Spazi di Hilbert e applicazioni

18 dicembre 2003

Abstract

In queste brevi note si riassumono gli aspetti salienti della teoria degli spazi di Hilbert. Quindi si considerano alcune applicazioni della teoria, in particolare ad un problema di apprendimento da dati sparsi. Segue qualche riferimento a risorse didattiche in rete.

1 Spazi di Hilbert

Definizione 1. (Spazio di Hilbert). *Sia H uno spazio vettoriale munito di un prodotto scalare*[1] $< \cdot, \cdot >$. *Se H è uno spazio metrico completo*[2] *rispetto alla distanza indotta*[3] *da* $< \cdot, \cdot >$, *allora H si dice spazio di Hilbert.*

Un primo esempio di spazio di Hilbert è costituito da $\mathbb{R}^N$, dotato del prodotto scalare $< x, y >= \sum_{i=1}^N x_i y_i$, che induce la norma euclidea $||x|| = \sqrt{\sum_{i=1}^N x_i^2}$.

Particolarmente interessanti per le applicazioni risultano essere gli spazi di Hilbert di funzioni, generalmente infinito-dimensionali. Un esempio fondamentale sono gli spazi L^2: sia $I \subset \mathbb{R}$ un intervallo, si pone

$$L^2(I) = \{f : I \to \mathbb{R} \quad \int_I f(x)^2 dx < +\infty\}.$$

Analogamente si definisce[4] $L^2(\Omega)$, per un dominio multidimensionale $\Omega \subset \mathbb{R}^N$. Il prodotto scalare in $L^2(\Omega)$ è dato da

$$< f, g >_{L^2} = \int_\Omega f(x)g(x)dx$$

[1] Un prodotto scalare (reale) su H è un'applicazione $< \cdot, \cdot >: H \times H \to \mathbb{R}$ bilineare simmetrica definita positiva.

[2] Uno spazio metrico X si dice completo se tutte le successioni $\{x_n\}_{n\in\mathbb{N}}$ di Cauchy in X (quelle per cui $\forall\varepsilon > 0$ si ha $\operatorname{dist}(x_n, x_m) < \varepsilon \ \forall n, m \geq n_\varepsilon$) convergono in X.

[3] Un prodotto scalare definisce una norma $|| \cdot ||$ su H ponendo $||x|| = \sqrt{< x, x >}$. A sua volta, una norma induce la distanza $\operatorname{dist}(x, y) = ||x - y||$. Tra le proprietà fondamentali della norma indotta da un prodotto scalare vi sono l'identità del parallelogramma $\frac{1}{2}||x + y||^2 + \frac{1}{2}||x - y||^2 = ||x||^2 + ||y||^2$ e la diseguaglianza di Cauchy-Schwarz $| < x, y > | \leq ||x|| \cdot ||y||$.

[4] Sia in questo caso che nel caso monodimensionale $\Omega = I \subset \mathbb{R}$, gli integrali sono intesi nel senso di Lebesgue (si tratta di una generalizzazione dell'integrale di Riemann): ciò è necessario per assicurare la completezza di L^2.

e la norma corrispondente è

$$||f||_{L^2} = \left(\int_\Omega f(x)^2 dx \right)^{\frac{1}{2}}.$$

Definizione 2. (Spazio metrico separabile). *Uno spazio metrico X si dice separabile se esiste un un sottoinsieme $Q \subset X$ numerabile e denso*[5] *in X.*

Gli spazi $\mathbb{R}^N$ ed $L^2(\Omega)$ sono spazi di Hilbert separabili[6]. Ecco una rassegna di alcuni spazi di Hilbert separabili che incontreremo in seguito.

Esempio 1. Spazi L^2 con peso. Ad esempio, data una funzione peso $p(x) > 0$ definita sull'intervallo $[a, b]$, si definisce il prodotto scalare

$$< f, g >_{L^2_p} = \int_a^b f(x)g(x)p(x)dx.$$

Esempio 2. Spazi di Sobolev[7]. Ad esempio

$$H^1([a, b]) = \{f \in L^2([a, b]),\ f' \in L^2([a, b])\},$$

dotato del prodotto scalare

$$< f, g >_{H^1} = < f, g >_{L^2} + < f', g' >_{L^2} = \int_a^b [f(x)g(x) + f'(x)g'(x)]dx,$$

e le varianti, utili per considerare problemi con date condizioni al contorno, $H^1_0([a, b]) = \{f \in H^1([a, b]),\ f(a) = f(b) = 0\}$, $H^1_{\text{per}}([a, b]) = \{f \in H^1([a, b]),\ f(a) = f(b)\}$, entrambi considerati relativamente al prodotto scalare $< \cdot, \cdot >_{H^1}$.

Esempio 3. Spazi indotti da prodotti scalari con peso nello spazio di Fourier. Ad esempio, date $f, g : \mathbb{R} \to \mathbb{R}$ regolari, dette $\hat{f}(\omega)$, $\hat{g}(\omega)$ le rispettive trasformate di Fourier, e data una funzione peso[8] $K(\omega)$, si pone

$$< f, g >_H = \int_{\mathbb{R}} K(\omega)\hat{f}(\omega)\hat{g}(\omega)d\omega.$$

Ad esempio, nel caso $K(\omega) = \omega^2 + 1$, lo spazio di Hilbert H indotto da K coincide con $H^1(\mathbb{R})$.[9]

[5] Un sottoinsieme Q di uno spazio metrico X si dice denso in X se ogni elemento di X è approssimabile mediante elementi di Q, ovvero $\forall\, x \in X\ \exists\, \{q_n\}_{n \in \mathbb{N}} \subset Q$ tale che $\text{dist}\,(x_n, x) \to 0$ per $n \to +\infty$. Ad esempio, l'insieme $\mathbb{Q}$ dei numeri razionali è denso in $\mathbb{R}$.

