenbrock - Eulero che fa uso degli integratori esponenziali per risolvere problemi differenziali con dato iniziale e lo confronteremo con i metodi visti nel capitolo precedente, in particolare in riferimento alla convergenza e alla stabilità.

Nel quarto capitolo applicheremo questi metodi per la risoluzione dell'equazione alle derivate parziali di Allen - Cahn.

Indice

| | |
|---|-----------|
| Introduzione | i |
| Prerequisiti | ix |
| 1 Esponenziale di matrice | 1 |
| 1.1 Funzioni di matrice | 1 |
| 1.2 Proprietà di base | 2 |
| 1.3 Condizionamento | 6 |
| 1.4 Approssimante di Padé | 8 |
| 1.5 Metodo di <i>Scaling and squaring</i> | 9 |
| 2 Metodi numerici per la risoluzione di equazioni differenziali | 13 |
| 2.1 Metodi numerici a un passo | 15 |
| 2.1.1 Metodo di Eulero esplicito | 16 |
| 2.1.2 Metodo di Eulero esponenziale | 17 |
| 2.2 Consistenza, convergenza e A-stabilità di un metodo | 18 |
| 3 Integratori esponenziali | 21 |
| 3.1 Metodo di Rosenbrock - Eulero | 22 |
| 3.2 Convergenza dei metodi | 23 |
| 3.3 Aspetti implementativi | 24 |
| 4 Applicazione alle equazioni differenziali alle derivate parziali | 27 |
| 4.1 Equazione di Allen Cahn | 27 |

Conclusioni

39

Bibliografia

41

Elenco delle figure

| | | |
|-----|--|----|
| 4.1 | Metodo delle linee | 28 |
| 4.2 | $N = 100, M = 200, \epsilon = 0.01$ | 32 |
| 4.3 | $N = 500, M = 600, \epsilon = 0.01$ | 33 |
| 4.4 | Eulero esplicito con $\epsilon = 0.01$ | 33 |
| 4.5 | $N = 100, M = 200, \epsilon = 0.05$ | 36 |
| 4.6 | $N = 500, M = 600, \epsilon = 0.05$ | 36 |
| 4.7 | Non convergenza del metodo di Eulero esplicito con $\epsilon = 0.05$ | 37 |

Elenco delle tabelle

| | | |
|-----|--|----|
| 4.1 | Differenza in $\ \cdot\ _2$ delle soluzioni usando i vari metodi | 34 |
| 4.2 | Tempi computazionali in secondi dei vari metodi | 34 |
| 4.3 | Confronto in termini di tempo computazionale del metodo di Rosenbrock - Eulero usando la Jacobiana come <i>handle</i> e la funzione <i>Jacobian</i> | 35 |

Prerequisiti

Definizione (Norma di matrice). Una norma di matrice su $\mathbb{C}^{m \times n}$ è una funzione $\|\cdot\| : \mathbb{C}^{m \times n} \rightarrow \mathbb{R}$ tale che:

1. $\|A\| \geq 0$ per ogni $A \in \mathbb{C}^{m \times n}$ e $\|A\| = 0$ se e solo se $A = 0$
2. $\|\alpha A\| = |\alpha| \|A\|$, per ogni $\alpha \in \mathbb{C}$, $A \in \mathbb{C}^{m \times n}$
3. $\|A + B\| \leq \|A\| + \|B\|$, per ogni $A, B \in \mathbb{C}^{m \times n}$
4. $\|AB\| \leq \|A\| \|B\|$, per ogni $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{n \times p}$

Esempio (Norme). Sia $A = (a_{ij})_{i=1, \dots, m; j=1, \dots, n}$:

- Norma indotta

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}, \quad x \in \mathbb{C}^n$$

- Norma di Frobenius

$$\|A\|_F = \text{tr}(A^*A)^{1/2} = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2},$$

dove $\text{tr}(M)$ è la traccia

- Norma p , con $1 \leq p \leq \infty$

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}, \quad x \in \mathbb{C}^n,$$

dove $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ se $1 \leq p < \infty$ e $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$

Definizione (Numero di condizionamento). Il numero di condizionamento di $A \in \mathbb{C}^{n \times n}$ invertibile è

$$\kappa(A) = \|A\| \|A^{-1}\|$$

Esempio. Per matrici simmetriche reali con autovalori λ_i vale

$$\kappa(A) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$$

Notazione. Denotiamo con $\Lambda(A) = \{\lambda_1, \dots, \lambda_n\}$ l'insieme degli autovalori di $A \in \mathbb{C}^{n \times n}$ e con $\alpha(A) = \max_{i=1, \dots, n} \Re(\lambda_i)$ l'ascissa spettrale di A .

Definizione (Matrice normale). $A \in \mathbb{C}^{n \times n}$ si dice matrice normale se $A^*A = AA^*$, dove A^* indica la trasposta coniugata di A .

Definizione (Forma canonica di Jordan). Sia $A \in \mathbb{C}^{n \times n}$ con autovalori distinti $\lambda_1, \dots, \lambda_p$; A si può esprimere in forma canonica di Jordan

$$Z^{-1}AZ = J = \text{diag}(J_1, \dots, J_p), \quad (1)$$

$$J_k = J_k(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k \times m_k},$$

con Z non singolare e $m_1 + \dots + m_p = n$. La matrice di Jordan J è unica a meno di riordinare i blocchi J_i , ma la matrice Z non è unica. Denotiamo con n_i l'ordine del blocco di Jordan più grande in cui compare λ_i , che è detto l'indice di λ_i .

Notazione. Indichiamo con $\mathcal{R}_{k,m}$ lo spazio delle funzioni razionali con numeratore e denominatore polinomi di grado rispettivamente al più k e m .

Capitolo 1

Esponenziale di matrice

In questo capitolo definiamo brevemente le funzioni di matrice e ne studiamo una in particolare, l'esponenziale di matrice. Evidenziamo alcune sue proprietà notevoli, tra cui l'applicazione ai sistemi di equazioni differenziali, che vedremo meglio nei capitoli successivi. Introduciamo poi l'approssimante di Padé e il metodo di *scaling and squaring*, che sono metodi numerici per approssimare l'esponenziale.

1.1 Funzioni di matrice

Il termine funzione di matrice può avere molteplici significati: noi siamo interessati a prendere una funzione scalare f e una matrice $A \in \mathbb{C}^{n \times n}$ e calcolare $f(A)$ che è una matrice delle stesse dimensioni di A .

Definizione 1.1. La funzione f si dice definita sullo spettro di A se esistono i valori

$$f^{(j)}(\lambda_j), \quad j = 0, \dots, n_i - 1, \quad i = 1, \dots, p$$

che sono detti valori della funzione f sullo spettro di A .

Definizione 1.2 (Funzione di matrice). Sia ora f definita sullo spettro di $A \in \mathbb{C}^{n \times n}$ con forma canonica di Jordan come in (1). Definiamo allora la funzione di matrice mediante forma canonica di Jordan come

$$f(A) = Zf(J)Z^{-1} = Z\text{diag}(f(J_1), \dots, f(J_p))Z^{-1}, \quad (1.1)$$

$$f(J_k) = \begin{bmatrix} f(\lambda_k) & f'(\lambda_k) & \cdots & \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f(\lambda_k) & \ddots & \vdots \\ & & \ddots & f'(\lambda_k) \\ & & & f(\lambda_k) \end{bmatrix}$$

Osserviamo che $f(A)$ richiede solamente la valutazione di f e delle sue derivate nei suoi autovalori e nessun'altra informazione su f .

Proposizione 1.1. *La definizione (1.2) non dipende dalla matrice Z .*

Dimostrazione. Consideriamo due forme canoniche di Jordan

$$A = ZJZ^{-1} = WJW^{-1} \quad (1.2)$$

A meno di moltiplicare W per una matrice di permutazione possiamo assumere senza perdita di generalità che J sia la stessa in entrambi i casi. Dalla definizione (1.1) abbiamo $f_1(A) = Zf(J)Z^{-1}$, $f_2(A) = Wf(J)W^{-1}$. Proviamo che $f_1(A) = f_2(A)$, cioè $W^{-1}Zf(J)Z^{-1}W = f(J)$. Poniamo $X = Z^{-1}W$ da cui dobbiamo provare $X^{-1}f(J)X = f(J)$. Ma ora da (1.2), $X^{-1}JX = J$ che implica $f(J) = f(X^{-1}JX) = X^{-1}f(J)X$, da cui la tesi. \square

Osservazione. Se A è diagonalizzabile, ossia $A = ZDZ^{-1}$ con $D = \text{diag}(\lambda_i)$ e le colonne di Z autovettori di A . Dalla definizione (1.1) otteniamo $f(A) = Zf(D)Z^{-1} = Z\text{diag}(f(\lambda_1), \dots, f(\lambda_p))Z^{-1}$. Pertanto, per matrici diagonalizzabili $f(A)$ ha gli stessi autovettori di A e autovettori ottenuti applicando f a quelli di A .

È possibile dare altre definizioni equivalenti per funzioni di matrice, ad esempio tramite interpolazione di Hermite oppure tramite il Teorema Integrale di Cauchy.

1.2 Proprietà di base

Definiamo l'esponenziale di matrice e osserviamo alcune sue proprietà importanti.

Definizione 1.3 (Esponenziale di matrice). Sia $A \in \mathbb{C}^{n \times n}$, definiamo l'*esponenziale di matrice* mediante sviluppo in serie di potenze

$$e^A = \exp(A) = \sum_{k=0}^{+\infty} \frac{A^k}{k!} = I + A + \frac{A^2}{2} + \cdots$$

Osservazione. La definizione è ben posta, in quanto la serie di potenze è assolutamente convergente con raggio di convergenza infinito. Nel caso $n = 1$ si ritrova la definizione formale della funzione esponenziale.

Definizione 1.4. Un altro modo di rappresentare l'esponenziale di matrice è

$$e^A = \lim_{s \rightarrow +\infty} (I + A/s)^s \quad (1.3)$$

Questa formula è il limite dello sviluppo di Taylor al primo ordine di A/s elevato alla potenza s . Più in generale, possiamo considerare i primi r termini nello sviluppo di Taylor di A/s elevati alla potenza s e prendere il limite per $r \rightarrow +\infty$ o $s \rightarrow +\infty$.

Il Teorema di Suzuki, dimostrato in seguito, garantisce che tale formula dà effettivamente e^A e stima l'errore in funzione di r e s . Prima di dimostrarlo, enunciamo un altro teorema sulla stima dell'errore di troncamento in serie di Taylor.

