Capitolo 10

Cenni sulla risoluzione di equazioni alle derivate parziali

In questo capitolo, vedremo brevemente come sia possibile utilizzare l'apparato metodologico, visto per le equazioni differenziali ordinarie, per il trattamento numerico di problemi evolutivi descritti da equazioni alle derivate parziali lineari. Per la precisione, esamineremo i problemi di tipo parabolico ed iperbolico, con un cenno alle equazioni di trasporto e diffusione, che si incontrano comunemente in molte applicazioni. In entrambi i casi, utilizzeremo la stessa tecnica di risoluzione, denominata *metodo delle linee*, che consiste nella semidiscretizzazione delle derivate spaziali del problema. Si ottiene, in questo modo, un sistema di equazioni differenziali ordinarie (lineari, nel nostro caso), caratterizzato dal fatto che, al tendere del parametro di discretizzazione spaziale a zero, tende ad avere dimensione infinita. Pertanto, nelle questioni di stabilità lineare, non si potranno utilizzare gli stessi argomenti visti per i sistemi a dimensione fissa. Per questo motivo, sarà necessario introdurre la nozione di spettro dell'operatore infinito limite, i cui autovalori determineranno le condizioni richieste per l'analisi di stabilità lineare. Per questo motivo, cominceremo trattando degli operatori infiniti.

10.1 Famiglie di matrici ed operatori di Toeplitz a banda

Una *matrice di Toeplitz* ha elementi costanti lungo le sue diagonali.[1] Particolari matrici di Toeplitz sono quelle *a banda*: una matrice di Toeplitz a banda di dimensione N avrà una struttura del tipo

$$T_N = \begin{pmatrix} a_0 & \dots & a_k & & \\ \vdots & \ddots & & \ddots & \\ a_{-m} & & \ddots & & a_k \\ & \ddots & & \ddots & \vdots \\ & & a_{-m} & \dots & a_0 \end{pmatrix}_{N\times N}, \tag{10.1}$$

in cui a_ℓ denota l'elemento (costante) sulla diagonale $\ell \equiv j - i$, dove (i, j) sono gli indici dell'elemento generico su questa diagonale. In particolare:

- se $\ell > 0$ si tratterà della ℓ-esima *sopradiagonale*;

[1] Abbiamo già incontrato le matrici di Toeplitz triangolari inferiori, nel Capitolo 4.

- se $\ell = 0$ si tratta della diagonale principale;
- se $\ell < 0$ si tratta della $(-\ell)$-esima *sottodiagonale* della matrice.

Pertanto, la matrice T_N avrà *ampiezza di banda superiore* pari a k ed *inferiore* pari ad m. È evidente che la *famiglia di matrici* $\{T_N\}$ è completamente caratterizzata dal seguente polinomio,

$$p(z) = \sum_{i=-m}^{k} a_i z^{i+m} \in \Pi_{m+k}, \tag{10.2}$$

definito dai coefficienti delle diagonali. Sarà altresì utile introdurre il corrispondente *simbolo*, che è la funzione razionale definita come:

$$g(z) = z^{-m} p(z) \equiv \sum_{i=-m}^{k} a_i z^i. \tag{10.3}$$

Ricordiamo, inoltre, la Definizione 4.5 di *tipo di un polinomio*: un polinomio $p(z) \in \Pi_k$ si dirà avere *tipo* (k_1, k_2, k_3) se esso ha:

- k_1 radici di modulo minore di 1;
- k_2 radici di modulo 1;
- k_3 radici di modulo maggiore di 1.

Chiaramente, dovrà aversi $k_1 + k_2 + k_3 = k$. Ricordiamo anche che i polinomi di tipo $(k, 0, 0)$ sono i polinomi di Schur, mentre quelli di tipo $(k_1, k_2, 0)$, con le radici di modulo 1 tutte semplici, sono i polinomi di Von Neumann. È possibile dimostrare il seguente risultato.

Teorema 10.1 *Se il polinomio (10.2) associato alla famiglia $\{T_N\}$ ha tipo $(m, 0, k)$, allora le matrici della famiglia sono invertibili (in generale, da un certo indice $\bar{N}$ in poi) e, inoltre, i loro numeri di condizionamento, $\{\kappa(T_N)\}$, sono uniformemente limitati. Ovvero, $\exists \gamma > 0$ tale che:*

$$\kappa(T_N) \leq \gamma, \qquad \forall N \geq \bar{N}. \tag{10.4}$$

Dimostrazione. Vedi [2, Teorema 3]. □

Osservazione 10.1 *Osservando che, facendo riferimento alle norme 1 o ∞,*

$$\|T_N\| = \sum_{i=-m}^{k} |a_i| \equiv K, \qquad \forall N \geq 1 + \max\{m, k\},$$

dalla (10.4) segue che, se il polinomio (10.2) ha tipo $(m, 0, k)$, allora

$$\|T_N^{-1}\| \leq \gamma K^{-1}, \qquad \forall N \geq \bar{N},$$

ovvero, le matrici della famiglia hanno inversa con norma uniformemente limitata rispetto alla loro dimensione.

Sotto le stesse condizioni, risulta essere *invertibile con inversa continua* l'operatore infinito

$$T = \lim_{N\to\infty} T_N. \tag{10.5}$$

Vediamo di rendere più precisa questa affermazione. La matrice infinita T può essere riguardata come un operatore lineare definito nello spazio di Banach delle successioni infinite,

$$\ell_1 = \left\{ \mathbf{x} = \begin{pmatrix} x_1 & x_2 & \dots \end{pmatrix}^T : \|\mathbf{x}\| < \infty \right\},$$

munito della norma

$$\|\mathbf{x}\| = \sum_{i=1}^{\infty} |x_i|.$$

Definizione 10.1 *Sia dato $T : \ell_1 \to \ell_1$, operatore lineare con norma limitata (ovvero, $\|T\| < \infty$). Esso di dirà* invertibile con inversa continua *se:*

1. $\forall \mathbf{y} \in \ell_1 \ \exists \mathbf{x} \in \ell_1 \ : \ T\mathbf{x} = \mathbf{y}$;
2. $\exists \mu > 0 \ : \ \|T\mathbf{x}\| \geq \mu \|\mathbf{x}\|, \ \forall \mathbf{x} \in \ell_1$.

In tal caso, T ammette un'inversa limitata:

$$\|T^{-1}\| \leq \mu^{-1}. \tag{10.6}$$

Osservazione 10.2 *La prima condizione richiesta nella precedente definizione garantisce, evidentemente, la* suriettività *dell'operatore. Similmente, la seconda condizione garantisce la sua* iniettività:[2] *infatti, se si avesse $T\mathbf{x}_1 = \mathbf{y} = T\mathbf{x}_2$, seguirebbe che*

$$0 = \|\mathbf{0}\| = \|T(\mathbf{x}_1 - \mathbf{x}_2)\| \geq \mu \|\mathbf{x}_1 - \mathbf{x}_2\| \qquad \Rightarrow \qquad \mathbf{x}_1 = \mathbf{x}_2.$$

Osserviamo che la proprietà (10.6) vale anche nel caso di matrici di dimensione finita, come si deduce dal seguente risultato.

Teorema 10.2 *Sia $T \in \mathbb{R}^{N\times N}$. Se $\exists \mu > 0$ tale che, per ogni vettore $\mathbf{x} \in \mathbb{R}^N$: $\|T\mathbf{x}\| \geq \mu \|\mathbf{x}\|$, allora T è nonsingolare e T^{-1} soddisfa la (10.6).*

Dimostrazione. La nonsingolarità di T segue dal fatto che, qualora si avesse $T\mathbf{x} = \mathbf{0}$, seguirebbe che

$$0 = \|T\mathbf{x}\| \geq \mu \|\mathbf{x}\| \qquad \Rightarrow \qquad \mathbf{x} = \mathbf{0}.$$

Pertanto, posto $\mathbf{x} = T^{-1}\mathbf{y}$, si ha che, per ogni $\mathbf{y} \neq \mathbf{0}$:

$$\|T\mathbf{x}\| = \|\mathbf{y}\| \geq \mu \|\mathbf{x}\| = \mu \|T^{-1}\mathbf{y}\| \qquad \Rightarrow \qquad \frac{\|T^{-1}\mathbf{y}\|}{\|\mathbf{y}\|} \leq \mu^{-1},$$

da cui discende, evidentemente, la (10.6). □

Ritornando all'operatore infinito (10.5), vale il seguente risultato:

Teorema 10.3 *Sia T l'operatore di Toeplitz a banda definito dalle (10.1)-(10.5). Allora T è invertibile con inversa continua se e solo se il polinomio (10.2) associato è di tipo $(m, 0, k)$.*

[2] Nel caso di matrici di dimensione finita, queste due proprietà sono infatti equivalenti alla nonsingolarità della matrice.

Dimostrazione. Vedi [9, Teorema 3.3.3]. □

Osservazione 10.3 *Pertanto, T sarà invertibile con inversa continua se e solo se il polinomio $p(z)$ associato ha un numero di radici nel cerchio aperto unitario pari al numero m delle sottodiagonali, e le rimanenti k radici al di fuori del cerchio unitario chiuso. Queste ultime saranno in numero pari a quello delle sopradiagonali di T. Analoga proprietà si applica alle matrici della corrispondente famiglia $\{T_N\}$.*

Data la nozione di invertibilità con inversa continua, possiamo definire lo *spettro* dell'operatore T come

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ non è invertibile con inversa continua}\}. \tag{10.7}$$

In virtù del Teorema 10.3, tenendo conto del fatto che il polinomio associato all'operatore $T - \lambda I$ risulta essere (vedi (10.2))

$$p_\lambda(z) = p(z) - \lambda z^m,$$

si ottiene, quindi, la seguente caratterizzazione:

$$\sigma(T) = \{\lambda \in \mathbb{C} : p_\lambda(z) \text{ non è di tipo } (m, 0, k)\}.$$

Ricordando che le radici di un polinomio sono funzioni analitiche dei suoi coefficienti e che, inoltre, il tipo del polinomio $p_\lambda(z)$ tende a $(m, 0, k)$, per $\lambda \to \infty$, è possibile dimostrare le seguenti proprietà:

1. $\sigma(T)$ è un insieme chiuso e limitato;
2. $\sigma(T)$ è un insieme connesso;
3. se $p(z)$ ha coefficienti reali, $\sigma(T)$ è simmetrico rispetto all'asse reale;
4. se T è Hermitiana, $\sigma(T)$ è un segmento dell'asse reale;
5. T è invertibile con inversa continua $\Leftrightarrow$ $0 \notin \sigma(T)$.

