

Capitolo 3

L'algoritmo regolarizzante del gradiente coniugato

3.1 Introduzione

Si considerino nuovamente X , Y spazi di Hilbert e $T \in \mathcal{B}(X, Y)$.
La determinazione di soluzioni ai minimi quadrati dell'equazione

$$Tx = y \quad (3.1)$$

può essere realizzata per mezzo della ricerca di punti di minimo del funzionale $f : X \rightarrow \mathbb{K}$

$$f(x) = \|Tx - y\|$$

In tal modo la teoria sviluppata per la minimizzazione di funzionali offre nuovi strumenti per la risoluzione delle equazioni (3.1).

In questo capitolo, dopo una breve illustrazione del metodo della massima discesa, verrà analizzato l'algoritmo del gradiente coniugato, uno dei metodi di minimizzazione tra i più utilizzati.

Seguendo un percorso proposto da W. J. Kammerer e M. Z. Nashed, verrà innanzitutto dimostrata la convergenza dell'algoritmo per operatori con range arbitrario.

Il risultato di Kammerer e Nashed è di grande importanza poiché, ricordando che gli operatori con range non chiuso conducono a problemi mal posti, colloca il gradiente coniugato nel contesto dei metodi per la risoluzione di problemi inversi.

Studi successivi hanno infatti dimostrato che l'algoritmo del gradiente coniugato appartiene alla classe dei metodi iterativi di regolarizzazione, come verrà illustrato nel terzo paragrafo. Più precisamente verrà esaminata una recente dimostrazione di M. Hanke del famoso teorema di Nemirovski, basata sulle proprietà di particolari polinomi associati all'algoritmo (cfr. [35], [17]) e successivamente verrà dimostrata l'appartenenza del metodo alla classe degli algoritmi regolarizzanti, come proposto da R. Plato (cfr. [38]).

Si osservi che l'algoritmo è effettivamente utilizzato per la regolarizzazione di problemi inversi, spesso associato a tecniche di preconditionamento (un settore in cui l'algoritmo ha visto un notevole sviluppo è l'elaborazione di immagini; a tal proposito si consulti [7], [33]).

3.2 Il metodo del gradiente (o della massima discesa)

Il metodo della massima discesa (steepest descent method) è il più antico metodo di minimizzazione di funzioni; per la sua semplicità, è stato per molto tempo il più utilizzato. Sebbene già a metà del secolo scorso Chanchy ne avesse intuito il funzionamento, il metodo è stato formulato in maniera precisa alla fine degli anni 30 (si consulti [23], [43]). Negli anni 50-60 il metodo è stato generalizzato alla risoluzione di equazioni lineari tra spazi di Hilbert del tipo (1.1).

Prima di dedicarci agli spazi di Hilbert, vediamo rapidamente l'algoritmo nella formulazione più semplice.

Sia $f : E^n \rightarrow \mathbb{R}$ una funzione definita su un generico spazio vettoriale euclideo n -dimensionale. Si supponga inoltre che la funzione sia dotata di derivate parziali prime continue su E^n . Sappiamo che, fissato $x \in E^n$, il vettore gradiente

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

individua la direzione di più ripida salita della funzione f in x , cioè la direzione in cui la derivata direzionale è massima.

L'idea alla base del metodo del gradiente è di cercare valori di minimo relativo iterativamente, muovendosi ad ogni passo nella direzione opposta a $\nabla f(x)$. Utilizzando le comuni notazioni dell'algebra lineare, questo dà luogo al seguente algoritmo:

sia x_k la soluzione ottenuta alla k -esima iterata;

supponendo $\nabla f(x_k) \neq 0$, l'iterata successiva è definita da

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)^t$$

dove $\alpha_k \in \mathbb{R}$ è il valore non negativo che minimizza la funzione nella direzione individuata da $-\nabla f(x_k)$. Il procedimento termina non appena $\nabla f(x_k) = 0$.

L'algoritmo ora formulato non sembra rivestire particolare interesse poiché risulta molto costoso (ad ogni passo occorre calcolare n derivate direzionali e minimizzare la funzione lungo la direzione individuata). Il metodo diventa efficiente se applicato a problemi quadratici. Si consideri la funzione $f : E^n \rightarrow \mathbb{R}$ definita nel seguente modo

$$f(x) = \frac{1}{2} \|Tx - y\|_2^2$$

dove $\|\cdot\|_2$ rappresenta la norma $l^2(E^n)$, $T \in M_{n,n}(\mathbb{R})$ è una matrice invertibile e $y \in E^n$. Si ha

$$f(x) = \frac{1}{2} x^t A x - x^t T^t y + \frac{1}{2} \|y\|_2^2 = \frac{1}{2} x^t A x - x^t b + Cost$$

con $A = T^t T$ matrice $(n \times n)$ simmetrica e definita positiva e $b = T^t y = T^t y$. Un semplice calcolo diretto permette di determinare il gradiente nella forma seguente

$$\nabla f(x)^t = Ax - b$$

Il valore del gradiente è ora immediatamente disponibile in ogni punto. L'algoritmo si semplifica ulteriormente poiché non solo il gradiente, ma anche il valore scalare α_k può essere calcolato esplicitamente. Si consideri

$$f(x_k - \alpha \nabla f(x_k)^t) = \frac{1}{2} (x_k - \alpha \nabla f(x_k)^t)^t A (x_k - \alpha \nabla f(x_k)^t) - (x_k - \alpha \nabla f(x_k)^t)^t b + Cost$$

Differenziando rispetto a α si ottiene il valore di α_k che minimizza l'espressione:

$$-\nabla f(x_k)^t A x_k + \alpha_k \nabla f(x_k)^t A \nabla f(x_k)^t + \nabla f(x_k)^t b = 0$$

$$\alpha_k = \frac{\nabla f(x_k)^t (A x_k - b)}{\nabla f(x_k)^t A \nabla f(x_k)^t} = \frac{\nabla f(x_k)^t \nabla f(x_k)^t}{\nabla f(x_k)^t A \nabla f(x_k)^t}$$

Possiamo riscrivere l'algoritmo della massima discesa nella forma seguente

Sia $x_0 \in E^n$.
Se $\nabla f(x_k) \neq 0$ allora

$$x_{k+1} = x_k - \begin{pmatrix} r_k^t & r_k^t \\ r_k^t & A r_k \end{pmatrix} r_k \quad k = 0, 1, 2, \dots$$

dove

$$r_k = \nabla f(x_k)^t = Ax - b$$

Supponendo che l'algoritmo si arresti quando $\nabla f(x_k) = 0$, si osservi che x_k è l'unico valore x^* che risolve il sistema lineare

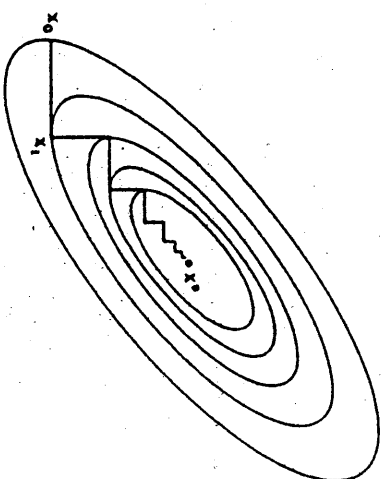
$$Ax = b$$

equivalente a $Tx = y$ (si osservi che T^t è invertibile, quindi $T^t T x = T^t y \Leftrightarrow Tx = y$). L'algoritmo della massima discesa si caratterizza come metodo iterativo per la risoluzione di sistemi lineari.

Il funzionamento dell'algoritmo può essere illustrato nel seguente modo.

Nel caso quadratico le linee di livello della funzione f sono ellissoidi concentrici n -dimensionali i cui assi sono nelle direzioni degli autovettori di A , quindi perpendicolari, e hanno lunghezza inversamente proporzionale ai rispettivi autovalori. Fissato un punto iniziale $x_0 \in E^n$, il gradiente è perpendicolare alla linea di livello $c = f(x_0)$. Lungo la direzione individuata dal gradiente, l'algoritmo determina il punto x_1 che risulta essere il punto in cui la retta $x_0 - \alpha r_0$, $\alpha > 0$ è tangente a una nuova linea di livello. Si ripete il procedimento partendo da x_1 e muovendosi lungo la direzione $r_1 = \nabla f(x_1)$. Si determina così il punto x_2 e si prosegue.

La situazione è illustrata in figura seguente.



La breve spiegazione svolta ha lo scopo di introdurre e chiarire in che modo e con quale semplice intuizione il metodo è stato sviluppato.

Il nostro interesse è però mirato alla formulazione del metodo per operatori lineari in spazi di Hilbert. Mostriamo che l'algoritmo della massima discesa permette di approssimare l'operatore inverso generalizzato (cfr. capitolo 1) e che quindi l'algoritmo può essere utilizzato per la risoluzione dell'equazione lineare (1.1).

Sia H uno spazio di Hilbert e $f: H \rightarrow \mathbb{R}$ un funzionale non negativo strettamente convesso.

Assumiamo f Fréchet-differenziabile in ogni punto $x \in H$ (cfr. Appendice C).

Analogamente al caso euclideo, lo scopo è minimizzare il funzionale, cioè determinare il punto $x^* \in H$ tale che

$$f(x^*) = \inf\{f(x) : x \in H\}$$

Partendo da un arbitrario punto iniziale $x_0 \in H$, cerchiamo di minimizzare f lungo la direzione di massima discesa.

Come dimostrato nell'appendice C, la derivata direzionale esiste in ogni punto e può essere espressa in funzione del gradiente nel seguente modo

$$Df(x_0, h) = (h, \nabla f(x_0))$$

dove $Df(x_0, h)$ rappresenta la derivata direzionale nel punto x_0 nella direzione individuata dal vettore $h \in H$.

Utilizzando la disuguaglianza di Schwartz $-\|h\| \|\nabla f(x_0)\| \leq (h, \nabla f(x_0))$, concludiamo che la derivata è minima per h multiplo positivo di $-\nabla f(x_0)$, ossia nell'unica direzione in cui la disuguaglianza di Schwartz vale come uguaglianza.

Questo risultato, ovviamente analogo a quello determinato nel caso euclideo, ci consente di estendere l'algoritmo ai funzionali lineari. Si ha il seguente schema iterativo:

Sia $x_0 \in H$;
Per $k = 0, 1, 2, \dots$, se $\nabla f(x_k) \neq 0$, allora

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

dove

$$\frac{d}{d\alpha} f(x_k - \alpha \nabla f(x_k)) \Big|_{\alpha=\alpha_k} = 0$$

altrimenti x_k è punto di minimo.

Questa formulazione per funzionali lineari permette di definire l'algoritmo per la risoluzione dell'equazione lineare (1.1), ossia dell'equazione

$$Tx = y \quad (3.2)$$

con X, Y spazi di Hilbert, $T \in B(X, Y)$, $x \in X$, $y \in Y$.

La soluzione dell'equazione lineare viene ricondotta alla minimizzazione del seguente funzionale quadratico

$$f(x) = \frac{1}{2} \|Tx - y\|^2$$

Dal teorema 1.1 sappiamo che le soluzioni ai minimi quadrati dell'equazione $Tx = y$ corrispondono ai punti di minimo del funzionale f :

applicando l'algoritmo della massima discesa al funzionale f si determina così una soluzione ai minimi quadrati.

Inoltre, come dimostrato nell'appendice C (cfr. lemma C.4),

$$\nabla f(x) = T^*Tx - T^*y$$

Questo mostra ulteriormente come i punti di stazionarietà determinati dal metodo corrispondano a soluzioni ai minimi quadrati $T^*Tx - T^*y = 0$ (cfr. teorema 1.1).

Analogamente a quanto fatto nel caso euclideo, il valore α_k che minimizza il funzionale nella direzione $-\nabla f(x_k)$ può essere determinato esplicitamente.

Ponendo $r = \nabla f(x)$ si ha

$$\begin{aligned} \frac{d}{d\alpha} f(x - \alpha r) &= \frac{d}{d\alpha} \frac{1}{2} \|Tx - \alpha Tr - y\|^2 \\ &= \frac{d}{d\alpha} \frac{1}{2} (Tx - \alpha Tr - y, Tx - \alpha Tr - y) \\ &= \frac{d}{d\alpha} \frac{1}{2} ((Tx - y, Tx - y) + \alpha^2 (Tr, Tr) - 2\alpha (Tx - y, Tr)) \\ &= \alpha \|Tr\|^2 - (Tr^*(Tx - y), r) \\ &= \alpha (\|Tr\|^2 - \|r\|^2) \end{aligned}$$

da cui

$$\frac{d}{d\alpha} f(x - \alpha r) \Big|_{\alpha=\bar{\alpha}} = 0 \quad \text{per} \quad \bar{\alpha} = \frac{\|r\|^2}{\|Tr\|^2}$$

Si ottiene così

Definizione 3.1 Metodo del gradiente per la risoluzione di equazioni lineari in spazi di Hilbert

Si consideri l'equazione (3.2) e sia $x_0 \in X$

Per $k = 0, 1, 2, \dots$, se $r_k = T^*Tx_k - T^*y \neq 0$ allora

$$x_{k+1} = x_k - \alpha_k r_k$$

dove

$$\alpha_k = \frac{\|r_k\|^2}{\|Tr_k\|^2}$$

Si osservi che nel corso delle iterazioni il denominatore di α_k non si annulla mai, infatti se $Tr_k = 0$ allora necessariamente $r_k = 0$ poiché $r_k \in R(T^*) = N(T)^\perp$; l'algoritmo è quindi ben definito.

Il seguente teorema, dovuto a Nashed (cfr. [34]), asserisce che il limite è un punto stazionario. Si osservi che non vi sono particolari ipotesi sull'operatore lineare.

Teorema 3.1 Se $T \in B(X, Y)$, allora

$$\lim_{k \rightarrow +\infty} r_k = 0$$

Dimostrazione.

$$\begin{aligned} f(x_{k+1}) &= \frac{1}{2} \|T(x_k - \alpha_k r_k) - y\|^2 \\ &= \frac{1}{2} \|Tx_k - \alpha_k Tr_k - y\|^2 \\ &= \frac{1}{2} (\|Tx_k - y\|^2 + \alpha_k^2 \|Tr_k\|^2 - 2\alpha_k (Tx_k - y, Tr_k)) \\ &= f(x_k) + \frac{1}{2} \alpha_k^2 (\|Tr_k\|^2 - \alpha_k \|r_k\|^2) \end{aligned}$$

$$= f(x_k) + \frac{1}{2} \frac{\|r_k\|^4}{\|T^* r_k\|^4} \|T^* r_k\|^2 - \frac{\|r_k\|^2}{\|T^* r_k\|^2} \|r_k\|^2$$

$$= f(x_k) - \frac{1}{2} \frac{\|r_k\|^4}{\|T^* r_k\|^2}$$

Usando questa formula ricorsivamente si ottiene il valore del funzionale al passo $(k+1)$ -esimo come differenza tra il valore iniziale e tutti i k decrementi fatti, ossia

$$f(x_{k+1}) = f(x_0) - \frac{1}{2} \sum_{i=0}^k \frac{\|r_i\|^4}{\|T^* r_i\|^2}$$

Poiché $f(x) \geq 0 \quad \forall x \in X$, possiamo concludere che $\|r_k\| \rightarrow 0$ infatti, considerando

$$\|T\|^{-2} \sum_{i=0}^{\infty} \|r_i\|^2 \leq \sum_{i=0}^{\infty} \frac{\|r_i\|^4}{\|T^* r_i\|^2} \leq 2f(x_0)$$

si ha che la serie delle norme al quadrato di r_i è convergente.

Si osservi che la generica iterata x_k contiene la componente di x_0 appartenente a $N(T)$ poiché $r_i \in R(T^*) \subset N(T)^\perp$.

Nel caso in cui l'operatore T^*T possiede range chiuso, allora $\forall x \in R(T^*)$, l'algoritmo converge con velocità almeno equivalente a quella di una successione geometrica di raggio $\frac{M-m}{M+m}$, cioè

$$\|x_k - T^* y - P_{N(T)} x_0\| \leq C \left[\frac{M-m}{M+m} \right]^k \quad (3.3)$$

dove m, M sono costanti positive tali che $m\|x\|^2 \leq (T^*T x, x) \leq M\|x\|^2 \quad \forall x \in R(T^*) = N(T)^\perp$ (per una dimostrazione si consulti [34]).

Si può fornire una semplice spiegazione di tale risultato nel caso euclideo visto all'inizio del paragrafo. In questo contesto la relazione (3.3) è stata formulata precedentemente da Kantorovich.

Qui m e M corrispondono al più piccolo e al più grande autovalore della matrice Q . Ponendo $s = \frac{M}{m}$, equivalente al valore del condizionamento della matrice Q , si ha $\frac{M-m}{M+m} = \frac{s-1}{s+1}$. Questo mostra immediatamente che la convergenza cresce al crescere di s .

Quindi migliore è il condizionamento della matrice (cioè più vicini sono tutti i suoi autovalori), migliore è la convergenza del metodo.

Da un punto di vista grafico, minore è l'eccentricità delle ellissoidi di livello $f(x) = Cost$, migliore è la convergenza. Nel caso limite in cui $s = 1$, le linee di livello diventano sferiche e l'algoritmo converge al primo passo (per maggiori dettagli si consulti [25]). Una dimostrazione della convergenza del metodo con raggio $(\frac{M-m}{M+m})^{\frac{1}{2}}$, si trova in [36] (questa dimostrazione è più semplice di quella di Kantorovich; si osservi però che il risultato è meno fine di quello di Kantorovich poiché il raggio considerato è più grande, infatti $(\frac{M-m}{M+m})^{\frac{1}{2}} > \frac{M-m}{M+m} > \frac{M-m}{M+m}$).

Kammerer e Nasched (cfr. [21]) hanno dimostrato la convergenza del metodo nel caso di operatori con range non chiuso, sotto l'ipotesi che $x_0 = 0$ e $P_{R(T)} y \in R(TT^*)$.

La dimostrazione, che non riportiamo, fa uso di strumenti che gli autori utilizzarono in seguito per un teorema analogo enunciato per il metodo del gradiente coniugato; quest'ultimo, con dimostrazione, si trova nel capitolo 3.3 (teorema 3.9).

Teorema 3.2 Se $P_{R(T)} y \in R(TT^*)$, allora la sequenza $\{x_k\}$ generata dal metodo della massima discesa (definizione 3.1) con vettore iniziale $x_0 = 0$, converge monotonamente alla inversa generalizzata $x^1 = T^1 y$.

Si ha

$$\|x_k - x^1\|^2 \leq \frac{\|T\|^2 \|x^1\|^2 \|(T^* T^*)^k y\|^2}{\|T\|^2 \|(T^* T^*)^k y\|^2 + k \|x^1\|^2}$$

Si osservi che nella disuguaglianza si hanno tutte costanti eccetto il valore k . Successivamente Mc.Cormick e Rodriguez (cfr. [28]) hanno esteso il risultato di convergenza a tutti gli $y \in D(T^1)$ e per ogni vettore iniziale x_0 .

Teorema 3.3 Se $y \in D(T^1) = R(T) \oplus R(T)^\perp$, allora, $\forall x_0 \in X$, l'algoritmo della massima discesa converge monotonamente a

$$T^1 y + P_{N(T)} x_0$$

Inoltre se $P_{R(T)} y \in R(TT^*)$ e $x_0 \in R(T^*) \oplus N(T)$ allora

$$\|x_k - T^1 y - P_{N(T)} x_0\|^2 \leq \frac{\|T\|^2 \|x_0 - z^*\|^2 \|x_0 - T^1 y - P_{N(T)} x_0\|^2}{\|T\|^2 \|x_0 - z^*\|^2 + k \|x_0 - T^1 y - P_{N(T)} x_0\|^2}$$

dove z_0 è l'unico vettore in $\overline{R(T)}$ tale che

$$T^* z_0 = (I_X - P_{N(T)}) x_0$$

e z^* è l'unico vettore in $\overline{R(T)}$ tale che

$$T^* z^* = T^1 y$$

Si osservi che in virtù delle ipotesi su x_0 e y , i vettori z_0 e z^* esistono infatti $(I_X - P_{N(T)}) x_0$ e $T^1 y$ appartengono a $\overline{R(T)}$.

Concludiamo questa breve analisi del metodo della massima discesa precisando che il metodo è ormai poco utilizzato a causa della sua ridotta velocità di convergenza. Il metodo del gradiente coniugato, che verrà analizzato nel resto del capitolo, ha caratteristiche di convergenza migliori.

3.3 Il metodo del gradiente coniugato

Il metodo del gradiente coniugato è un metodo di minimizzazione proposto nel 1952 da Hestenes e Stiefel (cfr. [18]) come algoritmo per la risoluzione di sistemi lineari di più variabili.

Analogamente a quanto fatto per il metodo della massima discesa, il gradiente coniugato è stato successivamente generalizzato alla risoluzione di equazioni lineari in spazi di Hilbert (cfr. [16]).

In letteratura il metodo viene presentato in diversi modi. Nel nostro contesto risulta utile considerare il gradiente coniugato come particolare algoritmo nella classe dei cosiddetti metodi delle direzioni coniugate. Vedremo in seguito come il gradiente coniugato appartenga anche alla famiglia dei metodi introdotti da Samarskii (cfr. [41]) e tra questi risulti essere quello con le migliori caratteristiche.

Introduciamo la nozione di direzione coniugata esaminando la formulazione del problema di minimizzazione in uno spazio euclideo n -dimensionale. Analogamente al paragrafo precedente, sia allora $f: E^n \rightarrow \mathbb{R}$ il funzionale definito da

$$f(x) = \frac{1}{2} \|Tx - y\|^2$$

dove T è una matrice invertibile. Abbiamo già visto che le linee di livello del funzionale sono ellissoidi del tipo $\frac{1}{2}x^T Ax - x^T b = Cost$ dove $A = T^T T$ e $b = T^T y$.

Al fine di esplicitare il centro delle ellissoidi, consideriamo la decomposizione $A = UDU^T$ dove U è matrice ortogonale e D diagonale definita positiva. Le linee di livello possono pertanto essere scritte come

$$\frac{1}{2}x^T UDU^T x - x^T U^T b = Cost, \text{ cioè}$$

$$\frac{1}{2}w^T Dw - w^T \bar{b} = Cost \quad \text{o analogamente}$$

$$\sum_{i=1}^n (\frac{1}{2}D)_{ii} (w_i)^2 - \bar{b}_i w_i = Cost$$

dove $w = U^T x$ e $\bar{b} = U^T b = U^T T^T y$.

Si vede così (cfr. [12]) che il centro delle ellissoidi è il punto individuato dal vettore

$$D^{-1}\bar{b} = D^{-1}U^T T^T y = U^* u$$

dove u è l'unica soluzione del sistema $Ax = b$.