Proposizione 1.2. *Supponiamo di scrivere f in serie di Taylor centrata in α con raggio di convergenza r nel seguente modo*

$$f(z) = \sum_{k=0}^{+\infty} a_k (z - \alpha)^k, \quad a_k = \frac{f^{(k)}(\alpha)}{k!}. \quad (1.4)$$

Se $A \in \mathbb{C}^{n \times n}$ allora $f(A)$ è definita da

$$f(A) = \sum_{k=0}^{+\infty} a_k (A - \alpha I)^k \quad (1.5)$$

se e solo se ogni autovalore distinto $\lambda_1, \dots, \lambda_s$ di A soddisfa una delle seguenti condizioni:

1. $|\lambda_i - \alpha| < r$
2. $|\lambda_i - \alpha| = r$ e la serie di $f^{(n_i-1)}(\lambda)$ (dove n_i rappresenta l'indice di λ_i) converge a $\lambda = \lambda_i$, $i = 1, \dots, s$.

Dimostrazione. Dalla definizione (1.1) vediamo che è sufficiente dimostrare il teorema per un blocco di Jordan $A = J(\lambda) = \lambda I + N \in \mathbb{C}^{n \times n}$, dove N è una matrice strettamente

triangolare superiore. Poniamo $f_m(z) = \sum_{k=0}^m a_k(z - \alpha)^k$. Abbiamo

$$\begin{aligned}
f_m(J(\lambda)) &= \sum_{k=0}^m a_k((\lambda - \alpha)I + N)^k \\
&= \sum_{k=0}^m \sum_{i=0}^k \binom{k}{i} (\lambda - \alpha)^{k-i} N^i \\
&= \sum_{i=0}^m N^i \sum_{k=i}^m a_k \binom{k}{i} (\lambda - \alpha)^{k-i} \\
&= \sum_{i=0}^m \frac{N^i}{i!} \sum_{k=1}^m a_k k(k-1)\dots(k-i+1) (\lambda - \alpha)^{k-i} \\
&= \sum_{i=0}^m \frac{N^i}{i!} f_m^{(i)}(\lambda) = \sum_{i=0}^{\min(m, n-1)} \frac{N^i}{i!} f_m^{(i)}(\lambda).
\end{aligned} \tag{1.6}$$

Vediamo che $\lim_{m \rightarrow \infty} f_m(J(\lambda))$ esiste se e solo se esiste $\lim_{m \rightarrow \infty} f_m^{(i)}(\lambda)$ per $i = 1, \dots, n-1$: questo è essenzialmente il punto (2) della dimostrazione, perché differenziando la serie di f termine a termine $n_i - 1$ volte, questa converge a λ ; allora anche la serie differenziata j volte per $j = 0, \dots, n_i - 1$ converge. Il punto (1) segue da risultati di analisi complessa, in cui una serie di potenze derivata termine a termine converge all'interno del raggio di convergenza della serie originaria. \square

Teorema 1.3 (Errore di troncamento per serie di Taylor). *Supponiamo che f abbia sviluppo in serie di Taylor come in (1.4) con raggio di convergenza r . Se $A \in \mathbb{C}^{n \times n}$ con $\rho(A - \alpha I) < r$ allora*

$$\|f(A) - \sum_{k=0}^{s-1} a_k(A - \alpha I)^k\| \leq \frac{1}{s!} \max_{0 \leq t \leq 1} \|(A - \alpha I)^s f^{(s)}(\alpha I + t(A - \alpha I))\|. \tag{1.7}$$

Dimostrazione. Vedere Mathias [409, 1993, Cor.2]. \square

Teorema 1.4 (Suzuki). *Per $A \in \mathbb{C}^{n \times n}$, sia*

$$T_{r,s} = \left[\sum_{i=0}^r \frac{1}{i!} \left(\frac{A}{s}\right)^i \right]^s \tag{1.8}$$

Allora per ogni norma di matrice

$$\|e^A - T_{r,s}\| \leq \frac{\|A\|^{r+1}}{s^r(r+1)!} e^{\|A\|} \tag{1.9}$$

e $\lim_{s \rightarrow +\infty} T_{r,s}(A) = \lim_{r \rightarrow +\infty} T_{r,s}(A) = e^A$.

Dimostrazione. Sia $T = \sum_{i=0}^r \frac{1}{i!} \left(\frac{A}{s}\right)^i$ e sia $B = e^{A/s}$. Poiché T e B commutano, abbiamo che

$$e^A - T_{r,s} = B^s - T^s = (B - T)(B^{s-1} + B^{s-2}T + \dots + T^{s-1}), \text{ da cui}$$

$$\|e^A - T_{r,s}\| \leq \|B - T\| s \max_{i=0, \dots, s-1} \|B\|^i \|T\|^{s-i-1}$$

Ora, $\|T\| \leq \sum_{i=0}^r \frac{1}{i!} \left(\frac{\|A\|}{s}\right)^i \leq e^{\|A\|/s}$ e $\|B\|$ soddisfa la stessa disuguaglianza, allora

$$\|e^A - T_{r,s}\| \leq s \|e^{A/s} - T\| e^{\frac{s-1}{s}\|A\|}$$

Dal Teorema (1.3)

$$\|e^{A/s} - T\| \leq \frac{1}{(r+1)!} \left(\frac{\|A\|}{s}\right)^{r+1} e^{\|A\|/s}, \text{ da cui}$$

$$\|e^A - T_{r,s}\| \leq \frac{\|A\|^{r+1}}{s^r (r+1)!} e^{\|A\|}$$

Le verifiche dei limiti sono immediate. \square

Lemma 1.5 (Proprietà commutativa dell'esponenziale di matrice). *Siano $A, B \in \mathbb{C}^{n \times n}$.*

Allora

$$e^{tA} e^{tB} = e^{t(A+B)} \text{ per ogni } t \iff AB = BA.$$

Dimostrazione. Se $AB = BA$ allora tutti i termini nella serie di potenze di $e^{t(A+B)}$ e $e^{tA} e^{tB}$ commutano e quindi le matrici sono uguali. Se $e^{t(A+B)} = e^{tA} e^{tB}$ allora uguagliando le serie di potenze per ogni t e uguagliando i coefficienti con il termine di secondo grado t^2 nella serie di potenze, abbiamo $(AB + BA)/2 = AB$ da cui $AB = BA$. \square

Esempio. La commutatività di A e B non è necessaria affinché valga $e^{A+B} = e^A e^B$. Consideriamo ad esempio le seguenti matrici

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 2\pi i \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 2\pi i \end{bmatrix}$$

Vediamo subito che $e^{A+B} = e^A = e^B = I$ ma A e B non commutano.

Vale però il seguente teorema che non dimostriamo.

Teorema 1.6 (Wermuth). *Siano $A, B \in \mathbb{C}^{n \times n}$ con elementi algebrici. Allora*

$$e^A e^B = e^B e^A \iff AB = BA.$$

Osservazione. L'esponenziale di matrice è sempre invertibile: poiché A e $-A$ commutano, dal teorema precedente $e^A e^{-A} = e^0 = I_n$, dove 0 indica la matrice nulla, quindi $(e^A)^{-1} = e^{-A}$.

Osservazione. In generale, se A è una matrice sparsa, non abbiamo informazioni sulla sparsità di e^A : nel caso diagonale, allora l'esponenziale della matrice è anch'essa una matrice diagonale. Tuttavia, già nel caso di matrici tridiagonali, l'esponenziale è una matrice piena: vedremo un esempio nel Capitolo 4, in cui useremo l'esponenziale di una matrice tridiagonale per risolvere un sistema lineare ed essendo tale matrice piena il costo per la risoluzione è notevole.

1.3 Condizionamento

Definizione 1.5 (Derivata di Fréchet). Sia $f : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$ una funzione di matrice, la derivata di Fréchet di f in $X \in \mathbb{C}^{n \times n}$ è una mappa lineare

$$L : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}, \quad E \mapsto L(X, E) \tag{1.10}$$

tale che, per ogni $E \in \mathbb{C}^{n \times n}$

$$f(X + E) - f(X) - L(X, E) = o(\|E\|)$$

Vediamo ora alcune proprietà di tale operatore, in particolare la linearità e l'unicità.

Lemma 1.7 (Linearità). *La derivata di Fréchet è lineare.*

Dimostrazione. La verifica è immediata dalla definizione. □

Lemma 1.8. *La derivata di Fréchet potrebbe non esistere, ma se esiste è unica.*

Dimostrazione. Se L e M sono entrambe derivate di Fréchet di f allora

$$f(X + E) - f(X) = L(X, E) + o(\|E\|) = M(X, E) + o(\|E\|)$$

da cui $L(X, E) - M(X, E) = O(\|E\|)$. Sia ora, $E = t\tilde{E}$, con \tilde{E} fissata. Per linearità della derivata di Fréchet, $L(X, E) - M(X, E) = t(L(X, \tilde{E}) - M(X, \tilde{E})) = o(t\|\tilde{E}\|)$, cioè $L(X, \tilde{E}) - M(X, \tilde{E}) = t^{-1}o(t\|\tilde{E}\|)$. Mandando al limite per $t \rightarrow 0$ otteniamo $L(X, \tilde{E}) = M(X, \tilde{E})$. Dall'arbitrarietà di \tilde{E} , segue che $L(X) = M(X)$. \square

L'esponenziale di matrice soddisfa

$$e^{t(A+E)} = e^{tA} + \int_0^t e^{(t-s)A} E e^{s(A+E)} ds \quad (1.11)$$

essendo soluzione del Problema di Cauchy

$$\begin{cases} X'(t) = AX(t) + EX(t) \\ X(0) = I, \end{cases} \quad (1.12)$$

che vedremo meglio nel capitolo successivo. Dunque, per l'esponenziale di matrice esiste la derivata di Fréchet (ed è quindi unica per (1.8)). Da (1.11) sostituendo nell'integrale otteniamo

$$e^{t(A+E)} = e^{tA} + \int_0^t e^{(t-s)A} E e^{sA} ds + \mathcal{O}(\|E\|^2) \quad (1.13)$$

Se prendiamo $t = 1$ in (1.13) otteniamo la derivata di Fréchet dell'esponenziale di matrice.

Definizione 1.6 (Derivata di Fréchet dell'esponenziale di matrice). La derivata di Fréchet in A nella direzione di E è

$$L(A, E) = \int_0^1 e^{(1-s)A} E e^{sA} ds \quad (1.14)$$

Se A e E commutano allora $L(A, E) = Ee^A = e^A E$.

