Sapienza Università di Roma
Anno accademico 2016/17

Approssimazione di Galerkin agli elementi finiti stabilizzati per l'equazione di diffusione-trasporto

Candidato:
Simone Casalvieri
matricola n. 633719

Relatore:
Elisabetta Carlini

Master in Calcolo Scientifico

Indice

Introduzione

Nel cuore della maggior parte dei fenomeni di mescolamento si incontra l'equazione di diffusione-trasporto, che descrive il processo fisico di una quantità scalare passiva che diffonde mentre al contempo è agitata e mossa da un flusso. Più in particolare il nostro obiettivo è risolvere numericamente l'equazione di diffusione-trasporto per il profilo di concentrazione di una sostanza in un dominio Ω, avente la seguente espressione:

$$u_t - \frac{1}{\mathbb{P}_e}\Delta u + \vec{b} \cdot \nabla u = 0, \quad \vec{x} \in \Omega,\ t > 0, \tag{1}$$

dove $\vec{b}(\vec{x}, t)$ è il campo vettoriale della velocità di trasporto (a divergenza nulla), t è il tempo e $\mathbb{P}_e$ è il numero di Péclet. Tipicamente si impongono condizioni al bordo di Neumann omogenee (flusso nullo) o periodiche, le quali assicurano che si conservi nel tempo la massa totale della sostanza studiata, $\int_\Omega u(\vec{x}, t) d\vec{x}$.

Nelle applicazioni si incontra una vasta gamma di numeri di Péclet: tipicamente $\mathbb{P}_e \approx 10^2$ per i flussi laminari, $\mathbb{P}_e \approx 10^3 - 10^5$ per coloranti molecolari in soluzioni microfluidiche di acqua/glicerolo, $\mathbb{P}_e \approx 10^5$ per materiali granulari nei tamburi rotanti ed infine $\mathbb{P}_e \approx 10^{10}$ per flussi reattivi turbolenti.

Per risolvere l'equazione (1) utilizziamo un metodo agli elementi finiti di Galerkin ma modificato rispetto a quello standard, che è denominato GLS, acronimo per *Galerkin Least Squares*: come risulterà evidente dall'esposizione che segue, l'idea che è alla base di questo metodo consiste nell'introdurre opportunamente nell'operatore L associato all'equazione di diffusione-trasporto una diffusione artificiale tramite l'inserimento di un parametro δ, al fine di smorzare le eventuali oscillazioni o instabilità a cui è soggetta la soluzione numerica al crescere del numero di Péclet e di conseguenza al diminuire del coefficiente di diffusione μ (visto che $\mu := 1/\mathbb{P}_e$). Una volta poi trasformati con tale tecnica la forma bilineare ed il funzionale lineare associati ad L, la discretizzazione temporale del problema dovuta al fatto che siamo in presenza di un'equazione evolutiva è stata ottenuta con l'ausilio del θ-metodo. Come verrà ampiamente richiamato nel seguito, esso consiste nel discretizzare la derivata temporale per mezzo di un semplice rapporto incrementale e

nel rimpiazzare gli altri termini della formulazione debole del problema mediante una combinazione del valore al tempo t^k e di quello al tempo t^{k+1}, in base al valore scelto per θ ($0 \leq \theta \leq 1$).

La soluzione numerica ottenuta con il metodo GLS verrà poi confrontata con quella ottenuta mediante una tecnica differente, variante del cosiddetto metodo *mapping*, sviluppata dagli autori dell'articolo di riferimento [7], al quale rimandiamo per approfondimenti, non trattata in questo lavoro ma che sostanzialmente fa uso di un operatore detto di *splitting* che consente di trattare indipendentemente le parti diffusive e di trasporto del problema assegnato.

Il presente lavoro è articolato nel modo seguente: nel capitolo 1 saranno presentati alcuni richiami teorici sulla formulazione debole delle equazioni evolutive e sul θ-metodo, con un'analisi dettagliata della sua stabilità e convergenza; nel capitolo 2, dopo ulteriori richiami teorici sulle equazioni di diffusione-trasporto-reazione introduciamo alcuni metodi FE stabilizzati per la risoluzione ed affrontiamo poi in dettaglio il metodo FE GLS, riportandone i risultati di stabilità e convergenza; il capitolo si conclude con la presentazione dei due problemi differenziali test, che vengono affrontati e risolti numericamente tramite il GLS. Infine nel terzo ed ultimo capitolo viene presentato il problema dell'applicazione a cui abbiamo fatto riferimento all'inizio di questa introduzione, vengono riportati i risultati numerici del problema mediante l'utilizzo del metodo GLS e vengono infine confrontati, soprattutto a livello grafico, con i risultati del metodo *mapping* dell'articolo.

Il software utilizzato per effettuare le simulazioni è *Freefem++*, che è uno strumento open-source scaricabile gratuitamente dal sito www.freefem.org, concepito proprio per implementare il metodo degli elementi finiti per la risoluzione di problemi alle derivate parziali e soggetto a continua evoluzione grazie al contributo dei suoi autori (Olivier Pironneau, Jacques Morice, Antoine le Hyaric, Kohji Ohtsuka e Pierre Jolivet) e di altri sviluppatori del settore. A tal proposito riportiamo in tre appendici al termini del lavoro i codici numerici rispettivamente del test 1, tratto dal libro di Quarteroni [3], del test 2 proposto dalla relatrice, dott.ssa Elisabetta Carlini ed infine dell'applicazione che è contenuta nell'articolo di riferimento.

In questo lavoro non viene presentata un'esposizione dettagliata ed esplicita del metodo agli elementi finiti, anche se di volta in volta sono stati inseriti nel testo dei riferimenti ai suoi aspetti principali.

Desidero infine esprimere la mia sincera gratitudine a tutti coloro che hanno preso parte con me all'esperienza del master e che mi hanno sostenuto nella fatica quotidiana: in primo luogo i miei compagni di corso, con i quali ho condiviso molto, riflessioni matematiche ma anche esperienze ed emozioni profonde, che hanno sicuramente reso meno pesante l'impegno di studio,

soprattutto nei momenti di maggiore stress dovuto anche ad impegni lavorativi; mia madre, per le opportunità che mi ha dato da quando sono venuto al mondo e per l'incitamento costante, la mia relatrice, la dottoressa Elisabetta Carlini, per la serietà ed impegno con i quali mi ha accompagnato e per i suoi preziosissimi suggerimenti soprattutto nell'elaborazione del codice numerico; ed infine un ringraziamento speciale a mio fratello Christian, che mi ha sempre aiutato a stemperare le ansie da prestazione, quando stavano per divorarmi, a vedere sempre comunque il bello ed il posiivo nella vita e per aver sempre creduto nelle mie capacità , oltre ad avermi dato un aiuto essenziale nell'elaborazione digitale del rapporto finale di stage.

Sono consapevole di essermi imbarcato in un'impresa alquanto ardua per me seguendo il master, sia per le mie conoscenze pregresse di matematica applicata, sia per altri impegni lavorativi; l'augurio che vorrei fare a chiunque dovesse mai imbattersi in queste note è di cogliere la bellezza e la gioia che derivano dal mettersi costantemente in gioco nella vita e dall'affrontare sempre nuove sfide, sia esse di natura gnoseologica sia più in generale umana.

Capitolo 1

Equazioni paraboliche

Per una trattazione rigorosa e chiara dell'equazione in questione procediamo per gradi e cominciamo richiamando alcuni risultati teorici relativi alle equazioni paraboliche e alla loro risoluzione numerica e alle equazioni di diffusione-trasporto-reazione.

Un'equazione parabolica si presenta in generale nella forma:

$$\frac{\partial u}{\partial t} + Lu = f, \qquad \vec{x} \in \Omega, \;\; t > 0, \tag{1.1}$$

dove Ω è un dominio di $\mathbb{R}^d$, con $d = 1, 2, 3$ ed $f = f(\vec{x}, t)$ è una data funzione, $L = L(\vec{x})$ è un generico operatore ellittico che agisce sull'incognita $u = u(\vec{x}, t)$.

Se risolviamo tale equazione in un certo intervallo di tempo limitato $0 < t < T$, la regione $Q_T = \Omega \times (0, T)$ è detta cilindro nello spazio $\mathbb{R}^d \times \mathbb{R}^+$. L'equazione data deve essere completata assegnando una condizione iniziale

$$u(\vec{x}, 0) = u_0(\vec{x}), \quad \vec{x} \in \Omega, \tag{1.2}$$

e condizioni al bordo, che possiamo esprimere nel modo seguente:

$$u(\vec{x}, t) = \phi(\vec{x}, t), \qquad \vec{x} \in \Gamma_D, \;\; t > 0, \tag{1.3}$$

oppure

$$\frac{\partial u(\vec{x}, t)}{\partial n} = \psi(\vec{x}, t), \qquad \vec{x} \in \Gamma_N, \;\; t > 0, \tag{1.4}$$

dove u_0, ϕ e ψ sono funzioni assegnate e $\{\Gamma_D, \Gamma_N\}$ rappresenta una partizione della frontiera di Ω, cioè $\Gamma_D \cup \Gamma_N = \partial\Omega$ e $\mathring{\Gamma}_D \cap \mathring{\Gamma}_N = \emptyset$. Γ_D è chiamata *frontiera di Dirichlet* e Γ_N *frontiera di Neumann.*

In una dimensione, il problema:

$$\begin{cases} \frac{\partial u}{\partial t} - \nu \frac{\partial^2 u}{\partial x^2} = f, & 0 < x < d, \;\; t > 0, \\ u(x,0) = u_0(x), & 0 < x < d, \\ u(0,t) = u(d,t) = 0, & t > 0 \end{cases} \tag{1.5}$$

descrive l'evoluzione della temperatura $u(x,t)$ in un punto x al tempo t di una barra metallica di lunghezza d che occupa l'intervallo $[0,d]$, di conduttività termica ν, i cui estremi sono tenuti alla temperatura costante di 0 gradi. Il termine u_0 descrive la temperatura iniziale della barra, mentre f rappresenta la produzione di calore (per unità di lunghezza) fornita dalla barra. Per tale ragione all'equazione (1.5) si dà in nome di *equazione del calore*.

1.1 Formulazione debole, sua approssimazione e θ-metodo

Per risolvere il problema (1.1), (1.2), (1.3) o (1.4), introduciamo una *formulazione debole* del problema. A tal fine si moltiplica, per ciascun $t > 0$, l'equazione differenziale (1.1) per una certa funzione test $v(\vec{x})$ e si integra su Ω. Posto $V := H^1_{\Gamma_D}(\Omega)$, per ciascun $t > 0$ cerchiamo una funzione $u(t) \in V$ tale che

$$\int_\Omega \frac{\partial u(t)}{\partial t} v d\Omega + a(u(t), v) = \int_0^t f(t) v d\Omega, \quad \forall v \in V, \tag{1.6}$$

dove $u(0) = u_0$, $a(\cdot,\cdot)$ è la forma bilineare associata all'operatore ellittico L e dove abbiamo posto $\phi = 0$ e $\psi = 0$ per semplicità. Sappiamo che una condizione sufficiente per l'esistenza ed l'unicità della soluzione del problema (1.6) è che valga l'ipotesi di continuità e *debole coercitività* di $a(\cdot,\cdot)$, cioè che

$$\exists \lambda \geq 0, \; \exists \alpha > 0 \; : \; a(v,v) + \lambda ||v||^2_{L^2(\Omega)} \geq \alpha ||v||^2_V, \quad \forall v \in V,$$

e nel caso particolare di $\lambda = 0$ ritroviamo la condizione standard di coercitività. Richiediamo che $u_0 \in L^2(\Omega)$ e che $f \in L^2(\Omega)$. Allora il problema (1.6) ammette un'unica soluzione $u \in L^2(\mathbb{R}^+; V) \cap C^0(\mathbb{R}^+; L^2(\Omega))$, con $V = H_{\Gamma_D}(\Omega)$.

Consideriamo ora l'approssimazione di Galerkin del problema (1.6): per ogni $t > 0$, si deve determinare una $u_h(t) \in V_h$ tale che

$$\int_\Omega \frac{\partial u_h(t)}{\partial t} v_h d\Omega + a(u_h(t), v_h) = \int_\Omega f(t) v_h d\Omega, \quad \forall v_h \in V_h, \tag{1.7}$$

con $u_h(0) = u_{0h}$, dove $V_h \subset V$ è un opportuno spazio di dimensione finita N_h e u_{0h} è una approssimazione conveniente di u_0 nello spazio V_h. Abbiamo

cosí ottenuto il *problema semidiscretizzato* (1.7) del problema originario (1.6), cioè discretizzato solo nella variabile spaziale, dal momento che la variabile temporale non è ancora stata discretizzata. Indicando con $\{\phi_j\}$ una base per V_h osserviamo che è sufficiente che (1.7) sia verificato per ϕ_j perché lo sia per tutte le funzioni del sottospazio. In particolare, dal momento che per ciascun $t > 0$ la soluzione del problema di Galerkin appartiene anch'esso al sottospazio V_h, possiamo scrivere:

$$u_h(\vec{x}, t) = \sum_{j=1}^{N_h} u_j(t)\phi_j(\vec{x}),$$

dove i coefficienti $\{u_j(t)\}$ rappresentano le incognite del problema (1.7). Indicando con $\dot{u}_j(t)$ le derivate della funzione $u_j(t)$ rispetto al tempo, il problema (1.7) diventa:

$$\sum_{j=1}^{N_h} \dot{u}_j(t) \underbrace{\int_\Omega \phi_j\phi_i d\Omega}_{m_{ij}} + \sum_{j=1}^{N_h} u_j(t) \underbrace{a(\phi_j, \phi_i)}_{a_{ij}} = \underbrace{\int_\Omega f(t)\phi_i d\Omega}_{f_i(t)}, \quad i = 1, 2, ..., N_h. \tag{1.8}$$

Se definiamo il vettore delle incognite $\vec{u} = (u_1(t), u_2(t), ..., u_{N_h}(t))^T$, con $M = (m_{ij})$ la *matrice di massa*, con $A = (a_{ij})$ la *matrice di stifness* e con $\vec{f} = (f_1(t), f_2(t), ..., f_{N_h}(t))^T$ il vettore del termine forzante, il sistema (1.8) può essere riscritto più compattamente

$$M\dot{\vec{u}}(t) + A\vec{u}(t) = \vec{f}(t).$$

Per risolvere numericamente tale sistema di equazioni differenziali ordinarie utilizziamo il *θ-metodo*. Esso consiste nel discretizzare la derivata rispetto al tempo per mezzo di un semplice rapporto incrementale e rimpiazzare gli altri termini con una combinazione lineare del valore al tempo t^k e del valore al tempo t^{k+1}, usando il parametro reale θ $(0 \leq \theta \leq 1)$,

$$M\frac{\vec{u}^{k+1} - \vec{u}^k}{\Delta t} + A\left[\theta u^{k+1} + (1-\theta)u^k\right] = \theta f^{k+1} + (1-\theta)f^k. \tag{1.9}$$

Il parametro reale positivo $\Delta t = t^{k+1} - t^k$, $k = 0, 1, ...$ indica il passo di discretizzazione (che per semplicità assumeremo costante), mentre l'apice k denota che la quantità in esame si riferisce al tempo t^k. Alcuni valori particolarmente significativi di θ sono i seguenti:
per $\theta = 0$ otteniamo il metodo di *Eulero in avanti* (o *Eulero esplicito*)

$$M\frac{\vec{u}^{k+1} - \vec{u}^k}{\Delta t} + A\vec{u}^k = \vec{f}^k,$$

che ha un'accuratezza al primo ordine rispetto a Δt;
per $\theta = 1$ otteniamo il metodo di *Eulero all'indietro* (o *Eulero implicito*)

$$M\frac{\vec{u}^{k+1} - \vec{u}^k}{\Delta t} + A\vec{u}^k = \vec{f}^{k+1},$$

anch'esso con un'accuratezza al primo ordine rispetto a Δt;
per $\theta = \frac{1}{2}$ otteniamo il metodo di *Crank-Nicolson* (o *trapezoidale*)

$$M\frac{\vec{u}^{k+1} - \vec{u}^k}{\Delta t} + \frac{1}{2}A\left(\vec{u}^{k+1} + \vec{u}^k\right) = \frac{1}{2}\left(\vec{f}^{k+1} + \vec{f}^k\right),$$

che ha una accuratezza al secondo ordine rispetto a Δt.

Se ora focalizziamo la nostra attenzione sui casi estremi $\theta = 0$ e $\theta = 1$, per entrambi otteniamo un sistema di equazioni lineari: nel primo caso il sistema da risolvere ha matrice $\frac{M}{\Delta t}$, nel secondo caso $\frac{M}{\Delta t} + A$, con M matrice definita positiva e quindi invertibile. Nel caso $\theta = 0$, eseguendo la procedura nota come *mass lumping* sulla matrice di massa (cfr. [3]), possiamo diagonalizzare la matrice M, e quindi disaccoppiare le equazioni del sistema. Tuttavia tale schema non è *incondizionatamente stabile* (cfr. [3]) e, nel caso in cui V_h sia un sottospazio degli elementi finiti, si ottiene la condizione di stabilità (cfr. [3]).

$$\exists c > 0 \ : \ \Delta t < ch^2 \quad \forall h > 0,$$

che non consente una scelta arbitraria di Δt rispetto ad h.

Nel caso $\theta > 0$, il sistema sarà del tipo

$$K\vec{u}^{k+1} = \vec{g}, \tag{1.10}$$

dove $\vec{g}$ rappresenta il termine noto e $K = \frac{M}{\Delta t} + \theta A$. Nel caso in cui la matrice K è invariante nel tempo, se L, e quindi la matrice A, è indipendente dal tempo, e se la *space mesh* non cambia, essa può essere fattorizzata una volta per tutte all'inizio del processo. Dal fatto che M è simmetrica, se anche A lo fosse allora K associata al sistema sarà simmetrica. Da qui, possiamo usare, per esempio, la fattorizzazione di Cholesky $K = H \cdot H^T$, essendo H triangolare inferiore per risolvere il sistema (1.10). In ciascun passo temporale dovremo perciò risolvere due sistemi triangolari in N_h incognite (cfr. [3] e [4]):

$$H\vec{y} = \vec{g}$$

$$H^T\vec{u}^{k+1} = \vec{y}.$$

1.2 Stime a priori

Dato il problema (1.6), osserviamo che, poichè le corrispondenti equazioni devono valere per ogni $v \in V$, sarà in particolare lecito porre $v = u(t)$ (fissato t), soluzione di (1.6):

$$\underbrace{\int_\Omega \frac{\partial u(t)}{\partial t} d\Omega}_{(1)} + \underbrace{a(u(t), u(t))}_{(2)} = \underbrace{\int_\Omega f(t)u(t)d\Omega}_{(3)}, \quad \forall t > 0. \tag{1.11}$$

In particolare il termine (1) in (1.11) può essere riscritto nel modo seguente:

$$\frac{1}{2}\frac{\partial}{\partial t}\int_\Omega |u(t)|^2 d\Omega = \frac{1}{2}\frac{\partial}{\partial t}||u(t)||^2_{L^2(\Omega)}. \tag{1.12}$$

Se per semplicità assumiamo che la forma bilineare sia coerciva (con costante di coercitività pari ad α), si ha che

$$a(u(t), u(t)) \geq \alpha ||u(t)||^2_V,$$

mentre infine, usando la disuguaglianza di Cauchy-Schwarz, per (3) in (1.11) si ha:

$$(f(t), u(t)) \leq ||f(t)||_{L^2(\Omega} \cdot ||u(t)||_{L^2(\Omega)}. \tag{1.13}$$

dove $(f, u) = \int_\Omega f u d\Omega$ è il prodotto scalare in $L^2(\Omega)$. Nei calcoli seguenti faremo uso della disuguaglianza di Young:

$$\forall a,\ b \in \mathbb{R}; \qquad ab \leq \epsilon a^2 + \frac{1}{4\epsilon}b^2, \quad \forall \epsilon > 0,$$

che deriva dalla disuguaglianza elementare:

$$\left(\sqrt{\epsilon} - \frac{1}{2\sqrt{\epsilon}}b\right)^2 \geq 0.$$

Usando dapprima la disuguaglianza di Poincaré e poi la disuguaglianza di Young, otteniamo:

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial t}||u(t)||^2_{L^2(\Omega)} + \alpha||\nabla u(t)||^2_{L^2(\Omega)} &\leq ||f(t)||_{L^2(\Omega)} \cdot ||u(t)||_{L^2(\Omega)} \\ &\leq \frac{C^2_\Omega}{2\alpha}||f(t)||^2_{L^2(\Omega)} + \frac{\alpha}{2}||\nabla u(t)||^2_{L^2(\Omega)} \end{aligned} \tag{1.14}$$

Integrando nel tempo otteniamo, $\forall t > 0$, la cosiddetta *prima stima a priori* dell'energia:

$$||u(t)||^2_{L_2(\Omega)} + \alpha \int_0^t ||\nabla u(s)||^2_{L^2(\Omega)} ds \leq ||u_0||^2_{L^2(\Omega)} + \frac{C^2_\Omega}{2}\int_0^t ||f(s)||^2_{L^2(\Omega)} ds.$$

Altre stime a priori differenti si possono ottenere con il seguente ragionamento. Osserviamo che

$$\frac{1}{2}\frac{d}{dt}||u(t)||^2_{L^2(\Omega)} = ||u(t)||_{L^2(\Omega)} \cdot \frac{d}{dt}||u(t)||_{L^2(\Omega)}.$$

Allora, usando le (1.11), (1.12) e (1.13), ancora sfruttando la disuguaglianza di Poincaré, otteniamo, per $t > 0$, che

$$||u(t)||_{L^2(\Omega}\cdot\frac{d}{dt}||u(t)||_{L^2(\Omega)}+\frac{\alpha}{c_\Omega}||u(t)||_{L^2(\Omega)}\cdot||\nabla u(t)||_{L^2(\Omega)} \leq ||f(t)||_{L^2(\Omega)}||u(t)||_{L^2(\Omega)}.$$

Se $||u(t)||_{L^2(\Omega)} \neq 0$ possiamo dividere tutto per $||u(t)||_{L^2(\Omega)}$ ed integrare nel tempo per ottenere la *seconda stima a priori*:

$$||u(t)||_{L^2(\Omega)} \leq ||u_0||_{L^2(\Omega)} + \int_0^t ||f(s)||_{L^2(\Omega)}ds, \quad t > 0,$$

(se $||u(t)||_{L^2(\Omega)}$ fosse stata nulla avremmo dovuto procedere in modo differente, ma alla fine il risultato sarebbe stato lo stesso).

Usiamo ora la prima disuguaglianza in (1.14) ed integriamo nel tempo per ricavare:

$$\begin{aligned}
&||u(t)||^2_{L^2(\Omega)} + 2\alpha\int_0^t ||\nabla u(s)||^2 ds \leq ||u_0||^2_{L^2(\Omega)} + 2\int_0^t ||f(s)||_{L^2(\Omega)} \cdot ||u(s)||^2_{L^2(\Omega)} ds \\
\leq\ & ||u_0||^2_{L^2(\Omega)} + 2\int_0^t ||f(s)||_{L^2(\Omega)} \cdot \left(||u_0||^2_{L^2(\Omega)} + \int_0^s ||f(\tau)||_{L^2(\Omega)}d\tau\right) ds \\
=\ & ||u_0||^2_{L^2(\Omega)} + 2\int_0^t ||f(s)||_{L^2(\Omega)} \cdot ||u_0||_{L^2(\Omega)} + 2\int_0^t ||f(s)||_{L^2(\Omega)} \int_0^s ||f(\tau)||_{L^2(\Omega)}d\tau \\
=\ & \left(||u_0||_{L^2(\Omega)} + \int_0^t ||f(s)||ds\right)^2. \qquad (1.15)
\end{aligned}$$

L'ultima uguaglianza segue notando che

$$||f(s)||_{L^2(\Omega)} \cdot \int_0^s ||f(\tau)||_{L^2(\Omega)}d\tau = \frac{d}{ds}\left(\int_0^s ||f(\tau)||_{L^2(\Omega)}d\tau\right)^2.$$

Perciò possiamo concludere con la *terza stima a priori*:

$$\left(||u(t)||^2_{L^2(\Omega)} + 2\alpha\int_0^t ||\nabla u(s)||^2_{L^2(\Omega)} ds\right)^{\frac{1}{2}} \leq ||u_0||_{L^2(\Omega)}+\int_0^t ||f(s)||_{L^2(\Omega)}ds, \quad t > 0.$$

Abbiamo visto che possiamo formulare il problema di Galerkin (1.7) per il problema (1.8) e che esso, sotto opportune ipotesi, ammette un'unica soluzione.