[6] Nel caso $\Omega = [a, b] \subset \mathbb{R}$, una famiglia numerabile densa in $L^2(\Omega)$ è data ad esempio dalle funzioni costanti su sottointervalli di Ω aventi estremi razionali, e che assumono solo valori razionali.

[7] Intervengono tra l'altro nella teoria delle equazioni differenziali alle derivate parziali.

[8] denominata kernel (nucleo)

[9] Ciò è conseguenza del fatto che la trasformata di Fourier è un'isometria rispetto al prodotto scalare L^2, o equivalentemente $||f||_{L^2(\mathbb{R})} = ||\hat{f}||_{L^2(\mathbb{R})}$ (Teorema di Plancherel)

Basi hilbertiane

Una proprietà notevole di uno spazio di Hilbert separabile H consiste nel fatto di possedere **basi hilbertiane**,[10] ovvero basi ortonormali[11] numerabili, che permettono di rappresentare ogni elemento $x \in H$ in modo unico come $x = \sum_{i=0}^{\infty} c_i e_i$, dove $\{e_i\}_{i\in\mathbb{N}} \subset H$ è la base Hilbertiana ed i $c_i \in \mathbb{R}$ sono detti coefficienti di Fourier di x rispetto alla base $\{e_i\}$, e sono dati esplicitamente dalla formula

$$c_i =< x, e_i >_H \qquad \forall\, i \in \mathbb{N}.$$

Quindi le informazioni relative ad x sono codificate nella successione $\{c_i\}_{i\in\mathbb{N}}$,[12] ed in particolare x è approssimabile con precisione arbitraria (nella distanza di H) mediante le somme parziali della serie $\sum_{i=0}^{\infty} c_i e_i$.[13]

Alcuni esempi:

Esempio 4. In $L^2([-1,1])$ una base hilbertiana è data dalle funzioni

$$\frac{1}{\sqrt{2}}, \quad \sin n\pi x, \quad \cos n\pi x, \qquad n \geq 1.$$

Per $f \in L^2([-1,1])$ si ha dunque la rappresentazione in serie (trigonometriche) di Fourier

$$f(x) = \frac{a_0}{2} + \sum_{i=1}^{\infty} a_i \cos(n\pi x) + \sum_{i=1}^{\infty} b_i \sin(n\pi x),$$

con i coefficienti di Fourier dati da

$$a_0 = \int_{-1}^{1} f(x)dx, \qquad a_i = \int_{-1}^{1} f(x)\cos(n\pi x)dx, \qquad b_i = \int_{-1}^{1} f(x)\sin(n\pi x)dx.$$

Esempio 5. In generale, negli spazi L^2 con peso p esistono basi ortonormali rispetto a $< \cdot, \cdot >_{L^2_p}$ del tipo $\{Q(x)P_n(x)\}_{n\in\mathbb{N}}$, con $P_n(x)$ un polinomio di grado n e $Q(x)$ una funzione legata al peso $p(x)$. A seconda del peso p, si ottengono le note famiglie dei polinomi di Legendre, Laguerre, Hermite, Chebyčev...

2 Proiezione ortogonale e rappresentazione

Nel quadro degli spazi di Hilbert separabili è possibile rispondere costruttivamente a problemi di approssimazione di funzioni mediante polinomi, polinomi trigonometrici o altre funzioni speciali (ad es. elementi finiti), per via dell'esistenza della proiezione ortogonale.

[10] altrimenti dette sistemi ortonormali completi

[11] Un insieme $\{v_j\}_{j\in J} \subset H$ si dice ortonormale se $||v_j||_H = 1\ \forall\, j \in J$ e $< v_i, v_j >_H = 0$ se $i \neq j$.

[12] ad esempio, $||x||_H^2 = \sum_{i=0}^{\infty} c_i^2$ (identità di Parseval)

[13] Su questo principio si basa ad esempio il sistema di codifica di immagini JPEG.

Sia H uno spazio di Hilbert separabile e consideriamo un suo sottospazio N-dimensionale V_N. Supponiamo per comodità che $V_N = \text{span}(e_1, ..., e_N)$ sia lo spazio vettoriale generato dai primi N elementi di una base Hilbertiana $\{e_i\}_{i \in \mathbb{N}}$ di H. Consideriamo il seguente problema di approssimazione:

Problema (dei minimi quadrati): Dato $f \in H$, determinare (se esiste) il punto $u \in V_N$ a minima distanza da f.

Soluzione: Si tratta di un problema di minimo in N variabili. Consideriamo il problema equivalente di minimizzare il quadrato della distanza:

$$\min\{||f - v||_H^2, \ v \in V_N\}.$$

Sfruttando la decomposizione ortogonale $v = \sum_{i=1}^N y_i e_i$ si ha

$$\begin{aligned} G(v) = ||f - v||^2 &= < f - v, f - v > = < f, f > -2 < f, v > + < v, v > \\ &= ||f||^2 - 2 \sum_{i=1}^N y_i < f, e_i > + \sum_{i=1}^N y_i^2. \end{aligned}$$

La funzione $G(y_1, ..., y_N) = G(v)$ è un polinomio di secondo grado, cresce a $+\infty$ quando $|v| = |(y_1, ..., y_N)| \to +\infty$, ed ha come unico punto critico il punto u di coordinate $y_i = < f, e_i >$, che risulta pertanto essere il minimo di G.

