UNIVERSITÀ DEGLI STUDI DI GENOVA

FACOLTÀ DI SCIENZE
MATEMATICHE FISICHE E NATURALI

CORSO DI LAUREA IN MATEMATICA

Tesi di laurea

REGOLARIZZAZIONE DI TIKHONOV GENERALE E SVD GENERALIZZATA

Relatore
Prof. Paola Brianzi

Correlatore
Dott. Fabio Di Benedetto

Candidato
Giuseppe Patané

Anno Accademico 1998/1999

Indice

Introduzione

Introduciamo l' argomento della tesi richiamando preliminarmente il concetto di problema diretto e inverso sottolineando le problematiche di quest' ultimo a cui abbiamo dato una risposta attraverso lo studio del metodo di regolarizzazione di Tikhonov generale e la SVD Generalizzata ($GSVD$).
Nell' ambito dei problemi reali vi è una distinzione naturale tra problema diretto e inverso. Per problema diretto si intende la determinazione del comportamento di un sistema fisico a partire dal suo stato iniziale e dalle leggi fisiche che ne regolano l' evoluzione.
Per problema inverso si intende la determinazione della configurazione iniziale del sistema a partire dalla conoscenza del suo stato negli istanti successivi (evoluzione a ritroso nel tempo). In questo ordine di idee, seguendo [1], possiamo dire che un problema diretto è "orientato lungo la sequenza causa-effetto mentre il corrispondente problema inverso è caratterizzato dall' inversione dei ruoli di questi ultimi.
Dal punto di vista matematico questa situazione può essere modellizzata considerando un' applicazione, nel nostro caso lineare,

$$K : X \to Y \tag{1}$$

tra due spazi X e Y di Hilbert definendo come

- problema diretto la determinazione, a partire dal dato $x \in X$, del risultato $y = Kx$;
- problema inverso la determinazione del valore $x \in X$ la cui immagine secondo K è $y = Kx$.

Una proprietà tipica dei problemi inversi è quella di poter risultare mal-posti (Hadamard) nel senso che la soluzione può non esistere, non essere unica o non

dipendere con continuità dai dati.
Poichè in ogni applicazione si hanno a disposizione dati finiti lo studio teorico del problema va affiancato da un approccio numerico opportunamente costruito in grado di sostituire i metodi standard dell' analisi numerica resi inefficaci dalle peculiarità dei problemi inversi. Questo è dovuto al fatto che questi metodi amplificano i disturbi (*noise*) di cui sono affetti i dati di ogni problema concreto generando una funzione altamente oscillante che altera la soluzione corrispondente al problema non affetto da noise.
Scopo di un qualunque metodo di regolarizzazione è pertanto quello di trattare i problemi inversi ristabilendo esistenza, unicità e continuità della soluzione dai dati.
Un metodo classico utilizzato nello studio dei problemi inversi è rappresentato dalla regolarizzazione di Tikhonov la cui validità è dovuta principalmente a un buon compromesso tra risultati teorici e numerici fondamentale per la trattazione di ogni problema concreto. Prenderemo questo metodo come punto di partenza proponendoci, come scopo della tesi, di studiarne una sua generalizzazione.
Spieghiamo più in dettaglio questa differenza tra i due metodi evidenziando il passaggio che porta dalla regolarizzazione di Tikhonov standard alla definizione di quella generale.
Il metodo di Tikhonov standard ha come fine la determinazione di una soluzione x che soddisfa le relazioni

- $\|Kx - y\|_Y \leq \epsilon$ che esprime l' approssimazione del dato entro un certo margine di errore;
- $\|x\|_X \leq E$ che individua un vincolo di regolarità sulla soluzione

minimizzando il funzionale associato

$$\tilde{G}_\lambda(x, y) = \|Kx - y\|_Y^2 + \lambda^2 \|x\|_X^2$$

dipendente dal parametro λ.
Il metodo di regolarizzazione di Tikhonov generale è ottenuto da quello standard imponendo nuovamente la condizione $\|Kx - y\|_Y \leq \epsilon$ e sostituendo il vincolo $\|x\|_X \leq E$ con la condizione $\Omega(x) = \|Lx\|_Y^2 \leq E$ dove

$$L : D(L) \subseteq X \to Y \tag{2}$$

è, per esempio, un operatore differenziale (cfr. def 2.2.1). Il funzionale $\tilde{G}(\lambda)$ è pertanto sostituito da

$$G_\lambda(x) = \|Kx - y\|_Y^2 + \lambda^2 \|Lx\|_Y^2 \quad \forall x \in D(L).$$

Nonostante la regolarizzazione di Tikhonov generale sia una naturale estensione di quella standard al caso di due operatori i problemi che ne derivano sono tutt' altro che banali. Tra questi i più importanti, con i quali ci siamo confrontati, sono:

- determinazione delle condizioni di esistenza e unicità della soluzione regolarizzata;
- trattazione teorica che fornisca condizioni di convergenza della soluzione regolarizzata e un suo sviluppo in funzioni singolari;
- trattazione numerica di effettiva applicabilità dovuta all' assenza di routine in grado di sostituire, in modo altrettanto valido, la SVD utilizzata nel caso standard.

Queste problematiche sono state affrontate proponendoci l' estensione dei risultati più caratteristici della regolarizzazione di Tikhonov standard al caso generale mantenendo una sostanziale simmetria tra approccio teorico e numerico motivato dalle seguenti considerazioni.
L'approccio teorico permetterà una caratterizzazione dei risultati di convergenza più forte dovuta alla possibiltà di introdurre metriche, e quindi topologie, più adatte al problema in esame.
L' approccio numerico, imposto dalla necessità di risolvere problemi concreti tipici della Matematica Applicata in cui i dati a disposizione sono di tipo discreto, è stato affrontato utilizzando la $GSVD$ delle matrici $A \in M_{m,n}(\mathbb{R})$, $L \in M_{p,n}(\mathbb{R})$ di discretizzazione di (1) e (2) rappresentata da [1]:

$$A = U\Sigma X^{-1} = U \begin{pmatrix} \Sigma_p & 0 \\ 0 & I_{n-p} \end{pmatrix} X^{-1}, \quad L = VMX^{-1} = V(M_p, 0)X^{-1},$$

con $U \in M_{m,n}(\mathbb{R})$ e $V \in M_{p,p}(\mathbb{R})$ matrici ortogonali e $X \in Gl_n(\mathbb{R})$.
Le matrici Σ_p e M_p sono matrici diagonali di ordine p:

$$\Sigma_p = diag(\sigma_1, ..., \sigma_p), \quad M_p = diag(\mu_1, ..., \mu_p)$$

[1] $Gl_n(\mathbb{R}) = \{Z \in M_{n,n}(\mathbb{R}) : det(Z) \neq 0\}$.

in cui gli elementi diagonali sono ordinati come segue

$$0 \leq \sigma_1 \leq ... \leq \sigma_p \leq 1, \quad 1 \geq \mu_1 \geq ... \geq \mu_p \geq 0,$$

e normalizzati in modo tale che

$$\sigma_i{}^2 + \mu_i{}^2 = 1 \quad i = 1, ..., p.$$

Con questo ordine di idee abbiamo così suddiviso lo studio del metodo di regolarizzazione.
Nel capitolo 1 introdurremo la decomposizione $GSVD$ utilizzata per lo studio discreto del metodo di regolarizzazione di Tikhonov generale.
Nel capitolo 2 presenteremo uno studio teorico della regolarizzazione in forma generale: la struttura che gli abbiamo dato è fatta al fine di renderlo il più compatibile possibile con i capitoli successivi.
Nel capitolo 3 abbiamo trattato lo stesso problema dal punto di vista discreto estendendo alcuni risultati della regolarizzazione standard al caso generale.
I risultati, ottenuti in modo autonomo, confermano la possibilità di unificare la trattazione numerica e teorica confermando la validità della trattazione e delle scelte fatte nel capitolo 2.
Presenteremo nel capitolo 4 un legame tra regolarizzazione standard e generale verificando che è possibile ridurre il problema generale, attraverso un opportuno cambiamento di variabili, a uno standard. Utilizzeremo poi questi risultati per proporre una routine [6] alternativa a quella [24] presentata nel capitolo 1 e di uguale efficacia per la trattazione del nostro problema.
Utile lo scambio di idee con il Prof. P.C.Hansen (Institute of Mathematical Modelling, University of Denmark, Lyngby) che ci ha permesso alcune osservazioni di carattere numerico sull' algoritmo *gsvd*, di cui è autore, implementato nell' ultima versione del Matlab [20]. Dallo studio dell' applicabilità di questo algoritmo alla regolarizzazione generale abbiamo inoltre ottenuto un nuovo approccio alla $TGSVD$ alternativo e concorde a quello proposto in [11].
Concludendo abbiamo dimostrato che è possibile utilizzare la regolarizzazione di Tikhonov generale come metodo di risoluzione dei problemi inversi al pari di quella standard.

Capitolo 1

Algoritmo $GSVD$

Introduciamo in questo capitolo l' algoritmo per il calcolo della $GSVD$ di (A, L). Nonostante la sua stabilità numerica non sia stata ancora dimostrata in maniera definitiva abbiamo riportato, nella sezione 1.3, alcune considerazioni che motivano le scelte fatte nell' algoritmo evidenziando le difficoltà connesse con la sua analisi completa.
Le prove da noi eseguite, forzando le dimensioni e soprattutto il condizionamento delle matrici A, L, forniscono buoni risultati di stabilità rispetto a perturbazioni sulle matrici e una capacità di ricostruzione altrettanto valida. Alcune osservazioni di particolare interesse relative all' applicabilità della $GSVD$ alla regolarizzazione di Tikhonov generale verranno fornite alla fine del capitolo 4.

1.1 Decomposizione CS (Coseno-Seno)

Sia [1]

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$$

una matrice ortogonale:

$$Q^T Q = Q_1^T Q_1 + Q_2^T Q_2 = I_p. \tag{1.1}$$

con $Q_1 \in M_{n_1,p}(\mathbb{R})$, $Q_2 \in M_{n_2,p}(\mathbb{R})$ e $n_1 \geq p$.
La decomposizione CS di Q consiste nel determinare tre matrici ortogonali

$$U_1 \in M_{n_1,n_1}(\mathbb{R}),\ U_2 \in M_{n_2,n_2}(\mathbb{R}),\ V \in M_{p,p}(\mathbb{R})$$

[1] Faremo uso di diverse fattorizzazioni tipiche dell' Analisi Numerica per la cui trattazione dettagliata rimandiamo a [3,7].

tali che

$$\begin{cases} U_1^T Q_1 V = C = diag(c_1, c_2, ..., c_p) \in M_{n_1,p}(\mathbb{R}) \\ U_2^T Q_2 V = S = diag(s_1, s_2, ..., s_q) \in M_{n_2,p}(\mathbb{R}) \end{cases} \tag{1.2}$$

con $q = min\{p, n_2\}$. Possiamo supporre che i valori singolari $\{c_i\}_{i=1}^p$, $\{s_i\}_{i=1}^q$ siano ordinati come segue:

$$0 \leq c_1 \leq c_2 \leq ... \leq c_q \leq c_{q+1} = ... = c_p = 1,$$

$$1 \geq s_1 \geq s_2 \geq ... \geq s_q \geq 0.$$

Osserviamo che (1.2) puó essere riscritta nella forma

$$\begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}^T \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} V = \begin{pmatrix} C \\ S \end{pmatrix} \in M_{n_1+n_2,p}(\mathbb{R}). \tag{1.3}$$

Da (1.2) otteniamo le relazioni

$$\begin{cases} Q_1 = U_1 C V^T \\ Q_2 = U_2 S V^T \end{cases}$$

che sostituite in (1.1) ci permettono di scrivere:

$$C^T C + S^T S = I_p$$

e quindi

$$\begin{cases} c_i^2 + s_i^2 = 1 \quad \forall i = 1, 2, ..., q \\ |c_i| = 1 \quad \forall i = q+1, ..., p. \end{cases}$$

Da queste otteniamo che i valori singolari di Q_1 e Q_2 rappresentano i coseni e i seni di opportuni angoli motivando il nome di tale decomposizione.
Andiamo ora a studiare l' algoritmo [24] per la CS Decomposizione che verrà poi utilizzato per determinare, in modo semplice e diretto, quello per il calcolo della $GSVD$ di (A, L) introdotto nel seguente paragrafo.

1° passo

Calcolo $U_2 \in M_{n_2,n_2}(\mathbb{R})$ e $V \in M_{p,p}(\mathbb{R})$ matrici ortogonali tali che

$$U_2^T Q_2 V = (0, \Delta), \quad q = min\{n_2, p\} \tag{1.4}$$

con $\Delta = diag(\delta_1, \delta_2, ..., \delta_q) \in M_{n_2,q}(\mathbb{R})$ (ovvero determino la SVD per la matrice Q_2 ordinando nel modo prescelto i valori singolari). Sia k tale che

$0 \leq \delta_1 \leq \delta_2 \leq ...\delta_k \leq \frac{1}{\sqrt{2}} \leq \delta_{k+1} \leq ... \leq \delta_q$: la scelta di questo parametro è fatta al fine di migliorare la stabilità dell' algoritmo e sarà in parte chiarita al passo 5. Per una trattazione più dettagliata di questo aspetto si consulti [24].

2° passo

Aggiorno Q_1 e Q_2 ponendo

$$\begin{cases} Q_1 := Q_1 V \\ Q_2 := U_2^T Q_2 V = (0, \Delta) \end{cases}. \tag{1.5}$$

3° passo

Utilizzando la fattorizzazione QR di Q_1 calcolo $U_1 \in M_{n_1,n_1}(\mathbb{R})$ ortogonale tale che:

$$U_1^T Q_1 = \begin{pmatrix} R \\ 0 \end{pmatrix} \in M_{n_1,p}(\mathbb{R}) \tag{1.6}$$

con $R \in M_{p,p}(\mathbb{R})$ matrice triangolare superiore e elementi diagonali non negativi.

Dimostriamo che $R^T R = diag(1, 1, ..., 1, 1 - \delta_1^2, ..., 1 - \delta_q^2) \in M_{p,p}(\mathbb{R})$: a tal fine utilizzeremo, solo per questa verifica, i simboli $\tilde{Q}_1$ e $\tilde{Q}_2$ per indicare i valori di Q_1 e Q_2 aggiornati (come fatto al passo 2)

$$\tilde{Q}_1 = Q_1 V \quad \text{e} \quad \tilde{Q}_2 = U_2^T Q_2 V.$$

Abbiamo allora, utilizzando (1.1), (1.5) e (1.6)

$$R^T R = (U_1^T \tilde{Q}_1)^T (U_1^T \tilde{Q}_1) = \tilde{Q}_1^T \tilde{Q}_1 = (Q_1 V)^T (Q_1 V) = V^T Q_1^T Q_1 V =$$

$$V^T (I_p - Q_2^T Q_2) V = V^T V - V^T Q_2^T Q_2 V = I_p - V^T Q_2^T Q_2 V = I_p - (Q_2 V)^T (Q_2 V) =$$

$$I_p - (\, 0, U_2\Delta \,)^T \, (\, 0, U_2\Delta \,) = I_p - \begin{pmatrix} 0 \\ \Delta^T U_2^T \end{pmatrix} (\, 0, U_2\Delta \,) = I_p - \begin{pmatrix} 0 & 0 \\ 0 & \Delta^T U_2^T U_2 \Delta \end{pmatrix} =$$

$$I_p - \begin{pmatrix} 0 & 0 \\ 0 & \Delta^T \Delta \end{pmatrix} = diag(1, ..., 1, 1 - \delta_1^2, ..., 1 - \delta_q^2).$$

4° passo

Aggiorniamo nuovamente Q_1 ponendo

$$Q_1 := U_1^T Q_1, \quad \gamma_i = 1 - \delta_i^2 \ \forall i = 1, 2, ..., q.$$

Dalla precedente uguaglianza otteniamo, essendo R triangolare e $1 - \delta_i^2 > 0 \ \forall i = 1, 2, ..., k$ [2],

$$\begin{pmatrix} R \\ 0 \end{pmatrix} = U_1^T Q_1 =: Q_1 = \begin{pmatrix} I_{p-q} & 0 & 0 \\ 0 & diag(\gamma_1, ..., \gamma_k) & 0 \\ 0 & 0 & R_1 \\ 0 & 0 & 0 \end{pmatrix} \in M_{n_1,p}(\mathbb{R})$$

con $R_1^T R_1 = diag(\gamma_{k+1}^2, ..., \gamma_q^2) \in M_{q-k,q-k}(\mathbb{R})$.

5° passo

Calcolo la SVD di R_1 determinando $\tilde{U}_1 \in M_{q-k,q-k}(\mathbb{R})$ e $\tilde{V} \in M_{q-k,q-k}(\mathbb{R})$ ortogonali tali che

$$\tilde{U}_1^T R_1 \tilde{V} = diag(\gamma_{k+1}, ..., \gamma_q) \tag{1.7}$$

e aggiorno le matrici nel seguente modo:

$$U_1 := U_1 diag(I_{p-q+k}, \tilde{U}_1, I_{n_1-p})$$

$$V := V diag(I_{p-q+k}, \tilde{V})$$

$$Q_1 := diag(I_{p-q+k}, \tilde{U}_1^T, I_{n_1-p}) Q_1 diag(I_{p-q+k}, \tilde{V})$$

$$Q_2 := Q_2 diag(I_{p-q+k}, \tilde{V}).$$

Osserviamo che per poter scrivere (1.7) abbiamo usato la relazione ottenuta al passo precedente che fornisce i valori singolari di R_1. A questo punto dell' algoritmo abbiamo

$$Q_2 = Q_2 diag(I_{p-q+k}, \tilde{V}) = Q_2 \begin{pmatrix} I_{p-q+k} & 0 \\ 0 & \tilde{V} \end{pmatrix} = (0, \Delta) \begin{pmatrix} I_{p-q+k} & 0 \\ 0 & \tilde{V} \end{pmatrix} =$$

$$(0, \Delta) \begin{pmatrix} I_{p-q} & 0 & 0 \\ 0 & I_k & 0 \\ 0 & 0 & \tilde{V} \end{pmatrix} = \begin{pmatrix} 0 & diag(\delta_1, ..., \delta_k) & 0 \\ 0 & 0 & dig(\delta_{k+1}, ..., \delta_q) \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} I_{p-q} & 0 & 0 \\ 0 & I_k & 0 \\ 0 & 0 & \tilde{V} \end{pmatrix} =$$

[2] Il parametro k è quello definito al passo 1.

$$\begin{pmatrix} 0 & diag(\delta_1,...,\delta_k) & 0 \\ 0 & 0 & diag(\delta_{k+1},...,\delta_q)\tilde{V} \\ 0 & 0 & 0 \end{pmatrix}.$$

Posto allora

$$W = diag(\delta_{k+1},...,\delta_q)\tilde{V} \in M_{q-k,q-k}(\mathbb{R})$$

si ha

$$W^TW = (diag(\delta_{k+1},...,\delta_q\tilde{V})^T(diag(\delta_{k+1},...,\delta_q)\tilde{V}) = diag(\delta_{k+1}^2,...,\delta_q^2) =$$

$$I_{q-k} - diag(1-\delta_{k+1}^2,...,1-\delta_q^2) = I_{q-k} - diag(\gamma_{k+1}^2,...,\gamma_q^2).$$

