Alma Mater Studiorum · Università di Bologna

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

Corso di Laurea in Matematica

METODO DEGLI ELEMENTI FINITI PER EQUAZIONI ELLITTICHE

Tesi di Laurea in Analisi Matematica

Relatore:
Chiar.mo Prof.
Bruno Franchi

Presentata da:
Chiara Vespasiano

Seconda Sessione
Anno Accademico 2010-2011

Al Prof. Tommaso Iurisci.

Introduzione

Molti fenomeni fisici o problemi di interesse applicativo possono essere formulati attraverso equazioni differenziali a derivate parziali (PDE) o sistemi di PDE.
Solo in un numero molto ristretto di casi è possibile trovare la soluzione esatta (per via analitica) di tali problemi. Per questo c'è l'esigenza di trovare delle soluzioni approssimate.
Tuttavia, si ha la necessità di stimare l'errore di approssimazione numerica per fornire validità ai risultati delle simulazioni su calcolatore.
In questa tesi tratteremo della soluzione numerica di equazioni a derivate parziali con condizione al contorno attraverso il metodo di Galerkin. Illustreremo poi, come caso particolare, il metodo degli elementi finiti. In particolare, affronteremo il problema della soluzione numerica del problema di Dirichlet per mezzo del metodo degli elementi finiti.
Nel primo e secondo capitolo vengono, infatti, mostrati i punti fondamentali su cui si basa questa tecnica che permette di passare dalla risoluzione di un problema definito in uno spazio continuo alla risoluzione di tale problema in uno spazio discreto al fine di determinare una soluzione numerica approssimata.
Nel terzo capitolo viene applicato al problema di Dirichlet, caratterizzato dalla condizione al bordo, e cercheremo di chiarire tutte le questioni fondamentali relative all'esistenza e all'unicitá della soluzione in opportuni spazi funzionali.
È importante, dunque, ricercare la soluzione piú efficiente per il problema in

esame, dove per efficienza si intende il miglior compromesso tra facilitá di formulazione e di risolubilitá teorica, e sufficiente generalitá e addattabilitá ai metodi numerici.
In questi termini la soluzione piú conveniente è la soluzione debole o variazionale, cioè ricercata tra le funzioni nello spazio di Sobolev H^1. Tale nozione di soluzione è una formulazione molto flessibile con un elevato grado di generalitá e con una teoria basata, sostanzialmente su un unico teorema di Analisi Funzionale, il teorema di Lax-Milgram. Inoltre la formulazione debole è quella naturale per implementare il metodo di Galerkin-Elementi Finiti.
Il problema ellittico puó dunque essere riscritto come un problema astratto. La risoluzione del problema di Dirichlet è stata quindi articolata nei seguenti passi:

1. derivazione della formulazione debole del problema;

2. verifica dell'esistenza e dell'unicità della soluzione per mezzo del teorema di Lax-Milgram;

3. approssimazione numerica tramite metodo di Galerkin-Elementi finiti.

Indice

Capitolo 1

Metodo di Galerkin

Dato un problema definito su uno spazio di Hilbert V, data una forma bilineare $A(\cdot,\cdot) : V \times V \to \mathbb{R}$ ed una forma lineare $F(\cdot) : V \to \mathbb{R}$ si vuole risolvere l'equazione

$$A(u,v) = F(v) \qquad \forall v \in V \ . \tag{1.1}$$

A questo scopo enunciamo il Teorema di Lax-Milgram che prova l'esistenza e l'unicità della soluzione $u \in V$.

1.1 Teorema di Lax-Milgram

Il seguente risultato è di fondamentale importanza, in quanto molti modelli alle equazioni a derivate parziali si possono ricondurre alla formulazione generale (1.1).

Teorema 1.1.1 (Teorema di Lax-Milgram).
Sia V uno spazio di Hilbert, dotato di norma $||\cdot||$, sia $A(u,v) : V \times V \to \mathbb{R}$ una forma bilineare e $F(v) : V \to \mathbb{R}$ un funzionale lineare continuo, ossia $F \in V'$ dove V' denota lo spazio duale di V.
Assumiamo $A(\cdot,\cdot)$ continua,ovvero:

$$\exists \gamma > 0 : |A(w,v)| \leq \gamma ||w||||v|| \qquad \forall w, v \in V \tag{1.2}$$

e corciva,ovvero:

$$\exists \alpha > 0 : A(u,v) \geq \alpha ||v||^2 \qquad \forall v \in V \ . \tag{1.3}$$

Allora, esiste un'unica soluzione $u \in V$ al problema (1.1) e inoltre

$$||u|| \leq \frac{1}{\alpha} ||F||_{V'} \ . \tag{1.4}$$

Dimostrazione. Dal teorema di rappresentazione di Riesz possiamo scrivere

$$F(v) = (RF, v)_V \qquad \forall v \in V$$

e per ogni fissato $w \in V$

$$A(w,v) = (Aw, v)_V \qquad \forall v \in V$$

dove $(\cdot,\cdot)_V$ è il prodotto scalare in V e le biiezioni $R : V' \to V$ e $A : V \to V'$ sono operatori lineari continui. Più precisamente, R è un operatore isometrico poichè

$$||RF|| = \sup_{\substack{v \in V \\ v \neq 0}} \frac{(RF, v)_V}{||v||} = \sup_{\substack{v \in V \\ v \neq 0}} \frac{F(v)}{||v||} = ||F||_{V'} \ ;$$

inoltre

$$||Aw|| = \sup_{\substack{v \in V \\ v \neq 0}} \frac{(Aw, v)_V}{||v||} = \sup_{\substack{v \in V \\ v \neq 0}} \frac{A(w,v)}{||v||} \leq \gamma ||w|| \ ,$$

da (1.2).

Il problema di trovare $u \in V$ tale che $A(u,v) = F(v)$ per ogni $v \in V$ è ora equivalente al seguente problema: per ogni $F \in V'$ trovare un'unica $u \in V$ tale che

$$Au = RF \ ,$$

ovvero, bisogna dimostare che A è una biezione.

Innanzitutto mostriamo che A è iniettiva. Noi abbiamo

$$||v||^2 \leq \frac{1}{\alpha}(Av, v)_V \leq \frac{1}{\alpha}||Av|| ||v|| \ , \tag{1.5}$$

e conseguentemente $||v|| \leq \frac{1}{\alpha}||Av||$: l'unicità è così provata. Noi possiamo ora mostrare che il campo $R(A)$ di A è chiuso e $R(A)^{\perp} = 0$, che è equivalente a dimostrare che $R(A) = V$. Supponiamo che $Av_n \to w$ in V. Da (1.5)

$$||v_n - v_m|| \leq \frac{1}{\alpha}||Av_n - Av_m|| \ ,$$

quindi v_n è una successione di Cauchy e sia $v = \lim v_n$. Poichè A è continua, $Av = w$, e $R(A)$ è chiuso. Ora prendiamo $z \in R(A)^{\perp}$; allora

$$0 = (Az, z)_V = A(z, z) \geq \alpha||z||^2 \ ,$$

ossia, $z = 0$.
Infine, dobbiamo dimostrare (1.4) . Scegliendo $v = u$ nel problema (1.1) , otteniamo

$$||u||^2 \leq \frac{1}{\alpha}A(u, u) = \frac{1}{\alpha}F(u) \leq \frac{1}{\alpha}||F||_{V'}||u|| \ ,$$

che ci da

$$||u|| \leq \frac{1}{\alpha}||F||_{V'} \ .$$

□

Nel caso in cui la forma bilineare è simmetrica:

$$A(w, v) = A(v, w) \qquad \forall w, v \in V,$$

allora $A(\cdot, \cdot)$ definisce un prodotto scalare su V, e il teorema di rappresentazione di Riesz è sufficiente per la deduzione dell'esistenza e unicità della soluzione del problema (1.1). Inoltre, in questo caso la soluzione al problema (1.1) puó essere considerata la soluzione unica al problema di ottimizzazione:

$$\text{trova } u \in V \text{ tale che } J(u) \leq J(v) \qquad \forall v \in V,$$

dove:

$$J(v) := \frac{1}{2}A(v, v) - F(v)$$

è un funzionale quadratico.

1.2 Metodo di approssimazione di Galerkin

Consideriamo la forma debole di un generico problema ellittico posto in un dominio $\Omega \in \mathbb{R}^d$; $d \leq 3$:

$$\text{trovare } u \in V \ : \ A(u,v) = F(v) \qquad \forall v \in V \tag{1.6}$$

essendo V un opportuno spazio di Hilbert, $A(\cdot,\cdot)$ una forma bilineare continua e coerciva su $V \times V$ in $\mathbb{R}$, F un funzionale lineare e continuo da V in $\mathbb{R}$. Sotto tali ipotesi il lemma di Lax-Milgram assicura esistenza e unicità della soluzione.

Il metodo di Galerkin per l'approssimazione di questo problema consiste nel cercare una soluzione approssimata $u_h \in V_h$, essendo V_h una famiglia di spazi dipendente dal parametro positivo h tali che

$$V_h \subset V \quad , \quad dim V_h = N_h < \infty \qquad \forall h > 0 \ .$$

Il problema approssimato assume allora la forma:

$$\text{trovare } u_h \in V_h \ : \ A(u_h, v_h) = F(v_h) \qquad \forall v_h \in V_h \tag{1.7}$$

e viene detto problema di Galerkin.