In modo analogo a quanto fatto per il problema (1.8) possiamo provare le seguenti stime a priori (di stabilità) per la soluzione del problema (1.7):

$$||u_h(t)||^2_{L^2(\Omega)}+\alpha \int_0^t ||\nabla u_h(s)||^2_{L^2(\Omega)} \leq ||u_{0h}(t)||^2_{L^2(\Omega)}+\frac{C^2_\Omega}{\alpha}\int_0^t ||f(s)||^2_{L^2(\Omega)}ds, \quad t>0. \tag{1.16}$$

Per dimostrarle possiamo prendere, per ogni $t > 0$, $v_h = u_h(t)$ e procedere esattamente come abbiamo fatto pe ottenere la (1.14). Successivamente, ricordando che il dato iniziale è $u_h(0) = u_{0h}$, possiamo infine dedurre le seguenti controparti discrete della seconda e terza stima a priori:

$$||u_h(t)||^2_{L^2(\Omega)} \leq ||u_{0h}(t)||^2_{L^2(\Omega)} + \int_0^t ||f(s)||_{L^2(\Omega)}ds, \quad t>0,$$

e

$$\left(||u_h(t)||^2_{L^2(\Omega)} + 2\alpha \int_0^t ||\nabla u_h(s)||^2_{L^2(\Omega)}ds\right)^{\frac{1}{2}} \leq ||u_{0h}||_{L^2(\Omega)}+\int_0^t ||f(s)||_{L^2(\Omega)}ds, \quad t>0.$$

1.3 Analisi di convergenza del problema semidiscreto

Ora riprendiamo il problema (1.6) e la sua approssimazione numerica (1.8). Vogliamo perciò condurre un'analisi di convergenza del problema semidiscreto e dimostrare la convergenza di u_h ed u in opportune norme. Dall'ipotesi di coercitività possiamo scrivere

$$\alpha||u-u_h||^2_{H^1(\Omega)} \leq \alpha(u-u_h,\, u-u_h) = a(u-u_h,\, u-v_h)+a(u-u_h,\, v_h-u_h), \quad \forall v_h \in V_h.$$

Sottraendo l'equazione (1.7) dall'equazione (1.8) e ponendo $w_h = v_h - u_h$ otteniamo:

$$\left(\frac{\partial(u-u_h)}{\partial t},\, w_h\right) + a(u-u_h,\, w_h) = 0.$$

Allora

$$\alpha||u-u_h||^2_{H^1(\Omega)} \leq \underbrace{a(u-u_h,\, u-v_h)}_{(1')} - \underbrace{\left(\frac{\partial(u-u_h)}{\partial t},\, w_h\right)}_{(2')}. \tag{1.17}$$

Analizziamo i due termini di destra separatamente. Per $(1')$, usando la contintuità di $a(\cdot,\cdot)$ e la disuguaglianza di Young, otteniamo

$$a(u-u_h, u-v_h) \leq M||u-u_h||_{H^1(\Omega)}\cdot||u-v_h||_{H^1(\Omega)} \leq \frac{\alpha}{2}||u-u_h||^2_{H^1(\Omega)}+\frac{M^2}{2\alpha}||u-v_h||^2_{H^1(\Omega)};$$

per $(2')$, scrivendo w_h nella forma $w_h = (v_h - u) + (u - u_h)$, otteniamo:

$$-\left(\frac{\partial(u-u_h)}{\partial t},\, w_h\right) = \left(\frac{\partial(u-u_h)}{\partial t},\, u - v_h\right) - \frac{1}{2}\frac{d}{dt}||u-u_h||^2_{L^2(\Omega)}.$$

Sostituendo tali risultati nella (1.17) si ha che

$$\frac{1}{2}\frac{d}{dt}||u-u_h||^2_{L^2(\Omega)} + \frac{\alpha}{2}||u-u_h||^2_{H^1(\Omega)} \leq \frac{M^2}{2\alpha}||u-v_h||^2_{H^1(\Omega)} + \left(\frac{\partial(u-u_h)}{\partial t},\, u - v_h\right).$$

Moltiplicando poi entrambi i membri per 2 ed integrando nel tempo tra 0 e t troviamo:

$$\begin{aligned} &||(u-u_h)(t)||^2_{L^2(\Omega)} + \alpha\int_0^t ||(u-u_h)(s)||^2_{H^1(\Omega} ds \leq ||(u-u_h)(0)||^2_{L^2(\Omega)} \\ +\ &\frac{M^2}{\alpha}\int_0^t ||u(s)-v_h||^2_{H^1(\Omega)} ds + 2\int_0^t \left(\frac{\partial}{\partial t}(u-u_h)(s),\, u(s)-v_h\right) ds \end{aligned} \tag{1.18}$$

Integrando per parti ed utilizzando la disuguaglianza di Young otteniamo:

$$\begin{aligned} &\int_0^t \left(\frac{\partial}{\partial t}(u-u_h)(s),\, u(s)-v_h\right) ds = -\int_0^t \left((u-u_h)(s),\, \frac{\partial}{\partial t}(u(s)-v_h\right) ds \\ +\ &((u-u_h)(t),\, (u-v_h)(t)) - ((u-u_h)(0),\, (u-v_h)(0)) \leq \frac{1}{4}\int_0^t ||(u-u_h)(s)||^2_{L^2(\Omega)} ds \\ +\ &\int_0^t \left|\left|\frac{\partial(u(s)-v_h)}{\partial t}\right|\right|^2_{L^2(\Omega)} ds + \frac{1}{4}||(u-u_h)(t)||^2_{L^2(\Omega)} + ||(u-v_h)(t)||^2_{L^2(\Omega} \\ +\ &||(u-u_h)(0)||_{L^2(\Omega)} \cdot ||(u-u_h)(0)||_{L^2(\Omega)}. \end{aligned} \tag{1.19}$$

Dalla (2.14) perciò ricaviamo

$$\begin{aligned} &\frac{1}{2}||(u-u_h)(t)||^2_{L^2(\Omega)} + \alpha\int_0^t ||(u-u_h)(s)||^2_{H^1(\Omega)} \\ \leq\ &\frac{M^2}{\alpha}\int_0^t ||u(s)-v_h||^2_{H^1(\Omega)} ds + 2\int_0^t \left|\left|\frac{\partial(u(s)-v_h)}{\partial t}\right|\right|^2_{L^2(\Omega)} \\ +\ &2||(u-u_h)(t)||^2_{L^2(\Omega)} + ||(u-u_h)(0)||^2_{L^2(\Omega)} \\ +\ &2||(u-u_h)(0)||^2_{L^2(\Omega)} \cdot ||(u-v_h)(0)||_{L^2(\Omega)} + \frac{1}{2}\int_0^t ||(u-u_h)(s)||^2_{L^2(\Omega)} ds \end{aligned} \tag{1.20}$$

Supponiamo ora che V_h sia lo spazio degli elementi finiti di grado r, più precisamente $V_h = \left\{v_h \in X_h^r \quad v_h|_{\Gamma_D} = 0\right\}$ e scegliamo, per ciascuna t, $v_h =$

$\Pi_h^r u(t)$, l'interpolante di $u(t)$ in V_h (vedi paragrafo 4.20 in [3]). Grazie alla stima

$$|v - \Pi_h^r v|_{H^m(\Omega)} \leq C \cdot h^{r+1-m} |v|_{H^1(\Omega)} \qquad \forall v \in H^{r+1}(\Omega)$$

(cit), abbiamo, assumendo che u sia sufficientemente regolare, che

$$h \left\| u - \Pi_h^r u \right\| + \left\| u - \Pi_h^r u \right\|_{L^2(\Omega)} \leq C_2 h^{r+1} |u|_{H^{r+1}(\Omega)}.$$

Da ciò segue che gli addendi del membro di destra della disuguaglianza (1.20) sono limitati come segue:

$$\begin{aligned}
E_1 &= \frac{M}{\alpha^2} \int_0^t \|u(s) - v_h\|^2_{H^1(\Omega)} \, ds \leq C_1 \cdot h^{2r} \int_0^t |u(s)|^2_{H^{r+1}(\Omega)} ds, \\
E_2 &= 2 \int_0^t \left\| \frac{\partial (u - v_h)}{\partial t}(s) \right\|_{L^2(\Omega)} ds \leq C_2 \cdot h \int_0^t \left| \frac{\partial u(s)}{\partial t} \right|^2_{H^r(\Omega)} ds, \\
E_3 &= 2 \left\| (u - v_h)(t) \right\|^2_{L^2(\Omega)} \leq C_3 \cdot h^{2r} |u|^2_{H^r(\Omega)}, \\
E_4 &= \left\| (u - u_h)(0) \right\|^2_{L^2(\Omega)} + 2 \left\| (u - u_h)(0) \right\|_{L^2(\Omega)} \cdot \left\| (u - v_h)(0) \right\|_{L^2(\Omega)} \leq C_4 \cdot h^{2r} |u(0)|^2_{H^r(\Omega)}.
\end{aligned}$$

Di conseguenza,

$$E_1 + E_2 + E_3 + E_4 \leq C \cdot h^{2r} \cdot N(u),$$

dove $N(u)$ è un'opportuna funzione che dipende da u e da $\frac{\partial u}{\partial t}$. Procedendo il tal modo otteniamo la disuguaglianza

$$\frac{1}{2} \left\| (u - u_h)(t) \right\|^2_{L^2(\Omega)} + \alpha \int_0^t \left\| (u - u_h)(s) \right\|^2_{H^2(\Omega)} \leq C \cdot H^{2r} \cdot N(u) + \frac{1}{2} \int_0^t \left\| (u - u_h)(s) \right\|^2_{L^2(\Omega)},$$

ed infine, applicando il lemma di Gronwall:

Lemma 1.1. *Sia $A \in L^1(t_0, T)$ una funzione non negativa, g e ϕ due funzioni continue in $[t_0, T]$. Se ϕ è tale che*

$$\phi(t) \leq g(t) + \int_{t_0}^t A(\tau)\phi(\tau) d\tau, \quad \forall t \in [t_0, T],$$

allora se g è non decrescente,

$$\phi(t) \leq g(t) \exp\left(\int_{t_0}^t A(\tau) d\tau \right) \quad \forall t \in [t_0, T].$$

possiamo scrivere una stima a priori dell'errore per ciascun $t > 0$:

$$\left\| u(t) - u_h(t) \right\|^2_{L^2(\Omega)} + 2\alpha \int_0^t \left\| u(s) - u_h(s) \right\|^2_{H^2(\Omega)} \leq C \cdot h^{2r} \cdot N(u) \cdot e^t. \quad (1.21)$$

Utilizzando una tecnica dimostrativa differente che non ricorre al lemma di Gronwall si può arrivare ad una stima dell'errore simile alla (1.21) nella quale non compare il fattore esponenziale a secondo membro:

$$\begin{aligned} &\|u(t) - u_h(t)\|^2_{L^2(\Omega)} + \alpha \int_0^t \|\nabla u(s) - \nabla u_h(s)\|^2_{L^2(\Omega)} \\ \leq\ & Ch^{2r} \left(|u_0|^2_{H^r(\Omega)} + \int_0^t |u(s)|^2_{H^{r+1}(s)} ds + \int_0^t \left| \frac{\partial u(s)}{\partial t} \right|^2_{H^{r+1}(\Omega)} ds \right). \end{aligned}$$

Per i dettagli sulla tecnica dimostrativa per arrivare alla (1.22) (vedi, Sezione 5.3 in [3]; ulteriori stime dell'errore sono dimostrate ad esempio in [5]).

1.4 Analisi di stabilità dei θ-metodi

Analizziamo ora la stabilità del problema *fully discretized*. Applicando il θ-metodo al problema di Galerkin (1.8) otteniamo:

$$\left(\frac{u_h^{k+1} - u_h^k}{\Delta t},\, v_h \right) + a\left(\theta u_h^{k+1} + (1-\theta) u_h^k,\, v_h\right) = \theta f^{k+1}(v_h) + (1-\theta) f^k(v_h), \quad \forall v_h \in V_h,\, \theta \in [0,1] \tag{1.22}$$

per ciascun $k \geq 0$, con $u_h^0 = u_{0h}$, u_h^k la soluzione u_h calcolata al tempo t^k; f^k indica che il funzionale è valutato al tempo t^k.

Nell'analisi che segue per semplicità ci limitiamo al caso in cui $f = 0$ e cominciamo considerando il metodo di Eulero implicito ($\theta = 1$), cioè

$$\left(\frac{u_h^{k+1} - u_h^k}{\Delta t},\, v_h \right) + a\left(u_h^{k+1},\, v_h\right) = 0 \quad \forall v_h \in V_h.$$

Scegliendo $v_h = u_h^{k+1}$, otteniamo

$$\left(u_h^{k+1},\, u_h^{k+1}\right) + \Delta t a\left(u_h^{k+1},\, u_h^{k+1}\right) = (u_h^k,\, u_h^{k+1}).$$

Sfruttando la coercitività della forma bilineare $a(\cdot,\cdot)$ si ha:

$$a(u_h^{k+1},\, u_h^{k+1}) \geq \alpha \left\| u_h^{k+1} \right\|^2_V, \tag{1.23}$$

mentre utilizzando la disuguaglianza di Cauchy-Schwarz e di Young, ricaviamo:

$$(u_h^k,\, u_h^{k+1}) \leq \frac{1}{2} \left\| u_h^k \right\|^2_{L^2(\Omega)} + \frac{1}{2} \left\| u_h^{k+1} \right\|^2_{L^2(\Omega)}. \tag{1.24}$$

Usando sia la (1.23) che la (1.24) otteniamo:

$$\left\| u_h^{k+1} \right\|_{L^2(\Omega)} + 2\alpha \Delta t \left\| u_h^{k+1} \right\|^2_V \leq \left\| u_h^2 \right\|^2_{L^2(\Omega)}. \tag{1.25}$$

Sommando sia sull'indice k da 0 a $n-1$ deduciamo che

$$\left\|u_h^k\right\|_{L^2(\Omega)}^2 + 2\alpha\Delta t \sum_{k=0}^{n-1} \left\|u_h^{k+1}\right\|_V^2 \leq \|u_0 h\|_{L^2(\Omega)}^2 .$$

Nel caso in cui $f \neq 0$, usando il lemma di Gronwall discreto:

Lemma 1.2. *Sia k_n una successione non negativa e ϕ_n tale che*

$$\phi_0 \leq g_0,\ \phi_n \leq g_0 + \sum_{m=0}^{n-1} p_m + \sum_{m=0}^{n-1} k_m \phi_m, \quad n \geq 1.$$

Se $g_0 \geq 0$ and $p_m \geq 0$ for $m \geq 0$, allora

$$\phi \leq \left(g_0 + \sum_{m=0}^{n-1} p_m \right) \exp\left(\sum_{m=0}^{n-1} k_m \right), \quad n \geq 1$$

si può provare similmente che

$$\|u_h^n\|_{L^2(\Omega)}^2 + 2\alpha\Delta t \sum_{k=1}^{n} \left\|u_h^k\right\|_V^2 \leq C(t^n) \left(\|u_{0h}\|_{L^2(\Omega)}^2 + \sum_{k=1}^{n} \Delta t \left\|f^k\right\|_{L^2(\Omega} \right). \tag{1.26}$$

Tale relazione è simile alla (1.16), eccetto per il fatto che gli integrali $\int_0^t ds$ sono approssimati mediante una formula di integrazione numerica composta con passo Δt.

Infine, osservando che $\left\|u_h^{k+1}\right\|_V \geq \left\|u_h^{k+1}\right\|_{L^2(\Omega)}$, dalla (1.25) deduciamo che, per ciascun dato $\Delta t > 0$,

$$\lim_{k \to +\infty} \left\|u_h^k\right\|_{L^2(\Omega)} = 0,$$

il che significa che il metodo di Eulero all'indietro è *assolutamente stabile* senza alcuna restrizione sul passo temporale Δt.

Prima di analizzare il caso generale in cui θ è un parametro qualunque compreso tra 0 ed 1, introduciamo la seguente:

Definizione 1.1. *Lo scalare λ si dice essere un autovalore della forma bilineare $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ e $w \in V$ si dice essere il suo corrispondente autovettore se risulta*

$$a(w, v) = \lambda(w, v), \quad \forall v \in V.$$

Se la forma bilineare $a(\cdot,\cdot)$ è simmetrica e coerciva ha autovalori tutti reali e positivi, che formano una successione infinita; inoltre le sue autofunzioni formano una base dello spazio V.

Gli autovalori e le autofunzioni di $a(\cdot,\cdot)$ possono essere approssimate trovando delle coppie $\lambda_h \in \mathbb{R}$ e $w_h \in V_h$ che soddisfano

$$a(w_h, v_h) = \lambda_h(w_h, v_h), \quad \forall v_h \in V_h. \tag{1.27}$$

Da un punto di vista algebrico il problema (1.27) può essere riformulato come segue:

$$A\vec{w} = \lambda_h M\vec{w},$$

essendo A la matrice di stiffness ed M la matrice di massa. Questo problema è detto *problema agli autovalori generalizzato*. Tali autovalori sono tutti positivi e tanti quanti N_h, cioè la dimensione del sottospazio V_h; dopo averli individuati in senso ascendente $\lambda_h^1 \leq \lambda_h^2 \leq ... \leq \lambda_h^{N_h}$ si ha

$$\lambda_h^{N_h} \to \infty \quad \text{per } N_h \to \infty.$$

Inoltre le corrispondenti autofunzioni formano una base del sottospazio V_h e possono essere scelte *ortonormali* rispetto al prodotto scalare di $L^2(\Omega)$, cioè

$$(w_h^i, w_h^j) = \delta_{ij}, \quad \forall i,j = 1, ..., N_h,$$

essendo w_h^i e w_h^j le autofunzioni corrispondenti a λ_h^i e λ_h^j. Perciò ciascuna funzione $v_h \in V_h$ si può rappresentare come segue:

$$v_h(\vec{x}) = \sum_{j=1}^{N_h} v_j w_h^j(\vec{x})$$

e grazie all'ortonormalità delle autofunzioni,

$$\|v_h\|^2_{L^2(\Omega)} = \sum_{j=1}^{N_h} v_j^2. \tag{1.28}$$

Consideriamo ora un arbitrario $\theta \in [0,1]$ e limitiamoci al caso in cui la forma bilineare sia *simmetrica* (il che ci assicura che le autofunzioni formano una base e quindi non sarebbe applicabile la dimostrazione che segue, sebbene risulterebbe ancora valida in generale la stabilità finale). Siano $\{w_h^i\}$ le autofunzioni discrete ortonormali di $a(\cdot,\cdot)$. Dal momento che $u_h^k \in V_h$, possiamo scrivere

$$u_h^k(\vec{x}) = \sum_{j=1}^{N_h} u_j^k w_h^j(\vec{x}).$$

Osserviamo che in tale espressione i coefficienti u_j^k non rappresentano più i valori di u_h^k nei nodi. Se ora poniamo $f = 0$ nella (1.22) e prendiamo $v_h = w_h^i$, troviamo

$$\frac{1}{\Delta t}\sum_{j=1}^{N_h}\left[u_j^{k+1} - u_j^k\right](w_h^i, w_h^j) + \sum_{j=1}^{N_h}\left[\theta u_j^{k+1} + (1-\theta)u_j^k\right]a(w_h^j, w_h^i) = 0,$$

per ciascun $i = 1, ..., N_h$. Per ciascuna coppia $i, j = 1, ..., N_h$ abbiamo:

$$a(w_h^i, w_h^j) = \lambda_h^j(w_h^j, w_h^i) = \lambda_h^j\delta_{ij} = \lambda_h^i,$$

e perciò per ciascun indice $i = 1, ..., N_h$,

$$\frac{u_i^{k+1} - u_i^k}{\Delta t} + \left[\theta u_i^{k+1} + (1-\theta)u_i^k\right]\lambda_h^i = 0.$$

Risolvendo rispetto a u_i^{k+1}, troviamo:

$$u_i^{k+1} = u_i^k \cdot \frac{1 - (1-\theta)\lambda_h^i\Delta t}{1 + \theta\lambda_h^i\Delta t}.$$

Richiamando la (1.28) concludiamo che, affinché il metodo sia assolutamente stabile, deve valere la disuguaglianza

$$\left|\frac{1 - (1-\theta)\lambda_h^i\Delta t}{1 + \theta\lambda_h^i\Delta t}\right| < 1,$$

cioè

$$-1 - \theta\lambda_h^i\Delta t < 1 - (1-\theta)\lambda_h^i\Delta t < 1 + \theta\lambda_h^i\Delta t.$$

Da qui,

$$-\frac{2}{\lambda_h^i\Delta t} - \theta < \theta - 1\theta.$$

La seconda disuguaglianza è sempre verificata, mentre la prima può essere riscritta nel seguente modo

$$2\theta - 1 > -\frac{2}{\lambda_h^i\Delta t}.$$

Se $\theta \geq \frac{1}{2}$ il membro di sinistra è non negativo, mentre il lato destro è negativo e perciò la disuguaglianza vale per ogni Δt. Se invece $\theta < \frac{1}{2}$ la disuguaglianza (e perciò la stabilità) è soddisfatta solo se

$$\Delta t < \frac{2}{(1-2\theta)\lambda_h^i}.$$

Poichè tale relazione deve valere per tutti gli autovalori λ_h^i della forma bilineare, sarà sufficiente richiedere che essa valga per il più grande tra di essi, che abbiamo supposto essere $\lambda_h^{N_h}$.

Per sintetizzare abbiamo:
-se $\theta \geq \frac{1}{2}$, il θ-metodo è *incondizionatamente stabile*, cioè stabile $\forall \Delta t$;
-se $\theta < \frac{1}{2}$, il θ-metodo è *stabile* solo per

$$\Delta t \leq \frac{2}{(1-2\theta)\lambda_h^{N_h}}.$$

In virtù della definizione di autovalore (1.27) e della continuità della $a(\cdot,\cdot)$ deduciamo che

$$\lambda_h^{N_h} = \frac{a\left(w_{N_h}, w_{N_h}\right)}{\|w_{N_h}\|^2_{L^2(\Omega)}} \leq \frac{M\|w_{N_h}\|^2_V}{\|w_{N_h}\|^2_{L^2(\Omega)}} \leq M\left(1 + C^2h^{-2}\right).$$

In particolare, la costante $C > 0$ che compare nell'ultima disuguaglianza deriva dalla seguente *disuguaglianza inversa*:

$$\exists C > 0 : \quad \|\nabla v_h\|_{L^2(\Omega)} \leq Ch^{-1}\|v_h\|_{L^2(\Omega)}, \quad \forall v_h \in V_h,$$

per i cui dettagli si rimanda al capitolo 3 di [5].

Prendendo allora h sufficientemente piccolo si ha che

$$\lambda_h^{N_h} \leq Ch^{-2}.$$

Infatti, possiamo provare che $\lambda_h^{N_h}$ è in effetti di ordine h^{-2}, cioè

$$\lambda_h^{N_h} = \max_i \lambda_h^i \approx Ch^{-2}.$$

Tenendo presente ciò otteniamo che, nel caso $\theta < \frac{1}{2}$, il metodo è assolutamente stabile solo se

$$\Delta t \leq C(\theta)h^2,$$

dove $C(\theta)$ indica una costante positiva che dipende da θ. L'ultima relazione implica che, per $\theta < \frac{1}{2}$, Δt non può essere scelto in modo arbitrario ma è limitato dalla scelta di h.

1.5 Analisi di convergenza dei θ-metodi

Vogliamo concentrare infine la nostra attenzione sull'analisi di convergenza del θ-metodo. A tal proposito vale il seguente:

Teorema 1.1. *Sotto l'ipotesi che u_0, f e la soluzione esatta siano sufficientemente regolari, vale la seguente stima a priori dell'errore:*

$$\forall n \geq 1 \quad \|u(t^n) - u_h^n\|^2_{L^2(\Omega)} + 2\alpha\Delta t \sum_{k=1}^{n} \left\|u(t^k) - u_h^k\right\|^2_V \leq C(u_0, f, u)\cdot\left(\Delta t^{p(\theta)} + h^{2r}\right),$$

dove $p(\theta) = 2$ se $\theta \neq \frac{1}{2}$, $p\left(\frac{1}{2}\right) = 4$ e C dipende dai suoi argomenti ma non da h e da Δt.