Riassumendo, il problema dei minimi quadrati ha un'unica soluzione u, che viene detta **proiezione ortogonale** di f su V_N, in quanto

$$u = \sum_{i=1}^N < f, e_i > e_i \qquad \Longleftrightarrow \qquad < f - u, u > = 0, \tag{1}$$

come si verifica facilmente.

Nota 1. Va osservato che la rappresentazione esplicita (1) della proiezione ortogonale di f è particolarmente semplice per il fatto di aver fissato una base **ortonormale** di V_N.

Nota 2. Per quanto visto sopra, la proiezione ortogonale di f sullo spazio V_N generato dai primi N elementi di una base hilbertiana $\{e_i\}_{i \in \mathbb{N}}$ di H coincide con la somma parziale N-esima di f nella sua rappresentazione rispetto alla base $\{e_i\}$.

In particolare, se il problema di partenza consisteva nel ricercare la migliore approssimazione in L^2 di una funzione $f : [-1, 1] \to \mathbb{R}$ mediante polinomi trigonometrici di grado al più N,[14] dato che lo spazio V_N generato da essi coincide con lo spazio generato da $\{\frac{1}{2}, \ \cos(n\pi x), \ \sin(n\pi x)\}_{1 \leq n \leq N}$, la risposta sarà data precisamente dalla somma parziale troncata all'indice N della sua serie (trigonometrica) di Fourier.

[14] ovvero polinomi di grado al più N nelle variabili $\cos(\pi x)$ e $\sin(\pi x)$

Abbiamo visto come si dimostra facilmente l'esistenza della proiezione ortogonale su un sottospazio V_N a dimensione finita. In realtà, in uno spazio di Hilbert c'è la proiezione ortogonale anche su particolari sottospazi di dimensione infinita, o particolari sottoinsiemi convessi. Questo fatto permette di stabilire tra l'altro risultati di esistenza e unicità per soluzioni di equazioni differenziali (ordinarie e/o alle derivate parziali, con eventuali dati al contorno) riformulando le cose in termini di problemi di minimo.

Teorema 1. (Proiezione ortogonale). *Sia H uno spazio di Hilbert, $K \subset H$ un sottoinsieme convesso[15] chiuso.[16] Allora, $\forall\, f \in H$ esiste un unico $u \in K$ tale che*

$$||f-u||_H = \min_{v\in K} ||f-v||_H, \tag{2}$$

ovvero

$$||f-u||_H^2 = \min_{v\in K} \left(||f||_H^2 - 2 < f, v >_H + ||v||_H^2 \right). \tag{3}$$

Inoltre, u è caratterizzato dalle proprietà

$$u \in K, \qquad < f-u, v-u > \le 0 \qquad \forall\, v \in K. \tag{4}$$

Nel caso in cui $K = M$ sia un sottospazio lineare chiuso di H, allora (4) equivale a

$$u \in M, \qquad < f-u, v > = 0 \qquad \forall\, v \in M. \tag{5}$$

L'elemento u si dice proiezione ortogonale di f su K.

Un secondo importante risultato che unitamente al teorema di proiezione ortogonale permette di stabilire risultati di esistenza nonché di rappresentazione di minimi di funzioni definite in spazi di Hilbert è il seguente.

Teorema 2. (di rappresentazione di Riesz-Fréchét). *Sia H uno spazio di Hilbert, $F : H \to \mathbb{R}$ una funzione lineare e continua (ovvero $\exists C > 0$ tale che $|F(v)| \le C||v||_H$ $\forall v \in H$). Allora esiste un unico elemento $k \in H$ tale che*

$$F(v) = < k, v >_H \qquad \forall\, v \in H.$$

Si dice che k rappresenta la funzione F.

Nota 3. In dimensione finita il teorema di rappresentazione di Riesz è immediato. Nel caso generale la dimostrazione si basa sul teorema di esistenza della proiezione ortogonale.

[15] Un sottoinsieme K di uno spazio vettoriale si dice convesso se $\forall\, x_1, x_2 \in K$, il segmento $\{tx_1 + (1-t)x_2 \ \ 0 \le t \le 1\}$ è contenuto in K.

[16] Un sottoinsieme K di uno spazio metrico X si dice chiuso se per ogni successione $\{x_n\} \subset K$ convergente a $x \in X$ (ovvero $\operatorname{dist}(x_n, x) \to 0$ per $n \to +\infty$), si ha che $x \in K$.

Esempio 6. Data $f \in L^2([a,b])$ consideriamo la funzione

$$F(v) = < f, v >_{L^2} = \int_a^b f(x)v(x)dx, \qquad v \in H^1([a,b]).$$

$F : H^1([a,b]) \to \mathbb{R}$ è ovviamente lineare. Inoltre è continua rispetto alla norma H^1: infatti vale

$$|F(v)| = | < f, v >_{L^2} | \leq ||f||_{L^2} ||v||_{L^2} \leq ||f||_{L^2} ||v||_{H^1},$$

dove si è utilizzata prima la diseguaglianza di Cauchy-Schwarz e poi il fatto che $||v||^2_{H^1} = ||v||^2_{L^2} + ||v'||^2_{L^2} \geq ||v||^2_{L^2}$.

3 Problemi variazionali

I risultati del paragrafo precedente ci permettono di stabilire risultati di esistenza e unicità di soluzioni per alcuni problemi di minimo. Inoltre, l'esistenza di basi hilbertiane associate ai medesimi problemi di minimo rende quasi-ottimali le approssimazioni finito-dimensionali del problema.