Per determinare "operativamente" $\sigma(T)$, si utilizzano argomenti simili a quelli visti per determinare la regione di assoluta stabilità di un metodo LMF (si veda la Sezione 4.5): si definisce il *boundary locus* associato all'operatore T, come il luogo dei punti λ del piano complesso in cui $p_\lambda(z)$ ha una radice di molulo 1, ovvero

$$z = e^{i\theta}, \qquad \theta \in [0, 2\pi].$$

Pertanto, si otterrà:

$$\sigma_+(T) = \left\{\lambda \in \mathbb{C} : p_\lambda(e^{i\theta}) = 0,\ \theta \in [0, 2\pi]\right\}.$$

Tuttavia, ricordando la definizione (10.3) del simbolo di T, si vede facilmente che il *boundary locus* dell'operatore T può essere, più convenientemente, definito come:

$$\sigma_+(T) = \left\{\lambda \in \mathbb{C} : \lambda = g(e^{i\theta}),\ \theta \in [0, 2\pi]\right\}.$$

Essendo le radici di $p_\lambda(z)$ funzioni analitiche di λ, $\sigma_+(T)$ dividerà il piano complesso in regioni connesse all'interno delle quali il tipo del polinomio è costante. Controllando il tipo

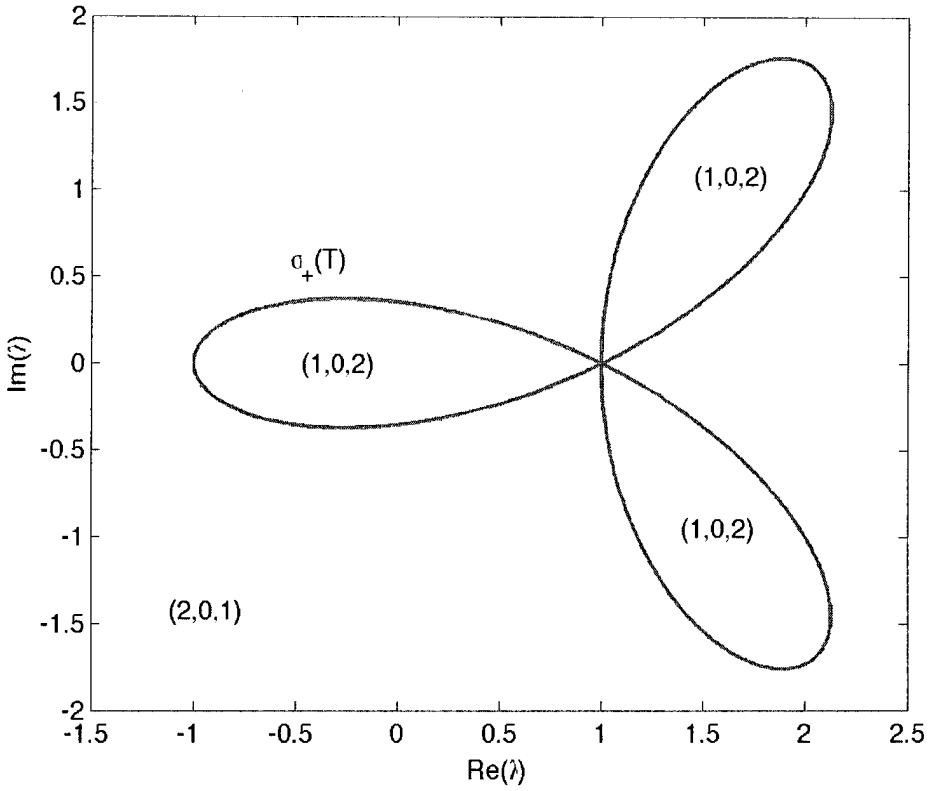

Figura 10.1: *Boundary locus* dell'operatore (10.8).

del polinomio in un qualunque punto in ciascuna di queste regioni, conduce ad una rapida individuazione di $\sigma(T)$, la cui frontiera è contenuta in $\sigma_+(T)$.

Si consideri, ad esempio, l'operatore:

$$T = \begin{pmatrix} 1 & 1 & & \\ 0 & \ddots & \ddots & \\ -1 & \ddots & \ddots & \ddots \\ & \ddots & & \end{pmatrix}. \tag{10.8}$$

Il suo *boundary locus* è graficato in Figura 10.20, dove è anche riportato il tipo del polinomio $p_\lambda(z)$ nelle varie regioni in cui esso divide il piano complesso. Le tre regioni limitate, inclusa la frontiera, costituiscono lo spettro $\sigma(T)$ dell'operatore. Al di fuori di questo, il tipo del polinomio $p_\lambda(z)$ è (2,0,1) e, pertanto, $T - \lambda I$ sarà sempre invertibile con inversa continua. In particolare, T non è invertibile con inversa continua, poiché $0 \in \sigma(T)$: infatti, il polinomio associato all'operatore è

$$p(z) \equiv p_0(z) = z^3 + z^2 - 1,$$

il cui tipo è $(1, 0, 2)$.

Osservazione 10.4 *Si può dimostrare che lo spettro $\sigma(T)$ dell'operatore (10.5) coincide, essenzialmente, con il cosiddetto* spettro della famiglia $\{T_N\}$, *costituito dall'unione degli*

autovalori delle matrici della famiglia (che è un insieme al più numerabile), e dei cosiddetti quasi-autovalori *della famiglia di matrici (si vedano, a riguardo, [10] e [9, Cap. 3]).*

10.2 L'equazione del calore

Fissata la *striscia* infinita

$$\Omega = [0,1] \times [0,\infty), \tag{10.9}$$

che verrà utilizzata per tutti i problemi trattati in questo capitolo, consideriamo un classico esempio di equazione *parabolica*, e cioé l'*equazione del calore*,

$$\begin{aligned} u_t(x,t) &= u_{xx}(x,t), && (x,t) \in \Omega, && (10.10)\\ u(x,0) &= \omega_0(x), && x \in [0,1], && (10.11)\\ u(0,t) &= u(1,t) = 0, && t \geq 0. && (10.12) \end{aligned}$$

Essa descrive la *diffusione* del calore in una sbarra conduttrice di lunghezza unitaria che ha una distribuzione di calore iniziale data da $\omega_0(x)$, ed è a temperatura nulla agli estremi.[3] Pertanto, le (10.11) e (10.12) saranno, rispettivamente, la *condizione iniziale* e le *condizioni al bordo* per l'equazione (10.10). Condizioni al bordo, come le (10.12), in cui si assegna il valore della soluzione, sono dette *condizioni di Dirichlet*, (*omogenee* nello specifico caso). È tuttavia possibile utilizzare diverse tipologie di condizioni al bordo: in Sezione 10.2.1 analizzeremo il caso di condizioni al bordo di Dirichlet *non omogenee*, mentre in Sezione 10.2.3 analizzeremo le condizioni di *Neumann* che, assieme alle condizioni di Dirichlet, sono tra le più utilizzate nella pratica. In ogni caso, assumeremo che le funzioni, che definiscono le condizioni iniziali ed al bordo, soddisfino appropriate condizioni di consistenza (e.g., $\omega_0(0) = \omega_0(1) = 0$, nel caso delle (10.11)-(10.12)).

Fissato, dunque, un passo di discetizzazione

$$\Delta x = \frac{1}{N} \tag{10.13}$$

per la variabile spaziale, è possibile discretizzare il secondo membro della (10.10), sulla *mesh*

$$x_i = i\Delta x \equiv \frac{i}{N}, \qquad i = 0, \ldots, N, \tag{10.14}$$

definendo le funzioni

$$u_i(t) \approx u(x_i,t), \qquad i = 0, \ldots, N, \tag{10.15}$$

ed i vettori

$$\mathbf{u}(t) = \begin{pmatrix} u_1(t) \\ \vdots \\ u_{N-1}(t) \end{pmatrix}, \qquad \mathbf{u}_0 = \begin{pmatrix} \omega_0(x_1) \\ \vdots \\ \omega_0(x_{N-1}) \end{pmatrix}, \tag{10.16}$$

trasformando (10.10)–(10.12) nel seguente problema ai valori iniziali per equazioni differenziali ordinarie:

$$\mathbf{u}'(t) = \frac{1}{\Delta x^2} T_N \mathbf{u}(t), \qquad t \geq 0, \qquad \mathbf{u}(0) = \mathbf{u}_0, \tag{10.17}$$

[3] Chiaramente, un qualunque coefficiente (positivo) della derivata spaziale in (10.10) potrà essere normalizzato a 1, mediante una trasformazione temporale.

dove

$$T_N = \begin{pmatrix} -2 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{pmatrix} \in \mathbb{R}^{N-1 \times N-1}. \tag{10.18}$$

Infatti, questo equivale ad utilizzare la seguente approssimazione (del secondo ordine) della derivata seconda:

$$\frac{\partial^2}{\partial x^2} u(x_i, t) = \frac{u(x_{i-1}, t) - 2u(x_i, t) + u(x_{i+1}, t)}{\Delta x^2} + O(\Delta x^2), \quad i = 1, \ldots, N-1. \tag{10.19}$$

Evidentemente, quando il parametro di discretizzazione Δx tende a 0, la dimensione del sistema (vedi (10.13)) tende ad infinito. Pertanto, ai fini dell'analisi di stabilità lineare, cosidereremo lo spettro dell'operatore infinito,

$$T = \lim_{N \to \infty} T_N,$$

che, utilizzando i risultati esposti in Sezione 10.1, si vede facilmente essere dato da

$$\sigma(T) = [-4, 0]. \tag{10.20}$$

Essendo la soluzione del problema continuo asintoticamente stabile (per $t \to \infty$, infatti, la temperatura della sbarra tenderà a quella degli estremi, che è nulla), richiederemo la stessa cosa all'approssimazione numerica ottenuta applicando un metodo alla differenze applicato al problema semidiscreto (10.17). Fissata la discretizzazione temporale

$$t_j = j\Delta t, \qquad j = 0, 1, \ldots, \tag{10.21}$$

utilizzando il metodo di Eulero esplicito per risolvere (10.17), si ottiene lo schema iterativo

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \frac{\Delta t}{\Delta x^2} T_N \mathbf{u}_k, \qquad k = 0, 1, \ldots,$$

dove, al solito, $\mathbf{u}_k \approx \mathbf{u}(t_k)$. Affinché la soluzione discreta sia stabile, qualunque sia la dimensione N utilizzata, dovrà aversi, considerando la (10.20),

$$4\frac{\Delta t}{\Delta x^2} \leq 2 \qquad \Rightarrow \qquad \Delta t \leq \frac{1}{2}\Delta x^2.$$

