Confronto sui vari FEM

Si consideri l'equazione differenziale lineare del secondo ordine:

$$\mathcal{L}(u) = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + u + x = 0$$

definita sul dominio [0,1] e con condizioni al contorno omogenee $u(0) = u(1) = 0$.
Determinare la soluzione numerica mediante i seguenti metodi variazionali:

- collocation sui punti $x = 1/4$ e $x = 1/2$;
- minimi quadrati;
- Galerkin.

Si consideri come base dello spazio funzionale in cui u viene approssimata la successione delle potenze crescenti di x e si scelga la funzione approssimante $\hat{u}$ in modo da soddisfare le condizioni al contorno. Confrontare la soluzione numerica ottenuta mediante i tre metodi variazionali con quella analitica sui punti $x = 1/4$, $x = 1/2$ e $x = 3/4$ e commentare i risultati.

Soluzione analitica

Dalla teoria delle equazioni differenziali, u è esprimibile come somma della soluzione $u^{(o)}$ dell'equazione differenziale omogenea associata a quella data e di una soluzione particolare u^* avente la forma del termine noto. La soluzione $u^{(o)}$ di $u'' + u = 0$ è una combinazione lineare di seni e coseni, mentre si verifica immediatamente che l'unica soluzione particolare nella forma del termine noto, (quindi nella forma di un polinomio di primo grado), è $-x$. Pertanto u risulta:

$$u = C\text{sen}x + D\text{cos}x - x$$

Imponendo le condizioni al contorno si determinano le due costanti C e D. La soluzione analitica è quindi:

$$u = \frac{\text{sen}x}{\text{sen}1} - x$$

Approssimazione per u

Scegliamo ora la funzione $\hat{u}$ che approssima u. Utilizziamo come funzioni base la successione delle potenze di x, cioè:

$$x^0, x^1, x^2, \ldots, x^n$$

Dovendo inoltre $\hat{u}$ soddisfare alle condizioni al contorno, aggiungiamo il fattore $x(1-x)$, ottenendo la seguente espressione:

$$\hat{u} = x(1-x) \sum_{i=1}^{n} a_i x^{i-1}$$

Il problema consiste nel determinare i coefficienti a_i che consentono di approssimare u mediante $\hat{u}$ secondo i metodi variazionali di collocation, minimi quadrati e Galerkin.

Metodo di collocation su 2 punti

Il metodo di collocation consente di determinare i coefficienti incogniti a_i imponendo l'ortogonalità fra il residuo dell'equazione differenziale calcolata utilizzando $\hat{u}$ anziché u e la delta di Dirac sugli n punti di collocation. Nel caso in esame, il residuo corrisponde a $\mathcal{L}(\hat{u})$ ed i nodi di collocation sono 2:

$$\int_0^1 \mathcal{L}(\hat{u})\delta(x - x_i)\mathrm{d}x = 0 \qquad i = 1, 2$$

Poiché si impongono 2 condizioni, i coefficienti incogniti determinabili sono a_1 ed a_2 e l'espressione di $\hat{u}$ risulta:

$$\hat{u} = x(1 - x)(a_1 + a_2 x)$$

Le espressioni integrali del metodo di collocation corrispondono ad annullare direttamente il residuo sui punti di collocation.

L'espressione esplicita di $\mathcal{L}(\hat{u})$ è:

$$\begin{aligned}
\mathcal{L}(\hat{u}) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}[x(1-x)(a_1+a_2x)] + x(1-x)(a_1+a_2x) + x \\
&= 2(a_2-a_1) - 6a_2x + a_1x + (a_2-a_1)x^2 - a_2x^3 + x \\
&= a_1(-x^2+x-2) + a_2(-x^3+x^2-6x+2) + x
\end{aligned}$$