Osservando che $\nabla f(w) = Dw - \bar{b}$, se v è un vettore tangente alla linea di livello nel punto w allora vale la seguente relazione

$$(\nabla f(w), v) = 0 \Leftrightarrow v^T (Dw - \bar{b}) = 0 \Leftrightarrow v^T D(w - D^{-1}\bar{b}) = 0$$

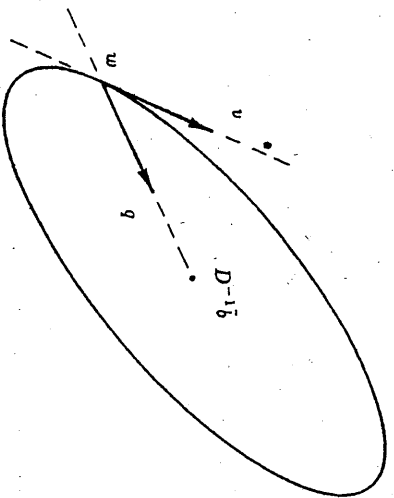
$$(w - D^{-1}\bar{b}, Dv) = 0$$

L'ultima relazione ci permette di introdurre il concetto di direzione coniugata.

Si osservi infatti che $q = -(w - D^{-1}\bar{b})$ è un vettore applicato nel punto w e diretto verso il centro $D^{-1}\bar{b}$ dell'ellissoide.

Quindi che in ogni punto $w \in E^n$, due vettori v e q , uno tangente alla linea di livello passante per w e l'altro diretto verso il centro della ellissoide, verificano la relazione

$$(q, Dv) = 0$$



Generalizzando questa proprietà si definiscono le direzioni coniugate.

Definizione 3.2 Data una matrice B simmetrica, due vettori p_1 e p_2 che verificano la relazione

$$p_1^T p_2 = 0$$

si dicono coniugati rispetto a B (o B -ortogonali).

Si osservi che nel caso in cui $B = I$, la B -ortogonalità corrisponde all'usuale nozione di ortogonalità.

Proposizione 3.1 A è definita positiva, ogni insieme di vettori mutuamente A -ortogonali risulta formato da vettori linearmente indipendenti.

Dimostrazione. Si consideri l'insieme $\{p_0, p_1, \dots, p_{n-1}\}$ formato da vettori non nulli mutuamente A -ortogonali, cioè tale che $p_i^T A p_j = 0$, $i \neq j$. Si supponga che $\beta_0 p_0 + \beta_1 p_1 + \dots + \beta_{n-1} p_{n-1} = 0$; dimostriamo che $\beta_i = 0$ $\forall i = 1, \dots, n-1$. Fissato i moltiplichiamo la combinazione lineare nulla per $p_i^T A$ e usiamo la A -ortogonalità. Otteniamo

$$\beta_i p_i^T A p_i = 0$$

Essendo A definita positiva l'ultima relazione implica $\beta_i = 0$. Ripetendo il ragionamento per ogni i si ottiene la tesi.

L'indipendenza dei vettori A -coniugati è la proprietà che consente di definire il seguente algoritmo per la minimizzazione di f , il quale determina la soluzione al più in n passi.

Teorema 3.4 Metodo delle direzioni coniugate per sistemi lineari

Sia A una matrice simmetrica definita positiva di dimensione $n \times n$ e sia $\{d_i\}_{i=0}^{n-1}$ una base di vettori A -ortogonali non nulli. Allora per ogni scelta del vettore iniziale $z_0 \in E^n$, la sequenza $\{z_k\}$ generata dall'algoritmo seguente

$$z_{k+1} = z_k + \alpha_k d_k \quad k = 0, 1, 2, \dots$$

dove

$$\alpha_k = \frac{r_k^T d_k}{d_k^T A d_k}$$

$$r_k = \nabla f(z_k) = A z_k - b$$

converge, dopo al più n passi, a x^* , unico punto di minimo del funzionale f , unica soluzione dell'equazione $Ax = b$.

Dimostrazione. Sfruttando l'indipendenza lineare dei vettori coniugati, possiamo scrivere:

$$x^* - z_0 = \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{n-1} d_{n-1}$$

per una particolare scelta dei α_i $i = 0, \dots, n-1$.

Moltiplicando per $d_i^T A$ si ottiene

$$\alpha_i = \frac{d_i^T A (x^* - z_0)}{d_i^T A d_i}$$

Esplicitando tutte le iterate, fino alla i -esima, si ha

$$z_i - z_0 = \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{i-1} d_{i-1}$$

Dimostriamo che gli α_i determinati dall'algoritmo, corrispondono ai coefficienti α_i di $(x^* - z_0)$.

Osservando che $d_i^T A(x_i - x_0) = 0$, si ha

$$\begin{aligned} \alpha_i &= \frac{d_i^T A(x^* - x_0)}{d_i^T A d_i} \\ &= \frac{d_i^T A(x^* - x_i)}{d_i^T A d_i} + \frac{d_i^T A(x_i - x_0)}{d_i^T A d_i} \\ &= -\frac{d_i^T r_i}{d_i^T A d_i} = \alpha_i \end{aligned}$$

L'algoritmo presentato ha una importante particolarità:

l'iterata k -esima minimizza il funzionale sulla varietà generata da $\{d_0, d_1, \dots, d_{k-1}\} \cup \{x_0\}$.

Questa proprietà, che verrà studiata in seguito nel contesto degli spazi di Hilbert, mostra che il metodo è monotonicamente convergente, ossia $f(x_{k+1}) < f(x_k)$.

Si sottolinea che non è stato presentato alcun procedimento per la determinazione di un generico insieme di vettori A -ortogonali. Vedremo infatti che il metodo del gradiente coniugato determina automaticamente la base $\{d_0, \dots, d_{n-1}\}$. Si osservi comunque che utilizzando il prodotto scalare

$$\langle \bullet, \bullet \rangle_A = (\bullet, A\bullet)$$

è possibile ottenere basi A -ortogonali semplicemente utilizzando l'algoritmo di ortogonalizzazione di Graham-Schmidt. Un metodo analogo è stato formulato da Hestenes (cfr. [19]).

Generalizziamo quanto fatto finora alla risoluzione di equazioni lineari in spazi di Hilbert. Siano X, Y spazi di Hilbert e $f: X \rightarrow \mathbb{R}$ il funzionale definito nel seguente modo

$$f(x) = \frac{1}{2} \|Tx - y\|^2 \quad (3.4)$$

con $T: X \rightarrow Y$ operatore lineare invertibile.

Definizione 3.3 $p_1, p_2 \in X$ si dicono coniugati rispetto all'operatore autoaggiunto

$C: X \rightarrow X$, 0 C -coniugati, se

$$(p_1, Cp_2) = 0$$

Ripercorrendo passaggi analoghi a quanto fatto nella proposizione 3.1, si dimostra che, se C è definito positivo, un insieme di elementi C -coniugati risulta essere formato da vettori linearmente indipendenti.

Analogamente a quanto fatto nel caso matriciale, per praticità notazionale spesso indicheremo con A l'operatore T^*T ossia

$$A = T^*T$$

Applicando la definizione 3.3 all'operatore A , consideriamo l'algoritmo delle direzioni coniugate per la minimizzazione del funzionale (3.4), ossia per la risoluzione dell'equazione invertibile

$$T^*Tx = T^*y$$

(cfr. teorema 1.1).

Come fatto per gli spazi euclidei, consideriamo l'algoritmo generato dal seguente procedimento iterativo

$$z_{n+1} = z_n + \alpha_n d_n$$

dove $\{d_i\}$ è formato da vettori A -coniugati e α_n è scelto in maniera ottimale, cioè

$$\frac{d}{d\alpha} f(z_n + \alpha d_n) \Big|_{\alpha=\alpha_n} = 0$$

Un calcolo analogo a quello svolto nel paragrafo precedente per il metodo della massima discesa fornisce il valore

$$\alpha_n = -\frac{(d_n, r_n)}{(d_n, Ad_n)}$$

con $r_n = Az_n - b$.

Posiamo ora definire l'algoritmo delle direzioni coniugate in spazi di Hilbert.

Definizione 3.4 Metodo delle direzioni coniugate in spazi di Hilbert

Usando le notazioni introdotte, $A = T^*T$, $b = T^*y$,

$\{d_n\}$ insieme di vettori A -coniugati, l'algoritmo iterativo definito nel seguente modo $z_0 \in X$

$$z_{n+1} = z_n + \alpha_n d_n \quad (3.5)$$

con

$$\alpha_n = -\frac{(d_n, r_n)}{(d_n, Ad_n)} \quad (3.6)$$

$$r_n = Az_n - b \quad (3.7)$$

viene detto metodo delle direzioni coniugate per la minimizzazione del funzionale $f(x) = \frac{1}{2} \|Tx - y\|^2$.

Il seguente risultato, già citato precedentemente, dà indicazioni sulla qualità del metodo.

Teorema 3.5 Se T è operatore invertibile, allora il vettore z_k determinato dal metodo delle direzioni coniugate, minimizza il funzionale f sull'insieme

$$B_k = \{z_0 + z : z \in \text{span}\{d_0, d_1, \dots, d_{k-1}\}\}$$

Prima di procedere con la dimostrazione di quest'ultimo teorema, enunciamo e verifichiamo una proprietà delle direzioni coniugate che verrà utile in seguito.

Lemma 3.1

$$(d_i, r_j) = 0 \quad \text{per } i < j$$

Dimostrazione. (Lemma 3.1)

Procediamo per induzione.

Iniziamo, per $j = i + 1$, si ha $(d_i, r_{i+1}) = 0$, infatti, per (3.6)

$$\begin{aligned} (d_i, r_{i+1}) &= (d_i, Az_{i+1} - b) = (d_i, Az_i + \alpha_i Ad_i - b) \\ &= (d_i, r_i + \alpha_i Ad_i) = (d_i, r_i) + \alpha_i (d_i, Ad_i) = 0 \end{aligned}$$

Utilizzando questa relazione come ipotesi induttiva, basta dimostrare che, se $(d_i, r_j) = 0$ per un generico $j > i$, allora $(d_i, r_{j+1}) = 0$. Per la A -coniugatività dei $\{d_i\}$, si ha $(d_i, r_{j+1}) = (d_i, r_j) + \alpha_j(d_i, Ad_j) = \alpha_j(d_i, Ad_j) = 0$

Possiamo ora dimostrare il teorema precedente.

Dimostrazione. (teorema 3.5)

Verifichiamo che $f(x) > f(x_k) \quad \forall x \in B_k, x \neq x_k$.

Sappiamo che $x = x_0 + \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{k-1} d_{k-1}$ per un'opportuna scelta dei coefficienti $\alpha_0, \dots, \alpha_{k-1}$. Si ottiene

$$\begin{aligned} (x - x_k, r_k) &= \left(x_0 + \sum_{i=0}^{k-1} \alpha_i d_i - \left(x_0 + \sum_{i=0}^{k-1} \alpha_i d_i \right), r_k \right) \\ &= \sum_{i=0}^{k-1} (\alpha_i - \alpha_i) (d_i, r_k) = 0 \end{aligned}$$

infatti, per il lemma appena enunciato, $(d_i, r_k) = 0 \quad i = 0, \dots, k-1$.
Si ha così

$$\begin{aligned} f(x) - f(x_k) &= \frac{1}{2} \|Tx - y\|^2 - \frac{1}{2} \|Tx_k - y\|^2 \\ &= \frac{1}{2} (Ax, x) - \frac{1}{2} (Ax_k, x_k) - (x, T^*y) + (x_k, T^*y) \\ &= \frac{1}{2} (Ax, x) - \frac{1}{2} (Ax_k, x_k) - (x - x_k, b) \\ &= \frac{1}{2} (Ax, x) - \frac{1}{2} (Ax_k, x_k) - (x - x_k, b - Ax_k + Ax_k) \\ &= \frac{1}{2} (Ax, x) - \frac{1}{2} (Ax_k, x_k) - (x - x_k, -r_k + Ax_k) \\ &= \frac{1}{2} (Ax, x) - \frac{1}{2} (Ax_k, x_k) - (x - x_k, Ax_k) \\ &= \frac{1}{2} ((Ax, x) + (Ax_k, x_k) - 2(x, Ax_k)) \\ &= \frac{1}{2} ((A(x - x_k), (x - x_k))) \end{aligned}$$

Poiché A è definita positiva, si ottiene

$$f(x) > f(x_k) \quad \forall x \neq x_k$$

Questo teorema conduce a importanti considerazioni.

La prima è che, sempre nel caso in cui T sia invertibile, $r_k \perp B_k$. Se così non fosse si avrebbe una componente di r_k appartenente a B_k e quindi x_k non sarebbe il punto di minimo su B_k (si osservi che sotto questa ipotesi su T , il funzionale f risulta strettamente convesso). Questo è consistente con il lemma 1.1, il quale asserisce proprio che $r_k \perp B_k$.

La seconda considerazione è che il metodo converge in n passi ogni volta che lo spazio X ha dimensione n , ossia ogni volta che

$$X = \text{span}\{d_0, d_1, \dots, d_{n-1}\}$$

3.3 - Il metodo del gradiente coniugato

V è un risultato più forte che è la generalizzazione di quest'ultimo.

Sia $X_k = \text{span}\{d_0, \dots, d_{k-1}\}$. Se lo spazio X , di dimensione infinita, è la chiusura dell'unione dei successivi X_k , ossia

$$X = \overline{\bigcup_{k=0}^{+\infty} X_k}$$

allora il metodo converge. Si osservi che questa condizione sullo spazio X non è particolarmente restrittiva; ad esempio ogni spazio di Hilbert separabile ammette tale rappresentazione (infatti, dalla definizione di spazio separabile, esiste un sottoinsieme denso di X numerabile, il quale sottintende ammette quindi basi numerabili).

Si ha il seguente risultato per operatori invertibili:

Teorema 3.6 Sia $T \in \mathcal{B}(X, Y)$ invertibile con inversa limitata.

Se

$$X = \overline{\bigcup_{n=0}^{+\infty} X_n}$$

con $X_n = \text{span}\{d_0, d_1, \dots, d_{n-1}\}$ e $\{d_i\}_{i=0}^{+\infty}$ insieme di elementi (T^*T) -coniugati, allora per ogni scelta del punto iniziale $x_0 \in X$, la successione $\{x_n\}$ determinata dal metodo delle direzioni coniugate (definizione 3.4) converge monotonamente al punto di minimo x_* del funzionale $f(x) = \frac{1}{2} \|Tx - y\|^2$, unica soluzione dell'equazione $Tx = y$.

Dimostrazione. Sebbene la scelta $x_0 = 0$ non sia in genere la migliore, possiamo assumere, senza perdere in generalità, che $x_0 = 0$.

La risoluzione dell'equazione $Ax = T^*y$, con $x_0 \neq 0$, è equivalente alla risoluzione dell'equazione $Ax' = T^*y - Ax_0$, con $x'_0 = 0$.

Sia allora $\{x'_i\}$ la successione determinata dal metodo delle direzioni coniugate applicato all'equazione

$$Ax' = T^*y - Ax_0$$

con $x'_0 = 0$.

Poiché X è chiuso e $X = \overline{\bigcup_{n=0}^{+\infty} X_n}$, allora esiste una successione $\{z_n\} \subset X$, con $z_n \in X_n$, tale che $\lim_{n \rightarrow +\infty} z_n = w$, dove w è l'unica soluzione dell'equazione

$$Aw = T^*y - Ax_0$$

Per il teorema 3.5, x'_n è punto di minimo sullo spazio X_n ; si ha così

$$f(z_n) \geq f(x'_n) \geq 0$$

e, per la continuità del funzionale f , allora $\lim_{n \rightarrow +\infty} f(z_n) = f(w)$.

Quindi

$$\lim_{n \rightarrow +\infty} f(x'_n) = f(w)$$

Per le ipotesi su T , A è invertibile e definito positivo; inoltre, poiché $x'_n \in X_n$ e $z_n \in X_n$, operando come nella dimostrazione del teorema 3.5, si ha

$$\begin{aligned} f(z_n) - f(x'_n) &= \frac{1}{2} (A(z_n - x'_n), z_n - x'_n) \\ &\geq \frac{m}{2} \|z_n - x'_n\| \end{aligned}$$

dove $A \geq mI$, $m > 0$

Ne segue che

$$w = \lim_{n \rightarrow +\infty} z_n = \lim_{n \rightarrow +\infty} (x'_n)$$

Poiché $Aw = Az - Az_0$, l'ultima uguaglianza implica che

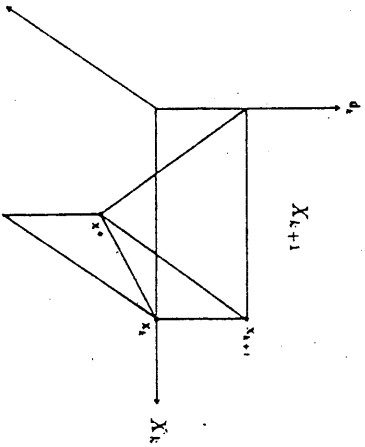
$$\lim_{n \rightarrow +\infty} x_n = x.$$

che conclude la dimostrazione.

Considerando il seguente prodotto scalare indotto dall'operatore A ,

$$\langle \bullet, \bullet \rangle_A = (A\bullet, \bullet)$$

la coniugatività può essere vista come ortogonalità rispetto a $\langle \bullet, \bullet \rangle_A$. Supponendo per semplicità $x_0 = 0$, possiamo rappresentare la iterata k -esima nel seguente modo



dove l'ortogonalità va intesa rispetto a $\langle \bullet, \bullet \rangle_A$.

Va sottolineato che le direzioni coniugate del metodo (3.5) vanno determinate prima di effettuare le iterazioni. Ciò significa che l'algoritmo, applicato a equazioni diverse, considera la stessa sequenza di direzioni di minimizzazione. Questa caratteristica mostra che l'algoritmo non si modella alla particolare equazione considerata, ma è invece "rigido", ossia ha troppi parametri già fissati a priori.

Consideriamo il metodo della massima discesa: sappiamo che esso determina le successive direzioni di minimizzazione al procedere dell'algoritmo. Abbiamo visto infatti che ogni direzione corrisponde al gradiente dell'iterata precedente e quindi viene determinata passo dopo passo. Questo metodo, a differenza del (3.5), risente in maniera diretta delle proprietà del funzionale che si sta minimizzando (le iterare determinate applicando l'algoritmo a due equazioni diverse non hanno nulla in comune). Il lato negativo del metodo della massima discesa è però proprio questa sua forte dipendenza dalle caratteristiche del funzionale, le quali spesso portano a una lenta convergenza, causata da oscillazioni sempre più fitte man mano che l'iterata si avvicina al punto di minimo.

Il metodo del gradiente coniugato, che ci apprestiamo a introdurre, unisce le proprietà migliori di questi due algoritmi: la rapida convergenza (eventualmente finita) del metodo delle direzioni coniugate e la flessibilità del metodo della massima discesa.

Fissato un punto $x_0 \in X$, la direzione di massima discesa $-\nabla f(x_0) = -r_0$ è normale alla curva di livello $x(t)$, $t \in [0, 1]$ tale che $f(x(t)) = f(x_0)$ $t \in [0, 1]$. Infatti, differenziando, si ottiene

$$0 = \frac{df(x(t))}{dt} = \frac{df}{dx} = (\nabla f(x(t)), x'(t))$$

3.3 - Il metodo del gradiente coniugato

dove $x' = \frac{dx}{dt}$ è tangente alla curva. Essendo normale alla ellissoide di livello, in genere la direzione individuata dal gradiente non passa per il centro (eccetto per i punti individuati dalle intersezioni con gli assi dell'ellissoide).

L'idea che conduce al metodo del gradiente coniugato è di considerare, dopo aver determinato il punto x_1 per mezzo dell'algoritmo della massima discesa, la direzione d_1 che passa per il centro dell'ellisse e appartiene al piano individuato da r_0 e r_1 .

All'inizio del paragrafo è stato dimostrato che, fissato un punto, la retta tangente alla linea di livello e la retta che unisce il punto al centro della ellisse sono coniugate. La direzione d_1 soddisfa quindi alla relazione seguente

$$(x_1 - x_0, Ad_1) = 0 \quad \text{ossia} \quad (r_0, Ad_1) = 0$$

dove $d_1 = \beta r_1 + \beta_0 r_0$.

Questa relazione permette di determinare d_1 . Assumendo in tutta generalità $\beta = 1$, si ottiene il valore β_0 che determina la direzione coniugata su cui minimizzare $(r_0, Ad_1) = (r_0, A(r_1 + \beta_0 r_0)) = 0$ da cui

$$\beta_0 = -\frac{(r_0, Ar_1)}{(r_0, Ar_0)}$$

Minizzando sulla retta $x_1 + \alpha d_1$ e determinato il punto di minimo x_2 , la successiva direzione d_2 sarà quella passante per il centro dell'ellisse e appartenente al piano individuato dal gradiente r_2 nel punto e dalla direzione coniugata d_1 ; il tutto è riassunto dalle relazioni seguenti

$$d_2 = r_2 + \beta_1 d_1 \quad \text{e} \quad (d_1, Ad_2) = 0.$$

Iterando il procedimento si ottiene:

Definizione 3.5 Metodo del gradiente coniugato in spazi di Hilbert

Siano X, Y spazi di Hilbert, $T \in B(X, Y)$.

Il metodo del gradiente coniugato per la minimizzazione del funzionale $\frac{1}{2} \|Tx - y\|^2$ è definito dal seguente algoritmo:

Sia

$$\begin{aligned} x_0 &\in X \\ r_0 &= \nabla f(x_0) = T^*Tx_0 - T^*y \\ d_0 &= r_0 \end{aligned}$$

Si ponga $i = 0$.

*

Se $r_i \neq 0$ allora

$$\alpha_i = \frac{(r_i, d_i)}{\|T d_i\|^2} \tag{3.8}$$

$$x_{i+1} = x_i - \alpha_i d_i \tag{3.9}$$

$$r_{i+1} = T^*Tx_{i+1} - T^*y = r_i - \alpha_i T^*T d_i \tag{3.10}$$

$$\beta_i = -\frac{(r_{i+1}, T^*T d_i)}{\|T d_i\|^2} \tag{3.11}$$

$$d_{i+1} = r_{i+1} + \beta_i d_i \tag{3.12}$$

$i := i + 1$ e si ritorni a *

altrimenti x_i è punto di minimo.

Ogni successiva direzione di ricerca di minimo è combinazione lineare del gradiente nel punto e della direzione precedente: l'algoritmo, a commento di quanto già detto, può essere visto come una variante del metodo della massima discesa. Le direzioni d_k dipendono dalla particolare equazione risolta; infatti vengono determinate al procedere delle iterate, in corrispondenza delle caratteristiche del funzionale nel punto. Inoltre il metodo appartiene alla classe dei metodi delle direzioni coniugate, come dimostrato dal seguente lemma.