Chiaramente, questa condizione è assai restrittiva, perché impone una limitazione molto severa sul passo temporale, per $\Delta x \to 0$. Sarà, pertanto, conveniente utilizzare un metodo implicito. In particolare, se usiamo il metodo dei trapezi, si ottiene uno schema del secondo ordine anche nel tempo, noto come *metodo di Crank-Nicolson*. Ponendo

$$\alpha = \frac{1}{2}\frac{\Delta t}{\Delta x^2},$$

il metodo diviene:

$$(I_{N-1} - \alpha T_N)\mathbf{u}_{k+1} = (I_{N-1} + \alpha T_N)\mathbf{u}_k, \qquad k = 0, 1, \ldots. \tag{10.22}$$

In questo caso, essendo il metodo dei trapezi A-stabile, non vi sono restrizioni sul passo temporale: si parla, in questo ambito, di *stabilità incondizionata* per il metodo. Tuttavia,

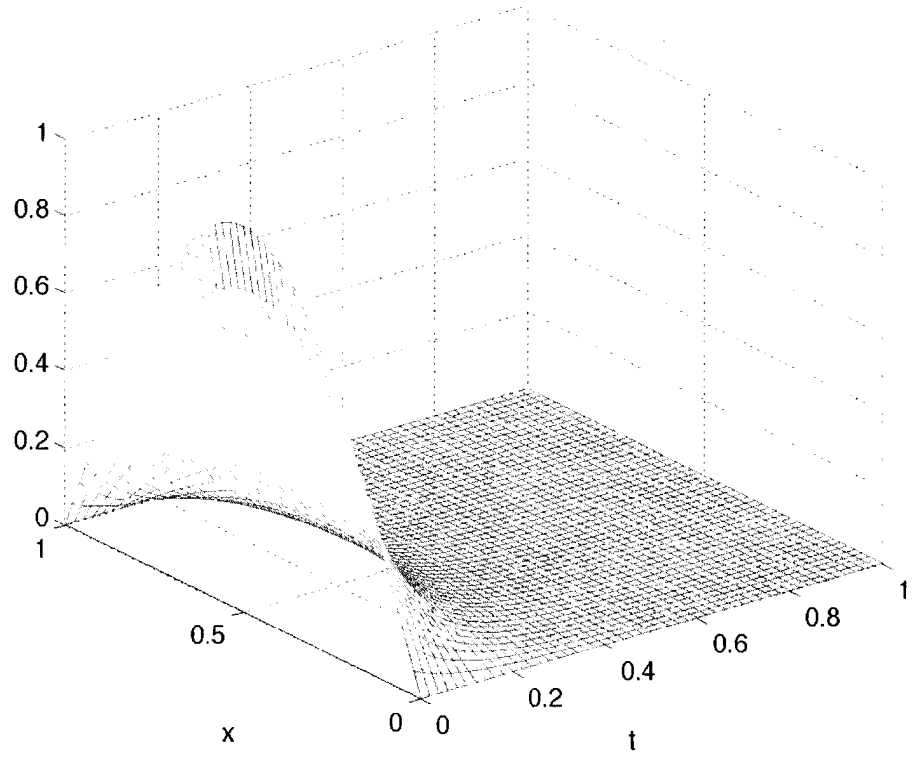

Figura 10.2: Metodo di Crank-Nicolson per approssimare la soluzione del problema (10.10) (10.12), $\omega_0(x) = \sin(\pi x)$, con passi $\Delta x = \Delta t = 1/50$.

per mere questioni di accuratezza, sarà conveniente utilizzare $\Delta t = \Delta x$, in modo da avere la stessa accuratezza sia nello spazio che nel tempo. In Figura 10.2 riportiamo il risultato dell'applicazione del metodo dei trapezi per risolvere (10.10) (10.12), con $\omega_0(x) = \sin(\pi x)$, utilizzando come passi di discretizzazione

$$\Delta x = \Delta t = \frac{1}{50}.$$

Osserviamo che, per ottenere un risultato qualitativamente simile, il metodo di Eulero esplicito avrebbe richiesto l'utilizzo di un passo

$$\Delta t = \frac{1}{2\Delta x^2} = \frac{1}{5000}.$$

10.2.1 Il caso di condizioni al bordo non omogenee

L'analisi di stabilità fatta per i metodi, applicati alla semidiscretizzazione del problema (10.10) (10.12), può essere estesa al caso di condizioni al bordo più generali. Consideriamo, quindi, il problema (vedi (10.9)):

$$u_t(x,t) = u_{xx}(x,t), \qquad (x,t) \in \Omega, \tag{10.23}$$

$$u(x,0) = \omega_0(x), \qquad x \in [0,1], \tag{10.24}$$
$$u(0,t) = \phi_0(t), \qquad u(1,t) = \phi_1(t), \qquad t \geq 0. \tag{10.25}$$

Se ne perturbiamo la condizione iniziale (10.24),

$$u(x,0) = \tilde{\omega}_0(x), \qquad x \in [0,1], \tag{10.26}$$

otterremo un nuovo problema, la cui soluzione sarà $\tilde{u}(x,t)$. Ponendo

$$z(x,t) = \tilde{u}(x,t) - u(x,t),$$

la differenza tra le due, si ottiene che z soddisfa il problema con condizioni al bordo omogenee (10.10)–(10.12), e condizione iniziale data da

$$z(x,0) = \tilde{\omega}_0(x) - \omega_0(x), \qquad x \in [0,1].$$

Questo costituisce il *problema variazionale* associato alla (10.23)–(10.25) che, pertanto, sarà della forma (10.10) (10.12).

La semidiscretizzazione del problema (10.23) (10.25), utilizzando la notazione (10.13) (10.16) introdotta in precedenza, porta al problema semidiscreto

$$\mathbf{u}'(t) = \frac{1}{\Delta x^2}\left[T_N \mathbf{u}(t) + \boldsymbol{\phi}(t)\right], \qquad t \geq 0, \qquad \mathbf{u}(0) = \mathbf{u}_0, \tag{10.27}$$

in cui T_N è definita come in (10.18) e

$$\boldsymbol{\phi}(t) = \begin{pmatrix} \phi_0(t) \\ 0 \\ \vdots \\ 0 \\ \phi_1(t) \end{pmatrix}.$$

Applicando un metodo numerico tra quelli esaminati in precedenza al problema (10.27), ottenendo quindi una soluzione approssimata $\{\mathbf{u}_k\}$, ed al problema ottenuto da questo considerando una condizione iniziale perturbata $\tilde{\mathbf{u}}_0$ (derivata da (10.26)), avendo quindi una nuova soluzione approssimata (perturbata rispetto alla prima) $\{\tilde{\mathbf{u}}_k\}$, e definendo la differenza

$$\mathbf{z}_k = \tilde{\mathbf{u}}_k - \mathbf{u}_k, \qquad k = 0, 1, \ldots,$$

si verifica facilmente che $\{\mathbf{z}_k\}$ altri non è che la soluzione ottenuta dallo stesso metodo se applicato al problema variazionale. Pertanto, l'analisi di stabilità per i metodi fatta in precedenza, fornisce le eventuali restrizioni sul passo temporale Δt, che garantiscono l'asintotica stabilità della $\{\mathbf{z}_k\}$, ovvero l'asintotica stabilità della soluzione di riferimento $\{\mathbf{u}_k\}$ (rispetto a perturbazioni della condizione iniziale).

10.2.2 Complessità computazionale del metodo di Crank-Nicolson

Analizziamo il costo computazionale dell'implementazione del metodo di Crank-Nicolson (10.22), tenendo conto della struttura della matrice (10.18). Si tratta, in effetti, di risolvere un sistema di equazioni *tridiagonale*, della forma

$$T(\alpha)\mathbf{u}_{k+1} = \mathbf{b},$$

con, in generale,

$$\mathbf{u}_k = \begin{pmatrix} u_1^k \\ \vdots \\ u_{N-1}^k \end{pmatrix}, \qquad u_i^k \approx u_i(t_k) \approx u(x_i, t_k),$$

il termine noto

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_{N-1} \end{pmatrix}, \qquad b_i = (1-2\alpha)u_i^k + \alpha(u_{i+1}^k + u_{i-1}^k), \qquad i = 1, \ldots, N-1, \tag{10.28}$$

in cui u_0^k e u_N^k (e, quindi, anche u_0^{k+1} e u_N^{k+1}) sono forniti dalle condizioni al bordo, e la matrice dei coefficienti

$$T(\alpha) = \begin{pmatrix} 1+2\alpha & -\alpha & & \\ -\alpha & \ddots & \ddots & \\ & \ddots & \ddots & -\alpha \\ & & -\alpha & 1+2\alpha \end{pmatrix} \in \mathbb{R}^{N-1 \times N-1}. \tag{10.29}$$

Evidentemente, la costruzione del termine noto ha complessità lineare, sia in termini di operazioni algebriche ($\approx 4N$) che di memoria (un vettore di lunghezza $N-1$). Riguardo alla matrice $T(\alpha)$, si verifica facilmente che

$$T(\alpha) = LD^{-1}L^T, \tag{10.30}$$

con

$$L = \begin{pmatrix} d_1 & & \\ -\alpha & \ddots & \\ & \ddots & \ddots \\ & & -\alpha \quad d_{N-1} \end{pmatrix}, \qquad D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_{N-1} \end{pmatrix}, \tag{10.31}$$

e gli elementi diagonali di D dati da:

$$d_1 = 1 + 2\alpha, \qquad d_i = (1+2\alpha) - \frac{\alpha^2}{d_{i-1}}, \qquad i = 2, \ldots, N-1. \tag{10.32}$$

Osserviamo che la fattorizzazione (10.30) necessita, per essere memorizzata, solo di N posizioni di memoria: $N-1$ per gli elementi diagonali di D e lo scalare α. Anche il numero di operazioni per ottenere la fattorizzazione (10.32) è lineare in N: $\approx 2N$ operazioni algebriche elementari. Il sistema $T(\alpha)\mathbf{u} = \mathbf{b}$ si risolve, pertanto, mediante il seguente algoritmo, in cui il vettore $\mathbf{u} = (u_1, \ldots, u_{N-1})^T$ contiene, in ingresso, il termine noto $\mathbf{b}$, ed è riscritto con la soluzione del sistema lineare:

$$\begin{aligned} u_1 = u_1/d_1, \qquad u_i &= (u_i + \alpha u_{i-1})/d_i, \qquad i = 2, \ldots, N-1, \\ u_i &= d_i u_i, \qquad i = 1, \ldots, N-2, \\ u_{N-i} &= (u_{N-i} + \alpha u_{N-i+1})/d_{N-i}, \qquad i = 2, \ldots, N-1, \end{aligned}$$

per un totale di $\approx 7N$ operazioni algebriche elementari.