Definizione 1.7 (Condizionamento dell'esponenziale). Il numero di condizionamento relativo dell'esponenziale è

$$\kappa_{exp}(A) = \frac{\|L(A)\| \|A\|}{\|e^A\|}, \quad (1.15)$$

dove

$$\|L(A)\| = \max_{Z \neq 0} \frac{\|L(A, Z)\|}{\|Z\|} \quad (1.16)$$

Lemma 1.9 (Stima del condizionamento di A). *Sia $A \in \mathbb{C}^{n \times n}$. Per ogni norma di matrice indotta, vale*

$$\|A\| \leq \kappa_{exp}(A) \leq \frac{e^{\|A\|} \|A\|}{\|e^A\|} \quad (1.17)$$

Dimostrazione. Da (1.6) abbiamo

$$\|L(A, E)\| \leq \|E\| \int_0^1 e^{(1-s)\|A\|} e^{s\|A\|} ds = \|E\| \int_0^1 e^{\|A\|} ds = \|E\| e^{\|A\|}$$

Da cui è immediato $\|L(A)\| \leq e^{\|A\|}$. Inoltre, $\|L(A)\| \geq \|L(A, I)\| = \|\int_0^1 e^A ds\| = \|e^A\|$. \square

Teorema 1.10 (Van Loan). *Se $A \in \mathbb{C}^{n \times n}$ è normale, allora in norma-2, vale*

$$\kappa_{exp}(A) = \|A\|_2$$

Dimostrazione. Modifichiamo leggermente la stima precedentemente ottenuta; notiamo che per una matrice normale B abbiamo che $\|e^B\|_2 = e^{\alpha(B)}$. Dunque,

$$\begin{aligned} \|L(A, E)\|_2 &\leq \|E\|_2 \int_0^1 e^{(1-s)\alpha(A)} e^{s\alpha(A)} ds \\ &= \|E\|_2 \int_0^1 e^{\alpha(A)} ds = e^{\alpha(A)} \|E\|_2 = \|e^A\|_2 \|E\|_2. \end{aligned}$$

Quindi $\|L(A)\|_2 \leq \|e^A\|_2$, da cui la tesi. \square

1.4 Approssimante di Padé

Per una funzione scalare $f(x)$, la funzione razionale $r_{km}(x) = p_{km}(x)/q_{km}(x)$ si dice $[k/m]$ *approssimante di Padé* di f se $r_{km} \in \mathcal{R}_{k,m}$, $q_{km}(0) = 1$, e

$$f(x) - r_{km}(x) = \mathcal{O}(x^{k+m+1}) \quad (1.18)$$

Proposizione 1.11 (Unicità dell'approssimante di Padé). *Se un approssimante di Padé esiste, allora è unico. Parleremo quindi dell'approssimante di Padé di una funzione f .*

Dimostrazione. Siano $r_{km}(x) = p_{km}(x)/q_{km}(x)$ e $\tilde{r}_{km}(x) = \tilde{p}_{km}(x)/\tilde{q}_{km}(x)$ due $[k/m]$ approssimanti di Padé. Allora da (1.18)

$$r_{km} - \tilde{r}_{km} = \frac{p_{km}\tilde{q}_{km} - \tilde{p}_{km}q_{km}}{q_{km}\tilde{q}_{km}} = \mathcal{O}(x^{k+m+1}) \quad (1.19)$$

Ma $q_{km}\tilde{q}_{km} = \mathcal{O}(1)$, in quanto $q_{km}(0) = \tilde{q}_{km}(0) = 1$ e $p_{km}\tilde{q}_{km} - \tilde{p}_{km}q_{km}$ è un polinomio di grado al più $k + m$. Dunque (1.19) vale solo se $p_{km}\tilde{q}_{km} - \tilde{p}_{km}q_{km} = 0$, ossia $p_{km}/q_{km} = \tilde{p}_{km}/\tilde{q}_{km}$. \square

1.5 Metodo di *Scaling and squaring*

Il metodo di *scaling and squaring* sfrutta la relazione $e^A = (e^{A/\sigma})^\sigma$, per $A \in \mathbb{C}^{n \times n}$ e $\sigma \in \mathbb{C}$, unitamente al fatto che e^A si può ben approssimare con un approssimante di Taylor o Padé vicino all'origine, cioè per $\|A\|$ piccola. L'idea consiste nel prendere σ una potenza intera di 2, ad esempio $\sigma = 2^s$, così A/σ ha norma vicina a 1: approssimiamo poi $e^{A/2^s} \approx r(A/2^s)$, dove r è l'approssimante di Taylor o Padé dell'esponenziale. Infine, prendiamo $e^A \approx r(A/2^s)^{2^s}$, dove l'approssimazione è costituita da s elevamenti ripetuti.

Trattiamo il caso dell'approssimante di Padé, che è da preferire rispetto all'approssimazione mediante serie di Taylor, in quanto ha un'accuratezza migliore con minor costo computazionale.

Usando (1.18), consideriamo il $[k/m]$ approssimante di Padé dell'esponenziale: deve valere $e^x - r_{km}(x) = \mathcal{O}(x^{k+m+1})$. In tale caso, gli approssimanti di Padé sono noti esplicitamente per ogni k e m :

$$p_{km}(x) = \sum_{j=0}^k \frac{(k+m-j)!k!}{(k+m)!(k-j)!} \frac{x^j}{j!}, \quad q_{km}(x) = \sum_{j=0}^m \frac{(k+m-j)!m!}{(k+m)!(m-j)!} \frac{(-x)^j}{j!} \quad (1.20)$$

Notiamo che $p_{km}(x) = q_{mk}(-x)$, che riflette la proprietà che $1/e^x = e^{-x}$. Analogamente, è possibile osservare che $p_{mm}(x)$ e $q_{mm}(x)$ approssimano rispettivamente $e^{x/2}$ e $e^{-x/2}$, ma l'approssimazione è decisamente meno accurata di quanto $r_{mm} = p_{mm}/q_{mm}$ approssimi e^x .

Lemma 1.12. *Osserviamo che r_{km} è effettivamente l'approssimante di Padé per l'esponenziale.*

Dimostrazione. Verifichiamo che $q_{km}(0) = 1$ e per l'errore

$$e^x - r_{km}(x) = (-1)^m \frac{k!m!}{(k+m)!(k+m+1)!} x^{k+m+1} + \mathcal{O}(x^{k+m+2}) \quad (1.21)$$

□

Per le matrici, abbiamo un'esatta espressione dell'errore: sia $A \in \mathbb{C}^{n \times n}$ allora

$$e^A - r_{km}(A) = \frac{(-1)^m}{(k+m)!} A^{k+m+1} q_{km}(A)^{-1} \int_0^1 e^{tA} (1-t)^k t^m dt \quad (1.22)$$

Gli approssimanti diagonali ($k = m$) sono da preferire, poiché r_{km} con $k \neq m$ è meno accurata di r_{jj} con $j = \max(k, m)$. Inoltre, gli approssimanti diagonali sfruttano la proprietà che se gli autovalori di A giacciono nel semipiano aperto sinistro allora gli autovalori di $r_{mm}(A)$ hanno modulo minore di 1, che è importante nelle applicazioni alle equazioni differenziali.

Per brevità, scriviamo gli approssimanti diagonali come $r_m(x) = p_m(x)/q_m(x)$.

L'obiettivo è prendere s nella scalatura iniziale da A a $A/2^s$ e il grado di m dell'approssimante di Padé, in modo tale che l'esponenziale venga calcolato computazionalmente con un errore all'indietro limitato dall'unità di roundoff (cioè la precisione della macchina) e costo minimo.

La scelta di s si basa su $\|A\|$, dove $\|\cdot\|$ è una qualunque norma di matrice.

Gli obiettivi principalmente sono due:

1. limitare l'errore all'indietro in termini di $\|2^{-s}A\|$
2. determinare, per ogni m , il massimo $\|2^{-s}A\|$ per cui r_m possa garantire l'errore all'indietro voluto

Poniamo

$$e^{-A}r_m(A) = I + G = e^H \quad (1.23)$$

dove assumiamo $\|G\| < 1$, in modo tale che $H = \log(I + G)$ esista. Denotiamo con \log il logaritmo principale, ossia $\log(I + X) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} X^k$.

Lemma 1.13. *Se $\|G\| < 1$ allora $\|H\| \leq -\log(1 - \|G\|)$ e se $\|G\| < \frac{2}{3}$ allora $\|H\| \geq \|G\| - \frac{\|G\|^2}{2(1-\|G\|)}$.*

Dimostrazione. $H = \log(I + G) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} G^k$. Passando alle norme:

$$\|\log(I + G)\| \leq \sum_{k=1}^{\infty} \frac{\|G\|^k}{k} = -\log(1 - \|G\|) \leq \|G\| \sum_{k=0}^{\infty} \|G\|^k = \frac{\|G\|}{1 - \|G\|}$$

Similmente,

$$\begin{aligned} \|\log(I + G)\| &\geq \|G\| - \frac{\|G^2\|}{2} - \frac{\|G^3\|}{3} - \dots \\ &\geq \|G\| - \frac{\|G^2\|}{2}(1 + \|G\| + \|G\|^2 + \dots) = \|G\| - \frac{\|G\|^2}{2(1 - \|G\|)} \end{aligned}$$

□

G è una funzione di A e quindi lo è anche H , quindi H e A commutano tra loro. Ne segue che

$$r_m(A) = e^A e^H = e^{A+H} \quad (1.24)$$

Ora, sostituiamo A con $A/2^s$, con s intero positivo ed eleviamo ambo i membri per 2^s : otteniamo

$$r_m(A/2^s)^{2^s} = e^{A+E}, \quad (1.25)$$

dove $E = 2^s H$ soddisfa

$$\|E\| \leq -2^s \log(1 - \|G\|)$$

e G soddisfa (1.23) con A sostituita da $A/2^s$.

Da quanto osservato, otteniamo il seguente teorema.

Teorema 1.14. *Sia r_m l'approssimante diagonale di Padé che soddisfa*

$$e^{-2^{-s}A} r_m(2^{-s}A) = I + G, \quad (1.26)$$

con $\|G\| < 1$ in ogni norma di matrice. Allora

$$r_m(2^{-s}A)^{2^s} = e^{A+E},$$

dove E commuta con A e

$$\frac{\|E\|}{\|A\|} \leq -\frac{\log(1 - \|G\|)}{\|2^{-s}A\|} \quad (1.27)$$

Dimostrazione. Segue da quanto visto in precedenza. \square

Tale teorema dà un risultato sull'errore all'indietro: l'errore di troncamento nell'approssimante di Padé è equivalente alla perturbazione della matrice originaria A . Osserviamo che il risultato vale per ogni approssimazione razionale r_m , poiché non abbiamo usato proprietà specifiche dell'approssimante di Padé. La stima di tale errore all'indietro tiene in considerazione gli effetti della scalatura nell'approssimante di Padé e, a confronto con una stima dell'errore in avanti, non considera il condizionamento del problema.

Riportiamo in conclusione l'algoritmo per approssimare l'esponenziale di matrice mediante il metodo di scaling and squaring:

1. Determinare il minimo intero m tale che $\|A\|/m < 1$
2. Calcolare e porre $B := A/m$
3. Calcolare $p_m(B)$ e $q_m(B)$
4. Calcolare $e^B = p_m(B)/q_m(B)$ risolvendo il sistema lineare $q_m(B)e^B = p_m(B)$ mediante eliminazione di Gauss
5. Calcolare e porre $e^A := (e^B)^m$

In particolare, è conveniente prendere $m = 2^k$ per un certo intero k in modo tale da ridurre il numero di moltiplicazioni di matrici: in tal caso, l'esponenziale di matrice $e^A := (e^B)^{2^k}$ è ottenuto effettuando solo k moltiplicazioni di matrici, invece che 2^k .