Lemma 3.2 Si consideri il metodo del gradiente coniugato (definizione 3.5).

Allora i vettori $\{d_k\}$ sono A -coniugati e i gradienti $\{\tau_k\}$ sono ortogonali, cioè

$$(i) \quad (d_i, Ad_j) = 0 \quad i \neq j$$

$$(ii) \quad (\tau_i, \tau_j) = 0 \quad i \neq j$$

Inoltre

$$(iii) \quad (\tau_i, d_k) = \|\tau_i\|^2$$

Dimostrazione. Dimostriamo (i) e (ii) simultaneamente per induzione. Si osservi che basta verificare le due relazioni per $i < j$, e successivamente utilizzare la commutatività del prodotto scalare e la simmetria dell'operatore $A = T^*T$.

Evidentemente i risultati valgono per $i = 0, j = 1$; infatti d_1 è scelto in modo da soddisfare (i); inoltre il primo passo dell'algoritmo equivale al primo passo del metodo della massima discesa, quindi soddisfa (ii).

Supponendo come ipotesi induttive

$$(d_i, Ad_j) = 0 \quad (\tau_i, \tau_j) = 0$$

$$\text{per } i < j \quad j \leq k$$

dimostriamo che si ha

$$(d_i, Ad_{k+1}) = 0 \quad (\tau_i, \tau_{k+1}) = 0$$

$$\text{per } i \leq k$$

Per $1 \leq i < k$, osservando che $\tau_i = d_i - \beta_{i-1}d_{i-1}$,

$$(\tau_i, \tau_{k+1}) = (\tau_i, \tau_k - \alpha_k Ad_k) = (\tau_i, \tau_k) - \alpha_k (\tau_i, Ad_k)$$

$$= (\tau_i, \tau_k) - \alpha_k (d_i, Ad_k) + \alpha_k \beta_{i-1} (d_{i-1}, Ad_k) = 0$$

poiché, per le ipotesi induttive, tutti i tre prodotti scalari sono nulli.

Per $1 \leq i = k$,

$$(\tau_i, \tau_{k+1}) = (\tau_k, \tau_{k+1}) = (\tau_k, \tau_k) - \alpha_k (d_k - \beta_{k-1}d_{k-1}, Ad_k)$$

$$= (\tau_k, \tau_k) - \alpha_k (d_k, Ad_k) = (\tau_k, \tau_k) - \frac{(\tau_k, d_k)}{(d_k, Ad_k)} (d_k, Ad_k) =$$

$$= -\beta_{k-1} (\tau_k, d_{k-1}) = -\beta_{k-1} (\tau_{k-1} - \alpha_{k-1} Ad_{k-1}, d_{k-1})$$

$$= -\beta_{k-1} [(\tau_{k-1}, d_{k-1}) - \alpha_{k-1} (Ad_{k-1}, d_{k-1})] = 0$$

per definizione di α_{k-1} ; (ii) è stata così dimostrata.

Si osservi che nell'ultima serie di uguaglianze è stata dimostrata anche la relazione (iii); infatti

$$(\tau_k, \tau_k) - \frac{(\tau_k, d_k)}{(d_k, Ad_k)} (d_k, Ad_k) = (\tau_k, \tau_k) - (\tau_k, d_k) = 0$$

Per quanto riguarda (i), considerando $i < k$, si ha

$$\begin{aligned} (d_i, Ad_{k+1}) &= (d_i, A\tau_{k+1}) + \beta_k (d_i, Ad_k) = (d_i, A\tau_{k+1}) \\ &= (Ad_i, \tau_{k+1}) = \frac{1}{\alpha_i} (\tau_i - \tau_{i+1}, \tau_{k+1}) \\ &= \frac{1}{\alpha_i} (\tau_i, \tau_{k+1}) - \frac{1}{\alpha_i} (\tau_{i+1}, \tau_{k+1}) = 0 \end{aligned}$$

per il punto (ii) appena dimostrato. Si osservi che necessariamente $\alpha_i \neq 0$, poiché, se per assurdo $\alpha_i = 0$, allora $\tau_i = 0$ e l'algoritmo sarebbe terminato al passo precedente. Considerando $i = k$

$$(d_k, Ad_{k+1}) = (d_k, Ad_k) + \beta_k (d_k, Ad_k) = 0$$

per definizione di β_k .

Analogamente a quanto fatto nel lemma precedente, si possono determinare altre utili relazioni tra i vettori τ_i, d_i e i coefficienti α_i, β_i ; si consulti [36] - capitolo V.3, lemma (6).

Corollario 3.1 Il metodo del gradiente coniugato (definizione 3.5) appartiene alla classe dei metodi alle direzioni coniugate (definizione 3.4).

Dal teorema 3.6 si ottiene immediatamente il seguente

Teorema 3.7 Sia $T \in B(X, Y)$ invertibile con inversa limitata.

Se

$$X = \bigcup_{n=0}^{+\infty} X_n$$

dove $X_n = \text{span}\{d_0, \dots, d_{n-1}\}$ e $\{d_k\}_{k=1}^{+\infty}$ insieme di elementi (T^*T) -coniugati, allora per ogni scelta del punto iniziale $x_0 \in X$, la successione $\{x_n\}$ determinata dal metodo del gradiente coniugato (definizione 3.5) converge monotonamente al punto di minimo del funzionale $f(x) = \frac{1}{2} \|Tx - y\|^2$, unica soluzione dell'equazione $Tx = y$.

Si richiama l'attenzione all'ipotesi di invertibilità di T : i risultati fin qui ottenuti valgono solo sotto questa ipotesi molto restrittiva.

Vedremo in seguito che, utilizzando questo teorema, la convergenza verrà estesa ad operatori non invertibili il cui range sia però chiuso e, successivamente, ad operatori lineari limitati qualsiasi.

Analogamente a quanto fatto per il metodo della massima discesa, Daniel (cf. [9]) ha dimostrato che, se T è invertibile, il metodo del gradiente coniugato converge almeno come una progressione geometrica di raggio

$$\left(\frac{M-m}{M+m} \right)$$

dove m e M sono numeri reali positivi tali che $mI \leq T^*T \leq MI$. Più precisamente si ha

$$\|x_n - x_*\| \leq C \left(\frac{M-m}{M+m} \right)^n \quad (3.13)$$

dove $C = \frac{\|Tx_0 - y\|}{\sqrt{m}}$.

Seguendo il percorso proposto da Kammerer e Nasched (cfr. [22]), dimostriamo la convergenza del metodo per operatori con range chiuso. Si ricorda che in questo caso, come visto nel capitolo 2, la risoluzione dell'equazione lineare non presenta grandi difficoltà; infatti il problema è ben posto nel senso della soluzione generalizzata. Si tratta di considerare l'operatore ristretto unicamente allo spazio $N(T)^\perp$ e di trascurare le componenti appartenenti al nucleo. L'operatore $T|_{N(T)^\perp}$ risulta così invertibile ed è possibile ricondursi al caso precedente (teorema 3.7).

Si osservi che questa procedura è comunemente utilizzata per estendere ad operatori non invertibili con $R(T)$ chiuso, risultati ottenuti per operatori invertibili.

Teorema 3.8 Si supponga X soddisfacente alle ipotesi del teorema 3.7 e $T \in \mathcal{B}(X, Y)$ operatore dotato di range chiuso.

Allora il metodo del gradiente coniugato converge monotonicamente alla soluzione ai minimi quadrati:

$$u = T^+b + (I - P_H)x_0$$

dell'equazione $Tx = y$, dove P_H è la proiezione ortogonale di X sul sottospazio $H = N(T)^\perp$.

Dimostrazione. Si osservi che, poiché $R(T)$ è chiuso, anche $R(T^*)$ risulta chiuso (si consulti [15]). Si ha quindi $R(T^*) = N(T)^\perp = H$.

Dalla definizione del metodo è immediato verificare che $\{r_i\} \subset H$ e $\{d_i\} \subset H$, infatti $r_i = T^*(u)$ e $d_i = T^*(w)$ per opportuni $u \in Y$ e $w \in Y$.

Si ha così che $\{x_i\} \subset z_0 + H$.
Si consideri l'operatore $R : z_0 + H \rightarrow H$ definito come la restrizione di P_H sull'insieme $z_0 + H$, cioè $R = P_H|_{z_0+H}$. R è evidentemente una isometria su H .

Come già osservato, l'operatore $S : N(T)^\perp \rightarrow R(T)$ definito come $S = T|_{N(T)^\perp}$ è operatore invertibile con inversa limitata (si consulti [15]).

Applicando l'algoritmo all'operatore S , con punto iniziale $z_0' = Rz_0$, si ottiene la sequenza $\{x_n'\}$, che risulta in relazione con $\{x_n\}$. Si ha

$$x_n' = Rx_n$$

poiché, considerando S al posto di T , si è solamente escluso l'insieme $N(T)^\perp$, il quale, come già osservato, interviene nelle iterate x_i solo attraverso l'elemento z_0 . Per il teorema 3.7 applicato a S , il che è possibile per le ipotesi su X , si ottiene che $\{x_n'\}$ converge monotonicamente all'elemento $v' \in H$, unica soluzione dell'equazione

$$Sx = Qy$$

con Q proiezione ortogonale di Y su $R(T)$.

Si osservi che $v' = T^+y$ (cfr. capitolo 1).

Applicando l'isometria inversa R^{-1} si ha

$$R^{-1}v' = R^{-1}T^+y = \hat{T}^+y + (I - P_H)x_0 = u$$

$$R^{-1}x_n' = x_n' + (I - P_H)x_0 = x_n$$

Le ultime due relazioni e la convergenza monotona di $\{x_n'\}$ permettono di concludere che la sequenza $\{x_n\}$ converge monotonicamente all'elemento u .

Si osservi che allo stesso modo si può estendere il risultato di Daniel (3.13). Applicando l'isometria R , si ottiene

$$\|x_n - u\| = \|x_n' - v'\| \leq C \left(\frac{M - m}{M + m} \right)^n \tag{3.14}$$

dove $C = \frac{\|Tx_0 - Qy\|}{\sqrt{m}}$; m e M numeri positivi tali che $mI \leq S^*S = T^*T|_H \leq MI$ con l'operatore identità I riferito al solo sottospazio H .

Vale così anche il seguente

Corollario 3.2 Nelle ipotesi del teorema precedente, se $R(T)$ ha dimensione k , allora il metodo converge al più in k passi.

Dimostrazione. La dimostrazione è immediata tenendo presente che $\{d_0, \dots, d_{k-1}\}$ è un insieme di vettori linearmente indipendenti che generano un sottospazio di dimensione k contenuto in $H = R(T^*) = N(T)^\perp$.

Poiché $\dim R(T) = k$, necessariamente $\dim H = k$, da cui segue la tesi.

Si osservi che il teorema 3.8 consente l'uso del gradiente coniugato a una grande classe di operatori, tra i quali vanno menzionati quelli a rango finito. Il gradiente coniugato si caratterizza come metodo per la determinazione della soluzione generalizzata di sistemi lineari singolari.

Sappiamo che le equazioni (3.2) definite per mezzo di operatori dotati di range chiuso, conducono a problemi ben posti; quindi non abbiamo ancora alcuna indicazione sulle qualità del metodo in presenza di problemi mal posti. Vedremo nel prossimo paragrafo come il metodo del gradiente coniugato si collochi nel contesto dei metodi di regolarizzazione.

Introduciamo per ora un risultato, pubblicato insieme al precedente da Kammerer e Nasched (cfr. [22]), di notevole importanza poiché è stato la base di tutti gli sviluppi successivi, culminati nei risultati di Nemirovskii e Plato che analizzeremo nel prossimo paragrafo. Basso afferma che il gradiente coniugato converge monotonicamente alla soluzione generalizzata anche nel caso in cui l'operatore abbia range non chiuso e conduca quindi ad un problema mal posto. Si osservi che in questo paragrafo non si farà alcun riferimento alla presenza di dati perturbati, argomento centrale nella teoria della regolarizzazione. Il risultato è comunque significativo poiché afferma che in presenza di dati esatti, e quindi almeno a livello teorico, il gradiente coniugato ha un buon funzionamento.

La convergenza è dimostrata sotto ipotesi particolari sul dato y e sul punto iniziale z_0 .

Per il dato, si richiede che

$$Qy \in R(TT^*T) \tag{3.15}$$

dove Q è la proiezione ortogonale di Y su $R(T)$.

Si osservi però che per ogni operatore limitato K si ha $\overline{R(K)} = \overline{R(KK^*)}$.

Questo, applicato a TT^* , fornisce $R(TT^*) = \overline{R(TT^*TT^*)}$; considerando che $R(TT^*) \subset R(TT^*T) \subset R(TT^*)$, si conlude che

$$\overline{R(T)} = \overline{R(TT^*)} = \overline{R(TT^*T)}$$

Questo mostra che l'insieme di appartenenza del dato y è un insieme denso nello spazio $D(T) = R(T) \oplus R(T)^\perp$, usuale dominio dell'inversa generalizzata.

L'ipotesi (3.15), necessaria nel corso della dimostrazione, è poco restrittiva e, nel caso di dati perturbati, non comporta grandi complicazioni (si può sempre approssimare a piacere il dato con elementi in $R(TT^*T)$).

Circa l'elemento iniziale, si richiede

$$z_0 \in R(T^*T) \tag{3.16}$$

Questa condizione è comunque rispettata assumendo $x_0 = 0$.
Si osservi che $R(T^*T) \subset N(T)^\perp$ ed è ivi denso.

Prima di fornire il teorema, riportiamo alcuni lemmi, anche questi proposti da Kammerer e Nashed, necessari per la dimostrazione.

Lemma 3.3 Con riferimento alle notazioni utilizzate nella definizione 3.5 di gradiente coniugato, si ha che:

- (a) $\|T d_k\| \leq \|T r_k\|$
 (b) $\beta_k = \frac{\|r_{k+1}\|^2}{\|r_k\|^2}$
 (c) $(d_k, d_k) = \|r_k\|^2 \frac{\|d_k\|^2}{\|r_k\|^2}$
 (d) $d_k = \|r_k\|^2 \sum_{j=0}^k \frac{r_j}{\|r_j\|^2}$
 (e) $\|r_k\| \leq \|d_k\|$

La verifica delle relazioni precedenti non è difficile; essa poggia sulla definizione per ricorrenza di r_k e d_k , e sulle loro rispettive proprietà di ortogonalità e di coniugatività annunciate nel lemma 3.2 (per la dimostrazione si consulti [19], [18]).

Abbiamo visto che la risoluzione dell'equazione $Tx = y$, viene ricondotta alla minimizzazione del funzionale $f(x) = \frac{1}{2} \|Tx - y\|^2$. D'ora in poi verrà trascurato il coefficiente $\frac{1}{2}$, il quale ha solo lo scopo di rendere la relazione $\nabla f(x) = 0$ equivalente all'equazione (3.2).

Inoltre poniamo:

$$\begin{aligned} f(x) &= \|Tx - y\|^2 & (3.17) \\ g(x) &= \|Tx - Qy\|^2 & (3.18) \end{aligned}$$

Si osservi che

$$\begin{aligned} f(x) &= \|Tx - y\|^2 = \|Tx - Qy - (y - Qy)\|^2 \\ &= \|Tx - Qy\|^2 - 2(Tx - Qy, y - Qy) + \|y - Qy\|^2 \\ &= g(x) + \|y - Qy\|^2 = g(x) + \text{Cost} \end{aligned}$$

infatti $Tx - Qy \in \overline{R(T)}$ e $y - Qy \in R(T)^\perp$.

Ne segue che minimizzare f equivale a minimizzare g .

Definiamo il vettore errore

$$e_k = x_k - u \quad (3.19)$$

dove

$$u = T^{-1}y + (I - P_{N(T)^\perp})x_0 \quad (3.20)$$

è la soluzione ai minimi quadrati a cui è stato dimostrato converge il metodo nel caso di operatore con range chiuso.

Le relazioni annunciate nei prossimi lemmi vengono dimostrate considerando e_k come definito in (3.19); si osservi però che nel teorema di Kammerer e Nashed, per l'ipotesi (3.16), l'errore (3.19) si semplifica in $e_k = x_k - x^1$ poiché $u = x^1$.

Lemma 3.4 $\forall i, j = 0, 1, 2, \dots$

- (i) $g(x_i) = (r_i, e_i)$
 (ii) $g(x_i) - g(x_{i+1}) = \alpha_i \|r_i\|^2$
 (iii) $g(x_i) = (r_i, e_k) = (e_i, r_k) \quad k \leq i$
 (iv) $(d_i, e_k) \|r_i\|^2 = g(x_i) \|d_i\|^2$
 (v) (d_i, e_i) e (e_i, e_j) sono non negativi.

Dimostrazione. Ricordando che Q è la proiezione su $\overline{R(T)} = N(T)^\perp$

(i)

$$\begin{aligned} (r_i, e_i) &= (T^*T x_i - T^*y, x_i - u) = (T^*T x_i - T^*Qy, x_i - u) \\ &= (T x_i - Qy, T x_i - T u) = \|T x_i - Qy\|^2 = g(x_i) \end{aligned}$$

infatti, (cfr. teorema 1.1 (i)),

$$T u = T T^{-1}y + T(I - P_{N(T)^\perp})x_0 = P_{\overline{R(T)}}y + 0 = Qy$$

(ii) Utilizzando (i), il lemma 3.2 (iii), osservando che $r_{i+1} \perp \text{span}\{d_0, \dots, d_i\}$ (lemma 3.1) e che

$$T^*T e_i = T^*T x_i - T^*T u = T^*T x_i - T^*y = r_i, \quad (3.21)$$

possiamo scrivere:

$$\begin{aligned} g(x_i) - g(x_{i+1}) &= (r_i, e_i) - (r_{i+1}, e_{i+1}) = (r_i, e_i) - (r_{i+1}, x_i - \alpha_i d_i - u) \\ &= (r_i, e_i) - (r_{i+1}, e_i) + \alpha_i (r_{i+1}, d_i) = \alpha_i (T^*T d_i, e_i) + \alpha_i (r_{i+1}, d_i) \\ &= \alpha_i (d_i, T^*T e_i) + 0 = \alpha_i (d_i, r_i) = \alpha_i \|r_i\|^2 \end{aligned}$$

(iii) Per $k = i$ la tesi è già stata dimostrata in (i).

Per $k < i$, utilizzando (i), il lemma 3.1 e osservando che

$$e_i = x_i - u = x_i - \sum_{j=k}^{i-1} \alpha_j d_j - u = e_k - \sum_{j=k}^{i-1} \alpha_j d_j \quad (3.22)$$

si ottiene:

$$\begin{aligned} g(x_i) &= (r_i, e_i) = (r_i, e_k - \sum_{j=k}^{i-1} \alpha_j d_j) \\ &= (r_i, e_k) - \sum_{j=k}^{i-1} \alpha_j (r_i, d_j) = (r_i, e_k) \end{aligned}$$

Inoltre per (3.21),

$$(r_i, e_k) = (T^*T e_i, e_k) = (e_i, T^*T e_k) = (e_i, r_k)$$

(iv) Utilizzando (iii), il lemma 3.3 (d) e l'ortogonalità degli r_j ,

$$\begin{aligned} (d_i, e_i) \|r_i\|^2 &= \|r_i\|^4 \sum_{j=0}^i \frac{r_j}{\|r_j\|^2} (r_j, e_i) \\ &= g(x_i) \|r_i\|^4 \sum_{j=0}^i \|r_j\|^{-2} \\ &= g(x_i) \left(\|r_i\|^2 \sum_{l=0}^i \frac{r_l}{\|r_l\|^2}, \|r_i\|^2 \sum_{m=0}^i \frac{r_m}{\|r_m\|^2} \right) = g(x_i) (d_i, d_i) \end{aligned}$$

(v) $(d_i, e_i) \geq 0$ segue da (iv).

Per $(e_i, e_j) \geq 0$, la tesi è ovvia se $i = j$.

Supponiamo, senza perdere in generalità, $i < j$. Si osservi innanzitutto che, per il lemma 3.2 (iii)

$$\alpha_i = \frac{(r_i, d_i)}{\|r_i d_i\|^2} = \frac{\|r_i\|^2}{\|r_i d_i\|^2} \geq 0$$

Utilizzando (iii), il lemma 3.3 (d) e (3.22)

si ottiene:

$$\begin{aligned} (e_i, e_j) &= (e_i, e_i + \sum_{m=i+1}^{j-1} \alpha_m d_m) \\ &= (e_i, e_i) + \sum_{m=i+1}^{j-1} \alpha_m (e_i, d_m) \\ &= (e_i, e_i) + \sum_{m=i+1}^{j-1} \alpha_m \left(e_i, \|r_m\|^2 \sum_{l=0}^m \frac{r_l}{\|r_l\|^2} \right) \\ &= (e_i, e_i) + \sum_{m=i+1}^{j-1} \alpha_m \|r_m\|^2 \sum_{l=0}^m \frac{(e_i, r_l)}{\|r_l\|^2} \\ &= \|e_i\|^2 + \sum_{m=i+1}^{j-1} \alpha_m \|r_m\|^2 \left(\sum_{l=0}^{i-1} \frac{g(x_l)}{\|r_l\|^2} + \sum_{l=i}^m \frac{g(x_m)}{\|r_l\|^2} \right) \geq 0 \end{aligned}$$

Lemma 3.5

$$\|e_{i+1}\|^2 \leq \|e_i\|^2 - \alpha_i g(x_i)$$

Dimostrazione. Consideriamo i lemma 3.4 (iv), (ii) e 3.3 (e).

$$\begin{aligned} \|e_{i+1}\|^2 &= \|e_i - \alpha_i d_i\|^2 = \|e_i\|^2 - 2\alpha_i (e_i, d_i) + \alpha_i^2 \|d_i\|^2 \\ &= \|e_i\|^2 - 2\alpha_i g(x_i) \frac{\|d_i\|^2}{\|r_i\|^2} + \alpha_i^2 \|d_i\|^2 \\ &= \|e_i\|^2 - \alpha_i (2g(x_i) - \alpha_i \|r_i\|^2) \frac{\|d_i\|^2}{\|r_i\|^2} \\ &= \|e_i\|^2 - \alpha_i (g(x_i) + g(x_{i+1})) \frac{\|d_i\|^2}{\|r_i\|^2} \\ &\leq \|e_i\|^2 - \alpha_i g(x_i) \frac{\|d_i\|^2}{\|r_i\|^2} \\ &\leq \|e_i\|^2 - \alpha_i g(x_i) \end{aligned}$$

Prima di proseguire, analizziamo quali sono le proprietà caratteristiche delle iterate del metodo sotto le condizioni (3.15) e (3.16).