10.2.3 Condizioni di Neumann

Cosideriamo adesso il caso del problema (vedi (10.9)):

$$\begin{aligned} u_t(x,t) &= u_{xx}(x,t), && (x,t) \in \Omega, && (10.33)\\ u(x,0) &= \omega_0(x), && x \in [0,1], && (10.34)\\ u_x(0,t) &= u_x(1,t) = 0, && t \geq 0, && (10.35) \end{aligned}$$

in cui le condizioni al bordo sono determinate assegnando la *derivata normale* della soluzione: si parla, in questo caso, di *condizioni di Neumann* (*omogenee*, nel caso della (10.35)). Utilizzando nuovamente la discretizzazione (10.13) (10.15), e l'approssimazione (10.19) per la derivata seconda, si perviene ad un problema semidiscreto formalmente ancora dato da (10.16)-(10.17). In tal caso, però, la matrice T_N ha una struttura leggermente diversa, generata dalle approssimazioni (del primo ordine nella variabile spaziale x):

$$u(0,t) \approx u(x_1,t) - \Delta x \overbrace{u_x(0,t)}^{=0}, \qquad u(1,t) \approx u(x_{N-1},t) + \Delta x \overbrace{u_x(1,t)}^{=0},$$

da cui si deduce

$$u_0^k = u_1^k, \qquad u_N^k = u_{N-1}^k, \qquad k = 0,1,\ldots,$$

che portano alla seguente famiglia di matrici,

$$T_N = \begin{pmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & 1 & \ddots & \ddots & & \\ & & \ddots & \ddots & 1 & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{N-1 \times N-1},$$

invece della (10.18). Considerazioni analoghe a quelle esaminate per il caso di Dirichlet valgono nel caso delle condizioni di Neumann, riguardo alle questioni di stabilità e sui metodi utilizzabili per risolvere (10.33) (10.35), nonché sulla loro complessità computazionale. Chiaramente, argomenti analoghi a quelli esaminati in Sezione 10.2.1 valgono nel caso in cui le condizioni al bordo (10.35) non siano omogenee.

10.3 L'equazione delle onde

Considereremo solo la forma del primo ordine di tale problema *iperbolico* dato da (vedi (10.9)):

$$\begin{aligned} u_t(x,t) &= -v u_x(x,t), && (x,t) \in [0,1] \times \Omega, && (10.36)\\ u(x,0) &= \omega_0(x), && x \in [0,1], && (10.37)\\ u(0,t) &= 0, && t \geq 0, && (10.38) \end{aligned}$$

in cui la *velocità*

$$v > 0 \qquad (10.39)$$

è assegnata. La soluzione di questo problema si verifica facilmente essere data da:

$$u(x,t) = \begin{cases} \omega_0(x - vt), & \text{se} \quad x \leq vt, \\ 0, & \text{se} \quad x < vt, \end{cases} \qquad (x,t) \in \Omega. \tag{10.40}$$

Pertanto, l'*onda* con profilo iniziale $\omega_0(x)$ si sposterà verso *destra*, per $t > 0$, con velocità (10.39) e, per $t > v^{-1}$, la soluzione sarà nulla, poiché l'onda "fuoriesce" dalla striscia Ω (vedi (10.9)). Fissata una partizione dell'intervallo spaziale analoga a quella vista per l'equazione del calore (vedi (10.13)-(10.14)), e continuando ad utilizzare la notazione (10.15), è possibile approssimare la derivata spaziale con delle *differenze all'indietro*, in considerazione che il dato al bordo (10.38) è posto in $x = 0$:

$$\frac{\partial}{\partial x} u(x_i,t) = \frac{u(x_i,t) - u(x_{i-1},t)}{\Delta x} + O(\Delta x).$$

Questo porta a definire il seguente sistema lineare di equazioni differenziali ordinarie:

$$\mathbf{u}'(t) = -\frac{v}{\Delta x} L_N \mathbf{u}(t), \qquad t \geq 0, \tag{10.41}$$

con

$$\mathbf{u}(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_N(t) \end{pmatrix}, \qquad L_N = \begin{pmatrix} 1 & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{N \times N}, \tag{10.42}$$

e condizione iniziale

$$\mathbf{u}(0) = \begin{pmatrix} \omega_0(x_1) \\ \vdots \\ \omega_0(x_N) \end{pmatrix}.$$

Applicando il metodo di Eulero esplicito per la sua approssimazione, e avendo posto

$$\alpha = v \frac{\Delta t}{\Delta x}, \tag{10.43}$$

si ottiene:

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \alpha L_N \mathbf{u}_k \equiv (I_N - \alpha L_N)\, \mathbf{u}_k, \qquad k = 0, 1, \ldots. \tag{10.44}$$

In questo caso, per quanto visto in Sezione 10.1, lo spettro dell'operatore

$$-L = \lim_{N \to \infty} -L_N$$

è dato dal cerchio unitario di centro -1 del piano complesso. Richiedendo che questo sia contenuto nella chiusura della regione di assoluta stabilità del metodo di Eulero esplicito, si ottiene che

$$\alpha \leq 1 \qquad \Rightarrow \qquad v\Delta t \leq \Delta x. \tag{10.45}$$

Questa restrizione non è molto penalizzante e, in effetti, la scelta

$$v\Delta t = \Delta x \tag{10.46}$$

è assai raccomandabile: infatti, con questa scelta, il metodo di Eulero esplicito fornisce la soluzione *esatta* del problema, proiettata ai nodi della *mesh* discreta.

Osservazione 10.5 *Se avessimo usato gli autovalori di L_N, invece che quelli dell'operatore limite L, avremmo ottenuto una limitazione sul passo temporale data da $v\Delta t \leq 2\Delta x$, invece della (10.45), che è errata. La condizione (10.45) è anche nota come condizione di* Courant-Friedrichs-Lewy (CFL).

In un modo del tutto simile, per il problema:

$$\begin{aligned} u_t(x,t) &= v u_x(x,t), && (x,t) \in \Omega, && (10.47)\\ u(x,0) &= \omega_0(x), && x \in [0,1], && (10.48)\\ u(1,t) &= 0, && t \geq 0, && (10.49) \end{aligned}$$

in cui v soddisfa la (10.39), la soluzione si verifica essere data da:

$$u(x,t) = \begin{cases} \omega_0(x+vt), & \text{se} \quad x \leq vt, \\ 0, & \text{se} \quad x < vt, \end{cases} \qquad (x,t) \in \Omega. \tag{10.50}$$

Pertanto, l'*onda* con profilo iniziale $\omega_0(x)$ si sposterà verso *sinistra*, per $t > 0$, con velocità v e, per $t > v^{-1}$, la soluzione sarà nulla, poiché l'onda "fuoriesce" da Ω. Fissata una partizione dell'intervallo spaziale analoga a quella vista precedentemente, è possibile approssimare la derivata spaziale con delle *differenze in avanti*, in considerazione che il dato al bordo (10.38) è posto in $x = 1$:

$$\frac{\partial}{\partial x} u(x_i,t) = \frac{u(x_{i+1},t) - u(x_i,t)}{\Delta x} + O(\Delta x).$$

Questo porta a definire il seguente sistema lineare di equazioni differenziali ordinarie:

$$\mathbf{u}'(t) = \frac{v}{\Delta x} U_N \mathbf{u}(t), \qquad t \geq 0, \tag{10.51}$$

con

$$\mathbf{u}(t) = \begin{pmatrix} u_0(t) \\ u_1(t) \\ \vdots \\ u_{N-1}(t) \end{pmatrix}, \qquad U_N = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & -1 \end{pmatrix} \in \mathbb{R}^{N \times N}, \tag{10.52}$$

e condizione iniziale

$$\mathbf{u}(0) = \begin{pmatrix} \omega_0(x_0) \\ \vdots \\ \omega_0(x_{N-1}) \end{pmatrix}.$$

Applicando il metodo di Eulero esplicito per la sua approssimazione, e utilizzando la notazione (10.43), si ottiene:

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \alpha U_N \mathbf{u}_k \equiv (I_N + \alpha U_N)\, \mathbf{u}_k, \qquad k = 0, 1, \ldots. \tag{10.53}$$

Anche in questo caso, lo spettro dell'operatore

$$U = \lim_{N \to \infty} U_N$$

è dato dal cerchio unitario di centro -1 del piano complesso. Pertanto, si perviene alla stessa condizione (10.45) vista innanzi. Anche in questo caso, utilizzando un passo temporale tale

che valga la (10.46), la soluzione discreta risulterà coincidere con quella esatta proiettata ai nodi della *mesh*.

Dal punto di vista del costo computazionale, trattandosi di un metodo esplicito, si ha complessità lineare, sia per quanto riguarda il numero di operazioni per passo temporale, che per quanto riguarda l'occupazione di memoria. Di seguito si riportano gli algoritmi per (10.44) ed (10.53), rispettivamente, dove si è posto

$$u_i^k \approx u(x_i, t_k), \qquad u_i^0 = \omega_0(x_i), \quad i = 0, \ldots, N, \tag{10.54}$$

ed α è definito dalla (10.43). Per (10.44) si ottiene:

$$u_0^{k+1} = 0, \qquad u_i^{k+1} = (1-\alpha)u_i^k + \alpha u_{i-1}^k, \qquad i = 1, \ldots, N. \tag{10.55}$$

Similmente, per (10.53) si ottiene:

$$u_N^{k+1} = 0, \qquad u_{N-i}^{k+1} = (1-\alpha)u_{N-i}^k + \alpha u_{N-i+1}^k, \qquad i = 1, \ldots, N. \tag{10.56}$$

In Figura 10.3 si riporta il risultato dell'applicazione del metodo (10.44) al problema (10.36)–(10.38), con $v = 1$ e passi temporale e spaziale uguali a 1/50. Similmente, in Figura 10.4 si riporta il risultato dell'applicazione del metodo (10.53) al problema (10.47) (10.49), con $v = 1$ e utilizzando gli stessi passi temporale e spaziale. Per entrambi i casi si è scelto $\omega_0(x) = \sin(\pi x)$.