Capitolo 2

Metodi numerici per la risoluzione di equazioni differenziali

In questo capitolo, vediamo come risolvere sistemi di equazioni differenziali, in cui è di fondamentale utilità l'esponenziale di matrici, e approssimiamo la soluzione mediante uso di metodi numerici a un passo.

Enunciamo, senza dimostrare il Teorema di esistenza e unicità locali per Problemi di Cauchy e il Teorema di esistenza globale.

Teorema 2.1 (Esistenza e unicità della soluzione per Problemi di Cauchy). *Sia $I \subseteq \mathbb{R}$ intervallo aperto, sia $\Omega \subseteq \mathbb{R}^n$ aperto. Sia $f : \Omega \times I \rightarrow \mathbb{R}^n$, $(t, y) \mapsto f(t, y)$ continua e localmente lipschitziana in y . Sia dato $(t_0, y_0) \in \Omega \times I$ e si consideri il Problema di Cauchy*

$$\begin{cases} y'(t) = f(t, y) \\ y(t_0) = y_0 \end{cases}$$

Allora esiste un'unica soluzione locale al Problema di Cauchy, cioè esiste $\delta > 0$ tale che $y : (t_0 - \delta, t_0 + \delta) \subseteq I \rightarrow \mathbb{R}^n$ risolve il Problema di Cauchy.

Teorema 2.2. *Sia f come nel teorema precedente. Se per ogni intervallo chiuso e limitato $[a, b] \subseteq I \exists M_1, M_2 \geq 0$ tali che*

$$\|f(t, y)\| \leq M_1 + M_2\|y\|, \quad \forall t \in [a, b]$$

Allora la soluzione massimale del Problema di Cauchy è definita su tutto I .

Lemma 2.3. *Osserviamo che se $f(t, y) = A(t)y(t) + b(t)$ con $A(t)$ matrice a coefficienti continui su I e b funzione continua su I allora valgono il Teorema 2.1 e il Teorema 2.2.*

Dimostrazione. f è continua, poiché somma e prodotto di funzioni continue, e localmente lipschitziana in y , in quanto

$$\|f(t, y_1) - f(t, y_2)\| = \|A(t)y_1(t) + b(t) - A(t)y_2(t) - b(t)\| \leq \max_{s \in [a, b]} \|A(s)\| \|y_1(t) - y_2(t)\|$$

Osserviamo che valgono anche le ipotesi del Teorema 2.2 infatti

$$\|f(t, y)\| \leq \max_{s \in [a, b]} \|A(s)\| \|y(t)\| + \max_{s \in [a, b]} \|b(s)\|$$

Da cui $M_1 = \max_{s \in [a, b]} \|b(s)\|$ e $M_2 = \max_{s \in [a, b]} \|A(s)\|$. □

Definizione 2.1 (Problema autonomo). Un Problema di Cauchy con dato iniziale $y(t_0) = y_0$ si dice autonomo se $f = f(y)$, cioè f non dipende esplicitamente dal tempo.

Consideriamo ora due Problemi di Cauchy autonomi, uno lineare e l'altro semi-lineare con matrice A costante e invertibile, $b : \Omega \rightarrow \mathbb{R}^n$, con $\Omega \subseteq \mathbb{R}^n$ aperto e $(t_0, u_0) \in \mathbb{R} \times \Omega$:

1. lineare

$$\begin{cases} u'(t) = Au(t) \\ u(t_0) = u_0 \end{cases} \quad (2.1)$$

2. semi-lineare

$$\begin{cases} u'(t) = Au(t) + b(u(t)) \\ u(t_0) = u_0 \end{cases} \quad (2.2)$$

Teorema 2.4. *La soluzione di (2.1) è definita su \mathbb{R} ed è*

$$u(t) = e^{(t-t_0)A}u_0 \quad (2.3)$$

Dimostrazione. Verifichiamo immediatamente che $u(t_0) = u_0$ e $u'(t) = Ae^{(t-t_0)A}u_0 = Au(t)$. La soluzione è unica e definita su \mathbb{R} per il Teorema 2.1 e il Teorema 2.2. □

Teorema 2.5. *La soluzione di (2.2) è*

$$u(t) = e^{(t-t_0)A}u_0 + \int_{t_0}^t e^{(t-s)A}b(u(s))ds \quad (2.4)$$

Dimostrazione. Verifichiamo immediatamente che $u(t_0) = u_0$ e $u'(t) = Ae^{(t-t_0)A}u_0 + A \int_{t_0}^t e^{(t-s)A}b(u(s))ds + b(u(t)) = Au(t) + b(u(t))$. L'unicità è garantita da (2.1). \square

Osserviamo che se b è costante, ossia $b(u(t)) = b$ allora la soluzione è

$$\begin{aligned} u(t) &= e^{(t-t_0)A}u_0 + \int_{t_0}^t e^{(t-s)A}ds b \\ &= e^{(t-t_0)A}u_0 + A^{-1}[e^{(t-s)A}]_{s=t_0}^{s=t} b \\ &= e^{(t-t_0)A}u_0 + A^{-1}b - A^{-1}e^{(t-t_0)A} b \end{aligned} \quad (2.5)$$

2.1 Metodi numerici a un passo

Cerchiamo ora dei metodi numerici che approssimino la soluzione di un Problema di Cauchy: ad esempio, nel caso semi-lineare (2.2) non conosciamo la soluzione in forma chiusa, poiché l'integrale dipende dalla soluzione stessa e quindi dobbiamo trovare un metodo per approssimarla.

Fissiamo $[t_0, T]$, intervallo di tempo in cui esiste la soluzione e discretizziamo tale intervallo in N_h nodi equispaziati di $h = \frac{T-t_0}{N_h-1}$, detto *passo di discretizzazione*, e otteniamo quindi

$$\{t_i\}_{i=\{0,\dots,N_h-1\}}, \quad t_n = t_0 + nh, \quad n = 1, \dots, N_h - 1, \quad (2.6)$$

da cui $t_{N_h-1} = t_0 + (N_h - 1)h = T$. Indichiamo con u_n la soluzione approssimata nei nodi di discretizzazione e $u(t)$ la soluzione esatta definita in $[t_0, T]$, cioè $u_n \approx u(t_n)$, $n = 1, \dots, N_h - 1$ e $u_0 = u(t_0)$ essendo il dato iniziale noto.

Consideriamo ora il problema (2.2), ricordando la soluzione

$$u(t) = e^{(t-t_0)A}u_0 + \int_{t_0}^t e^{(t-s)A}b(u(s))ds.$$

Supponendo nota la soluzione approssimata u_n al passo t_n possiamo ricavare la soluzione al passo successivo t_{n+1} , considerando come nuovo Problema di Cauchy quello con dato iniziale $u(t_n) = u_n$ quindi possiamo riscrivere la soluzione di (2.4) nel seguente modo:

$$\begin{aligned} u(t_{n+1}) &= e^{(t_{n+1}-t_n)A}u_0 + \int_{t_n}^{t_{n+1}} e^{(t_{n+1}-s)A}b(u(s))ds \\ &= e^{hA}u_0 + \int_{t_n}^{t_{n+1}} e^{(t_{n+1}-s)A}b(u(s))ds, \quad n = 0, \dots, N_h - 2 \end{aligned} \quad (2.7)$$

Notiamo che l'errore di approssimazione $\|u(t_n) - u_n\|$ ha effetti sulla soluzione al passo successivo, in quanto $u(t_n)$ non è nota esplicitamente, ma solo una sua approssimazione.

Trattiamo ora alcuni metodi a un passo molto basilari, cioè il *metodo di Eulero esplicito* (o *Forward Euler*) e il *metodo di Eulero esponenziale*.

2.1.1 Metodo di Eulero esplicito

Vediamo il metodo di Eulero esplicito, inizialmente nel caso generale, successivamente nel caso autonomo.

Osservazione. Data una funzione $f : I \rightarrow \mathbb{R}$ con $I \subseteq \mathbb{R}$ intervallo aperto; sia $x_i \in I$ tale che esista $f'(x_i)$. Possiamo approssimare con differenze finite in avanti e all'indietro tale derivata, per $h > 0$:

- in avanti: $f_i^{FD} = \frac{f(x_i+h) - f(x_i)}{h}$
- all'indietro: $f_i^{BD} = \frac{f(x_i) - f(x_i-h)}{h}$

L'errore di approssimazione dipende linearmente da h se f è di classe C^2 : infatti, dal Teorema di Taylor con resto di Lagrange, possiamo scrivere

$$f(x_i + h) = f(x_i) + hf'(x_i) + \frac{h}{2}f''(\xi), \quad h \rightarrow 0 \text{ e } \xi \in (x_i, x_i + h)$$

Dalla regolarità di f possiamo quindi limitare $f''(\xi)$ e quindi avere un'approssimazione lineare in h . Per le differenze finite all'indietro è analogo.

Consideriamo la discretizzazione nei nodi (2.6) e il seguente Problema di Cauchy:

$$\begin{cases} y'(t) = f(t, y(t)) \\ y(t_0) = y_0 \end{cases} \quad (2.8)$$

Mediante differenze finite in avanti, otteniamo il Metodo di Eulero esplicito

$$y_{n+1} = y_n + hf(t_n, y_n), \quad (2.9)$$

dove y_n è la soluzione approssimata nel nodo t_n .

Nel caso autonomo,

$$\begin{cases} y'(t) = f(y(t)) \\ y(t_0) = y_0 \end{cases} \quad (2.10)$$

otteniamo

$$y_{n+1} = y_n + hf(y_n), \quad (2.11)$$

con le solite notazioni.