La condizione

$$Qy \in R(TT^*T)$$

implica $\bar{y} = Qy = TT^*\bar{z}$, dove $\bar{z} \in R(T)$. Ricordando che la soluzione generalizzata soddisfa all'equazione $Tx^1 = Qy$, ne segue che $Tx^1 = TT^*\bar{z}$. Sapendo che $x^1 \in N(T)^\perp$ e che $R(T^*) \subset N(T)^\perp$, si ha

$$x^1 = T^1 y = T^* \bar{z} \quad (3.23)$$

cioè la soluzione generalizzata appartiene a $R(T^*T) \subset R(T^*) \subset N(T)^\perp$.

Inoltre, poiché $\bar{z} \in R(T) \subset N(T^*)^\perp = N(TT^*)^\perp$ e $TT^*\bar{z} = Qy$ si ha $\bar{z} = (TT^*)^1 y$.

La condizione $z_0 \in R(T^*T)$ fa sì che $z_i \in R(T^*T)$ $\forall i$; infatti z_i è combinazione lineare di z_0 e $\{d_j\}_{j=0}^{i-1}$, mentre ogni d_j è combinazione lineare di $\{r_l\}_{l=0}^j$, i quali appartengono a $R(T^*T)$ (cfr. relazione (3.10)).

Esiste quindi una successione di vettori $\{z_i\}$, con $z_i \in R(T)$ $\forall i$ tale che

$$z_i = T^* z_i \quad (3.24)$$

In più, dato che $z_i, \bar{z} \in R(T)$, si ha

$$e_i = z_i - u = z_i - x^1 = T^* (z_i - \bar{z}) \in R(T^*T) \quad (3.25)$$

Naturalmente queste proprietà di regolarità di z_i e e_i (cioè l'appartenenza delle iterate a $R(T^*T)$, ecc.) non sarebbero garantite senza le ipotesi (3.15) e (3.16).

Nel caso in cui $R(T)$ è chiuso, abbiamo visto come sia stato facile generalizzare la convergenza dimostrata per operatori invertibili. Questo si è reso possibile poiché $T^*T|_{N(T)^\perp}$ è definito positivo, quindi invertibile con inversa limitata. Nel caso in cui $R(T)$ non è chiuso, questo non si verifica più, poiché $T|_{N(T)^\perp}$ non ha inversa limitata (cfr. teorema 1.5 (iii)).

Introduciamo allora il concetto di operatore aggiunto di operatore non limitato.

Indichiamo con $S: D(S) \rightarrow Y$ un generico operatore lineare definito su un sottospazio denso $D(S)$ di X .

Definiamo l'insieme

$$D(S^*) = \{ g \in Y : (Sf, g) = (f, h) \quad h \in X, \quad \forall f \in D(S) \}$$

$D(S^*)$ è sottospazio lineare di Y , infatti

$$(i) \quad w_1, w_2 \in D(S^*) \Rightarrow \exists h_1, h_2 \in X \quad \text{t.c.} \quad (Sf, w_1) = (f, h_1) \quad \text{e} \quad (Sf, w_2) = (f, h_2)$$

Allora

$$(Sf, w_1 + w_2) = (f, h_1 + h_2)$$

che implica $w_1 + w_2 \in D(S^*)$ con $h = h_1 + h_2$.

(ii) analogamente si dimostra che $\alpha w_1 \in D(S^*)$, $\alpha \in \mathbb{R}$

Definiamo allora l'operatore $S^* : D(S^*) \rightarrow X$, ponendo:

$$S^*g = h \quad \forall g \in D(S^*)$$

Si osservi che h è unico: infatti, se per assurdo esistessero h_1, h_2 tali che $(Sf, g) = (f, h_1)$ e $(Sf, g) = (f, h_2)$, allora si avrebbe $(f, h_1 - h_2) = 0 \quad \forall f \in D(S)$ denso in X . Questo implica $h_1 = h_2$.

S^* è lineare e risulta

$$(Sf, g) = (f, S^*g) \quad \forall f \in D(S), \quad \forall g \in D(S^*) \quad (3.26)$$

Si sottolinea che, nel caso in cui S sia limitato, questa definizione coincide con quella usuale di operatore aggiunto, con $D(S) = X$, $D(S^*) = Y$.

Si consideri l'operatore $W : \overline{R(T^*)} = N(T)^\perp \rightarrow R(T)$ definito da

$$W = T|_{\overline{R(T^*)}}$$

W è iniettivo con inversa generalmente non limitata.

Analogamente consideriamo $V : R(T) = N(T^*)^\perp \rightarrow R(T^*)$,

$$V = T^*|_{\overline{R(T)}}$$

con inversa non necessariamente limitata.

Applichiamo allora la definizione di operatore aggiunto (3.26) all'operatore $V^{-1} : R(T^*) \rightarrow R(T)$ con $R(T^*)$ denso nello spazio di Hilbert $R(T^*)$.

Se $x \in N(T)^\perp$ e $f \in R(T^*)$, cioè $f = T^*z$ con $z \in N(T^*)^\perp$, si ha

$$(V^{-1}f, Tx) = (V^{-1}T^*z, Tx) = (z, Tx)$$

e

$$(f, W^{-1}Tx) = (f, x) = (T^*z, x) = (z, Tx)$$

da cui

$$(V^{-1}f, g) = (f, W^{-1}g)$$

dove $g = Tx \in R(T)$.

Otteniamo così l'operatore V^{-1} , con dominio contenente $R(T)$, tale che

$$V^{-1}y = W^{-1}y \quad \forall y \in R(T)$$

Riassumendo abbiamo ottenuto che W e V , restrizioni rispettivamente di T e T^* a $N(T)^\perp$ e $N(T^*)^\perp$, soddisfano alla relazione

$$V^{-1} = W^{-1}$$

in $R(T)$.

Utilizzando questi operatori dimostriamo il lemma più importante, che sarà il centro, insieme al successivo, della dimostrazione di Kammerer e Nashed.

Lemma 3.6 Considerando le notazioni introdotte, se $Qy \in R(T^*T)$ e $z_0 \in R(T^*T)$, si ha

$$(i) \quad \|z_{k+1} - \bar{z}\| \leq \|z_k - \bar{z}\| \leq \|z_0 - \bar{z}\|$$

$$(ii) \quad \|e_{k+1}\|^2 \leq (1 - B\|e_k\|^2)\|e_k\|^2$$

$$\text{dove } B = B(z_0) = \|z_0 - \bar{z}\|^{-2} \|T\|^{-2}.$$

Dimostrazione. Come già accennato, si osservi innanzitutto che, per l'ipotesi su z_0 , la

$$(3.19) \text{ si caratterizza nel seguente modo } e_k = z_k - u = z_k - \bar{z}.$$

Da $z_{k+1} = z_k - \alpha_k d_k$, cioè $T^*z_{k+1} = T^*z_k - \alpha_k d_k$, con $z_k, z_{k+1} \in R(T)$ e $d_k \in R(T^*T) \subset R(T^*)$, si ha

$$z_k - z_{k+1} = \alpha_k V^{-1}d_k$$

Ne segue che

$$\begin{aligned} \|z_{k+1} - \bar{z}\|^2 &= \|z_k - \bar{z} - \alpha_k V^{-1}d_k\|^2 \\ &= \|z_k - \bar{z}\|^2 - 2\alpha_k (V^{-1}d_k, z_k - \bar{z}) + \alpha_k^2 \|V^{-1}d_k\|^2 \\ &= \|z_k - \bar{z}\|^2 - \alpha_k [2(V^{-1}d_k, z_k - \bar{z}) - (V^{-1}d_k, z_k - z_{k+1})] \\ &= \|z_k - \bar{z}\|^2 - \alpha_k (V^{-1}d_k, (z_k - \bar{z}) + (z_{k+1} - \bar{z})) \\ &= \|z_k - \bar{z}\|^2 - \alpha_k (V^{-1}d_k, V^{-1}e_k + V^{-1}e_{k+1}) \end{aligned}$$

Si osservi che l'ultima uguaglianza è valida poiché $z_k - \bar{z} \in R(T)$ e quindi

$$V^{-1}e_k = V^{-1}T^*(z_k - \bar{z}) = z_k - \bar{z}$$

Si ha così

$$\|z_{k+1} - \bar{z}\|^2 = \|z_k - \bar{z}\|^2 - \alpha_k (W^{-1}V^{-1}d_k, e_k + e_{k+1}) \quad (3.27)$$

dove V^{-1} coincide con W^{-1} (infatti $d_k \in R(T^*T)$ e quindi $V^{-1}d_k \in R(T)$).

Si osservi che, essendo $\alpha_k > 0$ (vedi definizione di α_k e lemma 3.2 (iii)), basta dimostrare che $(W^{-1}V^{-1}d_k, e_k + e_{k+1}) \geq 0$ per avere la tesi (i).

Dal lemma 3.3 (d) e 3.4 (v) segue

$$\begin{aligned} (W^{-1}V^{-1}d_k, e_k + e_{k+1}) &= \|r_k\|^2 \sum_{j=0}^k \frac{1}{\|r_j\|^2} (W^{-1}V^{-1}r_j, e_k + e_{k+1}) \\ &= \|r_k\|^2 \sum_{j=0}^k \frac{1}{\|r_j\|^2} (e_j, e_k + e_{k+1}) \geq 0 \end{aligned}$$

L'ultima uguaglianza

$$\begin{aligned} W^{-1}V^{-1}r_j &= W^{-1}V^{-1}(T^*Tx_j - T^*y) \\ &= W^{-1}V^{-1}T^*Tx_j - W^{-1}V^{-1}T^*y \\ &= W^{-1}V^{-1}T^*Tx_j - W^{-1}V^{-1}T^*Qy \\ &= W^{-1}Tx_j - W^{-1}Qy \\ &= W^{-1}Tx_j - W^{-1}Tx_j^1 \\ &= x_j - x_j^1 = e_j \end{aligned}$$

è valida poiché $x_j \in R(T^*T) \subset R(T^*)$ e $x_j^1 \in R(T^*T) \subset R(T^*)$.
È così dimostrato (i).

Usando le relazioni (i) appena dimostrata, verifichiamo (ii).
Si ha

$$\begin{aligned} \|e_i\|^4 &= |(x_i - x^1, x_i - x^1)|^2 = |(x_i - x^1, T^* x_i - T^* z)|^2 \\ &= |(T(x_i - x^1), z_i - z)|^2 \leq \|T(x_i - x^1)\|^2 \|z_i - z\|^2 \\ &\leq g(x_i) \|z_0 - z\|^2 \end{aligned}$$

Inoltre, considerando il lemma 3.5, segue

$$\begin{aligned} \|e_{i+1}\|^2 &\leq \|e_i\|^2 - \alpha_i g(x_i) \\ &\leq \|e_i\|^2 - \alpha_i \frac{\|e_i\|^4}{\|z_0 - z\|^2} \\ &\leq \|e_i\|^2 \left(1 - \frac{1}{\|T\|^2} \frac{\|e_i\|^2}{\|z_0 - z\|^2}\right) \\ &= \|e_i\|^2 (1 - B \|e_i\|^2) \end{aligned}$$

con

$$B = \|z_0 - z\|^{-2} \|T\|^{-2}$$

Si osservi che l'ultima maggiorazione è valida poiché, per il lemma 3.2 (iii) e 3.3 (a),

$$\begin{aligned} \alpha_i &= \frac{(r_i, d_i)}{\|T d_i\|^2} = \frac{\|r_i\|^2}{\|T d_i\|^2} \geq \frac{\|r_i\|^2}{\|T^* r_i\|^2} \\ &\geq \frac{\|r_i\|^2}{\|T\|^2 \|r_i\|^2} = \frac{1}{\|T\|^2} \end{aligned}$$

La dimostrazione è così completata. ■

Si osservi che i lemmi 3.4 e 3.5 sono stati enunciati per permettere la verifica del 3.6, che, come già detto, insieme al successivo completerà la dimostrazione della convergenza del metodo per operatori con range non chiuso.

Lemma 3.7 Sia $\{C_i\}_{i=0}^{+\infty} \subset \mathbb{R}$ una successione di numeri tale che $C_i \geq 0$ e

$$C_{i+1} \leq (1 - BC_i)C_i \quad i = 0, 1, 2, \dots \quad (3.28)$$

dove $B > 0$ e $0 < BC_0 \leq 1$.

Allora

$$C_i \leq \frac{C_0}{1 + iBC_0} \quad i = 0, 1, 2, \dots$$

Dimostrazione. Si osservi che, poiché $BC_0 > 0$ con $B > 0$, allora per (3.28) si ha $C_0 > 0$.

Possiamo considerare $C_i > 0 \quad \forall k$; infatti se $\exists i$ tale che $C_i = 0$, allora necessariamente $C_j = 0 \quad \forall j > i$ e la tesi risulta banalmente verificata.

Innanzitutto verifichiamo per induzione che

$$C_{i+1} < C_i \quad i = 0, 1, 2, \dots \quad (3.29)$$

Per $i = 0$ ciò è vero per l'ipotesi su C_0 , la quale implica $0 \leq 1 - BC_0 < 1$.

Allora, supposta la relazione vera per $i = k$, dimostriamo che essa vale anche per $i = k+1$.

Sappiamo dalla ipotesi (3.28) che $C_{k+1} \leq (1 - BC_k)C_k$. Basta allora dimostrare che $0 \leq BC_k < 1$.

Si ha, per ipotesi induttiva, $0 < BC_k \leq BC_{k-1} \leq BC_{k-2} \leq \dots \leq BC_0 \leq 1$, quindi $0 \leq 1 - BC_k < 1$, da cui segue che $C_{k+1} \leq (1 - BC_k)C_k < C_k$, che dimostra (3.29).

Utilizzando la relazione appena dimostrata e l'ipotesi (3.28), otteniamo

$$\begin{aligned} \frac{1}{C_{i+1}} - \frac{1}{C_i} &= \frac{C_i - C_{i+1}}{C_i C_{i+1}} \\ &\geq \frac{BC_i^2}{C_i C_{i+1}} > \frac{BC_i^2}{C_i^2} = B \end{aligned}$$

Sommando iterativamente si determina

$$\frac{1}{C_i} - \frac{1}{C_0} = \sum_{j=0}^{i-1} \left(\frac{1}{C_{j+1}} - \frac{1}{C_j} \right) > iB$$

cioè

$$\begin{aligned} \frac{C_0 - C_i}{C_i C_0} > iB &\Leftrightarrow C_0 - C_i > iBC_i C_0 \Leftrightarrow C_i(1 + iBC_0) < C_0 \\ &\Leftrightarrow C_i < \frac{C_0}{1 + iBC_0} \end{aligned}$$

Abbiamo tutti gli strumenti necessari per poter enunciare il teorema di convergenza di Kammerer e Nashed.

Teorema 3.9 Siano X e Y spazi di Hilbert, $T: X \rightarrow Y$ operatore lineare limitato.

Si consideri l'equazione lineare $Tx = y$.

Se $Qy \in R(T^*T)$, dove Q è l'operatore di proiezione ortogonale di Y su $\overline{R(T)}$, allora la successione di punti ottenuta col metodo del gradiente coniugato (definizione 1.5) con punto iniziale $z_0 \in R(T^*T) \subset R(T^*) = N(T)^\perp$, converge monotonamente alla soluzione generalizzata (cfr. capitolo 1)

$$z_* = T^1 y$$

Inoltre esiste una costante $B > 0$ tale che

$$\|z_i - z_*\|^2 \leq \frac{\|z_0 - z_*\|^2}{1 + iB\|z_0 - z_*\|^2} \quad i = 1, 2, \dots \quad (3.30)$$

Dimostrazione. Si osservi innanzitutto che, poiché $z_0 \in N(T)^\perp$, allora $z_* = T^1 y = T^1 y + (I - P_{N(T)^\perp})z_0 = u$, con u definito in (3.20).

Si confronti il risultato (ii) del lemma 3.6 e la ipotesi (3.28) del successivo 3.7; evidentemente ponendo

$$C_i = \|e_i\|^2, \quad B = \|z_0 - z\|^{-2} \|T\|^{-2}$$

se si verifica che $0 < BC_0 = B\|e_0\|^2 \leq 1$, la tesi del teorema segue immediatamente

dal lemma 3.7.

$$\begin{aligned} 0 < \|e_0\|^2 &= B\|x_0 - x_*\|^2 = B\|T^k x_0 - T^k z\|^2 \\ &\leq B\|T^k\|^2 \|z_0 - z\|^2 = B\|T^k\|^2 \|T\|^{-2} B^{-1} \\ &= \|T^k\|^2 \|T\|^{-2} = 1 \end{aligned}$$

Otteniamo così

$$\|e_i\|^2 \leq \frac{\|e_0\|^2}{1 + iB\|e_0\|^2} \quad i = 1, 2, \dots$$

Il risultato di Kammerer e Nashed è formulato dagli autori nella forma seguente

$$\|x_i - T^i y\| \leq \frac{\|T\|^2 \|x_0 - T^i y\| \|T^i x_0 - (T^i T^*)^i y\|^2}{\|T\|^2 \|T^i x_0 - (T^i T^*)^i y\|^2 + i \|x_0 - T^i y\|^2}$$

Questa, sebbene appaia più complicata, è equivalente a quella del teorema esposto qui. Per la verifica si osservi che, come mostrato precedentemente,

$$z_0 = T^*{}^1 x_0, \quad \bar{z} = (T^i T^*)^i y$$

da cui

$$B = \|T\|^{-2} \|z_0 - \bar{z}\|^{-2} = \|T\|^{-3} \|T^*{}^1 x_0 - (T^i T^*)^i y\|^{-2}$$

Abbiamo così terminato l'analisi della convergenza del metodo del gradiente coniugato.

Prima di procedere con le caratteristiche regolarizzanti del metodo, enunciamo alcune importanti proprietà.

Lemma 3.8 Sia $r_k \neq 0$. Allora

- (a) $\text{span}\{r_0, r_1, r_2, \dots, r_k\} = \text{span}\{r_0, (T^* T)^r r_0, (T^* T)^2 r_0, \dots, (T^* T)^k r_0\}$
 (b) $\text{span}\{d_0, d_1, d_2, \dots, d_k\} = \text{span}\{r_0, (T^* T)^r r_0, (T^* T)^2 r_0, \dots, (T^* T)^k r_0\}$

Dimostrazione. Dimostriamo per induzione simultaneamente (a) e (b).

Le due relazioni sono ovvie per $k = 0$.

Supponiamo allora che siano verificate per $k = j$ e dimostriamo che valgono per $k = j+1$.

Si consideri (3.10), ossia $r_{j+1} = r_j - \alpha_j T^* T d_j$

r_j e $T^* T d_j$ appartengono a $\text{span}\{r_0, (T^* T)^r r_0, \dots, (T^* T)^j r_0\}$.

Quindi $r_{j+1} \in \text{span}\{r_0, (T^* T)^r r_0, \dots, (T^* T)^{j+1} r_0\}$.

Inoltre, per il lemma 3.1, r_{j+1} è linearmente indipendente da $\text{span}\{r_0, (T^* T)^r r_0, \dots, (T^* T)^j r_0\}$, d_1, d_2, \dots, d_j , quindi $\text{span}\{r_0, r_1, r_2, \dots, r_{j+1}\} = \text{span}\{r_0, (T^* T)^r r_0, (T^* T)^2 r_0, \dots, (T^* T)^{j+1} r_0\}$ che verifica (a).

Per (ii), ricordando che $d_{j+1} = r_{j+1} + \beta_j d_j$, utilizzando la relazione (a) appena dimostrata e l'ipotesi induttiva, si ottiene immediatamente la tesi, ossia che $\text{span}\{d_0, d_1, d_2, \dots, d_{j+1}\} = \text{span}\{r_0, (T^* T)^r r_0, (T^* T)^2 r_0, \dots, (T^* T)^{j+1} r_0\}$ ■

Teorema 3.10 Sia $\bar{n} > 0$ e

$$H_{\bar{n}} = \left\{ y \in Y : P_{R(T)} y - T x_0 \text{ è una combinazione lineare di } \bar{n} \text{ autovettori di } T T^* \right\}$$

Allora se $y \in H_{\bar{n}}$, il metodo del gradiente coniugato (definizione 3.5) determina la soluzione al più in \bar{n} passi.

Dimostrazione. Dalla relazione (3.9) z_k è combinazione lineare di elementi appartenenti a $\{x_0\} \cup \{d_j\}_{j=0}^{k-1}$, ossia

$$z_k \in B_k = \{x_0 + z : z \in \text{span}\{d_0, \dots, d_{k-1}\}\}$$

Per il lemma precedente si ottiene

$$\begin{aligned} z_k &\in \{x_0 + \text{span}\{d_0, \dots, d_{k-1}\}\} \\ &= \{x_0 + \text{span}\{r_0, \dots, r_{k-1}\}\} \\ &= \{x_0 + \text{span}\{r_0, (T^* T)^r r_0, \dots, (T^* T)^{k-1} r_0\}\} \end{aligned}$$

Si osservi che

$$\begin{aligned} &\text{span}\{r_0, (T^* T)^r r_0, \dots, (T^* T)^{k-1} r_0\} \\ &= \text{span}\{(T^* T)^r x_0 - T^* y, (T^* T)^r (T^* T x_0 - T^* y), \dots, (T^* T)^{k-1} (T^* T x_0 - T^* y)\} \\ &= \text{span}\{(T^* T)^r x_0 - T^* y, (T^* T)^2 x_0 - (T^* T)^2 y, \dots, (T^* T)^k x_0 - (T^* T)^{k-1} T^* y\} \\ &= \text{span}\{T^* (T x_0 - Qy), T^* [(T T^*)^r (T x_0 - Qy)], \dots, T^* [(T T^*)^{k-1} (T x_0 - Qy)]\} \\ &= \text{span}\{T^* (T x_0 - Qy), T^* [(T T^*)^r (T x_0 - Qy)], \dots, T^* [(T T^*)^{k-1} (T x_0 - Qy)]\} \\ &= T^* \text{span}\{(T x_0 - Qy), [(T T^*)^r (T x_0 - Qy)], \dots, [(T T^*)^{k-1} (T x_0 - Qy)]\} \end{aligned}$$

con Q proiezione ortogonale di Y su $\overline{R(T)} = N(T^*)^\perp$.

Poiché per ipotesi, $(T x_0 - Qy)$ è combinazione lineare di \bar{n} autovettori di $T T^*$, si ha che $\forall k > 0$

$$(T x_0 - y), [(T T^*)^r (T x_0 - y)], \dots, [(T T^*)^{k-1} (T x_0 - y)]$$

sono anch'essi in $H_{\bar{n}}$.

