Modellistica Numerica

Università di Pavia

18 febbraio 2020 — ore 10:00 — aula E9

Punti totali: 34 Durata dell'esame: 2 ore.

Problema 1. Equazione ellittica a coefficienti variabili [17 punti]

Consideriamo il problema al bordo

$$-\big(\mathrm{e}^{-x^2} u'(x)\big)' = x^3 \qquad x \in (-1,1), \qquad u(-1) = u(1) = 0. \tag{1}$$

(a) Derivare una formulazione debole del problema (1).

(b) Scrivere il problema in forma variazionale astratta.
La forma bilineare ottenuta è simmetrica?

(c) Mostrare che questo problema variazionale soddisfa le tre ipotesi del teorema di Lax–Milgram.
Quali conseguenze per il problema al bordo se ne possono dedurre?

(d) Derivare una maggiorazione (stima di stabilità) della soluzione u, misurata in una norma appropriata.

(e) Consideriamo il metodo degli elementi finiti lineari per questo problema, usando una griglia di nodi equispaziati.
Descrivere gli elementi della matrice $\underline{\underline{\mathbf{A}}}$ e del termine noto $\vec{\mathbf{F}}$ del sistema lineare corrispondente.
È possibile calcolare gli elementi di $\underline{\underline{\mathbf{A}}}$ esattamente?
La matrice ottenuta è sparsa?

Soluzione: (a) Moltiplicando per una funzione test $w \in H_0^1(-1,1)$, integrando per parti e usando che $w(\pm 1) = 0$ poiché $w \in H_0^1(-1,1)$, si ottiene

$$\int_{-1}^{1} x^3\, w \,\mathrm{d}x = \int_{-1}^{1} \Big(- \big(\mathrm{e}^{-x^2} u'\big)' w \Big) \mathrm{d}x = \int_{-1}^{1} \Big(\mathrm{e}^{-x^2} u' w' \Big) \mathrm{d}x - \mathrm{e}^{-1} u'(1) \underbrace{w(1)}_{=0} + \mathrm{e}^{-1} u'(-1) \underbrace{w(-1)}_{=0}$$
$$= \int_{-1}^{1} \Big(\mathrm{e}^{-x^2} u' w' \Big) \mathrm{d}x,$$

quindi la formulazione debole è

$$\text{trovare } u \in H_0^1(-1,1) \quad \text{tale che} \quad \int_{-1}^{1} \mathrm{e}^{-x^2} u' w' \,\mathrm{d}x = \int_{-1}^{1} x^3\, w \,\mathrm{d}x \quad \forall w \in H_0^1(-1,1).$$

(È possibile scrivere altri problemi variazionali calcolando la derivata del prodotto $\mathrm{e}^{-x^2} u'$ e perdendo la simmetria.)

(b) Il problema variazionale astratto generico consiste nel trovare $u \in V$ tale che $\mathcal{A}(u,w) = \mathcal{F}(w)$ per ogni $w \in V$. Qui semplicemente

$$V = H_0^1(-1,1), \qquad \mathcal{A}(u,w) = \int_{-1}^{1} \mathrm{e}^{-x^2} u'(x) w'(x) \,\mathrm{d}x, \qquad \mathcal{F}(w) = \int_{-1}^{1} x^3\, w(x) \,\mathrm{d}x.$$

La forma è simmetrica: $\mathcal{A}(u,w) = \mathcal{A}(w,u)$ per ogni $u, w \in H_0^1(-1,1)$.

(c) Mostriamo le tre ipotesi del Teorema di Lax–Milgram usando la disuguaglianza di Cauchy–Schwarz:

$$|\mathcal{F}(w)| = \left| \int_{-1}^{1} x^3\, w \,\mathrm{d}x \right| \le \left\| x^3 \right\|_{L^2(-1,1)} \|w\|_{L^2(-1,1)} = \sqrt{\frac{2}{7}}\, \|w\|_{L^2(-1,1)},$$

$$|\mathcal{A}(u,w)| = \left| \int_{-1}^{1} \mathrm{e}^{-x^2} u' w' \,\mathrm{d}x \right| \le \left\| \mathrm{e}^{-x^2} \right\|_{L^\infty(-1,1)} \left\| u' \right\|_{L^2(-1,1)} \left\| w' \right\|_{L^2(-1,1)} \le \|u\|_{H^1(-1,1)} \|w\|_{H^1(-1,1)},$$