Indicando con $\{\phi_j\}$, per $j = 1, 2, \ldots, N_h$, una base di V_h, è sufficiente che il problema approssimato sia verificato per ogni funzione della base, in quanto tutte le funzioni di V_h sono combinazione lineare delle ϕ_j. Si ottiene quindi :

$$A(u_h, \phi_i) = F(\phi_i) \quad , \quad i = 1, 2, \ldots, N_h \ .$$

Naturalmente, essendo $u_h \in V_h$, è possibile esprimere u_h come combinazione lineare delle funzioni di base ovvero

$$u_h = \sum_{j=1}^{N_k} u_j \phi_j(x)$$

dove gli u_j, per $j = 1, \ldots, N_h$, sono coefficienti incogniti. Le equazioni di cui sopra assumono allora la forma

$$\sum_{j=1}^{N_h} u_j A(\phi_j, \phi_i) = F(\phi_i) \quad , \quad i = 1, \ldots, N_h \ .$$

Indichiamo con $\mathcal{A}$ la matrice di rigidità (o di stiffness) con elementi a_{ij} dati da

$$a_{ij} = \mathcal{A}(\phi_j, \phi_i)$$

e con f il vettore di componenti

$$f_i = F(\phi_i)\ .$$

Indicando con u il vettore avente come componenti i coefficienti incogniti u_i, ci si riconduce al sistema lineare

$$\mathcal{A}u = f\ .$$

Tratteremo le proprietà della matrice di rigidità nell'ultimo paragrafo di questo capitolo.

1.3 Analisi del metodo di Galerkin

1.3.1 Esistenza e unicità

Il lemma di Lax-Milgram vale per ogni spazio di Hilbert e quindi vale in particolare per lo spazio V_h (essendo esso un sottospazio chiuso dello spazio di Hilbert V).

Siccome la forma bilineare e il funzionale sono gli stessi del problema variazionale astratto, allora sono soddisfatte le ipotesi del lemma, pertanto la soluzione del problema di Galerkin esiste ed è unica.

1.3.2 Stabilità

Definizione 1.1. Il problema di Galerkin è stabile uniformemente rispetto ad h valendo la seguente maggiorazione della soluzione

$$||u_h||_V \leq \frac{1}{\alpha}||F||_{V'}$$

dove $||F||_{V'}$ indica la norma nello spazio duale di V.

La stabilità del metodo garantisce che la norma della soluzione rimane limitata al tendere di h a zero, uniformemente rispetto ad h.

1.3.3 Convergenza

L'obiettivo del metodo di approssimazione di Galerkin consiste nella convergenza della soluzione u_h alla soluzione u di (1.6) quando h tende a zero. Ció significa che, a patto di prendere h sufficientemente piccolo, si puó approssimare bene quanto si vuole la soluzione esatta.
Per dimostrare la convergenza del metodo, iniziamo con il verificare la proprietà di forte consistenza:

Lemma 1.3.1. *Il metodo di Galerkin è fortemente consistente ovvero*

$$A(u-u_h, v_h) = 0 \qquad \forall v_h \in V_h \ . \tag{1.8}$$

Dimostrazione. Essendo $V_h \subset V$, la soluzione esatta u soddisfa il problema debole per ogni elemento di V_h e quindi

$$A(u, v_h) = F(v_h) \qquad \forall v_h \in V_h \ . \tag{1.9}$$

Sottraendo membro a membro la (1.8) dalla (1.9), si ottiene

$$A(u, v_h) - A(u_h, v_h) = 0 \qquad \forall v_h \in V_h \ . \tag{1.10}$$

dalla quale, grazie alla bilinearità della forma $A(\cdot,\cdot)$, segue la tesi. □

Il lemma esprime il fatto che il metodo di Galerkin è un metodo di proiezione ortogonale. Infatti, se $A(\cdot,\cdot)$ fosse il prodotto scalare euclideo, u e u_h dei vettori e V_h un sottospazio della spazio euclideo V, la proposizione esprimerebbe l'ortogonalità dell'errore $u - u_h$ rispetto al sottospazio V_h, ovvero u_h sarebbe la migliore approssimazione di u in V_h.
Tale interpretazione vale solo se la forma bilineare è simmetrica. Se ció non si verifica, vale comunque la seguente disuguaglianza (nota come Lemma di Ceà)

$$||u-u_h||_V \leq \frac{M}{\alpha} \inf_{v_h \in V_h} ||u - v_h||_V \ . \tag{1.11}$$

Dimostrazione. Consideriamo il valore che la forma bilineare assume quando entrambi i suoi argomenti sono pari a $u - u_h$. Se v_h è un arbitrario elemento di V_h si ottiene

$$A(u-u_h, u-u_h) = A(u-u_h, u-v_h) + A(u-u_h, v_h-u_h) = A(u-u_h, u-v_h)$$

essendo nullo il termine $A(u-u_h, v_h-u_h)$ grazie alla (1.8) in quanto $v_h - u_h \in V_h$. Infine:

$$|A(u-u_h, u-v_h)| \le M||u-u_h||_V||u-v_h||_V \ ,$$

avendo sfruttato la continuità della forma bilineare. Inoltre, per la coercività di $A(\cdot,\cdot)$ deve essere

$$A(u-u_h, u-v_h) \ge \alpha||u-u_h||_V^2$$

per cui si ha:

$$||u-u_h||_V \le \frac{M}{\alpha}||u-u_h||_V \qquad \forall v_h \in V_h \ .$$

Tale disuguaglianza vale per tutte le funzioni v_h e dunque varrà anche prendendone l'estremo inferiore. Si trova perció

$$||u-u_h||_V \le \frac{M}{\alpha} \inf_{v_h \in V_h} ||u-v_h||_V \tag{1.12}$$

□

È allora evidente che, affinchè il metodo converga, basta richiedere che al tendere di h a zero, lo spazio V_h tenda ad occupare tutto lo spazio V.
In tal caso il metodo di Galerkin è convergente e si puó scrivere

$$\lim_{h \to 0} ||u-u_h||_V = 0 \ .$$

In conclusione, se $\frac{M}{\alpha}$ è dell'ordine dell'unità, l'errore del metodo di Galerkin è identificabile con l'errore di migliore approssimazione possibile.
In ogni caso, i due errori sono legati l'uno all'altro, avendo lo stesso ordine di infinitesimo rispetto ad h.

1.4 Matrice di rigidità

Tornando all'approssimazione di Galerkin al problema (1.1), sia dal punto di vista algebrico, $\{\varphi_j | j = 1, ..., N_h\}$ una base dello spazio vettoriale V_h ,così che noi siamo in grado di porre

$$u_h(x) = \sum_{j=1}^{N_h} u_j \varphi_j(x) \ . \tag{1.13}$$

Allora da (1.13) deduciamo il seguente sistema lineare di dimensione N_h:

$$\mathcal{A}u = F \ , \tag{1.14}$$

con $u = (u_j)$, $F := (F(\varphi_j))$, $a_{ij} := \mathcal{A}(\varphi_j, \varphi_i)$ per $i, j = 1, ..., N_h$. La matrice $\mathcal{A}$ è chiamata *matrice di rigidità*.

1.4.1 Proprietà della matrice di rigidità

Evidenziamo alcune proprietà della matrice di rigidità che sono indipendenti dalla base scelta per V_h, ma che dipendono esclusivamente dalle caratteristiche del problema debole che si sta approssimando. Ad esempio, basi formate da funzioni con supporto piccolo saranno preferibili in quanto tutti gli elementi a_{ij} relativi a funzioni di base che hanno supporti con intersezione nulla, risulteranno nulli.

Proposizione 1.4.1. *La matrice $\mathcal{A}$ definita in (1.14) è definita positiva, ovvero, per ogni $u \in \mathbb{R}^{N_h}$, $u \neq 0$, $(Au, u) \geq 0$ dove $(\cdot, \cdot)$ denota il prodotto scalare.*

Infatti, sia $u_h \in V_h$ è la funzione definita come

$$u_h(x) = \sum_{j=1}^{N_h} u_j \varphi_j(x)$$

Allora

$$(Au, u) = \sum_{i,j=1}^{N_h} u_j A(\varphi_j, \varphi_i) u_j = A(u_h, u_h) \ ,$$

e si ha la conclusione utlizzando la coercività di A. In particolare, ogni autovalore di $\mathcal{A}$ ha parte reale positiva, quindi l'esistenza e l'unicità di una soluzione per (1.14) puó anche essere dimostrata per mezzo di un argomento puramente algebrico senza dover ricorrere al teorema di Lax-Milgram.

Proposizione 1.4.2. *La matrice $\mathcal{A}$ è simmetria se e solo se la forma bilineare A è simmetrica.*

Capitolo 2

Metodo degli elementi finiti

Il metodo di approssimazione degli elementi finiti (FEM), uno dei metodi di Galerkin, è una tecnica numerica atta a cercare soluzioni approssimate di problemi descritti da equazioni differenziali alle derivate parziali riducendo queste ultime ad un sistema di equazioni algebriche.
In questo paragrafo mostriamo le proprietà fondamentali di questo metodo di approssimazione: l'esistenza di una tringolazione di Ω, la costruzione di un sottospazio di dimensione finita costituito da polinomiali a tratti e l'esistenza di una base di funzioni avente supporto piccolo.
Questi aspetti forniscono le basi per studiare la stima dell'errore per l'approssimazione delle equazioni alle derivate parziali nel caso del problema di Dirichlet, che sarà approfondito nel prossimo paragrafo.