Inoltre $T^* (H_{\bar{n}})$ è di dimensione \bar{n} ; infatti considerando $u, v \in H_{\bar{n}}$, autovettori di $T T^*$ tali che $\langle u, v \rangle = 0$, si ha

$$(T^* u, T^* v) = (T T^* u, v) = \lambda_u \langle u, v \rangle = 0.$$

Per il teorema 3.5, l'iterata z_k del gradiente coniugato, che appartiene alla classe dei metodi alle direzioni coniugate, minimizza il funzionale sullo spazio

$$B_k = \{x_0 + z : z \in \text{span}\{d_0, \dots, d_{k-1}\}\} \subset \{x_0 + T^* (H_{\bar{n}})\}$$

Quindi, poiché B_k ha dimensione al più \bar{n} , necessariamente $\exists k \leq \bar{n}$ tale che

$$z_k = u$$

e l'algoritmo si arresta in un numero finito di passi. ■

Teorema 3.11 Il metodo del gradiente coniugato (definizione 3.5) appartiene alla classe dei metodi di Samarskii (cfr. [41]), formata dagli algoritmi della forma

$$z_{n+1} = x_0 - P_n (T^* T)^r x_0$$

dove $P_n(\lambda)$ è un polinomio reale di grado al più n .

La dimostrazione può essere fatta per induzione, sulla base dei risultati del lemma 3.8, sostituendo iterativamente la formula (3.12): una semplice dimostrazione può comunque essere trovata in [36].

La caratteristica importante è che tra tutti gli algoritmi di Samarskii, cioè tra tutti gli algoritmi iterativi che determinano la soluzione esatta nello spazio

$$K^{(i)}(T^* T, r_0) = \text{span}\{r_0, (T^* T)^r r_0, (T^* T)^2 r_0, \dots, (T^* T)^{i-1} r_0\}$$

detto spazio di Krilov di ordine (i), il gradiente coniugato è quello che, ad ogni passo, approssima la soluzione nel modo migliore possibile.

Questo è diretta conseguenza diretta del teorema 3.5, il quale asserisce che l'iterata n-esima di ogni algoritmo delle direzioni coniugate minimizza il funzionale $f(x)$ sullo spazio

$$x_0 + \text{span}\{d_0, d_1, \dots, d_{i-1}\} = x_0 + K^{(i)}(T^*T, r_0)$$

Abbiamo così dimostrato il seguente teorema.

Teorema 3.12 Ottimalità della iterata del gradiente coniugato

Sia $x_i = x_0 - P_{i-1}(T^*T)r_0$ l'iterata n-esima dell'algoritmo del gradiente coniugato.

Allora

$$\|e_i\| = \|x_i - x_*\| = \min_{\tilde{P}_i} \|x_0 - \tilde{P}_i(T^*T)r_0 - x_*\|$$

dove il minimo è inteso rispetto a qualsiasi polinomio reale $\tilde{P}_i(\lambda)$ di grado i.

Inoltre Samarskii, servendosi della generalizzazione da lui introdotta, determina un ordine di convergenza migliore di (3.13). Più precisamente, in presenza di operatore lineare invertibile con inversa limitata, si ha

$$\|x_n - u\| \leq \tilde{C} \left(\frac{\sqrt{M} - \sqrt{m}}{\sqrt{M} + \sqrt{m}} \right)^n \quad (3.31)$$

con \tilde{C} costante opportuna. Si osservi che

$$\frac{\sqrt{M} - \sqrt{m}}{\sqrt{M} + \sqrt{m}} \leq \frac{M - m}{M + m}$$

dove l'uguaglianza vale solo per $m = M$, caso di scarso interesse.

3.4 Regolarizzazione mediante gradiente coniugato

In numerose applicazioni numeriche il metodo del gradiente coniugato è risultato un ottimo strumento per il trattamento di problemi inversi.

Come è stato visto nel paragrafo precedente, agli inizi degli anni 70 la convergenza del metodo è stata dimostrata anche per operatori con range non chiuso, i quali conducono a problemi mal posti (cfr. proposizione 2.1 e teorema 3.9).

L'analisi delle proprietà di regolarizzazione del metodo viene però sviluppata più tardi, a metà degli anni 80. Nel 1990 Robert Plato, utilizzando famosi risultati di Nemirovskii datati 1986, ha pubblicato un articolo dove dimostra che l'algoritmo, con una precisa regola d'arresto, è effettivamente un metodo di regolarizzazione. Si osservi che, sebbene venisse già utilizzato per il trattamento di problemi mal posti, l'appartenenza o meno del gradiente coniugato alla classe dei metodi di regolarizzazione è stato fino al 1990 un problema non risolto.

In questo paragrafo, utilizzando la teoria della risoluzione spettrale (vedi appendice B) e seguendo il percorso proposto nel 1994 da M. Hanke, affronteremo i risultati di Nemirovskii, fornendo tutte le definizioni e gli strumenti necessari; successivamente verrà illustrata la dimostrazione di Plato.

Sappiamo che la risoluzione dei problemi non ben posti dipende fortemente dalla precisione con cui si dispone del dato. Utilizzeremo dati a livello ϵ , cioè, come già visto nel capitolo 2, approssimazioni $y_\epsilon \in Y$ del dato $y \in R(T)$ tali che

$$\|y - y_\epsilon\| \leq \epsilon, \quad \epsilon > 0$$

Si consideri l'equazione (2.3) del paragrafo 2.3.

Come visto al paragrafo 2.3, un algoritmo regolarizzante è una famiglia di operatori che, in presenza di dati perturbati che tendono al dato $y \in R(T)$, restituisce soluzioni che tendono alla inversa generalizzata $T^\dagger y$ (la quale esiste poiché $y \in R(T) \oplus R(T)^\perp$, cfr. proposizione 1.1).

La definizione 2.2, considerando operatori lineari, risulta utile per la verifica delle proprietà della regolarizzazione enunciate al capitolo 2. Si osservi però che, come già accennato, la linearità non riveste carattere di necessità, ossia si possono definire famiglie di operatori regolarizzanti non lineari. Diamo quindi questa nuova definizione che generalizza la 2.2 (si consulti [38]).

Definizione 3.6 Sia $T \in B(X, Y)$. La famiglia di operatori $\{R_\epsilon\}_{\epsilon > 0}$, $R_\epsilon : Y \rightarrow X$, $\epsilon > 0$, si dice algoritmo regolarizzante per T , se

$$\sup_{y_\epsilon \in Y, \|T x_* - y_\epsilon\| \leq \epsilon} \|x_* - R_\epsilon y_\epsilon\| \rightarrow 0 \quad (\epsilon \rightarrow 0) \quad (3.32)$$

$$V x_* \in N(T)^\perp.$$

Si osservi che il parametro di regolarizzazione α del capitolo 2, sebbene non sia in relazione esplicita con la tolleranza ϵ , ha però lo stesso comportamento al limite, ossia $\alpha \rightarrow 0$ per $\epsilon \rightarrow 0$ (si veda il corollario 2.2).

La definizione 3.6 risulta quindi analoga alla 2.2, con la sostituzione $R_\epsilon = R_{\alpha(\epsilon)}$, dove $\alpha(\epsilon) \rightarrow 0$ ($\epsilon \rightarrow 0$).

Inoltre nel caso iterativo la definizione 3.6 va letta come

$$R_\epsilon y_\epsilon = \tilde{R}_{n(\epsilon, y_\epsilon)} y_\epsilon$$

dove \tilde{R}_n è l'operatore ottenuto iterando n volte e $n(\epsilon, y_\epsilon)$ è l'indice di arresto dell'algoritmo (si considerino le notazioni del capitolo 2.7).

Come già visto nel capitolo 2.7, la scelta del numero di iterazioni da effettuare è fondamentale poiché rappresenta il compromesso tra la approssimazione del dato e la regolarità della soluzione.

Il problema è ricondotto quindi alla determinazione della funzione

$$n = n(\epsilon, y_\epsilon) \quad \forall \epsilon > 0$$

$$y_\epsilon \in \{y \in Y : \exists x_* \in N(T)^\perp \text{ t.c. } \|T x_* - y\| < \epsilon\}$$

che chiameremo regola d'arresto.

Una delle scelte possibili è quella fornita dal principio della discrepanza (cfr. capitoli 2.5 e 2.7), che nel caso iterativo si caratterizza nel modo seguente

Definizione 3.7 Sia $\{x_n\}$ la sequenza ottenuta per mezzo di un metodo iterativo, con $x_0 = 0$.

Sia $\tau > 1$, una costante reale.

La regola d'arresto $n(\epsilon, y_\epsilon) = \bar{n}$, dove \bar{n} è il primo indice tale che

$$\|Tx_n - y_\epsilon\| \leq \tau \epsilon \quad (3.33)$$

viene detta principio della discrepanza.

Si osservi che il principio della discrepanza, applicato al metodo del gradiente coniugato, risulta ben definito. Infatti, poiché x_n converge a $T^1 y_\epsilon$ (cfr. teorema 3.8 e 3.9), $y \in R(T)$ e $\|y - y_\epsilon\| \leq \epsilon$, allora si ha

$$\begin{aligned} \|Tx_n - y_\epsilon\| &\longrightarrow \|T^1 y_\epsilon - y_\epsilon\| = \|P_{\overline{R(T)}} y_\epsilon - y_\epsilon\| \\ &= \text{dist}(\overline{R(T)}, y_\epsilon) \leq \epsilon < \tau \epsilon \end{aligned}$$

ossia il residuo è monotonicamente convergente ad un valore minore di $\tau \epsilon$.

La dimostrazione di Plato, come accennato, poggia su una proprietà del gradiente coniugato determinata e dimostrata da Nemirovskii (si consulti [35]) nel 1986. Come hanno osservato diversi ricercatori, la dimostrazione di Nemirovskii è molto complicata; analizzeremo una dimostrazione successiva, più semplice, proposta nel 1994 da Hanke (si consulti [17]).

Le seguenti definizioni permetteranno di enunciare il risultato di Nemirovskii.

Sia $M \subset X$ e $\epsilon > 0$.

Si definisce errore massimo $E_R(M, \epsilon)$ di un generico operatore $R: Y \rightarrow X$, rispetto al livello d'errore $\epsilon > 0$ e all'insieme M , il numero reale seguente

$$E_R(M, \epsilon) = \sup \{ \|x_* - Rx_*\| : x_* \in M, y_* \in Y, \|Tx_* - y_*\| < \epsilon \}$$

Considerando tutti i possibili operatori "di inversione", si definisce migliore errore massimo $E(M, \epsilon)$ rispetto a $\epsilon > 0$ e a M , il numero reale seguente

$$E(M, \epsilon) = \inf_{R: Y \rightarrow X} E_R(M, \epsilon)$$

Il risultato di Nemirovskii è riferito a soluzioni appartenenti a classi con particolari proprietà di regolarità.

Per descrivere tali classi consideriamo l'insieme $M = M_{\mu, \rho}$, $\mu > 0$, $\rho \geq 0$ dove

$$M_{\mu, \rho} = \{ (T^* T)^\mu z : z \in X, \|z\| \leq \rho \} \subset X \quad (3.34)$$

Come fa notare Hanke (cfr. [17]), al variare di μ , gli insiemi $R((T^* T)^\mu)$, essendo $R(T)$ non chiuso, formano una successione decrescente rispetto a μ ; le soluzioni corrispondenti a μ più grandi hanno maggiori proprietà di regolarità (euristicamente si pensi, per $\mu \rightarrow 0$, all'operatore identità e per $\mu \rightarrow +\infty$, con T compatto tale che $\|T\| = 1$, all'operatore di proiezione sull'autospazio corrispondente all'autovettore massimo).

Negli anni 70, matematici sovietici hanno affrontato il problema di determinare $E(M_{\mu, \rho}, \epsilon)$, giungendo alla stima seguente

$$E(M_{\mu, \rho}, \epsilon) \leq C(\rho \epsilon^{2\mu})^{\frac{1}{2\mu+1}}$$

dove $C > 0$ è una costante (si consulti [20]).

Definizione 3.8 Sia $T \in B(X, Y)$, tale che $R(T)$ non sia chiuso.

Sia $\{R_\mu\}_{\mu > 0}$ una famiglia di operatori $R_\mu: Y \rightarrow X$, $\epsilon > 0$.

Fissato $\mu > 0$, se esiste una costante $C_\mu > 0$ tale che

$$\forall \rho \geq 0, \quad \epsilon > 0, \quad x_* \in M_{\mu, \rho} \text{ e } y_* \in Y \text{ con } \|Tx_* - y_*\| \leq \epsilon$$

si abbia

$$\|x_* - R_\mu y_*\| \leq C_\mu (\rho \epsilon^{2\mu})^{\frac{1}{2\mu+1}} \quad (3.35)$$

allora $\{R_\mu\}_{\mu > 0}$ si dice di ordine ottimo rispetto a $\mu > 0$.

Si osservi che più μ è grande, migliore è la convergenza; questo è in accordo con quanto appena detto circa le soluzioni corrispondenti a diversi μ .

Possiamo quindi enunciare il teorema di Nemirovskii riguardante l'algoritmo del gradiente coniugato.

Teorema 3.13 Nemirovskii

Sia $T \in B(X, Y)$, con $R(T)$ non chiuso.

Se $y \in R(T^*)$ e

$$x_* = T^1 y \in M_{\mu, \rho} \quad (3.36)$$

allora il metodo del gradiente coniugato, con regola d'arresto in accordo con il principio della discrepanza (3.33), è un metodo di ordine ottimo rispetto a $\mu > 0$, cioè esiste $C_\mu > 0$ tale che

$$\|T^1 y - x_{n(\epsilon, y_\epsilon)}\| \leq C_\mu (\rho \epsilon^{2\mu})^{\frac{1}{2\mu+1}} \quad (3.37)$$

dove $n(\epsilon, y_\epsilon)$ è l'ultima iterata effettuata.

Fissato $\mu > 0$, questo risultato assicura la convergenza solo per soluzioni $x_* \in M_{\mu, \rho}$. Ne segue che la convergenza è garantita sul sottoinsieme

$$\cup_{\mu > 0} R((T^* T)^\mu) \subset \overline{R(T^*)}$$

Si osservi che la definizione 3.6 di algoritmo regolarizzante richiede invece la convergenza $\forall x_* \in N(T)^\perp = R(T^*)$. Come già accennato, il risultato di Nemirovskii non garantisce l'appartenenza del gradiente coniugato all'insieme degli algoritmi regolarizzanti.

Dedichiamoci al gradiente coniugato, supponendo, per semplicità, $\|T\| \leq 1$ e $x_0 = 0$. Indichiamo con $\{E_\lambda\}$ e $\{F_\lambda\}$ le famiglie spettrali associate rispettivamente agli operatori autoaggiunti semidefiniti positivi $T^* T$ e $T T^*$ (vedi appendice B).

Il teorema 3.11, nel caso in cui $x_0 = 0$, consente di scrivere

$$x_k = q_{k-1}(T^* T)^k T^* y \quad (3.38)$$

dove q_{k-1} è un opportuno polinomio di grado $(k-1)$. Infatti, con le notazioni del teorema 3.11, per $x_0 = 0$

$$\begin{aligned} P_{k-1}(T^* T)^{k_0} &= P_{k-1}(T^* T)(T^* T x_0 - T^* y_\epsilon) \\ &= -P_{k-1}(T^* T) T^* y_\epsilon = q_{k-1}(T^* T) T^* y_\epsilon \end{aligned}$$

Ponendo

$$p_k(\lambda) = 1 - \lambda q_{k-1}(\lambda) \quad (3.39)$$

si ha

$$\begin{aligned} y_k - Tx_k &= y_k - Tq_{k-1}(T^*T)y_k \\ &= y_k - q_{k-1}(T^*T)T^*y_k \\ &= p_k(T^*T)y_k \end{aligned}$$

Per questa loro caratteristica i polinomi p_k vengono detti polinomi residui.

Si dimostra che la sequenza $\{p_k\}$ è formata da polinomi ortogonali rispetto al prodotto scalare

$$[\phi, \psi] = \int_0^{+\infty} \phi(\lambda)\psi(\lambda)d\|F_\lambda y_k\|^2 \quad (3.40)$$

La misura è associata alla distribuzione spettrale del dato y_k rispetto a $\{F_\lambda\}$ e l'integrale, sebbene rappresentato su $[0, +\infty)$, in realtà ha significato solo tra $[0, 1]$ poiché $\|T^*T\| \leq 1$ (ossia $F_\lambda = I$ per $\lambda \geq 1$).

Osservando che $t : \lambda \rightarrow \|F_\lambda y_k\|^2$ è monotona non decrescente (cfr. appendice B), indichiamo con \bar{k} il numero di punti in cui $\|F_\lambda y_k\|^2$ cresce positivamente, cioè

$$\bar{k} < +\infty \quad \text{se } \|F_\lambda y_k\|^2 \text{ ha un numero finito di discontinuità a salto ed è costante a tratti}$$

$$\bar{k} = +\infty \quad \text{se } \|F_\lambda y_k\|^2 \text{ ha un intervallo in cui è strettamente crescente o un numero infinito di discontinuità a salto}$$

È importante notare che, in accordo con il teorema 3.11, se $\bar{k} < +\infty$, l'algoritmo termina dopo \bar{k} iterazioni; infatti $P_{R(T^*T)}y_k$ è una combinazione lineare di \bar{k} autovettori di T^*T e appartiene a un sottospazio di dimensione finita.

Per definizione (3.39), la base di polinomi ortogonali $\{p_k\}$ è formata da $\bar{k} + 1$ polinomi tali che $p_k(0) = 1$, $0 \leq k \leq \bar{k}$.

Indichiamo con Π_k^c l'insieme formato da tutti i polinomi di grado al più k , con la normizzazione $p_k(0) = 1 \quad \forall k \leq \bar{k}$.

Utilizzando il formalismo introdotto, la proprietà di ottimalità enunciata nel teorema 3.12 può essere scritta nel seguente modo

$$\|y_k - Tx_k\| = [p_k, \frac{1}{\lambda} p_k]^{1/2} \leq [p, \frac{1}{\lambda} p]^{1/2} \quad \forall p \in \Pi_k^c \quad (3.41)$$

infatti, essendo $y_k - Tx_k = p_k(T^*T)y_k$, si ha

$$\|y_k - Tx_k\|^2 = \int_0^{+\infty} p_k^2(\lambda) d\|F_\lambda y_k\|^2 = [p_k, \frac{1}{\lambda} p_k] \quad (3.42)$$

Possiamo anche determinare un insieme di polinomi $\{s_k\}$ in relazione con le direzioni $\{s_k\}$ T^*T -conjugate.

Ricordando che $\alpha_k d_k = z_k - z_{k+1} = [q_{k-1}(T^*T) - q_k(T^*T)]T^*y_k$, poniamo

$$s_k(\lambda) = q_{k-1}(\lambda) - q_k(\lambda) = \frac{p_{k+1}(\lambda) - p_k(\lambda)}{\lambda} \quad (3.43)$$

quindi

$$d_k = \frac{1}{\alpha_k} s_k(T^*T)T^*y_k$$

Per l'ortogonalità del sistema $\{p_k\}$, si ottiene

$$[s_k, \lambda p] = [p_{k+1}, p] - [p_k, p] = 0 \quad (3.44)$$

$\forall p$ polinomio di grado $k' \leq k-1$.

Un ulteriore commento merita il caso in cui l'algoritmo determina la soluzione $T^1 y_k$ in un numero finito (ossia \bar{k}) di passi, che come abbiamo già detto, si verifica quando $P_{R(T^*T)}y_k$ è combinazione lineare di un numero finito di autovettori di T^*T . In questo caso la discrepanza $y_k - Tx_k$ è la minima, in norma, raggiungibile dal gradiente coniugato; si osservi infatti che $y_k - Tx_k$ contiene solo le componenti in $R(T)^\perp = N(T^*)$.

Si ha così $T^*(y_k - Tx_k) = 0$, ossia

$$T^* p_k(T^*T)y_k = 0$$

Questo si verifica poiché il polinomio p_k ha tutte le \bar{k} radici nei punti di crescita positiva di $\|F_\lambda y_k\|^2$, unici punti in cui la misura indotta in (3.40) è positiva (si osservi che in tutti gli altri punti $\|F_\lambda y_k\|^2$ è costante, da cui $d\|F_\lambda y_k\|^2 = 0$).

Con l'estensione dei concetti del paragrafo precedente al nuovo formalismo, basato sul prodotto scalare (3.40), possiamo procedere con l'analisi della dimostrazione di Hanke del teorema di Nemrovskii 3.13.

I prossimi due lemmi, utili in seguito, riconducono la stima del residuo e dell'errore alla valutazione di $|p_k'(0)|$. Questo numero gioca un ruolo fondamentale nella dimostrazione proposta da Hanke e il suo significato verrà illustrato più avanti.

Spesso verranno indicate con la stessa lettera C le diverse costanti che compaiono nelle equazioni.