$$\begin{aligned}\mathcal{A}(w,w) &= \int_{-1}^{1} \mathrm{e}^{-x^2}(w')^2\,\mathrm{d}x \geq \inf_{x\in(-1,1)} \mathrm{e}^{-x^2} \int_{-1}^{1}(w')^2\,\mathrm{d}x \\ &= \mathrm{e}^{-1}\int_{-1}^{1}(w')^2\,\mathrm{d}x = \mathrm{e}^{-1}\,|w|^2_{H^1(-1,1)} \geq \mathrm{e}^{-1}\frac{1}{1+C_P^2}\,\|w\|^2_{H^1(-1,1)}\end{aligned}$$

dove C_P è la costante della disuguaglianza di Poincaré e abbiamo usato $(w')^2 \geq 0$. Possiamo prendere

$$C_{\mathcal{F}} = \left\|x^3\right\|_{L^2(-1,1)} = \sqrt{\frac{2}{7}} \approx 0.5345, \qquad C_{\mathcal{A}} = 1, \qquad \gamma_{\mathcal{A}} = \frac{\mathrm{e}^{-1}}{1+C_P^2}.$$

Abbiamo anche $C_P = \frac{b-a}{\sqrt{2}} = \sqrt{2}$ quindi $\gamma_{\mathcal{A}} = \frac{1}{3\mathrm{e}} \approx 0.1226$.

Deduciamo che il problema al bordo ammette un'unica soluzione.

$$\text{(d)} \qquad \|u\|_{H^1(-1,1)} \leq \frac{C_{\mathcal{F}}}{\gamma_{\mathcal{A}}} = \sqrt{\frac{2}{7}}\,\mathrm{e}(1+C_P^2) \approx 4.359.$$

(e) Siano $x_j = -1 + hj$ i nodi della griglia, per $h = \frac{2}{n+1}$ e $j = 0, \ldots, n+1$, $K_j = [x_{j-1}, x_j]$ per $j = 1, \ldots, n+1$ gli elementi, e $\varphi_j(x) = \max\{0, 1 - \frac{|x-x_j|}{h}\}$ per $j = 1, \ldots, n$ le funzioni di base "a tenda" corrispondenti. La matrice $\underline{\underline{\mathbf{A}}} \in \mathbb{R}^{n\times n}$ ed il termine noto $\vec{\mathbf{F}} \in \mathbb{R}^n$ del sistema lineare degli elementi finiti hanno componenti

$$\begin{aligned}A_{j,k} =& \mathcal{A}(\varphi_k,\varphi_j) = \int_{-1}^{1} \mathrm{e}^{-x^2}\varphi_j'(x)\varphi_k'(x)\,\mathrm{d}x = \begin{cases} -\frac{1}{h^2}\int\limits_{K_j} \mathrm{e}^{-x^2}\,\mathrm{d}x & k=j-1, \\ \frac{1}{h^2}\int\limits_{K_j\cup K_{j+1}} \mathrm{e}^{-x^2}\,\mathrm{d}x & k=j, \\ -\frac{1}{h^2}\int\limits_{K_{j+1}} \mathrm{e}^{-x^2}\,\mathrm{d}x & k=j+1, \\ 0 & |k-j|>1, \end{cases} \qquad j,k=1,\ldots,n, \\ F_j =& \mathcal{F}(\varphi_j) = \int_{-1}^{1} x^3\varphi_j(x)\,\mathrm{d}x = \frac{1}{h}\int_{x_{j-1}}^{x_j} x^3(x-x_{j-1})\,\mathrm{d}x + \frac{1}{h}\int_{x_j}^{x_{j+1}} x^3(x_{j+1}-x)\,\mathrm{d}x.\end{aligned}$$

Non è possibile calcolare $A_{j,k}$ esattamente poiché contiene l'integrale di e^{-x^2}, quindi richiede l'uso di una formula di quadratura.

Nonostante e^{-x^2} sia diversa da zero su tutto il dominio, le funzioni di base sono supportate solo in due elementi della griglia, quindi la matrice $\underline{\underline{\mathbf{A}}}$ è sparsa e tridiagonale.

Problema 2. DFT & FFT [17 punti]

(a) Dare la definizione della trasformata di Fourier discreta (DFT_n) in $\mathbb{C}^n$.

(b) Calcolare $DFT_4(\vec{\mathbf{z}})$ per il vettore $\vec{\mathbf{z}} = (1000, 100, 10, 1)^\top$.