Annullando $\mathcal{L}(\hat{u})$ sui punti di collocation si ottiene il seguente sistema:

$$\begin{pmatrix} -116 & 35 \\ -14 & -7 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -16 \\ -4 \end{pmatrix}$$

la cui soluzione dà $a_1 = \frac{42}{217} \simeq 0.19355$ e $a_2 = \frac{40}{217} \simeq 0.18433$.
L'espressione approssimata della soluzione con il metodo di collocation su due punti è dunque:

$$u_{c,2} = x(1-x)(0.19355 + 0.18433x)$$

Metodo dei minimi quadrati

I coefficienti incogniti a_1 e a_2 vengono in questo caso scelti in modo da minimizzare il funzionale determinato integrando il quadrato del residuo:

$$\Omega(\hat{u}) = \int_0^1 \mathcal{L}(\hat{u})^2 \mathrm{d}x = \mathrm{min}$$

Deriviamo rispetto ad a_1 ed a_2 ed introduciamo l'espressione esplicita di $\mathcal{L}(\hat{u})$, ottenendo:

$$\frac{\partial \Omega}{\partial a_1} = 2\int_0^1 \left[a_1(-2+x-x^2) + a_2(2-6x+x^2-x^3) + x\right](-2+x-x^2)\mathrm{d}x = 0$$
$$\frac{\partial \Omega}{\partial a_2} = 2\int_0^1 \left[a_1(-2+x-x^2) + a_2(2-6x+x^2-x^3) + x\right](2+6x+x^2-x^3)\mathrm{d}x = 0$$

Risolvendo gli integrali così ricavati, si ha il sistema:

$$\begin{pmatrix} 202 & 101 \\ 101 & \frac{1572}{7} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 55 \\ 57 \end{pmatrix}$$

la cui soluzione dà $a_1 \simeq 0.18754$ e $a_2 \simeq 0.16947$. L'espressione approssimata della soluzione con il metodo dei minimi quadrati assume pertanto la seguente forma:

$$u_Q = x(1-x)(0.18754 + 0.16947x)$$

Si osservi che il calcolo di a_1 ed a_2 con il metodo dei minimi quadrati è decisamente più laborioso di quello del metodo di collocation, a causa degli integrali risultanti dal processo di minimizzazione del funzionale Ω.

Metodo di Galerkin

I coefficienti a_1 ed a_2 vengono determinati imponendo l'ortogonalità del residuo $\mathcal{L}(\hat{u})$ alle n funzioni forma ξ_i utilizzate per descrivere $\hat{u}$:

$$\int_0^1 \mathcal{L}(\hat{u})\xi_i \mathrm{d}x = 0 \qquad i = 1, 2$$

Le funzioni base ξ_i sono le prime due potenze di x moltiplicate per il fattore $x(1-x)$:

$$\begin{aligned} \xi_1 &= x(1-x) \\ \xi_2 &= x^2(1-x) \end{aligned}$$

Scrivendo esplicitamente $\mathcal{L}(\hat{u})$, si ha il sistema lineare:

$$\begin{cases} a_1 \int_0^1 x(1-x)(-2+x-x^2)\mathrm{d}x + a_2 \int_0^1 x(1-x)(2-6x+x^2-x^3)\mathrm{d}x = \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \int_0^1 x(1-x)x\mathrm{d}x \\ \\ a_1 \int_0^1 x^2(1-x)(-2+x-x^2)\mathrm{d}x + a_2 \int_0^1 x^2(1-x)(2-6x+x^2-x^3)\mathrm{d}x = \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \int_0^1 x^2(1-x)x\mathrm{d}x \end{cases}$$

$$\Downarrow$$

$$\begin{pmatrix} 18 & 9 \\ 63 & 52 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 21 \end{pmatrix}$$

la cui soluzione è $a_1 \simeq 0.19241$ e $a_2 \simeq 0.17073$. L'espressione approssimata della soluzione con il metodo di Galerkin è dunque:

$$u_G = x(1-x)(0.19241 + 0.17073x)$$

Si noti che la complessità di calcolo del metodo di Galerkin risulta paragonabile a quella del metodo dei minimi quadrati.

Confronto dei risultati numerici

Confrontiamo a questo punto i risultati numerici ottenuti mediante i tre metodi variazionali con la soluzione analitica sui punti $x = 1/4$, $x = 1/2$ e $x = 3/4$. Si ricava la seguente tabella:

	u	$u_{c,2}$	u_Q	u_G
$x = 1/4$	0.04401	0.04493	0.04311	0.04408
$x = 1/2$	0.06975	0.07143	0.06807	0.06944
$x = 3/4$	0.06006	0.06221	0.05900	0.06009

Si osservi come tutti e tre i metodi variazionali consentano di ottenere approssimazioni soddisfacenti della soluzione vera.