Poichè $\sigma_{min}(W) = \delta_{k+1} \geq \frac{\sqrt{2}}{2}$ è possibile diagonalizzare W, in modo numericamente stabile, normalizzando le sue colonne: calcoliamo pertanto $\tilde{U}_2 \in M_{q-k,q-k}(\mathbb{R})$ ortogonale tale che $\tilde{U}_2^TW$ sia triangolare superiore.
Poniamo quindi

$$Q_2 := diag(I_k, \tilde{U}_2^T I_{n_2-q})Q_2$$

$$U_2 := U_2 diag(I_k, \tilde{U}_2, I_{n_2-q}).$$

Dalle relazioni precedenti otteniamo

$$Q_1 = diag(I_{p-q+k}, \tilde{U}_1^T, I_{n_1-p})Q_1 diag(I_{p-q+k}, \tilde{V}) =$$

$$\begin{pmatrix} I_{p-q+k} & 0 & 0 \\ 0 & \tilde{U_1}^T & 0 \\ 0 & 0 & I_{n_1-p} \end{pmatrix} \begin{pmatrix} I_{p-q} & 0 & 0 \\ 0 & diag(\gamma_1,...,\gamma_k) & 0 \\ 0 & 0 & R_1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} I_{p-q+k} & 0 \\ 0 & \tilde{V} \end{pmatrix} =$$

$$\begin{pmatrix} I_{p-q} & 0 & 0 & 0 \\ 0 & I_k & 0 & 0 \\ 0 & 0 & \tilde{U}_1^T & 0 \\ 0 & 0 & 0 & I_{n_1-p} \end{pmatrix} \begin{pmatrix} I_{p-q} & 0 & 0 & 0 \\ 0 & diag(\gamma_1,...,\gamma_k) & 0 & 0 \\ 0 & 0 & R_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} I_{p-q} & 0 & 0 & 0 \\ 0 & I_k & 0 & 0 \\ 0 & 0 & \tilde{V} & 0 \end{pmatrix} =$$

$$\begin{pmatrix} I_{p-q} & 0 & 0 & 0 \\ 0 & diag(\gamma_1,...,\gamma_k) & 0 & 0 \\ 0 & 0 & \tilde{U}_1^T R_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} I_{p-q} & 0 & 0 & 0 \\ 0 & I_k & 0 & 0 \\ 0 & 0 & \tilde{V} & 0 \end{pmatrix} =$$

$$\begin{pmatrix} I_{p-q} & 0 & 0 & 0 \\ 0 & diag(\gamma_1,...,\gamma_k) & 0 & 0 \\ 0 & 0 & \tilde{U}_1^T R_1 \tilde{V} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} I_{p-q} & 0 & 0 & 0 \\ 0 & diag(\gamma_1,...,\gamma_k) & 0 & 0 \\ 0 & 0 & diag(\gamma_{k+1},...,\gamma_q) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

e

$$Q_2 = \begin{pmatrix} I_k & 0 & 0 \\ 0 & \tilde{U}_2^T & 0 \\ 0 & 0 & I_{n_2-q} \end{pmatrix} \begin{pmatrix} 0 & diag(\delta_1,...,\delta_k) & 0 \\ 0 & 0 & W \\ 0 & 0 & 0 \end{pmatrix} =$$
$$\begin{pmatrix} 0 & diag(\delta_1,...,\delta_k) & 0 \\ 0 & 0 & \tilde{U}_2^T W \\ 0 & 0 & 0 \end{pmatrix}.$$

Poichè $\tilde{U}_2^T W$ è triangolare superiore e

$$Q_2^T Q_2 = I_p - Q_1^T Q_1 = diag(1,...,1,1-\gamma_1^2,...,1-\gamma_q^2)$$

otteniamo $\tilde{U}_2^T W = diag(\delta_{k+1},...,\delta_q)$ e quindi

$$Q_2 = \begin{pmatrix} 0 & diag(\delta_1,...,\delta_k) & 0 \\ 0 & 0 & diag(\delta_{k+1},...,\delta_q) \\ 0 & 0 & 0 \end{pmatrix}.$$

A questo punto le matrici Q_1 e Q_2, ridotte in forma diagonale, vanno riportate nella forma (1.2) applicando opportunamente le matrici elementari; fatto ciò basta porre $C = Q_1$ e $S = Q_2$.

1.2 SVD Generalizzata

Prima di introdurre la SVD Generalizzata richiamiamo brevemente la SVD: riprenderemo questa decomposizione nel capitolo 3 in cui evidenzieremo la sua utilità per la risoluzione del metodo di regolarizzazione di Tikhonov standard.

Proposizione 1.2.1 *Sia $A \in M_{m,n}(\mathbb{R})$. Allora esistono due matrici ortogonali $U \in M_{m,m}(\mathbb{R})$ e $V \in M_{n,n}(\mathbb{R})$ tali che*

$$U^T A V = diag(\sigma_1,...,\sigma_p), \quad p = min\{m,n\}$$

con $\sigma_1 \geq ... \geq \sigma_p \geq 0$.

Dimostrazione. Si consulti [7]. ∎

Date due matrici $A \in M_{n_1,t}(\mathbb{R})$, $n_1 \geq t$, $L \in M_{n_2,t}(\mathbb{R})$, dimostriamo che esistono $U \in M_{n_1,n_1}(\mathbb{R})$, $V \in M_{n_2,n_2}(\mathbb{R})$ ortogonali e $X \in Gl_t(\mathbb{R})$ tali che

$$\begin{cases} A = U\Sigma X^T \\ L = VMX^T \end{cases} \tag{1.8}$$

con

$$\Sigma = diag(\sigma_1, ..., \sigma_t) \in M_{n_1,t}(\mathbb{R}), M = diag(\mu_1, ..., \mu_r) \in M_{n_2,t}(\mathbb{R})$$

e $r = min\{t, n_2\}$.
A tale scopo calcoliamo la SVD di $\begin{pmatrix} A \\ L \end{pmatrix}$:

$$\begin{pmatrix} A \\ L \end{pmatrix} = Q\Delta Z^T \tag{1.9}$$

dove $Q \in M_{n_1+n_2,n_1+n_2}(\mathbb{R})$, $\Delta \in M_{n_1+n_2,t}(\mathbb{R})$, $Z \in M_{t,t}(\mathbb{R})$.
Detto $p = rank(\begin{pmatrix} A \\ L \end{pmatrix})$ poniamo $m = n_1 + n_2 - p$ e partizioniamo le matrici precedenti come segue:

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$$

$$\Delta = \begin{pmatrix} \Delta_p & 0 \\ 0 & 0 \end{pmatrix}$$

$$Z = (Z_1, Z_2)$$

con $Q_{11} \in M_{p,p}(\mathbb{R})$, $Q_{22} \in M_{m,m}(\mathbb{R})$, $\Delta_p \in M_{p,p}(\mathbb{R})$, $Z_1 \in M_{t,p}(\mathbb{R})$.
Per l' ortogonalità di Q otteniamo

$$Q_{11}^T Q_{11} + Q_{21}^T Q_{21} = I_p$$

e quindi, applicando la CS Decomposizione a

$$\begin{pmatrix} Q_{11} \\ Q_{21} \end{pmatrix} \in M_{n_1+n_2,p}(\mathbb{R}),$$

si deducono le relazioni

$$\begin{cases} Q_{11} = U_1 C V^T \\ Q_{21} = U_2 S V^T \end{cases}.$$

Pertanto da (1.9) si ha, per l' ortogonalità di Z,

$$\begin{pmatrix} A \\ L \end{pmatrix} Z = Q\Delta = \begin{pmatrix} Q_{11}\Delta_p & 0 \\ Q_{21}\Delta_p & 0 \end{pmatrix} = \begin{pmatrix} U_1 C V^T \Delta_p & 0 \\ U_2 S V^T \Delta_p & 0 \end{pmatrix}$$

e quindi

$$\begin{pmatrix} A \\ L \end{pmatrix} = \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} \begin{pmatrix} C & 0 \\ S & 0 \end{pmatrix} \begin{pmatrix} V^T\Delta_p & 0 \\ 0 & W \end{pmatrix} Z^T$$

con $W \in M_{t-p,t-p}(\mathbb{R})$ matrice arbitraria.
Scelta W, non singolare, affinchè valga (1.8) poniamo

$$U = U_1,\ \Sigma = (\,C, 0\,)$$

$$V = U_2,\ M = (\,S, 0\,)$$

$$X^T = \begin{pmatrix} V^T\Delta_p & 0 \\ 0 & W \end{pmatrix} Z^T \quad (det(X) = \left(\prod_{j=1}^{p} \sigma_j\right) det(W) \neq 0).$$

L' algoritmo precedente, che permette di calcolare la $GSVD$ utilizzando decomposizioni standard, è definito in ogni sua parte tranne che per la scelta della matrice W.
Indichiamo con K_2 il condizionamento di una matrice rispetto alla norma 2.
Supponendo $\Sigma_p = diag(\sigma_1, ..., \sigma_p)$ con $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p$ abbiamo
$K_2(X) = K_2(X^T) = K_2(\begin{pmatrix} V^T\Sigma_p & 0 \\ 0 & W \end{pmatrix} Z^T) = max\{K_2(V^T\Sigma_p Z^T), K_2(WZ^T)\} =$
$max\{K_2(\Sigma), K_2(W)\} \geq K_2(\Sigma) = \frac{\sigma_1}{\sigma_p}$
e quindi la scelta migliore è data da

$$W = \sigma I \ \text{ con } \ \sigma_p < \sigma < \sigma_1$$

poichè in tal caso

$$K_2(X) = \frac{\sigma_1}{\sigma_p}.$$

1.3 Stabilità della CS Decomposizione e della SVD Generalizzata

Per poter confrontare due diversi algoritmi per il calcolo della decomposizione CS dobbiamo introdurre il concetto di stabilità numerica ad esso associato. Tale definizione deve essere formulata rileggendo le precedenti uguaglianze a

meno della precisione di macchina ϵ.
L' algoritmo per la CS Decomposizione è stabile se

$$\|\tilde{U}_1^T U_1 - I_{n_1}\|_2 \approx \epsilon \ (\tilde{U}_1 \text{ ortogonale rispetto a } \epsilon \text{ e } \| \ \|_2),$$

$$\|\tilde{U}_2^T U_2 - I_{n_2}\|_2 \approx \epsilon \ (\tilde{U}_2 \text{ ortogonale rispetto a } \epsilon \text{ e } \| \ \|_2),$$

$$\|V^T V - I_p\|_2 \approx \epsilon \ (V \text{ ortogonale rispetto a } \epsilon \text{ e } \| \ \|_2),$$

$$\tilde{C} = diag(\tilde{c_1}, ..., \tilde{c_p}) = \tilde{U_1}^T (Q_1 + E_1)\tilde{V} \text{ con } \|E_1\|_2 \approx \epsilon \|Q_1\|_2,$$

$$\tilde{S} = diag(\tilde{s_1}, ..., \tilde{s_q})\tilde{U_2}^T (Q_2 + E_2)\tilde{V} \text{ con } \|E\|_2 \approx \epsilon \|Q_2\|_2.$$

Premettiamo un risultato che ci servirà nel seguito.

Teorema 1.3.1 *Sia $X \in M_{m,k}(\mathbb{R})$, $k = rank(X)$ e*

$$X = (\, x_1, x_2, ..., x_k \,)$$

la corrispondente decomposizione per colonne. Supponiamo $X^T X = D^2 + E$ con $D = diag(\|x_1\|_2, \|x_2\|_2, ..., \|x_k\|_2)$. Detta $X = QR$ la sua fattorizzazione QR si ha che

$$|r_{ij}| \leq min\{\|x_j\|_2, \frac{\|E\|_2}{\sigma_{min}(X_j)}\} \ \ \forall j > i.$$

($r_{ij} = 0 \ i < j$) dove X_j è la sottomatrice principale di X e $\sigma_{min}(X_j)$ il corrispondente valore singolare più piccolo.

Dimostrazione. Poniamo $G = X^T X$ e R_i la sottomatrice principale di R (i.e $(R_i)_{p,q} = r_{pq} \ \ p, q = 1, 2, ..., i$): allora $G = X^T X = (QR)^T (QR) = R^T R$ e quindi $\forall j = 1, 2, ..., k$ si ha

$$(X^T X)_j = (R^T R)_j$$

ovvero $X^T x_j = R^T r_j$. Dalle precedenti relazioni otteniamo $\forall i = 1, 2, ..., k$, $\forall j = i + 1, ..., k$

$$R_i^T r_j = X_i^T x_j$$

e cioè

$$R_i^T \begin{pmatrix} r_{1j} \\ r_{2j} \\ . \\ . \\ r_{ij} \end{pmatrix} = X_i^T x_j.$$

Poichè per ipotesi $rank(X) = k$ si ha
$k = rank(X) = rank(QR) = rank(R) = rank(R^T R) \iff det(R^T R) \neq 0 \iff$ (essendo R triangolare superiore) $det(R^T R) = \prod_{j=1}^{k} r_{jj}^2 \neq 0 \iff r_{jj} \neq 0 \ \forall j = 1, 2, ..., k$.
Da tale relazione deduciamo che $R_i \in Gl_i(\mathbb{R}) \ \forall i = 1, 2, ..., k$ e quindi, invertendo R_i^T,

$$\begin{pmatrix} r_{1j} \\ r_{2j} \\ . \\ . \\ r_{ij} \end{pmatrix} = R_i^{-T} X_i^T x_j. \tag{1.10}$$

Passando alla norma 2 in (1.10), si deduce

$$|r_{ij}| \leq (r_{1j}^2 + ... + r_{ij}^2)^{1/2} = \|R_i^{-T} X_i^T x_j\|_2 \leq \|R_i^{-T}\|_2 \|X_i^T x_j\|_2 = \frac{1}{\sigma_{min}(R_i)} \|X_i^T x_j\|_2.$$

Poichè $\sigma_{min}(R_i) = \sigma_{min}(X_i)$ e $\|X_i^T x_j\|_2 \leq \|E\|_2$ otteniamo dalla relazione precedente

$$|r_{ij}| \leq \frac{\|E\|_2}{\sigma_{min}(X_i)}.$$

Utilizzando la maggiorazione $|r_{ij}| \leq (r_{i1}^2 + ... + r_{jj})^{1/2} = \|x_j\|_2$ si ottiene la tesi.
■

Con queste premesse vogliamo fare alcune osservazioni relative all' algoritmo per il calcolo della CS Decomposizione localizzando la nostra attenzione sul passo 3 che ne costituisce la parte più delicata. Infatti nelle altre parti dell' algoritmo si fa sostanzialmente uso della SVD la cui stabilità numerica è stata dimostrata dettagliatamente in [7].
Al passo 2 abbiamo introdotto la matrice $Q_1 V$, che da ora in poi indicheremo con X, di cui abbiamo calcolato (al passo 3) la fattorizzazione QR per poter determinare la matrice triangolare superiore R che nei passi successivi è stata ridotta in forma diagonale.
Indichiamo con $\tilde{X}$ la versione numerica di tale matrice e con $\tilde{R}$ quella di $\begin{pmatrix} R \\ 0 \end{pmatrix}$.
Allora, seguendo lo sviluppo fatto in [24], si ha

$$\tilde{X}^T \tilde{X} = diag(1, 1, ..., 1, 1 - \tilde{\delta}_1^2, ..., 1 - \tilde{\delta}_q^2) + E$$

con $\{\tilde{\delta}_i\}_{i=1}^{q}$ valori singolari di Q_2 effettivamente calcolati e $\|E\|_2 \approx \epsilon$.
La relazione (1.6) può essere pertanto riscritta nella forma

$$U_1^T(\tilde{X} + F) = \tilde{R}$$

dove U_1 è ortogonale e $\|F\|_2 \approx \epsilon\|\tilde{X}\|_2$.
Dal teorema 1.3.1 segue che $\forall j = i+1, ..., p$

$$|\tilde{r}_{ij}| \leq min\{\|\tilde{x}_j\|_2; \frac{\|E\|_2}{\sigma_{min}(\tilde{X}_j)}\} \approx min\{\|\tilde{x}_j\|_2^2, \frac{\epsilon}{\sqrt{1-\tilde{\delta}_i^2}}\}$$

e pertanto, essendo $\delta_{k+1} \geq \frac{\sqrt{2}}{2}$, si ha

$$|r_{ij}| \approx \epsilon\|R\|_2 \;\; i = 1, 2, ..., p-q+k, \quad j = i+1, ..., p.$$

Questa scelta di k in (1.4) permette di lavorare con una matrice $\tilde{R}$ effettivamente triangolare superiore garantendo la possibilità di applicare con successo la fattorizzazione QR in (1.6).
L' algoritmo per la CS Decomposizione soddisfa le proprietà di stabilità date all' inizio del paragrafo ed è pertanto numericamente stabile. La stabilità della $GSVD$, anche se non dimostrata, è essenzialmente dovuta all' uso di algoritmi quali la SVD e la CS Decomposizione che godono di buone proprietà di stabilità. Anche se utilizzeremo la $GSVD$ come decomposizione base per l' analisi del metodo di regolarizzazione di Tikhonov generale vogliamo evidenziare brevemente alcune proprietà dei valori singolari generalizzati

$$\gamma_i = \frac{\sigma_i}{\mu_i} \;\; \forall i = 1, 2, ..., p = rank(L)$$

e un legame tra $GSVD$ e SVD.

(1) $GSVD$ come generalizzazione della SVD:
Le matrici U, Σ, V utilizzate nella $GSVD$ di (A, L) sono diverse dalle corrispondenti matrici relative alla SVD di A. Se però L è la matrice identica le matrici U e V relative alla $GSVD$ di (A, L) sono uguali, a meno dell' ordine delle colonne, alle matrici U, V della SVD di A. Analogamente i valori singolari di A sono uguali ai valori singolari generalizzati γ_i di (A,I).
Va sottolineato però che in generale non vi è alcun legame tra valori singolari/vettori generalizzati e valori singolari/vettori ordinari.

(2) Legame tra lo spettro di A e spettro di (A, L):
Data una matrice $A \in M_{m,n}(\mathbb{R})$ si ha, utilizzando la proprietà di invarianza ortogonale del determinante e la SVD di A, che

$$det(A^TA - \lambda^2 I) = 0 \Longleftrightarrow \lambda \in \{\sigma_1, ..., \sigma_p\}.$$

Ci chiediamo ora se è possibile estendere questa proprietà dei valori singolari al caso dei valori generalizzati di (A,L).
In analogia con il caso precedente definiamo autovalori generalizzati di (A, L) gli elementi dell' insieme

$$\lambda(A, L) = \{\lambda : \lambda \geq 0, det(A^TA - \lambda^2 L^TL) = 0\}$$

con $L \in M_{p,n}(\mathbb{R})$.
Allora da (1.8) deduciamo che
$det(A^TA - \lambda^2 L^TL) = det[X(\Sigma^T\Sigma - \lambda^2 M^TM)X^T] =$
$det(X)^2 det(\Sigma^T\Sigma - \lambda^2 M^TM) = det(X)^2 \prod_{j=1}^{rank(L)}(\sigma_j^2 - \lambda^2\mu_j^2) \prod_{j=rank(L)+1}^{n} \sigma_j^2$
e quindi

$$\lambda(A, L) = \begin{cases} \mathbb{R}^+ & \text{se } \sigma_j = 0 \text{ per qualche } j \in \{rank(L) + 1, ..., n\} \\ \{\gamma_j, j = 1, 2, ..., r\} & \text{altrimenti} \end{cases}.$$

Osserviamo che $\lambda(A, L) = \mathbb{R}^+ \Longleftrightarrow \sigma_j = 0$ per qualche $j \in \{rank(L) + 1, ..., n\} \Longleftrightarrow N(A) \cap N(L) \neq \{0\}$. In tal caso, essendo $p = rank(\begin{pmatrix} A \\ L \end{pmatrix})$, abbiamo che il sottospazio vettoriale generato dalle prime p colonne di X è ortogonale al sottospazio generato dalle ultime $(n - p)$. Inoltre quest' ultimo sottospazio è proprio $N(A) \cap N(L) = N(\begin{pmatrix} A \\ L \end{pmatrix})$. Evidenziamo infine che il problema di determinare le radici dell' equazione (calcolo degli autovalori generalizzati)

$$det(A^TA - \lambda^2 L^TL) = 0$$

è una sintesi dei probleni $det(A^TA - \lambda^2 I) = 0$ e $det(A - \lambda L) = 0$. Pertanto il problema di determinare $\lambda(A, L)$ eredita le stesse difficoltà associate a ciascuno di questi problemi base.