Lemma 3.9 Se $x = T^1 y$ appartiene a $M_{\mu, \rho}$, con $y \in Y$ allora per $1 \leq k \leq \bar{k}$

$$\|y_k - Tx_k\| \leq \|y - y_k\| + C_1 |p_k'(0)|^{-\frac{1}{2}} \rho \quad (3.45)$$

dove C_1 è una costante

Dimostrazione. Ricordando che abbiamo supposto $\|T\| \leq 1$, il k -esimo polinomio residuo p_k , essendo un polinomio ortogonale (cfr. [30]), ha k radici distinte $\{\lambda_{j,k}\}_{j=1}^k$ tali che

$$0 < \lambda_{1,k} < \lambda_{2,k} < \dots < \lambda_{k,k} \leq 1$$

Ricordando che $p_k(0) = 1$, possiamo scrivere

$$p_k(\lambda) = \prod_{j=1}^k \left(1 - \frac{\lambda}{\lambda_{j,k}} \right) \quad (3.46)$$

da cui si ottiene, per calcolo diretto,

$$|p_k'(0)| = \sum_{j=1}^k \frac{1}{\lambda_{j,k}} \quad (3.47)$$

Si osservi inoltre che

$$\left[p_k, \frac{p_k}{\lambda - \lambda_{1,k}} \right] = \int_0^{+\infty} p_k(\lambda) \frac{p_k(\lambda)}{\lambda - \lambda_{1,k}} \lambda d\|F_\lambda y_k\|^2 = 0 \quad (3.48)$$

poiché $\frac{p_k^2}{\lambda - \lambda_{1;k}}$ è un polinomio di grado $(k-1)$ e $\{p_k\}$ è base ortogonale rispetto a $[e, e]$. L'ultima relazione può essere scritta nella seguente forma

$$\int_0^{\lambda_{1;k}} p_k^2(\lambda) \frac{\lambda}{\lambda_{1;k} - \lambda} d\|F_{\lambda} y_e\|^2 = \int_{\lambda_{1;k}}^{+\infty} p_k^2(\lambda) \frac{\lambda}{\lambda - \lambda_{1;k}} d\|F_{\lambda} y_e\|^2$$

Inoltre

$$\int_0^{\lambda_{1;k}} p_k^2(\lambda) \frac{\lambda}{\lambda_{1;k} - \lambda} d\|F_{\lambda} y_e\|^2 \geq \int_{\lambda_{1;k}}^{+\infty} p_k^2(\lambda) d\|F_{\lambda} y_e\|^2$$

infatti $\frac{\lambda}{\lambda - \lambda_{1;k}} > 1$ per $\lambda > \lambda_{1;k}$.
Si ottiene

$$\begin{aligned} \|y_e - T x_k\|^2 &= \int_0^{+\infty} p_k^2(\lambda) d\|F_{\lambda} y_e\|^2 \\ &= \int_0^{\lambda_{1;k}} p_k^2(\lambda) d\|F_{\lambda} y_e\|^2 + \int_{\lambda_{1;k}}^{+\infty} p_k^2(\lambda) d\|F_{\lambda} y_e\|^2 \\ &\leq \int_0^{\lambda_{1;k}} p_k^2(\lambda) \left(1 + \frac{\lambda}{\lambda_{1;k} - \lambda}\right) d\|F_{\lambda} y_e\|^2 \end{aligned}$$

Ponendo

$$\phi_k(\lambda) = p_k(\lambda) \left(\frac{\lambda_{1;k}}{\lambda_{1;k} - \lambda}\right)^{\frac{1}{2}} \quad 0 \leq \lambda \leq \lambda_{1;k}$$

riscriviamo l'ultima disuguaglianza nel seguente modo (cfr. appendice B)

$$\|y_e - T x_k\| \leq \|F_{\lambda_{1;k}} \phi_k(T T^*) y_e\| \quad (3.49)$$

Tenendo presente che, per ipotesi, $y = T x_*$, con $x_* \in M_{\mu, \rho}$, ossia $x_* = (T^* T)^{\mu} z$ con $\|z\| \leq \rho$

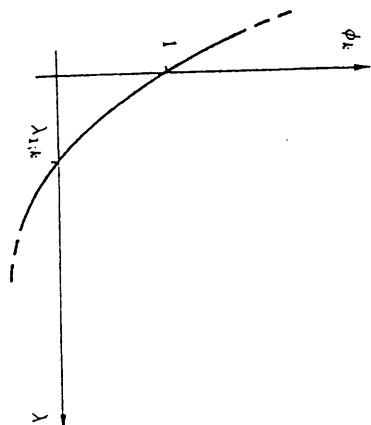
$$\begin{aligned} \|y_e - T x_k\| &\leq \|F_{\lambda_{1;k}} \phi_k(T T^*) (y_e - y + y)\| \\ &\leq \|F_{\lambda_{1;k}} \phi_k(T T^*) (y_e - y)\| + \|F_{\lambda_{1;k}} \phi_k(T T^*) y\| \\ &= \|F_{\lambda_{1;k}} \phi_k(T T^*) (y_e - y)\| + \|F_{\lambda_{1;k}} \phi_k(T T^*) T (T T^*)^{\mu} z\| \\ &= \|F_{\lambda_{1;k}} \phi_k(T T^*)\| \|y_e - y\| + \|F_{\lambda_{1;k}} \phi_k(T T^*) T (T T^*)^{\mu}\| \|z\| \\ &\leq \|\phi_k(\lambda)\| |c_{[0, \lambda_{1;k}]}| \|y - y_e\| + \|\lambda^{\mu + \frac{1}{2}} \phi_k(\lambda)\| |c_{[0, \lambda_{1;k}]}| \rho \end{aligned} \quad (3.50)$$

dove $\|\cdot\|$ è la usuale norma "del sup", detta anche norma della convergenza uniforme (si osservi che $[0, \lambda_{1;k}]$ è chiuso e limitato, e $\phi_k(\lambda)$ è ivi continua).

Cerchiamo di stimare le due norme "del sup" determinando i punti di massimo delle rispettive funzioni.

Studiamo l'andamento di $\phi_k(\lambda)^2$ nell'intervallo $[0, \lambda_{1;k}]$. È immediato verificare che $\phi_k(\lambda)^2$ ha $(2k-1)$ radici non distinte, tutte maggiori di $\lambda_{1;k}$. Inoltre, poiché $\phi_k(0) = 1$, si ha evidentemente

$$0 \leq \phi_k \leq 1 \quad 0 \leq \lambda \leq \lambda_{1;k} \quad (3.51)$$



Analizziamo ora l'andamento di $\phi_k^2(\lambda) \lambda^{2\mu+1}$ per un fissato $\mu > 0$, determinando il punto di massimo appartenente a $[0, \lambda_{1;k}]$.
Derivando rispetto a λ , determiniamo i punti di stazionarietà. Si ha:

$$\begin{aligned} (\lambda^{2\mu+1} \phi_k^2(\lambda))' &= (2\mu+1) \lambda^{2\mu} \phi_k^2(\lambda) + \lambda^{2\mu+1} 2\phi_k(\lambda) \phi_k'(\lambda) \\ &= \lambda^{2\mu} \phi_k(\lambda) \left\{ (2\mu+1) \phi_k(\lambda) + 2\lambda \left(p_k'(\lambda) \left(\frac{\lambda_{1;k}}{\lambda_{1;k} - \lambda} \right)^{\frac{1}{2}} \right. \right. \\ &\quad \left. \left. + p_k(\lambda) \frac{1}{2} \left(\frac{\lambda_{1;k}}{\lambda_{1;k} - \lambda} \right)^{-\frac{1}{2}} \lambda_{1;k} (\lambda_{1;k} - \lambda)^{-2} \right) \right\} \end{aligned}$$

da cui

$$\begin{aligned} (2\mu+1) p_k(\lambda) \left(\frac{\lambda_{1;k}}{\lambda_{1;k} - \lambda} \right)^{\frac{1}{2}} + \lambda \left(\frac{\lambda_{1;k}}{\lambda_{1;k} - \lambda} \right)^{\frac{1}{2}} \left(2p_k'(\lambda) + p_k(\lambda) \frac{1}{\lambda_{1;k} - \lambda} \right) &= 0 \\ \Leftrightarrow 2\mu+1 + 2\lambda \frac{p_k'(\lambda)}{p_k(\lambda)} + \lambda(\lambda_{1;k} - \lambda)^{-1} &= 0 \end{aligned}$$

Si osservi che è possibile esplicitare l'espressione di $\frac{p_k'(\lambda)}{p_k(\lambda)}$. Infatti da (3.46), per calcolo diretto si ha

$$p_k'(\lambda) = -\frac{1}{\lambda_{1;k}} \prod_{j \neq 1}^k \left(1 - \frac{\lambda}{\lambda_{j;k}} \right) - \frac{1}{\lambda_{2;k}} \prod_{j \neq 2}^k \left(1 - \frac{\lambda}{\lambda_{j;k}} \right) - \dots - \frac{1}{\lambda_{k;k}} \prod_{j \neq k}^k \left(1 - \frac{\lambda}{\lambda_{j;k}} \right) - \dots$$

da cui

$$\begin{aligned} \frac{p_k'(\lambda)}{p_k(\lambda)} &= -\frac{1}{\lambda_{1;k}} \left(1 - \frac{\lambda}{\lambda_{1;k}} \right)^{-1} - \frac{1}{\lambda_{2;k}} \left(1 - \frac{\lambda}{\lambda_{2;k}} \right)^{-1} - \dots \\ &= -\sum_{j=1}^k \frac{1}{\lambda_{j;k}} \left(\frac{\lambda_{j;k}}{\lambda_{j;k} - \lambda} \right) = -\sum_{j=1}^k \left(\frac{1}{\lambda_{j;k} - \lambda} \right) \end{aligned}$$

Si ottiene così che il punto di massimo λ_* in $[0, \lambda_{1;k}]$ soddisfa alla relazione

$$2\mu+1 + \lambda_* \left(\frac{1}{\lambda_{1;k} - \lambda_*} - \sum_{j=1}^k \frac{2}{\lambda_{j;k} - \lambda_*} \right) = 0 \quad (3.52)$$

Osservando che $\lambda_{j,k} - \lambda_k \geq 0 \quad \forall j = 1, \dots, k$, allora per (3.47) si ha

$$\begin{aligned} 2\mu + 1 &= \lambda_k \left(\sum_{j=1}^k \frac{2}{\lambda_{j,k} - \lambda_k} - \frac{1}{\lambda_{j,k} - \lambda_k} \right) \\ &\geq \lambda_k \sum_{j=1}^k \frac{1}{\lambda_{j,k} - \lambda_k} \geq \lambda_k \sum_{j=1}^k \frac{1}{\lambda_{j,k}} \\ &= \lambda_k |p_k'(0)| \end{aligned}$$

Quindi $\lambda_k \leq (2\mu + 1) |p_k'(0)|^{-1}$, da cui, ricordando (3.51), si ottiene

$$\lambda^{2\mu+1} \phi_k^2(\lambda) \leq \lambda^{2\mu+1} \phi_k^2(\lambda) \leq (2\mu + 1) |p_k'(0)|^{-2(2\mu+1)} \quad (3.53)$$

Utilizzando i risultati (3.51) e (3.53), rispettivamente per $\phi_k^2(\lambda)$ e $\lambda^{2\mu+1} \phi_k^2(\lambda)$, ritorniamo alla maggiorazione (3.50) di $\|y_k - Tx_k\|$. Si ha

$$\begin{aligned} \|y_k - Tx_k\| &\leq \|\phi_k(\lambda)\|_{C[0, \lambda_{1,k}]} \|y - y_k\| + \|\lambda^{\mu+\frac{1}{2}} \phi_k(\lambda)\|_{C[0, \lambda_{1,k}]} \rho \\ &\leq \|y - y_k\| + (2\mu + 1)^{\mu+\frac{1}{2}} |p_k'(0)|^{-\mu-\frac{1}{2}} \rho \end{aligned}$$

Ricordando che $\mu > 0$ è da considerarsi fissato, ponendo

$$C_1 = (2\mu + 1)^{\mu+\frac{1}{2}}$$

si ha la tesi.

Passiamo al secondo lemma, riguardante l'errore k -esimo.

Lemma 3.10 Nelle ipotesi del lemma precedente, si ha

$$\|T^l y - x_k\| \leq C_2 \left(\omega_k^{\frac{2l}{2\mu+1}} \rho^{\frac{2l}{2\mu+1}} + |p_k'(0)|^{\frac{1}{2}} \omega_k \right) \quad (3.54)$$

dove

$$\omega_k := \max \{ \|y_k - Tx_k\|, \|y - y_k\| \} \quad (3.55)$$

e C_2 è una opportuna costante.

Dimostrazione. Verifichiamo innanzitutto la tesi per $k = 0$. Si ricordi che, in accordo con la definizione 3.7, si è assunto $x_0 = 0$.

Le ipotesi sul (3.36) permettono di determinare la seguente disegugnanza, della quale non riportiamo la dimostrazione, detta disegugnanza interpolatoria

$$\|T^l y\| \leq \|y\| \omega^{\frac{2l}{2\mu+1}} \rho^{\frac{2l}{2\mu+1}}$$

Osservando che $p_0'(0) = 0$ e $\|y\| \leq \|y_k\| + \|y - y_k\| \leq 2\omega_0$, si ha

$$\|T^l y\| \leq 2 \omega_0^{\frac{2l}{2\mu+1}} \omega^{\frac{2l}{2\mu+1}} \rho^{\frac{2l}{2\mu+1}} \quad (3.56)$$

che prova la tesi per $k = 0$, $C_2 = 2 \omega_0^{\frac{2l}{2\mu+1}}$.

Sia allora $k > 0$; poniamo

$$0 < \delta \leq |p_k'(0)|^{-1}$$

Si osservi che, per (3.47),

$$|p_k'(0)| = \sum_{j=1}^k \frac{1}{\lambda_{j,k}} \geq \frac{1}{\lambda_{1,k}} \quad (3.57)$$

da cui

$$0 < \delta \leq |p_k'(0)|^{-1} \leq \lambda_{1,k} \quad (3.58)$$

Indicando con $x_k = q_{k-1}(T^*T)^T y_k$ l'iterata k -esima calcolata con dato perturbato, consideriamo

$$\tilde{x}_k = q_{k-1}(T^*T)^T y$$

Si osservi che il polinomio q_{k-1} che rappresenta l'iterata k -esima dipende dal particolare dato in ingresso; ne segue che \tilde{x}_k non è la k -esima iterata calcolata con dato esatto y , poiché questa sarebbe fornita da un differente polinomio $\tilde{q}_{k-1}(\lambda)$.

Si ha, per (3.39),

$$x_k - \tilde{x}_k = x_k - q_{k-1}(T^*T)^T y = p_k(T^*T)^T x_k = p_k(T^*T)^T y$$

Ricordando che, per ipotesi, $x_k \in M_{\mu,\rho}$, possiamo stimare l'errore k -esimo nel modo seguente

$$\begin{aligned} \|x_k - z_k\| &\leq \|E_\delta(x_k - z_k)\| + \|(I - E_\delta)(x_k - z_k)\| \\ &\leq \|E_\delta(x_k - \tilde{x}_k)\| + \|E_\delta(\tilde{x}_k - z_k)\| + \left(\int_\delta^{+\infty} d\|E_\lambda(x_k - z_k)\|^2 \right)^{\frac{1}{2}} \\ &= \|E_\delta p_k(T^*T)^T y - z_k\| + \|E_\delta q_{k-1}(T^*T)^T (y - y_k)\| + \left(\int_\delta^{+\infty} d\|E_\lambda(T^*T)^T (x_k - z_k)\|^2 \right)^{\frac{1}{2}} \\ &\leq \|E_\delta p_k(T^*T)^T y - z_k\| + \|E_\delta q_{k-1}(T^*T)^T (y - y_k)\| + \delta^{-\frac{1}{2}} \|(I - E_\delta)(y - Tx_k)\| \\ &\leq \|\lambda^\mu p_k(\lambda)\|_{C[0, \delta]} \rho + \|\lambda^{\frac{1}{2}} q_{k-1}(\lambda)\|_{C[0, \delta]} \|y - y_k\| + \delta^{-\frac{1}{2}} \|y - Tx_k\| \\ &\leq \|\lambda^\mu p_k(\lambda)\|_{C[0, \delta]} \rho + \|\lambda^{\frac{1}{2}} q_{k-1}(\lambda)\|_{C[0, \delta]} \|y - y_k\| + \delta^{-\frac{1}{2}} (\|y_k - Tx_k\| + \|y - y_k\|) \end{aligned} \quad (3.59)$$

Si osservi il passaggio da E_λ a F_λ nell'unico integrale esplicitato:

considerando solo gli autovalori $\lambda > \delta > 0$ l'operatore risultante è invertibile; si ha così, con abuso di linguaggio, $T^*T = T^{-1}(T^*T)^T$.

Il problema, come nel lemma precedente, è così ricondotto alla stima delle due norme

$$\| \bullet \|_{C[0, \delta]}$$

Per la prima norma, poiché $0 \leq p_k(\lambda) \leq 1$ in $[0, \lambda_{1,k}]$, si ha

$$0 \leq \lambda^\mu p_k(\lambda) \leq \delta^\mu \quad \lambda \in [0, \delta] \quad (3.60)$$

Vediamo come migliorare la seconda norma.

Poiché $p_k(0) = 1$ e $\lambda_{1,k}$ è la radice più vicina allo zero, il polinomio p_k in $[0, \lambda_{1,k}]$ è decrescente e concavo.

Questo implica che, in $[0, \lambda_{1,k}]$, la derivata massima, in valore assoluto, si ha per $\lambda = 0$. Si ha perciò

$$0 \leq \lambda q_{k-1}'(\lambda) = \frac{1 - p_k(\lambda)}{\lambda} (1 - p_k(\lambda)) \leq |p_k'(0)| \quad (3.61)$$

per $\lambda \in [0, \delta] \subset [0, \lambda_{1,k}]$.

Infatti, dal teorema di Lagrange, esiste $\tilde{\lambda} \in (0, \lambda_{1,k})$ tale che

$$p_k(\lambda) = p_k(0) + p_k'(\tilde{\lambda})\lambda \quad \forall \lambda \in [0, \lambda_{1,k}]$$

ossia $-p_k'(\lambda) = \frac{1-p_k(\lambda)}{\lambda}$

Considerando che, se $\lambda \in [0, \delta]$, allora $0 \leq 1 - p_k(\lambda) \leq 1$, otteniamo

$$\frac{1 - p_k(\lambda)}{\lambda} (1 - p_k(\lambda)) \leq |p_k'(0)| \quad \lambda \in [0, \delta] \quad (3.62)$$

Si osservi che il valore $\delta \in (0, |p_k'(0)|^{-1}]$, in (3.58), ha permesso, con il passaggio da $\{E_\lambda\}$ a $\{F_\lambda\}$ in (3.59), di mettere in relazione l'errore $\|x_n - x_k\|$ con il residuo $\|y_n - T x_k\|$ stimato nel lemma precedente.

Considerando (3.61) e (3.60), abbiamo la seguente maggiorazione

$$\|x_n - x_k\| \leq \delta^\mu \rho + |p_k'(0)|^{\frac{1}{2}} \omega_k + 2\delta^{-\frac{1}{2}} \omega_k \quad (3.63)$$

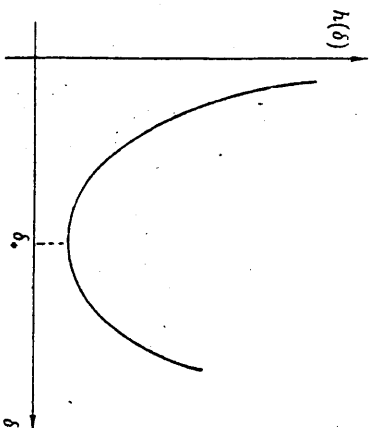
Studiamo le caratteristiche della funzione

$h(\delta) = \delta^\mu \rho + |p_k'(0)|^{\frac{1}{2}} \omega_k + 2\delta^{-\frac{1}{2}} \omega_k$

posto

$$\delta_0 = \left(\frac{1}{\mu} \frac{\omega_k}{\rho} \right)^{\frac{2\mu}{2\mu+1}} > 0$$

$h(\delta)$ è decrescente in $(0, \delta_0)$ e crescente in $(\delta_0, +\infty)$.



Infatti, calcolando la derivata prima, si ha $h'(\delta) = \mu \delta^{\mu-1} - \delta^{-\frac{1}{2}} \omega_k$

da cui $h'(\delta) = 0$ in $(0, +\infty)$ per $\delta^{-\frac{1}{2}} (\mu \delta^{\mu+\frac{1}{2}} - \omega_k) = 0$,

ossia per $\delta^{\mu+\frac{1}{2}} = \frac{\omega_k}{\mu}$.

Al fine di avere la maggiorazione più fine, distinguiamo due casi a seconda della posizione di δ_0 rispetto a $|p_k'(0)|^{-1}$.

(i) Se $|p_k'(0)|^{-1} \leq \delta_0$, il punto di minimo di $h(\delta)$ in $(0, |p_k'(0)|^{-1}]$ è $\delta = |p_k'(0)|^{-1}$, poiché, come già osservato, $h(\delta)$ è decrescente in $(0, \delta_0) \supset (0, |p_k'(0)|^{-1})$.

Poniamo allora $\delta = |p_k'(0)|^{-1}$; da (3.63) si ottiene

$$\|x_n - x_k\| \leq |p_k'(0)|^{-\mu} \rho + |p_k'(0)|^{\frac{1}{2}} \omega_k + 2|p_k'(0)|^{\frac{1}{2}} \omega_k$$

$$\leq \delta_0^\mu \rho + 3|p_k'(0)|^{\frac{1}{2}} \omega_k$$

$$= \mu^{-\frac{2\mu}{2\mu+1}} \rho^{-\frac{2\mu}{2\mu+1} + 1} \omega_k^{\frac{2\mu}{2\mu+1}} + 3|p_k'(0)|^{\frac{1}{2}} \omega_k$$

$$\leq C_2 \left(\omega_k^{\frac{2\mu}{2\mu+1}} \rho^{\frac{2\mu}{2\mu+1}} + |p_k'(0)|^{\frac{1}{2}} \omega_k \right)$$

con $C_2 = \max\{\mu^{-\frac{2\mu}{2\mu+1}}, 3\}$.

In questo caso la tesi è verificata.

(ii) Se $|p_k'(0)|^{-1} > \delta_0$, la maggiorazione risulta più semplice poiché possiamo proprio considerare il punto di minimo assoluto $\delta_0 \in (0, |p_k'(0)|^{-1})$.

Poniamo $\delta = \delta_0$; si ha così

$$\begin{aligned} \|x_n - x_k\| &\leq \delta_0^\mu \rho + |p_k'(0)|^{\frac{1}{2}} \omega_k + 2\delta_0^{-\frac{1}{2}} \omega_k \\ &= \left(\frac{1}{\mu} \frac{\omega_k}{\rho} \right)^{\frac{2\mu}{2\mu+1}} \rho + |p_k'(0)|^{\frac{1}{2}} \omega_k + 2 \left(\frac{1}{\mu} \frac{\omega_k}{\rho} \right)^{-\frac{1}{2\mu+1}} \omega_k \\ &\leq C_2 \left(\omega_k^{\frac{2\mu}{2\mu+1}} \rho^{\frac{2\mu}{2\mu+1}} + |p_k'(0)|^{\frac{1}{2}} \omega_k \right) \end{aligned}$$

con $C_2 = \max\{\mu^{-\frac{2\mu}{2\mu+1}} + 2\mu^{\frac{2\mu}{2\mu+1}}, 1\}$; che verifica le tesi anche in questo caso.

Prima di proseguire è utile fare alcune considerazioni sui due ultimi lemmi (3.9 e 3.10) e sul ruolo di $|p_k'(0)|$.