2.1.2 Metodo di Eulero esponenziale

Consideriamo ora il Problema di Cauchy semi-lineare

$$\begin{cases} u'(t) = Au(t) + b(u(t)) \\ u(t_0) = u_0 \end{cases} \quad (2.12)$$

Osserviamo che l'equazione differenziale è costituita da una parte lineare con matrice invertibile A costante nel tempo e una parte non lineare $b(u)$. La soluzione è data dalla somma di una soluzione omogenea e una particolare soluzione dell'equazione differenziale: abbiamo già mostrato che la soluzione è

$$u(t) = e^{A(t-t_0)}u_0 + \int_{t_0}^t e^{-A(s-t)}b(u(s))ds.$$

Consideriamo la discretizzazione (2.6) e approssimiamo al nodo t_n , $b(u(t_n)) \approx b(u_n)$; essendo $b(u_n)$ costante in s otteniamo quindi:

$$u_{n+1} = e^{hA}u_n + \int_{t_n}^{t_{n+1}} e^{A(t_{n+1}-s)}ds b(u_n). \quad (2.13)$$

L'integrale si può calcolare esplicitamente:

$$\int_{t_n}^{t_{n+1}} e^{A(t_{n+1}-s)}ds = [A^{-1}(e^{A(t_{n+1}-s)})]_{s=t_n}^{s=t_{n+1}} = A^{-1}(e^{hA} - I_n).$$

In conclusione, il *metodo di Eulero esponenziale* è

$$u_{n+1} = e^{hA}u_n + A^{-1}(e^{hA} - I_n)b(u_n), \quad n = 1, \dots, N_h - 2 \quad (2.14)$$

2.2 Consistenza, convergenza e A-stabilità di un metodo

Definiamo ora tre concetti fondamentali per metodi numerici per la risoluzione di equazioni differenziali. In generale, è dato il Problema di Cauchy

$$\begin{cases} u'(t) = f(t, u(t)) \\ u(t_0) = u_0 \end{cases} \quad (2.15)$$

Un generico metodo a un passo per la risoluzione è della forma

$$u_{n+1} = u_n + h\phi(t_n, t_{n+1}, u_n, u_{n+1}, h, f) \quad (2.16)$$

Ad esempio, in Eulero esplicito

$$\phi(t_n, t_{n+1}, u_n, u_{n+1}, h, f) = f(t_n, u_n)$$

Definizione 2.2 (Consistenza). Sia dato il metodo numerico a un passo (2.16) e sia $u(t)$ la soluzione esatta al Problema di Cauchy (2.15). Introduciamo l'errore di troncamento locale $\tau_n(h)$: $h\tau_n(h)$ è l'errore che si ottiene se forziamo la soluzione esatta $u(t)$ del Problema di Cauchy (2.15) a soddisfare il metodo numerico. Otteniamo:

$$u(t_{n+1}) = u(t_n) + h\phi((t_n, t_{n+1}, u(t_n), u(t_{n+1}), h, f) + h\tau_n(h)$$

Definiamo l'errore di troncamento globale (o anche errore di troncamento) come

$$\tau(h) = \max_{n=0, \dots, N_h-1} |\tau_n(h)|$$

Un metodo per il quale l'errore di troncamento globale tende a 0 per $h \rightarrow 0$ è detto consistente. Diciamo che è di ordine p se $\tau(h) = \mathcal{O}(h^p)$.

Esempio. Il metodo di Eulero esplicito è consistente di ordine 1.

Definizione 2.3 (Stabilità). Un metodo numerico per l'approssimazione del problema (2.15) su un intervallo limitato $I = [t_0, T]$ è zero-stabile se

$$\begin{aligned} &\exists h_0 > 0, \exists C > 0, \exists \epsilon_0 > 0 \text{ tale che } \forall h \in (0, h_0], \forall \epsilon \in (0, \epsilon_0], \\ &\text{se } |\rho_n| \leq \epsilon, 0 \leq n \leq N_h - 1, \text{ allora } |z_n - u_n| \leq C\epsilon, 0 \leq n \leq N_h - 1, \end{aligned} \quad (2.17)$$

dove:

- C è una costante che può dipendere dalla lunghezza dell'intervallo di integrazione I , ma non da h
- u_n è la soluzione ottenuta con il metodo numerico
- z_n è la soluzione che si ottiene con il sistema perturbato
- ρ_n è la perturbazione al passo n -esimo
- ϵ_0 è la grandezza massima della perturbazione

Ad esempio, nel caso di Eulero in avanti, u_n soddisfa il problema

$$\begin{cases} u_{n+1} = u_n + hf(t_n, u_n) \\ u_0 = u(t_0) \end{cases} \quad (2.18)$$

mentre z_n soddisfa il problema perturbato

$$\begin{cases} z_{n+1} = u_n + h[f(t_n, z_n) + \rho_{n+1}] \\ z_0 = u(t_0) + \rho_0 \end{cases} \quad (2.19)$$

Enunciamo ora senza dimostrare due teoremi che assicurano la convergenza di metodi a un passo.

Teorema 2.6. *Un metodo a un passo (2.16), la cui funzione ϕ soddisfi una proprietà di lipschitzianità rispetto a u_n e u_{n+1}*

$$\begin{aligned} \phi(t_n, t_{n+1}, u_n, u_{n+1}, h, f) &= 0, \\ \exists L_\phi > 0 : |\phi(t_n, t_{n+1}, u_n, u_{n+1}, h, f) - \phi(t_n, t_{n+1}, w_n, w_{n+1}, h, f)| & \\ \leq L_\phi \sum_{k=0,1} |u_{n+k} - w_{n+k}|, & \end{aligned} \quad (2.20)$$

dove L_ϕ è una costante che può dipendere da f , ma non da h , $\{u_n\}$, $\{w_n\}$, t_n , t_{n+1} . Allora il metodo è zero - stabile.

Teorema 2.7. *Un metodo a un passo (2.15) consistente e la cui funzione ϕ soddisfi (2.20) è convergente.*

Trattiamo infine l'assoluta stabilità: consideriamo il problema modello con $\lambda \in \mathbb{C}$:

$$\begin{cases} y'(t) = \lambda y(t), & t \in (0, +\infty) \\ y(0) = 1, \end{cases} \quad (2.21)$$

che ha come soluzione $y(t) = e^{\lambda t}$. La soluzione tende a 0 se $\Re(\lambda) < 0$.

Definizione 2.4. Definiamo \mathcal{A} regione di assoluta stabilità di un metodo numerico nel seguente modo

$$\mathcal{A} = \{z = h\lambda \in \mathbb{C} : |u_n| \rightarrow 0 \text{ per } t_n \rightarrow 0\} \quad (2.22)$$

Un metodo si dice assolutamente stabile o \mathcal{A} -stabile se $\mathcal{A} \supseteq \mathbb{C}^- = \{z \in \mathbb{C} : \Re(z) < 0\}$.

Tale definizione dice che non necessariamente per ogni passo di discretizzazione h la soluzione numerica si comporta come la soluzione esatta, ma ciò avviene solo se il passo h è sufficientemente piccolo:

Esempio. Per Eulero esplicito vale

$$h < -\frac{2\Re(\lambda)}{|\lambda|^2}$$

quindi non è un metodo \mathcal{A} -stabile.

Definizione 2.5 (Convergenza di ordine p). Un metodo numerico si dice convergente di ordine p se

$$\exists C \geq 0 \text{ tale che } \|u_n - u(t_n)\| \leq Ch^{p+1}, \quad n = 1, \dots, N_h - 1 \quad (2.23)$$

Per brevità, scriviamo anche

$$\|u_n - u(t_n)\| = \mathcal{O}(h^{p+1}), \quad n = 1, \dots, N_h - 1 \quad (2.24)$$

Capitolo 3

Integratori esponenziali

Gli integratori esponenziali costituiscono un'ampia classe di metodi numerici per l'integrazione nel tempo di equazioni differenziali autonome $u'(t) = F(u(t))$, $u(t_0) = u_0$: in particolare trattiamo metodi espliciti e ne studiamo la convergenza.

Definizione 3.1 (Integratore esponenziale). Definiamo l'integratore esponenziale di ordine k per una matrice A invertibile

$$\varphi_k(A) = \int_0^1 e^{(1-\tau)A} \frac{\tau^{k-1}}{(k-1)!} d\tau, \quad k \geq 1 \quad (3.1)$$

Definizione 3.2. Nel corso del capitolo, faremo spesso uso dell'integratore esponenziale del primo ordine

$$\varphi_1(A) = \int_0^1 e^{(1-\tau)A} = [-A^{-1}e^{(1-\tau)A}]_{\tau=0}^{\tau=1} = A^{-1}(e^A - I) \quad (3.2)$$

Osservazione. Gli integratori esponenziali fanno uso dell'esponenziale di matrice: in particolare, per φ_1 notiamo che A^{-1} e e^A commutano quindi è equivalente scrivere $\varphi_1(A) = A^{-1}(e^A - I) = (e^A - I)A^{-1}$.

Vediamo ora di applicare gli integratori esponenziali, in particolare quello del primo ordine, al Metodo di Eulero esponenziale per problemi differenziali semi-lineari e cerchiamo di estenderlo al caso autonomo non-lineare.

Consideriamo il Problema di Cauchy

$$\begin{cases} u'(t) = Au(t) + b(u(t)) \\ u(t_0) = u_0 \end{cases}$$

Consideriamo una discretizzazione in nodi più generale, ossia senza nodi equispaziati, e poniamo $h_n = t_{n+1} - t_n$. Otteniamo quindi una leggera modifica di (2.14) cioè

$$u_{n+1} = e^{h_n A} u_n + A^{-1}(e^{h_n A} - I)b(u_n), \quad n = 1, \dots, N_h - 2,$$

dove abbiamo sostituito il passo h con h_n . Osserviamo ora che

$$\varphi_1(h_n A) = \frac{A^{-1}}{h_n}(e^{h_n A} - I)$$

Otteniamo quindi il metodo di Eulero esponenziale

$$u_{n+1} = e^{h_n A} u_n + h_n \varphi_1(h_n A) b(u_n), \quad n = 1, \dots, N_h - 2 \quad (3.3)$$

3.1 Metodo di Rosenbrock - Eulero

Il metodo precedente è applicabile a problemi costituiti da un termine lineare e un termine non lineare: in generale, si cerca quindi linearizzare il problema, prendendo come termine lineare la matrice Jacobiana del problema.

Consideriamo il Problema di Cauchy autonomo e non lineare

$$\begin{cases} u'(t) = f(u(t)) \\ u(t_0) = u_0 \end{cases}$$

Discretizziamo l'intervallo temporale $[t_0, t_{N_h-1}]$ come in (2.6) e linearizziamo il problema:

$$u'(t) = J_n u(t) + g_n(u(t)), \quad (3.4)$$

dove $J_n(u(t)) = \frac{\partial F}{\partial u}(u(t))$ denota la derivata di Fréchet di f e $g_n(u(t)) = f(u(t)) - J_n u(t)$.

Applichiamo quindi il *metodo di Eulero esponenziale*, poiché il problema è costituito da una parte lineare e una non lineare che dipende da u :

$$u_{n+1} = e^{h_n J_n} u_n + h_n \varphi_1(h_n J_n) g_n(u_n), \quad n = 1, \dots, N_h - 2 \quad (3.5)$$

Usando la definizione di g_n e di φ_1 , possiamo anche scrivere:

$$\begin{aligned} u_{n+1} &= e^{h_n J_n} u_n + h_n \varphi_1(h_n J_n) f(u_n) - h_n \varphi_1(h_n J_n) J_n u_n \\ &= e^{h_n J_n} u_n + h_n \varphi_1(h_n J_n) f(u_n) - e^{h_n J_n} u_n + u_n \\ &= u_n + h_n \varphi_1(h_n J_n) f(u_n) \end{aligned} \quad (3.6)$$

Abbiamo ottenuto il *metodo di Rosenbrock - Eulero*:

$$u_{n+1} = u_n + h_n \varphi_1(h_n J_n) f(u_n), \quad n = 1, \dots, N_h - 2 \quad (3.7)$$

3.2 Convergenza dei metodi

Studiamo ora la convergenza del metodo di Eulero esponenziale per problemi semi-lineari.