Teorema 3. *Sia H uno spazio di Hilbert, $F : H \to \mathbb{R}$ una funzione lineare e continua, $K \subset H$ un sottospazio chiuso. Allora il problema*

$$\min_{v \in K} \mathcal{F}(v) = \frac{1}{2}||v||^2_H + F(v) \tag{6}$$

ammette un'unica soluzione $u \in K$ caratterizzata dalla proprietà

$$< u, v > + F(v) = 0 \qquad \textit{per ogni } v \in K. \tag{7}$$

Dimostrazione. Per il teorema di rappresentazione di Riesz-Fréchét, esiste $f \in H$ tale che $F(v) = < f, v >_H$. Pertanto il problema di minimo (6) equivale al problema

$$\min_{v \in K} \frac{1}{2}||v||^2_H + < f, v >_H + \frac{1}{2}||f||^2_H = \frac{1}{2}||v - (-f)||^2_H, \tag{8}$$

che, come abbiamo visto, ammette un'unica soluzione $u \in K$ caratterizzata dalla proprietà (5), ovvero la (7).

Nota 4. La relazione (7) si chiama **equazione di Eulero-Lagrange** del funzionale[17] $\mathcal{F}(v) = \frac{1}{2}||v||^2_H + F(v)$. Essa si ottiene considerando la derivata direzionale del funzionale $\mathcal{F}$ nella direzione v, e uguagliando a zero la derivata calcolata nel punto u (dato che u è un punto di minimo per $\mathcal{F}$). In altre parole, fissata un vettore (direzione)

[17] si chiamano così, in generale, le funzioni $\mathcal{F} : X \to \mathbb{R}$, dove X è uno spazio metrico i cui elementi sono essi stessi funzioni (ad es. $X = L^2([a,b])$)

$v \in K$, si considera la funzione $G(t)$ della variabile reale t definita da $G(t) = \mathcal{F}(u+tv)$, e si pone $G'(0) = 0$. Nel nostro caso,

$$\begin{aligned} G(t) &= \frac{1}{2}||u+tv||^2 + F(u+tv) \\ &= \frac{1}{2}||u||^2 + t < u,v > + \frac{t^2}{2}||v||^2 + F(u) + tF(v) \\ &= \mathcal{F}(u) + t(< u,v > + F(v)) + \frac{t^2}{2}||v||^2, \end{aligned}$$

ovvero un polinomio di secondo grado in t: la derivata in $t = 0$ restituisce l'equazione di Eulero-Lagrange (7).

Il Teorema 3 afferma tra l'altro che risolvere il problema (6) **equivale** a risolvere l'equazione di Eulero-Lagrange (7).

Il Teorema 3 si generalizza al caso di problemi variazionali del tipo

$$\min_{v \in K} \mathcal{F}(v) = \frac{1}{2}B(v,v) + \varphi(v), \tag{9}$$

con $B(\cdot,\cdot) : H \to \mathbb{R}$ una funzione bilineare simmetrica continua e coerciva[18] (ovvero un prodotto scalare su H "equivalente" a $< \cdot,\cdot >_H$), e φ una funzione convessa[19] e continua. La soluzione u del problema di minimo su un sottospazio chiuso $K \subset H$ esiste, è unica, e soddisfa l'equazione di Eulero-Lagrange

$$B(u,v) + < \varphi'(u), v > = 0 \qquad \text{per ogni } v \in K. \tag{10}$$

Nel caso in cui φ sia lineare, tale risultato è noto come Teorema di Lax-Milgram.

Un esempio particolare è dato dal problema di minimo

$$\min_{v \in H} \mathcal{F}(v) = \lambda||v||_H^2 + ||v-y||_L^2, \tag{11}$$

dove $\lambda > 0$ e $y \in H$ sono prefissati, e la norma $||\cdot||_L$ continua rispetto alla norma $||\cdot||_H$, ovvero $||v||_L \le M||v||_H$ per ogni $v \in H$. Ritroveremo questo funzionale nell'esempio 8.

Il metodo di Ritz - Galerkin

Accenniamo al problema dell'approssimazione della soluzione del problema di minimo (9), o equivalentemente, dell'equazione (10), nel caso $\varphi = F$ sia lineare e continua, e dunque sia rappresentata da un certo elemento $f \in H$.

[18] una funzione bilineare $B(v,v)$ è continua se $|B(v,v)| \le M||v||_H^2$ per ogni $v \in H$, ed è coerciva se esiste $m > 0$ tale che $m||v||_H^2 \le B(v,v)$ per ogni $v \in H$. Nel caso in cui H abbia dimensione finita, B è rappresentata da una matrice simmetrica definita positiva.

[19] ovvero $\varphi(tx_1 + (1-t)x_2) \le t\varphi(x_1) + (1-t)\varphi(x_2)$ per ogni $x_1, x_2 \in H$ e per ogni $0 \le t \le 1$. Esempi di funzioni convesse: le funzioni lineari, una norma $||\cdot||$, il suo quadrato,...