Il lemma 3.10 riconduce la stima dell'errore alla conoscenza di $|p_k'(0)|^{-1}$ e del residuo (presente in ω_k). Del resto il residuo è stato maggiorato nel lemma 3.9. Inserendo quest'ultima maggiorazione in (3.54), si ottiene

$$\begin{aligned} \|T^t y - x_k\| &= \|x_n - x_k\| \\ &\leq C_2 \left\{ \|y - y_e\| \frac{\omega_k}{\rho^{\frac{2\mu}{2\mu+1}}} + C_1 \left(|p_k'(0)|^{-\frac{2\mu}{2\mu+1}} \rho \right)^{\frac{2\mu}{2\mu+1}} \rho^{\frac{2\mu}{2\mu+1}} \right. \\ &\quad \left. + |p_k'(0)|^{\frac{1}{2}} \|y - y_e\| + |p_k'(0)|^{\frac{1}{2}} C_1 |p_k'(0)|^{-\frac{2\mu}{2\mu+1}} \rho \right\} \\ &\leq C \left(\|y - y_e\| \frac{\omega_k}{\rho^{\frac{2\mu}{2\mu+1}}} + |p_k'(0)|^{-\mu} \rho + |p_k'(0)|^{\frac{1}{2}} \|y - y_e\| \right) \end{aligned}$$

dove C è una costante positiva.

Possiamo giungere a una espressione più semplice osservando che, come suggerisce Hanke, fissati $a, b \in \mathbb{R}$, positivi, esiste una costante $C > 0$ tale che

$$\tilde{C} a^{\frac{2\mu}{2\mu+1}} b^{\frac{2\mu}{2\mu+1}} \leq t^{-\mu} b + t^{\frac{1}{2}} a \quad \forall t > 0$$

(per la verifica si osservi che l'espressione a sinistra è una costante, mentre la funzione di t a destra è continua, positiva, tendente a $+\infty$ per $t \rightarrow 0$ e per $t \rightarrow +\infty$).

Applicando questa relazione con $a = \|y - y_e\|$, $b = \rho$ e considerando $t = |p_k'(0)|$, si ottiene

$$\begin{aligned} \|x_n - x_k\| &\leq C \left(\|y - y_e\| \frac{\omega_k}{\rho^{\frac{2\mu}{2\mu+1}}} + |p_k'(0)|^{-\mu} \rho + |p_k'(0)|^{\frac{1}{2}} \|y - y_e\| \right) \\ &\leq C \left(|p_k'(0)|^{-\mu} \rho + |p_k'(0)|^{\frac{1}{2}} \|y - y_e\| + |p_k'(0)|^{-\mu} \rho + |p_k'(0)|^{\frac{1}{2}} \|y - y_e\| \right) \\ &\leq 2C \left(|p_k'(0)|^{-\mu} \rho + |p_k'(0)|^{\frac{1}{2}} \|y - y_e\| \right) \quad (3.64) \end{aligned}$$

Si vede chiaramente che $|p_k'(0)|$ regola la convergenza e la divergenza del metodo. Infatti, poiché

$$|p_k'(0)| = \sum_{j=1}^k \frac{1}{\lambda_j} \geq \sum_{j=1}^k 1 = k$$

per $0 < k \leq \bar{k}$, si ha che, nel caso in cui $\bar{k} = +\infty$, $|p_k'(0)| \rightarrow +\infty$, per $k \rightarrow +\infty$.
 È facile constatare come in (3.64) compaiano i due termini tipici della teoria della regolarizzazione (cfr. capitolo 2) :

(i) $|p_k'(0)|^{-\mu} \rho$ convergente a zero, che riassume l'errore dovuto alla approssimazione del dato;

(ii) $|p_k'(0)|^{\frac{1}{2}} \|y - y_k\|$ divergente, che riassume l'errore dovuto alla perturbazione sul dato, il quale sappiamo, tende ad "esplodere" per k crescenti.

Si osservi che, in virtù della definizione (3.39), $|p_k'(0)| = q_{k-1}(0)$. Questo aiuta a capire il significato di $|p_k'(0)|$.

Sappiamo che la perturbazione sul dato si manifesta maggiormente nelle componenti corrispondenti ad autovalori piccoli (cfr. capitolo 2.2). Poiché q_{k-1} è il polinomio che esprime l'iterata $(k-1)$ -esima e $q_{k-1}(0)$ cresce al crescere di k , si ha che le componenti corrispondenti ad autovalori prossimi allo zero acquistano via via peso sempre maggiore.

Questo è proprio il comportamento che hanno gli algoritmi iterativi per problemi mal posti. Le prime iterate sono governate dalle componenti corrispondenti ad autovalori grandi, solitamente poco perturbate, mentre le successive iterate risentono delle componenti più oscillanti corrispondenti ad autovalori piccoli. In questo modo la perturbazione sul dato comporta errori sulla soluzione sempre più grandi.

Analogamente a quanto visto per l'algoritmo di Landweber-Fridman, esisterà un valore del parametro di regolarizzazione, detto n_{opt} , in relazione con i valori di $|p_k'(0)| = q_k(0)$, tale che, per $n > n_{opt}$, l'errore dovuto alla perturbazione non è più compensato dalla migliore precisione con cui l'algoritmo determina la soluzione.

Per dimostrare il teorema di Nemirovskii dovremo porre in relazione $|p_k'(0)|$ con ρ e ϵ . Il prossimo e ultimo lemma consentirà proprio di stimare $|p_k'(0)|$.

Lemma 3.11 Per $0 < k \leq \bar{k}$, si definisce

$$\pi_{k-1} = p_{k-1}'(0) - p_k'(0)$$

Allora

$$\pi_{k-1} = p_{k-1}'(0) - p_k'(0) = \frac{[s_{k-1}, s_{k-1}]}{[p_{k-1}, \frac{1}{\lambda} p_{k-1}] - [p_k, \frac{1}{\lambda} p_k]}$$

dove $[e, e]$ è il prodotto scalare (3.40)

Dimostrazione. Ricordando che $s_{k-1}(\lambda) = \frac{p_k(\lambda) - p_{k-1}(\lambda)}{\lambda}$ e utilizzando De l'Hospital,

si determina

$$s_{k-1}(0) = p_k'(0) - p_{k-1}'(0) = -\pi_{k-1}$$

Quindi s_{k-1} ammette la rappresentazione seguente

$$s_{k-1}(\lambda) = -\pi_{k-1} + \lambda e_{k-2}(\lambda)$$

con e_{k-2} opportuno polinomio di grado $k-2$.

Inoltre, per la proprietà (3.44) di ortogonalità dei polinomi s_k , si ha

$$[s_{k-1}, s_{k-1}] = -\pi_{k-1} [s_{k-1}, 1] + [s_{k-1}, \lambda e_{k-2}] = -\pi_{k-1} [s_{k-1}, 1] = \pi_{k-1} [p_{k-1} - p_k, \frac{1}{\lambda}]$$

Per $j = k-1$ o $j = k$, essendo, per definizione (3.39), $q_{j-1}(\lambda) = \frac{1}{\lambda} - \frac{p_j(\lambda)}{p_{j-1}(\lambda)}$, e utilizzando la proprietà di ortogonalità dei polinomi $\{p_j\}$ si ha

$[p_j, \frac{1}{\lambda}] = [p_j, q_{j-1} + \frac{1}{\lambda} p_j] = [p_j, \frac{1}{\lambda} p_j]$
 Possiamo così verificare la tesi :

$$\begin{aligned} [s_{k-1}, s_{k-1}] &= \pi_{k-1} \left([p_{k-1}, \frac{1}{\lambda}] - [p_k, \frac{1}{\lambda}] \right) \\ &= \pi_{k-1} \left([p_{k-1}, \frac{1}{\lambda} p_{k-1}] - [p_k, \frac{1}{\lambda} p_k] \right) \end{aligned}$$

Utilizzando i lemmi 3.9, 3.10 e 3.11 possiamo verificare la dimostrazione del teorema 3.13.

Dimostrazione. Teorema 3.13 Nemirovskii

Si osservi innanzitutto la stretta relazione tra (3.54) del lemma 3.10 e la tesi di Nemirovskii.

Indichiamo con $\bar{n} = n(\epsilon, y)$ la regola d'arresto basata sul principio della discrepanza (definizione 3.7), con $\tau > 1$ fissato.

Per (3.55), si ha

$$w_n \leq \tau \epsilon$$

infatti $\|y_k - T x_n\| \leq \tau \epsilon$ e $\|y - y_k\| \leq \epsilon < \tau \epsilon$.

La tesi del lemma 3.10 si caratterizza nel modo seguente

$$\|T^n y - x_n\| \leq C \left(\epsilon^{\frac{2n}{2n+1}} \rho^{\frac{1}{2n+1}} + |p_n'(0)|^{\frac{1}{2}} \epsilon \right) \quad (3.65)$$

Confrontando (3.65) con il teorema 3.13, se

$$|p_n'(0)|^{\frac{1}{2}} \epsilon \leq C \left(\epsilon^{\frac{2n}{2n+1}} \rho^{\frac{1}{2n+1}} \right), \text{ ossia}$$

$$|p_n'(0)|^{\frac{1}{2}} \leq C \left(\frac{\rho}{\epsilon} \right)^{\frac{1}{2n+1}} \quad (3.66)$$

con C costante opportuna, allora il lemma verifica la tesi del teorema. La dimostrazione è così ricondotta alla verifica della (3.66).

Si osserva che (3.66) è sicuramente verificata per $\bar{n} = 0$ (il polinomio è una costante); consideriamo quindi $\bar{n} \geq 1$.

Il lemma 3.9 come è stato fatto in (3.65) per il successivo 3.10, si caratterizza per $k = \bar{n} - 1$ nel modo seguente

$$\tau \epsilon < \|y_k - T x_{\bar{n}-1}\| \leq \epsilon + C |p_{\bar{n}-1}'(0)|^{-\mu - \frac{1}{2}} \rho \quad (3.67)$$

(per $\bar{n} = 1$ si consideri simbolicamente l'espressione a destra, dove si avrebbe l'inversione di 0, come $+\infty$).

Il residuo, per $k = \bar{n} - 1$, non soddisfa alla condizione di discrepanza; si ha la seguente maggiorazione

$$(\tau - 1) \epsilon < C |p_{\bar{n}-1}'(0)|^{-\mu - \frac{1}{2}} \rho, \text{ ossia}$$

$$|p_{\bar{n}-1}'(0)| < C_A \left(\frac{\rho}{\epsilon} \right)^{\frac{2}{2\mu+1}} \quad (3.68)$$

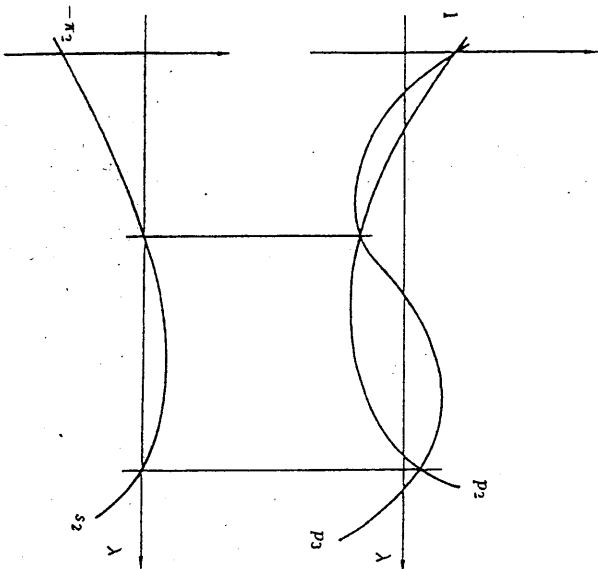
con $C_A = (C(\tau - 1))^{-\frac{2}{2\mu+1}}$.

La relazione (3.66) vale quindi per $k = \bar{n} - 1$.

Sulla base di questo risultato, utilizzando il lemma 3.11, che lega p_k' con p_{k-1}' , dimostremo (3.66).

A tale scopo studiamo le proprietà delle radici di p_n, p_{n-1}, s_{n-1} . Si osservi che, essendo $\{p_k\}_{k=0}^n$ formato da polinomi ortogonali, le radici di un generico polinomio p_j sono separate da quelle di p_{j-1} (cfr. [30]; questa proprietà può essere facilmente verificata per distribuzioni spettrali tali che $\|F_j y\| = \beta \|y\|, \beta \in (0, 1]$, le quali inducono una misura equivalente all'usuale misura di Lebesgue in \mathbb{R}).

Dalla definizione (3.43), studiamo il segno di s_{n-1} , ossia di $p_n - p_{n-1}$, negli zeri di p_{n-1} . Si veda l'esempio in figura per $\bar{n} = 3$



Le $\bar{n} - 1$ radici di $s_{\bar{n}-1}$, che indicheremo con $\lambda_{j,\bar{n}-1}$, si alternano alle $\bar{n} - 1$ radici $\lambda_{j,\bar{n}-1}$ di $p_{\bar{n}-1}$, cioè

$$0 < \lambda_{1,\bar{n}-1} < \lambda_{1,\bar{n}-1} < \lambda_{2,\bar{n}-1} < \lambda_{2,\bar{n}-1} < \dots < \lambda_{\bar{n}-1,\bar{n}-1} < \lambda_{\bar{n}-1,\bar{n}-1}$$

Per $0 < \delta \leq \lambda_{1,\bar{n}-1}$, si ha

$$0 \leq |s_{\bar{n}-1}(\lambda)| \leq |s_{\bar{n}-1}(0)| = \pi_{\bar{n}-1} \quad 0 \leq \lambda \leq \delta \leq \lambda_{1,\bar{n}-1} \quad (3.69)$$

Dalle relazioni (3.68) e (3.57), segue

$$\left(\frac{\rho}{\epsilon}\right)^{-\frac{\pi_{\bar{n}-1}}{\tau}} \leq C |p_{\bar{n}-1}'(0)|^{-1} \leq C \lambda_{1,\bar{n}-1}$$

Esiste quindi una costante C_δ , sufficientemente piccola, tale che posto

$$\delta = \left(C_\delta \frac{\epsilon}{\rho}\right)^{\frac{\tau}{\pi_{\bar{n}-1}}} \quad (3.70)$$

allora

$$0 < \delta < \lambda_{1,\bar{n}-1}$$

Dalla proprietà di ottimalità (3.41) applicata a polinomi $-\frac{s_{n-1}}{\pi_{n-1}} \in \Pi_{\bar{n}-1}^0$, si ha

$$\begin{aligned} \pi_{n-1} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]|^{\frac{1}{2}} &\leq \pi_{n-1} \left| -\frac{s_{n-1}}{\pi_{n-1}} - \frac{1}{\lambda} \frac{s_{n-1}}{\pi_{n-1}} \right|^{\frac{1}{2}} = |s_{n-1}(T T^n) y_\epsilon| \\ &\leq \|F_\delta s_{n-1}(T T^n) y_\epsilon\| + \|(I - F_\delta) s_{n-1}(T T^n) y_\epsilon\| \\ &\leq \|F_\delta s_{n-1}(T T^n) y_\epsilon\| + \delta^{-\frac{1}{2}} \|(I - E_\delta) T^n s_{n-1}(T T^n) y_\epsilon\| \end{aligned}$$

dove l'ultima maggiorazione è stata fatta in modo analogo a (3.59).

Otteniamo così

$$\begin{aligned} \pi_{n-1} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]|^{\frac{1}{2}} &\leq \|s_{n-1}(\lambda)\|_{C[0,\delta]} \|F_\delta y_\epsilon\| + \delta^{-\frac{1}{2}} \|T^n s_{n-1}(T T^n) y_\epsilon\| \\ &\leq \|s_{n-1}(\lambda)\|_{C[0,\delta]} \|F_\delta y_\epsilon\| + \delta^{-\frac{1}{2}} \left(\int_0^{+\infty} (\lambda^{\frac{1}{2}} s_{n-1}(\lambda))^2 d\|F_\lambda y_\epsilon\|^2 \right)^{\frac{1}{2}} \\ &\leq \|s_{n-1}(\lambda)\|_{C[0,\delta]} \|F_\delta y_\epsilon\| + \delta^{-\frac{1}{2}} [s_{n-1}, s_{n-1}]^{\frac{1}{2}} \end{aligned} \quad (3.71)$$

Valutiamo ogni singolo addendo dell'ultima relazione.

(i) Per $s_{n-1}(\lambda)$ in $[0, \delta]$, utilizziamo quanto determinato dall'analisi del segno (3.69); si determina così

$$\|s_{n-1}(\lambda)\|_{C[0,\delta]} \leq \pi_{n-1} \quad (3.72)$$

Per il secondo termine, ricordando che $T^i y = z, z \in M_{n,\rho}$:

$$\begin{aligned} \|F_\delta y_\epsilon\| &\leq \|F_\delta(y_\epsilon - y)\| + \|F_\delta y\| \\ &= \|F_\delta(y_\epsilon - y)\| + \|F_\delta T x_n\| \\ &= \|F_\delta(y_\epsilon - y)\| + \|F_\delta T(T^n T^i) z\| \\ &\leq \|y_\epsilon - y\| + \|F_\delta T(T^n T^i) z\| \\ &\leq \epsilon + \delta^{\mu+\frac{1}{2}} \rho = (1 + C_\delta) \epsilon \\ &< \frac{1 + C_\delta}{\tau} \|y_\epsilon - T x_{n-1}\| = \frac{1 + C_\delta}{\tau} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]|^{\frac{1}{2}} \end{aligned} \quad (3.73)$$

dove si è fatto uso di (3.67), (3.70) e (3.42).

(ii) Per il secondo addendo utilizziamo il lemma 3.11.

$$|s_{n-1}, s_{n-1}| = \pi_{n-1} |p_{n-1}[\frac{1}{\lambda} p_{n-1}] - \pi_{n-1} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]| \leq \pi_{n-1} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]|$$

Possiamo così ritornare a (3.71).

$$\begin{aligned} \pi_{n-1} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]|^{\frac{1}{2}} &\leq \pi_{n-1} \frac{1 + C_\delta}{\tau} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]|^{\frac{1}{2}} \\ &+ \delta^{-\frac{1}{2}} \pi_{n-1}^{\frac{1}{2}} |p_{n-1}[\frac{1}{\lambda} p_{n-1}]|^{\frac{1}{2}} \\ \Leftrightarrow \left(1 - \frac{1 + C_\delta}{\tau}\right) \pi_{n-1}^{\frac{1}{2}} &\leq \delta^{-\frac{1}{2}} \end{aligned} \quad (3.74)$$

$$\Leftrightarrow \left(\frac{\tau-1-C_\delta}{\tau} \right) \pi_{\tau-1}^{-\frac{1}{2}} \leq \left(\frac{C_\delta}{\rho} \right)^{-\frac{1}{2\tau+1}}$$

$$\Leftrightarrow (\tau-1-C_\delta) \pi_{\tau-1}^{-\frac{1}{2}} \leq \tau C_\delta^{-\frac{1}{2\tau+1}} \left(\frac{\epsilon}{\rho} \right)^{-\frac{1}{2\tau+1}}$$

Scegliendo la costante di (3.70) in $(0, \tau-1)$, si ottiene

$$\pi_{\tau-1} \leq C_B \left(\frac{\rho}{\epsilon} \right)^{\frac{2}{2\tau+1}}$$

dove C_B è una costante dipendente solo da τ e μ fissati.

Dalla definizione di $\pi_{\tau-1}$ del lemma 3.11 e dalla (3.68), utilizzando l'ultima relazione si ha che (3.66) è verificata anche per \bar{n} :

$$p_n'(0) = p_{n-1}'(0) - \pi_{n-1} \leq C_A \left(\frac{\rho}{\epsilon} \right)^{\frac{2}{2\tau+1}} + C_B \left(\frac{\rho}{\epsilon} \right)^{\frac{2}{2\tau+1}}$$

Questa, in accordo con la (3.66), conclude la dimostrazione. ■

Riassumendo, Nemirovskii ha dimostrato che, in presenza di un dato $y \in R(T)$ tale che $x_* = T^1 y = (T^* T)^{\mu} z$, con $\mu > 0$, allora il metodo del gradiente coniugato è di ordine ottimo rispetto a μ .

Questa proprietà non assicura che il metodo sia di regolarizzazione (nel senso della definizione 3.6) poiché, come è stato già detto, non si hanno indicazioni sulla convergenza del metodo per soluzioni appartenenti a tutto $N(T)^\perp$ (infatti $\bigcup_{\mu>0, \rho \geq 0} M_{\mu, \rho} = \bigcup_{\mu>0} R(T^* T)^\mu$ è propriamente contenuto in $N(T)^\perp$).

Concludiamo l'analisi delle proprietà di regolarizzazione del metodo riportando il risultato di R. Plato (cfr. [38]) riguardante la regolarizzazione mediante algoritmi di ordine ottimo. Plato ha dimostrato che ogni famiglia di operatori di ordine ottimo rispetto ad un generico $\mu_0 > 0$ conduce ad un algoritmo di regolarizzazione.

Il teorema di Plato non fa riferimento esclusivamente al gradiente coniugato, ma ha una valenza del tutto generale, ossia vale per qualsiasi famiglia di ordine ottimo. Però, come spiega l'autore, la più importante conseguenza del suo teorema è proprio la dimostrazione che il gradiente coniugato è un metodo di regolarizzazione.

Prima di enunciare il teorema, sarà utile accennare brevemente la dimostrazione.

Osservando che $R((T^* T)^{\mu_0})$ è denso in $N(T)^\perp$ (cfr. lemma 2.1 e [15]), per ogni elemento $x_* \in N(T)^\perp$, si determinerà una successione contenuta in $R((T^* T)^{\mu_0})$ e convergente a x_* . Il buon funzionamento del metodo per gli elementi della successione, conseguenza della proprietà di ordine ottimo, conduce alla convergenza anche per x_* .

Teorema 3.14 Plato

Sia $T \in B(X, Y)$ tale che $R(T)$ sia non chiuso.

Se $\{R_\epsilon\}_{\epsilon>0}$ è di ordine ottimo rispetto a $\mu_0 > 0$, allora la famiglia di operatori $\{\bar{R}_\epsilon\}_{\epsilon>0}$ definita da

$$\bar{R}_\epsilon = R_{b\epsilon} \quad b > 1$$

(I) è di ordine ottimo rispetto ad ogni $0 < \mu \leq \mu_0$

(II) è algoritmo di regolarizzazione nel senso della definizione 3.6.

Dimostrazione. Sia $x_* = T^1 y \in N(T)^\perp$. Determiniamo innanzitutto una famiglia di elementi $(x_\epsilon)_{\epsilon>0}$, con $x_\epsilon \in R((T^* T)^{\mu_0})$, $\epsilon > 0$, convergente a x_* ($\epsilon \rightarrow 0$). Tale famiglia dovrà avere le seguenti proprietà:

- (i) $x_\epsilon = (T^* T)^{\mu_0} z_\epsilon$ con $z_\epsilon \in X$, $\|z_\epsilon\| \leq \rho_\epsilon$
- (ii) $\|T(x_\epsilon - x_*)\| \leq (b-1)\epsilon$
- (iii) $x_\epsilon \rightarrow x_*$ ($\epsilon \rightarrow 0$)
- (iv) $\rho_\epsilon \epsilon^{2\mu_0} \rightarrow 0$ ($\epsilon \rightarrow 0$)

Se inoltre $x_* \in M_{\mu, \rho}$ per $0 < \mu \leq \mu_0$ e $\rho \geq 0$, allora

- (v) $(\rho_\epsilon \epsilon^{2\mu_0})^{\frac{1}{2\tau+1}} \leq C_1 (\rho_\epsilon \epsilon^{2\mu})^{\frac{1}{2\tau+1}}$
- (vi) $\|x_* - x_\epsilon\| \leq C_2 (\rho_\epsilon \epsilon^{2\mu})^{\frac{1}{2\tau+1}}$

con C_1, C_2 costanti, dipendenti unicamente da μ_0 .