Dimostrazione 1.1. *La dimostrazione si ottiene confrontando la soluzione del problema fully discretized (1.22) con la soluzione del problema semidiscreto (1.8), usando il risultato di stabilità (1.26) cosí come il tasso di decadimento dell'errore di troncamento della discretizzazione del tempo. Per semplicità ci limitiamo a considerare il metodo di Eulero all'indietro (corrispondente a $\theta = 1$)*

$$\frac{1}{\Delta t}\left(u_h^{k+1} - u_h^k, v_h\right) + a(u_h^{k+1}, v_h) = (f^{k+1}, v_h), \quad \forall v_h \in V_h. \tag{1.29}$$

Per la dimostrazione nel caso più generale si consulti (cit.)

Sia $\Pi^r_{1,h}$ l'operatore di proiezione ortogonale seguente:

$$\Pi^r_{1,h} : V \to V_h \ : \ \forall w \in V,\ a\left(\Pi^r_{1,h} w - w, v_h\right) = 0, \quad \forall v_h \in V_h. \tag{1.30}$$

(Si ricordi che tale operatore è definito supponendo che $a(\cdot,\cdot)$ sia simmetrica). Allora

$$\left\|u(t^k) - u_h^k\right\|_{L^2(\Omega)} \leq \left\|u(t^k) - \Pi^r_{1,h} u(t^k)\right\|_{L^2(\Omega)} + \left\|\Pi^r_{1,h} u(t^k) - u_h^k\right\|_{L^2(\Omega)}. \tag{1.31}$$

Il primo termine si può stimare per mezzo della seguente disuguaglianza (per un approfondimento vedi la sezione 3.5 in [5]):

$$\begin{aligned} &\forall w \in V \cap H^{r+1}(\Omega)\ \exists C > 0 \ : \\ &\left\|\Pi^r_{1,h} w - w\right\|_{H^1(\Omega)} + h^{-1}\left\|\Pi^r_{1,h} w - w\right\|_{L^2(\Omega)} \leq C h^p |w|_{H^{p+1}(\Omega)}, \quad 0 \leq p \leq r \end{aligned} \tag{1.32}$$

Al fine di analizzare il secondo termine, definendo $\epsilon_h^k := u_h^k - \Pi^r_{1,h} u(t^k)$, otteniamo

$$\frac{1}{\Delta t}\left(\epsilon_h^{k+1} - \epsilon_h^k, v_h\right) + a\left(\epsilon_h^{k+1}, v_h\right) = \left(\delta^{k+1}, v_h\right), \quad \forall v_h \in V_h, \tag{1.33}$$

avendo posto

$$\left(\delta^{k+1}, v_h\right) = \left(f^{k+1}, v_h\right) - \frac{1}{\Delta t}\left(\Pi^r_{1,h}\left(u(t^{k+1}) - u(t^k)\right), v_h\right) - a\left(u(t^{k+1}), v_h\right) \tag{1.34}$$

ed avendo esplicitato nell'ultimo addendo l'ortogonalità (1.30) dell'operatore $\Pi^r_{1,h}$. La successione $\left\{\epsilon^k_h,\ k=0,1,...\right\}$ soddisfa il problema (1.33) che è simile al problema (1.29) (eccetto che abbiamo sostituito f^{k+1} con δ^{k+1}). Adattando la stima di stabilità (1.26) otteniamo, per ciascun $n \geq 1$, che

$$\|\epsilon^n_h\|^2_{L^2(\Omega)} + 2\alpha\Delta t \sum_{k=1}^{n} \left\|\epsilon^k_h\right\|^2_V \leq C(t^n) \cdot \left(\left\|\epsilon^0_h\right\|^2_{L^2(\Omega)} + \sum_{k=1}^{n} \Delta t \left\|\delta^k\right\|^2_{L^2(\Omega)} \right). \tag{1.35}$$

Si può facilmente stimare la norma associata al livello di tempo iniziale, per esempio se $u_{0h} = \Pi^r_h u_0$ è l'elemento finito interpolante di u_0, utilizzando in modo opportuno le stime:

$$|v - \Pi^r_h(v)|_{H^m(\Omega)} \leq C \cdot h^{r+1-m} |v|_{H^{r+1}(\Omega)} \quad \forall v \in H^{r+1}(\Omega) \tag{1.36}$$

e la (1.32), otteniamo:

$$\left\|\epsilon^0_h\right\|_{L^2(\Omega)} = \left\|u_{0h} - \Pi^r_{1,h} u_0\right\| \leq \|\Pi^r_h u_0 - u_0\|_{L^2(\Omega)} + \left\|u_0 - \Pi^r_{1,h}\right\|_{L^2(\Omega)} \leq C h^r |u_0|_{H^r(\Omega)}. \tag{1.37}$$

Focalizziamo ora la nostra attenzione sulla norma $\left\|\delta^k\right\|_{L^2(\Omega)}$. Notiamo che, grazie alla (1.7),

$$\left(f^{k+1}, v_h\right) - a\left(u\left(t^{k+1}\right), v_h\right) = \left(\frac{\partial u\left(t^{k+1}\right)}{\partial t}, v_h\right).$$

Ciò ci consente di riscrivere la (1.34) nel seguente modo:

$$\begin{aligned} \left(\delta^{k+1}, v_h\right) &= \left(\frac{\partial u\left(t^{k+1}\right)}{\partial t}\right) - \frac{1}{\Delta t}\left(\Pi^r_{1,h}\left(u\left(t^{k+1}\right) - u(t^k)\right), v_h\right) \\ &= \left(\frac{\partial u\left(t^{k+1}\right)}{\partial t} - \frac{u\left(t^{k+1}\right) - u\left(t^k\right)}{\Delta t}, v_h\right) + \left(\left(I - \Pi^r_{1,h}\right)\left(\frac{u\left(t^{k+1}\right) - u\left(t^k\right)}{\Delta t}\right), v_h\right) \end{aligned} \tag{1.38}$$

Usando la formula di Taylor con il resto nella forma integrale, abbiamo:

$$\frac{\partial u\left(t^{k+1}\right)}{\partial t} - \frac{u^{t^{k+1}} - u\left(t^k\right)}{\Delta t} = \frac{1}{\Delta t}\int_{t^k}^{t^{k+1}} (s - t^k) \frac{\partial^2 u}{\partial t^2}(s) ds, \tag{1.39}$$

avendo richiesto un'opportuna regolarità della funzione u rispetto alla variabile temporale. Utilizzando poi il teorema fondamentale di integrazione ed esplicitando la commutatività tra l'operatore di proiezione $\Pi^r_{1,h}$ e la derivata temporale, otteniamo:

$$\left(I - \Pi^r_{1,h}\right)\left(u\left(t^{k+1}\right) - u(t^k)\right) = \int_{t^k}^{t^{k+1}} \left(I - \Pi^r_{1,h}\right)\left(\frac{\partial u}{\partial t}\right)(s) ds. \tag{1.40}$$

Scegliendo $v_h = \delta^{k+1}$ nella (1.38), grazie alle (1.39) e (1.40), possiamo dedurre una stima dall'alto:

$$\begin{aligned}\left\|\delta^{k+1}\right\|_{L^2(\Omega)} &\leq \left\|\frac{1}{\Delta t}\int_{t^k}^{t^{k+1}}(s-t^k)\frac{\partial^2 u}{\partial t^2}(s)ds\right\|_{L^2(\Omega)} + \left\|\frac{1}{\Delta t}\int_{t^k}^{t^k+1}\left(I-\Pi_{1,h}^r\right)\left(\frac{\partial u}{\partial t}(s)\right)\right\|_{L^2(\Omega)} \\ &\leq \int_{t^k}^{t^{k+1}}\left\|\frac{\partial^2 u}{\partial t^2}(s)\right\|_{L^2(\Omega)} + \frac{1}{\Delta}\int_{t^k}^{t^{k+1}}\left\|\left(I-\Pi_{1,h}^r\right)\left(\frac{\partial u}{\partial t}(s)\right)\right\|_{L^2(\Omega)} ds. \qquad (1.41)\end{aligned}$$

Ritornando alla stima di stabilità (1.35) ed esplicitando la (1.37) e la (1.41) con indici opportunamente riscalati, abbiamo

$$\begin{aligned}\|\epsilon_h^n\|^2_{L^2(\Omega)} \leq C(t^n)\Bigg\{h^{2r}|u_0|^2_{H^r(\Omega)} + \sum_{k=1}^{n}\Delta t\Bigg[\left(\int_{t^{k-1}}^{t^k}\left\|\frac{\partial^2 u}{\partial t^2}(s)\right\|_{L^2(\Omega)} ds\right) \\ + \frac{1}{\Delta t^2}\left(\int_{t^{k-1}}^{t^k}\left\|\left(I-\Pi_{1,h}^r\right)\left(\frac{\partial u}{\partial t}\right)(s)\right\|_{L^2(\Omega)} ds\right)^2\Bigg]\Bigg\}. \qquad (1.42)\end{aligned}$$

Quindi, usando la disuguaglianza di Cauchy-Schwarz e la stima (1.32) per l'operatore di proiezione $\Pi_{1,h}^r$, otteniamo:

$$\begin{aligned}\|\epsilon_h^n\|^2_{L^2(\Omega)} &\leq C(t^n)\Bigg\{h^{2r}|u_0|^2_{H^r(\Omega)} + \sum_{k=1}^{n}\Delta t\Bigg[\Delta t\int_{t^{k-1}}^{t^k}\left\|\frac{\partial^2 u}{\partial t^2}(s)\right\|^2_{L^2(\Omega} ds \\ &+ \frac{1}{\Delta t^2}\left(\int_{t^{k-1}}^{t^k} h^r\left|\frac{\partial u}{\partial t}(s)\right|_{H^r(\Omega)}\right)^2\Bigg]\Bigg\} \\ &\leq C(t^n)\Bigg(h^{2r}|u_0|^2_{H^r(\Omega)} + \Delta t^2\sum_{k=1}^{n}\int_{t^{k-1}}^{t^k}\left\|\frac{\partial^2 u}{\partial t^2}(s)\right\|^2_{L^2(\Omega)} ds \\ &+ \frac{1}{\Delta t}h^{2r}\sum_{k=1}^{n}\Delta t\int_{t^{k-1}}^{t^k}\left|\frac{\partial u}{\partial t}(s)\right|^2_{H^r(\Omega)} ds\Bigg). \qquad (1.43)\end{aligned}$$

Il risultato ora segue usando la (1.31) e la stima (1.32). Per altre stime di stabilità si può consultare [8].

Capitolo 2

Equazioni di diffusione-trasporto-reazione

Il nostro prossimo obiettivo consiste nell'enunciare un metodo numerico per trattare e approssimare la parte di diffusione-trasporto dell'equazione evolutiva introdotta all'inizio: il suo nome è *Galerkin Least Squares* (abbreviato in GLS). Prima però di analizzarlo in dettaglio e di considerarlo in un'applicazione concreta, ripercorriamo gli aspetti fondamentali dell'equazione di diffusione-trasporto-reazione e la loro trattazione numerica mediante l'approccio degli elementi finiti.

Consideriamo problemi della forma seguente:

$$\begin{cases} -\mathrm{div}(\mu\nabla u) + \vec{b}\cdot\nabla u + \sigma u = f & \text{in } \Omega \\ u = 0 & \text{su } \partial\Omega \end{cases}, \tag{2.1}$$

dove μ, σ, e $\vec{b}$ sono funzioni o costanti assegnate. Nel caso più generale supponiamo che $\mu \in L^\infty(\Omega)$ con $\mu(\vec{x}) \geq \mu_0 > 0$, $\sigma \in L^2(\Omega)$ con $\sigma(x) \geq 0$ quasi ovunque in Ω, $\vec{b} \in [L^\infty(\Omega)]^2$ con $\mathrm{div}(\vec{b}) \in L^2(\Omega)$ ed $f \in L^2(\Omega)$. In molte applicazioni pratiche il termine *diffusivo* $-\mathrm{div}(\mu\nabla u)$ è dominato dal termine di *trasporto* (o *convettivo*) $\vec{b}\cdot\nabla u$ o dal termine di *reazione* σu (anche detto *termine di assorbimento* quando σ è non negativo). In tali casi la soluzione può dar luogo ai cosiddetti *strati di bordo*, cioè regioni, generalmente vicine alla frontiera di Ω, dove la soluzione è caratterizzata da forti gradienti.

2.1 Formulazione debole del problema

Ora introduciamo la formulazione debole del problema (2.1), consideriamo il metodo di Galerkin per approssimare la soluzione, illustriamo le sue difficoltà

nel fornire soluzioni stabili in presenza di strati di bordo ed infine proponiamo come metodo di discretizzazione per approssimare il problema (2.1) il metodo GLS. Sia $V = H_0^1(\Omega)$. Introduciamo la forma bilineare $a : V \times V \to \mathbb{R}$,

$$a(u,v) = \int_\Omega \mu \nabla u \cdot \nabla v d\Omega + \int_\Omega v\vec{b} \cdot \nabla d\Omega + \int_\Omega \sigma uv d\Omega, \quad \forall u, v \in V,$$

la formulazione debole del problema (2.1) diventa:

$$\text{trovare } u \in V \ : \ a(u,v) = (f,v), \quad \forall v \in V. \tag{2.2}$$

Si può dimostrare che tale forma bilineare soddisfa le ipotesi del teorema di Lax-Milgram, essendo sia coerciva in quanto si verifica che, per un'opportuna costante $C_\Omega > 0$ indipendente da v,

$$a(v,v) \geq \alpha \|v\|^2_{H^1(\Omega)} \quad \forall v \in V, \text{ con } \alpha = \frac{\mu_0}{1 + C_\Omega^2},$$

che anche continua in quanto si verifica che esiste una costante $M > 0$ (ad esempio si può prendere $M = \|\mu\|_{L^\infty(\Omega)} + \|b\|_{L^\infty(\Omega)} + \|\sigma\|_{L^2(\Omega)}$) tale che

$$|a(u,v)| \leq M \|u\|_{H^1(\Omega)} \cdot \|v\|_{H^1(\Omega)}, \quad \forall u, v \in V.$$

D'altra parte il lato destro della (2.2) definisce un funzionale limitato e lineare, grazie alla disuguaglianza di Cauchy-Schwarz ed alla disuguaglianza di Poincaré (vedi paragrafo 2.13 in [3]). Poiché sono verificate le ipotesi del lemma di Lax-Milgram (vedi lemma 3.1 in [3]), segue che esiste ed è unica la soluzione del problema debole (2.2). Inoltre si può dimostrare che valgono le seguenti stime:

$$\|u\|_{H^1(\Omega)} \leq \frac{1}{\alpha} \|f\|_{L^2(\Omega)}, \quad \|\nabla u\|_{L^2(\Omega)} \leq \frac{C_\Omega}{\mu_0} \|f\|_{L^2(\Omega)}.$$

L'approssimazione di Galerkin del problema (2.2) è data da:

$$\text{trovare } u_h \in V_h \ : \ a(u_h, v_h) = (f, v_h), \quad \forall v_h \in V_h,$$

dove $\{V_h,\, h > 0\}$ è un'opportuna famiglia di sottospazi di $H_0^1(\Omega)$. Replicando le stime svolte per il problema esatto (2.2) si possono provare le seguenti stime:

$$\|u_h\|_{H^1(\Omega)} \leq \frac{1}{\alpha} \|f\|_{L^2(\Omega)}, \quad \|\nabla u_h\|_{L^2(\Omega)} \leq \frac{C_\Omega}{\mu_0} \|f\|_{L^2(\Omega)}. \tag{2.3}$$

Queste ultime provano in particolare che il gradiente della soluzione discreta (cosí come quella della soluzione debole u) potrebbe essere molto grande, qualora μ_0 fosse piccola. In aggiunta, la disuguaglianza di Galerkin per l'errore (vedi paragrafo 4.10 in [3]) fornisce:

$$\|u - u_h\|_V \leq \frac{M}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V. \tag{2.4}$$

Per le definizioni di α ed M, la costante $\frac{M}{\alpha}$ diventa tanto più alta (e quindi la stima (2.4) meno significativa) quanto più il rapporto $\frac{\|b\|_{L^\infty}}{\|\mu\|_{L^\infty(\Omega)}}$ (o il rapporto $\frac{\|\sigma\|_{L^\infty(\Omega)}}{\|\mu\|_{L^\infty(\Omega)}}$) cresce, cioè quando il termine convettivo (o reattivo) domina su quello diffusivo. In tali casi il metodo di Galerkin può portare a soluzioni non accurate, a meno che non venga utilizzato un passo di discretizzazione h estremamente piccolo.

2.2 Analisi di un problema di diffusione-trasporto unidimensionale

Per valutare più precisamente il comportamento della soluzione numerica fornita dal metodo di Galerkin, spendiamo qualche parola ad analizzare un problema unidimensionale di interesse per ciò che segue. Il problema di interesse è di tipo diffusione-trasporto:

$$\begin{cases} -\mu u'' + bu' = 0, \quad 0 < x < 1 \\ u(0) = 0, \quad u(1) = 1 \end{cases}, \tag{2.5}$$

con μ e b costanti positive. La sua formulazione debole è :

$$\text{trovare } u \in H^1(0,1) \ : \ a(u,v) = 0 \quad \forall v \in H_0^1(0,1), \quad \text{con } u(0) = 0, \ u(1) = 1 \tag{2.6}$$

con

$$a(u,v) = \int_0^1 (\mu u'v' + bu'v)\, dx.$$

Il problema (2.6) può essere riformulato introducendo un opportuno *lifting* o estensione dei dati al bordo (per i dettagli vedi la sezione 3.2.2 in [3]). In questo caso possiamo scegliere $R_g = x$. Ponendo $\mathring{u} = u - R_g = u - x$ il problema (2.6) si può riformulare cosí:

$$\text{trovare } \mathring{u} \in H_0^1(0,1) \ : \ a\,(u,v) = F(v), \quad \forall v \in H_0^1(0,1), \tag{2.7}$$

essendo

$$F(v) = -a(x,v) = -\int_0^1 bv dx$$

il contributo dovuto al lifting.

Definiamo il *numero di Péclet globale* il rapporto

$$\mathbb{P}_{e_g} = \frac{|\vec{b}|L}{2\mu},$$

dove L è la dimensione lineare del dominio (nel nostro caso $L = 1$). Tale rapporto fornisce una misura di quanto il termine di trasporto domina su quello diffusivo. La soluzione esatta di tale problema è dato da:

$$u(x) = c_1 + c_2 e^{\frac{b}{\mu}x}.$$

Imponendo le condizioni al bordo si calcolano le costanti c_1 e c_2 e perciò la soluzione

$$u(x) = \frac{e^{\frac{b}{\mu}x} - 1}{e^{\frac{b}{\mu}} - 1}.$$

Espandendo poi con Taylor gli esponenziali nell'ipotesi di $\frac{b}{\mu} << 1$ otteniamo:

$$u(x) = \frac{1 + \frac{b}{\mu}x + ... - 1}{1 + \frac{b}{\mu} + ... - 1} \approx \frac{\frac{b}{\mu}x}{\frac{b}{\mu}} = x.$$

Perciò la soluzione è vicina alla retta che interpola i dati al bordo (che è la soluzione corrispondente al caso $b = 0$).

Di contro, se $\frac{b}{\mu} >> 1$, gli esponenziali sono molto grandi e dunque

$$u(x) \approx \frac{e^{\frac{b}{\mu}x}}{e^{\frac{b}{\mu}}} = e^{-\frac{b}{\mu}(1-x)},$$

da cui segue che la soluzione è vicina a 0 in quasi tutto l'intervallo, eccetto che in un intorno del punto $x = 1$ dove tende ad 1 in modo esponenziale. Tale intorno ha un'ampiezza dell'ordine di $\frac{\mu}{b}$ ed è perciò molto piccolo: la soluzione mostra una soglia di bordo di ampiezza $O\left(\frac{\mu}{b}\right)$ vicino ad $x = 1$, dove la derivata si comporta come $\frac{b}{\mu}$ ed è quindi illimitata per $\mu \to 0$.

Supponiamo ora di utilizzare il metodo degli elementi finiti di Galerkin lineare per approssimare il problema (2.6), cioè

$$\text{trovare } u_h \in X_h^1 \ : \begin{cases} a(u_h, v_h) = 0 \quad \forall v_h \in \mathring{X}_h^1 \\ u_h(0) = 0, \quad u_h(1) = 1 \end{cases}, \tag{2.8}$$

dove, denotando con x_i per $i = 0, ..., M$, i vertici della partizione introdotta su $(0, 1)$, abbiamo posto

$$X_h^r = \left\{ v_h \in C^0([0,1]) \ : \ v_h|_{[x_{i-1}, x_i]} \in \mathbb{P}_r, \ i = 1, ..., M \right\},$$

$$\mathring{X}_h^r = \left\{ v_h \in X_h^r \ : \ v_h(0) = v_h(1) = 0 \right\},$$

per $r \geq 1$. Avendo scelto, per ciascun $i = 1, ..., M-1$, $v_h = \phi_i$ (la i-ma funzione della base di X_h^1), abbiamo

$$\int_0^1 \mu u_h' \phi_i' dx + \int_0^1 b u_h' \phi_i dx = 0,$$

che, riscrivendo $u_h = \sum_{j=1}^{M-1} u_j \phi_j(x)$ e poiché il supporto di ϕ_i è uguale ad $[x_{i-1}, x_{i+1}]$, si può esprimere equivalentemente come

$$\begin{aligned} &\mu \left[u_{i-1} \int_{x_{i-1}}^{x_i} \phi_{i-1}' \phi_i' dx + u_i \int_{x_{i-1}}^{x_{i+1}} (\phi_i')^2 dx + u_{i+1} \int_{x_i}^{x_{i+1}} \phi_{i+1}' \phi_i' dx \right] \\ &+b \left[u_{i-1} \int_{x_{i-1}}^{x_i} \phi_{i-1}' \phi_i dx + u_i \int_{x_{i-1}}^{x_{i+1}} \phi_i' \phi_i dx + u_{i+1} \int_{x_i}^{x_{i+1}} \phi_{i+1}' \phi_i dx \right] \\ &= 0, \end{aligned} \tag{2.9}$$

$\forall i = 1, ..., M-1$. Se la partizione è uniforme, cioè $x_i = x_{i-1} + h$ con $i = 1, ..., M$, osservando che $\phi_i'(x) = \frac{1}{h}$ se $x_{i-1} < x < x_i$, $\phi_i'(x) = -\frac{1}{h}$ se $x_i < x < x_{i+1}$, per $i = 1, ..., M-1$ otteniamo

$$\mu \left(-u_{i-1} \frac{1}{h} + u_i \frac{2}{h} - u_{i+1} \frac{1}{h} \right) + b \left(-u_{i-1} \frac{1}{h} \frac{h}{2} + u_{i+1} \frac{1}{h} \frac{h}{2} \right) = 0,$$

cioè

$$\frac{\mu}{h} \left(-u_{i-1} + 2u_i - u_{i+1} \right) + \frac{b}{2} \left(u_{i+1} - u_{i-1} \right) = 0, \quad i = 1, ..., M-1.$$

Riordinando i termini, troviamo:

$$\left(\frac{b}{2} - \frac{\mu}{h} \right) u_{i+1} + \frac{2\mu}{h} u_i - \left(\frac{b}{2} + \frac{\mu}{h} \right) u_{i-1} = 0, \;\; i = 1, ..., M-1.$$

Dividendo tutto per $\frac{\mu}{h}$ e definendo il *numero di Péclet locale*

$$\mathbb{P}_e = \frac{|\vec{b}| h}{2\mu},$$

abbiamo infine:

$$(\mathbb{P}_e - 1) u_{i+1} + 2u_i - (\mathbb{P}_e + 1) u_{i-1} = 0, \;\; i = 1, ..., M-1. \tag{2.10}$$

Si tratta di una equazione differenziale che ammette soluzioni della forma $u_i = \rho^i$ (per maggiori dettagli vedi [4]). Sostituendo tale espressione nella (2.10) abbiamo:

$$(\mathbb{P}_e - 1) \rho^2 + 2\rho - (\mathbb{P}_e + 1) = 0, \;\; i = 1, ..., M-1,$$

da cui ricaviamo le due radici:

$$\rho_{1,2} = \frac{-1 \pm \sqrt{1+\mathbb{P}_l^2} - 1}{\mathbb{P}_e - 1} = \begin{cases} \frac{1+\mathbb{P}_e}{1-\mathbb{P}_e} \\ 1 \end{cases} .$$

Grazie alla linearità di (2.10), l'integrale generale di tale equazione assume la forma:

$$u_i = A_1\rho_1^i + A_2\rho_2^i,$$

essendo A_1 ed A_2 due costanti arbitrarie. Imponendo le condizioni al bordo $u_0 = 0$ ed $u_M = 1$, troviamo che

$$A_2 = -A_1, \qquad A_1 = \left[1 - \left(\frac{1+\mathbb{P}_e}{1-\mathbb{P}_e}\right)^M\right]^{-1}.$$

In conclusione, la soluzione del problema (2.8) ha i seguenti valori sui nodi:

$$u_i = \frac{1 - \left(\frac{1+\mathbb{P}_e}{1-\mathbb{P}_e}\right)^i}{1 - \left(\frac{1+\mathbb{P}_e}{1-\mathbb{P}_e}\right)^M}, \quad i = 1, ..., M.$$

Osserviamo che, se $\mathbb{P}_e > 1$, l'esponenziale a numeratore possiede una base della potenza negativa, perciò la soluzione approssimata diventa oscillatoria, contrariamente alla soluzione esatta che è monotona. Tale fenomeno è rappresentato nella figura (2.1),

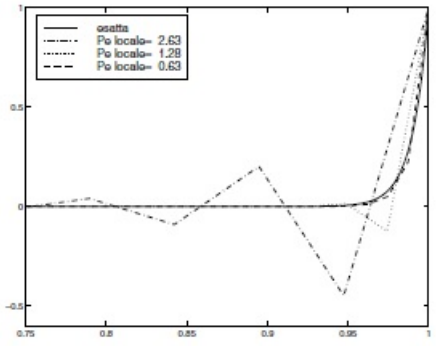


Figura 2.1: Soluzione agli elementi finiti del problema di diffusione-trasporto con $\mathbb{P}_{e_g} = 50$ per differenti valori del numero di Péclet locale.

dove la soluzione di (2.10), per differenti numeri di Péclet locali è confrontata con la soluzione esatta nel caso in cui il numero di Péclet globale è uguale a 50. Come si può notare quanto più grande è il numero di Péclet tanto più il

comportamento della soluzione approssimata differisce dalla soluzione esatta, con oscillazioni che diventano più apprezzabili in prossimità del limite della frontiera. Il rimedio più ovvio a tale discrepanza consisterebbe nello scegliere una taglia h della griglia sufficientemente piccola, per assicurarsi che $\mathbb{P}_e < 1$. Tuttavia tale strategia non sempre conviene: ad esempio, se $b = 1$ e $\mu = \frac{1}{5\,000}$, dovremmo scegliere $h < 10^{-4}$, il che introdurrebbe 10 000 intervalli in $(0, 1)$! In particolare, tale strategia richiederebbe un irragionevolmente alto numero di nodi per problemi con valori al bordo in più dimensioni.