Proposizione 3.1 (convergenza del metodo di Eulero esponenziale). *Per il metodo di Eulero esponenziale (2.14)*

- se $u(t_n) = u_n$, ossia al passo n la soluzione approssimata coincide con quella esatta, allora

$$\|u(t_{n+1}) - u_{n+1}\| = \mathcal{O}(h^2)$$

- se $b(u(t)) = b(u_0) = b$, ossia il termine b è costante, allora

$$u(t_n) = u_n$$

Dunque il metodo è esatto per problemi lineari autonomi a coefficienti costanti, ed è di ordine 1 altrimenti.

Dimostrazione. Ricordiamo che

$$u_{n+1} = e^{hA} u_n + \int_{t_n}^{t_{n+1}} e^{(t_{n+1}-s)A} b(u(t_n)) ds$$

e

$$u(t_{n+1}) = e^{hA} u_n + \int_{t_n}^{t_{n+1}} e^{(t_{n+1}-s)A} b(u(s)) ds.$$

Valutiamo l'errore e poniamo $r(t) = b(u(t))$

$$\begin{aligned} u(t_{n+1}) - u_{n+1} &= \int_{t_n}^{t_{n+1}} e^{(t_{n+1}-s)A} (r(s) - r(t_n)) ds = \\ &= \int_{t_n}^{t_{n+1}} e^{(t_{n+1}-s)A} (r(t_n) + r'(\epsilon)(s - t_n) - r(t_n)) ds = \\ &= h^2 \varphi_2(hA) r'(\epsilon) = \mathcal{O}(h^2) \end{aligned}$$

□

Proposizione 3.2 (convergenza del metodo di Rosenbrock - Eulero). *Il metodo di Rosenbrock - Eulero è convergente di ordine 2, cioè*

$$\|u(t_{n+1}) - u_{n+1}\| = \mathcal{O}(h^3)$$

Dimostrazione. Accenniamo brevemente alla dimostrazione: consideriamo $\tilde{u}_{n+1} = e^{h_n J_n} + h_n \varphi(h_n J_n) g_n(u_{n+1})$. Il termine dominante nel controllo dell'errore è

$$\begin{aligned} \tilde{u}_{n+1} - u_{n+1} &= h_n \varphi(h_n J_n) (g_n(u_{n+1}) - g_n(u_n)) = \\ &= h_n \varphi(h_n J_n) \left(\frac{1}{2} \frac{\partial^2 f}{\partial u^2}(u_n) (u_{n+1} - u_n)^2 + \dots \right) \end{aligned}$$

□

3.3 Aspetti implementativi

Per semplificare l'implementazione in MATLAB[®] e alleggerire la notazione considereremo il passo di discretizzazione costante, quindi $h_n = h$: ciò permette inoltre di ridurre il costo computazionale del problema. Confrontiamo ora i metodi di Eulero esponenziale e di Rosenbrock - Eulero: il primo ha convergenza lineare in h e il secondo quadratica in h .

Eulero esponenziale

Riscriviamo brevemente l'iterazione:

$$u_{n+1} = e^{h_n A} u_n + h_n \varphi_1(h_n A) b(u_n), \quad n = 1, \dots, N_h - 2$$

Osserviamo che a ogni iterazione è necessario risolvere un sistema lineare $Az = b(u_n)$, che costa in generale $\mathcal{O}(n^3)$ flops: se la matrice ha proprietà particolari, ad esempio è a banda, si può abbassare il costo; se la matrice è mal condizionata il calcolo di z potrebbe essere lontano dalla soluzione esatta. Se h_n varia, a ogni iterazione è necessario calcolare l'esponenziale di matrice $e^{h_n A}$ e si hanno inoltre 2 prodotti matrice per vettore: il primo $e^{h_n A} u_n$ e il secondo $(e^{h_n A} - I_n)z = e^{h_n A} z - z$. Spesso risulta conveniente dal punto di vista computazionale lavorare con passo h_n costante, in modo tale da non dover ricalcolare a ogni iterazione l'esponenziale di matrice, ma solamente i prodotti matrice per vettore e risolvere il sistema lineare. In generale, per tale metodo otteniamo un moderato costo

computazionale e un ordine di convergenza lineare.

Rosenbrock - Eulero

Riscriviamo brevemente l'iterazione:

$$u_{n+1} = u_n + h_n \varphi_1(h_n J_n) f(u_n), \quad n = 1, \dots, N_h - 2$$

Osserviamo che a ogni iterazione è necessario calcolare J_n o una sua approssimazione, ad esempio mediante differenze finite; di conseguenza, l'esponenziale della matrice $h_n J_n$ viene quindi ricalcolato a ogni iterazione. Inoltre, a ogni iterazione si risolve un sistema lineare $J_n z = f(u_n)$ (costo $\mathcal{O}(n^3)$ flops), si valuta $f(u_n)$ e si ha 1 prodotto matrice per vettore $e^{h_n J_n} z$. In generale, per tale metodo otteniamo un alto costo computazionale e un ordine di convergenza quadratico.

A parità di nodi, Rosenbrock - Eulero impiega decisamente molto più tempo computazionale rispetto a Eulero esplicito (ad esempio, nella tabella (4.2)), a maggior ragione se la Jacobiana deve essere approssimata mediante differenze finite: confronteremo nel capitolo successivo i tempi di esecuzione del codice. Tuttavia il vantaggio consiste nel fatto che un ordine di convergenza quadratico permette di utilizzare un numero minore di nodi nella discretizzazione, cioè un passo h più grande, rispetto a un metodo convergente solo linearmente.

Teorema 3.3. *Il metodo di Rosenbrock - Eulero è convergente di ordine 2 e \mathcal{A} -stabile.*

Capitolo 4

Applicazione alle equazioni differenziali alle derivate parziali

In questo capitolo applichiamo il metodo di Rosenbrock - Eulero con gli integratori esponenziali e lo confrontiamo con il metodo di Eulero esplicito. In particolare, studiamo un modello classico di un problema parabolico, che richiede la discretizzazione sia in una dimensione spaziale che nel tempo.

4.1 Equazione di Allen Cahn

Consideriamo l'equazione differenziale evolutiva di Allen Cahn con condizioni omogenee di Dirichlet al bordo con $x \in [-1, 1]$, $t \in [0, 3]$:

$$\begin{cases} \frac{\partial}{\partial t}u(x, t) = \epsilon \frac{\partial^2}{\partial x^2}u(x, t) + (u + x) - (u + x)^3 \\ u(x, 0) = s(x) = 0.53x + 0.47\sin(-\frac{3}{2}\pi x) - x \\ u(-1, t) = u(1, t) = g(t) = 0 \end{cases} \quad (4.1)$$

L'equazione di Allen Cahn è un modello fisico matematico di tipo reattivo - diffusivo e descrive il processo di separazione di fase in sistemi di leghe multicomponente: la variabile x rappresenta lo spazio, la t il tempo. Scegliamo inoltre il parametro $\epsilon = 0.01$. Osserviamo anche che le condizioni sono compatibili al bordo, infatti $s(-1) = s(1) = 0 = g(0) = g(3)$.

Fissiamo con N numero di nodi nello spazio e con M quello nel tempo e discretizziamo in nodi equispaziati:

- l'intervallo spaziale x_0, x_1, \dots, x_{N-1} da cui $h = \frac{x_{N-1} - x_0}{N-1}$
- l'intervallo temporale t_0, t_1, \dots, t_{M-1} da cui $k = \frac{t_{M-1} - t_0}{M-1}$

Approssimiamo la soluzione nei nodi (x_i, t_j) con il metodo delle linee: in Figura 4.1 è rappresentato uno schema di approssimazione. La soluzione lungo le linee colorate è già nota dalle condizioni iniziali (in blu) e da quelle al bordo (in verde) quindi dobbiamo ricavare la soluzione approssimata solo nei nodi interni alla griglia.

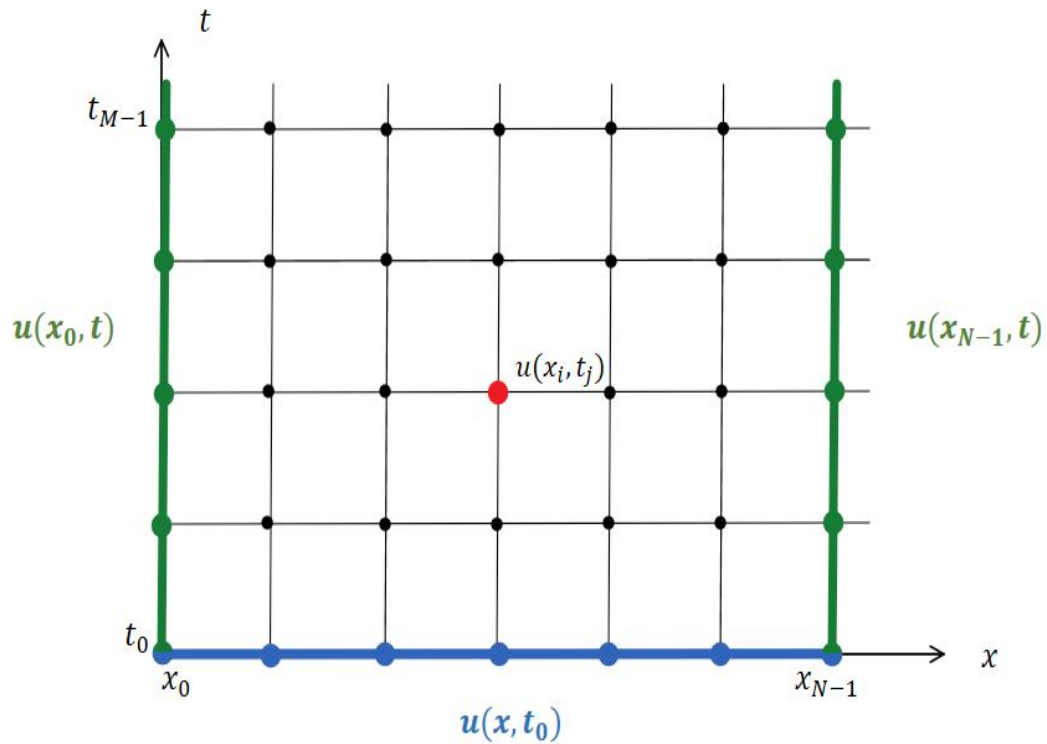


Figura 4.1: Metodo delle linee

Fissiamo un nodo temporale t_j e approssimiamo la derivata parziale seconda spaziale mediante differenze finite:

$$\frac{\partial^2}{\partial x^2} u(x_i, t_j) \approx \frac{u(x_{i-1}, t_j) - 2u(x_i, t_j) + u(x_{i+1}, t_j))}{h^2} \quad i = 1, \dots, N-2 \quad (4.2)$$

Per $i = 1$ è noto esplicitamente il termine $u(x_0, t_j)$ e analogamente $i = N - 1$ si ha $u(x_N, t_j)$.