Osservazione 10.6 *Chiaramente, i metodi (10.44) ed (10.53) hanno ordine 1, sia nello spazio che nel tempo.*

Definizione 10.2 *La discretizzazione con le differenze all'indietro o in avanti, in base al segno della derivata spaziale, è denominata* discretizzazione "up-wind".

10.3.1 Il caso di condizioni al bordo non omogenee

Argomenti del tutto analoghi a quelli visti per l'equazione del calore, valgono per l'equazione delle onde, in caso di condizione al bordo di Dirichlet non omogenea. Esaminiamo il caso corrispondente al problema (10.36)–(10.38) ed analogamente si ragionerà per il problema, con condizione al bordo non omogenea, corrispondente a (10.47)-(10.49). Supponiamo, quindi, che la (10.38) sia sostituita da:

$$u(0,t) = \phi_0(t), \qquad t \geq 0. \tag{10.57}$$

In tal caso la soluzione di vede essere sempre data formalmente dalla (10.40), con la convenzione che

$$u(x,t) = \begin{cases} \omega_0(x - vt), & \text{se} \quad x \leq vt, \\ \phi_0(t - v^{-1}x), & \text{se} \quad x < vt, \end{cases} \qquad (x,t) \in \Omega.$$

Tuttavia, considerando il problema variazionale relativo a perturbazioni del dato iniziale, si vede che questo è della forma (10.36) (10.38) e, pertanto, le condizioni sul passo temporale Δt, applicando il metodo di Eulero esplicito al problema semidiscreto, risultano essere le stesse ottenute in (10.45).[4]

Il problema semidiscreto si verifica facilmente essere dato da:

$$\mathbf{u}'(t) = -\frac{1}{\Delta x}\left[L_N \mathbf{u}(t) - \boldsymbol{\phi}(t)\right], \qquad t \geq 0, \tag{10.58}$$

[4]Questa conclusione si ottiene mediante argomenti del tutto analoghi a quelli visti in Sezione 10.2.1 per l'equazione del calore.

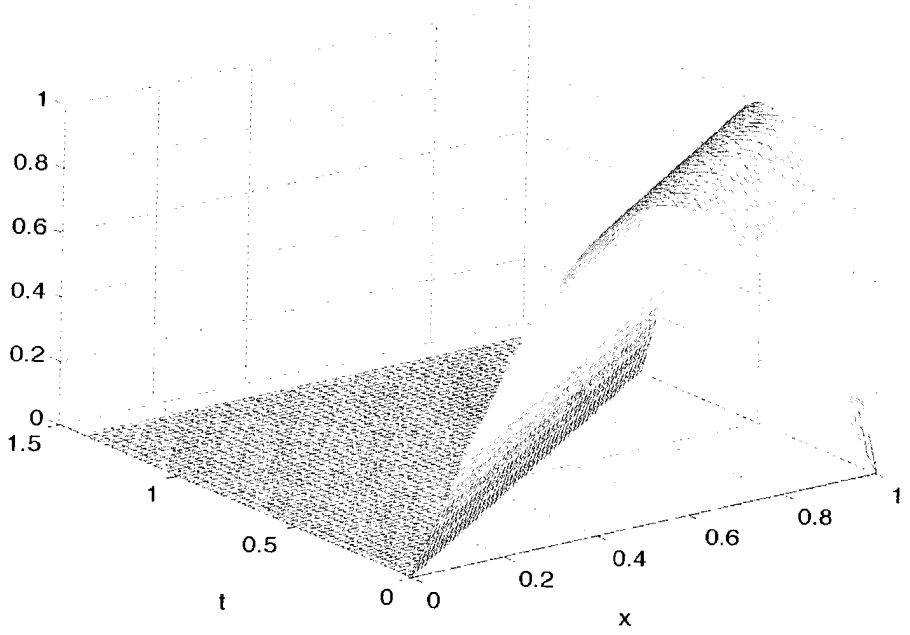

Figura 10.3: Metodo di Eulero esplicito per approssimare la soluzione del problema (10.36)-(10.38), $v = 1$, $\omega_0(x) = \sin(\pi x)$, con passi $\Delta x = \Delta t = 1/50$.

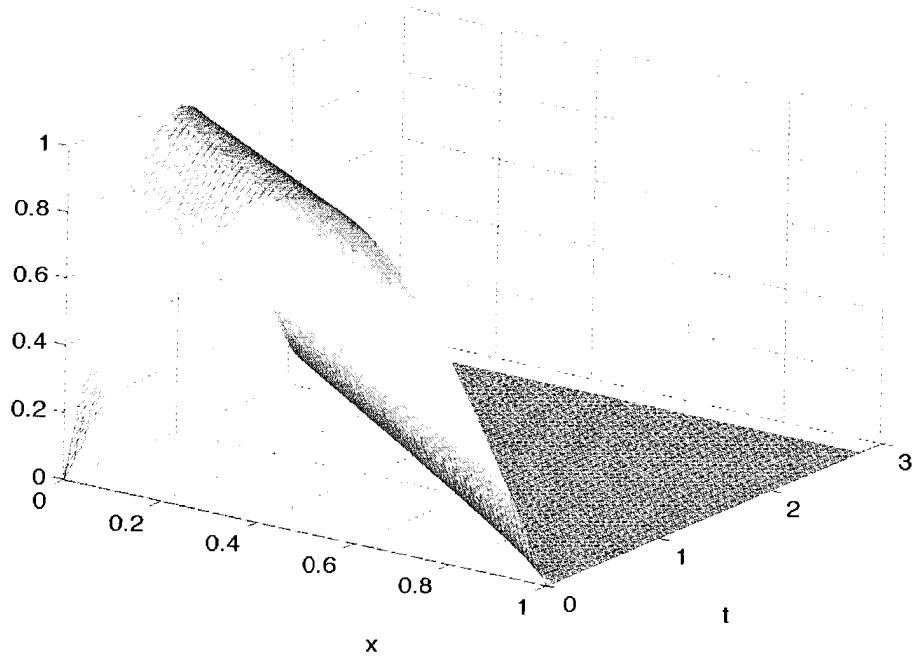

Figura 10.4: Metodo di Eulero esplicito per approssimare la soluzione del problema (10.47) (10.49), $v = 1$, $\omega_0(x) = \sin(\pi x)$, con passi $\Delta x = \Delta t = 1/50$.

con

$$\boldsymbol{\phi}(t) = \begin{pmatrix} \phi_0(t) & 0 & \dots & 0 \end{pmatrix}^T \in \mathbb{R}^{N-1}.$$

Conseguentemente, applicando il metodo di Eulero esplicito, si ottiene lo schema iterativo:

$$\mathbf{u}_{k+1} = (I_N - \alpha L_N)\,\mathbf{u}_k + \alpha\boldsymbol{\phi}_k, \qquad k = 0, 1, \dots, \tag{10.59}$$

dove si è posto $\boldsymbol{\phi}_k = \boldsymbol{\phi}(t_k)$ e α è definito da (10.43). In termini di componenti, l'algoritmo risultante coincide con (10.55), tenendo conto che, in questo caso, $u_0^k = \phi_0(t_k)$ e $u_0^{k+1} = \phi_0(t_{k+1})$.

Argomenti del tutto simili si applicano quando le condizioni al bordo sono di tipo Neumann, ovvero quando la (10.38) diviene

$$u_x(0,t) = \phi_0(t), \qquad t \geq 0.$$

In questo caso, considerando che, al primo ordine, si ha:

$$\frac{u_1(t) - u_0(t)}{\Delta x} \approx u_x(0,t) \equiv \phi_0(t) = -v^{-1}u_t(0,t),$$

dove l'ultima eguaglianza deriva dalla (10.36), l'algoritmo (10.55) si modifica come segue, utilizzando, ad esempio, il metodo dei trapezi:

$$u_0^{k+1} = u_0^k - \frac{v\Delta t}{2}\left[\phi_0(t_k) + \phi_0(t_{k+1})\right], \qquad u_i^{k+1} = (1-\alpha)u_i^k + \alpha u_{i-1}^k, \qquad i = 1, \dots, N,$$

dove, al solito, si è usata la notazione (10.54).

Osservazione 10.7 *Con argomenti analoghi, si ottengono le varianti dell'algoritmo (10.56), qualora si abbiano condizioni al bordo di Dirichlet non omogeneee, o di Neumann, per il problema (10.47) (10.49): ovvero considerando, al posto della (10.49), condizioni al bordo del tipo*

$$u(1,t) = \phi_1(t), \qquad \textit{oppure} \qquad u_x(1,t) = \phi_1(t), \qquad t \geq 0.$$

In questo caso, infatti, in (10.56) si sostituisce $u_N^{k+1} = 0$, con

$$u_N^{k+1} = \phi_1(t_{k+1}), \qquad o \qquad u_N^{k+1} = u_N^k + \frac{v\Delta t}{2}\left[\phi_1(t_k) + \phi_1(t_{k+1})\right],$$

rispettivamente.

10.4 Equazione di trasporto e diffusione

Un problema di trasporto e diffusione ha, nel caso più semplice, la seguente forma:[5]

$$u_t(x,t) = u_{xx}(x,t) + vu_x(x,t), \qquad (x,t) \in \Omega, \tag{10.60}$$

$$u(x,0) = \omega_0(x), \qquad x \in [0,1], \tag{10.61}$$

$$u(0,t) = u(1,t) = 0, \qquad t \geq 0, \tag{10.62}$$

con $v \in \mathbb{R}$, costante nota.

[5] In questo caso, si normalizza a 1 il coefficiente della derivata spaziale di ordine massimo al secondo membro della (10.60), ed il segno della derivata prima non viene specificato.

Questo modello descrive la propagazione di un'onda (termine di trasporto, o convezione) che si smorza per effetto della diffusione: $v \in \mathbb{R}$ è la velocità di propagazione dell'onda, che si dirigerà verso *destra*, se $v < 0$, o verso *sinistra*, se $v > 0$; il coefficiente di diffusione è invece normalizzato a 1.