Un rimedio più opportuno consiste invece nell'usare una procedura a priori adattiva che raffina la griglia solo in prossimità del limite delle frontiera. Si possono seguire a tal proposito diverse strategie, con griglie di tipo Bakhvâlov o di tipo Shishkin (per maggiori dettagli vedi [1]).

2.3 Schemi alle DF decentrate e diffusione artificiale

Un'analisi comparativa con il metodo alle differenze finite ci consente di trovare un rimedio al comportamento oscillatorio delle soluzioni agli elementi finiti nel caso del problema diffusione-trasporto (2.5). Se consideriamo le differenze finite, le oscillazioni nella soluzione numerica sorgono qualora si usi uno schema alle differenze finite centrato (CFD) per la discretizzazione del termine del trasporto. Dal momento che l'ultimo è non simmetrico, ciò suggerisce di discretizzare la derivata prima in un punto x_i mediante un rapporto incrementale decentrato in cui il valore in x_{i-1} interviene se il campo è positivo e quello in x_{i+1} interviene nel caso opposto. Questa tecnica è detta *upwinding* e lo schema che ne risulta, chiamato *upwind* (FDUP in breve) nel caso $b > 0$ si può scrivere:

$$-\mu\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + b\frac{u_i - u_{i-1}}{h} = 0, \quad i = 1, ..., M-1. \tag{2.11}$$

Il prezzo da pagare è una riduzione dell'ordine di convergenza, dal momento che il rapporto incrementale decentrato introduce un errore di discretizzazione locale che ha ordine $O(h)$ invece di $O(h^2)$ come nel caso CFD. Ora osserviamo che

$$\frac{u_i - u_{i-1}}{h} = \frac{u_{i+1} - u_{i-1}}{2h} - \frac{h}{2}\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2},$$

cioè il rapporto incrementale decentrato per approssimare la derivata prima si può scrivere come la somma del rapporto incrementale centrale più un termine proporzionale alla discretizzazione della derivata seconda, ancora

con un rapporto incrementale centrato. Perciò lo schema upwind si può reinterpretare come uno schema centrato alle differenze finite nel quale si sia introdotto un termine di diffusione artificiale proporzionale ad h. Di fatto, (2.11) è equivalente a

$$-\mu_h \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + b\frac{u_{i+1} - u_{i-1}}{2h} = 0, \quad i = 1, ..., M-1, \tag{2.12}$$

dove $\mu_h = \mu(1 + \mathbb{P}_e)$, essendo $\mathbb{P}_e$ il numero di Péclet locale introdotto precedentemente.

Lo schema (2.12) corrisponde alla discretizzazione usando uno schema del *problema perturbato*

$$-\mu_h u'' + bu' = 0.$$

La viscosità correttiva $\mu_h - \mu = \mu\mathbb{P}_e = \frac{bh}{2}$ è chiamata *viscosità numerica* o *viscosità artificiale*. Il nuovo numero di Péclet associato allo schema (2.12) è :

$$\mathbb{P}_e^* = \frac{bh}{2\mu_h} = \frac{\mathbb{P}_e}{(1 + \mathbb{P}_e)},$$

e perciò verifica: $\mathbb{P}_e^* < 1$ per tutti i possibili valori di $h > 0$. Tale interpretazione, come vedremo fra poco, permette di estendere la tecnica upwind agli elementi finiti e anche al caso bidimensionale, dove la nozione di differenziazione decentrata non è ovvia.

Più in generale, in uno schema CFD della forma (2.12) si può utilizzare il coefficiente di viscosità numerica

$$\mu_h = \mu\left(1 + \phi\left(\mathbb{P}_e\right)\right), \tag{2.13}$$

con ϕ opportuna funzione del numero di Péclet locale che deve soddisfare la proprietà : $\lim_{t\to 0^+} \phi(t) = 0$. Osserviamo che, se $\phi = 0$, otteniamo il metodo CFD

$$\begin{cases} -\mu\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + b\frac{u_{i+1} - u_{i-1}}{2h}, \quad i = 1, ..., M-1 \\ u_0 = 0, \;\; u_M = 1 \end{cases}$$

mentre se $\phi(t) = t$, otteniamo il metodo upwind FDUP (2.11) o (2.12). Osserviamo che il numero di Péclet locale associato al coefficiente (2.13) è

$$\mathbb{P}_e^* = \frac{bh}{2\mu_h} = \frac{\mathbb{P}_e}{(1 + \phi\left(\mathbb{P}_e\right))},$$

ed è perciò sempre minore di 1 per ciascun valore di h.

2.4 Autovalori dell'equazione di diffusione-trasporto

Consideriamo ora l'operatore $Lu = -\mu u'' + bu'$ associato al problema (2.5) in un generico intervallo (α, β). I suoi autovalori λ soddisfano il problema

$$Lu = \lambda u, \quad \alpha < x < \beta, \quad u(\alpha) = u(\beta) = 0$$

essendo u una autofunzione. Tali autovalori in generale sono complessi a causa della presenza del termine del primo ordine bu'. Supponendo che $\mu > 0$ sia costante (e b variabile a priori), abbiamo:

$$\mathrm{Re}(\lambda) = \frac{\int_\alpha^\beta Lu \cdot \bar{u} dx}{\int_\alpha^\beta |u|^2 dx} = \frac{\mu \int_\alpha^\beta |u'|^2 dx - \frac{1}{2}\int_\alpha^\beta b'|u|^2 dx}{\int_\alpha^\beta |u|^2 dx}.$$

Si può dedurre che se μ è piccolo e b' strettamente positivo, $\mathrm{Re}(\lambda)$ non è necessariamente positiva. Tuttavia, grazie alla disuguaglianza di Poincarè

$$\int_\alpha^\beta |u'|^2 dx \geq C_{\alpha,\beta} \int_\alpha^\beta |u|^2 dx,$$

con $C_{\alpha,\beta}$ costante positiva che dipende da $\beta - \alpha$, deduciamo che

$$\mathrm{Re}(\lambda) \geq C_{\alpha,\beta}\mu - \frac{1}{2} b'_{max},$$

essendo $b'_{max} = \max_{\alpha \leq s \leq \beta} b'(s)$. Dunque soltanto un numero finito di autovalori può avere parte reale negativa. In particolare osserviamo che $\mathrm{Re}(\lambda) > 0$ se b è costante o se $b'(x) \leq 0 \ \forall x \in [\alpha, \beta]$. Il medesimo tipo di limite inferiore si può ottenere per gli autovalori associati all'approssimazione agli elementi finiti di Galerkin del problema in esame. Gli ultimi sono la soluzione del problema:

$$\text{trovare } \lambda_h \in \mathbb{C},\ u_h \in V_h : \int_\alpha^\beta \mu u_h' v_h' dx + \int_\alpha^\beta b u_h' v_h dx = \lambda_h \int_\alpha^\beta u_h v_h dx, \ \ \forall v_h \in V_h, \tag{2.14}$$

dove

$$V_h = \{v_h \in V_h \ : \ v_h(\alpha) = v_h(\beta) = 0\}.$$

Per dimostrarlo è sufficiente prendere nuovamente $v_h = \bar{u}_h$ nella (2.14) e procedere come fatto in precedenza.

Per quanto concerne invece un limite superiore, scegliendo di nuovo $v_h = \bar{u}_h$ nella (2.14) e passando ai moduli di entrambi i membri, scriviamo:

$$|\lambda_h| \leq \frac{\mu \|u_h'\|^2_{L^2(\alpha,\beta)} + \|b\|_{L^\infty(\alpha,\beta)} \|u_h\|_{L^2(\alpha,\beta)}}{\|u_h\|^2_{L^2(\alpha,\beta)}}.$$

Usando poi la *disuguaglianza inversa* seguente (per i dettagli sulla dimostrazione vedi ad esempio [5]), che vale sotto opportune ipotesi di regolarità della triangolazione τ_h

$$\exists C_I > 0 \ : \ \forall v_h \in V_h, \ \|\nabla v_h\|_{L^2(\Omega)} \leq C_I h^{-1} \|v\|_{L^2(\Omega)} ,$$

con C_I indipendente da h, nel caso unidimensionale possiamo scrivere

$$\exists C_I = C_I(r) > 0 \ : \ \forall v_h \in X_h^r, \ \|v_h'\|_{L^2(\alpha,\beta)} \leq C_I h^{-1} \|v_h\|_{L^2(\alpha,\beta)} ,$$

e trovare facilmente che

$$|\lambda_h| \leq \mu C_I^2 h^{-2} + \|b\|_{L^\infty(\alpha,\beta)} C_I h^{-1}.$$

2.5 Metodi FE stabilizzati

Il metodo di Galerkin introdotto in precedenza fornisce un'approssimazione centrata del termine di trasporto. Un possibile modo di decentrare o desimmetrizzare tale approssimazione consiste nello scegliere delle funzioni test v_h in uno spazio differente da quello al quale appartiene u_h: facendo ciò si ottiene un metodo chiamato di *Petrov-Galerkin* per il quale l'analisi basata sul lemma di Céa non è più valido. Non è nostra intenzione analizzare le proprietà generali di tale metodo quanto piuttosto trattare i metodi *agli elementi finiti stabilizzati*, ed in particolare il metodo GLS. Tuttavia vogliamo sottolineare il fatto che tali metodi si possono vedere come casi particolari del metodi di Petrov-Galerkin.

L'approssimazione del problema (2.7) attraverso il metodo agli elementi finiti di Galerkin sarebbe:

$$\text{trovare } u_h \in V_h \ : \ a(u_h, v_h) = F(v_h) \quad \forall v_h \in V_h, \tag{2.15}$$

essendo $V_h = \mathring{X}_h^r$, con $r \geq 1$ ed u_h un'approssimazione di $\mathring{u}$. Invece di tale metodo, consideriamo il metodo di Galerkin generalizzato:

$$\text{trovare } u_h \in V_h \ : \ a_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h,$$

dove

$$a_h(u_h, v_h) = a(u_h, v_h) + b_h(u_h, v_h)$$

ed

$$F_h(v_h) = F(v_h) + G_h(v_h).$$

I termini aggiuntivi $b_h(u_h, v_h)$ e $G_h(v_h)$ hanno lo scopo di eliminare (o quantomeno ridurre) le oscillazioni numeriche prodotte dal metodo di Galerkin e

sono perciò chiamati *termini di stabilizzazione*. In particolare essi dipendono da h. Sottolineamo il fatto che il termine *stabilizzazione* non è esatto: il metodo di Galerkin è già stabile, nel senso della continuità della soluzione rispetto ai dati del problema (2.3). In questo contesto *stabilizzazione* deve essere inteso con lo scopo di ridurre (idealmente eliminare) le oscillazioni nella soluzione numerica quando $\mathbb{P}_e > 1$.

2.5.1 Consistenza ed errore di troncamento per i metodi di Galerkin e Galerkin generalizzato

Ora vogliamo concentrare la nostra attenzione in alcuni metodi di stabilizzazione molto accurati ed in particolare sul metodo di *Galerkin least squares* (GLS). Prima di ciò diamo qualche definizione utile. Per il problema di Galerkin (2.15) consideriamo la differenza tra il membro sinistro e destro quando si rimpiazza la soluzione approssimata u_h con quella esatta, cioè

$$\tau_h(u; v_h) = a_h(u, v_h) - F_h(v_h). \tag{2.16}$$

Quest'ultimo è un funzionale della variabile v_h, la cui norma

$$\tau_h(u) = \sup_{v_h \in V_h,\, v_h \neq 0} \frac{|\tau_h(u; v_h)|}{\|v_h\|_V}$$

definisce l'errore di troncamento associato al metodo (2.15). Diremo che il metodo di Galerkin generalizzato è *consistente* se $\lim_{h\to 0} \tau_h(u) = 0$. Inoltre esso è *fortemente (o pienamente) consistente* se l'errore di troncamento risulta essere nullo per ciascun valore di h. Il metodo di Galerkin standard, ad esempio, è fortemente consistente dal momento che $\forall v_h \in V_h$ si ha:

$$\tau_h(u; v_h) = a(u, v_h) - F(v_h) = 0.$$

Il metodo di Galerkin generalizzato è in generale convergente come segue dal lemma di Strang:

Lemma 2.1. *Si consideri il problema*

$$\textit{trovare } u \in V \ : \ a(u, v) = F(v) \quad \forall v \in V, \tag{2.17}$$

dove V è uno spazio di Hilbert con norma $\|\cdot\|_V$, con $F \in V'$ funzionale lineare e limitato su V e $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ una forma bilineare continua e coerciva su V. Inoltre si assuma che una approssimazione della (2.17) può essere formulata attraverso il seguente problema di Galerkin generalizzato:

$$\textit{trovare } u_h \in V_h \ : \ a_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h, \tag{2.18}$$

dove $\{V_h,\, h>0\}$ *è una famiglia di sottospazi finito-dimensionali di* V.

Supponiamo che la forma bilineare discreta $a_h(\cdot,\cdot)$ *sia continua su* $V_h \times V_h$, *e uniformemente coerciva su* V_h, *cioè che*

$$\exists \alpha^* \textit{indipendente da } h \,:\, a_h(v_h, v_h) \geq \alpha^* \left\|v_h\right\|_V^2 \quad \forall v_h \in V_h.$$

Infine, supponiamo che F_h *sia un funzionale lineare limitato su* V_h. *Allora*
1. *esiste una sola soluzione* u_h *del problema (2.18);*
2. *tale soluzione dipende in modo continuo dai dati e si ha*

$$\left\|u_h\right\|_V \leq \frac{1}{\alpha^*} \sup_{v_h \in V_h \setminus \{0\}} \frac{F_h(v_h)}{\left\|v_h\right\|_V};$$

3. *vale la seguente stima a priori*

$$\left\|u - u_h\right\|_V \leq \inf_{w_h \in V_h} \left\{ \left(1 + \frac{M}{\alpha^*}\right) \left\|u - w_h\right\|_V + \frac{1}{\alpha^*} \sup_{v_h \in V_h \setminus \{0\}} \frac{|a(w_h, v_h) - a_h(w_h, v_h)|}{\left\|v_h\right\|_V} \right\}$$
$$+ \frac{1}{\alpha^*} \sup_{v_h \in V_h \setminus \{0\}} \frac{|F(v_h) - F_h(v_h)|}{\left\|v_h\right\|_V}, \tag{2.19}$$

essendo M *la costante di continuità della forma bilineare* $a(\cdot,\cdot)$

a patto che $a_h - a$ ed $F_h - F$ tendano a zero quando h tende a zero.

2.5.2 Parte simmetrica ed antisimmetrica di un operatore

Sia ora V uno spazio di Hilbert e V' il suo duale. Diremo che un operatore $L : V \to V'$ è simmetrico se

$$_{V'} < Lu, v >_V = _V < u, Lv >'_V \quad \forall u, v \in V,$$

antisimmetrico se

$$_{V'} < Lu, v >_V = -_V < u, Lv >'_V \quad \forall u, v \in V.$$

Un operatore può essere suddiviso nella somma della sua parte simmetrica L_S e nella sua parte antisimmetrica L_{SS}, cioè

$$Lu = L_S u + L_{SS} u.$$

Consideriamo, ad esempio, l'operatore di diffusione-trasporto-reazione

$$Lu = -\mu \Delta u + \operatorname{div}(\vec{b} u) + \sigma u, \quad x \in \Omega \subset \mathbb{R}^d,\, d \geq 2,$$

sullo spazio di Hilbert $V = H_0^1(\Omega)$. Dal momento che

$$\mathrm{div}(\vec{b}u) = \frac{1}{2}\mathrm{div}(\vec{b}u) + \frac{1}{2}\mathrm{div}(\vec{b}u) = \frac{1}{2}\mathrm{div}(\vec{b}u) + \frac{1}{2}u\mathrm{div}(\vec{b}) + \frac{1}{2}\vec{b}\cdot\nabla u,$$

possiamo suddividere l'operatore nel modo seguente:

$$Lu = \underbrace{-\mu\Delta u + \left[\sigma + \frac{1}{2}\mathrm{div}(b)\right]u}_{L_S u} + \underbrace{\frac{1}{2}\left[\mathrm{div}(\vec{b}u) + \vec{b}\cdot\nabla u\right]}_{L_{SS}u}.$$

Si osservi che il coefficiente di reazione è divenuto:

$$\sigma^* = \sigma + \frac{1}{2}\mathrm{div}(\vec{b}).$$

Si può verificare che le due parti nelle quali è stato suddiviso l'operatore sono simmetriche e anti-simmetriche (la verifica è un semplice esercizio, per i dettagli vedi [3]).

2.5.3 Metodi fortemente consistenti

Consideriamo ora un problema di diffusione-trasporto-reazione che scriviamo nella forma astratta $Lu = f$ in Ω, con $u = 0$ su $\partial\Omega$. Scriviamo poi la corrispondente formulazione debole:

$$\text{trovare } u \in V = H_0^1(\Omega) \ : \ a(u,v) = (f,v) \quad \forall v \in V,$$

con $a(\cdot,\cdot)$ forma bilineare associata ad L. Si può ottenere un metodo stabilizzato e fortemente consistente aggiungendo un ulteriore termine alla sua approssimazione di Galerkin (2.15), considerando il problema seguente:

$$\text{trovare } u_h \in V_h \ : \ a(u_h, v_h) + \mathcal{L}_h(u_h, f; v_h) = (f, v_h) \quad \forall v_h \in V_h, \tag{2.20}$$

per un'opportuna forma $\mathcal{L}_h$ tale che

$$\mathcal{L}_h(u, f; v_h) = 0 \quad \forall v_h \in V_h. \tag{2.21}$$

Tale forma $\mathcal{L}_h$ dipende sia dalla soluzione approssimata u_h che dal termine forzante f. Una possibile scelta della forma che soddisfi (2.21) è :

$$\mathcal{L}_h(u_h, f; v_h) = \mathcal{L}_h^{(\rho)}(u_h, f; v_h) = \sum_{K\in\tau_h} \delta \left(Lu_h - f, S_K^{(\rho)}(v_h)\right)_{L^2(K)},$$

dove usiamo la notazione $(u,v)_{L^2(K)} = \int_K uvdK$, ρ e δ sono parametri da assegnare ed

$$S_K^{(\rho)}(v_h) = \frac{h_K}{|\vec{b}|}\left[L_{SS}v_h + \rho L_S v_h\right],$$

con b campo di trasporto, L_S ed L_{SS} rispettivamente parte simmetrica ed antisimmetrica dell'operatore L in esame ed h_K il diametro del generico elemento K della triangolazione.

Per verificare la consistenza di (2.20) poniamo:

$$a_h(u_h, v_h) = a(u_h, v_h) + \mathcal{L}_h^{(\rho)}(u_h, f; v_h).$$

In virtù della definizione (2.16) otteniamo:

$$\tau_h(u; v_h) = a_h(u, v_h) - (f, v_h) = a(u, v_h) + \mathcal{L}_h^{(\rho)}(u, f; v_h) - (f, v_h) = \mathcal{L}_h^{(\rho)}(u, f; v_h) = 0.$$

L'ultima identità deriva dal fatto che $Lu - f = 0$. Da qui, $\tau_h(u) = 0$ e perciò la proprietà (2.21) ci assicura che il metodo (2.20) è fortemente consistente. A seconda dei particolari valori assegnati al parametro ρ si ottengono metodi differenti:
- se $\rho = 1$ si ottiene il metodo detto *Galerkin Least Squares* (GLS), per il quale

$$S_K^{(1)}(v_h) = \frac{h_K}{|\vec{b}|_{eu}} L v_h;$$

se si prende $v_h = u_h$, vediamo che, su ciascun triangolo, è stato aggiunto un termine proporzionale a $\int_K (Lu_h)^2 dK$;
- se $\rho = 0$ otteniamo il metodo *Streamline Upwind Petrov-Galerkin* (SUPG) per il quale

$$S_K^{(0)}(v_h) = \frac{h_K}{|\vec{b}|_{eu}} L_{SS} v_h;$$

- se $\rho = -1$ otteniamo il cosiddetto metodo di *Douglas-Wang* (DW) per il quale

$$S_K^{(-1)}(v_h) = \frac{h_K}{|\vec{b}|_{eu}} \left(L_{SS} - L_S\right) v_h.$$

Notiamo che nel caso in cui $\sigma^* = 0$ ed usiamo gli elementi finiti $\mathbb{P}_1$ i tre metodi precedenti coincidono, poiché $-\Delta u_k|_K = 0 \; \forall K \in \tau_h$. Se ci limitiamo alle due procedure più classiche, cioè GLS ($\rho = 1$) e SUPG ($\rho = 0$), possiamo definire la *norma* ρ nel seguente modo:

$$\|v\|_\rho = \left\{ \mu \|\nabla v\|^2_{L^2(\Omega)} + \|\sqrt{\gamma} v\|^2_{L^2(\Omega)} + \sum_{K \in \tau_h} \delta \left((L_{SS} + \rho L_S)v, S_K^{(\rho)}(v) \right)_{L^2(K)} \right\}^{\frac{1}{2}},$$

dove γ è una costante positiva tale che $\frac{1}{2}\text{div}\vec{b} + \sigma \geq \gamma > 0$.

Vale la seguente disuguaglianza (riguardante la stabilità): $\exists \alpha^* = \alpha^*(\gamma, \alpha)$, con α costante di coercitività di $a(\cdot,\cdot)$, tale che

$$\|u_h\|_{(\rho)} \leq \frac{C}{\alpha^*} \|f\|_{L^2(\Omega)}, \tag{2.22}$$

con C opportuna costante (che vedremo meglio più avanti). Inoltre vale la seguente stima dell'errore:

$$\|u - u_h\|_{(\rho)} \leq C \cdot h^{r+\frac{1}{2}} |u|_{H^{r+1}(\Omega)}, \tag{2.23}$$

da cui segue che l'ordine di accuratezza del metodo cresce quando il grado r del polinomio che utilizziamo cresce, come nel metodo di Galerkin standard. La scelta del parametro di stabilizzazione δ, che misura l'ammontare della viscosità artificiale, è estremamente importante. A tal fine riportiamo nella prossima tabella il range ammesso per tale parametro in funzione dello schema stabilizzato che si sceglie.