Sono possibili altre formulazioni della $GSVD$ che però non aggiungono, a

nostro parere, nulla di più rispetto alla nostra trattazione. Questa routine è infatti di facile implementazione ed è stata recentemente introdotta in diversi pacchetti software [20]; ritorneremo a parlare più dettagliatamente di tale algoritmo nel capitolo 3 e 4 dove avremo modo di analizzare la sua utilità per la trattazione del metodo di regolarizzazione di Tikhonov generale.

Capitolo 2

Regolarizzazione in forma generale con operatori differenziali

Introduciamo in questo capitolo il metodo di regolarizzazione di Tikhonov con operatori differenziali per il problema lineare $Kx = y$ determinando le condizioni di esistenza e unicità della soluzione regolarizzata verificando che queste estendono in modo naturale quelle valide nel caso standard. Presenteremo nelle sezioni 2.3 i risultati di convergenza di maggiore rilievo e in 2.6 lo sviluppo in funzioni singolari della soluzione regolarizzata.
Sottolineiamo che mentre i risultati di esistenza e unicità sono stati studiati a fondo da più autori quelli di convergenza non sono altrettanto conosciuti: proprio questi ultimi, pur richiedendo un' approccio teorico più profondo imposto dalla necessità di lavorare con diverse strutture hilbertiane indotte dagli operatori che definiscono il metodo di regolarizzazione, permettono di avere una visione unificata della regolarizzazione riducendo il metodo standard a un caso particolare di quello generale. Per capire meglio questo abbiamo riportato nella prima sezione alcune peculiarità dei problemi mal posti e della regolarizzazione di Tikhonov standard per fornire un background di situazioni a cui tenteremo di dare un possibile e più generale metodo di analisi.

Riportiamo qui di seguito alcune notazioni e definizioni di base.

Definizione 2.0.2 *Un operatore* $K : X \mapsto Y$ *è detto* **operatore lineare** *se*

$$K(\alpha x + \beta y) = \alpha Kx + \beta Ky \quad \forall \alpha, \beta \in \mathbb{R}, \forall x, y \in X.$$

Indichiamo con:
- **dominio** di K l' insieme $D(K) = \{x \in X : \exists\ Kx\} \subseteq X$
- **nucleo** di K l' insieme $N(K) = \{x \in X : Kx = 0\} \subseteq X$
- **range** di K l' insieme $R(K) = \{Kx : x \in D(K)\} \subseteq Y$.

Dati due spazi vettoriali normati X e Y indichiamo con $\Lambda(X, Y)$ lo spazio vettoriale

$$\Lambda(X, Y) = \{K : X \mapsto Y : K \text{ lineare}\}$$

e $\forall K \in \Lambda(X, Y)$ diciamo che K è **continuo** se esiste una costante $M > 0$ tale che

$$\|Kx\|_Y \leq M\|x\|_X \quad \forall x \in X.$$

Nel caso in cui $X = Y$ indicheremo $\Lambda(X, X)$ con $\Lambda(X)$.

Definizione 2.0.3 *Definiamo* **norma di un operatore**

$$\|K\| = sup\{\|Kx\|_Y : \|x\|_X = 1\}.$$

Diremo che K è limitato se $\|K\| < +\infty$.
Le nozioni di limitatezza e continuità, per un operatore lineare, sono equivalenti negli spazi di Hilbert come specificato nel seguente teorema.

Teorema 2.0.4 *Dato* $K \in \Lambda(X, Y)$ *sono fatti equivalenti*
(1) K *è continuo in* $X \iff K$ *è continuo in* $x_0 \quad \forall x_0 \in X \iff$
$\forall x_0 \in X \ \lim_{x \to x_0} Kx = Kx_0$
(2) K *è limitato.*

Definito

$$B(X, Y) = \{K : X \to Y : K \text{ lineare e continuo}\}$$

possiamo considerare l' applicazione

$$B(X, Y) \to \mathbb{R}^+$$

$$K \to \|K\|$$

che definisce una norma su $B(X,Y)$ rendendolo uno spazio normato.
Dati due spazi di Hilbert X, Y e $K \in B(X,Y)$ definiamo **aggiunto** di K l' operatore $K^* : Y \to X$ definito, in modo unico, dalla relazione

$$(x, K^*y)_X = (Kx, y)_Y \quad \forall x \in X, \forall y \in Y.$$

Si può dimostare [8] che $K \in B(X,Y) \iff K^* \in B(Y,X)$.
Dato $K \in B(X,X) =: B(X)$ diciamo che K è autoaggiunto se $K = K^*$: esempi di operatori autoaggiunti sono K^*K e KK^*.

Definizione 2.0.5 *Dato un sottoinsieme S di uno spazio di Hilbert X definiamo complemento ortogonale di S, indicato con $S^\perp$, l' insieme*

$$S^\perp = \{x \in X : (x,y) = 0 \ \ \forall y \in S\} \subseteq X.$$

Ricordiamo che $S^\perp$ è un sottospazio di X chiuso indipendentemente da S; inoltre $S^{\perp\perp} = \overline{S}$.

Teorema 2.0.6 *Se S è un sottospazio chiuso di uno spazio di Hilbert X allora X può essere decomposto nella somma diretta di S e $S^\perp$ ($X = S \oplus S^\perp$) intendendo con questo che*

$$\forall x \in X \ \ \exists! x_1 \in S, \ \ \exists! x_2 \in S^\perp \text{ tali che } x = x_1 + x_2.$$

Passiamo ora a definire le proiezioni ortogonali.

Definizione 2.0.7 *Supposto S un sottospazio chiuso di uno spazio di Hilbert X. L' operatore $P_S : X \to S$ definito da $P_S x = x_1$, con $x = x_1 + x_2$ rappresentazione di X attraverso la decomposizione $X = S \oplus S^\perp$, è detto* **proiezione ortogonale** *di X su S.*

Riportiamo qui di seguito alcune proprietà fondamentali degli operatori di proiezione ortogonale:
(1) P_S è lineare e limitato ($\|P_S\| = 1$),
(2) $P_{S^\perp} = I - P_S$ (operatore di proiezione complementare rispetto a P_S),
(3) $P_S = P_S^*$ ovvero P_S è autoaggiunto.

Teorema 2.0.8 *Sia $K \in B(X,Y)$. Allora valgono le seguenti relazioni:*
(1) $N(K)^{\perp} = \overline{R(K^*)}$,
(2) $N(K^*)^{\perp} = \overline{R(K)}$,
(3) $N(K^*) = R(K)^{\perp}$,
(4) $N(K) = R(K^*)^{\perp}$.

2.1 Problemi mal posti e loro trattazione

Consideriamo una generica equazione

$$Kx = y \tag{2.1}$$

dove K è una applicazione tra gli spazi topologici X e Y, y è il dato assegnato e richiamiamo brevemente le peculiarità dei problemi mal posti attraverso la definizione data da Hadamard. Un problema è detto ben posto se sono soddisfatte le seguenti condizioni.

$$\text{Esiste una soluzione per ogni dato } y \in Y. \tag{2.2}$$

$$\text{La soluzione è unica per ogni dato } y \in Y. \tag{2.3}$$

$$\text{La soluzione dipende con continuità dai dati.} \tag{2.4}$$

Se una di queste condizioni non è soddisfatta il problema è detto mal-posto.
Per rendere le condizioni precedenti più precise dobbiamo specificare la nozione di soluzione e quali topologie dobbiamo utilizzare per valutare la continuità.
La condizione (2.2) equivale alla surgettività dell' operatore K, ed è quindi soddisfatta se $Y = K(X)$, mentre la condizione (2.3) equivale alla sua iniettività garantita da $N(K) = \{0\}$. Condizione necessaria e sufficiente per la validità di (2.2) e (2.3) è l' invertibilità di K. Supposte valide (2.2) e (2.3), la condizione (2.4) dipende dalle topologie introdotte in X e Y che possono rendere continua o meno l' applicazione che associa ad ogni dato $y \in Y$ la sua (unica) controimmagine $x \in X$.
Da qui nasce la necessità di ridefinire alcuni concetti, come quello di soluzione, introducendo un nuovo problema, opportunamente legato a (2.1), per cui valgano le condizioni (2.2), (2.3), (2.4).

A tale scopo diamo ora alcune nozioni fondamentali relative agli operatori generalizzati tra spazi di Hilbert evidenziandone le caratteristiche più importanti. Per una trattazione più dettagliata si consulti [2,8].

Definizione 2.1.1 *Sia $K : X \to Y$ un operatore lineare e limitato.*
$x_{LS} \in X$ è detta soluzione ai minimi quadrati (o pseudosoluzione) di $Kx = y$ se

$$\|Kx_{LS} - y\| = inf\{\|Kz - y\| : z \in X\}.$$

$x^\dagger \in X$ è detta soluzione generalizzata di $Kx = y$ se è la soluzione ai minimi quadrati di $Kx = y$ di norma minima; cioè

$$\|x^\dagger\| = inf\{\|z\| : z \ e' \ soluzione \ ai \ minimi \ quadrati \ di \ Ku = y\}.$$

Per definizione di pseudosoluzione otteniamo che questa è l' elemento dello spazio X che approssima al meglio il dato dell' equazione (2.1) in quanto minimizza la distanza $\|Kx - y\|$ su tutto X.
Osserviamo inoltre che il vettore $x^\dagger$ è determinato in modo unico poichè
l' insieme delle pseudosoluzioni di $Kx = y$ è un sottoinsieme chiuso e convesso di X.
Definiamo ora l' operatore inversa generalizzata di K restringendo il dominio di quest' ultimo in modo tale che l' operatore ottenuto sia invertibile; il suo inverso verrà quindi esteso al suo dominio massimale.
Indichiamo con $K|S$, $S \subset X$, la restrizione dell' operatore K al sottoinsieme S di X. Senza ledere la generalità della trattazione supporremo fin da subito che il dominio di K sia tutto X ($D(K) = X$).

Definizione 2.1.2 *L' inversa generalizzata (o pseudoinversa) $K^\dagger$ di $K \in \Lambda(X, Y)$ è definita come l' unica estensione lineare di $\tilde{K}^{-1}$ a :*

$$D(K^\dagger) := R(K) \oplus R(K)^\perp$$

con

$$N(K^\dagger) = R(K)^\perp$$

dove $\tilde{K}$ è l' isomorfismo definito da

$$\tilde{K} := K|N(K)^\perp : N(K)^\perp \to R(K).$$

Verifichiamo che la definizione appena data è ben formulata dimostrando inoltre che è possibile esprimere $K^\dagger$ a partire da $\tilde{K}$ e P con P operatore di proiezione su $\overline{R(K)}$.
Innanzitutto dobbiamo dimostrare che $\tilde{K}^{-1}$ esiste: a tale scopo verifichiamo che

$$N(\tilde{K}) = \{0\} \quad \text{e} \quad R(\tilde{K}) = R(K).$$

Infatti:
-$x \in N(\tilde{K}) \Longleftrightarrow \tilde{K}x = 0 \Longleftrightarrow x \in N(K)^\perp$ e $Kx = 0 \Longleftrightarrow$
$x \in N(K) \cap N(K)^\perp = \{0\} \Longleftrightarrow x = 0$.
-$R(\tilde{K}) = R(K|N(K)^\perp) = R(K)$.
La linearità di $\tilde{K}^{-1}$ si deduce dalla linearità di K.
A questo punto dobbiamo dimostrare che $\forall y \in R(K) \oplus R(K)^\perp$ si ha

$$K^\dagger y = \tilde{K}^{-1} P y.$$

Infatti per l' ipotesi $y \in R(K) \oplus R(K)^\perp = D(K^\dagger)$ possiamo scrivere

$$y = y_1 + y_2 \ , \ y_1 \in R(K), \quad y_2 \in R(K)^\perp = N(K^\dagger)$$

da cui segue

$$K^\dagger y = K^\dagger(y_1 + y_2) = K^\dagger y_1 = \tilde{K}^{-1} P y$$

dove nell' ultimo passaggio si è utilizzata la relazione $K^\dagger | R(K) = \tilde{K}$. L' unicità di $K^\dagger$ discende dalla relazione appena trovata.
Un' altra caratterizzazione dell' operatore $K^\dagger$ è data dalle condizioni di Moore-Penrose indicate nella proposizione seguente.

Proposizione 2.1.3 *L' operatore $K^\dagger : D(K^\dagger) \to X$ è l' unico operatore per cui valgono le equazioni di Moore-Penrose:*

$$K^\dagger K K^\dagger = K^\dagger,$$

$$K K^\dagger K = K,$$

$$K^\dagger K = P_{N(K)^\perp},$$

$$K K^\dagger = P_{\overline{R(K)}} | D(K^\dagger).$$

Inoltre $R(K^\dagger) = N(K)^\perp$.

Il legame tra $K^\dagger$ e soluzione ai minimi quadrati di (2.1) è fornito dal seguente teorema.

Teorema 2.1.4 *Sia $y \in D(K^\dagger)$. Allora, $Kx = y$ ha un' unica soluzione generalizzata data da*

$$x^\dagger \equiv K^\dagger y. \tag{2.5}$$

L' insieme di tutte le soluzioni ai minimi quadrati è rappresentato da

$$x^\dagger + N(K) = \{x^\dagger + y : \ y \in N(K)\}. \tag{2.6}$$

Teorema 2.1.5 *Se $y \in Y$ con $y = \bar{y} + z$ dove $\bar{y} \in \overline{R(K)}$ e $z \in R(K)^\perp$, allora sono fatti equivalenti:*
(1) $y \in D(K^\dagger) = R(K) \oplus R(K)^\perp$.
(2) $\bar{y} \in R(K)$.
(3) Esiste $x \in X$ soluzione ai minimi quadrati di (2.1):

$$\|Kx - y\| = inf_{u \in X}\|Ku - y\|. \tag{2.7}$$

(4) Esiste $x \in X$ soluzione dell' equazione normale:

$$K^*Kx = K^*y. \tag{2.8}$$

(5) Esiste $x \in X$ soluzione dell'equazione proiettata:

$$Kx = P_{\overline{R(K)}}y = \bar{y}. \tag{2.9}$$

Inoltre se $x \in X$ soddisfa una delle relazioni (2.7), (2.8), (2.9) verifica le altre due.

Se l' operatore K ha range chiuso, l' equazione $Kx = P_{\overline{R(K)}}y$ ammette sempre almeno una soluzione: infatti, essendo $P_{\overline{R(K)}}y \in \overline{R(K)} = R(K)$, si ha $\{x \in X : Kx = P_{\overline{R(K)}}y\} \neq \emptyset$. Poichè il teorema precedente considera un generico operatore $K \in B(X, Y)$ con range arbitrario, se $R(K)$ non è chiuso, l' elemento $(P_{\overline{R(K)}}y)$ puó non appartenere a $R(K)$ e in tal caso, per la relazione (2.9), non esistono pseudosoluzioni del problema (2.1). Possiamo pertanto riassumere quanto detto nella seguente proposizione.

Proposizione 2.1.6 *Il problema (2.1) ammette pseudosoluzioni se e solo se $y \in R(K) \oplus R(K)^\perp$.*

Ricordando che $R(K) \oplus R(K)^{\perp}$ è denso in Y e coincide con quest' ultimo se $R(K)$ è chiuso otteniamo che, se l' operatore K ha range chiuso, la pseudosoluzione di (2.1) esiste sempre.
Introdotte le proprietà della pseudoinversa (proposizione 2.1.3) e le caratterizzazioni della soluzione ai minimi quadrati (proposizione 2.1.5) ci chiediamo quale condizione su K assicura la continuità di $K^{\dagger}$.

Teorema 2.1.7 *Sia $K \in B(X,Y)$. Allora:*
(1) $K^{\dagger}$ è lineare
(2) $K^{\dagger}$ è limitato $\Longleftrightarrow$ $R(K)$ è chiuso.

Dimostrazione. Si consulti [10]. ∎

Possiamo pertanto dire che se $K \in B(X,Y)$ il problema $Kx = y$ è ben posto, nel senso della soluzione generalizzata, se e solo se $R(K)$ è chiuso: infatti in queste ipotesi si ha $D(K^{\dagger}) = R(K) \oplus R(K)^{\perp} = Y$ e $K^{\dagger}$ continuo garantendo così la validità di (2.2) e di (2.4).

2.1.1 Operatori lineari compatti: rappresentazione in funzioni singolari

In questo paragrafo daremo alcune nozioni relative alla teoria spettrale [9].

Definizione 2.1.8 *Sia $K : X \mapsto Y$ un operatore lineare tra X e Y spazi vettoriali normati (in particolare spazi di Hilbert). K è detto* **compatto** *se $\forall S \subset X$ limitato $\overline{K(S)}$ è compatto in Y.*

Definizione 2.1.9 *$\lambda \in \mathbb{C}$ è detto* **autovalore** *di un operatore $K : X \to Y$ se $(K - \lambda I)$ non è iniettivo. Definiamo inoltre* **spettro** *di K l' insieme*

$$\sigma(K) = \{\lambda \in \mathbb{C} : K - \lambda I \text{ non } e' \text{ invertibile}\}.$$

Gli elementi di $N(K-\lambda I)$ sono detti **autovettori** *di K relativi all' autovalore λ.*

Fondamentali per lo studio degli operatori compatti sono i teoremi riportati qui di seguito.

Teorema 2.1.10 *Sia K un operatore compatto su uno spazio di Hilbert di dimensione infinita. Allora*
(1) $0 \in \sigma(K)$
(2) ogni elemento di $\sigma(K) - \{0\}$ *è un autovalore di* K
(3) se $\sigma(K)$ *non è finito, allora* $\sigma(K)$ *è numerabile e ha zero come unico punto di accumulazione*
(4) se K *è autoaggiunto* $\sigma(K) \subseteq \mathbb{R}$ *e autovettori corrispondenti ad autovalori distinti sono ortogonali.*

Teorema 2.1.11 *Sia* $K \in B(X)$ *un operatore compatto, autoaggiunto e sia* $\lambda_1, \lambda_2, ...$ *una enumerazione dei suoi autovalori distinti e non nulli. Allora* $\forall x \in X$

$$Kx = \sum_{n=1}^{+\infty} \lambda_n P_n x$$

dove P_n *è l' operatore di proiezione di* X *sul sottospazio vettoriale* $N(K - \lambda_n I)$.