Fissiamo una base hilbertiana $\{e_i\}$ di H, e sia $V_N = \text{span}(e_1, ..., e_n)$ lo spazio generato dai primi N vettori della base. Consideriamo il problema finito-dimensionale

$$\min_{v \in V_N} \mathcal{F}(v) = \frac{1}{2}B(v,v) + F(v), \tag{12}$$

ovvero l'equazione

$$B(u,v) + < f, v >= 0 \qquad \text{per ogni } v \in V_N. \tag{13}$$

Sia u_N la soluzione di (12), ovvero di (13). Allora $u_N = \sum_{i=1}^N c_i e_i$, e, posto $f = \sum_{i=1}^\infty f_i e_i$ e $v = \sum_{i=1}^N v_i e_i$, la (10) diventa la relazione

$$\sum_{i,j=1}^N B(e_j, e_i) c_j v_i + \sum_{i=1}^N f_i v_i = 0 \qquad \text{per ogni } (v_1, ..., v_N) \in V_N, \tag{14}$$

che implica che i coefficienti c_i sono soluzioni del sistema lineare

$$\sum_{j=1}^N B(e_i, e_j) c_j = f_i \qquad \text{per } i = 1, ...N. \tag{15}$$

La situazione più favorevole da un punto di vista computazionale per risolvere il sistema (15) si avrà evidentemente quando la matrice $B_N = (B(e_i, e_j))$ è diagonale, ossia $B(e_i, e_j) = \lambda_i < e_i, e_j >$ per $i, j = 1, ..., N$, in altre parole, quando la base ortonormale di V_N è formata da autovettori di B_N.[20] In tal caso (15) diventa

$$\lambda_j c_j = f_j, \qquad \text{ossia } u_N = \sum_{j=1}^N \frac{< f, e_j >}{B(e_j, e_j)} e_j, \tag{16}$$

da confrontarsi con la formula (1), e la successiva Nota 1. In molti casi, per $B(\cdot, \cdot)$ vale un teorema spettrale, ovvero esiste una base hilbertiana $\{e_i\}$ di H composta di autovettori di B, che sono automaticamente autovettori di B_N per ogni N.

A questo punto ci si chiede se u_N converge alla soluzione u del problema originario. Anche qui, in molti casi la risposta è affermativa, ed avendo fissato una base hilbertiana $\{e_i\}$ di autovettori di B, per la (16) la soluzione u si rappresenterà mediante la serie

$$u = \sum_{j=1}^\infty \frac{< f, e_j >}{B(e_j, e_j)} e_j. \tag{17}$$

Per quanto riguarda le stime di convergenza di u_N a u, in alcuni casi[21] si ha la stima di quasi-ottimalità

$$||u - u_N||_H \leq C \min_{v \in V_N} ||u - v||_H$$

per la soluzione u_N del problema approssimato.

[20] Questo si può sempre fare grazie al teorema spettrale, dato che B_N è simmetrica e definita positiva per le ipotesi su $B(\cdot, \cdot)$. Gli autovettori dipenderanno tuttavia dalla matrice B_N.

[21] ad esmpio nei problemi classici ambientati in H^1, si veda l'esempio 7

Esempi di applicazioni

Esempio 7. (Problema di Sturm-Liouville). Vogliamo risolvere il problema con dati al contorno omogenei[22]

$$-(p(x)y')' + q(x)y = f(x) \qquad y(a) = 0, y(b) = 0, \tag{18}$$

dove $p(x)$ e $q(x)$ sono funzioni regolari definite sull'intervallo $[a, b]$, e si richiede $p(x) > 0$ $\forall\, x \in [a, b]$. Ci limitiamo per semplicità al caso $[a, b] = [0, 1]$, $p(x) = 1$ e $q(x) = 1$ per ogni $0 \leq x \leq 1$. Sia dunque $f \in L^2([0, 1])$ assegnata, e consideriamo il problema, nella sua *formulazione forte*,

$$-y'' + y = f(x), \qquad y(0) = y(1) = 0. \tag{19}$$

Moltiplicando ambo i membri di (19) per una funzione $v(x)$ tale che $v(0) = v(1) = 0$ ed integrando per parti, otteniamo la *formulazione debole*[23]

$$\int_0^1 y'v' + yv = \int_0^1 fv, \qquad \text{per ogni } v \in H_0^1([0, 1]), \tag{20}$$

ben definita per $y \in H_0^1([0, 1])$.[24] L'equazione (20) corrisponde all'equazione

$$< y, v >_{H^1} - < f, v >_{L^2} = 0 \qquad \text{per ogni } v \in H_0^1([0, 1]),$$

che è del tipo (7), ed ammette un'unica soluzione[25] $y \in H_0^1([0, 1])$ (per via dell'esempio 6 e del Teorema 3 applicato al caso $H = H^1([0, 1])$ e $K = H_0^1([0, 1])$), caratterizzata dalla proprietà (si confronti con (8))

$$\mathcal{F}(y) = \min_{v \in H_0^1([0,1])} \mathcal{F}(v) = \frac{1}{2}\int_0^1 (v')^2 + v^2 - \int_0^1 fv = \frac{1}{2}||v||_{H^1}^2 - < f, v >_{L^2} .$$

Si può dimostrare che esiste una base ortonormale $\{e_n\}$ di $L^2([0, 1])$ di autovettori della forma bilineare $B(v, v) = ||v||_{H^1}^2$: in particolare gli e_n saranno soluzioni del problema agli autovalori

$$-y'' + y = \lambda y \qquad y(0) = y(1) = 0,$$

le cui soluzioni sono date esplicitamente da $e_n = \frac{1}{\sqrt{2}} \sin(\pi n x)$, corrispondenti agli autovalori $\lambda_n = \pi^2 n^2 + 1$, per $n \in \mathbb{N}$.

---

[22] tali problemi intervengono ad esempio nello studio degli equilibri e dei modi vibratorî di strutture elastiche aventi gli estremi vincolati (corde vibranti, travi, membrane circolari e/o rettangolari, ecc.)