SUPG	$0 < \delta < \frac{1}{C_0}$
GLS	$0 < \delta$
DW	$0 < \delta < \frac{1}{2C_0}$

Tabella 2.1: Valori ammissibili del parametro di stabilizzazione δ

La costante C_0 compare nella disuguaglianza inversa:

$$\sum_{K \in \tau_h} h_K^2 \int_K |\Delta v_h|^2 dK \leq C_0 \|\nabla v_h\|^2_{L^2(\Omega)} \quad \forall v_h \in X_h^r. \tag{2.24}$$

2.5.4 Analisi del metodo GLS

Ora vogliamo concentrarci più approfonditamente sul metodo GLS e sulle sue proprietà di stabilità (2.22) e di convergenza (2.23). Supponiamo che l'operatore differenziale L abbia la forma

$$Lu = -\mu\Delta u + \text{div}(\vec{b}u) + \sigma u$$

con $\mu > 0$ e $\sigma \geq 0$ costante, con condizioni al bordo di Dirichlet omogenee assegnate. La forma bilineare $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ associata all'operatore L è perciò

$$a(u,v) = \mu \int_\Omega \nabla u \cdot \nabla v d\Omega + \int_\Omega \text{div}(\vec{b}u) v d\Omega + \int_\Omega \sigma u v d\Omega,$$

con $V = H_0^1(\Omega)$. Per maggiore semplicità supponiamo in ciò che segue che esistano due costanti γ_0 e γ_1 tali che

$$0 < \gamma_0 \leq \gamma(\vec{x}) = \frac{1}{2}\mathrm{div}(\vec{b}(\vec{x})) + \sigma \leq \gamma_1 \quad \forall \vec{x} \in \Omega. \tag{2.25}$$

In tal caso la forma bilineare $a(\cdot,\cdot)$ è coerciva, dal momento che

$$a(v,v) \geq \mu \left\|\nabla v\right\|^2_{L^2(\Omega)} + \gamma_0 \left\|v\right\|^2_{L^2(\Omega)}.$$

Consideriamo le parti simmetrica ed antisimmetrica associati ad L, cioè

$$L_S u = -\mu\Delta u + \gamma u, \qquad L_{SS} u = \frac{1}{2}\left(\mathrm{div}(\vec{b}u) + \vec{b}\cdot\nabla u\right).$$

Inoltre, riscriviamo la formulazione stabilizzata (2.20) dividendo $\mathcal{L}_h(u_h, f; v_h)$ in due termini, uno contenente u_h, l'altro f:

$$\text{trovare } u_h \in V_h \ : \ a_h^{(1)}(u_h, v_h) = f_h^{(1)}(v_h) \quad \forall v_h \in V_h, \tag{2.26}$$

con

$$a_h^{(1)}(u_h, v_h) = a(u_h, v_h) + \sum_{K\in\tau_h} \delta \left(Lu_h, \frac{h_K}{|\vec{b}|} Lv_h \right)_{L^2(K)}$$

e

$$f_h^{(1)}(v_h) = (f, v_h) + \sum_{K\in\tau_h} \delta \left(f, \frac{h_K}{|\vec{b}|} Lv_h \right)_{L^2(K)}.$$

Osserviamo che, con tali notazioni, la proprietà di consistenza forte si esprime attraverso l'uguaglianza

$$a_h^{(1)}(u, v_h) = f_h^{(1)}(v_h) \quad \forall v_h \in V_h.$$

Dimostriamo il seguente risultato preliminare:

Lemma 2.2. *Per ciascun $\delta > 0$, la forma bilineare $a_h^{(1)}(\cdot,\cdot)$ soddisfa la seguente relazione:*

$$a_h^{(1)}(v_h, v_h) = \mu \left\|\nabla v_h\right\|^2_{L^2(\Omega)} + \left\|\sqrt{\gamma} v_h\right\|^2_{L^2(\Omega)} + \sum_{K\in\tau_h} \delta \left(\frac{h_K}{|\vec{b}|} Lv_h, Lv_h \right)_{L^2(K)} \qquad \forall v_h \in V_h.$$

Dimostrazione: *Tale risultato segue dalla definizione di $a_h^{(1)}(u_h, v_h)$ (avendo posto $v_h = u_h$) e dall'ipotesi (2.25). Nel caso sotto esame la norma $\|\cdot\|_{(1)}$, che denotiamo per convenienza con il simbolo $\|\cdot\|_{GLS}$, diviene:*

$$\|v_h\|^2_{GLS} = \mu \left\|\nabla v_h\right\|^2_{L^2(\Omega)} + \left\|\sqrt{\gamma} v\right\|^2_{L^2(\Omega)} + \sum_{K\in\tau_h} \delta \left(\frac{h_K}{|\vec{b}|} Lv_h, Lv_h \right)_{L^2(K)}. \tag{2.27}$$

Vale il seguente risultato di stabilità:

Lemma 2.3. *Sia u_h la soluzione fornita dallo schema GLS. Allora esiste una costante $C > 0$ che è indipendente da h, tale che*

$$\|u_h\|_{GLS} \leq C \|f\|_{L^2(\Omega)}.$$

Dimostrazione: *Scegliamo $v_h = u_h$ nella (2.26). Sfruttando il lemma (2.2) e la definizione (2.27) possiamo dapprima scrivere*

$$\|u_h\|^2_{GLS} = a_h^{(1)}(u_h, u_h) = f_h^{(1)}(u_h) = (f, u_h) + \sum_{K\in\tau_h} \delta \left(f, \frac{h_K}{|\vec{b}|} L u_h \right)_{L^2(K)}. \tag{2.28}$$

I due termini a secondo membro possono essere limitati dall'alto usando opportunamente le disuguaglianze di Cauchy-Schwarz e di Young; più precisamente si ha:

$$(f, u_h) \leq \frac{1}{4} \|\sqrt{\gamma} v\|^2_{L^2(\Omega)} + \left\| \frac{1}{\sqrt{\gamma}} f \right\|^2_{L^2(\Omega)},$$

$$\sum_{K\in\tau_h} \delta \left(f, \frac{h_K}{|\vec{b}|} L u_h \right)_K \leq \sum_{K\in\tau_h} \delta \left(\frac{h_K}{|\vec{b}|} f, f \right)_K + \frac{1}{4} \sum_{K\in\tau_h} \delta \left(\frac{h_K}{|\vec{b}|} L u_h, L u_h \right)_{L^2(K)}.$$

Sommando i due limiti superiori ed esplicitando di nuovo la definizione (2.27) otteniamo:

$$\|u_h\|^2_{GLS} \leq \left\| \frac{1}{\sqrt{\gamma} f} \right\|^2_{L^2(\Omega)} + \sum_{K\in\tau_h} \delta \left(\frac{h_K}{|\vec{b}|} f, f \right)_{L^2(K)} + \frac{1}{4} \|u_h\|^2,$$

che, ricordando che $h_K \leq h$ e definendo

$$C = \left(\frac{4}{3} \max_{x\in\Omega} \left(\frac{1}{\gamma} + \delta \frac{h}{|\vec{b}|} \right) \right)^{\frac{1}{2}},$$

diventa infine:

$$\|u_h\|^2_{GLS} \leq \frac{4}{3} \left[\left\| \frac{1}{\sqrt{\gamma}} f \right\|^2_{L^2(\Omega)} + \sum_{K\in\tau_h} \delta \left(\frac{h_K}{|\vec{b}|} f, f \right)_{L^2(K)} \right] \leq C^2 \|f\|^2_{L^2(\Omega)}.$$

Osservazione 2.1. *La precedente disuguaglianza è valida con l'unico vincolo che il parametro di stabilizzazione δ sia positivo. Infatti tale parametro potrebbe anche variare per ciascun elemento K della triangolazione τ_h. In questo caso avremmo δ_K invece di δ nelle formule precedenti, mentre la costante δ che compare nella definizione della costante C avrebbe il significato di $\max_{K\in\tau_h} \delta_K$.*

Ora presentiamo un risultato che concerne la convergenza del metodo GLS.

Teorema 2.1. *Supponiamo che lo spazio V_h soddisfi la seguente proprietà di approssimazione locale: $\forall v \in V \cap H^{r+1}(\Omega) \exists$ una funzione $\hat{v}_h \in V_h$ tale che*

$$\|v - \hat{v}_h\|_{L^2(K)} + h_K \|v - \hat{v}\|_{H^1(K)} + h_k^2 |v - \hat{v}_h|_{H^2(K)} \leq C \cdot h_K^{r+1} |v|_{H^{r+1}(K)}$$

per ciascun $K \in \tau_h$. Inoltre supponiamo che il numero di Péclet locale di K soddisfi la proprietà :

$$\mathbb{P}_{l_K}(x) = \frac{|b(x)| h_K}{2\mu} > 1 \quad \forall x \in K.$$

Infine ipotizziamo che valga la disuguaglianza inversa (2.24) e che il parametro di stabilizzazione soddisfi la relazione:

$$0 < \delta \leq 2C_0^{-1}.$$

Allora vale la seguente stima dell'errore associato allo schema GLS:

$$\|u - u_h\|_{GLS} \leq C \cdot h^{r+\frac{1}{2}} |u|_{H^{r+1}(\Omega)},$$

fintanto che $u \in H^{r+1}(\Omega)$.

Per lo scopo della presente tesina decidiamo di omettere la dimostrazione del teorema, rimandando il lettore curioso ad alcuni testi di approfondimento (vedi sezione 11.2 di [3]).

2.5.5 Due testi numerici del metodo GLS

In questo capitolo consideriamo due problemi test che risolviamo con l'ausilio del metodo GLS; in entrambi i casi richiamiamo brevemente l'aspetto teorico, implementiamo lo schema numerico mediante il software *Freefem++* e concludiamo con una serie di osservazioni sui risultati ottenuti, facendo variare i parametri più significativi in gioco.

test 1

Consideriamo il seguente problema parabolico in dimensione 2:

$$\begin{cases} u_t - \mu \Delta u + \vec{b} \cdot \nabla u = 1 & \text{in } \Omega \times [0,T] \\ u(x,y,0) = 0 & \text{in } \Omega \\ u(x,y,t) = 0 & \text{in } \partial\Omega \times [0,T] \end{cases}, \tag{2.29}$$

dove $\Omega = (0,1) \times (0,1)$ è il quadrato unitario, $\vec{b}$ è il campo $(1,1)^T$ ed $f \equiv 1$. Il metodo risolutivo agli elementi finiti di Galerkin stabilizzato con i Least Squares consente di determinare una $u_h \in V_h$:

$$a(u_h, v_h) + \mathcal{L}_h(u_h, f; v_h) = (f, v_h) \quad \forall v_h \in V_h, \tag{2.30}$$

per un'opportuna

$$\mathcal{L}_h \mid \mathcal{L}_h(u, f; v_h) = 0 \quad \forall v_h \in V_h.$$

Ricordiamo che una possibile scelta per $\mathcal{L}_h$ era data da:

$$\mathcal{L}_h(u_h, f; v_h) = \mathcal{L}_h^{(\rho)}(u_h, f; v_h) = \sum_{K \in \tau_h} \delta \left(Lu_h - f, S_K^{(\rho)}(v_h) \right)_{L^2(K)},$$

essendo L l'operatore del problema di diffusione-trasporto tale che:

$$\begin{cases} Lu = f & \text{in } \Omega \\ u = 0 & \text{in } \partial\Omega \end{cases},$$

con ρ e δ parametri da assegnare ed infine

$$S_K^{(\rho)} = \frac{h_K}{|\vec{b}|} \left[L_{SS} v_h + \rho L_S v_h \right].$$

Il metodo di GLS si ottiene scegliendo $\rho = 1$; in tal caso (2.30) diventa: determinare $u_h \in V_h$ tale che

$$\begin{aligned} (f, v_h) = F(v_h) &= a(u_h, v_h) + \mathcal{L}_h^{(1)}(u_h, f; v_h) = a(u_h, v_h) + \sum_{K \in \tau_h} \delta \left(Lu_h - f, S_K^{(1)}(v_h) \right)_{L^2(K)} \\ &= a(u_h, v_h) + \sum_{K \in \tau_h} \delta \left(Lu_h - f, \frac{h_K}{|\vec{b}|} Lv_h \right)_{L^2(K)} = a(u_h, v_h) + \frac{\delta}{\sqrt{2}} \sum_{K \in \tau_h} h_K \left(Lu_h - f, Lv_h \right)_{L^2(K)} \\ &= a(u_h, v_h) + \frac{\delta}{\sqrt{2}} \left[\sum_{K \in \tau_h} h_K \left(Lu_h, Lv_h \right)_{L^2(K)} - \sum_{K \in \tau_h} h_K \left(f \cdot Lv_h \right)_{L^2(K)} \right]. \end{aligned} \tag{2.31}$$

Per una questione di semplicità e per mettere in risalto le varie componenti del metodo GLS supponiamo ora che $h_K = h \; \forall K \in \tau_h$. Allora,

$$\begin{aligned} F(v_h) = a(u_h, v_h) + \frac{\delta h}{\sqrt{2}} \Bigg[&\int_\Omega \mu^2 \Delta u_h \Delta v_h - \mu \int_\Omega \Delta u_h \mathrm{div} v_h - \mu \int_\Omega \mathrm{div} u_h \Delta v_h \\ &+ \int_\Omega \mathrm{div} u_h \mathrm{div} v_h + \mu \int_\Omega f \Delta v_h - \int_\Omega f \mathrm{div} v_h \Bigg]. \end{aligned} \tag{2.32}$$

Separando opportunamente le varie componenti otteniamo (qui stiamo ancora considerando un termine forzante f generico):

$$a(u_h, v_h) = \mu \int_\Omega \nabla u_h \cdot \nabla v_h d\Omega + \int_\Omega \text{div} u_h v_h d\Omega;$$

$$a_h(u_h, v_h) = a(u_h, v_h) + \bar{\mathcal{L}}_h^{(1)},$$

essendo

$$\bar{\mathcal{L}}_h^{(1)} = \frac{\delta h}{\sqrt{2}} \left[\int_\Omega \mu^2 \Delta u_h \Delta v_h - \mu \int_\Omega \Delta u_h \text{div} v_h - \mu \int_\Omega \text{div} u_h \Delta v_h + \int_\Omega \text{div} u_h \cdot \text{div} v_h \right];$$

$$F(v_h) = \int_\Omega f v_h d\Omega;$$

$$F_h(v_h) = F(v_h) + \bar{\bar{\mathcal{L}}}_h^{(1)},$$

essendo

$$\bar{\bar{\mathcal{L}}}_h^{(1)} = -\frac{\delta h}{\sqrt{2}} \left[\mu \int_\Omega \Delta v_h d\Omega - \int_\Omega \text{div} v_h d\Omega \right].$$

Quindi la formulazione debole del problema è dato da:

$$\text{trovare } u_h \in V_h \ : \ a(u_h, v_h) + \bar{\mathcal{L}}_h^{(1)}(u_h, f; v_h) = F(v_h) + \bar{\bar{\mathcal{L}}}_h^{(1)}(u_h, f; v_h),$$

cioè

$$a_h(u_h, v_h) = F_h(v_h).$$

Nel caso specifico in cui $f \equiv 1$, in particolare si ha:

$$F(v_h) = \int_\Omega v_h d\Omega,$$

$$F_h(v_h) = F(v_h) - \frac{\delta h}{\sqrt{2}} \left[\mu \int_\Omega \Delta v_h d\Omega - \int_\Omega \text{div} v_h d\Omega \right].$$

Il problema che dobbiamo risolvere è in più evolutivo, in quanto dipendente dal tempo, e dunque va ricritto nel modo seguente:

$$\int_\Omega u_t(\vec{x}, t) v(\vec{x}) d\Omega + a_h(u, v) = F_h(v).$$

Applicando il θ-metodo alla u_h nella forma bilineare e alla funzione f, ed approssimando la derivata di u_h rispetto a t mediante un semplice rapporto incrementale, otteniamo finalmente:

$$\int_\Omega \frac{u_h^{n+1} - u_h^n}{dt} d\Omega + a_h \left((1-\theta) u_h^n + \theta u_h^{n+1}, v_h \right) = F_{h,\theta}(v).$$

Svolgendo tutti i calcoli e ricordando che in questo caso $f = 1$, otteniamo infine lo schema cercato (che scriveremo nel caso di h_K variabile):

$$\int_\Omega u_h^{n+1} v_h - \int_\Omega u_h^n v_h + \int_\Omega dt\mu(1-\theta)\nabla u_h^n \nabla v_h + \int_\Omega dt\mu\theta\nabla u_h^{n+1}\cdot\nabla v_h$$
$$+ \int_\Omega dt(1-\theta)\mathrm{div}u_h^n v_h + \int_\Omega dt\theta\mathrm{div}u_h^{n+1}v_h + \frac{\delta}{\sqrt{2}}\sum_{K\in\tau_h}\left[\int_K dt\, h_K\mu^2(1-\theta)\Delta u_h^n\Delta v_h\right.$$
$$+ \int_K dt\, h_K\mu^2\theta\Delta u_h^{n+1}\Delta v_h - \int_K dt\, h_K\mu(1-\theta)\Delta u_h^n\mathrm{div}v_h - \int_K dt\, h_K\mu\theta\Delta u_h^{n+1}\mathrm{div}v_h$$
$$- \int_K dt\, h_K\mu(1-\theta)\mathrm{div}u_h^n\Delta v_h - \int_K dt\, h_K\mu\theta\mathrm{div}u_h^{n+1}\Delta v_h + \int_K dt\, h_K(1-\theta)\mathrm{div}u_h^n\cdot\mathrm{div}v_h^n$$
$$\left.+ \int_K dt\, h_K\theta\mathrm{div}u_h^{n+1}\cdot\mathrm{div}v_h\right] = \int_\Omega dt v_h - \frac{\delta}{\sqrt{2}}\sum_{K\in\tau_h}\left[\int_K h_K dt\mu\Delta v_h - \int_K dt\, h_K\mathrm{div}v_h\right].$$

Lo schema GLS per il primo problema test in esame è implementato in *Freefem++* ed il codice è riportato nell'appendice 1 della presente tesina.

Procediamo riportando i risultati e facendo alcune considerazioni sui parametri in gioco. Riportiamo nella figura (2.2) il grafico della mesh prodotta dal software per completezza; in essa si evidenzia la triangolazione (scelta uniforme) in base alla quale opera il metodo FE.

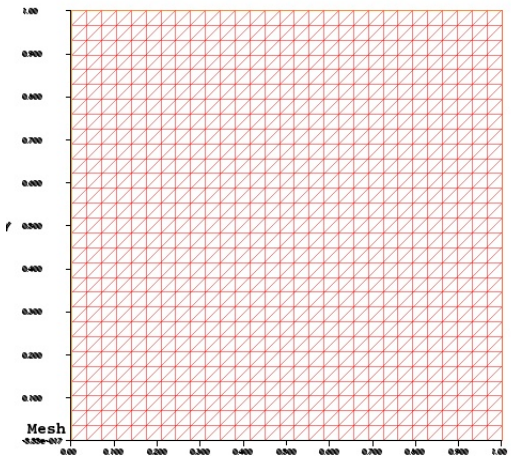


Figura 2.2: Rappresentazione di una griglia quadrata di lato 1 con passo di discretizzazione uniforme $h = \frac{1}{20}$.

Nella terna di figure (2.3) mostriamo come, nel caso di un coefficiente di viscosità μ dell'ordine di 10^{-3}, scegliendo un δ pari ad 1, la soluzione del problema mediante il metodo di Galerkin Least Squares è stabilizzata per un tempo finale T piuttosto basso, cioè $T = 1$.

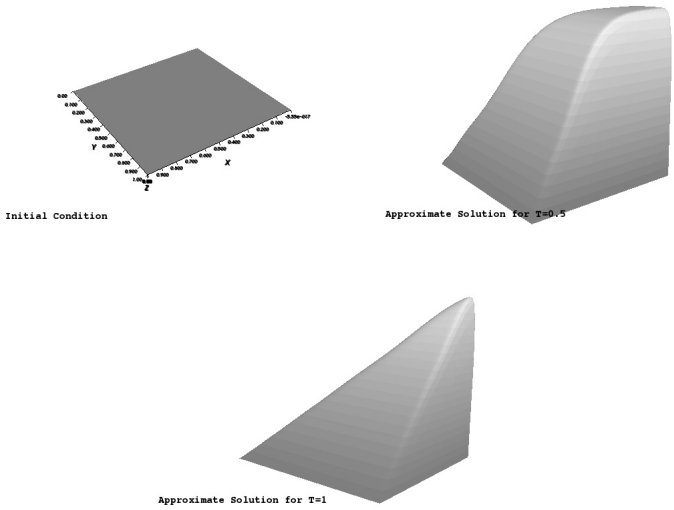


Figura 2.3: Nella terna di figure qui riportata si è scelto un passo di discretizzazione spaziale $h = \frac{1}{80}$ e temporale $dt = 0.1$ ed un valore del parametro per il θ-metodo pari a 0.5. Già al tempo $T = 1$, nel caso $\mu = 10^{-3}$ e $\delta = 1$, si ha la regolarizzazione della soluzione approssimata.

Risulta molto interessante confrontare l'approssimazione del problema (2.29), al variare del parametro μ e del passo di discretizzazione spaziale h, utilizzando il metodo di Galerkin standard ($\delta = 0$) ed il metodo di Galerkin GLS (qui si è scelto $\delta = 1.5$). Le tre coppie di grafici (2.4), (2.5), (2.6) riportano a sinistra il primo metodo ed a destra il secondo metodo e nella didascalia sotto tali grafici sono precisati tutti gli altri parametri in gioco.

Dall'analisi dei due metodi ci si può convincere, anche solo osservando i grafici prodotti mediante la simulazione in *Freefem++*, del fatto che, grazie all'introduzione del parametro di stabilizzazione δ, la soluzione fornita dal metodo di Galerkin Least Squares è depurata di tutte le fluttuazioni alle quali è invece soggetta la soluzione fornita dal metodo di Galerkin standard, al crescere dei numeri di Péclet ottenuti combinando i differenti valori di μ e di h.

Aggiungiamo infine un'ultima coppia di grafici (2.7) per enfatizzare tale diverso comportamento dei due metodi.

Laddove Galerkin standard fallisce, GLS è in grado di fornire una soluzione numerica accettabile anche per numeri di Péclet estremamente alti. Osserviamo, a conclusione della trattazione del primo esempio test, che tale situazione si ripresenta anche utilizzando lo spazio $P2$ o $P3$ in FE e pertanto ne omettiamo i dettagli.

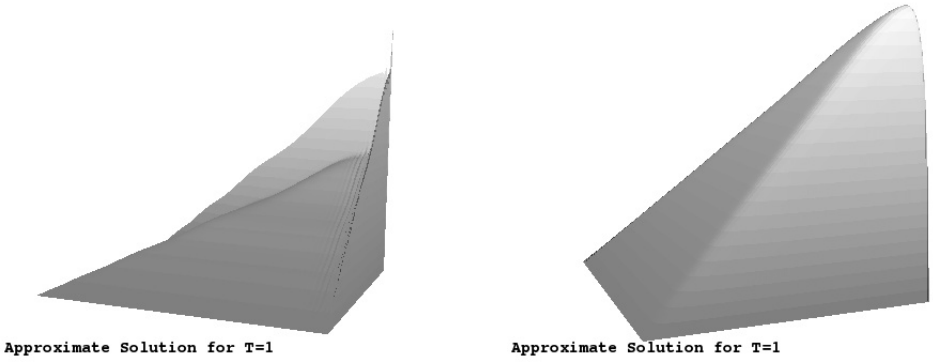


Figura 2.4: Approssimazione del problema (2.29) con $\mu = 10^{-3}$, $h = \frac{1}{80}$, $dt = 0.1$ usando Galerkin standard (sinistra) e Galerkin Least Squares (destra). Il corrispondente numero di Péclet è $\mathbb{P}_e = 8.84$.