Otteniamo quindi per i nodi interni:

$$\frac{\partial^2}{\partial x^2} u(x, t_j) \approx \frac{1}{h^2} \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 \end{bmatrix} \begin{bmatrix} u(x_1, t_j) \\ u(x_2, t_j) \\ \vdots \\ u(x_{N-2}, t_j) \\ u(x_{N-1}, t_j) \end{bmatrix} + \frac{1}{h^2} \begin{bmatrix} u(x_0, t_j) \\ 0 \\ \vdots \\ 0 \\ u(x_N, t_j) \end{bmatrix}$$

Da (4.1) le condizioni al bordo sono nulle, quindi il vettore a destra è nullo.

Sia

$$A = \frac{\epsilon}{h^2} \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 \end{bmatrix} \quad (4.3)$$

La matrice A è simmetrica, definita negativa e ha un condizionamento di $\mathcal{O}(\frac{1}{h^2})$.

Fissiamo ora un un nodo spaziale x_i , per cui il problema diventa

$$\begin{cases} u'(x_i, t) = f(u(x_i, t)) = Au(x_i, t) + (u(x_i, t) + x_i) - (u(x_i, t) + x_i)^3 \\ u(x_i, t_0 = 0) = s(x_i) \end{cases} \quad (4.4)$$

Abbiamo ridotto il problema nella forma (2.2) e possiamo quindi applicare i metodi visti in precedenza, in particolare utilizziamo gli integratori esponenziali, discretizzando nel tempo con passo k , cioè

$$u(x_i, t_{j+1}) = u(x_i, t_j) + k\varphi_1(kJ_n)f(u(x_i, t_j)). \quad (4.5)$$

Riportiamo di seguito il codice MATLAB[®] per la risoluzione: osserviamo che la Jacobiana viene approssimata mediante differenze finite a ogni iterazione, ma nel caso dell'esempio è possibile anche calcolarla a mano (confronteremo in seguito con altri metodi e tempo computazionale). La funzione *Jacobian* approssima numericamente la

matrice Jacobiana di f : prende in input un punto x , la funzione f e la valutazione di f in x , chiamata f_0 . La funzione *Directional derivative* approssima la derivata direzionale di f lungo un vettore in un punto: x , f e f_0 sono come sopra e w indica il vettore lungo cui si deriva. La funzione *phi1* prende in input una matrice M e f_0 e restituisce $\varphi(M)f_0$: usiamo due comandi predefiniti in MATLAB[®] per calcolare l'esponenziale di matrice **expm** (che utilizza l'approssimante di Padé visto nella sezione 1.4) e risolvere sistemi lineari **backslash** \. Riportiamo qui di seguito il codice.

```

1  epsilon = 0.01;
2  x0 = -1; xfinal = 1; N = 100; %intervallo spaziale
3  t0 = 0; tfinal = 3; M = 200; %intervallo temporale
4  alfa = zeros(1,M); beta = zeros(1,M); %condizioni al bordo
5  s = @(x)(0.53*x + 0.47*sin(-3*pi*x/2) - x);
6  hx = (xfinal-x0)/(N-1); %passo di discretizzazione spaziale
7  ht = (tfinal-t0)/(M-1); %passo di discretizzazione temporale
8
9  x = linspace(x0,xfinal,N)';
10 t = linspace(t0,tfinal,M)';
11
12 e = ones(N-2,1);
13 A = (epsilon/hx^2)*(spdiags([e, -2*e, e],-1:1,N-2,N-2));
14
15 u = zeros(N,M);
16 u(:,1) = s(x); %u(x,0) = s(x) condizione al bordo
17
18 x(1) = []; x(end) = [];
19 f = @(u)(A*u + (u+x) - (u+x).^3);
20
21 for j = 1:M-1
22     f0 = f(u(2:end-1,j));
23     J = Jacobian(u(2:end-1,j),f,f0);
24     u(2:end-1,j+1) = u(2:end-1,j) + ht*phi1(ht*J,f0);

```

```
25 end
26
27 u(1,:) = alfa;
28 u(end,:) = beta;
29
30 x = linspace(x0,xfinal,N)';
31
32 mesh(t,x,u);
33 xlabel('t'); ylabel('x'); zlabel('u(x,t)');
34
35 -----
36
37 function J = Jacobian(x,f,f0)
38
39 n = length(x);
40
41 for j = 1:n
42     z = zeros(n,1);
43     z(j) = 1; %base canonica
44     J(:,j) = Directional_Derivative(x,z,f,f0);
45 end
46
47 -----
48
49 function z = Directional_Derivative(x,w,f,f0)
50
51 n = length(x);
52
53 if norm(w) == 0
54     z = zeros(n,1);
55 return
```

```

56 end
57
58 h = 1e-7; %incremento differenze finite
59 h = h/norm(w); %normalizzazione di w
60
61 deltax = x + h*w;
62 f1 = f(deltax);
63
64 z = (f1 - f0)/h;
65
66 -----
67
68 function v = phi1(M,f0)
69
70 v = (expm(M)-speye(length(M)))*(M\f0);
71
72 end

```

Riportiamo ora la figura ottenuta, usando $N = 100$ nodi spaziali e $M = 200$ nodi temporali e otteniamo:

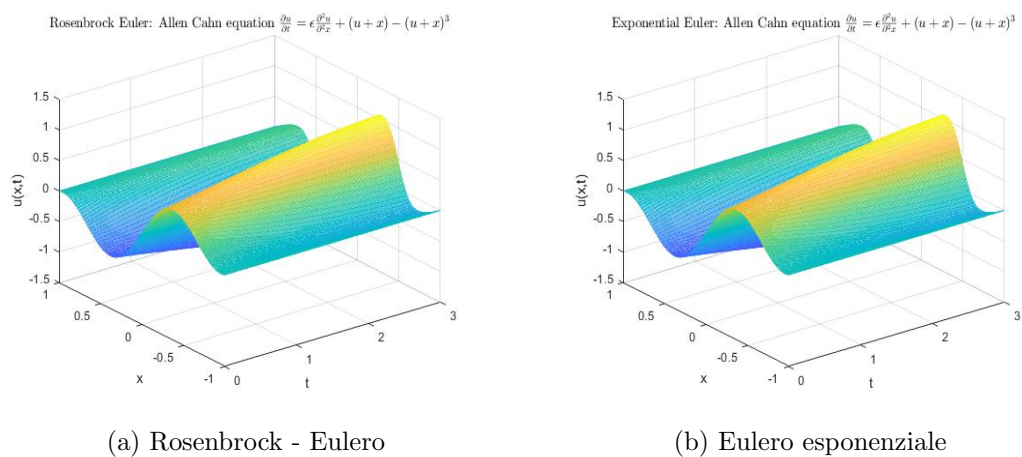
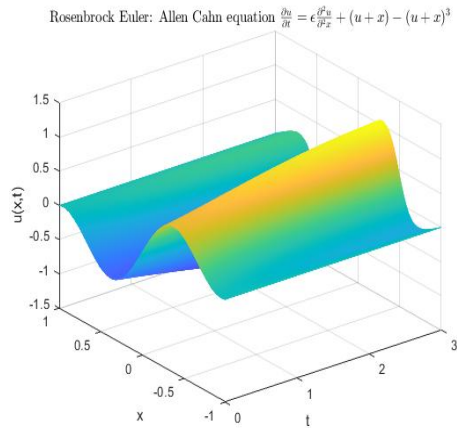
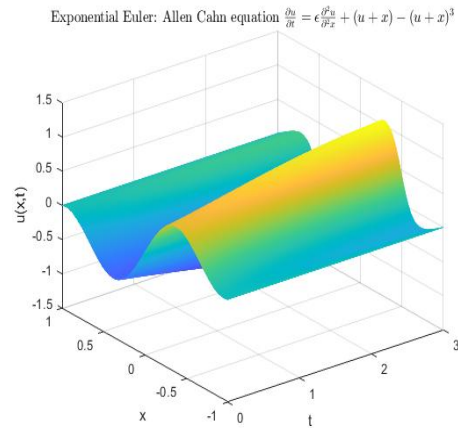


Figura 4.2: $N = 100$, $M = 200$, $\epsilon = 0.01$

Raffiniamo ora la griglia dei nodi:



(a) Rosenbrock - Eulero



(b) Eulero esponenziale

Figura 4.3: $N = 500, M = 600, \epsilon = 0.01$

Usando Eulero esplicito abbiamo problemi di stabilità: come osserviamo nella Figura 4.4 il grafico a destra non viene nemmeno disegnato dopo pochi passi di iterazione in quanto la soluzione esplose (vediamo valori di 10^{305} sugli assi).

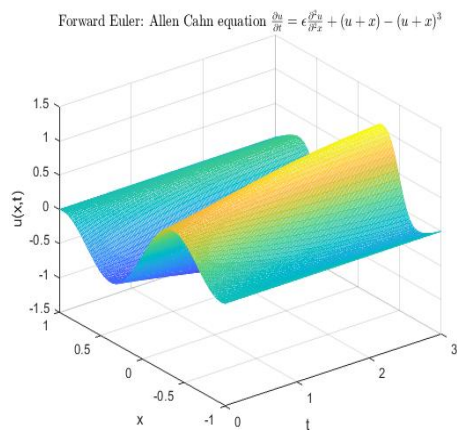
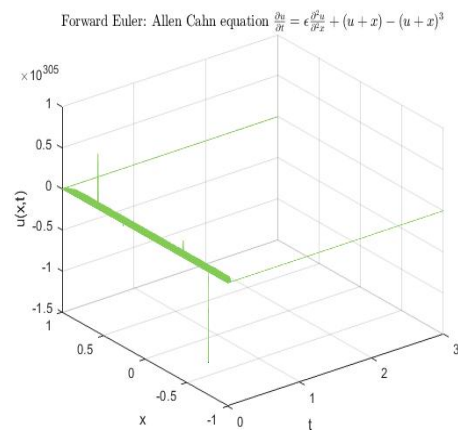
(a) $N = 100, M = 200$ (b) $N = 500, M = 600$

Figura 4.4: Eulero esplicito con $\epsilon = 0.01$

Riportiamo in tabella la differenza in norma-2 tra le varie soluzioni ottenute: nei casi in cui Eulero esplicito non converge, omettiamo i risultati ottenuti. Denotiamo con u_{RE} la soluzione ottenuta con Rosenbrock - Eulero, u_{EE} con Eulero esponenziale e u_{FE} con Eulero esplicito:

| N, M | $\ u_{RE} - u_{EE}\ $ | $\ u_{RE} - u_{FE}\ $ | $\ u_{FE} - u_{EE}\ $ |
|----------|-----------------------|-----------------------|-----------------------|
| 100, 200 | 6.347432e-02 | 3.451248e-02 | 7.622385e-02 |
| 500, 600 | 8.209779e-02 | - | - |
| 700, 200 | 1.691501e-01 | - | - |
| 200, 700 | 4.792188e-02 | 2.628842e-02 | 5.746533e-02 |

Tabella 4.1: Differenza in $\|\cdot\|_2$ delle soluzioni usando i vari metodi

Riportiamo anche i tempi computazionali ottenuti con i vari metodi, usando le funzioni predefinite *tic* e *toc* di MATLAB[®]; nei casi in cui non c'è convergenza, omettiamo il dato.

| N, M | t_{RE} | t_{FE} | t_{EE} |
|----------|--------------|--------------|--------------|
| 100, 200 | 1.325946e+00 | 9.282000e-03 | 1.801310e-02 |
| 500, 600 | 1.697712e+02 | - | 1.743231e+00 |
| 700, 200 | 1.514615e+02 | - | 1.204883e+00 |
| 200, 700 | 1.757340e+01 | 4.341180e-02 | 1.351012e-01 |

Tabella 4.2: Tempi computazionali in secondi dei vari metodi

Osservazione. I risultati ottenuti in norma-2 differiscono di 10^{-2} : questo significa che quando abbiamo convergenza i grafici ottenuti sono buoni e anche aumentando i nodi non si notano differenze. I tempi computazionali differiscono invece notevolmente, anche per molti ordini di grandezza. Certamente, Rosenbrock - Eulero impiega molto più tempo di Eulero esplicito ed esponenziale, però la convergenza è quadratica a differenza degli altri due metodi.