[23] Se la formulazione forte corrisponde all'equazione di Newton $F = ma$, la formulazione debole corrisponde al principio dei lavori virtuali: $\int F \cdot \delta r = \int ma \cdot \delta r$ per ogni spostamento virtuale δr.

[24] In particolare non sono necessarie ipotesi sull'esistenza della derivata seconda y''.

[25] detta *soluzione debole* di (19)

In particolare, posto $f = \sum_{n=1}^{\infty} < f, e_n > e_n$, la rappresentazione esplicita della soluzione y sarà data, ai sensi di (17), da

$$y(x) = \sum_{n=1}^{\infty} \frac{1}{2} \left(\int_0^1 f(t) \sin(n\pi t) dt \right) \frac{\sin(n\pi x)}{\pi^2 n^2 + 1},$$

ovvero da

$$\begin{aligned} y(x) &= \int_0^1 \left(\frac{1}{2} \sum_{n=1}^{\infty} \frac{\sin(n\pi t) \sin(n\pi x)}{n^2\pi^2 + 1} \right) \cdot f(t) dt \\ &= \int_0^1 G(x,t) \cdot f(t) dt =< G(x, \cdot), f >_{L^2}, \end{aligned} \tag{21}$$

dove si è evidenziato il ruolo della funzione (kernel)[26] $G(x,t)$ associata al problema (19) per la rappresentazione della soluzione in funzione del termine noto (sorgente) f.

Nel quadro del problema di Sturm-Liouville (18) si ripropone in generale una situazione analoga al caso (19), con la corrispondente formulazione debole in spazi dotati di norme L^2 con peso. In particolare, a seconda delle funzioni peso $p(x)$ e $q(x)$, si riotterranno le basi hilbertiane di L^2 composte dai polinomi di Legendre, Hermite, Chebičev..., visti ora come soluzioni del problema agli autovalori

$$(p(x)y')' + q(x)y = \lambda y,$$

con le opportune condizioni al contorno.

Esempio 8. (Apprendimento da dati sparsi). Consideriamo una collezione di dati $y_1, ..., y_N$, e cerchiamo una funzione f che abbia una certa regolarità, e tale che nei punti $x_1, ..., x_N$ interpoli questi dati, o perlomeno ne minimizzi la discrepanza, se questi sono troppo sparsi rispetto alla regolarità richiesta per f. Siamo condotti a considerare un problema variazionale del tipo

$$\min_{f \in H} \mathcal{F}_\lambda(f) = \lambda ||f||_H^2 + \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2 \tag{22}$$

dove $|| \cdot ||_H$ è una norma in un opportuno spazio di funzioni $f : \Omega \to \mathbb{R}$ aventi la regolarità desiderata e $\lambda > 0$ un parametro (moltiplicatore di Lagrange). Per ottenere un problema ben posto, cioè con esistenza e unicità della soluzione, possiamo ispirarci a (11) (o a (9)), e formulare (22) in un opportuno spazio di Hilbert H di cui $|| \cdot ||_H$ sia la norma, e per cui si abbia la proprietà

$$|F_{x_i}(f)| \leq M ||f||_H \qquad \text{per ogni } i = 1, ..., N, \tag{23}$$

dove si è indicata con $F_{x_i} : H \to \mathbb{R}$ la *funzione di valutazione* $F_{x_i}(f) = f(x_i)$, che è ovviamente lineare. La proprietà (23) equivale ad affermare che le funzioni di valutazione F_{x_i} sono continue rispetto alla norma $|| \cdot ||_H$, il che implica quindi la continuità della funzione convessa

$$\varphi(f) = \frac{1}{N} \sum_{i=1}^{N} (F_{x_i}(f) - y_i)^2,$$

[26]detta anche funzione o nucleo di Green

e garantisce l'esistenza e unicità della soluzione del problema (22) come caso particolare di (9) o (11). L'equazione di Eulero-Lagrange verificata dalla soluzione u di (22) si ricava, come al solito, derivando in $t = 0$ la funzione di una variabile $G(t) = \mathcal{F}(u+tv)$, dove $v \in H$, ed uguagliando a zero la derivata per ogni scelta della direzione $v \in H$. In questo caso, si ottiene:

$$\begin{aligned} G(t) &= \lambda||u+tv||_H^2 + \frac{1}{N}\sum_{i=1}^{N}(F_{x_i}(u+tv) - y_i)^2 \\ &= \lambda||u||_H^2 + 2\lambda t < u, v >_H + \lambda t^2||v||_H^2 \\ &+ \frac{1}{N}\sum_{i=1}^{N}(F_{x_i}(u) - y_i)^2 + 2t(F_{x_i}(u) - y_i)F_{x_i}(v) + t^2(F_{x_i})(v))^2, \end{aligned} \tag{24}$$

per cui $G'(0) = 0$ se e solo se

$$2\lambda < u, v >_H + \frac{2}{N}\sum_{i=1}^{N}(F_{x_i}(u) - y_i)F_{x_i}(v) = 0 \qquad \text{per ogni } v \in H. \tag{25}$$

La (25) si riscrive come

$$< u, v >_H + \frac{1}{\lambda N}\sum_{i=1}^{N}(u(x_i) - y_i) \cdot v(x_i) = 0. \tag{26}$$

La proprietà chiave per garantire la buona posizione del problema (22) è che su H valga la (23), ossia la continuità dei funzionali di valutazione $F_{x_i}(f) = f(x_i)$ in ogni punto di interpolazione $x_i \in \Omega$. Uno spazio di Hilbert H i cui elementi siano funzioni $f : \Omega \to \mathbb{R}$ per cui valga la (23) per ogni $x_i \in \Omega$ si dice **Reproducing Kernel Hilbert Space** (più brevemente, RKHS).