2.1 Triangolazione

Sia $\Omega \subset \mathcal{R}^d$, $d = 2{,}3$, un dominio poligonale, ossia, un aperto connesso e limitato tale che $\bar{\Omega}$ è l'unione di un numero finito di poliedri:

$$\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} K \tag{2.1}$$

dove

(i) Ogni K è un poliedro con $\mathring{K} \neq \emptyset$;

(ii) $\mathring{K}_1 \cap \mathring{K}_2 \neq \emptyset$ per ogni K_1 e $K_2 \in \mathcal{T}_h$ distinti;

(iii) se $F = K_1 \cap K_2 \neq \emptyset$ (K_1 e K_2 elementi distinti di $\mathcal{T}_h$) allora F è una faccia, lato o vertice comune di K_1 e K_2;

(iv) $diam(K) \leq h \ \forall K \in \mathcal{T}_h$;

$\mathcal{T}_h$ è chiamata *triangolazione* di Ω .
Per semplicità, nel seguito, assumiamo che ogni elemento K di $\mathcal{T}_h$ può essere ottenuto come $K = T_K(\hat{K})$, dove $\hat{K}$ è un riferimento poliedrico e T_K è una mappa affine, ossia, $T_K = B_K\hat{x} + b_K$, dove B_K è una matrice non singolare. Consideriamo due casi differenti:

Caso a. Il riferimento poliedrico $\hat{K}$ è l'unità *d-simplesso*, ossia il triangolo di vertici $(0,0), (1,0)$ e $(0,1)$ quando $d = 2$ o il tetraedo di vertici $(0,0,0), (1,0,0)$ $(0,1,0)$, e $(0,0,1)$ quando $d = 3$. Di conseguenza, ogni $K = T_K(\hat{K})$ è un triangolo o un tetraedo rispettivamente.

Caso b. Il riferimento poliedrico $\hat{K}$ è l'unità *d-cubo* $[0,1]^d$. Di conseguenza, ogni $K = T_K(\hat{K})$ è un parallelogramma, quando $d = 2$, o un parallelepipedo, quando $d = 3$.

Nel secondo caso la triangolazione è fatta da *d-rettangoli* se per ogni $K \in \mathcal{T}_h$ è la matrice B_K, che definisce la trasformazione lineare T_K è diagonale.

2.2 Sottospazi di funzioni polinomiali a tratti

Un secondo aspetto importante del metodo degli elementi finiti consiste nel determinare uno spazio di dimensione finita X_h, che poi dovrebbe servire come buona approssimazione dello spazio di dimensione infinita X.
Sia $\{X_h, h > 0\}$ una famiglia di funzioni polinomiali a tratti, ovvero, per ogni $K \in \mathcal{T}_h$ lo spazio

$$P_K := \left\{ v_{h|K} \mid v_h \in X_h \right\}$$

è costituito da polinomi algebrici.
Per essere più precisi, denotiamo con $\mathbb{P}_k$, $k \geq 0$, lo spazio dei polinomi di grado minore o uguale di k nelle variabili $x_1, \ldots, x_d$, e con $\mathbb{Q}_k$ lo spazio dei polinomi di grado minore o uguale a k rispetto ad ogni $x_1, \ldots, x_d$.

Definiamo gli spazi X_h più comunemente usati.
Nel caso a. impostiamo

$$X_h = X_h^k := \left\{ v_h \in C^0(\bar{\Omega}) \mid v_{h|K} \in \mathbb{P}_k \ \forall K \in \mathcal{T}_h \right\}, \ k \geq 1, \tag{2.2}$$

che sarà chiamato lo spazio degli *elementi finiti triangolari*.
Nel caso b. , invece, definiamo

$$X_h = X_h^k := \left\{ v_h \in C^0(\bar{\Omega}) \mid v_{h|K} \circ T_K \in \mathbb{Q}_k \ \forall K \in \mathcal{T}_h \right\}, \ k \leq 1, \tag{2.3}$$

che è chiamato lo spazio degli *elementi finiti parallelepipedi*.
È importante notare che, in entrambe i casi (2.2) e (2.3) vale:

$$X_h^k \subset H^1(\Omega) \qquad \forall k \geq 1.$$

Questo è una conseguenza del seguente risultato (per semplicità scriveremo $H^s(K)$ invece di $H^s(\dot{K})$) :

Proposizione 2.2.1. *Una funzione $v : \Omega \longrightarrow \mathbb{R}$ appartiene allo spazio $H^1(\Omega)$ se e solo se*

a. $v_K \in H^1(K)$ per ogni $K \in \mathcal{T}_h$;

b. per ogni faccia comune $F = K_1 \cap K_2$, $K_1, K_2 \in \mathcal{T}_h$, la traccia su F di $v_{|K_1}$ e $v_{|K_2}$ è la stessa.

Dimostrazione. Usando a., definiamo la funzione $w_j \in L^2(\Omega)$, $j = 1, \ldots, d$, attraverso

$$w_{j|K} := D_j(v_{|K}) \qquad \forall K \in \mathcal{T}_h \ .$$

Allo scopo di mostrare che $v \in H^1(\Omega)$, basta semplicemente dimostare che $w_j = D_j v$.

Usando la formula di Green, possiamo scrivere:

$$\int_\Omega w_j\phi = \sum_K \int_K w_j\phi = -\sum_K \int_K (v_{|K}) D_j\phi + \sum_K \int_{\partial K} v_{|K}\, \phi\, n_{K,j} \qquad \forall\phi \in \mathcal{D}(\Omega)\ ,$$

dove n_K è il vettore unitario su ∂K. Dato che ϕ si annulla su $\partial\Omega$ e $n_{k_1} = -n_{K_2} =: n$ su una faccia comune $F = K_1 \cap K_2$, si ottiene, utilizzando il punto b.

$$\int_\Omega w_j\phi = -\int_\Omega vD_j\phi + \sum_F \int_F (v_{|K_1} - v_{|K_2})\phi\, n_j = -\int_\Omega vD_j\phi\ ,$$

quindi, $w_j = D_j v$.
D'altra parte se noi assumiamo che $v \in H^1(\Omega)$ segue subito che il punto a. è vero. Inoltre, valendo l'uguaglianza appena verificata $w_j = D_j v$, allora procedendo come prima si ottiene:

$$\sum_F \int_F (v_{|K_1} - v_{|K_2})\, \phi\, n_j = 0 \qquad \forall\phi \in \mathcal{D}(\Omega)\ ,\ j = 1, \ldots, d,$$

quindi il punto b. è così soddisfatto. □

2.3 Gradi di libertà e funzioni di forma

È ora necessario costruire una base per lo spazio X_h tale che sia facilmente descrivibile.
Un punto importante riguarda la scelta di un insieme di *gradi di libertà* su ogni elemento K .

2.3.1 Elementi finiti triangolari

Iniziamo con il caso a., $d = 2$. Per identificare $v_{h|K}$, quando $K = 1$, dobbiamo scegliere tre gradi di libertà su ogni elemento K, con il vincolo aggiuntivo che $v_h \in C^0(\bar{\Omega})$. La scelta più semplice è prendere i valori ai vertici di ogni elemento K.
Notiamo che, se consideriamo:

$$Y_h^1 := \left\{ v_h \in L^2(\Omega) | v_{h|K} \in \mathbb{P}_1\ \forall K \in \mathcal{T}_h \right\}$$

invece di X_h^1 come definita in (2.2), siamo in questo modo liberi di scegliere i gradi di libertà su K come i valori in tre punti arbitrari (non necessariamente coincidenti con i vertici). Per esempio, si possono prendere come nodi i tre punti interni oppure i punti posti al centro di ogni lato. L'ultima scelta implica chiaramente continuità solo nei punti centrali, con la prima scelta invece le funzioni appartenenti a Y_h^1 hanno generalmente due definizioni differenti sui lati comuni di triangoli adiacenti.
Quando $k = 2$, assumiamo che i gradi di libertà siano dati dai valori ai vertici e ai punti centrali di ogni lato. Per dare un'idea della dimostrazione nel caso più generale, mostriamo questa proposizione. Denotiamo i vertici del triangolo K con $a^i, i = 1, 2, 3$ e i punti centrali di ogni lato con $a^{ij}, i < j$, $i, j = 1, 2, 3$.

Proposizione 2.3.1. *Una funzione $p \in \mathbb{P}_2$ è univocamente determinata da i sei valori $p(a^i), 1 \leq i < j \leq 3$.*

Dimostrazione. Poichè il numero di gradi di libertà è uguale alla dimensione di $\mathbb{P}_2(= 6)$, bisogna solo provare che se $p(a^i) = p(a^{ij}) = 0$ allora $p \equiv 0$.
Detto ció, notiamo che la restrizione di p su ogni lato, essendo p una funzione quadratica di una variabile che si annulla in tre punti distinti, p si annulla su ogni lato. Allora noi possiamo scrivere

$$p(x) = cp_1(x)p_2(x)p_3(x) ,$$

dove ogni $p_i(x)$ sono funzioni lineari e ognuna si annulla su un lato di K. Quindi $p \in \mathbb{P}_3$ e da questo segue che $c = 0$. $\square$

Inoltre, ricordiamo che tale scelta dei gradi di libertà garantisce che $v_h \in C^0(\bar{\Omega})$, poichè i gradi di libertà, su ogni lato, identificano univocamente la restrizione di v_h sullo stesso lato.
In maniera simile, si puó provare che i gradi di libertà per un triangolo - cubico $k = 3$ sono dati dai dieci valori ai nodi seguenti: i tre vertici; altri due nodi per ogni lato in modo tale che lo dividono in tre sottointervalli di ugual lunghezza ed infine il centro di gravità.