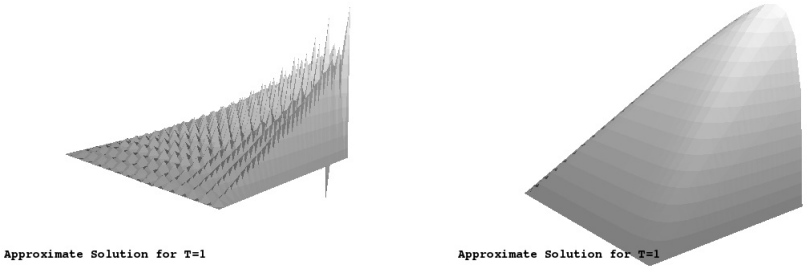


Figura 2.5: Approssimazione del problema (2.29) con $\mu = 10^{-3}$, $h = \frac{1}{20}$, $dt = 0.1$ usando Galerkin standard (sinistra) e Galerkin Least Squares (destra). Il corrispondente numero di Péclet è $\mathbb{P}_e = 35.35$.

test 2

Consideriamo il seguente problema parabolico in dimensione 2:

$$\begin{cases} u_t - \mu\Delta u + \vec{b}\cdot\nabla u = 0 & \text{in } Q = \Omega \times [0,T) \\ u(x,y,0) = 0 & \text{in } \Omega \\ u(x,y,t) = \phi(x,y,t) & \text{in } \partial_e\Omega \times [0,T) \\ \frac{\partial u}{\partial n} = 0 & \text{in } \partial_i\Omega \times [0,T) \end{cases}, \tag{2.33}$$

essendo $\Omega = (-1,1)\times(-1,1)$ il quadrato di lato 2 privato al suo interno del cerchio di raggio pari a 0.15 la cui circonferenza abbiamo indicato con il simbolo $\partial_i\Omega$ (mentre il bordo esterno del quadrato con $\partial_l\Omega$). Sulla circonferenza $\partial_i\Omega$ imponiamo una condizione al bordo di Neumann omogenea, ad indicare una situazione fisica di *parete buco isolante*, mentre in $\partial_l\Omega$ imponiamo le

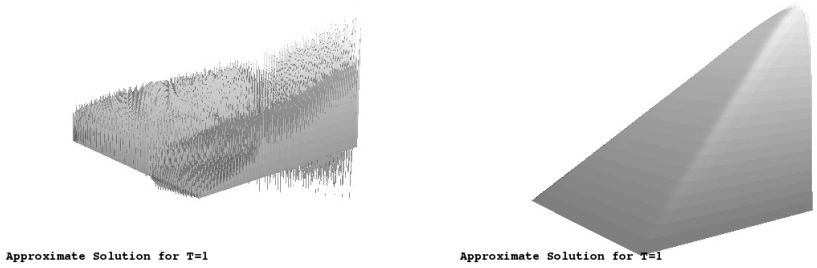


Figura 2.6: Approssimazione del problema (2.29) con $\mu = 10^{-3}$, $h = \frac{1}{80}$, $dt = 0.1$ usando Galerkin standard (sinistra) e Galerkin Least Squares (destra). Il corrispondente numero di Péclet è $\mathbb{P}_e = 883.88$.

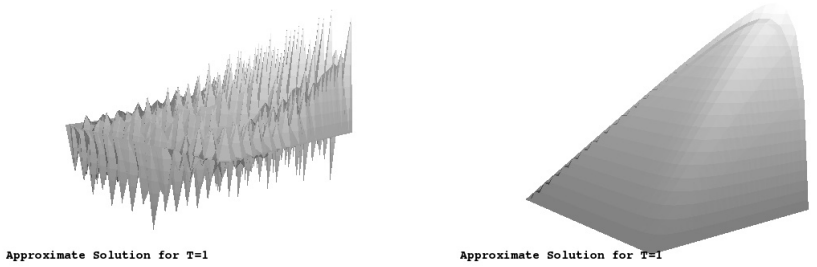


Figura 2.7: Approssimazione del problema (2.29) con $\mu = 10^{-3}$, $h = \frac{1}{20}$, $dt = 0.1$ usando Galerkin standard (sinistra) e Galerkin Least Squares (destra). Il corrispondente numero di Péclet è $\mathbb{P}_e = 3535.5$.

condizioni di Dirichlet miste fornite dalla seguente definizione:

$$\phi(x, y, t) = \begin{cases} 1 & \text{per } x = -1, \quad -1 \leq y \leq 1 \\ 0 & \text{altrove} \end{cases}$$

(che fisicamente si può interpretare come se inizialmente avessimo del calore tutto concentrato su una singola parete, mentre le altre tre pareti del quadrato sono fredde). Infine la parte convettiva è rappresentata dal seguente campo vettoriale:

$$\vec{b}(x, y) = \big(y(1 - x^2),\, -x(1 - x^2)\big)^T .$$

Ripetendo il ragionamento visto per il primo test si arriva alla formulazione debole del problema in esame, all'algoritmo di Galerkin stabilizzato (GLS) e alla sua implementazione in *Freefem++* (il cui codice è riportato nell'appendice 2). Procediamo come nel caso del primo test riportando i risultati e facendo alcune considerazioni sui parametri in gioco. In figura (2.8) abbiamo riportato, nel caso di un coefficiente di viscosità μ dell'ordine di 10^{-3},

scegliendo un δ pari a 5, la soluzione del problema mediante il metodo GLS per un tempo finale $T = 10$, avendo altresí costruito una griglia non strutturata con parametro di suddivisione della griglia $n = 100$, $dt = 0.1$. La terna di figura è il risultato dell'approssimazione del problema mediante lo spazio FE $P1$. Nella stessa figura riportiamo per completezza anche il grafico della soluzione approssimata in $3d$, laddove nella terna il terzo grafico riportato è in $2d$.

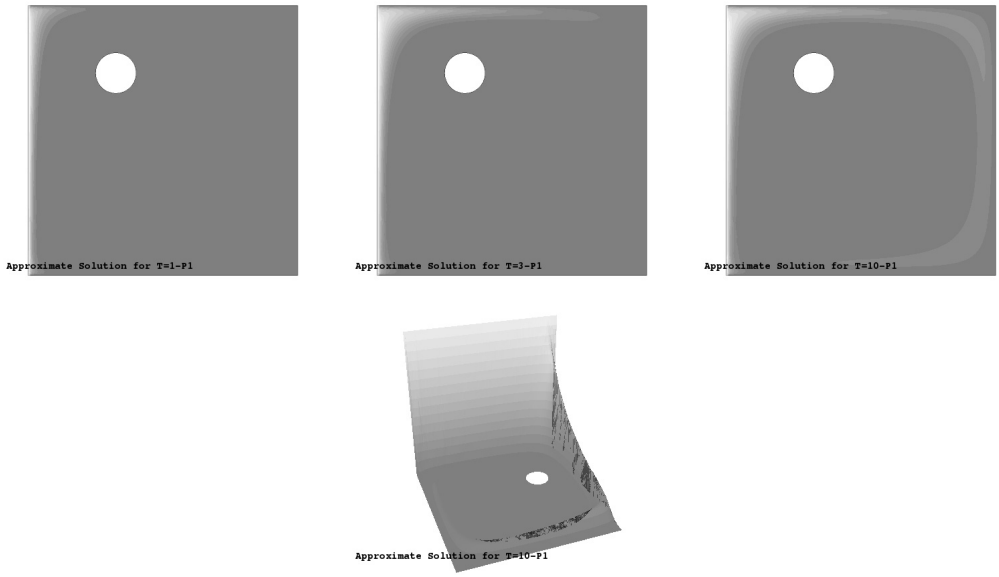


Figura 2.8: Soluzione approssimata del problema (2.33) con $\mu = 10^{-3}$, $\delta = 5$, $dt = 0.1$, suddivisione della griglia in $n = 100$ nodi, con rappresentazione finale in dimensione 2 e 3. Si osservi la prevalenza del comportamento convettivo, dato che $\mu << 1$.

Come si può notare dai grafici, essendo $\mu << 1$, il problema è a trasporto dominante e, a meno di scegliere un δ sufficientemente grande ($\delta > 10$), il comportamento della soluzione approssimata presenta i vortici tipici del caso; nella terna di figure (2.9) riportiamo gli stessi grafici prendendo però stavolta un valore di μ dell'ordine di 10^{-1} ed un $\delta = 10$ ed in tal caso prevale il comportamento diffusivo della soluzione numerica, con il calore che si diffonde più omogeneamente ad onde dalla parete inizialmente calda. Si è optato ancora per lo spazio FE $P1$, rimandando a considerazioni successive l'uso dello spazio FE $P2$. In figura (2.9) poniamo ancora in evidenza la differenza tra il caso del metodo di Galerkin standard (non stabilizzato con $\delta = 0$), un caso semiregolarizzato e il metodo stabilizzato, sia lavorando sullo spazio FE $P1$ che in FE $P2$. All'aumentare di δ si osserva una sempre maggiore prevalenza dell'effetto diffusivo rispetto a quello convettivo.

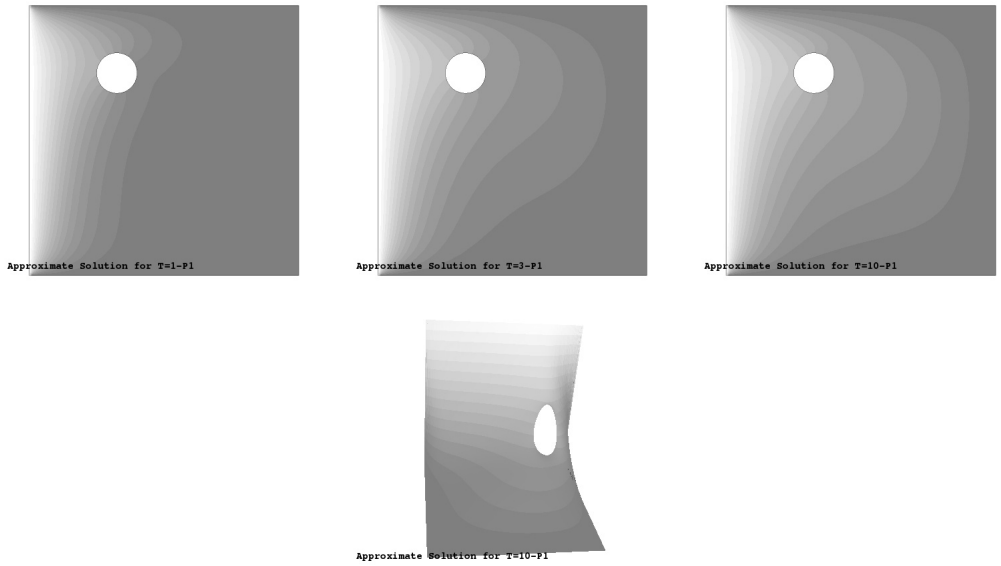


Figura 2.9: Soluzione approssimata del problema (2.33) con $\mu = 10^{-1}$, $\delta = 10$, $dt = 0.1$, suddivisione della griglia in $n = 100$ nodi, con rappresentazione finale in dimensione 2 e 3. Si osservi la prevalenza del comportamento diffusivo, dato che μ e δ sono molto più grandi rispetto al caso precedente ($\mu >> 1$).

Nelle figure (2.10), (2.11), (2.12), (2.13) abbiamo scelto un numero di suddivisione della griglia $n = 35$ per evidenziare l'effetto di stabilizzazione del metodo GLS.

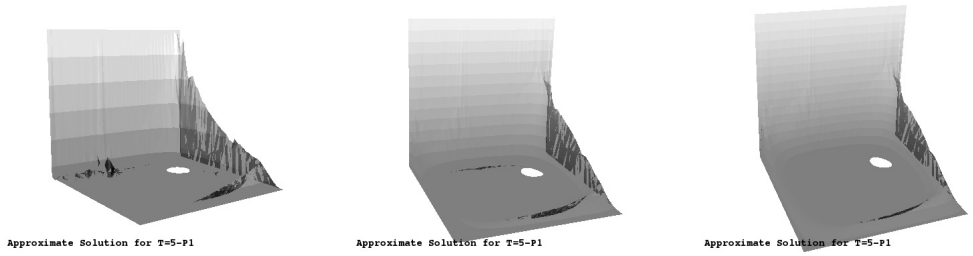


Figura 2.10: Soluzione approssimata del problema (2.33) con $\mu = 10^{-3}$, $T = 5$, $n = 35$, $dt = 0.1$, $\theta = 0.5$, spazio FE $P1$, nei casi $\delta = 0$, $\delta = 5$, $\delta = 10$, andando da sinistra verso destra in domensione 3.

Osserviamo anzitutto che, con una suddivisione della griglia usando $n = 35$ la soluzione numerica non presenta particolare instabilità per valori di μ dell'ordine di 10^{-3} ed è per questo motivo che abbiamo scelto tre valori di δ abbastanza distaccati tra loro nella prima terna di figure ($\delta = 0$, $\delta = 5$, $\delta = 10$), proprio per mettere in risalto la seppur poco percepibile regolarizzazione della suddetta soluzione.

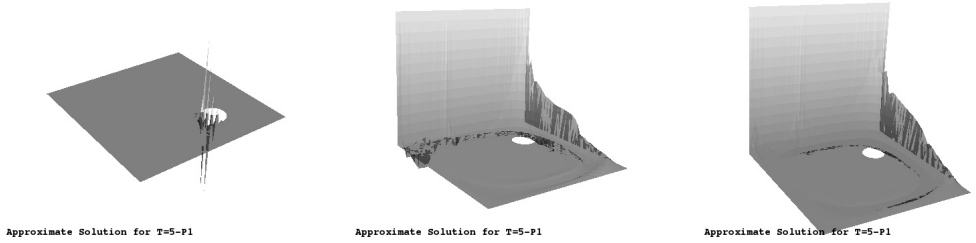


Figura 2.11: Soluzione approssimata del problema (2.33) con $\mu = 10^{-5}$, $T = 5$, $n = 35$, $dt = 0.1$, $\theta = 0.5$, spazio FE $P1$, nei casi $\delta = 0$, $\delta = 1$, $\delta = 3$, andando da sinistra verso destra in domensione 3.

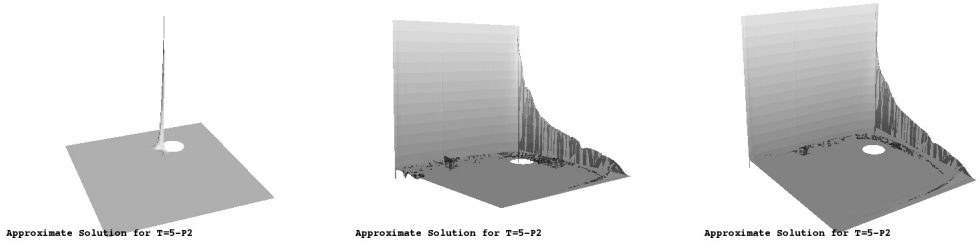


Figura 2.12: Soluzione approssimata del problema (2.33) con $\mu = 10^{-4}$, $T = 5$, $n = 35$, $dt = 0.1$, $\theta = 0.5$, spazio FE $P2$, nei casi $\delta = 0$, $\delta = 1$, $\delta = 3$, andando da sinistra verso destra in domensione 3.

Nella terna (2.11), con un $\mu \approx 10^{-5}$ la regolarizzazione della soluzione apportata dal metodo GLS diventa nettamente più evidente; si noti ad esempio come, in tal caso, anche solo passando da $\delta = 0$ al caso semiregolarizzato con $\delta = 1$, le fortissime oscillazioni vengano drasticamente smorzate.

Nelle terne delle figure (2.12) e (2.13) ripetiamo la simulazione lasciando invariato $\mu = 35$ e $\theta = 0.5$ e passiamo dallo spazio FE $P1$ ad FE $P2$. Consideriamo i due casi $\mu \approx 10^{-4}$ e $\mu \approx 10^{-5}$ perché, come abbiamo già avuto modo di sperimentare nel caso $P1$ per μ fino all'ordine di 10^{-3} le oscillazioni della soluzione numerica non risultano particolarmente apprezzabili sí da cogliere l'efficacia del metodo GLS. Ribadiamo che, aumentando il partizionamento della griglia prendendo $n > 35$, osserviamo che per vedere le oscillazioni della soluzione numerica e quindi l'effetto stabilizzante del metodo GLS occorre considerare un ordine di grandezza di μ ancora minore, ma il ragionamento concernente il comportamento della suddetta soluzione è analogo a quanto visto con $n = 35$.

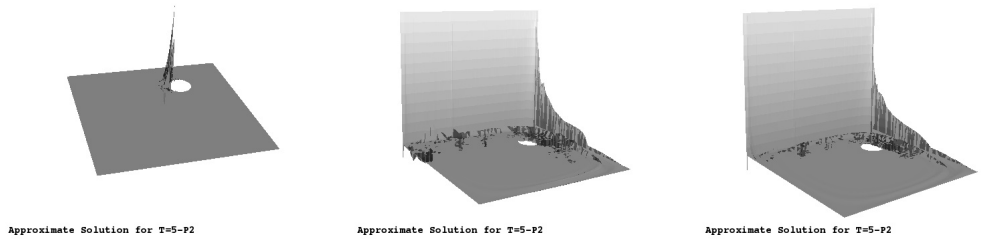


Figura 2.13: Soluzione approssimata del problema (2.33) con $\mu = 10^{-5}$, $T = 5$, $n = 35$, $dt = 0.1$, $\theta = 0.5$, spazio FE $P2$, nei casi $\delta = 0$, $\delta = 1$, $\delta = 3$, andando da sinistra verso destra in domensione 3.

Capitolo 3

Applicazione del metodo GLS

3.1 Presentazione del problema

L'obiettivo finale del presente lavoro consiste nel risolvere numericamente con il metodo GLS un'equazione evolutiva di tipo diffusione-trasporto per il profilo di concentrazione instabile di una certa sostanza in un dominio Ω. L'equazione alla quale facciamo riferimento è la seguente:

$$u_t - \frac{1}{\mathbb{P}_e}\Delta u + \vec{b}\cdot\nabla u = 0, \quad x \in \Omega, \;\; t > 0, \tag{3.1}$$

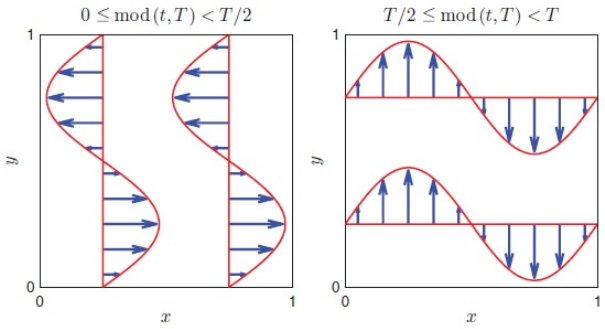
dove al solito $\vec{b}$ rappresenta il campo vettoriale che esprime la velocità di trasporto, t è il tempo e $\mathbb{P}_e$ è il numero di Péclet. In genere è utile nelle applicazioni considerare condizioni al bordo di Neumann omogenee (caso nel quale non si ha alcun flusso) o periodiche, condizioni queste ultime che ci assicurano che la massa totale della sostanza considerata nel dominio Ω data dall'integrale $\int_\Omega u dx$ rimanga costante nel tempo. Per quanto invece concerne il numero di Péclet è interessante considerare valori differenti, ai quali corrispondono condizioni fisiche estremamente diverse; tipicamente gli ordini di grandezza che si confrontano per tale parametro sono 10^2 per flussi laminari, da 10^3 a 10^5 escluso per dyes molecolari in tipiche soluzioni microfluidiche d'acqua/glicerolo, 10^5 per sostanze granulari in rotating tumblers ed infine 10^{10} per flussi reattivi turbolenti. Il dominio è $Q = \Omega \times [0, T]$, essendo Ω il quadrato $(0, 1)^2$ e T il periodo. In particolare prendiamo come campo della velocità di trasporto quello indicato con il nome *time-periodic sine flow* (abbreviato con la sigla TPSF), avente la seguente espressione analitica:

$$\vec{b}(x, y, t) = \begin{cases} \sin(2\pi y)\hat{x}, & 0 \leq \mathrm{mod}(t, T) < \frac{T}{2} \\ \sin(2\pi x)\hat{y}, & \frac{T}{2} \leq \mathrm{mod}(t, T) < T \end{cases},$$

dove $\mathrm{mod}(t,T) = t - qT$, con q numero intero tale che

$$0 \leq t - qT < T.$$

Il campo $\vec{b}$ considerato è un profilo di velocità che alterna tra componenti ortogonali unidirezionali; il periodo del flusso T può essere un qualunque numero positivo e $\frac{T}{2}$ rappresenta quanto a lungo il flusso agisce in una direzione prima di dirottarsi nella direzione perpendicolare. Nell'immagine in figura (3.1) riportiamo proprio il comportamento di tale campo vettoriale.



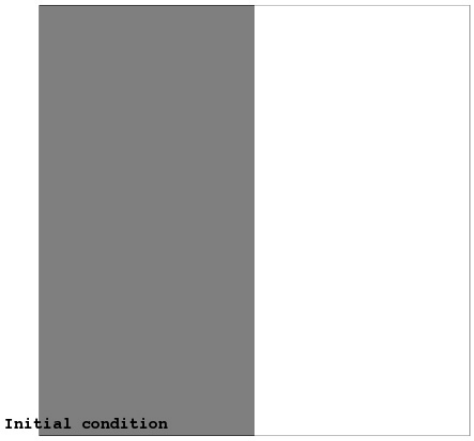
Nelle simulazioni che faremo partiremo da un dato iniziale relativo alla concentrazione della sostanza caratterizzato dal fatto che metà del dominio $\left(x > \frac{1}{2}\right)$ ha una concentrazione massima (pari ad 1, indicata in bianco), mentre l'altra metà ha una concentrazione minima (pari a 0, indicata in nero). Per quanto concerne il periodo T, consideriamo due casi: $T = 0.8$ e $T = 1.6$. La scelta di tali valori è legata al fatto che periodi più piccoli di 0.8 non permettono di ottenere un mescolamento significativo della sostanza, mentre periodi più lunghi di 1.6 non comportano un sostanziale incremento del mescolamento.

Utilizzeremo il metodo di Galerkin Least Squares per simulare il comportamento della sostanza in esame e poi confronteremo i grafici ottenuti con quelli ricavati applicando una classe di tecniche note come *mapping methods* (che è la tecnica utilizzata dagli autori dell'articolo di riferimento; per ulteriori dettagli vedi [7]).

3.2 Risultati numerici e confronto con il metodo *Mapping*

Cominciamo ora con il presentare i risultati numerici delle simulazioni sul problema in esame al variare di differenti parametri. Il primo grafico in figura (3.2) è ovviamente costituito dalla condizione iniziale, quella cioè che,

come abbiamo accennato poco fa, corrisponde fisicamente alla condizione di netta separazione tra la zona a concentrazione 1 (contrassegnata dal color bianco) a quella a concentrazione 0 (contrassegnata dal colore nero).



La prima simulazione consiste nel considerare il caso limite in cui il numero di Péclet è ∞; esso corrisponde alla situazione fisica di puro trasporto (cioè la sostanza non diffonde) e quindi $\mu = 0$; nella prima coppia di figure (3.1) è riportato il comportamento della sostanza nel caso in cui si fissi il periodo $T = 0.8$, il numero dei nodi della griglia ad $n = 25$, per $\delta = 0$ e tempo finale dapprima $T_{fin} = T$ e poi $T_{fin} = 5T$ nello spazio degli elementi finiti $P1$; nella seconda coppia di figura (3.2) abbiamo invece variato δ, che ora assume il valore 0.25, lasciando inalterati tutti gli altri parametri.