Il teorema precedente mostra che ogni operatore compatto e autoaggiunto è il limite di operatori con range di dimensione finita che rappresentano le proiezioni ortogonali sugli autospazi di K. Attraverso il teorema precedente è possibile definire, per un operatore compatto $K : X \to Y$, un sistema singolare $(\sigma_n, v_n, u_n)_n$ nel seguente modo. Indichiamo innanzitutto con $(\sigma_n^2)_n$ l' insieme degli autovalori non nulli dell' operatore lineare, compatto e autoaggiunto K^*K, scritti in ordine decrescente

$$\sigma_1 \geq \sigma_2 \geq ....$$

e con molteplicità. L' insieme $(v_n)_n$ rappresenta allora il sistema ortonormale completo di autovettori dell' operatore K^*K corrispondente al sistema di valori singolari $(\sigma_n)_n$ (con $\sigma_n > 0 \ \forall n$) e genera $\overline{R(K^*)} = \overline{R(K^*K)}$. L' insieme di vettori $(u_n)_n$ è costruito mediante la relazione:

$$u_n = \frac{Kv_n}{\|Kv_n\|}$$

e costituisce un sistema ortonormale completo di autovettori di KK^* e genera $\overline{R(K)} = \overline{R(KK^*)}$. Ricordiamo infine le relazioni:

$$Kv_n = \sigma_n u_n \tag{2.10}$$

$$K^* u_n = \sigma_n v_n \tag{2.11}$$

$$Kx = \sum_{n=1}^{+\infty} \sigma_n (x, v_n) u_n, \quad \forall x \in X \tag{2.12}$$

$$K^* y = \sum_{n=1}^{+\infty} \sigma_n (y, u_n) v_n \quad \forall y \in Y \tag{2.13}$$

dove le serie infinite convergono negli spazi di Hilbert X e Y. Pertanto un sistema singolare per K è costituito da una famiglia di autovalori $(\sigma_n)_n$ non nulli e da due sistemi ortonormali completi di autovettori $(u_n)_n$ e $(v_n)_n$ attraverso cui l' operatore K puó essere diagonalizzato.
Nel caso in cui vi siano infiniti valori singolari è possibile ordinarli in modo decrescente $\sigma_1 \geq \sigma_2 \geq ...$ in maniera tale che abbiano come unico punto di accumulazione zero (teorema 2.1.10):

$$\lim_{n \to +\infty} \sigma_n = 0.$$

Osserviamo esplicitamente che se K ha rango di dimensione finita, K ha un numero finito di valori singolari e le serie in (2.12), (2.13) diventano somme finite.
Ricordando ora che i soli sottospazi di uno spazio di Hilbert chiusi sono quelli di dimensione finita, dalla proposizine 2.1.7, segue che, nell' ipotesi K compatto, l' operatore $K^\dagger$ è continuo se e solo se $dim R(K) < +\infty$. Pertanto, se si esclude il caso banale, la soluzione generalizzata, qualora esista ($y \in R(K) \oplus R(K)^\perp$), non dipende con continuità dal termine noto e il problema corrispondente è mal posto. Usando il sistema singolare di K, è possibile trovare una rappresentazione esplicita dell' inversa generalizzata di un operatore lineare compatto che permette di mettere in relazione le proprietà spettrali di K con il carattere mal posto del problema (2.1).

Teorema 2.1.12 *Sia (σ_n, v_n, u_n) il sistema singolare per l' operatore lineare compatto K, $y \in Y$. Allora abbiamo:*
(1) $y \in D(K^\dagger) \Longleftrightarrow \sum_{n=1}^{+\infty} \frac{|(y,u_n)|^2}{\sigma_n^2} < +\infty$
(2) per ogni $y \in D(K^\dagger)$,

$$x^\dagger \equiv K^\dagger y = \sum_{n=1}^{+\infty} \frac{(y, u_n)}{\sigma_n} v_n. \tag{2.14}$$

La relazione (2.14) mostra come gli errori in y influenzino il risultato in $K^\dagger y$: le componenti degli errori in y rispetto alla base $(u_n)_n$, che corrispondono ai valori singolari "maggiori, sono ininfluenti mentre le componenti degli errori che corrispondono ai valori singolari σ_n più piccoli sono amplificati dai fattori $\frac{1}{\sigma_n}$ risultando così pericolosi per la costruzione di $x^\dagger$. Se $dimR(K) < +\infty$ esiste un numero finito di valori singolari cosicchè questi fattori di amplificazione sono limitati benchè possano risultare ampi e quindi inaccettabili dal punto di vista numerico. Se invece $dimR(K) = +\infty$, essendo $\lim_{n\to+\infty} \sigma_n = 0$, gli errori sui dati di una certa grandezza posso essere amplificati in modo arbitrario attraverso il fattore $\frac{1}{\sigma_n}$ che cresce senza limite superiore. Ad esempio, posto $y_{\delta,n} = y + \delta u_n$ si ha $\|y_{\delta,n} - y\| = \delta$ e per (2.14) vale

$$K^\dagger y - K^\dagger y_{\delta,n} = \frac{(\delta u_n, u_n)}{\sigma_n} v_n = \frac{\delta}{\sigma_n} v_n$$

da cui

$$\|K^\dagger y - K^\dagger y_{\delta,n}\| = \frac{\delta}{\sigma_n} \to +\infty \quad n \to +\infty.$$

Lo studio dei problemi mal posti ha acquistato notevole importanza con la necessità di affrontare i problemi inversi. Considerato il modello matematico (2.1) per problema diretto intendiamo la determinazione, a partire dal dato $x \in X$, del risultato $y = Kx$. Supposto l' operatore K compatto e con sistema singolare $(\sigma_n, v_n, u_n)_n$ la determinazione del valore $y = Kx$ è espressa da (2.12). Il problema inverso sarà la determinazione del valore x la cui immagine secondo K sia y. Utilizzando l' espressione dell' inversa generalizzata (2.14) si puó osservare che, mentre nel problema diretto i coefficienti dello sviluppo in funzioni singolari tendono a zero per $n \to +\infty$, i corrispondenti coefficienti divergono nel caso inverso. Una generica metodologia che permetta di affrontare un problema mal posto viene detta "regolarizzazione. Lo scopo di ogni metodo di regolarizzazione è quello di approssimare la soluzione cercando di fare in modo che soddisfi opportune proprietà di regolarità che hanno in genere le soluzioni corrispondenti ai dati non perturbati.

2.1.2 Regolarizzazione

Consideriamo nuovamente l' equazione (2.1)

$$Kx = y$$

con X e Y spazi di Hilbert, $x \in X$, $y \in Y$ e supponiamo che $R(K)$ non sia chiuso. L' equazione (2.1) dà luogo a un problema mal posto in quanto la non continuità dell'operatore $K^\dagger$ comporta che la soluzione non dipende con continuità dai dati.
L' idea alla base dei metodi di regolarizzazione è quella di definire una famiglia di operatori R_λ limitati che approssimino con continuità $K^\dagger$.

Definizione 2.1.13 *Una famiglia di operatori* $(R_\lambda)_\lambda$*, con* $R_\lambda : Y \mapsto X \ \ \forall \lambda \in \mathbb{R}^+$*, è detto* **algoritmo regolarizzante** *per il problema (2.1) se sono soddisfatte le seguenti condizioni:*
(1) R_λ *è lineare e continuo* $\forall \lambda > 0$
(2) $\forall y \in R(K) \oplus R(K)^\perp$ *si ha* $\lim_{\lambda \to 0} R_\lambda y = x^\dagger = K^\dagger y$.

Dalla definizione appena data, segue che ogni algoritmo regolarizzante converge alla soluzione generalizzata qualora questa esista ($y \in D(K^\dagger)$). Nell' ipotesi che il dato sia perturbato puó succedere che y non appartenga a $D(K^\dagger)$: in tal caso, mentre l' inversa generalizzata non è di alcuna utilità l' algoritmo regolarizzante $(R_\lambda)_\lambda$, essendo $D(R_\lambda) = Y$, permette di ottenere ugualmente una soluzione approssimata del problema non perturbato. Supposto infatti $y = \bar{y} + e$ con $\bar{y} \in D(K^\dagger)$ ed $e \in Y - D(K^\dagger)$ abbiamo

$$R_\lambda y = R_\lambda \bar{y} + R_\lambda e$$

con $\lim_{\lambda \to 0} R_\lambda \bar{y} = K^\dagger \bar{y}$: l' operatore R_λ, per piccoli valori di λ, approssima bene la soluzione del problema corrispondente al dato $\bar{y}$ ma, come dimostrato in [5], questo comporta un grosso errore dovuto al fatto che il valore $(R_\lambda e)$ diverge per $\lambda \to 0$. Pertanto la scelta del parametro λ dovrà essere tale da garantire il fatto che l' operatore di regolarizzazione approssimi in maniera soddisfacente $K^\dagger$ e contemporaneamente non risenta eccessivamente dell' errore sul dato.
L' idea che sta alla base della realizzazione di un algoritmo regolarizzante è quella di utilizzare informazioni aggiuntive che permettano di gestire il problema mal posto in maniera più regolare.
Considereremo soltanto una particolare famiglia di operatori regolarizzanti, proposta da Tikhonov, analizzando in dettaglio la sua estensione al caso generale. Il metodo della regolarizzazione di Tikhonov ha come fine la determinazione di una soluzione x che soddisfa, non proprio il problema $Kx = y$, bensì

le due relazioni
(1) $\|Kx - y\| \leq \epsilon$ (approssimazione del dato entro un margine di errore ϵ)
(2) $\|x\| \leq E$ (vincolo sulla soluzione)
minimizzando il funzionale, dipendente dal parametro λ,

$$\tilde{G}_\lambda(x, y) = \|Kx - y\|_Y^2 + \lambda^2 \|x\|_Y^2. \tag{2.15}$$

Attraverso lo studio di $\tilde{G}_\lambda$ si possono stabilire le condizioni di esistenza di una soluzione regolarizzata e, in caso affermativo, si puó dimostrare [5] che questa è ottenuta risolvendo l' equazione di Eulero

$$(K^*K + \lambda^2 I)x = K^*y.$$

Per un operatore compatto K con sistema singolare $(\sigma_n, v_n, u_n)_n$ la soluzione regolarizzata del problema standard

$$min_{x \in X}\{\tilde{G}(x)\} = min_{x \in X}\{\|Kx - y\|_Y^2 + \lambda^2 \|x\|_Y^2\}$$

è espressa da :

$$x_\lambda = \sum_{n=1}^{+\infty} \frac{\sigma_n}{\sigma_n^2 + \lambda^2}(y, u_n)v_n.$$

In tal caso, un semplice confronto con (2.14) mostra il miglioramento ottenuto in termini di stabilità: gli errori in (y, u_n) non vengono propagati, nel risultato, con i fattori $1/\sigma_n$, ma solo con i fattori $\sigma_n/(\sigma_n^2 + \lambda^2)$ che sono limitati superiormente da σ_1/λ^2 e tendono a zero per $n \to +\infty$.

2.2 Regolarizzazione di Tikhonov generale

Definizione 2.2.1 *Definiamo* **operatore di regolarizzazione** *una funzione* $L : D(L) \subseteq X \longrightarrow Y$ *tale che:*
-L *è lineare,*
-$D(L)$ *è un sottospazio di* X*,*
-$D(L)$ *denso in* X*,*
-il grafico di L, $\Gamma = \{(x, Lx) : x \in D(L)\}$, *è chiuso nello spazio prodotto* $X \otimes Y = \{(x, y) : x \in X , y \in Y\}$ *munito del prodotto scalare usuale*

$$< (x, y), (x', y') >= (x, x')_X + (y, y')_Y \quad \forall x, y \in X \ \forall x', y' \in X'. \tag{2.16}$$

Definiamo in $D(L)$ un nuovo prodotto scalare e la relativa norma indotta:

$$(x,y)_L = (x,y) + (Lx,Ly) \qquad \|x\|_L^2 = \|x\|^2 + \|Lx\|^2 \ \ \forall\, x,y \in D(L)$$

e verifichiamo che l' ipotesi L chiuso [1] assicura che $(D(L),\ (\ ,\ )_L)$ è uno spazio di Hilbert. Dobbiamo dimostrare che se $(x_n)_n \subseteq D(L)$ è una successione di Cauchy in $D(L),$ rispetto a $\|\ \|_L$, esiste $x \in D(L)$ tale che $\lim_{n\to+\infty} x_n = x$. Per ipotesi: $\forall \epsilon > 0\ \exists N$ tale che se $n,m \geq N$

$$\|x_n - x_m\|_L = \|x_n - x_m\| + \|Lx_n - Lx_m\| \leq \epsilon$$

e quindi $(x_n)_n$ e $(Lx_n)_n$ sono successioni di Cauchy in X e Y rispettivamente. Dalle ipotesi X e Y spazi di Hilbert e L è chiuso otteniamo che:

$$(x_n)_n \longrightarrow x \in D(L) \ \text{ e } \ (Lx_n)_n \longrightarrow y = Lx \ \text{ per } \ n \to +\infty$$

come volevemo dimostrare.

Il metodo di regolarizzazione di Tikhonov generale, in analogia con il suo corrispondente standard, ha come scopo la determinazione di una soluzione che soddisfi la condizione $\|Kx - y\| \leq \epsilon$ mentre il vincolo $\|x\| \leq E$ è sostituito dalla condizione di regolarità $\Omega(x) = \|Lx\| \leq E$ introdotta al fine di diminuire le oscillazioni nella soluzione. Tale soluzione è pertanto ottenuta studiando il funzionale

$$G_\lambda : D(L) \longrightarrow \mathbb{R} \tag{2.17}$$
$$x \longrightarrow \|Kx - y\|_Y^2 + \lambda^2 \|Lx\|_Y^2.$$

che estende quello definito in (2.15) al caso $L \neq I$.
Determiniamo ora le condizioni su K e L che garantiscono esistenza e unicità della soluzione regolarizzata nel caso generale.

Teorema 2.2.2 *Il funzionale G_λ ha un minimo nel punto $x \in D(L)$ se e solo se $x \in D(L^*L)$ e*

$$(K^*K + \lambda^2 L^*L)x = K^*y \tag{2.18}$$

[1] Si ricordi che L ha grafico chiuso $\iff$ L è chiuso $\iff$

$$\{x_n\}_n \subseteq D(L),\ \ x_n \to x \ \text{ e } \ Lx_n \to y$$

implica $x \in D(L)$ e $Lx = y$.

Dimostrazione. Siano $x, w \in D(L)$ e $f : \mathbb{R} \longrightarrow \mathbb{R}$ il funzionale quadratico definito a partire da G_λ come $f(\alpha) = G_\lambda(x + \alpha w)$.
Se x è un punto di minimo per G_λ allora $\alpha = 0$ è punto di minimo per f:infatti per ipotesi $\forall y \in D(L)\ \ G_\lambda(x) \leq G_\lambda(y)$ e quindi scelto $y = x + \alpha w\ \forall \alpha \in \mathbb{R}$ si ha $f(0) = G_\lambda(x) \leq G_\lambda(x + \alpha w) = f(\alpha)\ \forall\ \alpha \in \mathbb{R}$.
Poichè f è derivabile la condizione che f abbia un minimo in $\alpha = 0$ è equivalente a imporre:

$$0 = f^{'}(0) = 2[\lambda^2(Lx, Lw) + (Kx - y, Kw)]$$

da cui otteniamo, sfruttando la linearità del prodotto scalare e le proprietà dell'operatore aggiunto, la relazione cercata.
Dimostriamo ora che se $x \in D(L^*L)$ soddisfa (2.18) allora x è punto di minimo per G_λ. Sia $x \in D(L^*L) \subseteq D(L)$ tale che $(K^*K + \lambda^2 L^*L)x = K^*y$ allora $\forall u \in D(L)$ poniamo $w = u - x$ e $f(\alpha) = G_\lambda(x + \alpha w)$: poichè f è una funzione quadratica in α e $f^{'}(0) = (K^*K + \lambda^2 L^*L)x - K^*y = 0$ otteniamo che f ha un minimo in $\alpha = 0$.
In particolare:

$$G_\lambda(x) = f(0) \leq f(1) = G_\lambda(x + w) = G_\lambda(u)\ \ \forall u \in D(L)$$

come volevamo dimostrare. ∎

Da ora in poi indicheremo, per ogni λ, la soluzione di (2.18) con x_λ.

Osservazione 2.2.3

$$D(L^*L) = \{x \in D(L) : Lx \in D(L^*)\}.$$

$$N(K^*K + \lambda^2 L^*L) = N(K) \cap N(L)\ \ \forall \lambda \neq 0.$$

La prima uguaglianza discende dalla definizione di dominio di un operatore. Per la seconda dobbiamo verificare due inclusioni. Per " $\supseteq$" basta osservare che se $x \in N(K) \cap N(L)$ abbiamo subito

$$(K^*K + \lambda^2 L^*L)x = K^*Kx + \lambda^2 L^*Lx = 0$$

ovvero $x \in N(K^*K + \lambda^2 L^*L)$. Proviamo ora " $\subseteq$": $x \in N(K^*K + \lambda^2 L^*L) \Longrightarrow$
$((K^*K + \lambda^2 L^*L)^2 x, x) = 0 \Longleftrightarrow (K^*Kx, x) + \lambda^2 (L^*Lx, x) = 0 \Longleftrightarrow$
$\|Kx\|^2 + \lambda^2 \|Lx\|^2 = 0 \Longleftrightarrow x \in N(K) \cap N(L)$.

La precedente osservazione e il teorema 2.2.2 permettono di dedurre che il problema $min_{x\in D(L)}G_\lambda(x)$ ammette un' unica soluzione, qualora esista, se e solo se

$$N(K) \cap N(L) = \{0\}. \tag{2.19}$$

Basta infatti osservare che se x_1 e x_2 sono due soluzioni di (2.18) allora $(K^*K + \lambda^2 L^*L)(x_1 - x_2) = 0$ cioè $(x_1 - x_2) \in N(K^*K + \lambda^2 L^*L)$ e quindi, se vale (2.19), $x_1 = x_2$. Se invece $N(K) \cap N(L) \neq \{0\}$ allora detta x_λ una soluzione di (2.18) si possono definire infinite soluzioni mediante la relazione $x_{\beta,\lambda} = x_\lambda + \beta u$ con $u \in N(K) \cap N(L)$, $u \neq 0$ e $\beta \in \mathbb{R}$.
Si ricordi che nel caso standard la condizione di unicità della soluzione regolarizzata era data da

$$N(K) = \{0\}.$$

Osservazione 2.2.4 *Il teorema precedente generalizza l' equazione di Eulero* $(K^*K+\lambda^2 I)x = K^*y$ *che individua il minimo del funzionale* G_λ *nel caso in cui* $L = I$. *Sottolineiamo tuttavia che mentre nel caso standard* G_λ *era definito su tutto* X *ora è definito in* $D(L)$. *Pertanto l' equazione (2.18) è valida in* $D(L^*L) \subseteq D(L)$ *garantendo così una maggiore regolarità della soluzione.*
Più precisamente scegliamo $X = L^2([a,b])$, $Y = L^2([c,d])$, $K \in L^2([c,d] \times [a,b])$

$$(Kf)(x) = \int_a^b K(x,t)f(t)dt \quad \forall x \in [c,d]$$

operatore di Fredholm di prima specie e sia L *un operatore di derivata di ordine* n *definito sullo spazio di Sobolev*

$$D(L) = H^m([a,b]) = \{f \in L^2([a,b]) : D^p f \in L^2([a,b]) \ per \ \ 1 \leq p \leq m\} \subset L^2([a,b])$$

con prodotto scalare

$$(f,g) = \sum_{k=0}^{m} (D^k f, D^k g) = \sum_{k=0}^{m} \int_a^b D^k f \, D^k g d\omega \tag{2.20}$$

dove $D^p f = \frac{d^p}{d\omega^p} f$. *Il teorema precedente garantisce allora che*

$$x_\lambda \in D(L^*L) = H^{2m}([a,b]) \subset H^m([a,b])$$

ovvero una condizione di regolarità della soluzione x_λ.

Dalla relazione (2.18), otteniamo che una condizione sufficiente a garantire sia esistenza che unicità della soluzione x_λ per il problema $\min_{x\in D(L)} G_\lambda(x)$ è data dall' invertibilità dell' operatore

$$K^*K + \lambda^2 L^*L : D(L^*L) \subset X \to X. \tag{2.21}$$

Si ricordi infatti che la nozione di invertibilità di (2.21) è equivalente a imporre che sia iniettivo, garantendo l' unicità di x_λ, e che sia surgettivo, garantendo invece l' esistenza di tale soluzione. Stabiliamo ora quali ipotesi su K e L assicurano l' esistenza di $(K^*K + \lambda^2 L^*L)^{-1}$. A tale scopo assumiamo che K e L soddisfino le seguenti condizioni [18]:

$$N(K) \cap N(L) = 0 \tag{2.22}$$

$$R(L) \text{ è chiuso} \tag{2.23}$$

$$\exists \beta \in \mathbb{R}^+ : \quad \|Kx\| \geq \beta \|x\| \quad \forall x \in N(L) \tag{2.24}$$

e introduciamo una seconda struttura sul sottospazio $D(L)$ di X definendo [9, 18, 19]:

$$(x, y)_* = (Kx, Ky) + (Lx, Ly), \;\; \|x\|_*^2 = \|Kx\|^2 + \|Lx\|^2 \;\; \forall x \in D(L).$$

Tale struttura in $D(L)$ è detta *-struttura su $D(L)$.