Bisogna notare che la scelta dei gradi libertà riduce il problema dell'individuazione del polinomio $p \in \mathbb{P}_k$ dal caso tridimensionale a quello bidimensionale. Infatti, su ogni faccia i gradi di libertà sono gli stessi del corrispondente triangolo bidimensionale; quindi se un polinomio si annulla nei punti appena descritti allora si annulla su tutta la faccia. Conseguentemente, si resta con un polinomio p di grado $1, 2$ o 3, che si annulla su quattro piani distinti, il che implica $p \equiv 0$.
Una base per X_h^k è ora facilmente costuita. In particolare, denotando con $a_j, j = 1, \dots, N_h$ l'insieme globale dei nodi in $\bar{\Omega}$, è sufficiente scegliere delle funzioni tali che

$$\phi_i(a_j) = \delta_{ij}\ ,\ i, j = 1, \dots, N_h \quad (\delta_{ij} \text{ è il simbolo di Kronecker}). \tag{2.4}$$

le funzioni che formano la base sono chiamate *funzioni di forma* (shape functions).
Una delle proprietà delle funzioni di forma è che il supporto sia piccolo, ovvero, essa è data da pochi elementi della triangolazione.

2.3.2 Elementi finiti parallelepipedi

Passiamo ora a considerare il caso (2.15). In questo paragrafo descriveremo, prima di tutto, i gradi di libertà sul quadrato di riferimento $\hat{K} = [0, 1]^2$ e poi considereremo il caso generale. Verrà mostrato nella seconda proposizione di questo paragrafo che quando $k = 1$ i gradi di libertà sono i valori ai vertici del quadrato; quando $k = 2$ a questi si aggiungono i valori dei punti centrali di ogni lato e il centro di gravità; infine quando $k = 3$ si considerano i valori ai vertici e ai punti di coordinate $\frac{1}{3}$ e $\frac{2}{3}$.
Dimostriamo, innanzitutto, che una funzione in $\mathbb{Q}_k$ è univocamente determinata dai suoi valori in questi nodi appena descritti.

Proposizione 2.3.2. *Sia q una funzione appartenente allo spazio $\mathbb{Q}_k$ $(k = 1, 2, 3)$ che si annulla ai nodi detti precedentemente allora $q \equiv 0$.*

Dimostrazione. Iniziamo dal caso $k = 1$. La restrizione di q ad ogni lato è un polinomio lineare in una variabile, quindi q si annulla su ogni lato e puó essere, conseguentemente, scritto come

$$q(x) = c_1x_1(1 - x_1)x_2(1 - x_2) \ ,$$

tale equazione implica $c_1 = 0$.
Allo stesso modo, applicando tale ragionamento ai casi $k = 2$ e $k = 3$, comporta che q ha la forma

$$q(x) = c_2x_1(\frac{1}{2} - x_1)(1 - x_1)x_2(1 - x_2) \qquad (k = 2)$$

oppure

$$q(x) = c_3x_1(\frac{1}{3} - x_1)(\frac{2}{3} - x_1)(1 - x_1)x_2(1 - x_2) \qquad (k = 3).$$

Poichè $x_1^3x_2^3 \notin \mathbb{Q}_3$, e $x_1^4x_2^2 \notin \mathbb{Q}_3$, segue che $c_2 = c_3 = 0$. $\square$

Il caso tridimensionale puó essere trattato allo stesso modo. Prendiamo il cubo di riferimento $\hat{K} = [0, 1]^3$, anche in questo caso per $k = 1$ si considerano i valori ai vertici del cubo, se $k = 2$ oltre ai vertici anche i nodi posti al centro di ogni faccia ed infine per $k = 3$ abbiamo otto nodi interni posti nelle coordinate $\frac{1}{3}$ e $\frac{2}{3}$.
Anche in questo caso, come fatto nel paragrafo precendente, possiamo ridurre il problema dell'individuazione di un polinomio $p \in \mathbb{Q}_k$ dal caso tridimensionale al caso bidimensionale. Infatti, su ogni faccia i nodi sono gli stessi del corrispondente quadrato bidimensionale e una funzione $p \in \mathbb{Q}_k$ ristretta ad una faccia da luogo ad una funzione in $\mathbb{Q}_k$ dipendente solo da due variabili. Quindi, se $p \in \mathbb{Q}_k$ si annulla nei nodi indicati allora è identicamente uguale a 0 su ogni faccia del cubo.
Se $k = 1$, possiamo scrivere

$$p(x) = cx_1(1 - x_1)x_2(1 - x_2)x_3(1 - x_3).$$

Poichè $x_1^2x_2^2x_3^3 \notin \mathbb{Q}_1$ si puó concludere che $c = 0$, ovvero, $p \equiv 0$.
I casi $k = 2$ e $k = 3$ possono essere trattati allo stesso modo.
Assumiamo ora che $K = T_K(\hat{K})$, dove T_K è una mappa affine invertibile.

Ricordiamo che $v_h \in X_h^k$ se $v_h \in C^0(\bar{\Omega})$ e $v_{h|K} \circ T_K \in \mathbb{Q}_k$. Quindi i gradi di libertà in K sono i valori di v_h ai nodi $a_{j,K} = T_K(\bar{a}_j)$, dove $\bar{a}_j$ sono i nodi in $[0,1]^d$. Inoltre, il numero totale di gradi di libertà è dato dai valori di v_h sull'insieme globale dei nodi

$$\Sigma_h := \{a_{j,K} \mid K \in \mathcal{T}_h\} \subset \bar{\Omega}.$$

Denotiamo con a_j questi nodi, per $j = a, \dots, N_h$. Le funzioni di forma sono proprio questi polinomi a tratti $\phi_j \in X_h^k$ tali che:

$$\phi_j(a_i) = \delta_{ij}\ , \qquad i, j = 1, \dots, N_h.$$

2.4 L'operatore di interpolazione

L'individuazione dei gradi di libertà e delle funzioni di forma conduce facilmente alla definizione di un *operatore di interpolazione*, ovvero, un operatore definito sullo spazio delle funzioni continue e a valori nello spazio degli elementi finiti X_h^k definito in (2.2) e (2.3).
Per ogni $v \in C^0(\bar{\Omega})$ possiamo fissare

$$\pi_h^k(v) := \sum_{i=1}^{N_h} v(a_i)\phi_i\ , \tag{2.5}$$

dove a_i sono i nodi su $\bar{\Omega}$ e ϕ_i sono le corrispondenti funzioni di forma.
Il polinomio interpolante $\phi_h^k(v)$ è l'unica funzione in X_h^k che assume gli stessi valori della funzione di partenza v in corrispondenza di tutti i nodi a_i.
In maniera simile si puó introdurre l'operatore di interpolazione *locale*, un operatore che agisce sulle funzioni definite in K e che restituisce un polinomio in $\mathbb{P}_k$ o $\mathbb{Q}_k$. Se $a_{i,K}$ $i = 1, \dots, M_k$ sono i nodi in K, fissiamo

$$\pi_K^k(v) := \sum_{i=1}^{M_k} v(a_{i,K})\phi_{i|K} \qquad \forall v \in C^0(K), \tag{2.6}$$

dove ϕ_i sono le funzioni di forma.
È possibile verificare subito che

$$\pi_h^k(v)_{|K} = \pi_K^k(v_{|K}) \qquad \forall K \in \mathcal{T}_h\ ,\ \forall v \in C^0(\bar{\Omega}). \tag{2.7}$$

2.4.1 Errore di interpolazione

Forniamo, in questo paragrafo, una stima per l'errore di interpolazione $v-\pi_h^k(v)$. Per mezzo della proprietà (2.7) di localizzazione, che presenteremo nel corso della discussione, si puó ridurre il problema alla stima di $v - \pi_h^k(v)$ su ogni singolo elemento $K \in \mathcal{T}_h$. L'idea chiave è quella di ottenere stime più adeguate sull'elemento di riferimento $\hat{K}$ e poi convertirle su ogni elemento K, usando le proprietà della mappa affine T_K.
La scelta della norma per misurare l'errore di interpolazione ha anch'essa un ruolo importante. Innanzitutto, noi siamo interessati a stimare l'errore $v - \pi_K^k(v)$ rispetto alle seminorme $|\cdot|_{m,K}$ degli spazi di Sobolev $H^m(K)$, $m \geq 0$, dove la seminorma è definita come segue

$$|v|_{m,K} := \left(\sum_{|\alpha|=k} ||D^\alpha v||^2_{L^2(K)} \right)^{\frac{1}{2}} .$$

In seguito, assumeremo che $v \in C^0(K)$, così che $\pi_K^k(v)$ sia ben definita.
Per prima cosa mostriamo che ogni seminorma di Sobolev in K è limitata da sopra e da sotto dalla corrispondente seminorma in $\hat{K}$. Per essere più esaustivi scriviamo:

$$T_K(\hat{x}) = B_K\hat{x} + b_K \ , \qquad \hat{x} \in \hat{K},$$

dove B_K è una matrice di dimensione $(d \times d)$.
Data una matrice B di dimensione $(d \times d)$, definiamo norma matriciale associata alla norma euclidea:

$$||B|| = \left\{ ||Bx|| : x \in \mathbb{R}^d, ||x|| = 1 \right\} .$$

Proposizione 2.4.1. *Per ogni $v \in H^m(K)$, $m \geq 0$, definiamo $\hat{v} := v \circ T_K$. Allora $\hat{v} \in H^m(\hat{K})$, ed inoltre esiste una costante $C = C(m,d)$ tale che*

$$|\hat{v}|_{m,\hat{K}} \leq C||B_K||^m |detB_K|^{-\frac{1}{2}} |v|_{m,K} \qquad \forall v \in H^m(K) \tag{2.8}$$

e

$$|v|_{m,K} \leq C||B_K^{-1}||^m |detB_K|^{\frac{1}{2}} |\hat{v}|_{m,\hat{K}} \qquad \forall \hat{v} \in H^m(\hat{K}) \tag{2.9}$$

dove $||\cdot||$ *è la norma matriciale associata alla norma euclidea in* $\mathbb{R}^d$.