Semidiscretizzando la derivata seconda come visto per l'equazione del calore, e la derivata prima mediante *up-wind*, si ottiene il sistema di equazioni ordinarie

$$\begin{aligned} \mathbf{u}'(t) &= \left(\frac{1}{\Delta x^2} T_N + \frac{v}{\Delta x} L_{N-1}\right) \mathbf{u}(t) \qquad (10.63) \\ &\equiv \frac{1}{\Delta x^2} \left(T_N - 2\beta\, L_{N-1}\right) \mathbf{u}(t), \qquad \text{se} \qquad v < 0, \end{aligned}$$

$$\begin{aligned} \mathbf{u}'(t) &= \left(\frac{1}{\Delta x^2} T_N + \frac{v}{\Delta x} U_{N-1}\right) \mathbf{u}(t) \qquad (10.64) \\ &\equiv \frac{1}{\Delta x^2} \left(T_N + 2\beta\, U_{N-1}\right) \mathbf{u}(t), \qquad \text{se} \qquad v > 0, \end{aligned}$$

dove T_N, L_{N-1}, U_{N-1} sono definite in accordo alle (10.18), (10.42), (10.52), rispettivamente, il vettore $\mathbf{u}(t)$ e la condizione iniziale sono definiti in (10.16), mentre

$$\beta = \frac{1}{2}|v|\Delta x > 0 \qquad (10.65)$$

è denominato *numero di Péclet di cella* del problema semidiscreto [17, pag. 509].

Chiaramente, vi possono essere dei *boundary layer*, ovvero rapide variazioni della soluzione sul bordo del dominio, qualora $|v| \gg 1$, in quanto il profilo dell'onda, se non adeguatamente smorzato per effetto della diffusione, deve raccordarsi con la condizione omogenea (10.62).

Considerando lo spettro dell'operatore limite

$$T - 2\beta\, L \qquad \text{o} \qquad T + 2\beta\, U,$$

si verifica facilmente che esso è delimitato dal *boundary locus*

$$\sigma_+ = \left\{\lambda = -2(1+\beta) + 2\cos\theta + 2\beta \mathrm{e}^{i\theta},\ \theta \in [0, 2\pi]\right\},$$

che risulta essere un'ellisse, con asse maggiore $\Re(\lambda) \in [-4(1+\beta), 0]$ ed asse minore $\Im(\lambda) \in [-2\beta, 2\beta]$. Evidentemente,

$$\sigma_+ \to [-4, 0], \qquad \text{se} \qquad \Delta x \to 0 \quad \Rightarrow \quad \beta \to 0.$$

Tuttavia, questo non si può sempre assumere (lavorando con Δx piccolo ma finito), nei cosiddetti problemi "*transport dominated*", per i quali $|v| \gg 1$.

Da quanto visto in Sezione 10.2, il termine diffusivo è opportuno che venga discretzzato in modo implicito, mentre quello di trasporto può essere discretizzato mediante *up-wind*: in questo modo, l'utilizzo di un passo temporale

$$\Delta t = \frac{\Delta x}{|v|}, \qquad (10.66)$$

permette di risolvere la propagazione dell'onda, indipendentemente dal termine diffusivo:

$$\left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N\right) \mathbf{u}_{k+1} = \left(I_{N-1} + \frac{v\Delta t}{\Delta x} L_{N-1}\right) \mathbf{u}_k, \qquad \text{se } v < 0, \qquad (10.67)$$

$$\left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N\right) \mathbf{u}_{k+1} = \left(I_{N-1} + \frac{v\Delta t}{\Delta x} U_{N-1}\right) \mathbf{u}_k, \qquad \begin{array}{l} k = 0, 1, \ldots, \\ \text{se } v > 0. \end{array} \tag{10.68}$$

Evidentemente, questi schemi hanno ordine 1 sia nello spazio che nel tempo, e la loro complessità è lineare, sia in termini di operazioni e memoria richiesti per passo temporale, richiedendo la risoluzione di un sistema tridiagonale.

Esaminiamo adesso le proprietà di stabilità della soluzione discreta. Vale il seguente risultato.

Teorema 10.4 *L'operatore infinto*

$$T = \begin{pmatrix} (1+\alpha+\beta) & -\beta & \\ -\alpha & \ddots & \ddots \\ & \ddots & \end{pmatrix}, \qquad \alpha, \beta \geq 0,$$

è invertibile con inversa continua e $\|T\mathbf{x}\| \geq \|\mathbf{x}\|$, $\forall \mathbf{x} \in \ell_1$. *Pertanto, da (10.6) segue che*

$$\|T^{-1}\| \leq 1.$$

Dimostrazione. Sia

$$\mathbf{x} = (x_1,\ x_2,\ \ldots)^T \in \ell_1.$$

Ponendo $x_0 = 0$, si ottiene:

$$\begin{aligned} \|T\mathbf{x}\| &= \sum_{i\geq 1} |(1+\alpha+\beta)x_i - \alpha x_{i-1} - \beta x_i| \\ &\geq \sum_{i\geq 1} |(1+\alpha+\beta)x_i| - \alpha\,|x_{i-1}| - \beta\,|x_i| \\ &= (1+\alpha+\beta)\sum_{i\geq 1} |x_i| \ - \ (\alpha+\beta)\sum_{i\geq 1} |x_i| \\ &= (1+\alpha+\beta)\|\mathbf{x}\| - (\alpha+\beta)\|\mathbf{x}\| \ = \ \|\mathbf{x}\|. \ \square \end{aligned}$$

In modo del tutto simile, definendo

$$T_N = \begin{pmatrix} (1+\alpha+\beta) & -\beta & \\ -\alpha & \ddots & \ddots \\ & \ddots & \end{pmatrix} \in \mathbb{R}^{N\times N}, \qquad \alpha, \beta \geq 0, \tag{10.69}$$

si dimostra che:

$$\|T_N\mathbf{x}\| \geq \|\mathbf{x}\|, \qquad \forall \mathbf{x} \in \mathbb{R}^N. \tag{10.70}$$

Vale, allora, il seguente risultato.

Corollario 10.1 *Sia* T_N *la matrice definita in (10.69), soddisfacente (10.70). Segue che* T_N *è nonsingolare e, inoltre,*

$$\|T_N^{-1}\| \leq 1.$$

Dimostrazione. La dimostrazione segue immediatamente dal Teorema 10.2. □

Conseguentemente, per lo schema (10.67) si ottiene, utilizzando la norma 1 o la norma ∞:

$$\begin{aligned}\|\mathbf{u}_{k+1}\| &= \left\| \left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N \right)^{-1} \left(I_{N-1} + \frac{v\Delta t}{\Delta x} L_{N-1} \right) \mathbf{u}_k \right\| \\ &\leq \left\| \left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N \right)^{-1} \right\| \left\| \left(I_{N-1} + \frac{v\Delta t}{\Delta x} L_{N-1} \right) \right\| \|\mathbf{u}_k\| \\ &\leq \left(\left| 1 - \frac{|v|\Delta t}{\Delta x} \right| + \left| \frac{|v|\Delta t}{\Delta x} \right| \right) \|\mathbf{u}_k\| = \|\mathbf{u}_k\|,\end{aligned}$$

se

$$\Delta t \leq \frac{\Delta x}{|v|}, \tag{10.71}$$

che è in accordo con quanto osservato in (10.66). Ad analoga conclusione di perviene per lo schema (10.68), sostituendo L_{N-1} con U_{N-1}. Chiaramente, la condizione (10.71) assicura la stabilità della soluzione discreta generata da (10.67)-(10.68).

Osservazione 10.8 *È interessante osservare che la condizione (10.71), che è condizione solo* sufficiente *per la stabilità della soluzione discreta, garantisce anche la* positività *della stessa, nel caso in cui $\omega_0(x)$ sia positiva, come si dimostra facilmente osservando che la matrice $\left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N\right)$ è una M-matrice. Diversamente, la soluzione discreta potrebbe avere delle oscillazioni non fisiche nel tempo. Osserviamo, infine, che la (10.71) coincide con la condizione (10.45) ottenuta per lo schema* up-wind, *nel caso dell'equazione delle onde (in quel caso, infatti, si era usata la convenzione $v > 0$).*

A titolo di esempio, in Figura 10.5 riportiamo la soluzione discreta generata dallo schema (10.67), utilizzato per approssimare il problema (10.60) (10.62) con $v = -10$ e $\omega_0(x) = \sin(\pi x)$, utilizzando i passi:

$$\Delta x = 10^{-2}, \qquad \Delta t = \Delta x/|v| = 10^{-3}, \tag{10.72}$$

che sono in accordo con la (10.71). Similmente, in Figura 10.6 riportiamo la soluzione discreta ottenuta con lo schema (10.68), utilizzato per approssimare lo stesso problema, ma con $v = 10$, e con gli stessi passi (10.72) considerati prima.

È possibile rilassare il requisito (10.71) utilizzando la seguente discretizzazione della (10.63) (in modo del tutto analogo si procede per la (10.64)): per un fissato parametro $\theta \in [0, 1]$ si considera la discretizzazione (ricordiamo che $v < 0$):

$$\left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N - (1-\theta)\frac{v\Delta t}{\Delta x} L_{N-1} \right) \mathbf{u}_{k+1} = \left(I_{N-1} + \theta \frac{v\Delta t}{\Delta x} L_{N-1} \right) \mathbf{u}_k, \qquad k \geq 0,$$

che si riduce alla (10.67) per $\theta = 1$, e per cui la matrice a primo membro soddisfa le ipotesi del Corollario 10.1. Pertanto, procedendo in modo simile a quanto per il caso $\theta = 1$, si ottiene:

$$\|\mathbf{u}_{k+1}\| = \left\| \left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N - (1-\theta)\frac{v\Delta t}{\Delta x} L_{N-1} \right)^{-1} \left(I_{N-1} + \theta \frac{v\Delta t}{\Delta x} L_{N-1} \right) \mathbf{u}_k \right\|$$

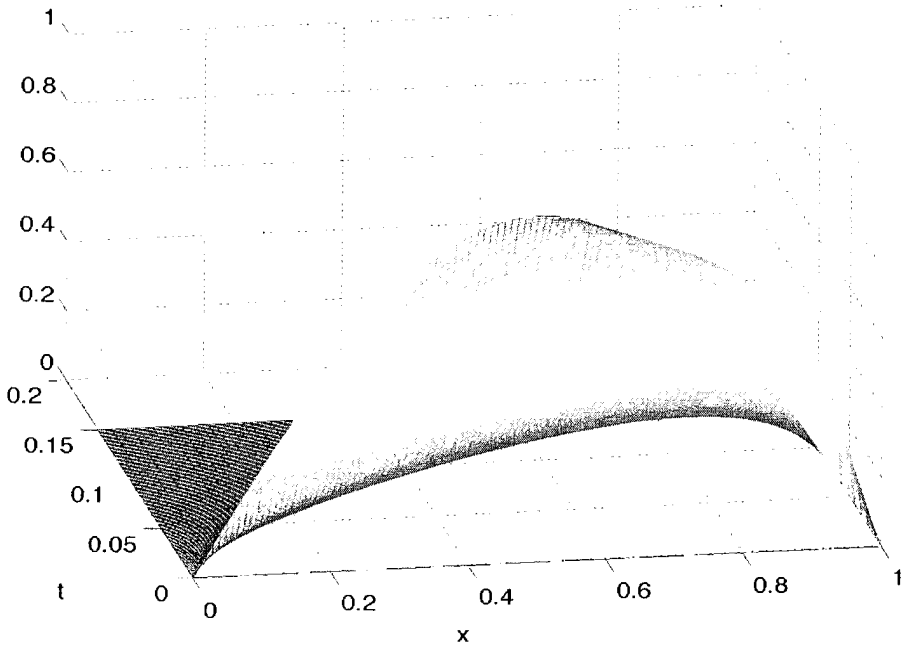

Figura 10.5: Metodo (10.67) la soluzione del problema (10.60)–(10.62), $v = -10, \omega_0(x) = \sin(\pi x)$, con passi $\Delta x = 10^{-2}, \Delta t = 10^{-3}$.