Lemma 2.2.5 *Si ha:*
(1) $(\ ,\ )_*$ *è un prodotto scalare in D(L),*
(2) $D(L)$ *è uno spazio di Hilbert rispetto al prodotto* $(\ ,\ )_*$*,*
(3) $\|\ \|_*$ *e* $\|\ \|_L$ *sono norme topologicamente equivalenti in D(L).*

Dimostrazione.
(1) Naturalmente $(\ ,\ )_*$ è un operatore bilineare poichè composto di K, L lineari e $(\ ,\ )_Y$ bilineare. Da dimostrare che $\|\ \|_*$ è una norma in $D(L)$:
$\|x\|_*^2 = \|Kx\|^2 + \|Lx\|^2 = 0 \iff Lx = Kx = 0 \iff x \in N(K) \cap N(L) = \{0\} \iff x = 0$. Le altre proprietà discendono dalle analoghe valide per $\|\ \|_Y$.
(2) Sia $(x_n)_n \subseteq D(L)$ una successione di Cauchy in $D(L)$ e dimostriamo che esiste $x \in D(L)$ tale che, per $n \to +\infty$, $(x_n)_n \longrightarrow x$ rispetto alla norma $\|\ \|_*$.
Per ipotesi $0 = \lim_{n,m\to+\infty} \|x_n - x_m\|_* = \lim_{n,m\to+\infty} [\|Kx_n - Kx_m\| + \|Lx_n - Lx_m\|] \iff \lim_{n,m\to+\infty} \|Kx_n - Kx_m\| = 0$ e $\lim_{n,m\to\infty} \|Lx_n - Lx_m\| = 0$.

Pertanto $(Kx_n)_n$ e $(Lx_n)_n$ sono successioni di Cauchy in Y e quindi esistono $\xi, \eta \in Y$ tali che:

$$\lim_{n\to+\infty} Kx_n = \xi \quad \text{e} \quad \lim_{n\to+\infty} Lx_n = \eta.$$

Poichè $(x_n)_n \subseteq D(L) = (D(L)\cap N(L))\oplus(D(L)\cap N(L))^\perp$ scriviamo $x_n = u_n+v_n$ con $u_n \in N(L)$ e $v_n \in N(L)^\perp$ da cui

$$Lx_n = Lv_n \longrightarrow \eta \;\; n \longrightarrow +\infty.$$

Poichè, per ipotesi, $R(L)$ è chiuso si ha che $L^\dagger$ è continuo, $\eta \in R(L)$ e quindi otteniamo

$$v_n = L^\dagger L x_n \longrightarrow L^\dagger\eta \in D(L) \cap N(L)^\perp.$$

Poichè K è continuo $Kv_n \longrightarrow KL^\dagger\eta$ e quindi

$$Ku_n = K(x_n - v_n) = Kx_n - Kv_n \longrightarrow \xi - KL^\dagger\eta \;\; n \to +\infty.$$

Essendo $(u_n)_n \subseteq N(L)$ posso applicare l' ipotesi (2.24) e scrivere:

$$\|u_n - u_m\| \leq \frac{1}{\beta}\|Ku_n - Ku_m\|$$

da cui si ottiene che $(u_n)_n$ è una successione di Cauchy in X e quindi esiste $u \in N(L)$ tale che $(u_n)_n \longrightarrow u$.
Posto allora $x = u + L^\dagger\eta \in D(L)$ si ha :
-$x_n = u_n + v_n \longrightarrow (u + L^\dagger\eta) =: x$
-$Kx_n = Ku_n + Kv_n \longrightarrow \xi$
-$Lx_n \longrightarrow \eta$.
Poichè L è chiuso $Lx = \eta$ e quindi

$$\|x_n - x\|_*^2 = \|Kx_n - Kx\|^2 + \|Lx_n - Lx\|^2 = \|Kx_n - \xi\|^2 + \|Lx_n - \eta\|^2 \longrightarrow 0$$

ovvero $(x_n)_n \longrightarrow_{\|\,\|_*} x \in D(L) \;\; n \to +\infty$.
(3) Basta dimostrare che $\forall x \in D(L)$ esiste $c \in \mathbb{R}^+ - \{0\}$ indipendente da x tale che [2]:

$$\|x\|_* \leq c\|x\|_L.$$

[2]Si ricordi che se X è uno spazio di Banach rispetto a due norme $\|\ \|$ e $\|\ \|'$ per cui esiste $\alpha > 0$ tale che

$$\|x\| \leq \alpha\|x\|' \;\; \forall x \in X$$

allora queste sono topologicamente equivalenti [8].

Infatti $\forall x \in D(L)$ abbiamo: $\|x\|_*^2 = \|Kx\|^2 + \|Lx\|^2 \leq \|K\|^2\|x\|^2 + \|Lx\|^2 \leq max\{\|K\|, 1\}[\|x\|^2 + \|Lx\|^2] = max\{\|K\|, 1\}\|x\|_L^2$. ∎

In seguito lavoreremo con entrambe le stutture di spazi di Hilbert definite nel lemma 2.2.5 tenendo presente che inducono la stessa topologia e pertanto la stessa nozione di convergenza

Lemma 2.2.6 *Per ogni $\lambda \neq 0$ l' operatore lineare $K^*K + \lambda^2 L^*L$ è autoaggiunto e invertibile. Inoltre $(K^*K + \lambda^2 L^*L)^{-1}$ è continuo.*

Dimostrazione.
-Per la disuguaglianza di Schwartz e la proprietà (3) del lemma 2.2.5 si ha
$\|(K^*K + \lambda^2 L^*L)x\|\|x\| \geq ((K^*K + \lambda^2 L^*L)x, x) = (K^*Kx, x) + \lambda^2(L^*Lx, x) = \|Kx\|^2 + \lambda^2\|Lx\|^2 \geq \min\{1, \lambda^2\}(\|Kx\|^2 + \|Lx\|^2) = min\{1, \lambda^2\}\|x\|_*^2$
$\geq m^2 min\{1, \lambda^2\}\|x\|_L^2 \geq m^2 min\{1, \lambda^2\}\|x\|^2$
da cui otteniamo:

$$\|(K^*K + \lambda^2 L^*L)x\| \geq m^2 min\{1, \lambda^2\}\|x\| \quad \forall x \in D(L^*L).$$

Da quest' ultima relazione si deduce:
-$(K^*K + \lambda^2 L^*L)$ è iniettiva:
infatti supposto $(K^*K + \lambda^2 L^*L)x_1 = (K^*K + \lambda^2 L^*L)x_2$ si ha
$0 = \|(K^*K + \lambda^2 L^*L)(x_1 - x_2)\| \geq m^2 min\{1, \lambda^2\}\|x_1 - x_2\| \Longrightarrow \|x_1 - x_2\| = 0 \Longleftrightarrow x_1 = x_2$.
-$R(K^*K + \lambda^2 L^*L)$ è chiuso:
dobbiamo dimostrare che se $((K^*K + \lambda^2 L^*L)x_n)_n \to y$ per $n \to +\infty$ allora $y \in R(K^*K + \lambda^2 L^*L)$. Utilizzando l' ipotesi, la maggiorazione precedente e l' invertibilità di $(K^*K + \lambda^2 L^*L)$ otteniamo
$0 \longleftarrow \|(K^*K + \lambda^2 L^*L)x_n - y\| = \|(K^*K + \lambda^2 L^*L)[x_n - (K^*K + \lambda^2 L^*L)^{-1}y]\| \geq m^2 min\{1, \lambda^2\}\|x_n - (K^*K + \lambda^2 L^*L)^{-1}y\|$
ovvero esiste $\lim_{n\to+\infty} x_n = (K^*K + \lambda^2 L^*L)^{-1}y := x \in D(L^*L) \subset D(L)$ e quindi

$$y = (K^*K + \lambda^2 L^*L)x \in R((K^*K + \lambda^2 L^*L))$$

come volevamo dimostrare.
-$(K^*K + \lambda^2 L^*L)$ è surgettivo:

$$R(K^*K + \lambda^2 L^*L) = \overline{R(K^*K + \lambda^2 L^*L)} = N(K^*K + \lambda^2 L^*L)^{\perp} = \{0\}^{\perp} = X.$$

Pertanto $(K^*K + \lambda^2 L^*L)$ è un operatore lineare, invertibile e

$$\|(K^*K + \lambda^2 L^*L)^{-1}\| \leq \frac{1}{m^2 min\{1, \lambda^2\}}$$

ovvero $(K^*K + \lambda^2 L^*L)^{-1}$ è limitato. ∎

Possiamo riunire i risultati di questa sezione nel seguente teorema:

Teorema 2.2.7 *Se K è un operatore lineare e L di regolarizzazione per cui valgono le ipotesi (2.22), (2.23), (2.24) allora: per ogni $\lambda \neq 0$ esiste un unico elemento $x_\lambda \in D(L)$ tale che*

$$\|Kx_\lambda - y\|^2 + \lambda^2 \|Lx_\lambda\|^2 = \inf_{x \in D(L)} [\|Kx - y\|^2 + \lambda^2 \|Lx\|^2].$$

Inoltre l' elemento x_λ è l' unico in $D(L^*L)$ tale che

$$(K^*K + \lambda^2 L^*L)x_\lambda = K^*y.$$

2.3 Analisi della convergenza

Dimostriamo ora la convergenza di x_λ, per $\lambda \to 0$, a una particolare soluzione ai minimi quadrati di (2.1) di cui daremo una caratterizzazione in termini dei metodi di regolarizzazione vincolati [2]. La convergenza è ottenuta rispetto alla norma $\| \ \|_L$, o equivalentemente rispetto a $\| \ \|_*$, risultando quindi più forte rispetto a quella indotta in $D(L^*L)$ dalla metrica definita su X.
Indichiamo con

$$A = K|D(L)$$

la restrizione di K a $D(L)$. Poichè A e L sono ancora operatori, da $(D(L), (\ ,\ )_*)$ a Y, lineari e limitati possiamo considerare i rispettivi operatori aggiunti, rispetto a tali strutture, che indicheremo con $A^\sharp$ e $L^\sharp$ definiti dalle relazioni

$$(Ax, y) = (x, A^\sharp y)_*, \qquad (Lx, y) = (x, L^\sharp y)_*,$$

$\forall x \in D(L),\ \forall y \in Y.$

Lemma 2.3.1 *(1)* $A^\sharp y = (K^*K + L^*L)^{-1}K^*y \ \ \forall y \in Y.$
(2) $L^\sharp y = (K^*K + L^*L)^{-1}L^*y \ \ \forall y \in D(L^*).$
(3) $(A^\sharp A + L^\sharp L) = I \quad in \quad D(L).$

Dimostrazione.
(1) Per il lemma precedente l' operatore $(K^*K + L^*L)$ è invertibile e quindi $\forall y \in Y$ possiamo porre $u = (K^*K + L^*L)^{-1}K^*y$ da cui si ha
-$u \in D(L^*L)$
-$(K^*K + L^*L)u = K^*y$
-$\forall x \in D(L)\ \ (x, A^\sharp y)_* = (Ax, y) = (Kx, y) = (x, K^*y) = (x, (K^*K + L^*L)u) = (x, K^*Ku) + (x, L^*Lu) = (Kx, Ku) + (Lx, Lu) = (x, u)_*$
ovvero, essendo $D(L)$ denso in X, $A^\sharp y = u = (K^*K + \lambda^2 L^*L)^{-1}K^*y$.
(2) Analoga alla precedente.
(3) $(A^\sharp A + L^\sharp L)x = A^\sharp Kx + L^\sharp Lx = (K^*K + L^*L)^{-1}K^*Kx + (K^*K + L^*L)^{-1}L^*Lx = Ix\ \ \forall x \in D(L^*L)$.
Estendiamo ora tale risultato a $D(L)$:

$$\forall y \in D(L)\ \ \exists (y_n)_n \subset D(L^*L) \text{ tale che } \lim_{n\to+\infty} y_n = y, \lim_{n\to+\infty} Ly_n = Ly$$

e quindi

$$(A^\sharp A + L^\sharp L)y = \lim_{n\to+\infty}(A^\sharp A + L^\sharp L)y_n = \lim_{n\to+\infty} y_n = y.$$

∎

Teorema 2.3.2 *Se K è un operatore lineare e L è un operatore differenziale per cui valgono le ipotesi (2.22), (2.23), (2.24) allora x_λ converge a qualche elemento $x_0 \in D(L)$ per $\lambda \to 0$ rispetto alla norma $\| \|_*$ (o equivalentemente $\| \|_L$) se e solo se :*

$$y \in D(A_*^\dagger) = R(A) \oplus R(A)^\perp = R(A) \oplus N(A^\sharp)^\perp$$

dove $A_*^\dagger$ rappresenta l' inversa generalizzata di A rispetto alla norma $\| \|_*$. In questo caso x_0 è la soluzione ai minimi quadrati di (2.1) data da

$$x_0 = \lim_{\lambda\to 0} x_\lambda = A_*^\dagger y. \tag{2.25}$$

Dimostrazione.
" $\Longrightarrow$"
Per ipotesi abbiamo: $\lim_{\lambda\to 0} \|x_\lambda - x_0\|_* = 0$ con $x_0 \in D(L)$.

Poichè $\|x_\lambda - x_0\|_* = [\|K(x_\lambda - x_0)\|^2 + \|L(x_\lambda - x_0)\|^2]^{1/2} \to 0 \iff^{\odot}$
$x_\lambda \to x_0$, $Kx_\lambda \to Kx_0$, $Lx_\lambda \to Lx_0$ abbiamo, per la continuità di $A^\sharp$ e $L^\sharp$,

$$A^\sharp Ax_\lambda \to A^\sharp Ax_0, \;\; L^\sharp Lx_\lambda \to L^\sharp Lx_0 \;\; \lambda \to 0.$$

Motiviamo brevemente l' equivalenza "$\iff^{\odot}$" dimostrando la validità, per $\lambda \to 0$, di

$$\|x_\lambda - x_0\|_* \to 0 \iff x_\lambda \to x_0, Kx_\lambda \to Kx_0, Lx_\lambda \to Lx_0.$$

L' implicazione " $\Longleftarrow$" discende dalla definizione di $\| \; \|_*$. Verifichiamo ora l' implicazione opposta.
Per il lemma 2.2.5 abbiamo, per $\lambda \to 0$,

$$\|x_\lambda - x_0\|_* \to 0 \iff \|x_\lambda - x_0\|_L \to 0 \iff \|x_\lambda - x_0\| \to 0 \Longrightarrow x_\lambda \to x_0$$

mentre

$$0 \leftarrow \|x_\lambda - x_0\|_*^2 = \|Kx_\lambda - Kx_0\|^2 + \|Lx_\lambda - Lx_0\|^2$$

è equivalente a

$$Kx_\lambda \to Kx_0 \;\; \text{e} \;\; Lx_\lambda \to Lx_0.$$

Per ogni $x_\lambda \in D(L^*L) \subset D(L)$ si ha

$$(K^*K + \lambda^2 L^*L)x_\lambda = K^*y \tag{2.26}$$

da cui, applicando $(K^*K + L^*L)^{-1}$ ad ambo i membri, si deduce:

$$A^\sharp Kx_\lambda + \lambda^2 L^\sharp Lx_\lambda = A^\sharp y$$

ovvero

$$(A^\sharp A + \lambda^2 L^\sharp L)x_\lambda = A^\sharp y. \tag{2.27}$$

Passando al limite per $\lambda \to 0$ in (2.26) e (2.27) otteniamo

$$K^*Kx_0 = K^*y \;\; \text{e} \;\; A^\sharp Ax_0 = A^\sharp y$$

e quindi per il teorema (2.1.5) x_0 è soluzione ai minimi quadrati per $Kx = y$ e $y \in D(A_*^\dagger) = R(A) + R(A)^\perp$.
" $\Longleftarrow$"
Sia ora $y \in R(A) \oplus R(A)^\perp$. Sostituendo l' identità $(A^\sharp A + L^\sharp L)x_\lambda = x_\lambda$, dedotta nel lemma 2.3.1, in $(A^\sharp Ax_\lambda + \lambda^2 L^\sharp L)x_\lambda = A^\sharp y$ e ricordando che

$x_\lambda \in D(L)$ otteniamo:
$A^\sharp A x_\lambda + \lambda^2(-A^\sharp A x_\lambda + x_\lambda) = A^\sharp y \iff (1-\lambda^2)A^\sharp A x_\lambda + \lambda^2 x_\lambda = A^\sharp y \iff$
$A^\sharp A x_\lambda + \frac{\lambda^2}{1-\lambda^2} x_\lambda = \frac{1}{(1-\lambda^2)} A^\sharp y \iff$

$$x_\lambda = (1-\lambda^2)^{-1}(A^\sharp A + \lambda^2(1-\lambda^2)^{-1} I)^{-1} A^\sharp y \ \ |\lambda| < 1 \ .$$

Pertanto, utilizzando la relazione

$$\lim_{\beta \to 0^+} (A^\sharp A + \beta^2 I)^{-1} A^\sharp y = A_*^\dagger y \ \ \forall y \in D(A_*^\dagger),$$

deduciamo che:

$$\lim_{\lambda \to 0} x_\lambda = A_*^\dagger y = x_0.$$

■

Sottolineiamo che l' elemento $x_0 = A_*^\dagger y$ è in generale diverso da $x^\dagger = K^\dagger y$. Verificheremo nel capitolo 3 che questi, nel caso discreto, coincidono come conseguenza dell' ipotesi (2.22).

Osservazione 2.3.3 *Dalla dimostrazione del teorema precedente deduciamo che l' operatore*

$$L: \quad (D(L), \| \ \|_*) \to (Y, \| \ \|_Y)$$

è continuo indipendentemente dal fatto che

$$L: \quad (D(L), \| \ \|_X) \to (Y, \| \ \|_Y)$$

lo sia o meno. Scelto, ad esempio,

$$L: \quad D(L) = H^1([0,\pi]) \to L^2([0,\pi])$$

$$f \to Lf := f'$$

si ha [8,22] che $D(L)$ è denso in $L^2([0,\pi])$ ma L non è continuo. Infatti la successione

$$f_n(x) := n^{-1/2} sen(nx) \in L^2([0,\pi]), \quad n \in \mathbb{N} - \{0\}$$

è tale che $(f_n)_n \to_{L^2([0,\pi])} 0$ ma $(Lf_n)_n \not\to_{L^2([0,\pi])} 0$.