(c) Per $n = 2m$, $m \in \mathbb{N}$, $\vec{\mathbf{x}} \in \mathbb{C}^n$, definiamo

$$\vec{\mathbf{x}}^{\text{dispari}} := (x_1, x_3, \cdots, x_{n-1})^\top \in \mathbb{C}^m, \qquad \vec{\mathbf{x}}^{\text{pari}} := (x_2, x_4, \cdots, x_n)^\top \in \mathbb{C}^m.$$

Dimostrare che, per $j = 1, \ldots, n$,

$$\big(DFT_n(\vec{\mathbf{x}})\big)_j = \big(DFT_m(\vec{\mathbf{x}}^{\text{dispari}})\big)_j + \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}(j-1)}\big(DFT_m(\vec{\mathbf{x}}^{\text{pari}})\big)_j. \tag{2}$$

Qui assumiamo di indicizzare gli elementi dei vettori $\vec{\mathbf{v}} \in \mathbb{C}^m$ (in particolare $\vec{\mathbf{v}} = DFT_m(\vec{\mathbf{x}}^{\text{dispari}})$ e $\vec{\mathbf{v}} = DFT_m(\vec{\mathbf{x}}^{\text{pari}})$) con indici periodici: $\vec{\mathbf{v}}_{m+j} = \vec{\mathbf{v}}_j$.

(d) Spiegare brevemente come usare la formula (2) per calcolare la trasformata di Fourier veloce (FFT) di un vettore in $\mathbb{C}^n$ con $n = 2^N$.

Qual è il costo computazionale della FFT, in funzione della dimensione n?

(e) Come nel punto (b), calcolare ancora $DFT_4(\vec{\mathbf{z}})$, questa volta usando la formula (2) in modo ricorsivo.

(f) Completare il seguente codice Matlab che calcola la FFT in $\mathbb{C}^{2^N}$ in modo ricorsivo.

Assumere che il vettore di input `y` è un vettore colonna la cui lunghezza è una potenza di 2.

```
function Y = fft_ricorsiva(y)
n = length (y);
if n == 1
    [            (I)            ]
else
    Ydisp = fft_ricorsiva(y(1:2:n));
    Ypari = fft_ricorsiva(y(2:2:n));
    [            (II)           ]
end
end
```

SOLUZIONE: (a) Per $n \in \mathbb{N}$, la DFT è la mappa lineare $DFT_n : \mathbb{C}^n \to \mathbb{C}^n$, $\vec{\mathbf{v}} \mapsto \underline{\underline{\mathbf{W}}}\vec{\mathbf{v}}$, dove la matrice $\underline{\underline{\mathbf{W}}} \in \mathbb{C}^{n\times n}$ ha elementi $\overline{W}_{j,k} = \overline{\omega}_n^{(j-1)(k-1)}$ per $\overline{\omega}_n = \mathrm{e}^{-\frac{2\pi\mathrm{i}}{n}}$.

(b) Usiamo che $\overline{\omega}_4 = \mathrm{e}^{-\frac{2\pi\mathrm{i}}{4}} = \mathrm{e}^{-\frac{\pi\mathrm{i}}{2}} = -\mathrm{i}$.

$$\underline{\underline{\mathbf{W}}} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -\mathrm{i} & -1 & \mathrm{i} \\ 1 & -1 & 1 & -1 \\ 1 & \mathrm{i} & -1 & -\mathrm{i} \end{pmatrix}, \quad DFT_4\begin{pmatrix} 1000 \\ 100 \\ 10 \\ 1 \end{pmatrix} = \underline{\underline{\mathbf{W}}}\begin{pmatrix} 1000 \\ 100 \\ 10 \\ 1 \end{pmatrix} = \begin{pmatrix} 1111 \\ 990-99\mathrm{i} \\ 909 \\ 990+99\mathrm{i} \end{pmatrix}$$

In alternativa, si può espandere la somma sugli indici invece di usare la matrice $\underline{\underline{\mathbf{W}}}$: per $j = 1, 2, 3, 4$

$$\begin{aligned}
\big(DFT_4(\vec{\mathbf{z}})\big)_j &= \sum_{k=1}^4 \overline{\omega}_4^{(j-1)(k-1)} z_k = \sum_{k=1}^4 (-\mathrm{i})^{(j-1)(k-1)} z_k = z_1 + (-\mathrm{i})^{j-1} z_2 + (-1)^{j-1} z_3 + \mathrm{i}^{j-1} z_4 \\
&= \begin{cases} z_1 + z_2 + z_3 + z_4 \\ z_1 - \mathrm{i}z_2 - z_3 + \mathrm{i}z_4 \\ z_1 - z_2 + z_3 - z_4 \\ z_1 + \mathrm{i}z_2 - z_3 - \mathrm{i}z_4 \end{cases} = \begin{cases} 1111 & j = 1, \\ 990 - 99\mathrm{i} & j = 2, \\ 909 & j = 3, \\ 990 + 99\mathrm{i} & j = 4. \end{cases}
\end{aligned}$$