$$\begin{aligned} &\leq \left\| \left(I_{N-1} - \frac{\Delta t}{\Delta x^2} T_N - (1-\theta)\frac{v\Delta t}{\Delta x} L_{N-1} \right)^{-1} \right\| \left\| \left(I_{N-1} + \theta \frac{v\Delta t}{\Delta x} L_{N-1} \right) \right\| \|\mathbf{u}_k\| \\ &\leq \left(\left| 1 - \theta \frac{|v|\Delta t}{\Delta x} \right| + \theta \left| \frac{|v|\Delta t}{\Delta x} \right| \right) \|\mathbf{u}_k\| \;=\; \|\mathbf{u}_k\|, \end{aligned}$$

se

$$\theta \Delta t \leq \frac{\Delta x}{|v|}.$$

Chiaramente, quest'ultima condizione sarà facilmente soddisfabile, scegliendo θ sufficientemente piccolo.[6] Anche in questo caso, si ottiene che la predetta condizione, che è sufficiente per la stabilità, garantisce la nonnegatività della soluzione, nel caso in cui

$$\omega_0(x) \geq 0, \qquad \forall x \in [0,1].$$

Va tuttavia precisato che, sebbene la limitazione sul passo temporale sia adesso meno stringente, l'accuratezza della soluzione sarà minore, utilizzando un passo Δt più grande.

Osservazione 10.9 *Per concludere, osserviamo che argomenti del tutto analoghi a quelli visti nelle sezioni precedenti, per le equazioni del calore e delle onde, si applicano quando le condizioni al bordo (10.62) sono non omogenee,*

$$u(0,t) = \phi_0(t), \qquad u(1,t) = \phi_1(t), \qquad t \geq 0,$$

[6] In altri termini, sarà possibile utilizzare un passo massimo $\Delta t = \Delta x/(\theta|v|)$, invece di (10.66).

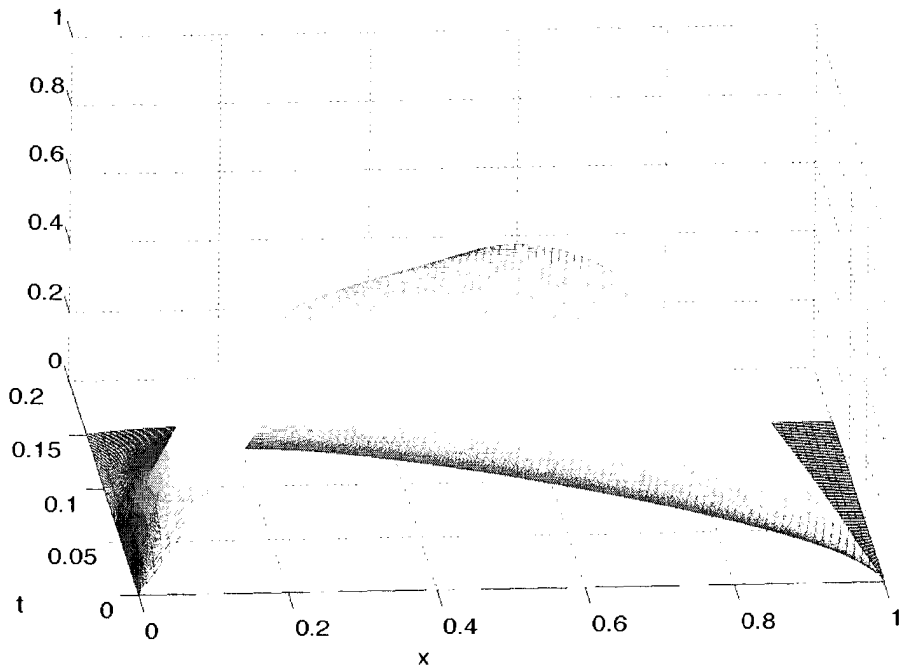

Figura 10.6: Metodo (10.68) la soluzione del problema (10.60)–(10.62), $v = 10, \omega_0(x) = \sin(\pi x)$, con passi $\Delta x = 10^{-2}, \Delta t = 10^{-3}$.

oppure sono di Neumann,

$$u_x(0,t) = \phi_0(t), \qquad u_x(1,t) = \phi_1(t), \qquad t \geq 0.$$

Questo sia per quanto riguarda l'analisi di stabilità, che porta alle stesse conclusioni appena viste, che per quanto riguarda l'implementazione dei metodi.

Appendice A

Il prodotto di Kronecker

In questa appendice enunciamo la definizione, ed alcune proprietà, del *prodotto tensoriale* o *prodotto di Kronecker* di matrici. Trattandosi di uno strumento algebrico assai potente, si esporranno, oltre alle definizioni e le proprietà strettamente indispensabili ai fini della trattazione degli argomenti del corso, alcune ulteriori proprietà ed applicazioni. Il materiale riportato è sostanzialmente tratto da [9, Sezione A.3].

A.1 Il prodotto di Kronecker

Siano $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ e $B \in \mathbb{R}^{k\times p}$: la matrice

$$A \otimes B \equiv \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}_{(m\,k)\times(n\,p)}$$

definisce il *prodotto di Kronecker*, o *prodotto tensoriale*, di A e B. Ad esempio:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \qquad \Rightarrow \qquad A \otimes B = \left(\begin{array}{ccc|ccc} 4 & 5 & 6 & 8 & 10 & 12 \\ 12 & 15 & 18 & 16 & 20 & 24 \end{array}\right).$$

Teorema A.1 *Siano A, B, C e D matrici le cui dimensioni siano compatibili con le operazioni di seguito indicate. Allora:*

i) $\forall \alpha \in \mathbb{R}$: $(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha(A \otimes B)$;

ii) $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$;

iii) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$;

iv) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$;

v) $(A \otimes B)^T = A^T \otimes B^T$;

vi) $(A \otimes B)(C \otimes D) = (A\,C) \otimes (B\,D)$;

vii) $\forall\, A \in \mathbb{R}^{n\times n}$ *e* $B \in \mathbb{R}^{m\times m}$: $A \otimes B = (A \otimes I_m)(I_n \otimes B)$;

viii) se A e B sono matrici nonsingolari: $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$;

ix) se $A \in \mathbb{R}^{n\times n}$ *e* $B \in \mathbb{R}^{m\times m}$: $\det(A \otimes B) = \det(A)^n \det(B)^m$.

Dimostrazione. La dimostrazione di *i)* *v)* segue facilmente dalla definizione di prodotto di Kronecker; *vii)* e *viii)* discendono da *vi)*. Il punto *vi)* può essere dimostrato come segue: sia $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ e $C = (c_{ij}) \in \mathbb{R}^{n\times k}$, allora

$$
\begin{aligned}
(A \otimes B)(C \otimes D) &= \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix} \begin{pmatrix} c_{11}D & \dots & c_{1k}D \\ \vdots & & \vdots \\ c_{n1}D & \dots & c_{nk}D \end{pmatrix} \\
&= \begin{pmatrix} \sum_{r=1}^n a_{1r}c_{r1}BD & \dots & \sum_{r=1}^n a_{1r}c_{rk}BD \\ \vdots & & \vdots \\ \sum_{r=1}^n a_{mr}c_{r1}BD & \dots & \sum_{r=1}^n a_{mr}c_{rk}BD \end{pmatrix} = \begin{pmatrix} (AC)_{11}BD & \dots & (AC)_{1k}BD \\ \vdots & & \vdots \\ (AC)_{m1}BD & \dots & (AC)_{mk}BD \end{pmatrix} \\
&= (AC) \otimes (BD).
\end{aligned}
$$

Dimostriamo, infine, il punto *ix)*. A questo fine, osserviamo che in generale $A \otimes B \neq B \otimes A$. Tuttavia, se $A = (a_{ij}) \in \mathbb{R}^{m\times m}$ e $B = (b_{ij}) \in \mathbb{R}^{n\times n}$, allora:

$$
A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mm}B \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & \dots & a_{11}b_{1n} & & a_{1m}b_{11} & \dots & a_{1m}b_{1n} \\ \vdots & & \vdots & \dots & \vdots & & \vdots \\ a_{11}b_{n1} & \dots & a_{11}b_{nn} & & a_{1m}b_{n1} & \dots & a_{1m}b_{nn} \\ & \vdots & & & & \vdots & \\ a_{m1}b_{11} & \dots & a_{m1}b_{1n} & & a_{mm}b_{11} & \dots & a_{mm}b_{1n} \\ \vdots & & \vdots & \dots & \vdots & & \vdots \\ a_{m1}b_{n1} & \dots & a_{m1}b_{nn} & & a_{mm}b_{n1} & \dots & a_{mm}b_{nn} \end{pmatrix}.
$$

Conseguentemente, definendo la matrice (ortogonale) di permutazione $Q_{m,n} \in \mathbb{R}^{(mn)\times(mn)}$ tale che

$$
\begin{aligned}
&Q_{m,n} \begin{pmatrix} 1 & 2 & 3 & \dots & mn \end{pmatrix}^T \\
&= \begin{pmatrix} 1 & n+1 & \dots & (m-1)n+1, & 2 & n+2 & \dots & (m-1)n+2, & \dots, & n & 2n & \dots & mn \end{pmatrix}^T,
\end{aligned}
$$

si ottiene:

$$
Q_{m,n}(A \otimes B)Q_{m,n}^T = \begin{pmatrix} a_{11}b_{11} & \dots & a_{1m}b_{11} & & a_{11}b_{1n} & \dots & a_{1m}b_{1n} \\ \vdots & & \vdots & \dots & \vdots & & \vdots \\ a_{m1}b_{11} & \dots & a_{mm}b_{11} & & a_{m1}b_{1n} & \dots & a_{mm}b_{1n} \\ & \vdots & & & & \vdots & \\ a_{11}b_{n1} & \dots & a_{1m}b_{n1} & & a_{11}b_{nn} & \dots & a_{1m}b_{nn} \\ \vdots & & \vdots & \dots & \vdots & & \vdots \\ a_{m1}b_{n1} & \dots & a_{mm}b_{n1} & & a_{m1}b_{nn} & \dots & a_{mm}b_{nn} \end{pmatrix} = B \otimes A.
$$

Da questo segue infine che:

$$
\begin{aligned}
\det(A \otimes B) &= \det(A \otimes I_n)\det(I_m \otimes B) \\
&= \det(Q_{m,n}(A \otimes I_n)Q_{m,n}^T)\det(I_m \otimes B) = \det(I_n \otimes A)\det(I_m \otimes B)
\end{aligned}
$$

$$= \det\begin{pmatrix} A & & \\ & \ddots & \\ & & A \end{pmatrix} \det\begin{pmatrix} B & & \\ & \ddots & \\ & & B \end{pmatrix} = \det(A)^n \det(B)^m. \square$$

Il prossimo risultato, che enunciamo senza dimostrare, evidenzia le connessioni tra gli autovalori/autovettori di due matrici A e B e quelli di $A \otimes B$.