Osservazione 2.3.4 *La convergenza trattata nel teorema precedente è equivalente a richiedere :*

$$x_\lambda \to x_0 \quad , Kx_\lambda \to Kx_0 \quad , Lx_\lambda \to Lx_0$$

per $\lambda \to 0$. *Nel caso in cui* L *sia scelto come operatore differenziale di ordine* n, *la convergenza rispetto a* $\| \ \|_L$ *corrisponde alla convergenza negli spazi di Sobolev* $H^n([a,b])$*: cioè alla convergenza uniforme in* $[a,b]$ *delle derivate di ordine* $0, 1, ..., n-1$ *e alla convergenza in* $L^2[a,b]$ *delle derivate di ordine* n. *Questo conferma come le ipotesi e le scelte fatte siano tra le più opportune per la regolarizzazione di problemi rappresentati da equazioni lineari di tipo integrale e/o differenziale in cui gli spazi di appartenenza della soluzione sono, per l' appunto, spazi di Sobolev.*

Il teorema precedente ci permette di dimostrare che anche nel caso generale è possibile definire un algoritmo regolarizzante per (2.1). Infatti l' equazione (2.18) implica che x_λ puó essere scritta nella forma

$$x_\lambda = R_\lambda y \tag{2.28}$$

dove

$$R_\lambda = (K^*K + \lambda^2 L^*L)^{-1} K^*. \tag{2.29}$$

è una approssimazione dell' operatore inverso generalizzato $A_*^\dagger$. A tale scopo osserviamo che valgono le seguenti proprietà:
(1) $\forall \lambda > 0$ R_λ è un operatore lineare e continuo come conseguenza del lemma 2.2.6,
(2) per ogni elemento non perturbato $\bar{y} \in D(A_*^\dagger) = R(A) \oplus R(A)^\perp$,

$$\lim_{\lambda \to 0} R_\lambda \bar{y} = A_*^\dagger \bar{y} \tag{2.30}$$

dove il limite è fatto rispetto a $\| \ \|_*$ o equivalentemente $\| \ \|_L$ (teorema 2.3.2). Si osservi peró che, rispetto alla definizione 2.1.13, la convergenza ora è ottenuta in $R(A) \oplus R(A)^\perp$ e non in $R(K) \oplus R(K)^\perp$ come conseguenza del fatto che $D(L) \subset X$. Inoltre $x^\dagger$ è sostituito da $x_0 = A_*^\dagger y$.

Osservazione 2.3.5

$$R(K)^\perp = N(A^\sharp).$$

Infatti: $x \in N(A^\sharp) \iff A^\sharp x = 0 \iff (K^*K + L^*L)^{-1}K^*x = 0 \iff K^*x = 0 \iff (K^*x, y) = 0\ \forall y \in X \iff (x, Ky) = 0\ \forall y \in X \iff x \in R(K)^\perp$.
In particolare $R(A) \oplus N(A^\sharp) = R(A) \oplus R(K)^\perp$.

Osservazione 2.3.6 *Diamo ora una nuova caratterizzazione a* $x_0 = A^\dagger_* y \quad \forall y \in D(A^\dagger_*)$. *Indichiamo con*

$$S_y = \{x \in D(L) : K^*Kx = K^*y\}$$

l' insieme delle soluzioni ai minimi quadrati di $Kx = y$ *appartenenti a* $D(L)$ *e dimostriamo che*

$$\|Lx_0\| \leq \|Lx\| \quad \forall x \in S_y.$$

In $D(L)$ *l' equazione* $K^*Kx = K^*y$ *è equivalente a* $A^\sharp Ax = A^\sharp y$ *e quindi, per il teorema (2.1.5),* $S_y \neq \emptyset$ *se e solo se* $y \in D(A^\dagger_*)$. *In tal caso*

$$S_y = N(A) + x_0 = N(K) \cap D(L) + x_0.$$

Supponiamo $y \in D(A^\dagger_*)$ *e per ogni* $x \in S_y \neq \emptyset$ *poniamo* $x = u + x_0$ *con* $u \in N(K) \cap D(L)$. *Allora* $x_0 = A^\dagger_* y$, $Kx_0 = Kx$, *e dalla definizione di* x_0 *abbiamo*

$$\|x_0\|^2_* = \|Kx_0\|^2 + \|Lx_0\|^2 \leq \|x\|_* = \|Kx\|^2 + \|Lx\|^2 = \|Kx_0\|^2 + \|Lx\|^2$$

ovvero

$$\|Lx_0\| \leq \|Lx\| \quad \forall x \in S_y.$$

Forniamo infine una stima, rispetto a $\| \ \|_*$, dell' errore che si commette approssimando x_0 con x_λ.

Proposizione 2.3.7 *Nelle ipotesi del teorema 2.3.2, sia* $y \in D(A^\dagger_*)$, $x_0 = A^\dagger_* y$ *e* $\forall R > 0$ *sia* δ_R *definito da*

$$\delta_R = inf\{\|A^\sharp \omega - x_0\|_* : \omega \in Y, \|\omega\| \leq R\}.$$

Allora

$$\|x_\lambda - x_0\|_* \leq \frac{\delta_R}{2} + \frac{1}{2}[\lambda^2(1-\lambda^2)^{-1}(R + \|Kx_0\|)^2 + \delta_R^2]^{1/2}.$$

Dimostrazione. La dimostrazione è ottenuta sfruttando alcune relazioni prima definite. Si consulti [18]. ∎

Abbiamo pertanto verificato che se K e L sono operatori per cui valgono le ipotesi del teorema 2.3.2 il problema (2.1) nel senso della regolarizzazione generale è ben posto.

2.4 Legame tra regolarizzazione e minimi quadrati

Nelle precedenti sezioni abbiamo studiato l' esistenza, l' unicità e la convergenza della soluzione regolarizzata generale x_λ del metodo di Tikhonov. Il precedente approccio è stato condotto utilizzando l' equazione di Eulero generalizzata (2.18) che fornisce una forte caratterizzazione della soluzione chiarendo, come osservato in 2.2.4 e 2.3.4, i motivi per cui la soluzione così ottenuta sia regolare e priva dell' instabilità tipica della soluzione generalizzata $x^\dagger = K^\dagger y$. Vogliamo ora considerare la regolarizzazione generale sotto una nuova prospettiva ottenuta trattando il problema (2.1) come "processo ai minimi quadrati in $Y \otimes Y$. A tale scopo consideriamo nello spazio prodotto

$$Y \otimes Y = \{(x, y) : x \in Y,\ y \in Y\}$$

il prodotto standard e la relativa norma già definita in (2.16).
Per ogni $\lambda \neq 0$ sia T_λ l' operatore lineare da X in $Y \otimes Y$ definito da:

$$D(T_\lambda) = D(L), \qquad T_\lambda(x) = (Kx, \lambda Lx).$$

L' operatore T_λ è chiuso e definito per densità in X con

$$N(T_\lambda) = N(K) \cap N(L),$$

e per ogni $x \in D(L)$ e $z = (\xi, \eta) \in Y \otimes Y$ si ha:

$$\|T_\lambda x - z\|^2 = \|Kx - \xi\|^2 + \lambda^2 \|Lx - \eta\|^2. \tag{2.31}$$

In particolare, se $y \in Y$ e posto $\tilde{y} = (y, 0) \in Y \otimes Y$, si ha

$$\|T_\lambda x - \tilde{y}\|^2 = \|Kx - y\|^2 + \lambda^2 \|Lx\|^2 = G_\lambda(x). \tag{2.32}$$

$\forall x \in D(T_\lambda) = D(L)$.
Dimostriamo ora che:

$$D(T_\lambda^*) = D(L^*L) \text{ e } T_\lambda^* = K^*K + \lambda^2 L^*L.$$

Per definizione di aggiunto si deve avere:
$< T_\lambda x, (x^{'}, y^{'}) >=< x, T_\lambda^*(x^{'}, y^{'}) > \forall x \in D(T_\lambda) = D(L), \forall (x^{'}, y^{'}) \in D(T_\lambda^*)$
ovvero $< T_\lambda x, (x^{'}, y^{'}) >=< (Kx, \lambda Lx), (x^{'}, y^{'}) >= (Kx, x^{'}) + (\lambda Lx, y^{'}) = (x, K^*x^{'}) + \lambda(x, L^*y') = (x, K^*x' + \lambda L^*y')$ da cui ottenimo la tesi.

Lemma 2.4.1 *Se gli operatori K e L soddisfano le ipotesi fatte allora $\forall \lambda \neq 0$ $R(T_\lambda)$ è chiuso in $Y \otimes Y$.*

Dimostrazione.
La dimostrazione è analoga a quella del lemma 2.2.5 ∎

Possiamo, attraverso questo lemma, ottenere la relazione tra regolarizzazione generale e approccio nel senso dei minimi quadrati.
Indichiamo con Q_λ l' operatore di proiezione ortogonale da $Y \otimes Y$ a $R(T_\lambda)$.

Teorema 2.4.2 *Se K e' un operatore lineare e L di regolarizzazione per cui valgono le ipotesi (2.22), (2.23), (2.24) allora valgono i seguenti fatti:*
(1) T_λ è iniettivo, $R(T_\lambda)$ è chiuso, $T_\lambda^\dagger = (T_\lambda)^{-1}Q_\lambda$ è definito in $Y \otimes Y$ ed è limitato.
(2) Per ogni $y \in Y$ e per ogni $\lambda \neq 0$ esiste un unico elemento $x_\lambda \in D(L)$ tale che

$$\|Kx_\lambda - y\|^2 + \lambda^2\|Lx_\lambda\|^2 = \inf_{x \in D(L)}[\|Kx - y\|^2 + \lambda^2\|Lx\|^2].$$

(3) L' elemento x_λ rappresenta anche l' unica soluzione ai minimi quadrati dell' equazione

$$T_\lambda x = \tilde{y} \tag{2.33}$$

dove $\tilde{y} = (y, 0)$ ovvero $x_\lambda = T_\lambda^\dagger \tilde{y}$.

Dimostrazione.
Basta applicare il teorema (2.1.5) all' operatore T_λ sfruttando l' espressione del suo aggiunto ottenuta precedentemente. ∎

2.5 Inverso L-generalizzato

Vogliamo ora estendere il concetto di operatore inverso generalizzato al caso della regolarizzazione generale. A tale scopo diamo alcune definizioni e risultati [22] che ci serviranno nel seguito.

Definizione 2.5.1 *$T \in \Lambda(X)$ è detto operatore semidefinito positivo se $(Tx, x) \geq 0 \ \forall x \in X$ e operatore definito positivo se $(Tx, x) > 0 \ \forall x \in X - \{0\}$.*

Naturalmente l' operatore T^*T è semi-definito positivo in quanto:

$$\forall x \in X \quad (T^*Tx, x) = \|Tx\|^2 \geq 0.$$

Teorema 2.5.2 *(1) Per ogni operatore positivo $T \in B(X)$ esiste un unico $S \in B(X)$ tale che $T = S^2$. Se T è invertibile allora lo è anche S. L' operatore S è detto radice quadrata di T e verrà indicata con $T^{1/2}$.*
(2) Se $T \in B(X)$, allora la radice quadrata di (T^*T) è l' unico operatore positivo in $B(X)$ che soddisfa l' uguaglianza:

$$\|(T^*T)^{1/2}x\| = \|Tx\| \ \ \forall x \in X \tag{2.34}$$

(3) Se $T \in B(X, Y)$ è invertibile, X e Y sono spazi di Hilbert, allora T ha un' unica decomposizione polare $T = UP$ intendendo con questo che

$$U : X \to Y \ \ e' \ \ un' \ \ isometria, P = (T^*T)^{1/2}. \tag{2.35}$$

Dimostrazione.
Per la dimostrazione si consulti [22]. ∎

Per poter definire l' operatore inverso L-generalizzato dobbiamo fare delle ipotesi aggiuntive sull' operatore di regolarizzazione L: più precisamente richiediamo che $L : D(L) \subset X \to Y$ soddisfi la relazione

$$\|Lx\| \geq \beta\|x\| \ \ \forall x \in D(L). \tag{2.36}$$

Indichiamo ora alcune brevi osservazioni che useremo nel seguito.

Osservazione 2.5.3 *(1) $L : D(L) \subset X \to R(L)$ è bigettivo. Naturalmente L è surgettivo mentre l' iniettività è garantita da (2.36). Infatti:*

$$0 = \|Lx\| \geq \beta\|x\| \Longrightarrow x = 0$$

ovvero $N(L) = \{0\}$
(2) L è continuo come operatore da $(D(L), \| \ \|_L)$ in Y.
(3) $L^{-1} : R(L) \to (D(L), \| \ \|_X)$ è continuo. Basta dimostrare che $\exists \alpha \in \mathbb{R}^+ - \{0\}$ tale che $\forall y \in R(L)$ $\|L^{-1}y\| \leq \alpha\|y\|$. Poichè $y \in R(L)$ si ha $y = Lx$ e quindi

$$\|L^{-1}y\| = \|x\| \leq 1/\beta\|Lx\| = 1/\beta\|y\|.$$

Posto $T = L^*L : (D(T), \| \ \|_*) \subset X \to R(T) \subset Y$ si ha che T è lineare, continuo (composto di operatori lineari e continui) e poichè $N(L^*L) = N(L) = \{0\}$ l' operatore T è invertbile. Applicando il teorema precedente otteniamo che T ammette la decomposizione polare: $T = U(L^*L)^{1/2}$ con $\|Lx\| = \|(L^*L)^{1/2}x\|$ $\forall x \in D(L)$.

Definizione 2.5.4 $x^\dagger_L$ *è detta* **pseudosoluzione L-generalizzata** *[2] per $Kx = y$ se risolve il problema ai minimi quadrati vincolati*

$$\|Lx^\dagger{}_L\| = \min_{x \in S_y}\{\|Lx\|\} \tag{2.37}$$

con $S_y = \{x \in D(L) : x$ soluzione ai minimi quadrati di $Kx = y\}$ $\forall y \in D(K^\dagger) = R(K) \oplus R(K)^\perp$.
Osserviamo che se $y \notin D(K^\dagger)$ si avrebbe $S_y = \emptyset$ e quindi non sarebbe definito $x^\dagger_L$.

Proposizione 2.5.5 *Esiste un' unica pseudosoluzione L-generalizzata del problema (2.6) se e solo se*

$$y \in D(A^\dagger_*) = R(K|D(L)) \oplus R(K)^\perp \tag{2.38}$$

Dimostrazione.
"$\Longrightarrow$
Per come è definita $x^\dagger{}_L$ si ha, usando (2.6)

$$x^\dagger{}_L = x^\dagger + u \quad u \in N(K)$$

da cui si ottiene:

$$Kx^{\dagger}{}_L = Kx^{\dagger} + Ku = Kx^{\dagger} = KK^{\dagger}y = P_{\overline{R(K)}}y$$

e quindi, essendo $x^{\dagger}{}_L \in D(L)$, otteniamo la tesi.
"$\Longleftarrow$
Per dimostrare l' implicazione opposta ricorriamo alla decomposizione polare di L data da

$$L = U(L^*L)^{1/2} \quad \text{con} \quad U : X \to Y \quad \text{operatore isometrico.}$$

Allora
-$D((L^*L)^{1/2}) = D(L)$,
-$R((L^*L)^{1/2}) = X$.
Pertanto $(L^*L)^{-1/2}$ è definito da X a $D(L)$ e $\forall u \in D(L)$ $\exists w \in X$ tale che $u = (L^*L)^{-1/2}w$ e quindi $\|Lu\| = \|L(L^*L)^{-1/2}w\| = \|Uw\| = \|w\|$. Posto infine $C = K(L^*L)^{-1/2}$ abbiamo

$$\|Ku - y\| = \|K(L^*L)^{-1/2}w - y\| = \|Cw - y\|.$$

Il problema è allora ricondotto a determinare la pseudosoluzione di

$$Cw = y.$$

Ricordando (2.5) otteniamo che esiste un' unica

$$w^{\dagger} = C^{\dagger}y$$

se $y \in R(C^{\dagger}) = R(C) \oplus R(C)^{\perp} = R(K|D(L)) \oplus R(K)^{\perp}$ da cui si ottiene che

$$x^{\dagger}{}_L = (L^*L)^{-1/2}C^{\dagger}y = (L^*L)^{-1/2}(K(L^*L)^{-1/2})^{\dagger}y. \tag{2.39}$$

■

Usando la proposizione precedente possiamo introdurre la nozione di operatore inverso L-generalizzato.

Definizione 2.5.6 *Diciamo inverso L-generalizzato dell' operatore lineare e continuo $K : X \to Y$, l' operatore*

$$K^{\dagger}{}_L : R(K|D(L)) \oplus R(K)^{\perp} \to X \tag{2.40}$$

$$K^{\dagger}{}_L y = x^{\dagger}_L \tag{2.41}$$

dove $x^{\dagger}_L$ è l' unica pseudosoluzione L-generalizzata del problema (2.1).

Osservazione 2.5.7 *Se l' operatore K è iniettivo allora*

$$K^{\dagger}{}_{L} = (K^{-1})|D(L)$$

infatti

$$K_L^{\dagger} = (L^*L)^{-1/2}(K(L^*L)^{-1/2})^{\dagger} = K^{-1}|D(L).$$

Inoltre la soluzione $x^{\dagger}{}_{L}$ coincide con x_0 se $y \in D(A_*^{\dagger})$. Infatti per l' osservazione (2.3.6) si ha

$$\|Lx_0\| \leq \|Lx\| \quad \forall x \in S_y$$

da cui discende la tesi.

2.6 La soluzione attraverso il sistema singolare

Vogliamo ora esprimere la soluzione regolarizzata x_λ di (2.18) utilizzando il sistema singolare di opportuni operatori costruiti a partire da K e L: tale espressione [9] trovata nel caso continuo verrà poi ricavata nel discreto a partire dalla $GSVD$ di (A, L) [3] mantenendo così il parallelismo tra approccio teorico e numerico alla regolarizzazione in analogia con il caso standard.
Supponiamo che K sia un operatore compatto e indichiamo con $L_*^{\dagger}$ l' inverso generalizzato di L rispetto alla norma $\| \ \|_*$. Allora [9] $(L_*^{\dagger})^*K^*KL_*^{\dagger}$ è compatto e esiste una base ortonormale $(w_j)_{j=1}^{+\infty}$ di $R(L)$ e una famiglia di valori singolari non negativi $(\gamma_j)_{j=1}^{+\infty}$ tali che:

$$(L_*^{\dagger})^*K^*KL_*^{\dagger}w_j = {\gamma_j}^2 w_j \quad \forall j = 1, 2, ..., +\infty. \tag{2.42}$$

Supponiamo d' ora in poi che $\gamma_1 \geq \gamma_2 \geq ... \geq 0$. Per ogni j sia $\mu_j \geq 0$ arbitrario e definiamo $v_j = \mu_j L^{-1}w_j$. Allora, $Lv_j = \mu_j w_j$ e ricordando che $L_*^{\dagger}w_j = L^{-1}w_j \quad \forall j$ si ha

$$KL_*^{\dagger}(L_*^{\dagger})^*K^*Kv_j = {\gamma_j}^2 Kv_j. \tag{2.43}$$

Detto $D_1 = R(KL_*^{\dagger})$ dimostriamo che $(Kv_j)_{j=1}^{+\infty}$ è una sua base ortogonale. La relazione $(Kv_j)_{j=1}^{+\infty} \subseteq D_1$ è vera per costruzione e dobbiamo pertanto verificare soltanto l' ortogonalità dei vettori. Moltiplicando entrambi i membri di (2.42) per $\mu_j L^*$ otteniamo

$$\mu_j L^*(L_*^{\dagger})^*K^*KL_*^{\dagger}w_j = {\gamma_j}^2 \mu_j L^*w_j \iff L^*(L_*^{\dagger})^*K^*Kv_j = {\gamma_j}^2 \mu_j L^*w_j = \gamma_j^2 L^*Lv_j$$

[3] Qui A rappresenta la matrice di discretizzazione dell' operatore K.