Definendo l'errore come $err = |u - \hat{u}|$, ricaviamo una nuova tabella degli errori sui punti $x = 1/4$, $x = 1/2$ e $x = 3/4$:

	$err_{c,2}$	err_Q	err_G
$x = 1/4$	92e-5	90e-5	7e-5
$x = 1/2$	168e-5	168e-5	31e-5
$x = 3/4$	215e-5	106e-5	3e-5

Il metodo di collocation dà i risultati peggiori, specialmente sul punto $x = 3/4$ nel quale il residuo non è stato annullato.

Il metodo dei minimi quadrati, nonostante il peso computazionale superiore, migliora l'approssimazione ottenuta solo sul terzo punto.

Il metodo di Galerkin si rivela il più accurato.

Sul collocation

Per migliorare il risultato ottenuto con il metodo di collocation possiamo pensare di utilizzare un terzo punto in $x = 3/4$. Poiché si vuole mantenere la medesima approssimazione $\hat{u}$ con soli due coefficienti incogniti a_1 ed a_2, anziché annullare il residuo sui tre punti di collocation minimizziamo il residuo complessivo nel senso nei minimi quadrati. Si vuole, cioè, determinare il valore di a_1 ed a_2 tali da rendere minima l'espressione:

$$\Lambda = [\mathcal{L}(\hat{u})]^2_{x=1/4} + [\mathcal{L}(\hat{u})]^2_{x=1/2} + [\mathcal{L}(\hat{u})]^2_{x=3/4} = \min$$

Calcoliamo l'espressione esplicita dei residui:

$$
\begin{aligned}
\epsilon_1 = \mathcal{L}(\hat{u})_{x=1/4} &= -\frac{29}{16}a_1 + \frac{35}{64}a_2 + \frac{1}{4} \\
\epsilon_2 = \mathcal{L}(\hat{u})_{x=1/2} &= -\frac{7}{4}a_1 - \frac{7}{8}a_2 + \frac{1}{2} \\
\epsilon_3 = \mathcal{L}(\hat{u})_{x=3/4} &= -\frac{29}{16}a_1 + \frac{151}{64}a_2 + \frac{3}{4}
\end{aligned}
$$

Introducendo $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}$, possiamo scrivere

$$
\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix} = \begin{pmatrix} -\frac{29}{16} & \frac{35}{64} \\ -\frac{7}{4} & -\frac{7}{8} \\ -\frac{29}{16} & -\frac{151}{64} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix}
$$

Λ si può dunque scrivere come $\Lambda = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 = \epsilon^T \epsilon$.

Minimizzare Λ equivale a minimizzare il prodotto scalare $\epsilon^T \epsilon$ dove

$$\epsilon^T = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} -\frac{29}{16} & -\frac{7}{4} & -\frac{29}{16} \\ \frac{35}{64} & -\frac{7}{8} & -\frac{151}{64} \end{pmatrix} + \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{pmatrix}$$

Calcoliamo le derivate di $\epsilon^T \epsilon$ rispetto a a_1 e a_2 e uguagliamo a zero. Eseguendo i calcoli otteniamo:

$$\begin{pmatrix} 9.66 & 4.82 \\ 4.82 & 6.61 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 2.69 \\ 2.06 \end{pmatrix}$$

la cui soluzione è $a_1 \simeq 0.19292$ e $a_2 \simeq 0.17204$. L'espressione approssimata della soluzione risulta pertanto:

$$u_{c,3} = x(1-x)(0.19292 + 0.17204x)$$

Confrontiamo quest'ultima approssimazione numerica con la soluzione analitica. Si ricava sui tre punti di collocation:

	u	$u_{c,3}$	$err_{c,3}$
$x = 1/4$	0.04401	0.04424	23e-5
$x = 1/2$	0.06975	0.06974	1e-5
$x = 3/4$	0.06006	0.06037	31e-5

Come atteso, il risultato ottenuto è superiore sia al metodo di collocation con 2 punti che a quello dei minimi quadrati, ed è paragonabile al metodo di Galerkin.