Teorema A.2 *Siano $\lambda_1, \ldots, \lambda_m$ gli autovalori di A e $\mu_1, \ldots, \mu_n$ quelli di B. Si consideri, inoltre, il polinomio*

$$p(x, y) = \sum_{i,j=0}^{k} c_{ij} x^i y^j,$$

e la matrice

$$C = \sum_{i,j=0}^{k} c_{ij} A^i \otimes B^j.$$

Allora gli autovalori mn autovalori, $\{\eta_{rs}\}$, di C sono dati da:

$$\eta_{rs} = p(\lambda_r, \mu_s), \qquad r = 1, \ldots, m, \quad s = 1, \ldots, n.$$

Inoltre, se $A\mathbf{u} = \lambda\mathbf{u}$ e $B\mathbf{v} = \mu\mathbf{v}$, segue che:

$$C(\mathbf{u} \otimes \mathbf{v}) = p(\lambda, \mu)(\mathbf{u} \otimes \mathbf{v}).$$

Si ottengono, di conseguenza, i seguenti semplici risultati:

1. gli autovalori di $A \otimes B$ sono gli mn numeri $\lambda_r \mu_s$, $r = 1, \ldots, m$, $s = 1, \ldots, n$.
2. gli autovalori di
$$(A \otimes I_n) + (I_m \otimes B), \tag{A.1}$$
detta *somma di Kronecker* di A e B, sono gli mn numeri
$$\lambda_r + \mu_s, \qquad r = 1, \ldots, m, \ s = 1, \ldots, n. \tag{A.2}$$

A.2 Risoluzione di equazioni matriciali

Siano $A \in \mathbb{R}^{k \times n}$, $B \in \mathbb{R}^{n \times s}$, e $C \in \mathbb{R}^{k \times s}$ matrici date. Una matrice $X \in \mathbb{R}^{n \times n}$ soddisfacente l'equazione

$$AXB = C, \tag{A.3}$$

è una *soluzione matriciale* di (A.3). Questo problema può essere formulato in modo più semplice da trattare, introducendo la seguente funzione *vec*:

$$Z \in \mathbb{R}^{m \times n} \qquad \Rightarrow \qquad \mathrm{vec}(Z) = \begin{pmatrix} Z_{*1} \\ Z_{*2} \\ \vdots \\ Z_{*n} \end{pmatrix} \in \mathbb{R}^{mn},$$

dove con Z_{*j} abbiamo denotato la j-esima colonna di Z. Valgono le seguenti proprietà.

Teorema A.3 *Siano A, B, X matrici definite come in (A.3), Y avente la stessa dimensione di X, e $\alpha, \beta \in \mathbb{R}$. Allora:*

i) $\mathrm{vec}(X) = \mathrm{vec}(Y)$ *se e solo se* $X = Y$;

ii) $\mathrm{vec}(\alpha X + \beta Y) = \alpha\,\mathrm{vec}(X) + \beta\,\mathrm{vec}(Y)$;

iii) $\mathrm{vec}(AXB) = (B^T \otimes A)\mathrm{vec}(X)$.

Dimostrazione. La dimostrazione di *i)* e *ii)* è banale. Riguardo al punto *iii)*, ponendo$B = (b_{ij})$, si ottiene:

$$\mathrm{vec}(AXB) = \mathrm{vec}\left([AX_{*1}, \ldots, AX_{*n}]B\right) = \mathrm{vec}\left(\left[\sum_{i=1}^{n} b_{i1}AX_{*i}, \ldots, \sum_{i=1}^{n} b_{is}AX_{*i}\right]\right)$$

$$= \begin{pmatrix} \sum_{i=1}^{n} b_{i1}AX_{*i} \\ \vdots \\ \sum_{i=1}^{n} b_{is}AX_{*i} \end{pmatrix} = \begin{pmatrix} b_{11}A & \ldots & b_{n1}A \\ \vdots & & \vdots \\ b_{1s}A & \ldots & b_{ns}A \end{pmatrix} \begin{pmatrix} X_{*1} \\ \vdots \\ X_{*n} \end{pmatrix} = (B^T \otimes A)\mathrm{vec}(X). \square$$

Pertanto, dalla proprietà *iii)* del precedente teorema segue, ponendo

$$\mathbf{x} = \mathrm{vec}(X) \qquad \text{e} \qquad \mathbf{c} = \mathrm{vec}(C), \tag{A.4}$$

che l'equazione (A.3) può essere equivalentemente riformulata come il sistema di equazioni lineari:

$$(B^T \otimes A)\mathbf{x} = \mathbf{c}.$$

Un'altra importante applicazione è la risoluzione dell'*equazione di Sylvester*,

$$AX + XB = C, \tag{A.5}$$

in cui $A, B, C, X \in \mathbb{R}^{n\times n}$, A, B, C sono note, e X è da determinare. Utilizzando gli stessi argomenti appena illustrati, e la notazione (A.4), si vede facilmente che (A.5) è equivalente al sistema lineare

$$\left[A \otimes I_n + I_n \otimes B^T\right]\mathbf{x} = \mathbf{c},$$

la cui matrice dei coefficienti è la somma di Kronecker di A e B^T (vedi (A.1)). Pertanto, dalla (A.2) segue che (A.5) ammette soluzione, e questa è unica, se e solo se A e $-B$ non hanno autovalori comuni.

Come caso particolare, quando in (A.5) $B = A^T$ a $C = C^T$, si ottiene la *equazione di Lyapunov continua*. Di quest'ultima esiste anche una *versione discreta*, della forma:

$$AXA^T - X = C,$$

con $C = C^T$, che, in virtù di quanto visto per la (A.3), si può formulare come il sistema lineare

$$\left[A \otimes A - I\right]\mathbf{x} = \mathbf{c},$$

in cui è stata, al solito, utilizzata la notazione (A.4). Evidentemente, questo sistema lineare ammette soluzione, e questa è unica, se e solo se $1 \notin \sigma(A)$.

Bibliografia

[1] L. Aceto, D. Trigiante. The matrices of Pascal and other greats. *Amer. Math. Monthly* **108**, no. 3 (2001) 232–245.

[2] P. Amodio, L. Brugnano. The Conditioning of Toeplitz Banded Matrices. *Mathematical and Computer Modelling* **23**(10) (1996) 29–42.

[3] L. Brugnano, F. Iavernaro. *Line Integral Methods and their application for the solution of conservative problems*, (2013) `arXiv:1301.2367 [math.NA]` (Stable URL at `http://arxiv.org/abs/1301.2367`)

[4] L. Brugnano, F. Iavernaro, D. Trigiante. *The Hamiltonian BVMs (HBVMs) Homepage*, (2010). (Stable URL at `http://arxiv.org/abs/1002.2757`)

[5] L. Brugnano, F. Iavernaro, D. Trigiante. A simple framework for the derivation and analysis of effective one-step methods for ODEs. *Applied Mathematics and Computation* **218** (2012) 8475–8485.

[6] L. Brugnano, C. Magherini, A. Sestini. *Calcolo Numerico, seconda edizione ampliata e corretta*. Master, Università & Professioni, Firenze, 2010 (x+158 pp.).

[7] L. Brugnano, F. Mazzia, D. Trigiante. Fifty Years of Stiffness. Chapter 1 in *Recent Advances in Computational and Applied Mathematics*, T.E. Simos Ed., Springer, 2011, pp. 1–21.

[8] L. Brugnano, D. Trigiante. On the characterization of *stiffness* for ODEs. *Dynamics of Continuous, Discrete and Impulsive Systems* **2**, 3 (1996) 317–335.

[9] L. Brugnano, D. Trigiante. *Solving differential problems by multistep initial and boundary value methods*. Gordon and Breach Science Publishers, Amsterdam, 1998. (xvi+418 pp.)

[10] G. Di Lena, D. Trigiante. On the spectrum of families of matrices with applications to stability problems. *Lecture Notes in Mathematics* **1386** (1989) 36–53.

[11] E. Hairer, S.N. Nørsett, G. Wanner. *Solving Ordinary Differential Equations I*, 3^{rd} corr. ed., Springer-Verlag, Berlin, 2008.

[12] E. Hairer, G. Wanner. *Solving Ordinary Differential Equations II*, 2^{nd} rev. ed., Springer-Verlag, Berlin, 1996.

[13] F. Iavernaro, F. Mazzia, D. Trigiante. Stability and Conditioning in Numerical Analysis. *Journal of Numerical Analysis, Industrial and Applied Mathematics* **1**, 1 (2006) 81–90.

[14] V. Lakshmikantham, D. Trigiante. *Theory of difference equations: numerical methods and applications. Second edition.* Monographs and Textbooks in Pure and Applied Mathematics, 251. Marcel Dekker, Inc., New York, 2002. (x+300 pp.)

[15] P. Lancaster, M. Tismenetsky. *The theory of matrices. Second edition.* Computer Science and Applied Mathematics. Academic Press, Inc., Orlando, FL, 1985. (xv+570 pp.)

[16] C. Sparrow. *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors.* Springer-Verlag, New York, 1982.

[17] G. Strang. *Computational Science and Engineering*, Wellesley-Cambridge Press, 2007

[18] D. Trigiante. *Dispense del corso di "Metodi di Approssimazione".*