In un RKHS, per il Teorema di rappresentazione di Riesz-Fréchét, si avrà dunque, fissato $x \in \Omega$,

$$f(x) = F_x(f) =< K(x, \cdot), f(\cdot) >_H \qquad \forall\, f \in H \tag{27}$$

per una certa funzione $K(x, \cdot) \in H$ indicizzata da $x \in \Omega$. Il nucleo $K(x, y)$ verifica la proprietà *riproducente*

$$K(x, y) =< K(x, \cdot), K(y, \cdot) >_H, \tag{28}$$

ed in particolare è simmetrico[27] e definito positivo.[28]

In effetti, il nucleo riproducente $K(x, y)$ determina univocamente lo spazio RKHS[29].

[27]la formula precedente, nel caso particolare $f(x) = K(y, x)$, dà $K(y, x) =< K(x, \cdot), K(y, \cdot) >_H$, e si conclude grazie alla simmetria di $< \cdot, \cdot >_H$

[28]ovvero, la matrice simmetrica $K(x_i, x_j)$ è definita positiva per ogni scelta di N-uple di punti $x_i \in \Omega$, $i = 1, ..., N$

[29]nel senso che dato un nucleo simmetrico e definito positivo $K : \Omega \times \Omega \to \mathbb{R}$, esiste un unico RKHS H avente $K(x, y)$ quale nucleo riproducente

Tramite il nucleo $K(x,y)$ è possibile ottenere una rappresentazione ottimale per la soluzione u del problema variazionale (22): infatti, ponendo $v = K(x,\cdot)$ nell'equazione di Eulero-Lagrange (26), si ottiene

$$< u, K(x,\cdot) >_H + \frac{1}{\lambda N} \sum_{i=1}^{N} (u(x_i) - y_i) \cdot K(x, x_i) = 0. \tag{29}$$

Per la (27) si avrà quindi la rappresentazione

$$u(x) = -\frac{1}{\lambda N} \sum_{i=1}^{N} (u(x_i) - y_i) \cdot K(x, x_i) = \sum_{i=1}^{N} c_i K(x, x_i). \tag{30}$$

La formula notevole (30) afferma che u coincide con la soluzione del problema **finito-dimensionale**

$$\min_{f \in V_N} \mathcal{F}(f), \qquad V_N = \text{span}\,(K(x_1,\cdot), ..., K(x_N,\cdot)). \tag{31}$$

Alcuni esempi di RKHS

A mo' di conclusione vediamo qualche esempio concreto di RKHS, iniziando con un controesempio.

Esempio 9. Gli spazi L^2 **non** sono RKHS, ovvero i funzionali di valutazione non sono continui rispetto alla norma L^2. Si consideri ad esempio su $L^2([0,1])$ la successione di funzioni $f_n \in L^2([0,1])$ definita da $f_n(x) = \sqrt{2n+1} \cdot x^n$: si ha

$$\frac{F_1(f_n)}{||f_n||_{L^2([0,1])}} = \frac{f_n(1)}{||f_n||_{L^2([0,1])}} = \sqrt{2n+1} \to +\infty \text{ per } n \to +\infty,$$

pertanto il funzionale F_1 non risulta continuo in $L^2([0,1])$.

L'esempio precedente dice, in sostanza, che non si può sperare di avere un controllo puntuale di una funzione (ovvero controlli sui funzionali di valutazione) se si dispone di controlli globali relativi alla sola media quadratica[30] della funzione data.

La situazione cambia se si controllano le derivate della funzione, come mostra il seguente

Esempio 10. Sia $f \in H_0^1([0,1])$. Allora, dato $0 \le x \le 1$, si ha

$$\begin{aligned} |f(x)| &= |\int_0^x f'(t)dt\,| \le \int_0^x |f'(t)|dt \\ &\le |x|^{1/2} \left(\int_0^1 |f'(t)|^2 \right)^{1/2} \le ||f'||_{L^2}, \end{aligned} \tag{32}$$

[30] o, in generale, alla sola p-media integrale $\int_0^1 |f(x)|^p dx$, per un certo $p \ge 1$.

dove si è usata la diseguaglianza di Cauchy-Schwarz nel terzo passaggio. La (32) afferma che su $H_0^1([0,1])$ i funzionali di valutazione sono continui rispetto alla norma $||f||_H = ||f'||_{L^2}$, che è una norma equivalente[31] alla norma H^1. Pertanto, $H_0^1([0,1])$ è un RKHS. Descriviamo il nucleo riproducente $K(x,y)$ relativo alla norma $||\cdot||_H$ testè introdotta. Per ogni $0 \le x \le 1$ si ponga $K_x(y) = K(x,y)$. Deve valere la relazione

$$f(x) = < K_x(\cdot), f(\cdot) >_H = \int_0^1 K_x'(y) f'(y) dy \qquad \forall f \in H_0^1([0,1]),$$

ovvero, integrando per parti,[32]

$$f(x) = -\int_0^1 K_x''(y) f(y) dy \qquad \forall f \in H_0^1([0,1]). \tag{33}$$