Dimostrazione. Poichè $C^\infty(K)$ è denso in $H^m(K)$, è sufficiente provare che (2.8) è vera per una funzione v regolare. Sia

$$|\hat{v}|^2_{m,\hat{K}} = \sum_{|\alpha|=m} \int_{\hat{K}} |D^\alpha \hat{v}|^2 \ ,$$

utilizzando la regola della catena per calcolare $D^\alpha \hat{v}$ si puó facilmente trovare

$$||D^\alpha \hat{v}||_{0,\hat{k}} \leq C ||B_K||^m ||(D^\alpha v) \circ T_K||_{0,\hat{K}} \ .$$

Facendo la funzione inversa e tornando su K, si ottiene

$$||D^\alpha \hat{v}||_{0,\hat{k}} \leq C ||B_K||^m |det B_K|^{-\frac{1}{2}} ||D^\alpha v||_{0,K} \ ,$$

sommando, poi, per $|\alpha| = m$ si arriva alla tesi (2.8).
La dimostrazione della seconda disuguaglianza (2.9) viene effettuata allo stesso modo. □

È ora necessario valutare $||B_K||$ e $||B_K^{-1}||$ in termini di quantità geometriche relative a K. Definiamo

$$h_K := \ \text{diam}(K) \ , \ \rho_K := \sup \{\text{diam}(S) \mid S \text{ è una palla contenuta in K } \} \, .$$

Le stesse quantità saranno denotate con $\hat{h}$ e $\hat{\rho}$ quando si riferiranno al dominio $\hat{K}$.

Proposizione 2.4.2. *Valgono le seguenti stime*

$$||B_K|| \leq \frac{h_K}{\hat{\rho}} \qquad , \qquad ||B_K^{-1}|| \leq \frac{\hat{h}}{\rho_K} \ . \tag{2.10}$$

Dimostrazione. Possiamo scrivere

$$||B_K|| = \frac{1}{\hat{\rho}} \sup \{|B_k u| \ : \ |u| = \hat{\rho}\} \ .$$

Per ogni u che soddisfa l'equazione $|u| = \hat{\rho}$, esistono due punti $\hat{x}, \hat{y} \in \hat{K}$ tale che $\hat{x} - \hat{y} = u$. Poichè $B_K u = T_K \hat{x} - T_K \hat{y}$, si puó dedurre $|B_K u| \leq h_K$, quindi si ottiene, conseguentemente, la prima disuguaglianza di (2.10). L'altra si puó provare in modo simile. □

Abbiamo ora bisogno di una stima per la seminorma di $[v-\pi_K^k(v)]\circ T_K$ in $H^m(\hat{K})$. Nel seguito denoteremo con $[\pi_K^k(v)]^\wedge$ l'elemento $\pi_K^k(v)\circ T_K$. Poichè i nodi $a_{i,K}$ in K coincidono con $T_K(\hat{a}_i)$, dove $\hat{a}_i$ sono i nodi in $\hat{K}$ e, allo stesso modo, le funzioni di forma in $\hat{K}$ sono date da $\hat{\phi}_i = \phi_i \circ T_K$ allora

$$[\pi_K^k(v)]^\wedge = \pi_K^k(v)\circ T_K = \sum_{i=1}^{M_K} v(a_{i,K})(\phi_i \circ T_K) = \sum_{i=1}^{M_K} v(T_K(\hat{a}_i))\hat{\phi}_i = \pi_{\hat{K}}^k(\hat{v}) \ . \tag{2.11}$$

Quindi dobbiamo stimare $\hat{v} - \pi_{\hat{K}}^k(\hat{v})$ in $H^m(\hat{K})$.

Proposizione 2.4.3 (Lemma di Bramble-Hilbert)**.**
Sia $\hat{l} : H^{k+1}(\hat{K}) \to H^m(\hat{K})$, $m \geq 0,\ k \geq 0$, *una funzione continua e lineare tale che*

$$\hat{l}(\hat{p}) = 0 \qquad \forall \hat{p} \in \mathbb{P}_k \ . \tag{2.12}$$

Allora per ogni $\hat{v} \in H^{k+1}(\hat{K})$

$$|\hat{l}(\hat{v})|_{m,\hat{K}} \leq ||\hat{l}||_{\mathcal{L}(H^{k+1}(\hat{K});H^m(\hat{K}))} \inf_{\hat{p}\in\mathbb{P}_k} ||\hat{v}+\hat{p}||_{k+1,\hat{K}} \ . \tag{2.13}$$

Dimostrazione. Sia $\hat{v} \in H^{k+1}(\hat{K})$, per ogni $\hat{p} \in \mathbb{P}_k$, da (2.12), si ottiene

$$|\hat{l}(\hat{v})|_{m,\hat{K}} = |\hat{l}(\hat{v}+\hat{p})|_{m,\hat{K}} \leq ||\hat{l}||_{\mathcal{L}(H^{k+1}(\hat{K});H^m(\hat{K}))}||\hat{v}+\hat{p}||_{k+1,\hat{K}} \ ,$$

quindi segue la tesi (2.13), poichè $\hat{p}$ è preso arbitrariamente. □

In particolare, abbiamo provato che

$$|[v-\pi_K^k(v)]^\wedge|_{m,\hat{K}} \leq ||I-\pi_{\hat{K}}^k||_{\mathcal{L}(H^{k+1}(\hat{K});H^m(\hat{K}))} \inf_{\hat{p}\in\mathbb{P}_k} ||\hat{v}+\hat{p}||_{k+1,\hat{K}}$$

per $0 \leq m \leq k+1,\ k \geq 1$.
La seguente Proposizione fornisce l'ultimo strumento di cui abbiamo bisogno per dimostrare la stima desiderata per l'errore di interpolazione.

Proposizione 2.4.4 (Lemma di Deny-Lions)**.**
Per ogni $k \geq 0$ *esiste una costante* $C = C(k,\hat{K})$ *tale che*

$$\inf_{\hat{p}\in\mathbb{P}_k} ||\hat{v}+\hat{p}||_{k+1,\hat{K}} \leq C|\hat{v}|_{k+1,\hat{K}} \qquad \forall \hat{v} \in H^{k+1}(\hat{K}) \ . \tag{2.14}$$

Dimostrazione. Prima di tutto, proviamo che esiste una costante $C = C(\hat{K})$ tale per cui

$$||\hat{v}||_{k+1,\hat{K}} \leq C \left\{ |\hat{v}|^2_{k+1,\hat{K}} + \sum_{|\alpha| \leq k} \left(\int_{\hat{K}} D^\alpha \hat{v} \right)^2 \right\}^{\frac{1}{2}} \tag{2.15}$$

per ogni $\hat{v} \in H^{k+1}(\hat{K})$.
Procediamo per assurdo.
Se (2.15) non valesse, allora potremmo trovare una successione $\hat{v}_s \in H^{k+1}(\hat{K})$ tale che

$$||\hat{v}_s||_{k+1,\hat{K}} = 1 \tag{2.16}$$

e

$$|\hat{v}_s|^2_{k+1,\hat{K}} + \sum_{|\alpha| \leq k} \left(\int_{\hat{K}} D^\alpha \hat{v}_s \right)^2 < \frac{1}{s^2}. \tag{2.17}$$

Dato che l'immersione $H^{k+1}(\hat{K}) \hookrightarrow H^k(\hat{K})$ è compatta, possiamo scegliere una sottosuccessione, sempre denotata con $\hat{v}_s$, convergente in $H^k(\hat{K})$.
Come conseguenza di (2.17) $\hat{v}_s$ è una successione di Cauchy in $H^{k+1}(\hat{K})$, quindi esiste una funzione $\hat{w}$ tale che $\hat{v}_s$ converge a $\hat{w}$ in $H^{k+1}(\hat{K})$ e $||\hat{w}||_{k+1,\hat{K}} = 1$. Inoltre, $\int_K D^\alpha \hat{w} = 0$ per $|\alpha| \leq k$ e $D^\alpha \hat{w} = 0$ per $|\alpha| = k+1$. La prima relazione produce $\hat{w} = 0$, mentre la seconda implica $\hat{w} \in \mathbb{P}_k$, il che è una contraddizione. Allora (2.15) è provata.
Ora, per ogni $\hat{v} \in H^{k+1}(\hat{K})$, possiamo costruire un'unica $\hat{q} \in \mathbb{P}_k$ tale che

$$\int_{\hat{K}} D^\alpha \hat{q} = -\int_{\hat{K}} D^\alpha \hat{v} \ , \qquad \forall |\alpha| \leq k.$$

Quindi da (2.15), applicata a $\hat{v} + \hat{q}$, si ottiene

$$\inf_{\hat{p} \in \mathbb{P}_k} ||\hat{v} + \hat{p}||_{k+1,\hat{K}} \leq ||\hat{v} + \hat{q}||_{k+1,\hat{K}} \leq C|\hat{v} + \hat{q}|_{k+1,\hat{K}} = C|\hat{v}|_{k+1,\hat{K}}$$

ossia la tesi. □

Siamo ora in grado di dimostrare il risultato principale nello studio dell'errore di interpolazione.