Proviamo ora a calcolare a mano la Jacobiana: si vede facilmente che la Jacobiana è una matrice tridiagonale.

Inseriamo nel codice la Jacobiana come *handle* di funzioni

```
1 Jac = @(u) (A + (1-3.*(u+x).^2).*speye(N-2));
```

e sostituiamo la riga 23 del codice con

```
1 J = Jac(u(2:end-1,j));
```

Confrontiamo ora i tempi computazionali ottenuti con il metodo di Rosenbrock - Eulero, usando la funzione *Jacobian* e la Jacobiana come *handle*: i dati sono riassunti nella tabella sottostante.

| N, M | Jacobiana come <i>handle</i> | Jacobiana con <i>Jacobian</i> |
|----------|------------------------------|-------------------------------|
| 100, 200 | 3.5155e-01 | 1.325946e+00 |
| 500, 600 | 2.7712e+01 | 1.697712e+02 |
| 700, 200 | 2.2783e+01 | 1.514615e+02 |
| 200, 700 | 3.4995e+00 | 1.757340e+01 |

Tabella 4.3: Confronto in termini di tempo computazionale del metodo di Rosenbrock - Eulero usando la Jacobiana come *handle* e la funzione *Jacobian*

Osservazione. I tempi computazionali vengono ridotti di 1 ordine di grandezza, quindi, come ci aspettavamo, approssimare numericamente la Jacobiana impiega molto più tempo che averla a disposizione come *handle*. Nonostante ciò, i tempi sono comunque maggiori rispetto a Eulero esplicito ed esponenziale: tuttavia il primo a volte non converge, il secondo invece ha convergenza lineare. Possiamo quindi usare molti meno nodi nel metodo di Rosenbrock - Eulero, avendo convergenza quadratica.

Consideriamo ora il caso $\epsilon = 0.05$, osserviamo in particolare eventuali differenze con il caso precedente. Maggiore è ϵ , maggiore è il termine stiff del problema: infatti, il termine lineare del problema con matrice A corrisponde alla parte stiff del problema, mentre il termine non lineare $b(u)$ corrisponde alla parte non stiff. Per tale motivo, anche nel caso $N = 100, M = 200$ Eulero esplicito non converge, poiché il passo di discretizzazione è troppo grande, come vediamo dalla Figura 4.7. Riportiamo di seguito le soluzioni ottenute con i due metodi convergenti per confrontarle; i tempi computazionali e gli errori

ottenuti sono dello stesso ordine di grandezza di quelli ottenuti nel caso precedente con $\epsilon = 0.01$.

Osserviamo infine che la matrice A è simmetrica reale, dunque ha autovalori reali ed è definita negativa, in particolare, all'aumentare di h l'autovalore più piccolo in modulo si avvicina a 0, il che rende il condizionamento della matrice sempre più elevato all'aumentare dei nodi.

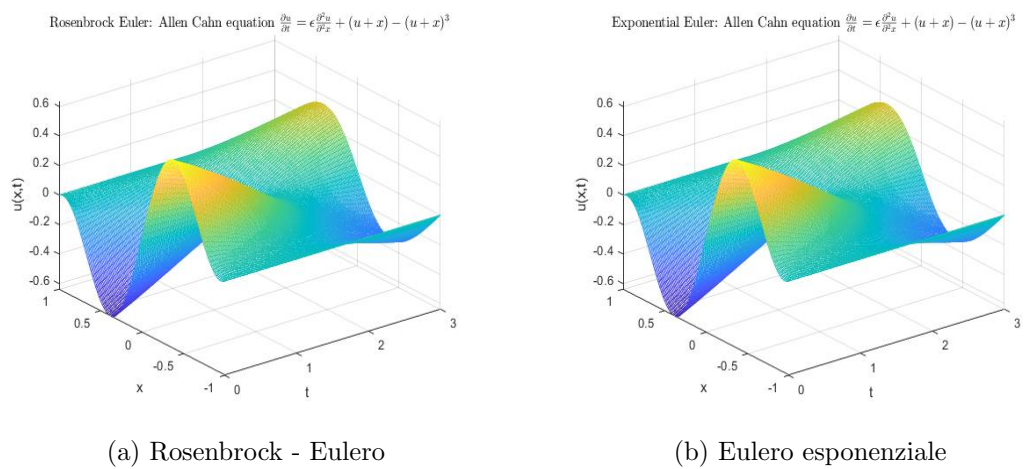


Figura 4.5: $N = 100, M = 200, \epsilon = 0.05$

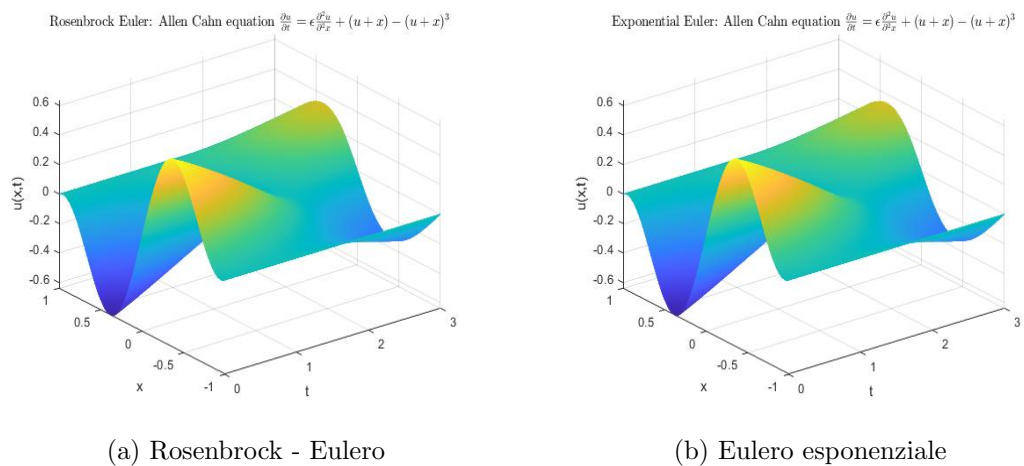
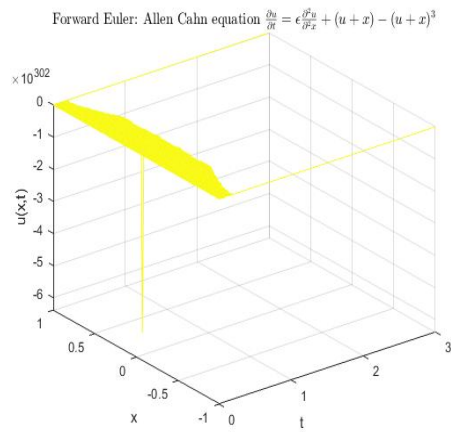
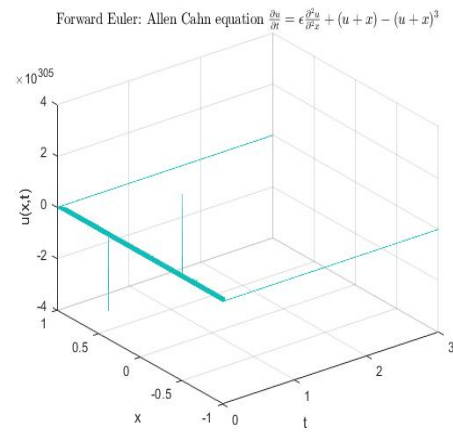


Figura 4.6: $N = 500, M = 600, \epsilon = 0.05$

Riportiamo infine i grafici per mostrare la non convergenza del metodo di Eulero esplicito con $\epsilon = 0.05$ sia nel caso $N = 100, M = 200$ sia in quello $N = 500, M = 600$:

(a) $N = 100, M = 200$ (b) $N = 100, M = 200$ Figura 4.7: Non convergenza del metodo di Eulero esplicito con $\epsilon = 0.05$

Conclusioni

Nel primo capitolo abbiamo studiato l'esponenziale di matrice, una funzione utile per risolvere sistemi di equazioni differenziali, in particolare abbiamo visto il metodo di *Scaling and Squaring* per approssimarlo. Successivamente, abbiamo applicato l'esponenziale di matrice a sistemi di equazioni differenziali e visto in generale alcuni metodi basilari per risolverli; abbiamo poi studiato un nuovo metodo numerico esplicito, stabile e con convergenza quadratica nel passo di discretizzazione per risolvere problemi differenziali. Infine, nell'ultimo capitolo abbiamo visto un esempio di problema alle derivate parziali, lo abbiamo discretizzato nel tempo e spazio, usato i metodi studiati in precedenza e notato le differenze, sia in termini di tempo computazionale sia in termini di convergenza. Esistono metodi numerici con ordine di convergenza ancora maggiore che sfruttano gli integratori esponenziali con $k \geq 2$, la cui implementazione sfrutta sempre l'esponenziale di matrice, ma il cui costo computazionale diventa sempre più elevato.

Bibliografia

- [1] [Higham, 2008] N.Higham, *Function of Matrices: Theory e Computation*, SIAM, Philadelphia, 2008
- [2] [Caliari, Ostermann, 2008] M. Caliari, A. Ostermann, *Implementation of exponential Rosenbrock-type integrators*, Applied Numerical Mathematics, 2008
- [3] [Quarteroni, Saleri, Gervasio, 2017] A. Quarteroni, F. Saleri, P. Gervasio, *Calcolo Scientifico*, Springer, 2017