Costruita tale famiglia, il teorema sarà dimostrato; infatti, osservando che:

- $\{R_\epsilon\}_{\epsilon>0}$ è di ordine ottimo rispetto a μ_0
- per (i), ogni elemento x_ϵ è contenuto in $M_{\mu_0, \rho}$.
- per (ii),

$$\|Tx_\epsilon - y_\epsilon\| \leq \|T(x_\epsilon - x_*)\| + \|Tx_* - y_\epsilon\|$$

$$\leq (b-1)\epsilon + \epsilon = b\epsilon$$

e considerando la definizione 3.8, si ottiene

$$\|x_* - \bar{R}_\epsilon y_\epsilon\| = \|x_* - R_{b\epsilon} y_\epsilon\|$$

$$\leq \|x_* - x_\epsilon\| + \|x_\epsilon - R_{b\epsilon} y_\epsilon\|$$

$$\leq \|x_* - x_\epsilon\| + b \frac{2\mu_0}{2\tau+1} C(\rho_\epsilon \epsilon^{2\mu_0})^{\frac{1}{2\tau+1}}$$

Quest'ultima relazione, con le proprietà (iii) e (iv), fornisce

$$\sup_{y_\epsilon \in Y, \|Tx_* - y_\epsilon\| \leq \epsilon} \|x_* - \bar{R}_\epsilon y_\epsilon\| \rightarrow 0 \quad (\epsilon \rightarrow 0) \quad (3.75)$$

che verifica la tesi (II), ossia che il metodo è di regolarizzazione.

Inoltre, se $x_* \in M_{\mu, \rho}$ $0 < \mu < \mu_0$, utilizzando (v) e (vi), si ottiene

$$\sup_{y_\epsilon \in Y, \|Tx_* - y_\epsilon\| \leq \epsilon} \|x_* - \bar{R}_\epsilon y_\epsilon\| \leq (C_2 + b \frac{2\mu_0}{2\tau+1} C C_1) (\rho_\epsilon \epsilon^{2\mu})^{\frac{1}{2\tau+1}} \quad (3.76)$$

Osservando che la (3.76) è valida per ogni arbitrario valore $0 < \mu < \mu_0$ tale che $x_* \in M_{\mu, \rho}$, allora si conclude che questa verifica la tesi (I), ossia che il metodo è di ordine ottimo rispetto ad ogni $0 < \mu \leq \mu_0$.

Per costruire la famiglia $(x_\epsilon)_{\epsilon>0}$, utilizzeremo un algoritmo di inversione basato sulla finestra spettrale rettangolare, il quale, come dimostra Vainikko (cfr. [44]), è algoritmo di regolarizzazione e di ordine ottimo rispetto ad ogni $\mu > 0$.

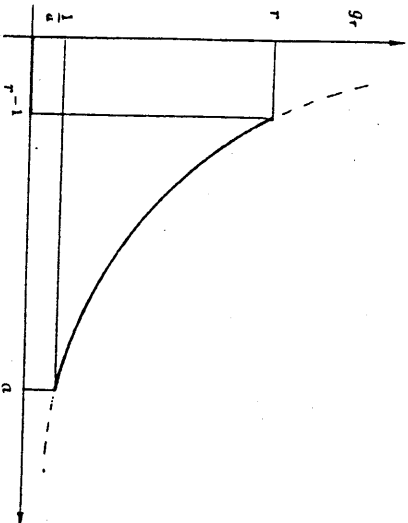
Consideriamo quindi la seguente famiglia di funzioni reali $(g_r)_{r \geq 0}$.

$$g_r : [0, a] \rightarrow \mathbb{R} \quad r \geq 0$$

$$g_0(t) \equiv 0$$

$$g_r(t) = \begin{cases} 0 & t < r^{-1} \\ \frac{1}{t} & t \geq r^{-1} \end{cases} \quad r > 0$$

con $a = \|T\|^2 = \|T^*T\|$.



Si osservi che questa famiglia di funzioni, in accordo col formalismo visto nel capitolo 2,6, definisce un algoritmo regolarizzante con finestra rettangolare

$$W_r(t) = \begin{cases} 0 & t < \frac{1}{r} \\ 1 & t \geq \frac{1}{r} \end{cases}$$

Si consideri la seguente famiglia di elementi $(x_r(y_\delta))_{r \geq 0} \subset X$ per ogni $y_\delta \in Y$, definita da

$$x_r(y_\delta) = g_r(T^*T)^T y_\delta \quad (3.77)$$

Come già accennato, Vainikko dimostra che l'algoritmo di inversione (3.77) è di regolarizzazione e soddisfa a particolari proprietà. Più precisamente:

Sia $K > 1$, costante. Allora, $\forall x \in N(T)^\perp$ e $\forall y_\delta \in Y$ tale che $\|Tx - y_\delta\| \leq \delta$, esiste un parametro di regolarizzazione $r(\delta, y_\delta)$, scelto per mezzo del criterio della discrepanza, tale che

(a)

$$\|Tx_r(\delta, y_\delta)(y_\delta) - y_\delta\| = \|Tg_r(\delta, y_\delta)(T^*T)^T y_\delta - y_\delta\| \leq K\delta$$

(b)

$$x_r(\delta, y_\delta)(y_\delta) \rightarrow x \quad (\delta \rightarrow 0)$$

(c)

$$r(\delta, y_\delta)^{\frac{1}{2}} \delta' \rightarrow 0 \quad (\delta \rightarrow 0)$$

Se in più $x \in M_{\mu, \rho}$, allora

(d)

$$\|x - x_r(\delta, y_\delta)(y_\delta)\| \leq C_3(\rho\delta^{2\mu})^{\frac{1}{2\mu+1}}$$

(e)

$$r(\delta, y_\delta) \leq C_4 \left(\frac{\delta}{\rho}\right)^{-\frac{2\mu}{2\mu+1}}$$

con C_3, C_4 dipendenti solo da μ .

Consideriamo l'operatore (3.77) con

$$y_\delta = Tx_\epsilon, \quad x = x_\epsilon, \quad \delta = \frac{b-1}{K}\epsilon$$

I risultati (a) - (e) di Vainikko garantiscono che esiste il parametro di regolarizzazione

$$r_\epsilon := r\left(\frac{b-1}{K}\epsilon, Tx_\epsilon\right)$$

tale che, per

$$x_\epsilon := x_{r_\epsilon}(Tx_\epsilon) = g_{r_\epsilon}(T^*T)^T T x_\epsilon \quad r_\epsilon \geq 0$$

(a')

$$\|Tx_\epsilon - Tx_\epsilon\| \leq K \frac{b-1}{K} \epsilon = (b-1)\epsilon$$

ossia $(x_\epsilon)_{\epsilon > 0}$ verifica (ii)

(b')

$$x_\epsilon \rightarrow x_\epsilon \quad (\epsilon \rightarrow 0)$$

ossia $(x_\epsilon)_{\epsilon > 0}$ verifica (iii)

(c')

$$r_\epsilon^{\frac{1}{2}} \frac{b-1}{K} \epsilon \rightarrow 0 \quad (\epsilon \rightarrow 0)$$

Se in più $x_\epsilon \in M_{\mu, \rho}$, allora

(d')

$$\|x_\epsilon - x_\epsilon\| \leq C_3 \left(\rho \left(\frac{b-1}{K}\epsilon\right)^{2\mu}\right)^{\frac{1}{2\mu+1}}$$

$$= C_3 \left(\frac{b-1}{K}\right)^{\frac{2\mu}{2\mu+1}} (\rho\epsilon^{2\mu})^{\frac{1}{2\mu+1}}$$

$$= C_2 (\rho\epsilon^{2\mu})^{\frac{1}{2\mu+1}}$$

ossia $(x_\epsilon)_{\epsilon > 0}$ verifica (iv) con $C_2 = C_3 \left(\frac{b-1}{K}\right)^{\frac{2\mu}{2\mu+1}}$.

(c')

$$r_\epsilon \leq C_5 \left(\frac{\epsilon}{\rho}\right)^{-\frac{2\mu_0}{2\mu_0+1}}$$

con $C_5 = C_4 \left(\frac{\rho-1}{\rho}\right)^{-\frac{2\mu_0}{2\mu_0+1}}$

Al fine di utilizzare le proprietà di convergenza che derivano dalla μ_0 -ottimalità dell'operatore $\{R_\epsilon\}_{\epsilon>0}$, occorre dimostrare (i), ossia che la famiglia $(z_\epsilon)_{\epsilon>0}$ è contenuta in $R((T^\mu T)^{\mu_0})$.

Verifichiamo questa condizione nel caso in cui $z_\epsilon \in M_{\mu_0}$, $0 < \mu \leq \mu_0$.
Si ha $z_\epsilon = (T^\mu T)^\mu z$ con $z \in X$, $\|z\| \leq \rho$, quindi

$$z_\epsilon = g_\epsilon \cdot (T^\mu T)^{\mu_0} (T^\mu T)^\mu z = g_\epsilon \cdot (T^\mu T)^{\mu_0} (T^\mu T)^{\mu_0} (T^\mu T)^\mu z \quad (3.78)$$

Verifichiamo che $\exists (z_\epsilon)_{\epsilon>0} \subset X$ tale che

$$z_\epsilon = (T^\mu T)^{\mu_0} z_\epsilon \quad \text{con } z_\epsilon \leq \rho_\epsilon \quad (3.79)$$

Consideriamo allora la famiglia di funzioni reali $(h_\epsilon)_{\epsilon>0}$

$$h_\epsilon : [0, a] \rightarrow \mathbb{R}, \quad \epsilon \geq 0$$

definita da

$$h_\epsilon(t) = t^{1+\mu-\mu_0} g_\epsilon \cdot (t)$$

e dimostriamo che

$$z_\epsilon = h_\epsilon (T^\mu T)^\mu z \quad (3.80)$$

soddisfa (3.79). Considerando (3.78) si ha

$$\begin{aligned} (T^\mu T)^{\mu_0} z_\epsilon &= (T^\mu T)^{\mu_0} (T^\mu T)^{1+\mu-\mu_0} g_\epsilon \cdot (T^\mu T)^\mu z \\ &= (T^\mu T) (T^\mu T)^\mu g_\epsilon \cdot (T^\mu T)^\mu z \\ &= g_\epsilon \cdot (T^\mu T) (T^\mu T) (T^\mu T)^\mu z = z_\epsilon \end{aligned}$$

che verifica l'appartenenza di z_ϵ a $R((T^\mu T)^{\mu_0})$ e quindi dimostra (i) nel caso in cui $z_\epsilon \in M_{\mu_0}$, $0 < \mu \leq \mu_0$.
Si osservi inoltre che

$$\sup_{t \in [0, a]} t^q g_\epsilon(t) \leq \left(\frac{1}{\tau}\right)^{q-1} \quad \tau \geq 0, \quad q \in (-\infty, +1]$$

La quale, applicata ponendo $q = 1 + \mu - \mu_0 \in (1 - \mu_0, 1]$, fornisce

$$h_\epsilon(t) \leq \tau^{-1-(1+\mu-\mu_0)} = \tau^{\mu_0-\mu}$$

L'ultima relazione consente di stimare ρ_ϵ , infatti

$$\rho_\epsilon = \|z_\epsilon\| = \|h_\epsilon (T^\mu T)^\mu z\| \leq \|h_\epsilon(t)\|_{C[0, a]} \|z\| \leq \rho \tau^{\mu_0-\mu} \quad (3.81)$$

Alla luce di quanto fatto all'inizio della dimostrazione in (i) - (vi) e (3.76), occorre solamente dimostrare che $(z_\epsilon)_{\epsilon>0}$ verifica (v), unica condizione non ancora considerata. Dalla proprietà (c') si ha

$$\begin{aligned} \rho_\epsilon &\leq \rho \tau^{\mu_0-\mu} \leq \rho \left(C_5 \left(\frac{\epsilon}{\rho}\right)^{-\frac{2\mu_0}{2\mu_0+1}}\right)^{\mu_0-\mu} \\ &= C_5^{\mu_0-\mu} \rho^{1+\frac{2\mu_0(\mu_0-\mu)}{2\mu_0+1}} \epsilon^{-\frac{2\mu_0(\mu_0-\mu)}{2\mu_0+1}} \\ &= C_6 (\rho^{2\mu_0+1} \epsilon^{2\mu-2\mu_0})^{\frac{1}{2\mu_0+1}} \end{aligned}$$

3.4 - Regolarizzazione mediante gradiente coniugato

con $C_6 = C_5^{\mu_0-\mu}$.
Possiamo così verificare (v)

$$\begin{aligned} \rho_\epsilon \epsilon^{2\mu_0} &\leq C_6 (\rho^{2\mu_0+1} \epsilon^{2\mu-2\mu_0})^{\frac{1}{2\mu_0+1}} \epsilon^{2\mu_0} \\ &= C_6 \rho^{\frac{2\mu_0}{2\mu_0+1}} \epsilon^{\frac{2\mu_0(2\mu-2\mu_0)}{2\mu_0+1} + 2\mu_0} \\ &= C_6 \rho^{\frac{2\mu_0}{2\mu_0+1}} \epsilon^{2\mu \frac{2\mu_0+1}{2\mu_0+1}} \\ &= C_6 (\rho \epsilon^{2\mu})^{\frac{2\mu_0}{2\mu_0+1}} \end{aligned}$$

dove qui la costante C_6 equivale a C_1 in (v).

La verifica di (v), insieme a quella di (i), (ii), (iii), (vi) precedente, permette, in accordo con (3.76), di dimostrare la tesi (I) del teorema, ossia che $(\bar{R}_\epsilon)_{\epsilon>0}$ è di ordine ottimo rispetto a $0 < \mu \leq \mu_0$.

Per la verifica di (II), occorre considerare $z_\epsilon \in N(T)^\perp$ e dimostrare (iv).

Generalissimo le relazioni determinate per la verifica di (I) al caso $\mu = 0$.
Si ha allora $\rho = \|z_\epsilon\|$; infatti l'operatore $(T^\mu T)^0$ corrisponde all'identità.

Dalla relazione (3.81), ossia $\rho_\epsilon \leq \rho \tau^{\mu_0-\mu}$, si ottiene

$$\rho_\epsilon \leq \rho \tau^{\mu_0} = \|z_\epsilon\| \tau^{\mu_0}$$

La verifica di (iv) risulta così conclusa poiché, per la proprietà (c'), osservando che $\frac{\rho-1}{K}$ è una costante positiva, si ha

$$\rho_\epsilon \epsilon^{2\mu_0} \leq \|z_\epsilon\| \tau^{\mu_0} \epsilon^{2\mu_0} = \|z_\epsilon\| \left(\tau \frac{\epsilon}{\rho}\right)^{2\mu_0} \rightarrow 0$$

per $\epsilon \rightarrow 0$.

R_ϵ risulta così anche algoritmo regolarizzante e la dimostrazione è conclusa. ■

Si precisa che la regolarizzazione mediante gradiente coniugato è basata sul criterio della discrepanza. Infatti il teorema 3.14 di Plato richiede $\{R_\epsilon\}$ di ordine ottimo e Nemirovskii in 3.13 ha dimostrato che il gradiente coniugato è di ordine ottimo se associato alla regola di arresto della discrepanza (3.33).

Abbiamo già visto che il parametro di regolarizzazione, ossia il numero di iterazioni da effettuare, è fortemente condizionato dal dato $y_\epsilon \in Y$, ossia $n = n(\epsilon, y_\epsilon)$ (cfr. capitolo 2.7). Ciò significa che, a parità di livello di errore $\epsilon > 0$, l'algoritmo può fornire soluzioni accettabili con un numero di iterazioni molto diverso a seconda del particolare dato. Un criterio che, sulla base della sola precisione che occorre raggiungere, non è in grado di fornire indicazioni sul numero di iterazioni da effettuare viene detto "a posteriori". Sappiamo che il criterio della discrepanza per il gradiente coniugato è "a posteriori". In pratica occorre calcolare la soluzione regolarizzata prima di poter valutare la sua affidabilità.

Concludiamo questo paragrafo mostrando che il gradiente coniugato non ammette scelte del parametro di regolarizzazione del tipo $n = n(\epsilon)$, che vengono dette "a priori", ma ogni scelta risulta necessariamente "a posteriori" (si consulti [11]).

Definizione 3.9 Un metodo di regolarizzazione $\{R_n\}_{n \geq 0}$, $R_n : Y \rightarrow X$, $n \geq 0$ è detto convergente se

$$\lim_{l \rightarrow +\infty} R_n(\varepsilon_l) y^j = T^l y$$

$\forall y \in R(T) \oplus R(T)^\perp$,

per ogni successione $(\varepsilon_l)_{l \in \mathbb{N}} \subset \mathbb{R}$ tale che $\varepsilon_l \rightarrow 0$

e per ogni successione $(y^j)_{j \in \mathbb{N}} \subset Y$ tale che $\|y - y^j\| \leq \varepsilon_l$.

In accordo con il teorema 3.10, il metodo del gradiente coniugato determina la soluzione in $\bar{n} < +\infty$ passi ogni volta che il dato y appartiene a

$$H_n = \{y \in Y : P_{R(T)^\perp} y - T^{\bar{n}} x_0 \text{ è combinazione lineare di } \bar{n} \text{ autovettori di } T^{\bar{n}}\}$$

Teorema 3.15 Se T è compatto con range di dimensione infinita, allora per $n \geq 1$, indicando con R_n l'operatore ottenuto effettuando n iterazioni del metodo del gradiente coniugato,

(i) R_n è continuo in $Y \setminus H_{n-1}$

(ii) R_n è discontinuo in H_{n-1} ,

cioè $\forall y \in H_{n-1}$ esiste una successione $(y^j)_{j \in \mathbb{N}}$, con $\lim_{l \rightarrow +\infty} y^j = y$, tale che

$$\lim_{l \rightarrow +\infty} \|R_n y^j - R_n y\| = +\infty$$

Dimostrazione.

(i) In $Y \setminus H_{n-1}$, l'algoritmo del gradiente coniugato non determina la soluzione dopo sole $(n-1)$ iterazioni, quindi $r_i \neq 0$ per $i = 0, \dots, n-1$.

Dalla definizione 3.5 dell'algoritmo segue che $\alpha_i, \beta_i, r_i, d_i, z_i$, per $i \leq n$, dipendono con continuità dal dato in $Y \setminus H_{n-1}$ (non si hanno "divisioni per zero").

(ii) In questo caso, per $y \in H_{n-1}$, il gradiente coniugato determina la soluzione

$$R_{n-1} y = R_n y = T^l y$$

Si consideri allora il sistema singolare $\{\mu_n, u_n, v_n\}$ di T (vedi appendice A).

Per $y \in H_{n-1}$ si definisca la successione

$$y^j = y + \mu_l^{-\frac{1}{2}} v_l$$

y^j differisce da y per il solo autovettore v_l di T^l , quindi, dato che $y \in H_{n-1}$, allora $y^j \in H_n$ e $R_n y^j = T^l y$.

Considerando che, per il teorema 1.10, $T^l v_l = \frac{1}{\mu_l} u_l$, $l = 0, 1, \dots$, si ha

$$\begin{aligned} \lim_{l \rightarrow +\infty} \|R_n y^j - R_n y\| &= \lim_{l \rightarrow +\infty} \|T^l y^j - T^l y\| \\ &= \lim_{l \rightarrow +\infty} \|T^l y + \mu_l^{-\frac{1}{2}} T^l v_l - T^l y\| \\ &= \lim_{l \rightarrow +\infty} \|\mu_l^{-\frac{1}{2}} \mu_l^{-1} u_l\| = \lim_{l \rightarrow +\infty} \mu_l^{-\frac{3}{2}} = +\infty \end{aligned}$$

ossia la successione y^j verifica la tesi.

Teorema 3.16 Se T è compatto con range di dimensione infinita, allora non esiste alcuna regola d'arresto $n = n(\varepsilon)$ "a priori" che renda l'algoritmo del gradiente coniugato convergente secondo la definizione 3.9.

Dimostrazione. Si consideri un generico $y \in H_{\bar{n}}$, con $\bar{n} \in \mathbb{N}^*$.

In accordo con le notazioni introdotte nella definizione 3.9, si consideri la successione di dati perturbati

$$y^j = y + \varepsilon_l v_l$$

con $\varepsilon_l = \mu_l^{-\frac{1}{2}} \rightarrow 0$ ($l \rightarrow +\infty$).

Si osservi che $y, (\varepsilon_l)_{l \in \mathbb{N}}, (y^j)_{j \in \mathbb{N}}$ soddisfano alla definizione 3.9; quindi se si verifica che la sequenza delle iterate non converge, la tesi risulta dimostrata.

Inoltre, per qualsiasi regola d'arresto $n^l(\varepsilon_l)$, si ha necessariamente $R_{n^l(\varepsilon_l)} y^j = R_{n^l+1} y^j = T^l y^j$; infatti $y^j \in H_{\bar{n}+1}$ e quindi il gradiente coniugato determina la soluzione esatta alla $(\bar{n}+1)$ -esima iterazione.

Analogamente al teorema precedente si ha

$$\begin{aligned} \lim_{l \rightarrow +\infty} \|R_{n^l(\varepsilon_l)} y^j - T^l y\| &= \lim_{l \rightarrow +\infty} \|R_{n^l+1} y^j - T^l y\| \\ &= \lim_{l \rightarrow +\infty} \|T^l y^j - T^l y\| \\ &= \lim_{l \rightarrow +\infty} \|T^l y - \varepsilon_l T^l v_l - T^l y\| \\ &= \varepsilon_l^{-1} \rightarrow +\infty \quad (l \rightarrow +\infty) \end{aligned}$$

che conclude la dimostrazione.

Si osservi che gli ultimi due teoremi sono conseguenza della proprietà del gradiente coniugato di essere metodo diretto ogni volta che il dato (o più precisamente la sua proiezione su $N(T)^\perp$) appartiene allo spazio di dimensione finita generato da un numero finito di autovettori di $T^* T$.

Ad esempio il metodo iterativo e lineare di Landweber-Fridman (cf. capitolo 2.7), che non soddisfa a tale proprietà, ammette regole d'arresto "a-priori" (per maggiori dettagli si consulti [11]).