e quindi usando la relazione $L^*(L_*^\dagger)^* K^* b = K^* P_1 b$ con $P_1 = P_{R(KL_*^\dagger)}$, si ha

$$K^* K v_j = \gamma_j^2 L^* L v_j \quad \forall j = 1, 2, ... \tag{2.44}$$

da cui segue che

$$(Kv_i, Kv_j) = (v_i, K^* K v_j) = {\gamma_j}^2 (v_i, L^* L v_j) = {\gamma_j}^2 \mu_j \mu_i (w_i, w_j) = \gamma_j^2 \mu_i \mu_j \delta_{i,j}$$

come volevamo dimostrare.
Introduciamo ora i valori singolari $\sigma_j = \gamma_j \mu_j \quad \forall j = 1, 2, ...$ e la successione $(y_j)_{j=1}^{+\infty}$ definita da

$$K v_j = \sigma_j y_j \quad \text{se} \quad \sigma_j \neq 0. \tag{2.45}$$

Poichè

$$\|Kv_i\|^2 = \gamma_i^2 \mu_i^2 = \sigma_i^2$$

deduciamo che $(y_j)_{j=1}^{+\infty}$ costituisce una base ortonormale di D_1. Scegliamo ora μ_j in modo unico imponendo che valga la relazione $\sigma_j^2 + \mu_j^2 = 1 \quad \forall j = 1, 2, ...$.
Poiche K è compatto esiste una base $(v_{-j})_{j=1}^{+\infty}$ di $N(L)$ tale che

$$(Kv_{-i}, Kv_{-j}) = \delta_{i,j}$$

e quindi posto

$$y_{-j} = K v_{-j} \quad \text{se} \quad \sigma_{-j} = 1, \quad j > 0$$

otteniamo che la relazione (2.45) è valida per ogni $j \neq 0$. Pertanto $(v_j)_{j \neq 0}$ è una base di $D(L)$ e $(y_j)_{j \neq 0}$ è una base ortonormale di $R(K|D(L))$.
Dalle relazioni precedenti segue che

$$L^* w_j = \mu_j v_j \quad \forall j > 0.$$

Abbiamo allora introdotto tutti gli elementi che ci permettono di scrivere la soluzione x_λ del problema

$$min_{x \in D(L)} G_\lambda(x)$$

in termini dei vettori precedentemente definiti e dei valori $\gamma_j = \sigma_j / \mu_j$ detti valori singolari generalizzati di (K, L).
Ricaviamo pertanto x_λ utilizzando l' equazione di Eulero generalizzata

$$(K^* K + \lambda^2 L^* L) x_\lambda = K^* y$$

a partire da

$$x_\lambda = \Sigma_{j\neq 0} a_j v_j \ \in D(L^*L).$$

con $(a_j)_{j\neq 0}$ da determinare. Allora utilizzando (2.44) e (2.45) otteniamo

$(K^*K + \lambda^2 L^*L)x_\lambda = \sum_{j<0} a_j(K^*K + \lambda^2 L^*L)v_j + \sum_{j>0} a_j(K^*K + \lambda^2 L^*L)v_j =$

$\sum_{j>0} a_{-j}(K^*Kv_{-j}) + \sum_{j>0} a_j(\gamma_j^2 + \lambda^2)L^*Lv_j =$

$\sum_{j>0} a_{-j}v_{-j} + \sum_{j>0} a_j(\gamma_j^2 + \lambda^2)L^*(\mu_j w_j) =$

$\sum_{j>0} a_{-j}v_{-j} + \sum_{j>0} a_j\mu_j^2(\gamma_j^2 + \lambda^2)v_j.$

Utilizzando l' ortogonalità di $(v_j)_{j\neq 0}$ si ottiene:

$$\begin{cases} a_{-j} = (K^*y, v_{-j}) = (y, Kv_{-j}) = (y, y_{-j}) & \text{se } j > 0 \\ a_j\mu_j^2(\gamma_j^2 + \lambda^2) = (K^*y, v_j) = (y, Kv_j) = \sigma_j(y, y_j) & \text{se } j > 0 \end{cases}$$

da cui deduciamo l' espressione

$$x_\lambda = \sum_{j>0}(y, y_{-j})v_{-j} + \sum_{j>0} \frac{\sigma_j}{\gamma_j^2 + \lambda^2}\frac{(y, y_j)}{\mu_j^2}v_j = \sum_{j>0}(y, y_{-j})v_{-j} + \sum_{j>0,\gamma_j\neq 0} \frac{\gamma_j^2}{\gamma_j^2 + \lambda^2}\frac{(y, y_j)}{\sigma_j}v_j. \tag{2.46}$$

Posto $\bar{x} = \sum_{j>0}(y, y_{-j})v_{-j}$ e supposto $y \in D(A_*^\dagger)$ le precedenti relazioni e $A_*^\dagger$ possono essere scritte nella forma:

$$A_*^\dagger y = \sum_{j<0}(y, y_j)v_j + \sum_{j>0,\sigma_j\neq 0} \frac{(y, y_j)}{\sigma_j}v_j = \bar{x} + \sum_{j>0,\sigma_j\neq 0} \frac{(y_j, y)}{\sigma_j}v_j \tag{2.47}$$

$$x_\lambda = \bar{x} + \sum_{j>0,\sigma_j\neq 0} \frac{(y_j, y)}{\sigma_j}\gamma_j^2 r(\gamma_j^2)v_j \tag{2.48}$$

con $r(x) = \frac{1}{\lambda^2 + x}$. Pertanto la funzione $F(x) = xr(x)$ svolge il ruolo di funzione di filtro in quanto elimina i contributi in (2.48) corrispondenti a valori singolari generalizzati troppo piccoli. Scegliendo

$$r(x) = \begin{cases} 0 & \text{se } x > \gamma_p^2 > 0 \\ \frac{1}{x} & \text{se } 0 < x \leq \gamma_p^2 \end{cases}$$

la (2.48) diventa

$$x_r = \sum_{j<0}(y_j, y)v_j + \sum_{j=1}^{p} \frac{(y_j, y)}{\sigma_j} \tag{2.49}$$

detta approssimazione troncata ai valori singolari generalizzati di $A_*^\dagger y$.
Vedremo nel capitolo 3 una espressione analoga a (2.46) per x_λ e nel capitolo 4 una analoga a (2.49) valide entrambe nel caso discreto.

Capitolo 3

Approccio numerico alla regolarizzazione generale

Nel capitolo 2 abbiamo trattato la regolarizzazione generale dal punto di vista teorico evidenziando tutti gli elementi necessari a fornire un quadro il più possibile generale. Affronteremo in questo capitolo lo stesso problema ma dal punto di vista discreto attraverso la $GSVD$ introdotta nel capitolo 1. Verranno quindi sottolineate e chiarite le analogie tra approccio teorico e discreto evidenziando che i risultati presentati nel capitolo precedente sono quelli necessari a uno studio concreto dei problemi inversi.

3.1 Derivazione del problema discreto

Per la trattazione numerica del problema $Kx = y$ e dei relativi metodi di regolarizzazione è necessario discretizzare il problema continuo riducendolo a un sistema finito di equazioni lineari.
Denotiamo con $X_n \subset X$ e $Y_m \subset Y$ due sottospazi vettoriali di dimensione n e m con basi ortogonali:

$$X_n = span\{\varphi_1, ..., \varphi_n\}, \quad Y_m = span\{\phi_1, ..., \phi_m\}$$

e definiamo $x_n \in X_n$ come la soluzione dell' equazione proiettata

$$Q_m K x_n = Q_m y \tag{3.1}$$

dove $Q_m : Y \to Y_m$ è un operatore di proiezione. In generale a x_n possiamo associare il vettore $\xi \in M_{n,1}(\mathbb{R})$ delle sue coordinate rispetto alla base $(\varphi_i)_{i=1}^n$

corrispondente alla rappresentazione

$$x_n = \sum_{i=1}^{n} \xi_i \varphi_i, \quad \xi = (\xi_i)_{i=1}^{n}. \tag{3.2}$$

L' equazione proiettata è pertanto equivalente al sistema lineare $m \times n$ con matrice dei coefficienti $A = ((Kx_i, y_j))_{i,j}$ e termine noto $b = ((y, y_j))_j$ nelle incognite $\xi_i \quad i = 1, 2, ..., n$.

Teorema 3.1.1 *Siano x_n e $\{Y_i\}_i$ definiti come sopra e tali che $Y_i \subset \overline{R(K)} = N(K^*)^{\perp}$. Se $y \in D(K^{\dagger})$ allora:*
(1) $x_n = P_n x^{\dagger}$ con P_n proiezione ortogonale su $X_n = K^* Y_n$
(2) $x_n \to x^{\dagger} \quad n \to +\infty$.

Il teorema precedente [10] afferma che, sotto opportune ipotesi, x_n converge a $x^{\dagger}$; pertanto, per la definizione di algoritmo regolarizzante 2.1.13 e la proprietà (2), il processo di discretizzazione rappresenta di per sè un metodo di regolarizzazione. Va tuttavia sottolineato che il ricorso a dimensioni via via maggiori di X_n crea forti problemi numerici dovuti al fatto che cresce il numero di condizionamento della matrice di discretizzazione dell' operatore $Q_m K$ ovvero la sensibiltà della soluzione x_n rispetto a perturbazioni sul dato. Pertanto la discretizzazione non può rappresentare di per sè un metodo sufficiente a determinare una soluzione significativa rendendo necessario il ricorso alla regolarizzazione.
Per i nostri esempi sceglieremo come L la discretizzazione di operatori di derivata prima o seconda per i legami che hanno con i concetti di tangente e curvatura.
La scelta di A e b è più delicata e dipende dal tipo di discretizzazione adottato. Per i test di questo capitolo utilizzeremo equazioni di Fredholm del primo tipo

$$K : L^2([a, b]) \to L^2([c, d])$$

$$(Kf)(s) = \int_a^b K(s, t) f(t) dt \tag{3.3}$$

con $K \in L^2([c, d] \times [a, b])$ in cui l' integrale è discretizzato attraverso le formule di quadratura

$$(Kf)(s) \approx I(s) = \sum_{i=1}^{n} \omega_i K(s, t_i) f(t_i) \tag{3.4}$$

con ω_i pesi opportuni. Utilizzando, ad esempio, le formule di quadratura con punto di mezzo si ha

$$\omega_i = \frac{b-a}{n}, \quad t_i = \frac{(i-1/2)(b-a)}{n} \quad \forall i = 1, 2, ..., n. \tag{3.5}$$

La soluzione del problema (3.5) è ricondotta, mediante il metodo di collocazione su un reticolo $\{s_i\}_{i=1}^m \subset [c, d]$, alla risoluzione del sistema lineare $Ax = b$ con

$$A = (a_{i,j})_{i,j} = (\omega_j K(s_i, t_j))_{i.j}, \quad b_i = g(s_i), \quad \forall i, j \tag{3.6}$$

in cui $x = (x_i)_{i=1}^n = (f(s_i))_{i=1}^n$.
In generale, il problema $Kx = y$ si caratterizza quindi come sistema lineare

$$Ax = b \tag{3.7}$$

dove la matrice dei coefficienti A è l' approssimazione numerica dell' operatore K e il termine noto b è ottenuto valutando il dato y nei punti di discretizzazione.
Per le osservazioni fatte nel capitolo 2 il concetto di problema mal posto è legato alla dimensione infinita di $R(K)$. Infatti, indicata con $\bar{x}^\dagger$ la pseudosoluzione del problema $Ax = \bar{y}$ e con $x^\dagger$ quella del problema perturbato $Ax = y := \bar{y} + e$, se $dim(R(K)) < +\infty$ il rapporto $\frac{\|\triangle x\|}{\|\triangle y\|} = \frac{\|x^\dagger - \bar{x}^\dagger\|_X}{\|e\|_Y}$ rimane finito, anche se molto grande, e quindi il problema (3.7) è sempre ben posto. Tuttavia alcuni problemi discreti di dimensione finita hanno proprietà molto simili ai problemi mal posti risultando molto sensibili alle perturbazioni dei dati. Più precisamente considerato il sistema lineare (3.7) diciamo che questo rappresenta un problema discreto " mal posto se sono soddisfatte entrambe le condizioni di seguito riportate:
(1) i valori singolari di A convergono a zero all' aumentare di n (i.e $\sigma_n \approx 0$ al crescere di n)
(2) il condizionamento di A è molto grande e cresce con n.
Sottolineiamo che la manifestazione tipica dei problemi discreti mal posti è determinata proprio dalla discretizzazione di quelli continui il cui studio è stato affrontato nel capitolo precedente. Motiviamo brevemente la proprietà 2.
Supponiamo di avere discretizzato K per mezzo di una matrice $A \in M_{n,n}(\mathbb{R})$

invertibile e si consideri la decomposizione in valori singolari della matrice A, ossia

$$A = \sum_{i=1}^{n} u_i \sigma_i {v_i}^T \tag{3.8}$$

con $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n > 0$.
Si supponga ora di realizzare una discretizzazione più fine che dia luogo a una matrice $\tilde{A} \in Gl_{n'}(\mathbb{R})$ $(n' > n)$. Considerate allora le rispettive inverse generalizzate $A^\dagger$ e $\tilde{A}^\dagger$ i valori singolari di $A^\dagger$ sono

$${\sigma_1}^{-1} \leq {\sigma_2}^{-1} \leq ... \leq {\sigma_n}^{-1}$$

mentre quelli di $\tilde{A}^\dagger$ sono

$${\tilde{\sigma_1}}^{-1} \leq {\tilde{\sigma_2}}^{-1} \leq ... \leq {\tilde{\sigma_{n'}}}^{-1}.$$

Si ottiene allora

$$cond_2(A^\dagger) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$$

e

$$cond_2(\tilde{A}^\dagger) = \|\tilde{A}\|_2 \|\tilde{A}^{-1}\|_2 = \frac{\tilde{\sigma}_1}{\tilde{\sigma}_{n'}}.$$

Supposto che σ_n e $\sigma_{n'}$ non differiscano di molto [5] essendo, per l' ipotesi (1), $\sigma_n >> \tilde{\sigma_{n'}}$ si ha

$$cond_2(\tilde{A}^\dagger) >> cond_2(A^\dagger)$$

come volevamo dimostrare.

3.2 Metodi numerici standard

Sia $A \in M_{m,n}(\mathbb{R})$ una matrice rettangolare con $m \geq n$. Allora la SVD di A è una decomposizione espressa da

$$A = U \Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T, \tag{3.9}$$

dove $U = (u_1, ..., u_m) \in M_{m,m}(\mathbb{R})$ e $V = (v_1, ..., v_n) \in M_{n,n}(\mathbb{R})$ sono matrici ortonormali, mentre $\Sigma = diag(\sigma_1, ..., \sigma_n) \in M_{m,n}(\mathbb{R})$ ha elementi diagonali non negativi ordinati in modo decrescente

$$\sigma_1 \geq ... \geq \sigma_n \geq 0. \tag{3.10}$$

In connessione con i problemi discreti mal-posti vanno evidenziati le seguenti caratteristiche della SVD di A:
(1) valori singolari di A convergono a zero al crescere di n. Ad un aumento delle dimensioni di A corrisponde un numero maggiore di valori singolari piccoli.
(2) i vettori singolari destri e sinistri u_i e v_i hanno oscillazioni in numero crescente con l' indice i.
Vediamo ora come la SVD permette di determinare la pseudoinversa di una generica matrice.
Considerato il problema LS $(Least - square)$

$$min\{\|Ax - b\|_2^2\} \tag{3.11}$$

tutto ciò che è stato detto nella sezione (2.1) può essere ripetuto in modo analogo tenendo conto che $X = \mathbb{R}^n$ e $Y = \mathbb{R}^m$ hanno dimensione finita. Inoltre, attraverso la SVD di A otteniamo che la soluzione generalizzata del problema (3.7) è esprimibile come

$$x^\dagger = \sum_{i=1}^{rank(A)} \frac{u_i^T b}{\sigma_i} v_i. \tag{3.12}$$

Questa relazione illustra, in analogia con il caso continuo, le difficoltà connesse alla soluzione di (3.11): quando i coefficienti di Fourier $|u_i^T b|$ relativi ai σ_i più piccoli non decadono velocemente come i valori singolari, la soluzione $x^\dagger$ è caratterizzata fortemente dai termini della sommatoria corrispondenti ai σ_i più piccoli e questo comporta che $x^\dagger$ presenta, per (2), molte oscillazioni risultando così ampia.
Se defininiamo la matrice $A^\dagger \in M_{n,m}(\mathbb{R})$ come $A^\dagger = V\Sigma^\dagger U^T$ dove

$$\Sigma^\dagger = diag(\frac{1}{\sigma_1}, ..., \frac{1}{\sigma_{rank(A)}}, 0, ..., 0) \in M_{n,m}(\mathbb{R}) \tag{3.13}$$

allora

$$x^\dagger = A^\dagger b \quad \text{e} \quad \rho_{LS} = \|Ax^\dagger - b\|_2 = \sum_{i=rank(A)+1}^{m} (u_i^T b)^2 = \|(I - A^\dagger A)b\|_2. \tag{3.14}$$

Così come nel caso continuo, $A^\dagger$ rappresenta la pseudoinversa di A ed è definita come l' unica matrice $X \in M_{n,m}(\mathbb{R})$ che soddisfa le quattro condizioni di Moore-Penrose:
(1) $AXA = A$ (2) $(AX)^T = AX$ (3) $XAX = X$ (4) $(XA)^T = XA$ con

$k = rank(A)$.
Queste condizioni equivalgono alla richiesta che $AA^\dagger$ e $A^\dagger A$ siano le proiezioni ortogonali sul range di A e A^T rispettivamente: abbiamo quindi $AA^\dagger = U_1 U_1^T$ con $U_1 = U(1 : m, 1 : k)$ e $A^\dagger A = V_1 V_1^T$ con $V_1 = V(1 : n, 1 : k)$ con $k = rank(A)$.

Definizione 3.2.1 *Definiamo $TSVD$ di A la matrice*

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T \tag{3.15}$$

con $k < rank(A)$. Per la $TSVD$ di A vale la seguente caratterizzazione.

Teorema 3.2.2

$$\forall k < rank(A) \quad \min_{rank(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}.$$

Dimostrazione. Si consulti [7]. ∎

Il teorema precedente permette di concludere che la matrice A_k è l' elemento in $\{X \in M_{m,n}(\mathbb{R}) : rank(X) = k\}$ che rende minima la distanza tra A e un generico elemento di questo insieme.

Abbiamo già evidenziato nel capitolo 2 che lo scopo di un qualunque metodo di regolarizzazione è quello di smorzare, mediante l' uso di opportuni filtri, i contributi sulla soluzione dovuti ai valori singolari più piccoli.
Nel nostro studio prenderemo in considerazione metodi di regolarizzazione per un problema mal posto discreto che producano una soluzione regolarizzata x_λ che possa essere scritta in uno dei seguenti modi:

$$x_\lambda = \sum_{i=1}^{n} f_i \frac{u_i^T b}{\sigma_i} v_i \tag{3.16}$$

$$x_\lambda = \sum_{i=1}^{p} f_i \frac{u_i^T b}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T b) x_i \tag{3.17}$$

a seconda che si stia considerando un metodo di regolarizzazione standard ($L = I_n$) oppure generalizzato ($L \neq I_n$). Osserviamo subito che i vettori $(u_i)_{i=1}^n$ e i valori $(\sigma_i)_{i=1}^n$ in (3.16) non hanno, per $L \neq I$, alcun legame con

i corrispondenti elementi in (3.17). In entrambi i casi i fattori di filtro, pur risultando diversi, dovranno essere scelti in modo tale che, quando σ_i decresce, il corrispondente fattore f_i tenda a zero affinchè i contributi $(u_i^T b/\sigma_i)x_i$ sulla soluzione vengano filtrati nel modo migliore possibile.
Più precisamente i valori $(f_i)_{i=1}^p$ in (3.16) e (3.17) sono detti filtri se soddisfano le relazioni:

- $|f_i| \leq 1 \quad \forall \lambda > 0$
- $lim_{\lambda \to 0} f_i = 0 \quad \forall \ \lambda \ \text{e} \ \sigma_i \neq 0$
- $\frac{f_i}{\sigma_i}$ sono limitati per ogni i e λ fissato.