Teorema 2.4.5. *Sia* $0 \leq m \leq k+1,\ k \geq 1$. *Allora esiste una costante* $C = C(\hat{K}, \pi^k_{\hat{K}}, k, m, d)$ *tale che*

$$|v - \pi^k_K(v)|_{m,K} \leq C \frac{h_K^{k+1}}{\rho_K^m} |v|_{k+1,K} \qquad \forall v \in H^{k+1}(K) \ . \tag{2.18}$$

Dimostrazione. Prima di tutto, ricordiamo che il teorema di inclusione di Sobolev ci dice che $H^{k+1}(K) \subset C^0(K)$ per $k \geq 1$. Quindi l'operatore di interpolazione π^k_K è definito in $H^{k+1}(K)$. Usando (2.9) e (2.11) si ottiene

$$|v - \pi^k_K(v)|_{m,K} \leq C \frac{1}{\rho_K^m} |detB_K|^{\frac{1}{2}} |\hat{v} - \pi^k_{\hat{K}}(\hat{v})|_{m,\hat{K}} \ . \tag{2.19}$$

Poichè $(I - \pi^k_{\hat{K}})(\hat{p}) = 0$ per ogni $\hat{p} \in \mathbb{P}_k$ (ricordiamo che, se $\hat{K} = [0,1]^d$, allora $\pi^k_{\hat{K}}$ è invariante su $\mathbb{Q}_k$) da (2.13) e (2.14) otteniamo

$$|\hat{v} - \pi^k_{\hat{K}}(\hat{v})|_{m,\hat{K}} \leq C |\hat{v}|_{k+1,\hat{K}} \ .$$

Infine applicando (2.10) e (2.14) troviamo

$$|\hat{v} - \pi^k_{\hat{K}}(\hat{v})|_{m,\hat{K}} \leq C h_K^{k+1} |detB_K|^{-\frac{1}{2}} |v|_{k+1,K} \ . \tag{2.20}$$

Quindi la tesi segue immediatamente da (2.19) e (2.20). □

Osservazione 1. Vale la pena notare che, se $1 \leq l < k,\ 0 \leq m \leq l+1$, allora otteniamo

$$|v - \pi^k_K(v)|_{m,K} \leq C \frac{h_K^{l+1}}{\rho_K^m} |v|_{l+1,K} \qquad \forall v \in H^{l+1}(K) \ , \tag{2.21}$$

ossia, la stessa stima dell'errore provata per l'operatore di interpolazione $\pi^l_K(v)$.
Questa osservazione mostra chiaramente che un' interpolazione di ordine alto su v non dá, in linea di principio, una miglior stima dell'errore se v non è abbastanza regolare.

Sottolineamo che risultati simili valgono anche per l'interpolazione negli spazi di Sobolev $W^{k+1,p}(\Omega)$, $p \in [1, \infty]$. La prossima osservazione è rivolta al caso $p = \infty$

Osservazione 2. È possibile stimare l'errore di interpolazione rispetto a norme differenti. Per esempio, procedendo come prima, si mostra facilmente che

$$|v - \pi_K^k(v)|_{m,\infty,K} \leq C[\mu(K)]^{-\frac{1}{2}} \frac{h_K^{l+1}}{\rho_K m} |v|_{l+1,K} \qquad \forall v \in H^{l+1}(K)$$

per $1 \leq l \leq k,\ 0 \leq m < l+1-\frac{d}{2},\ d = 2,3$ e

$$|v - \pi_K^k(v)|_{m,\infty,K} \leq C \frac{h_K^{l+1}}{\rho_K^m} |v|_{l+1,\infty,K} \qquad \forall v \in W^{l+1,\infty}(K)$$

per ogni $1 \leq l \leq k,\ 0 \leq m \leq l+1$. Qui $|\cdot|_{m,\infty,K}$ denota la seminorma in $W^{m+1,\infty(K)}$.

L'ultima stima è infatti banale, poichè sia il lemma di Bramble-Hilbert che quello di Deny-Lions continuano a valere anche se si sostituisce $H^{k+1}(\hat{K})$ e $H^m(\hat{K})$ con $W^{k+1,\infty}(\hat{K})$ e $W^{m,\infty}(\hat{K})$, rispettivamente.

Per dimostrare invece la prima stima, notiamo che per un indice α tale che $|\alpha| = m$ si ha

$$||D^\alpha v||_{\infty,K} \leq C||B_K^{-1}||^m ||D^\alpha \hat{v}||_{\infty,\hat{K}} \ .$$

Procedendo come nel Teorema (2.4.5) e ricordando che $H^{l+1}(\hat{K}) \subset W^{m,\infty}(\hat{K})$ per $0 \leq l+1-\frac{d}{2}$, si ottiene

$$|v - \pi_K^k(v)|_{m,\infty,K} \leq \frac{C}{\rho_K^m} |\hat{v}|_{l+1,\hat{K}} \ .$$

Poichè

$$|detB_K| = \frac{\mu(K)}{\mu(\hat{K})} \ ,$$

il risultato segue immediatamente da (2.8).

Prima di considerare l'errore di interpolazione globale su Ω, introduciamo una definizione che riguarda la famiglia di triangolazioni $\mathcal{T}_h$.

Definizione 2.1. Una famiglia di triangolazioni $\mathcal{T}_h,\ h > 0$ è chiamata *regolare* se esiste una costante $\sigma \geq 1$ tale che

$$\max_{K \in \mathcal{T}_h} \frac{h_K}{\rho_K} \leq \sigma \qquad \forall h > 0 \ . \tag{2.22}$$

Finalmente dimostriamo una stima per l'errore di interpolazione globale.

Teorema 2.4.6. *Sia $\mathcal{T}_h$ una famiglia regolare di triangolazioni e assumiamo che $m = 0, 1$, $k \geq 1$. Allora esiste una costante C, indipendente da h, tale che*

$$|v - \pi_h^k(v)|_{m,\Omega} \leq Ch^{l+1-m}|v|_{l+1,\Omega} \qquad \forall v \in H^{l+1}(\Omega) \ , \ 1 \leq l \leq k \ . \tag{2.23}$$

Dimostrazione. Identifichiamo la stima su K:

$$|v - \pi_h^k(v)|_{m,\Omega}^2 = \sum_{K \in \mathcal{T}_h} |v - \pi_h^k(v)|_{m,K}^2 \ .$$

Da (2.18), (2.21) e (2.22) possiamo scrivere

$$|v - \pi_h^k(v)|_{m,K} \leq Ch^{l+1-m}|v|_{l+1,K} \ ,$$

allora segue il risultato sommando su K. □

Notiamo che la restrizione sull'indice m è dovuta al fatto che l'inclusione $X_h^k \subset H^m(\Omega)$ è vera solo se $m \leq 1$. La costruzione di uno spazio di dimensione finita contenuto in $H^2(\Omega)$ richiederebbe continuità di ordine superiore su tutto il bordo.

Capitolo 3

Il problema di Dirichlet

3.1 Formulazione debole

Sia Ω un aperto limitato di $\mathbb{R}^d$, $d = 2, 3$, sia $\partial\Omega$ il bordo di Ω lipschitziano e denotiamo con $\boldsymbol{n}$ il vettore normale unitario uscente da $\partial\Omega$.
Consideriamo l'operatore lineare del secondo ordine L definito da:

$$L\omega := -\sum_{i,j=1}^{d} D_i(a_{ij}D_j w) + \sum_{i=1}^{d}[D_i(b_i w) + c_i D_i w] + a_0 w \ . \tag{3.1}$$

Abbiamo denotato con D_j le derivate parziali $\frac{\partial}{\partial x_j}$; $a_{ij} = a_{ij}(x)$,$b_i = b_i(x)$, $c_i = c_i(x)$,$a_0 = a_0(x)$ sono funzioni date. Se i coefficienti b_i e c_i sono abbastanza regolari, noi possiamo omettere o $D_i(b_i w)$ o $c_i D_i w$ in (3.1) senza perdere di generalità.

Definizione 3.1. L'operatore differenziale L è detto ellittico in Ω se esiste una costante $\alpha_0 > 0$ tale che

$$\sum_{i,j=1}^{d} a_{ij}(x)u_i u_j \geq \alpha_0|u|^2$$

per ogni $u \in \mathbb{R}^d$ e per quasi tutte le $x \in \Omega$.