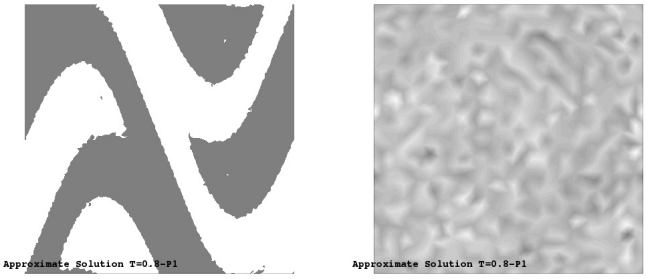


Figura 3.1: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = \infty$, $\delta = 0$, $T = 0.8$, FE $P1$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

Le successive due coppie di grafici (3.3), (3.4) rappresentano il comportamento della sostanza nel caso in cui si fissi $T = 1.6$, lasciando inalterate tutte le restanti grandezze in gioco e considerando come in precedenza i due casi distinti in cui il parametro di stabilizzazione δ sia la prima volta 0 e la seconda $\delta = 0.25$. La distinzione tra i due valori del periodo T (cioè $T = 0.8$

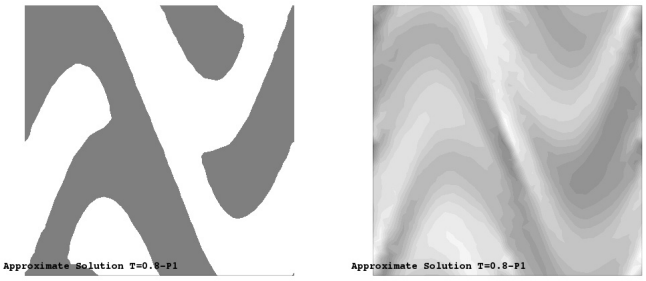


Figura 3.2: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = \infty$, $\delta = 0.25$, $T = 0.8$, FE $P1$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

e $T = 1.6$) implica fisicamente un comportamento differente della sostanza: per $T = 1.6$ il flusso è completamente caotico in tutto il dominio, mentre per $T = 0.8$ il flusso è parzialmente caotico, presentando sia regioni caotiche che regolari.

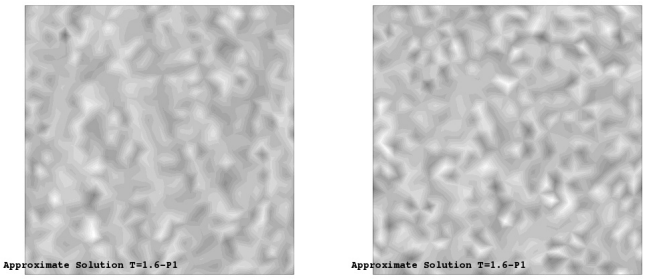


Figura 3.3: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = \infty$, $\delta = 0$, $T = 1.6$, FE $P1$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

Come si evidenzia in tutti i casi finora considerati, la presenza del coefficiente di stabilizzazione introduce una diffusività nella soluzione approssimata, a discapito della componente puramente convettiva. Il caso $\mathbb{P}_e = \infty$ rappresenta idealmente la condizione fisica propria dei flussi reattivi turbolenti. Tale condizione è stata approssimata nel codice *Freefem++* prendendo $\mu \approx 10^{-7}$.

La successiva simulazione nelle figure (3.5), (3.6) consiste nel testare il metodo GLS nel caso in cui il numero di Péclet è dell'ordine 10^5 (e quindi $\mu \approx 10^{-5}$), fisicamente incarnato dal comportamento delle sostanze granulari nei *rotating tumblers*, anche qui per i due valori di $\delta = 0$ e $\delta = 0.25$; questa volta abbiamo scelto di utilizzare il metodo GLS nello spazio FE $P2$.

Ripetiamo poi nelle figure (3.9)-(3.12) la simulazione con il metodo GLS per $\mathbb{P}_e = 10^3$ (e perciò $\mu \approx 10^{-3}$), corrispondente al comportamento fisico che si riscontra per le *dyes* molecolari in tipiche soluzioni di acqua/glicerolo

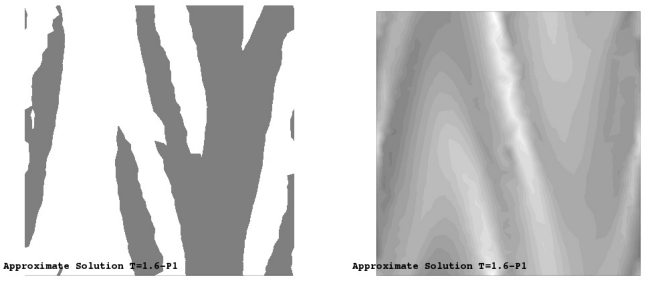


Figura 3.4: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = \infty$, $\delta = 0.25$, $T = 1.6$, FE $P1$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

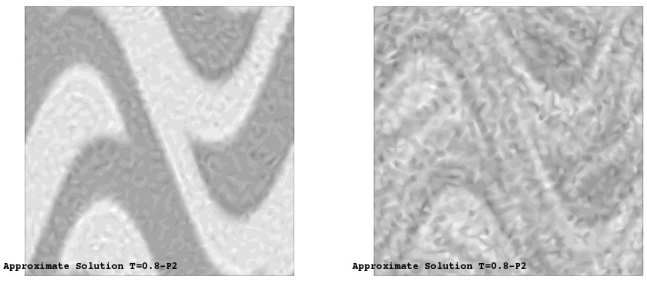


Figura 3.5: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^5$, $\delta = 0$, $T = 0.8$, FE $P2$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

microfluidiche, sempre per valori di δ pari a 0 e a 0.25; utilizziamo il metodo GLS ancora nello spazio FE $P2$.

Per maggior completezza riportiamo in figura (3.2) l'evoluzione temporale della soluzione numerica del problema (3.1) nel caso $T = 0.8$ e $T_{fin} = 5 \cdot 0.8$ effettuando la regolarizzazione.

L'ultima simulazione riguarda il test del metodo GLS per $\mathbb{P}_e \approx 10^{-2}$, cioè per $\mu \approx 10^{-2}$, che descrive il comportamento fisico dei flussi laminari.

Riportiamo in figura (3.17) non tanto l'evoluzione temporale della soluzione numerica del problema (3.1) quanto piuttosto la diversità della regolarizzazione della soluzione al tempo finale, al variare dei valori del coefficiente δ.

Chiaramente, al diminuire del numero di $\mathbb{P}_e$ le oscillazioni della soluzione numerica per $\delta = 0$ diminuiscono e conseguentemente anche le differenze con la soluzione numerica trovata con il metodo GLS per $\delta > 0$. Nelle figure (3.18), infine, riportiamo le soluzioni numeriche del problema (3.1) per $T = 0.8$ e $T_{fin} = 5 \cdot 0.8$, tenendo fisso il coefficiente di stabilizzazione $\delta = 0.5$ e facendo variare il numero di Péclet (e quindi il coefficiente di diffusione μ) tra i seguenti valori: $\mathbb{P}_e = \infty$, $\mathbb{P}_e = 10^5$, $\mathbb{P}_e = 10^3$, $\mathbb{P}_e = 10^2$.

Dal confronto della nostra simulazione ottenuta mediante l'utilizzo del

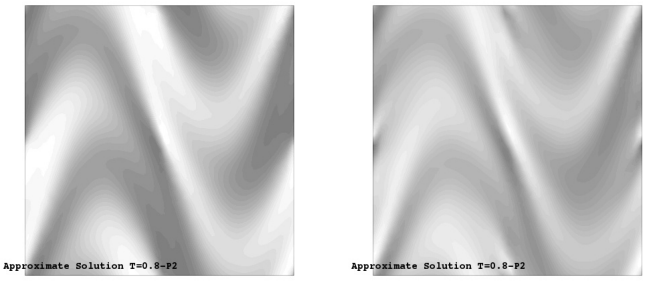


Figura 3.6: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^5$, $\delta = 0.25$, $T = 0.8$, FE $P2$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

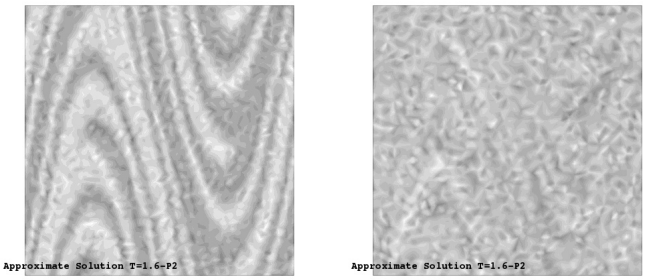


Figura 3.7: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^5$, $\delta = 0$, $T = 1.6$, FE $P2$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

metodo GLS con la tecnica di approssimazione numerica dagli autori dell'articolo possiamo concludere, soprattutto analizzando i grafici di entrambe le simulazioni (figure (3.19)-(3.23)), che i risultati numerici sono del tutto confrontabili; per sottolineare ciò riportiamo le più significative coppie di grafici dei due metodi.

A conclusione del presente lavoro possiamo mettere in risalto alcuni aspetti che, a mio avviso, sarebbe interessante, oltreché utile, approfondire: 1) il ruolo giocato dal parametro di regolarizzazione δ nel metodo GLS; in effetti, anche se la stabilità (addirittura assoluta) e la convergenza del metodo GLS sono garantiti per $\delta > 0$, andrebbe sicuramente delucidato come tale metodo varia al variare dei valori assunti da δ e capire meglio l'interazione di quest'ultimo con gli altri parametri in gioco, quali in primis μ ed il campo vettoriale $\vec{b}$.
2) Un confronto più approfondito con i differenti metodi regolarizzanti, quali lo Streamline Upwind Petrov-Galerkin e il metodo di Douglas-Wang ai quali abbiamo accennato precedentemente, è auspicabile a fine di comprendere i punti di forza dei vari metodi ed ottimizzare al massimo la stabilizzazione delle eventuali oscillazioni della soluzione numerica.
3) Infine, l'analisi del metodo GLS e di analoghi metodi può costituire il

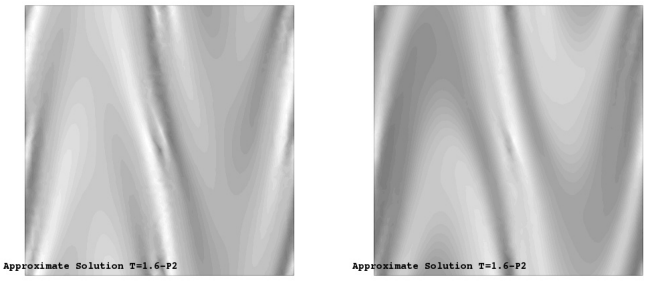


Figura 3.8: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^5$, $\delta = 0.25$, $T = 1.6$, FE $P1$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

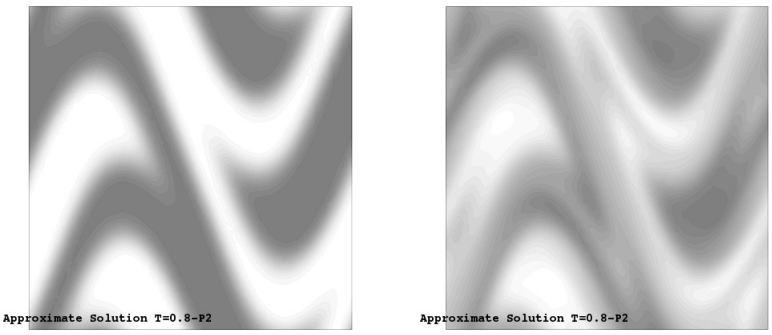


Figura 3.9: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^3$, $\delta = 0$, $T = 0.8$, FE $P2$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

punto di partenza per l'elaborazione e ottimizzazione dei metodi e conseguenti algoritmi che siano sempre più efficienti e duttili al fine di simulare e modellizzare la soluzione approssimata di problemi più complessi.

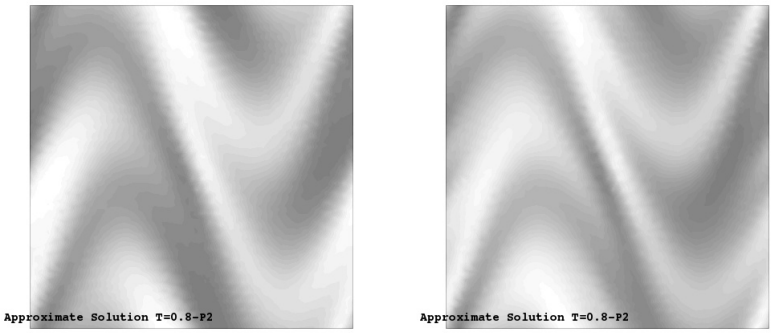


Figura 3.10: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^3$, $\delta = 0.25$, $T = 0.8$, FE $P2$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

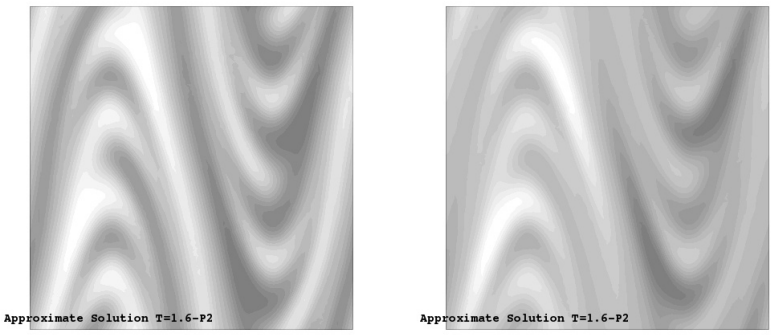


Figura 3.11: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^3$, $\delta = 0$, $T = 1.6$, FE $P2$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

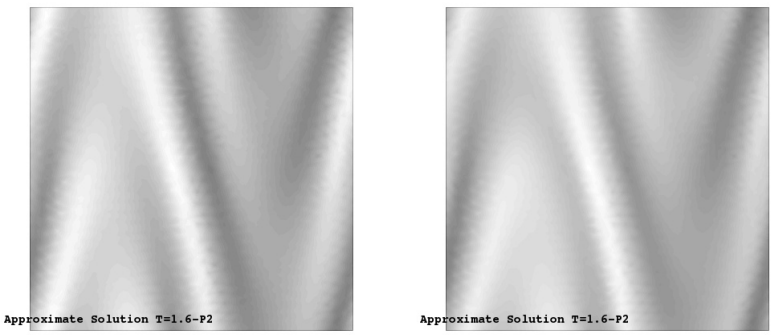


Figura 3.12: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^3$, $\delta = 0.25$, $T = 1.6$, FE $P2$, $n = 25$, $dt = 0.01$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

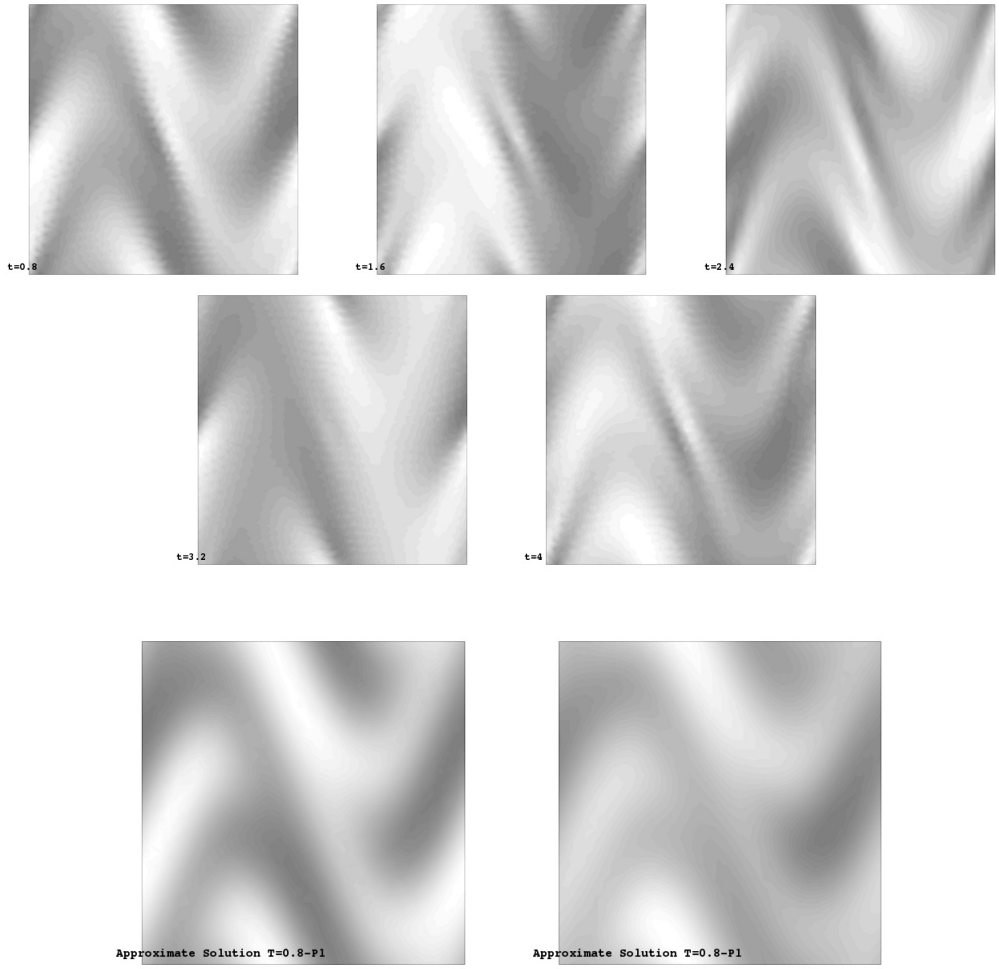


Figura 3.13: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^2$, $\delta = 0$, $T = 0.8$, $n = 45$, $dt = 0.01$, FE $P1$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

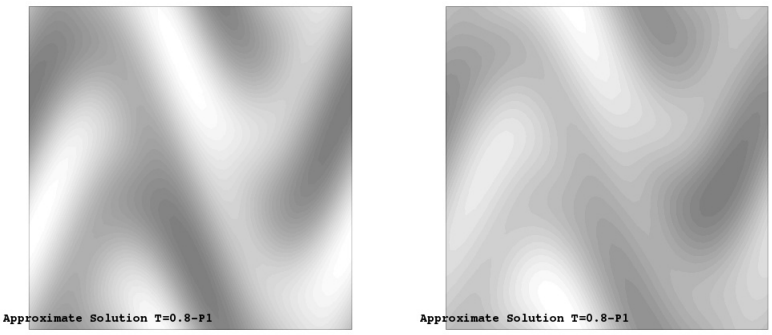


Figura 3.14: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^2$, $\delta = 0.25$, $T = 0.8$, $n = 45$, $dt = 0.01$, FE $P1$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

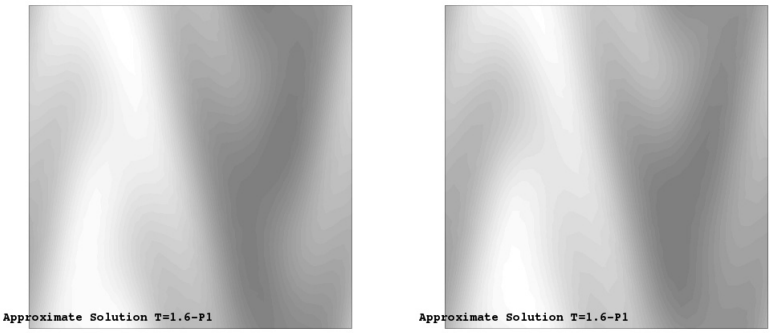


Figura 3.15: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^2$, $\delta = 0$, $T = 1.6$, $n = 45$, $dt = 0.01$, FE $P1$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

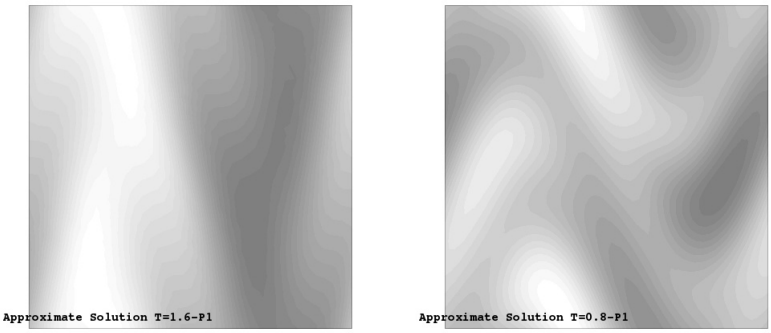


Figura 3.16: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^2$, $\delta = 0.25$, $T = 1.6$, $n = 45$, $dt = 0.01$, FE $P1$; nella figura di sinistra il tempo finale è T, in quella di destra $5T$.

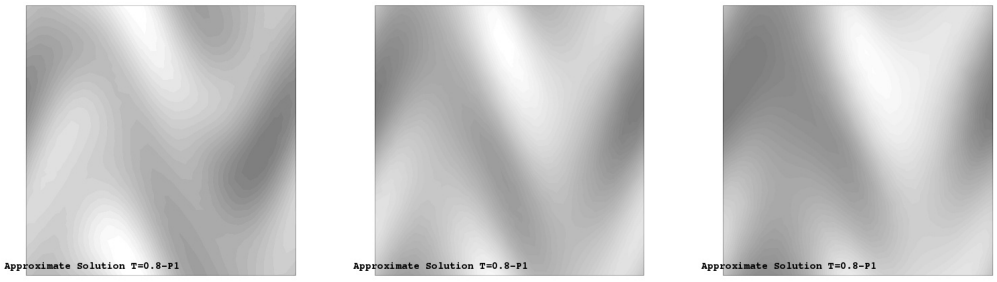


Figura 3.17: Confronto delle soluzioni numeriche del problema (3.1) per $T = 0.8$, $T_{fin} = 5T$, $n = 45$, $dt = 0.01$, FE $P1$ per valori crescenti di δ ($\delta = 0.25$, 1, 1.05), nel caso di flussi laminari.

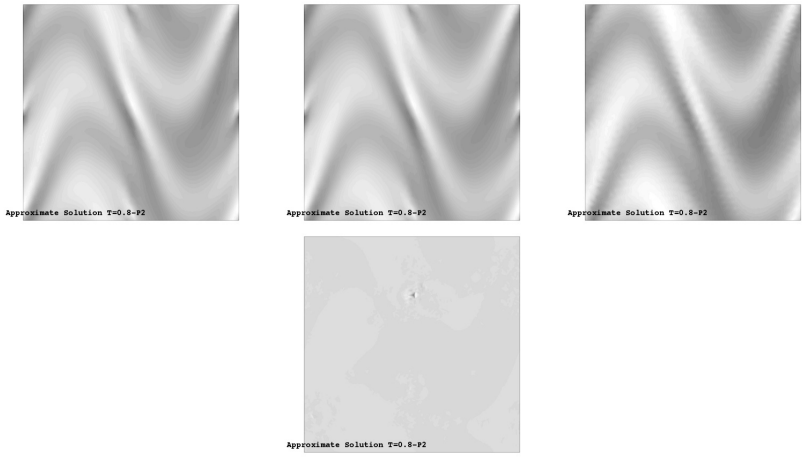


Figura 3.18: Confronto delle soluzioni numeriche del problema (3.1) per $T = 0.8$, $T_{fin} = 5T$, $n = 45$, $dt = 0.01$, FE $P2$, tenendo fissi $\delta = 0.25$ per i seguenti valori di Péclet: ∞, 10^5, 10^3, 10^2.

Figura 3.19: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = \infty$, $\delta = 0.25$, $T = 0.8$, $T_{fin} = T$ FE $P2$, $n = 80$, $dt = 0.01$; a sinistra con la tecnica *Mapping* e a destra con il metodo GLS.

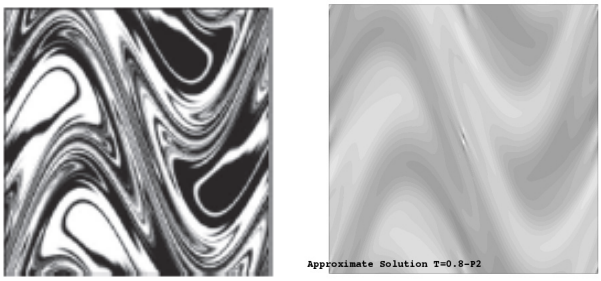


Figura 3.20: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = \infty$, $\delta = 0.25$, $T = 0.8$, $T_{fin} = 5T$ FE $P2$, $n = 80$, $dt = 0.01$; a sinistra con la tecnica *Mapping* e a destra con il metodo GLS.

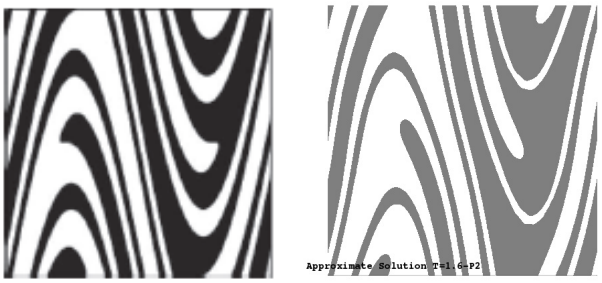


Figura 3.21: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = \infty$, $\delta = 0.25$, $T = 1.6$, $T_{fin} = T$ FE $P2$, $n = 80$, $dt = 0.01$; a sinistra con la tecnica *Mapping* e a destra con il metodo GLS.