La (33) sarà soddisfatta se $-K_x''(y) = \delta(x-y)$, dove $\delta(t)$ indica la delta (o massa) di Dirac[33] valutata nel punto t. In particolare si avrà $K_x(y) = -\frac{|x-y|}{2} + c_1(x)y + c_2(x)$, e la scelta di $c_1(x)$ e $c_2(x)$ è imposta dalla richiesta che $K_x \in H_0^1([0,1])$, ossia $K_x(0) = K_x(1) = 0$ per ogni $0 \le x \le 1$. Si ha quindi

$$\begin{aligned} K(x,y) = \frac{1}{2}(-|x-y| + (|x-1| - |x|)y + |x|) &= (1-x)y \qquad \text{se } x \ge y. \\ &= (1-y)x \qquad \text{se } y \ge x. \end{aligned}$$

Si osservi come $K(x,y)$ sia un nucleo simmetrico e definito positivo. Se avessimo considerato su $H_0^1([0,1])$ la norma $||\cdot||_{H^1}$, avremmo ottenuto un nucleo riproducente $\tilde{K}$ caratterizzato dalla relazione[34]

$$-\tilde{K}_x''(y) + \tilde{K}_x(y) = \delta(x-y)\,, \qquad \tilde{K}_x(0) = \tilde{K}_x(1) = 0\,, \qquad \forall\, 0 \le x, y \le 1.$$

Come utile esercizio si può trovare la relazione tra $\tilde{K}(x,y)$ ed il nucleo di Green $G(x,y)$ per l'equazione (19) (nell'esempio 7), definito nella formula (21).

In generale saranno RKHS spazi di funzioni dotati di norme che controllano le derivate (prime, o di ordine superiore). Vediamo un esempio (in dimensione uno) in cui la norma è data mediante una funzione peso nello spazio delle trasformate di Fourier delle funzioni date (si veda l'esempio 3).

Esempio 11. Data una funzione $G : \mathbb{R} \to \mathbb{R}^+$, sia H lo spazio di Hilbert associato alla norma

$$||f||_H^2 = \int_{\mathbb{R}} G(\omega) |\hat{f}(\omega)|^2 d\omega\,,$$

[31] Due norme $||\cdot||_1$ e $||\cdot||_2$ su di un medesimo spazio X si dicono equivalenti se esistono due costanti positive C_1 e C_2 tali che $||f||_1 \le C_1 ||f||_2$ e $||f||_2 \le C_2 ||f||_1$ per ogni $f \in X$. In altre parole, le due norme si controllano a vicenda. Nell'esempio considerato, dalla (32) si deduce $||f||_{L^2}^2 \le ||f'||_{L^2}^2$, ed in particolare si ha $||f'||_{L^2} \le ||f||_{H^1} \le \sqrt{2} ||f'||_{L^2}$ per ogni $f \in H_0^1([0,1])$.

[32] supponendo che ciò si possa fare: l'operazione è comunque giustificata nel senso delle distribuzioni

[33] altrimenti nota come impulso ideale in teoria dei segnali

[34] come sopra, si pone $\tilde{K}_x(y) = \tilde{K}(x,y)$

per $f : \mathbb{R} \to \mathbb{R}$. Se $|G(\omega)| \to +\infty$ abbastanza rapidamente per $|\omega| \to \infty$, (ad esempio, $\omega^2 \leq O(G(\omega))$, allora la norma $||f||_H$ controlla perlomeno la derivata prima di f (si veda l'esempio 3), e per quanto visto nel precedente esempio 8, lo spazio H è un RKHS.

Per descriverne il nucleo K in funzione del peso G, basterà esplicitare rispetto a K la relazione

$$f(x) = \int_{\mathbb{R}} G(\omega)\hat{f}(\omega) \cdot \hat{K}_x(\omega) d\omega.$$

Ora, per ogni f si ha

$$f(x) = \int_{\mathbb{R}} f(t)\delta(x-t)dt = \int_{\mathbb{R}} \hat{f}(\omega)e^{i\omega\cdot x}d\omega,$$

da cui si ricava la relazione puntuale

$$e^{i\omega\cdot x} = G(\omega) \cdot \hat{K}_x(\omega), \qquad \text{ovvero} \quad \hat{K}_x(\omega) = \frac{e^{i\omega\cdot x}}{G(\omega)}.$$

Si ricava finalmente l'espressione

$$K_x(y) = \int_{\mathbb{R}} \frac{e^{i\omega\cdot(x-y)}}{G(\omega)} d\omega = \tilde{K}(y-x),$$

dove $\tilde{K}$ è la trasformata di Fourier di $\frac{1}{G}$. In questo contesto il nucleo riproducente è invariante per traslazioni[35]. Nel caso particolare in cui $G(\omega) = e^{\sigma\omega^2}$, ovvero la funzione $\frac{1}{G}$ è una gaussiana, si può osservare come il nucleo riproducente K generi le funzioni di base radiali gaussiane.[36]

Risorse didattiche

Ecco qualche link utile per reperire risorse didattiche sull'argomento (e più in generale su metodi matematici), e sugli aspetti applicativi (problemi di apprendimento) maggiormente inerenti il corso:

Metodi matematici

http://www.math.utexas.edu/users/caffarel/m383.html
http://www.math.utexas.edu/users/caffarel/arbogast-bona-notes.ps
http://www.maths.unsw.edu.au/ForStudents/courses/math5115/lect.shtml
http://ejde.math.swt.edu/ /Monographs/01/abstr.html
http://gershwin.ens.fr/vdaniel/Doc-Locale/Cours-Mirrored
/Methodes-Maths/white/math/s0/outl539/outl539.html

Apprendimento

http://www.kernel-machines.org/tutorial.html
http://www-2.cs.cmu.edu/ tom/702.html

[35] *shift-invariant*
[36] *gaussian radial basis functions*