All'operatore L associamo la seguente forma bilineare

$$a(w, v) := \int_\Omega \left[\sum_{i,j=1}^{d} a_{ij}D_j w D_i v - \sum_{i=1}^{d} b_i w D_i v - c_i v D_i w) + a_0 wv\right] \ , \tag{3.2}$$

dove w, v sono funzioni definite in Ω.
Formalmente questa scrittura si ottiene moltiplicando (3.1) per v, integrando su Ω, applicando la formula di Green ed eliminando le condizioni al contorno. Un tipico esempio è l'operatore di Laplace $L = -\Delta$, in questo caso la forma bilineare associata è la forma di Dirichlet

$$a(w, v) = \int_\Omega \nabla w \cdot \nabla v \ .$$

Se scegliamo un sottospazio chiuso V di $H^1(\Omega)$ tale che

$$H_0^1(\Omega) \subset V \subset H^1(\Omega) \tag{3.3}$$

e assumiamo che

$$a_{ij}, b_i, c_i, a_0 \in L^\infty(\Omega) \tag{3.4}$$

allora $a(\cdot, \cdot)$ è ben definito in $V \times V$.
Siamo interessati, quindi, a risolvere il problema variazionale: dato F, funzionale continuo e lineare sullo spazio di Hilbert V,

$$\text{trova } u \in V \ : \ A(u, v) = F(v) \qquad \forall v \in V \ . \tag{3.5}$$

Come vedremo nel seguito, la forma bilineare $A(\cdot, \cdot)$ coincide con $a(\cdot, \cdot)$. Dipendendo dalla scelta del sottospazio V, (3.5) descrive la formulazione debole del problema con condizioni al bordo di Dirichlet.

3.2 Problema di Dirichlet

Consideriamo il problema di Dirichlet omogeneo, ovvero, data una funzione f definita in Ω, trovare u tale

$$\begin{cases} Lu = f & \text{in } \Omega \\ u = 0 & \text{su } \partial\Omega \ . \end{cases} \tag{3.6}$$

La formulazione debole si legge come segue: data $f \in L^2(\Omega)$,

$$\text{trova } u \in H_o^1(\Omega) \ : \ a(u, v) = f(v) \qquad v \in H_0^1(\Omega) \ , \tag{3.7}$$

dove $(\cdot,\cdot)$ è il prodotto scalare in $L^2(\Omega)$.
Dalla formula di integrazione per parti si verifica subito che, se u è una soluzione per (3.6), allora u è anche soluzione per (3.7).
Se u è sufficientemente regolare, scegliendo $v \in C_0^\infty(\Omega)$ e integrando per parti, si trova facilmente che

$$\int_\Omega Lu\, v = \int_\Omega fv \qquad \forall v \in C_0^\infty(\Omega)$$

e quindi

$$Lu = f\ .$$

Dato che $f \in L^2(\Omega)$, questa uguaglianza è vera quasi ovunque in Ω. Inoltre, la condizione al contorno di Dirichlet $u_{|\partial\Omega}$ si ottengono nel senso di $H_0^1(\Omega)$, ovvero, la traccia di u su $\partial\Omega$ si elimina.

3.2.1 Esistenza, unicità e stima a priori della soluzione

Il teorema di base per provare l'esistenza di una soluzione è il teorema di Lax Milgram visto nel primo capitolo, che a sua volta è una generalizzazione del teorema di rappresentazione di Riesz. Con l'obiettivo di applicare questo risultato, ci accingiamo a verificare che, sotto opportune ipotesi sui dati, le condizioni per applicare il teorema di Lax Milgram sono soddisfatte.
Notiamo, prima di tutto, che ogni $f \in L^2(\Omega)$ definisce un funzionale lineare continuo $v \to (f,v)$ in $H_0^1(\Omega)$.
Inoltre dall'assunzione di ellitticità si ottiene

$$\int_\Omega \sum_{i,j=1}^{d} a_{ij} D_i v D_j v \geq \alpha_0 \int_\Omega |\nabla v(x)|^2 \qquad \forall v \in H_0^1(\Omega)\ , \tag{3.8}$$

e scambiando $u = \nabla v(x)$ nella definizione (3.1) si ha

$$\sum_{i,j=1} a_{ij}(x) D_i v(x) D_j v(x) \geq \alpha_0 |\nabla v(x)|^2 \qquad \text{quasi ovunque in } \Omega.$$

Consideriamo ora il rimanente termine $a(v,v)$. Assumendo che $div(b-c) \in L^\infty(\Omega)$, possiamo scrivere

$$\int_\Omega \left[-\sum_{i=1}^d (b_i - c_i) v D_i + a_0 v^2 \right] = \int_\Omega \left[-\frac{1}{2} \sum_{i=1}^d (b_i - c_i) D_i(v^2) + a_0 v^2 \right] =$$
$$= \int_\Omega \left[\frac{1}{2} \mathrm{div}(b-c) + a_0 \right] v^2$$

per ogni $v \in H_0^1(\Omega)$. Se C_Ω è la costante della disuguaglianza di Poincarè, ovvero:

$$\int_\Omega v^2 \leq C_\Omega \int_\Omega |\nabla v|^2 \qquad v \in H_0^1(\Omega) \ , \tag{3.9}$$

possiamo allora concludere che $a(\cdot,\cdot)$ è coerciva purchè assumiamo che per quasi tutte le $x \in \Omega$ vale la seguente disequazione

$$\frac{1}{2} div[b(x) - c(x)] + a_0(x) \geq -\eta, \qquad -\infty < \eta < \frac{\alpha_0}{C_\Omega} \ .$$

Come abbiamo già visto, il metodo di approssimazione di Galerkin a (3.7) si legge:

$$\text{trova } u_h \in V_h \ : \ A(u_h, v_h) = F(v_h) \qquad \forall v_h \in V_h \ , \tag{3.10}$$

dove V_h è un sottospazio di V di dimensione finita.
Per garantire l'esistenza e l'unicità della soluzione u e u_h per i problemi (3.10) e (3.5), rispettivamente, assumiamo che la forma bilineare $A(\cdot,\cdot)$ sia continua e coerciva, ed inoltre che il funzionale lineare $F(\cdot)$ sia continuo. Inoltre, in quanto conseguenza delle assunzioni fatte, sappiamo che vale la stima dell'errore fatta nel primo capitolo

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

Inoltre, sempre nel capitolo 1 è stato dimostrato il punto principale che prova la convergenza di u_h ad u, ovvero

$$\lim_{h \to 0} \inf_{v_h \in V_h} ||v - v_h|| = 0 \qquad \forall v \in V, \tag{3.11}$$

dove noi abbiamo denotato con $|| \cdot ||$ la norma di V.
Dimostriamo ora un risultato più generale sulla convergenza del metodo di Galerkin.

Proposizione 3.2.1. *Assumiamo che la forma bilineare $A(\cdot,\cdot)$ sia continua e coerciva in V ed il funzionale lineare $F(\cdot)$ sia continuo in V. Sia V_h una famiglia di sottospazi di V di dimensione finita. Assumiamo che esista un sottospazio $\mathcal{V}$ denso in V, tale che*

$$\inf_{v_h \in V_h} ||v - v_h|| \to 0 \ \ per \ \ h \to 0 \qquad \forall v \in \mathcal{V}. \tag{3.12}$$

Allora il metodo di Galerkin è convergente, o meglio, la soluzione u_h di (3.10) converge in V alla soluzione u di(1.6), rispetto alla norma $||\cdot||$.

Dimostrazione. Poichè $\mathcal{V}$ è denso in V, per ogni $\epsilon > 0$ possiamo trovare $v \in \mathcal{V}$ tale che

$$||u - v|| < \epsilon \, .$$

Inoltre, valendo per ipotesi (3.2.1), esiste $h_0(\epsilon) > 0$ e, per ogni $h < h_0(\epsilon)$ positivo, esiste $v_h \in V_h$ tale che

$$||v - v_h|| < \epsilon \, .$$

Quindi, usando la stima dell'errore (1.11)

$$||u - u_h|| \leq \frac{M}{\alpha}||u - v_h|| \leq \frac{M}{\alpha}(||u - v|| + ||v - v_h||) \, ,$$

e da questo segue la tesi. □

3.3 Applicazione del metodo degli elementi finiti al problema di Dirichlet

Sia $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, un dominio poligonale con $\partial\Omega$ lipschitziano e sia $\mathcal{T}_h$ una famiglia di triangolazioni per $\bar{\Omega}$. Inoltre, sia:

$$V_h = X_h^k \cap H_0^1(\Omega) = \left\{ v_h \in X_h^k \, , v_h = 0 \text{ su } \partial\Omega \right\} , \qquad k \geq 1, \tag{3.13}$$

dove X_h^k, è definito in (2.2), se il riferimento poliedrico $\hat{K}$ è l'unità d-simplesso; mentre, quando $\hat{K} = [0,1]^d$, allora lo spazio X_h^k è definito da (2.3).

Assumiamo, inoltre, che i gradi di libertà e le funzioni di forma siano quelle descritte in (2.3). Di conseguenza, per ogni $v \in C^0(\bar{\Omega})$ la funzione di interpolazione $\pi_h^k(v)$ è quella definita in (2.4.1) .
Mostriamo quindi ora la convergenza del motodo degli elementi finiti e verifichiamo che (3.11) o (3.12) vale, utilizzando i risultati di approssimazione ottenuti nel paragrafo precedente.