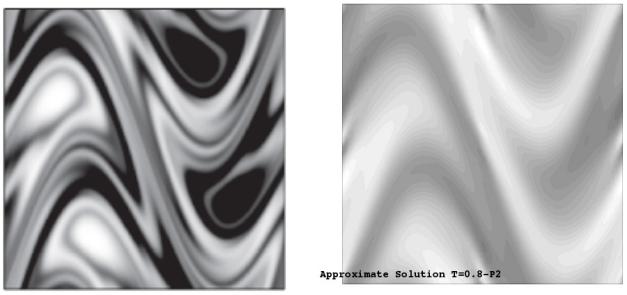


Figura 3.22: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^4$, $\delta = 0.25$, $T = 0.8$, $T_{fin} = 5T$ FE $P2$, $n = 45$, $dt = 0.01$; a sinistra con la tecnica *Mapping* e a destra con il metodo GLS.

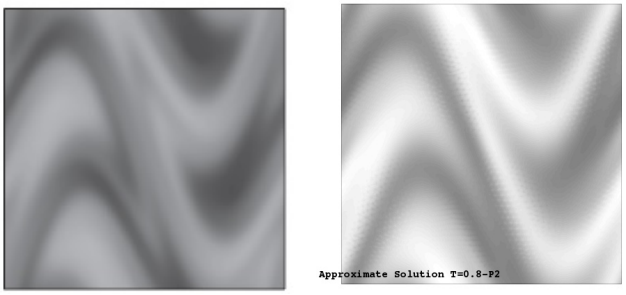


Figura 3.23: Soluzione approssimata del problema (3.1) con $\mathbb{P}_e = 10^3$, $\delta = 0.25$, $T = 0.8$, $T_{fin} = 5T$ FE $P2$, $n = 45$, $dt = 0.01$; a sinistra con la tecnica *Mapping* e a destra con il metodo GLS.

Appendice A

Codice test 1

```
////////////////////////////////////////////////////////////////////////////
//                                                                        //
//  Eq. di diffusione-trasporto evolutiva; Omega = quadrato 1 x 1         //
//                                                                        //
//      u_t - mu*Lap u + div u = 1   in Omega x [0,T]                     //
//           u(x,y,0) = 0   in Omega                                      //
//           u(x,y,t) = 0   on dOmega   x [0,T]                           //
//      div u = b * grad u, dove qui b = (1,1)^T                          //
////////////////////////////////////////////////////////////////////////////


//Generazione della griglia quadrata
real x0=0,x1=1;
real y0=0,y1=1;
int n=29, m=29;            // tale scelta corrisponde ad aver scelto un passo di discretizzazione uniforme
// in spazio h pari a circa 1/20; infatti h=sqrt(2)/(n-1)=1/20 mi dà 29,3.
//int n=114, m=114 ;       // tale scelta corrisponde ad aver scelto un passo di discretizzazione uniforme
// in spazio h pari a circa 1/80; infatti  h=sqrt(2)/(n-1)=1/80 mi dà 114,1.

mesh Th=square(n,m,[x0+(x1-x0)*x,y0+(y1-y0)*y]);

plot(Th,cmm="Mesh",wait=1);

//creazione di alcune macro utili per rendere più duttile e chiaro il codice:
```

```
macro Grad(u) [dx(u), dy(u)]       //gradiente di u
macro sumDerivateParziali(u) ( dx(u) + dy(u) )      //divergen-
za di u
macro Laplaciano(u) ( dxx(u) + dyy(u) )   //laplaciano di u


fespace Vh(Th,P1);

Vh u,v,uold,error,ues;
Vh h=hTriangle;          // h gives the size of the current triangle


//Parametri utili per effettuare la simulazione

real dt=0.1;
real t,T=1;
real mu=0.00001;   // mu è il coefficiente di viscosità.

real delta=1.5;          // delta  il parametro di stabilizzazione, che intro-
duce una diffusione
// (in generale sarà delta>0).


cout<<"Passo in spazio: "<< h[].max <<endl;
cout<<"Passo in tempo: "<< dt <<endl;

real H=h[].max;
func u0=0;

// forzante costante ed uguale ad 1
func f=1;

// passo temporale
int step=T/dt;

uold=u0;//interpolazione di u0 in Vh
plot(uold,WindowIndex = -1,fill = 1,value = 1,wait = 1,dim = 3,Sho-
wAxes = 1,cmm = "Initial Condition"); // plot Initial Datum 3d

//Parametro per Theta Metodo
real theta=0.5;       // ponendo theta=0 riotteniamo il metodo di Galer-
```

```
kin standard, che non  stabile

real eL2,eH1,eL2H1=0;


problem heat(u,v)=int2d(Th)(u*v)   //derivata in tempo
-int2d(Th)(uold*v)
+int2d(Th)(dt*mu*(1-theta)*Grad(uold)'*Grad(v))
+int2d(Th)(dt*mu*theta*Grad(u)'*Grad(v))
+int2d(Th)(dt*(1-theta)*sumDerivateParziali(uold)*v)
+int2d(Th)(dt*theta*sumDerivateParziali(u)*v)
+int2d(Th)((dt*mu^2*delta*H)*(1-theta)/sqrt(2)*Laplaciano(uold)*Laplaciano(v))
+int2d(Th)((dt*mu^2*delta*H)*theta/sqrt(2)*Laplaciano(u)*Laplaciano(v))
-int2d(Th)((dt*mu*delta*H)*(1-theta)/sqrt(2)*Laplaciano(uold)*sumDerivateParziali(v))
-int2d(Th)((dt*mu*delta*H)*theta/sqrt(2)*Laplaciano(u)*sumDerivateParziali(v))
-int2d(Th)((dt*mu*delta*H)*(1-theta)/sqrt(2)*sumDerivateParziali(uold)*Laplaciano(v))
-int2d(Th)((dt*mu*delta*H)*theta/sqrt(2)*sumDerivateParziali(u)*Laplaciano(v))
+int2d(Th)((dt*delta*H)*(1-theta)/sqrt(2)*sumDerivateParziali(uold)*sumDerivateParziali
+int2d(Th)((dt*delta*H)*theta/sqrt(2)*sumDerivateParziali(u)*sumDerivateParziali(v))
-int2d(Th)(dt*v)
+int2d(Th)((dt*delta*H)/sqrt(2)*mu*Laplaciano(v))
-int2d(Th)((dt*delta*H)/sqrt(2)*sumDerivateParziali(v))
+on(1,2,3,4,u=0);


//Iterazione in tempo [u=u(k+1), uold=u(k)]
for(int i=0;i<step;i++){
cout<<"iteration "<<i+1<<" time="<<(i+1)*dt<<endl;
t=(i+1)*dt;
heat;
uold=u;
//if ( !(i % 1))
//plot(u,fill=1,value=0,wait=0,dim=2,ShowAxes =0,cmm="t="+t+""); // plot so-
lution 3d
}


plot(u,fill = 1,value = 1,wait = 1,dim = 3,ShowAxes = 1,cmm = "Ap-
proximate Solution for T=1");
```

Appendice B

Codice test 2

```
////////////////////////////////////////////////////////////////////////////////
//                                                                            //
//  Eq. di diffusione-trasporto evolutiva; Omega = quadrato 2 x 2  con buco   //
//                                                                            //
//      u_t - mu*Lap u + b * grad u = 0  in Q=Omega x [0,T), con Omega=(-1,1)×(-1,1)      //
//       u(x,y,0) = 0       in Omega                //
//       u(x,y,t) = phi(x,y,t)            in dOmega  x [0,T)  Dirichlet miste //
//  du/dn = 0                        in buco x [0,T)     Neumann omogenea (PARETE BUCO ISOLANTE) //
//       con  b(x,y) = (y(1-x^2),-x(1-y^2))^T, //
//  con phi(x,y,t) =  1   per x=-1, -1<=y<=1 //
// =  0   altrove //
//                                                                            //
////////////////////////////////////////////////////////////////////////////////

//Generazione della griglia Omega quadrata 2 x 2 non strutturata (-1,1)x(-1,1)

border floor(t=-1,1){ x=t; y=-1; label=1; };
border right(t=-1,1){ x=1; y=t; label=1; };
border ceiling(t=-1,1) { x=-t; y=1; label=1; };
border left(t=-1,1) { x=-1; y=-t; label=2; };
border buco(t=0,2*pi) { x=0.15*cos(t)-0.35; y=0.15*sin(t)+0.5; label=3; };

int n=100;
```

```
mesh Th=buildmesh(floor(n) + right(n) + ceiling(n) + left(n) + buco(-n));

plot(Th, wait=true);
//savemesh(Th, "Th.msh");

//get data of the mesh
int nbtriangles = Th.nt;
cout << "number of triangles of the Delaunay triangulation:" << nbtriangles << endl;

//spazio degli elementi finiti
fespace Vh(Th,P2);
Vh u,v,uold,error,ues;

//creazione di alcune macro utili per rendere più duttile e chiaro il codice:
macro Grad(u) [dx(u), dy(u)]      //gradiente di u
macro sumDerivateParziali(u) ( dx(u) + dy(u) )      //divergenza di u
macro Laplaciano(u) ( dxx(u) + dyy(u) )   //laplaciano di u


//Parametri utili per effettuare la simulazione
real dt=0.1;
real t,T=5;
real mu=0.1;  // mu è il coefficiente di viscosità


real delta=5;          // parametro di stabilizzazione, che introduce una diffusione (in generale sarà delta¿0).
//cout<<"Passo in spazio: "<< h[].max <<endl;
cout<<"Passo in tempo: "<< dt << endl;

// definizione del dato iniziale:
func u0=0;

// forzante nulla;
func f=0;
```

```
// definizione della funzione phi(x,y,t):
func phi= ( (x=-1 && y>=-1 && y<=1 ) ? 1.0 : 0.0 );

// componenti del campo vettoriale  b(x,y) = (y(1-x^2),-x(1-y^2))^T e norma di b(x,y):
func b1=y*(1-x^2);
func b2=-x*(1-y^2);
func normab=sqrt(b1^2 + b2^2);

// passo temporale:
int step=T/dt;

uold=u0;//interpolazione di u0 in Vh

//Parametro per Theta Metodo
real theta=0.5;        // ponendo theta=0 riotteniamo il metodo di Galerkin standard, che non è stabile
real eL2,eH1,eL2H1=0;


// in ciò che segue hTriangle indica di volta in volta il diametro di ciascuno dei triangolini della
// triangolazione della mesh
problem heat(u,v)=int2d(Th)(u*v)   //derivata in tempo
-int2d(Th)(uold*v)
+int2d(Th)(dt*mu*(1-theta)*Grad(uold)'*Grad(v))
+int2d(Th)(dt*mu*theta*Grad(u)'*Grad(v))
+int2d(Th)(dt*b1*(1-theta)*dx(uold)*v)
+int2d(Th)(dt*b1*theta*dx(u)*v)
+int2d(Th)(dt*b2*(1-theta)*dy(uold)*v)
+int2d(Th)(dt*b2*theta*dy(u)*v)
+int2d(Th)(dt*mu^2*delta*hTriangle*(1-theta)/normab*Laplaciano(uold)*Laplaciano(v))
+int2d(Th)(dt*mu^2*delta*hTriangle*theta/normab*Laplaciano(u)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*b1*(1-theta)/normab*Laplaciano(uold)*dx(v))
-int2d(Th)(dt*mu*delta*hTriangle*b2*(1-theta)/normab*Laplaciano(uold)*dy(v))
+int2d(Th)(dt*mu*delta*hTriangle*b1*theta/normab*Laplaciano(u)*dx(v))
-int2d(Th)(dt*mu*delta*hTriangle*b2*theta/normab*Laplaciano(u)*dy(v))
-int2d(Th)(dt*mu*delta*hTriangle*b1*(1-theta)/normab*dx(uold)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*b1*theta/normab*dx(u)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*b2*(1-theta)/normab*dy(uold)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*b2*theta/normab*dy(u)*Laplaciano(v))
```

```
+int2d(Th)(dt*delta*hTriangle*b1^2*(1-theta)/normab*dx(uold)*dx(v))
+int2d(Th)(dt*delta*hTriangle*b1*b2*(1-theta)/normab*dx(uold)*dy(v))
+int2d(Th)(dt*delta*b1^2*hTriangle*theta/normab*dx(u)*dx(v))
+int2d(Th)(dt*delta*b1*b2*hTriangle*theta/normab*dx(u)*dy(v))
+int2d(Th)(dt*delta*b1*b2*hTriangle*(1-theta)/normab*dy(uold)*dx(v))
+int2d(Th)(dt*delta*b2^2*hTriangle*(1-theta)/normab*dy(uold)*dy(v))
+int2d(Th)(dt*delta*b1*b2*hTriangle*theta/normab*dy(u)*dx(v))
+int2d(Th)(dt*delta*b2^2*hTriangle*theta/normab*dy(u)*dy(v))
+on(1, u=0)
+on(2, u=1);



//Iterazione in tempo [u=u(k+1), uold=u(k)]
for(int i=0;i<step;i++){
cout<<"iteration "<<i+1<<" time="<<(i+1)*dt<<endl;
t=(i+1)*dt;
heat;
uold=u;
//if ( !(i % 1))
//plot(u,fill=1,value=0,wait=0,dim=2,ShowAxes =0,cmm="t="+t+""); // plot solution 3d
}


plot(u,fill = 1,value = 1,wait = 1,dim = 2,ShowAxes = 1,cmm = "Approximate Solution for T=5-P2");
```

Appendice C

Codice dell'applicazione

```
////////////////////////////////////////////////////////////////////////////////
//                                                                            //
// Eq. di diffusione-trasporto evolutiva; Omega = quadrato 1 x 1              //
//                                                                            //
//      u_t - (1/Pe)*Lap u + b * grad u = 0  in Q=Omega x [0,T), con Ome-
ga=(0,1)x(0,1)     //
//       u(x,y,0) = 0  (concentrazione nulla)    in (0,1/2] x (0,1) //
//        u(x,y,0) = 1  (concentrazione massima)          in (1/2,1) x (0,1)
//
//       u(x,y,t) = periodiche o Neumann omogeneo(no flux)           in dO-
mega  x [0,T)    //
//       con  b(x,y) = (sin(2*pi*y),0)^T se 0 <= mod(t,T) <= T/2,  //
//  = (0,sin(2*pi*x))^T se T/2 <= mod(t,T) <= T, //
//  per 0 <= x,y <= 1, //
// con mod(t,T) = t-q*T, essendo q un intero tale che 0 <= t-q*T <= T.
//
//       //
//                                                                          //
////////////////////////////////////////////////////////////////////////////////

//Generazione della griglia Omega quadrata 1 x 1 non strutturata (0,1)x(0,1)

border floor(t=0,1){ x=t; y=0; label=1; };
border right(t=0,1){ x=1; y=t; label=4; };
border ceiling(t=0,1) { x=1-t; y=1; label=3; };
border left(t=0,1) { x=0; y=1-t; label=2; };
```

```
int n=45;

mesh Th=buildmesh(floor(n) + right(n) + ceiling(n) + left(n), fixebor-
der=1);

plot(Th, wait=true);
//savemesh(Th, "Th.msh");

//get data of the mesh
int nbtriangles = Th.nt;
cout << "number of triangles of the Delaunay triangulation:" << nbtrian-
gles << endl;

//spazio degli elementi finiti con condizioni al bordo periodiche:
// label : 2 and 4 are left and right side with y the curve abscissa
// 1 and 3 are bottom and upper side with x the curve abscissa
fespace Vh(Th,P2, periodic=[[2,y],[4,y],[1,x],[3,x]]);
Vh u,v,uold,error,ues,u0,c1,c2,normab;

//creazione di alcune macro utili per rendere più duttile e chiaro il codi-
ce:
macro Grad(u) [dx(u), dy(u)]       //gradiente di u
macro sumDerivateParziali(u) ( dx(u) + dy(u) )      //divergen-
za di u
macro Laplaciano(u) ( dxx(u) + dyy(u) )   //laplaciano di u


//Parametri utili per effettuare la simulazione
real dt=0.01;
real t=0;
real T=0.8;     // a T assegneremo i seguenti valori: 0.8, 1.6.
real mu=0.01;      // mu is the viscosity coefficient (qui mu = 1/Pe. Ad es-
so assegneremo
// differenti valori in base ai seguenti valori del numero di Peclèt:
// Pe=10^2 => mu=0.01 (for laminar flames), Pe=10^3-10^5 => mu=0.0001 (for mo-
lecular dyes in
// typical microfluidic water/glycerol solutions), Pe=10^5 => mu=0.00001 (for gra-
nular materials in rotating tumblers),
// Pe=10^10   => mu=0.0000000001 (for turbulent reactive flows).
int q=0;      // q è il parametro che interviene nella definizione del cam-
po vettoriale (deve soddisfare
```

```
// la condizione: 0 <= t-q*T < T).

real delta=0.25;                    // parametro di stabilizzazione, che introduce una diffusione (in generale sarà delta>0).
// se mu è grande, allora la diffusione è dominante ed il delta posso sceglierlo piccolo,
// in quanto il metodo FE funziona bene e non necessita di stabilizzazione; se invece mu è piccolo,
// allora il trasporto è dominante e per stabilizzare il metodo devo scegliere un delta molto più grande.
real Tfin=5*T; // E' il tempo finale, multiplo del periodo T.

//cout<<"Passo in spazio: "<< h[].max <<endl;
cout<<"Passo in tempo: "<< dt << endl;

// definizione del dato iniziale:
u0=( (x>0.5 && x<1 && y>=0 && y<=1) ? 1.0 : 0.0 );

// plottiamo il grafico del dato iniziale (osserviamo che il plot viene fatto solo per le funzioni dello spazio degli elementi finiti Vh)
plot(u0, fill=1, value=1, wait=1, viso=viso(0:viso.n-1), dim=2, grey=1, ShowAxes=1, cmm = "Initial condition" );


// forzante nulla;
func f=0;

// componenti del campo vettoriale  b(x,y,t) = (sin(2*pi*y),0)^T se 0 <= mod(t,T) <= T/2,  //
//    = (0,sin(2*pi*x))^T se T/2 <= mod(t,T) <= T, //
// per 0 <= x,y <= 1, //
// con mod(t,T) = t-q*T, essendo q un intero tale che 0 <= t-q*T <= T.
func real b1(int q, real t, real T)  // q è il parametro che interviene nella definizione del campo vettoriale (deve soddisfare
{   // la condizione: 0 <= t-q*T <= T).
if(0 <= t-q*T && t-q*T  < T/2)
return sin(2*pi*y);
if(T/2 <= t-q*T && t-q*T < T)
return 0;
}

func real b2(int q, real t, real T)
```

```
{
if(0 <= t-q*T && t-q*T  < T/2)
return 0;
if(T/2 <= t-q*T && t-q*T  < T)
return sin(2*pi*x);
}


// passo temporale:
int step=Tfin/dt; //quando aggiungiamo Tfin=2*T opppure 5*T etc. qui mettiamo Tfin
//invece di T, che ricompare nel ciclo finale.
uold=u0;//interpolazione di u0 in Vh

//Parametro per Theta Metodo
real theta=0.5;        // ponendo theta=0 riotteniamo il metodo di Galerkin standard, che non è stabile



// in ci che segue hTriangle indica di volta in volta il diametro di ciascuno dei triangolini della
// triangolazione della mesh
problem heat(u,v)=int2d(Th)(u*v)   //derivata in tempo
-int2d(Th)(uold*v)
+int2d(Th)(dt*mu*(1-theta)*Grad(uold)'*Grad(v))
+int2d(Th)(dt*mu*theta*Grad(u)'*Grad(v))
+int2d(Th)(dt*c1*(1-theta)*dx(uold)*v)
+int2d(Th)(dt*c1*theta*dx(u)*v)
+int2d(Th)(dt*c2*(1-theta)*dy(uold)*v)
+int2d(Th)(dt*c2*theta*dy(u)*v)
+int2d(Th)(dt*mu^2*delta*hTriangle*(1-theta)/normab*Laplaciano(uold)*Laplaciano
+int2d(Th)(dt*mu^2*delta*hTriangle*theta/normab*Laplaciano(u)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*c1*(1-theta)/normab*Laplaciano(uold)*dx(v))
-int2d(Th)(dt*mu*delta*hTriangle*c2*(1-theta)/normab*Laplaciano(uold)*dy(v))
+int2d(Th)(dt*mu*delta*hTriangle*c1*theta/normab*Laplaciano(u)*dx(v))
-int2d(Th)(dt*mu*delta*hTriangle*c2*theta/normab*Laplaciano(u)*dy(v))
-int2d(Th)(dt*mu*delta*hTriangle*c1*(1-theta)/normab*dx(uold)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*c1*theta/normab*dx(u)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*c2*(1-theta)/normab*dy(uold)*Laplaciano(v))
-int2d(Th)(dt*mu*delta*hTriangle*c2*theta/normab*dy(u)*Laplaciano(v))
```

```
+int2d(Th)(dt*delta*hTriangle*c1^2*(1-theta)/normab*dx(uold)*dx(v))
+int2d(Th)(dt*delta*hTriangle*c1*c2*(1-theta)/normab*dx(uold)*dy(v))
+int2d(Th)(dt*delta*c1^2*hTriangle*theta/normab*dx(u)*dx(v))
+int2d(Th)(dt*delta*c1*c2*hTriangle*theta/normab*dx(u)*dy(v))
+int2d(Th)(dt*delta*c1*c2*hTriangle*(1-theta)/normab*dy(uold)*dx(v))
+int2d(Th)(dt*delta*c2^2*hTriangle*(1-theta)/normab*dy(uold)*dy(v))
+int2d(Th)(dt*delta*c1*c2*hTriangle*theta/normab*dy(u)*dx(v))
+int2d(Th)(dt*delta*c2^2*hTriangle*theta/normab*dy(u)*dy(v));
//+on(floor, right, ceiling, left, u=phi);  questo pezzo va tolto perchè ho im-
posto condizioni al bordo periodiche;
//  se volessi imporre condizioni di Neumann omogeneo sul bordo, bastereb-
be togliere
//  periodic dentro fespace Vh di sopra.
int step2=T/dt;
//Iterazione in tempo [u=u(k+1), uold=u(k)]
for(int i=0; i<step; i++)
{
cout<<"iteration "<< i+1 <<" time="<< (i+1)*dt <<endl;
t=(i+1)*dt;
q=(t/T); //aggiorno q assegnandole la parte intera di t/T;
cout <<" q ="<< q << endl;
c1=b1(q,t,T);
c2=b2(q,t,T);
//  plot(c1,fill = 1,value = 1,wait = 1,dim = 2,ShowAxes = 1, hsv=colorhsv, cmm = "Speed x");
//  plot(c2,fill = 1,value = 1,wait = 1,dim = 2,ShowAxes = 1, hsv=colorhsv, cmm = "Speed y");
normab=sqrt(c1^2 + c2^2);
if (normab==0)
normab=1;
heat;
uold=u;

if ( !((i+1) % step2))
plot(u,fill=1,value=0,wait=0,dim=2,ShowAxes =0,grey=1,cmm="t="+t+"");// plot so-
lution 3d

}

plot(u,fill = 1,value = 1,wait = 1,dim = 2,ShowAxes = 1, grey=1, cmm = "Ap-
proximate Solution T=0.8-P2");
```

Bibliografia

[1] C. Grossmann, H. Ross, M. Stynes *Numerical treatment of Partial Differential Equations*, Springer, Heidelberg, 2007.

[2] F. Hecht *Freefem++*, Third Edition, Version 3.7-1, 2005.

[3] A. Quarteroni *Numerical Models for Differential Problems*, vol. 2, MS&A (Modeling, Simulation & Applications), Sringer 2008.

[4] A. Quarteroni, F. Saleri, R. Sacco *Numerical Mathematics*, Springer, Brlin Heidelberg, II edition, 2007.

[5] A. Quarteroni, A. Valli *Numerical Approximation of Partial Differential Equations*, Springer, Berlin Keidellung 1994.

[6] H.G. Ross, M. Stynes, L. Tobiska *Numerical Methods for Singularly Perturbed Differential Equations. Connection-Diffusion and Flow Problems*, Springer-Verlag, Berlin Keidellung 1996.

[7] C.P. Schlick, I.C. Christov, P.B. Umbanhowar, J.M. Ottimo, R.M. Lueptow *A mapping method for distributive mixing with diffusion: Interplay between chaos and diffusion in time-periodic sine flow*, Physics of fluids 25, 052102 (2013).

[8] V. Thomee *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin and Heidelberg, 1984.