(c)
$$\begin{aligned}
y_j = \big(DFT_n(\vec{\mathbf{x}})\big)_j &= \sum_{k=1}^n x_k \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}(j-1)(k-1)} \\
&= \sum_{\ell=1}^m x_{2\ell-1} \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}(j-1)2(\ell-1)} + \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}(j-1)} \sum_{\ell=1}^m x_{2\ell} \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}(j-1)2(\ell-1)} \\
&= \sum_{\ell=1}^m x_{2\ell-1} \mathrm{e}^{-\mathrm{i}\frac{2\pi}{m}(j-1)(\ell-1)} + \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}(j-1)} \sum_{\ell=1}^m x_{2\ell} \mathrm{e}^{-\mathrm{i}\frac{2\pi}{m}(j-1)(\ell-1)}
\end{aligned}$$

$$= \left(DFT_m(\vec{\mathbf{x}}^{\text{dispari}})\right)_j + \mathrm{e}^{-\mathrm{i}\frac{2\pi}{n}(j-1)}\left(DFT_m(\vec{\mathbf{x}}^{\text{pari}})\right)_j, \qquad 1 \le j \le n.$$

(d) La formula (2) permette di calcolare la DFT_n come 2 $DFT_{n/2}$, che a sua volta sono calcolate come 4 $DFT_{n/4}$, ..., finché si calcolano n DFT_1, cioè identità. Questo può essere ottenuto in modo ricorsivo: la DFT_n chiama 2 $DFT_{n/2}$ etc.

Il costo è $O(n\log_2 n) = O(2^N N)$: ci sono $N = \log_2 n$ livelli e in ciascuno di questi si compiono $O(n) = O(2^N)$ operazioni.

(e) La DFT_4 richiede due DFT_2, che a loro volta richiedono quattro DFT_1. Queste ultime sono identità: $DFT_1 1000 = 1000$, $DFT_1 100 = 100$, $DFT_1 10 = 10$, $DFT_1 1 = 1$. Da queste calcoliamo le due DFT_2 usando (2) con $n = 2$:

$$DFT_2\begin{pmatrix}1000\\10\end{pmatrix} = \begin{pmatrix}DFT_1 1000 + \mathrm{e}^{-\mathrm{i}\pi 0}DFT_1 10\\ DFT_1 1000 + \mathrm{e}^{-\mathrm{i}\pi 1}DFT_1 10\end{pmatrix} = \begin{pmatrix}1000+10\\1000-10\end{pmatrix} = \begin{pmatrix}1010\\990\end{pmatrix},$$

$$DFT_2\begin{pmatrix}100\\1\end{pmatrix} = \begin{pmatrix}DFT_1 100 + \mathrm{e}^{-\mathrm{i}\pi 0}DFT_1 1\\ DFT_1 100 + \mathrm{e}^{-\mathrm{i}\pi 1}DFT_1 1\end{pmatrix} = \begin{pmatrix}100+1\\100-1\end{pmatrix} = \begin{pmatrix}101\\99\end{pmatrix}.$$

Usando ancora (2), questa volta con $n = 4$ otteniamo il risultato:

$$DFT_4\begin{pmatrix}1000\\100\\10\\1\end{pmatrix} = \begin{pmatrix}DFT_2\begin{pmatrix}1000\\10\end{pmatrix}\\ DFT_2\begin{pmatrix}1000\\10\end{pmatrix}\end{pmatrix} + \begin{pmatrix}1\\-\mathrm{i}\\-1\\\mathrm{i}\end{pmatrix} .* \begin{pmatrix}DFT_2\begin{pmatrix}100\\1\end{pmatrix}\\ DFT_2\begin{pmatrix}100\\1\end{pmatrix}\end{pmatrix} = \begin{pmatrix}1010+(1)101\\990+(-\mathrm{i})99\\1010+(-1)101\\990+(\mathrm{i})99\end{pmatrix} = \begin{pmatrix}1111\\990-\mathrm{i}99\\909\\990+\mathrm{i}99\end{pmatrix}.$$

(f)

```
function Y = fft_ricorsiva(y)
n = length (y);
if n == 1
    Y = y;
else
    Ydisp = fft_ricorsiva(y(1:2:n));
    Ypari = fft_ricorsiva(y(2:2:n));
    Y = [Ydisp; Ydisp] + (exp(-2*pi*1i/n).^((0:n-1)')).*[Ypari;Ypari];
end
end
```