3.3.1 Stima dell'errore

Teorema 3.3.1. *Sia Ω un dominio poligonale di $\mathbb{R}^d$, $d = 2, 3$, con la proprietà di Lipschitz al bordo e $\mathcal{T}_h$ una famiglia regolare di triangolazioni di $\bar{\Omega}$ associata al riferimento poliedrico $\hat{K}$, che è l'unità d-simplesso o $[0,1]^d$. Supponiamo che la forma bilineare $A(\cdot,\cdot)$ sia continua e coerciva in V e il funzionale lineare $F(\cdot)$ continuo in V. Sia V_h definito come indicato in (3.13). Sotto queste ipotesi il metodo degli elementi finiti è convergente. Se, inoltre, la soluzione esatta u appartiene allo spazio $H^s(\Omega)$ per alcuni $s \geq 2$, allora la stima dell'errore vale*

$$||u - u_h||_1 \leq Ch^l ||u||_{l+1} \ , \tag{3.14}$$

dove $l = \min(k, s-1)$.

Dimostrazione. Applichiamo la proposizione (3.2.1), poichè $C^\infty(\bar{\Omega})$ è denso in $H^1(\Omega)$, possiamo prendere $\mathcal{V} = C^\infty(\bar{\Omega}) \cap H_0^1(\Omega)$.
Inoltre, per ogni $v \in \mathcal{V}$

$$\inf_{v_h \in V_h} ||v - v_h||_1 \leq ||v - \pi_h^k(v)||_1 \ ,$$

pertanto questo converge a zero per (2.23).
Verifichiamo ora la disuguaglianza (3.14) . Sotto l'ipotesi che $u \in H^s(\Omega)$, $s \geq 2$, dal teorema di inclusione di Sobolev si ha $u \in C^0(\bar{\Omega})$, quindi la funzione di interpolazione $\pi_h^k(u)$ è ben definita. In più, $\pi_h^k(u) \in V_h$, dal momento che è facilmente verificabile che $\pi_h^k(u) \in H_0^1(\Omega)$ nel caso di Dirichlet. Dal teorema (2.4.6) e dall'osservazione (1) si ottiene

$$||u - \pi_h^k(u)||_1 \leq Ch^l ||u||_{l+1} \ . \tag{3.15}$$

Inoltre,vale la stima dell'errore (1.11), ovvero

$$||u - u_h||_1 \leq \frac{M}{\alpha} \inf_{v_h \in V_h} ||u - v_h||_1 \ . \tag{3.16}$$

A questo punto la tesi segue dalle ultime due disuguaglianze (3.15) e (3.16).

□

Il risultato di convergenza (3.14) risulta *ottimale* rispetto alla norma definita sullo spazio $H^1(\Omega)$, ossia, fornisce il più alto grado di convergenza possibile rispetto a tale norma permesso dal polinomio di grado k. Tuttavia, osservando la stima dell'errore di interpolazione (2.23) per $m = 0$, ci si potrebbe aspettare che la norma in $L^2(\Omega)$ sia infatti $O(h^{l+1})$. Questo, naturalmente, è vero sotto opportune ipotesi. Per chiarire questa affermazione consideriamo il problema seguente:

$$\text{trova } \phi(r) \in V : A(v, \phi(r)) = (r, v) \qquad \forall v \in V \ . \tag{3.17}$$

Se A è continua e coerciva, l'esistenza di $\phi(r)$ è assicurata dal Teorema di Lax-Milgram. Utilizzando un altro tipo di argomentazione, possiamo ottenere un otimo risultato di convergenza in $L^2(\Omega)$.

Proposizione 3.3.2. *Siano soddisfatte le ipotesi del Teorema (3.3.1). Supponiamo, inoltre, che per ogni $r \in L^2(\Omega)$ la soluzione $\phi(r)$ di (3.17) appartiene a $H^2(\Omega)$, così che, conseguentemente al teorema del grafico chiuso, esiste una costante $C > 0$ tale che*

$$||\phi(r)||_2 \leq C||r||_0 \qquad \forall r \in L^2(\Omega) \ . \tag{3.18}$$

Allora se $u \in H^s(\Omega)$, $s \geq 2$, vale la seguente stima dell'errore:

$$||u - u_h||_0 \leq Ch^{l+1}||u||_{l+1} \quad , \quad l = min(k, s-1) \ . \tag{3.19}$$

Dimostrazione. Usando (3.17), possiamo scrivere

$$||u - u_h||_0 = \sup_{r \in L^2(\Omega) r \neq 0} \frac{(r, u - u_h)}{||r||_0} = \sup_{r \in L^2(\Omega) r \neq 0} \frac{\mathcal{A}(u - u_h, \phi(r))}{||r||_0} \ .$$

Per ogni ϕ_h arbitrario tale che $\phi_h \in V_h$, si ha

$$\mathcal{A}(u - u_h, \phi(r)) = \mathcal{A}(u - u_h, \phi(r) - \phi_h) \ ,$$

quindi

$$(r, u - u_h) \leq \gamma ||u - u_h||_1 ||\phi(r) - \phi_h||_1 \ .$$

Poichè $\phi(r) \in H^2(\Omega) \subset C^0(\bar{\Omega})$, possiamo prendere $\phi_h = \pi_h^k(\phi(r))$ e da (2.23):

$$(r, u - u_h) \leq C\gamma ||u - u_h||_1 h ||\phi(r)||_2 \ .$$

Utilizzando ora (3.18) si ottiene

$$||u - u_h|_0 \leq Ch||u - u_h||_1 \ ,$$

in questo modo segue la tesi da (3.14). □

Stima dell'errore in L^∞ Ci accingiamo, ora, a fornire una stima dell'errore in $L^\infty(\Omega)$. Scriviamo

$$||u - \pi_K^k(u)||_{L^\infty(\Omega)} \leq ||u - \pi_h^k(u)||_{L^\infty(\Omega)} + ||\pi_h^k(u) - u_h||_{L^\infty(\Omega)} \ .$$

Nell'osservazione (2) abbiamo mostrato che per $1 \leq l \leq k$ vale

$$||u - \pi_K^k(u)||_{L^\infty(\Omega)} \leq Ch_K^{l+1}[\mu(K)]^{-\frac{1}{2}}|u|_{l+1,K} \qquad \forall u \in H^{l+1}(K) \ .$$

Se la famiglia di triangolazioni è regolare, abbiamo

$$[\mu(K)]^{-\frac{1}{2}} \leq Ch_K^{-\frac{d}{2}} \ ,$$

quindi

$$||u - \pi_h^k(u)||_{L^\infty(\Omega)} \leq Ch^{l+1-\frac{d}{2}}|u|_{l+1,\Omega} \qquad \forall u \in H^{l+1}(\Omega);.$$

Dall'altra parte, poichè, in uno spazio di dimensione finita tutte le norme sono equivalenti, possiamo dedurre che

$$||\hat{u}_h - \pi_{\hat{K}}^k(\hat{u})||_{L^\infty(\hat{K})} \leq C||\hat{u}_h - \pi_{\hat{K}}^k(\hat{u})||_{0,\hat{K}} \ .$$

Allora da (2.8) segue

$$||u_h - \pi_K^k(u)||_{L^\infty(K)} \leq C|detB_K|^{-\frac{1}{2}}||u_h - \pi_K^k(u)||_{0,K} \leq C[\mu(K)]^{\frac{1}{2}}||u_h - \pi_K^k(u)||_{0,K} \ .$$

Assumiamo ora che la famiglia di triangolazioni $\mathcal{T}_h$ sia quasi uniforme, nel senso che sia regolare ed esista una costante $\tau > 0$ tale che

$$\min_{K \in \mathcal{T}_h} h_K \geq \tau h \qquad \forall h > 0 \ .$$

Allora

$$[\mu(K)]^{-\frac{1}{2}} \leq Ch^{-\frac{d}{2}} \qquad \forall K \in \mathcal{T}_h \ , \ \forall h > 0 \ ,$$

quindi

$$||u_h - \pi_h^k(u)||_{L^\infty(\Omega)} \leq Ch^{-\frac{d}{2}}||u_h - \pi_h^k(u)||_{0,\Omega} \ .$$

Infine,

$$||u_h - \pi_h^k(u)||_{0,\Omega} \leq ||u - u_h||_{0,\Omega} + ||u - \pi_h^k(u)||_{0,\Omega} \ ,$$

e, dal Teorema (2.4.6) e dalla Proposizione (3.3.2), si ha

$$||u_h - \pi_h^k(u)||_{0,\Omega} \leq Ch^{l+1}|u|_{l+1,\Omega} \qquad \forall u \in H^{l+1}(\Omega) \ .$$

Riassumendo, abbiamo ottenuto la seguente stima dell'errore

$$||u - u_h||_{L^\infty(\Omega)} \leq Ch^{l+1-\frac{d}{2}}|u|_{l+1,\Omega} \qquad \forall u \in H^{l+1}(\Omega) \ .$$

Bibliografia

[1] Alfio Quarteroni, Riccardo Sacco, Fausto Saleri
Matematica Numerica
Springer, 2008.

[2] Alfio Quarteroni
Modellistica numerica per problemi differenziali
Springer, terza edizione, 2006.

[3] Sandro Salsa
Equazioni a derivate parziali - Metodi, modelli e applicazioni
Springer, 2004.

[4] Alfio Quarteroni, Alberto Valli
Numerical approximation of partial differential equations
Springer, 1994.