Se la soluzione, per esempio, è ottenuta implementando il metodo di regolarizzazione di Tikhonov standard

$$x_\lambda = min\{\|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2\} \tag{3.18}$$

si ha, utilizzando la SVD di A,

$$x_\lambda = \sum_{i=1}^{rank(A)} \frac{{\sigma_i}^2}{\sigma_i^2 + \lambda^2} \frac{u_i^T b}{\sigma_i} v_i \tag{3.19}$$

e pertanto vale

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \quad \forall i = 1, 2, ..., rank(A). \tag{3.20}$$

3.3 $GSVD$ e regolarizzazione generale

Supporremo da ora in poi

$$A \in M_{m,n}(\mathbb{R}), \quad L \in M_{p,n}(\mathbb{R}), \quad m \geq n \geq p = rank(L).$$

Per studiare la regolarizzazione generale utilizziamo una formulazione della $GSVD$ di (A, L) leggermente diversa da quella data nel capitolo 1 e più adatta al nostro problema.
Scriviamo la $GSVD$ di (A, L) nella forma [11,16] :

$$A = U\Sigma X^{-1} = U \begin{pmatrix} \Sigma_p & 0 \\ 0 & I_{n-p} \end{pmatrix} X^{-1}, \quad L = VMX^{-1} = V(M_p, 0)X^{-1}, \tag{3.21}$$

dove le colonne di $U \in M_{m,n}(\mathbb{R})$ e $V \in M_{p,p}(\mathbb{R})$ sono ortonormali e $X \in Gl_n(\mathbb{R})$. Osserviamo esplicitamente che la differenza sostanziale tra (1.8) e (3.21) è dovuta al fatto che in quest' ultima è presente X^{-1} in luogo di X^T .
Le matrici Σ_p e M_p sono matrici diagonali di ordine p:

$$\Sigma_p = diag(\sigma_1, ..., \sigma_p), \quad M_p = diag(\mu_1, ..., \mu_p) \tag{3.22}$$

in cui gli elementi diagonali sono ordinati come segue

$$0 \leq \sigma_1 \leq ... \leq \sigma_p \leq 1, \quad 1 \geq \mu_1 \geq ... \geq \mu_p \geq 0, \tag{3.23}$$

e normalizzati in modo tale che

$$\sigma_i{}^2 + \mu_i{}^2 = 1 \quad i = 1, ..., p \tag{3.24}$$

In generale non vi è alcun legame tra la $GSVD$ di (A, L) e la SVD di A; pertanto le matrici U, Σ, V relative alla $GSVD$ non vanno confuse con le corrispondenti matrici della SVD di A. È possibile ricavare la $GSVD$ di (A, L) a partire dalla SVD di una opportuna matrice $\bar{A}$ nel caso in cui L abbia rengo p. La trattazione di questa relazione sarà alla base del prossimo capitolo e permetterà la formulazione di un algoritmo per il calcolo della $GSVD$ di (A, L) alternativo a quello di Van Loan, e l' estensione della $TSVD$ al caso della $GSVD$.
Da (3.21) otteniamo alcune relazioni a cui ci riferiremo spesso nel seguito.
Dall' uguaglianza $AX = U\Sigma$ otteniamo

$$\begin{cases} Ax_i = \sigma_i u_i & i = 1, 2, ..., p \\ Ax_i = u_i & i = p+1, ..., n \end{cases}$$

mentre da $LX = VM$ si deduce

$$\begin{cases} Lx_i = \mu_i v_i & i = 1, 2, ..., p \\ Lx_i = 0 & i = p+1, ..., n \end{cases}$$

dove $\{u_i\}_{i=1}^n$, $\{v_i\}_{i=1}^p$, $\{x_i\}_{i=1}^n$ rappresentano le colonne di U, V, X rispettivamente.

Osservazione 3.3.1 *Vogliamo ricavare attraverso la $GSVD$ di (A, L) il nucleo e il range degli operatori A e L. A tale scopo poniamo*

$$I = \{i \in \{1, 2, ..., p\} : \sigma_i \neq 0\} \ e \ \ I^C = C_{\{1,2,...,p\}} I := \{1, 2, ..., p\} - I$$

Allora:
(1) $N(A) = \{x \in M_{n,1}(\mathbb{R}) : Ax = 0\} = span\{x_i : i \in I^C\}$
Infatti
$\forall x \in M_{n,1}(\mathbb{R})\ \ Ax = 0 \iff U\Sigma X^{-1}x = 0 \iff U\Sigma y = 0$ *con* $y = X^{-1}x \iff$
$\Sigma y = 0 \iff \begin{pmatrix} \Sigma_p & 0 \\ 0 & I_{n-p} \end{pmatrix} \begin{pmatrix} y_p \\ y_{n-p} \end{pmatrix} = 0$ *con* $y_p = (y_i)_{i=1}^p$ *e* $y_{n-p} = (y_i)_{i=1+p}^n \iff$

$$\begin{cases} \Sigma_p y_p = 0 \\ \\ y_{n-p} = 0 \end{cases} \iff \begin{cases} y_i = 0 \quad se\ i \in I \\ \\ y_i = 0 \quad se\ i = p+1, ..., n. \end{cases}$$

Pertanto, detto e_k *il* k*-esimo vettore canonico di* $\mathbb{R}^n$*, abbiamo* $\forall k \in I^C\ \ \Sigma e_k = 0$ *ovvero, essendo* $y = X^{-1}x$*,* $\{Xe_k\}_{k \in\ I^C}$ *è una base di* $N(A)$*.*

(2) $Range(A) = span\{u_i\}_{i \in I \cup \{p+1,...,n\}}$ *con* $\{u_i\}_{i=p+1}^n$ *sistema ortonormale ottenuto completando* $\{u_i\}_{i=1}^n$ *a base di* $\mathbb{R}^n$*.*
Infatti:
$y \in R(A) \iff \exists x \in M_{n,1}(\mathbb{R}) : Ax = y \iff U\Sigma X^{-1}x = y \iff \begin{pmatrix} \Sigma_p & p \\ 0 & I_{n-p} \end{pmatrix} z =$

$$U^T y \ \ , z = X^{-1}y \iff \begin{cases} \sigma_i z_i = u_i^T b & \forall i \in I \\ \\ 0 z_i = u_i^T b & \forall i \in I^C \\ \\ z_i = u_i^T b & \forall i = p+1, ..., n \end{cases}$$

e pertanto tale sistema è risolubile se e solo se $u_i^T b = 0\ \forall i \in I^C$*.*

(3) $N(L) = span\{x_{p+1}, ..., x_n\}$
Basta dimostrare che $Lx_i = 0\ \ \forall i = p+1..., n$ *in quanto i vettori* $(x_i)_{i=p+1}^n$ *sono linearmente indipendenti, poichè colonne di* $X \in GL_n(\mathbb{R})$*, e* $dim(N(L)) = n-p$ *essendo* $rank(L) = p$ *per ipotesi. Infatti*

$$L = V\,(\,M_p \quad 0\,)\,X^{-1} \iff LX = (\,VM_p \quad 0\,) \iff Lx_i = 0\,\forall i = p+1, ..., n.$$

(4) $R(L) = \mathbb{R}^p$*. Infatti* $p = rank(L)$*. Si noti come i risultati* $N(A) = span\{Xe_i : i \in I^C\}$ *e* $N(L) = span\{Xe_i : i \in I\}$ *siano in perfetto accordo con l' ipotesi* $N(A) \cap N(L)$*.*

Dalla relazione (3.24) otteniamo che sia σ_i che μ_i possono essere calcolati attraverso i valori singolari generalizzati $\gamma_i = \sigma_i/\mu_i$ mediante le relazioni:

$$\sigma_i = \gamma_i({\gamma_i}^2 + 1)^{-1/2} \quad \mu_i = ({\gamma_i}^2 + 1)^{-1/2} \quad \forall i = 1, 2..., p. \tag{3.25}$$

e pertanto

$$\gamma_i = \sigma_i(1 - {\sigma_i}^2) \approx \sigma_i \quad \text{per} \quad \sigma_i \approx 0$$

ovvero per piccoli valori dei σ_i i valori singolari generalizzati γ_i decadono a zero così come accade per i valori singolari " ordinari di A (ottenuti mediante la SVD).
L' equazione (2.1), il funzionale G_λ (2.17) e l' equazione di Eulero generalizzata (2.18) divengono:

$$Ax = b \tag{3.26}$$

$$G_\lambda : \mathbb{R}^n \to \mathbb{R}^n$$

$$x \longrightarrow G_\lambda(x) = \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2 \tag{3.27}$$

$$(A^T A + \lambda^2 L^T L)x = A^T b. \tag{3.28}$$

Poichè lo spazio $M_{s,1}(\mathbb{R})$, isomorfo a $\mathbb{R}^s$, è uno spazio di Hilbert per ogni $s \geq 0$ e le ipotesi (2.23) (2.24) sono banalmente verificate nel caso discreto, come conseguenza della finitezza dello spazio, l' unica ipotesi che faremo nel seguito sarà $N(A) \cap N(L) = \{0\}$ [1] garantendo così la validità del teorema (2.2.7) a cui ci riferiremo spesso nel seguito.
Ricaviamo ora la soluzione di (3.28). A partire da (3.28) usando (3.21) otteniamo:
$A^T A + \lambda^2 L^T L = X^{-T}(\Sigma^T \Sigma + \lambda^2 M^T M)X^{-1} = X^{-T} \begin{pmatrix} \Sigma_p^2 + \lambda^2 M_p^2 & 0 \\ 0 & I_{n-p} \end{pmatrix} X^{-1}$.
Pertanto:
$(A^T A + \lambda^2 L^T L)x = A^T b \iff X^{-T} \begin{pmatrix} \Sigma_p^2 + \lambda^2 M_p^2 & 0 \\ 0 & I_{n-p} \end{pmatrix} X^{-1} x = X^{-T} \Sigma U^T b \iff$

[1] L' ipotesi

$$N(A) \cap N(L) = \{0\}$$

garantisce che $rank(\begin{pmatrix} A \\ L \end{pmatrix}) = n$. Infatti :

$$rank(\begin{pmatrix} A \\ L \end{pmatrix}) = n \iff N(\begin{pmatrix} A \\ L \end{pmatrix}) = \{0\} \iff N(A) \cap N(L) = \{0\}.$$

$$\begin{pmatrix} \Sigma_p^2 + \lambda^2 M_p^2 & 0 \\ 0 & I_{n-p} \end{pmatrix} y = \Sigma U^T b \text{ con } y = X^{-1}x \Longleftrightarrow$$

$$y_i = \begin{cases} 0 & \forall i = 1, ..., p \text{ e } \sigma_i = 0 \\ \frac{\sigma_i}{\sigma_i^2 + \lambda^2 \mu_i^2} u_i^T b = \frac{\gamma_i^2}{\gamma_i^2 + \lambda^2} u_i^T b & \forall i = 1, 2, ..., p \text{ e } \sigma_i \neq 0 \\ u_i^T b & \forall i = p+1, ..., n \end{cases}$$

ovvero

$$x_\lambda = \sum_{i=1}^{p} \frac{\gamma_i^2}{\gamma_i^2 + \lambda^2} \frac{u_i^T b}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T b) x_i \tag{3.29}$$

dove x_i rappresenta la i^a componente di X.
In particolare i fattori di filtro sono

$$f_i = \frac{\gamma_i^2}{\gamma_i^2 + \lambda^2} \quad \forall i = 1, 2, ..., p \tag{3.30}$$

Osservazione 3.3.2 *Osserviamo esplicitamente che i fattori f_i sono effettivamente di filtro: infatti*

- $|f_i| = |\frac{\gamma_i^2}{\gamma_i^2 + \lambda^2}| \leq 1$
- $lim_{\lambda \to 0} f_i = 1$ *se* $\sigma_i \neq 0$
- $\frac{f_i}{\sigma_i}$ *sono limitati.*

Anche nel caso discreto vale la relazione

$$x_\lambda = R_\lambda b \tag{3.31}$$

con

$$R_\lambda = (A^T A + \lambda^2 L^T L)^{-1} A^T \tag{3.32}$$

dove, attraverso la $GSVD$ di (A, L), l' algoritmo regolarizzante (3.31) è dato da

$$R_\lambda = XF\Sigma^\dagger U^T = X \begin{pmatrix} F_p & 0 \\ 0 & I_{n-p} \end{pmatrix} \Sigma^\dagger U^T = X_p F_p \Sigma_p^\dagger U_p^T + X_0 U_0^T \tag{3.33}$$

con $o = (n-p)$, $F_p = diag(f_1, ..., f_p)$, $\Sigma_p^\dagger = diag({\sigma_1}^\dagger, ..., {\sigma_p}^\dagger)$ dove $\forall i = 1, 2, ..., p$

$${\sigma_i}^\dagger = \begin{cases} 0 & \text{se } \sigma_i = 0 \\ \frac{1}{\sigma_i} & \text{se } \sigma_i \neq 0. \end{cases}$$

e $U = (\, U_p, U_0\,)$. Dalle relazioni (3.31) e (3.33) segue che la soluzione x_λ può essere decomposta nel modo seguente

$$x_\lambda = x_\lambda^{(1)} + x_\lambda^{(2)}, \quad x_\lambda^{(1)} = X_p F_p \Sigma_p^\dagger U_p^T b, \quad x_\lambda^{(2)} = X_0 U_0^T b. \tag{3.34}$$

dove l' ultima componente, $x_\lambda^{(2)}$, appartiene a $N(L) = span\{x_{p+1}, ..., x_n\}$ ed è nulla se $p = n$. Consideriamo adesso la componente $x_\lambda^{(1)}$. L' effetto della regolarizzazione corrisponde a smorzare i termini $(u_i^T b)/\sigma_i$ in $x_\lambda^{(1)}$ corrispondenti ai σ_i piccoli, attraverso la matrice di filtro F. Il motivo per cui la soluzione x_λ è più regolare rispetto a (3.12) è dovuta al fatto che il numero di oscillazioni (o cambiamenti di segno) nei vettori generalizzati x_i aumenta al diminuire di σ_i, così che la matrice F filtra i contributi in $x_\lambda^{(1)}$ con oscillazioni di maggior rilievo.
Se la $GSVD$ è calcolata attraverso l' algoritmo di Van Loan, dalle identità

$$\begin{cases} A = U\Sigma X^T = U\Sigma (X^{-T})^{-1} \\ L = VMX^T = VM(X^{-T})^{-1} \end{cases}$$

si deduce che per potere applicare le relazioni (3.21) e (3.29) è necessario invertire la matrice X. Dobbiamo pertanto determinare sotto quali condizioni è possibile calcolare $(X^T)^{-1}$ in modo numericamente stabile: a tale scopo dimostreremo [11], di seguito, che se L è ben condizionata anche X è tale.

Teorema 3.3.3 *Posto* $Z = \begin{pmatrix} A \\ L \end{pmatrix}$ *si ha:*

$$\|X^{-1}\|_2 = \|Z\|_2 \leq \|A\|_2 + \|L\|_2, \quad \|X\|_2 \leq \|Z^\dagger\|_2 \leq \Pi_p^{-1} \tag{3.35}$$

con Π_p *definita da*

$$\Pi_p = \begin{cases} min\{\|L^\dagger\|_2^{-1}, \sigma_{min}(AP_{N(L)})\} & se\ p \quad n \\ \|L^\dagger\|_2^{-1} & se\ p{=}n \end{cases} \tag{3.36}$$

con $\sigma_{min}(AP_{N(L)})$ *il più piccolo valore singolare di* $AP_{N(L)}$ *non nullo.*

Dimostrazione. Dalla relazione

$$\begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix} \begin{pmatrix} A \\ L \end{pmatrix} = \begin{pmatrix} \Sigma \\ M \end{pmatrix} X^{-1}$$

otteniamo che lo spettro di X^{-1} è uguale a quello di Z e quindi:
$\|X^{-1}\|_2 = \sigma_{max}(X^{-1}) = \sigma_{max}(Z) = \|Z\|_2 = \| \begin{pmatrix} A \\ L \end{pmatrix} \|_2 \leq \| \begin{pmatrix} A \\ 0 \end{pmatrix} \|_2 + \| \begin{pmatrix} 0 \\ L \end{pmatrix} \|_2 \leq$

$\|A\|_2 + \|L\|_2$.
Essendo $rank(L) = p$ abbiamo

$$\sigma_i(Z) = \sigma_i(\begin{pmatrix} A \\ L \end{pmatrix}) \geq \sigma_i(\begin{pmatrix} 0 \\ L \end{pmatrix}) = \begin{cases} \sigma_i(L) & \text{se } i = 1, 2, ..., p \\ 0 & \text{se } i = p+1, ..., n \end{cases}.$$

Per ottenere una minorazione significativa per $i = p+1, ..., n$ consideriamo la matrice $Z = \begin{pmatrix} A \\ L \end{pmatrix}$ come la matrice ottenuta da $\begin{pmatrix} 0 \\ L \end{pmatrix}$ mediante la perturbazione $\begin{pmatrix} A \\ 0 \end{pmatrix}$: si deduce allora che $\forall i = 1, 2, ..., p$

$$\sigma_i(Z) = \sigma_i(\begin{pmatrix} A \\ L \end{pmatrix}) = \sigma_i(\begin{pmatrix} A \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ L \end{pmatrix}) \geq \sigma_{min}(\begin{pmatrix} A \\ 0 \end{pmatrix} P_{N\begin{pmatrix} 0 \\ L \end{pmatrix}}) = \sigma_{min}(AP_{N(L)})$$

Pertanto
-se $n = p$, essendo $rank(L) = p = n$ e quindi $N(L) = \{0\}$ $AP_{N(L)} = 0$, otteniamo

$$\sigma_i(Z) \geq \sigma_i(L) \quad \forall i = 1, 2, ..., n$$

ovvero

$$\|Z^\dagger\|_2 = \frac{1}{\sigma_{min}(Z)} = \frac{1}{\sigma_n(Z)} \leq \frac{1}{\sigma_n(L)} = \|L^\dagger\|_2.$$

-se $p < n$, $N(L) \neq \{0\}$ e quindi

$$\|Z^\dagger\|_2 \leq \|L^\dagger\|_2,$$

$$\|Z^\dagger\|_2 = \frac{1}{\sigma_{min}(Z)} = \frac{1}{\sigma_n(Z)} \leq \frac{1}{\sigma_{min}(AP_{N(L)})}$$

ovvero

$$\|Z^\dagger\|_2 \leq \min\{\|L^\dagger\|_2, \frac{1}{\sigma_{min}(AP_{N(L)})}\}.$$

■

Per poter utilizzare la seconda maggiorazione (3.35) è necessario conoscere lo spettro dell' operatore $AP_{N(L)}$. Poichè questo non è possibile utilizziamo l' approssimazione, indicata in [10],

$$AP_{N(L)} \approx \sum_{i=1}^{n-p} \mu_i(A)\sigma_i(A)v_i^T(A) \tag{3.37}$$

dove l' insieme $\{\sigma_i(A), u_i(A), v_i(A)\}_i$ è dedotto dalla SVD di A (cfr.(3.9)). Poichè $d = n - p$ è di solito un intero piccolo, in quanto rappresenta l' ordine della derivata discretizzata da L (tipicamente 1 o 2), otteniamo

$$\sigma_{min}(AP_{N(L)}) \approx \sigma_{n-p}(A) \approx \sigma_1(A) = \|A\|_2.$$

e supposto $\|A\|_2 \approx \|L\|_2$, si ha che

$$\sigma_p(L) = \|L^\dagger\|_2^{-1} \leq \|A\|_2 \approx \sigma_{min}(AP_{N(L)})$$

e, per il teorema 3.3.3, $\Pi_p^{-1} \approx \|L^\dagger\|_2$.
Queste relazioni sostituite in (3.35) permettono di ottenere

$$\|X\|_2 \leq \Pi_p^{-1} \approx \|L^\dagger\|_2 \tag{3.38}$$

come volevamo dimostrare.
Supposta valida la relazione $\|A\|_2 \approx \|L\|_2$ da (3.35) segue che
$K_2(X) = \|X^{-1}\|_2\|X\|_2 \approx \|L^\dagger\|_2\|X\|_2^{-1} \leq \|L^\dagger\|_2(\|A\|_2+\|L\|_2) \approx 2\|L\|_2\|L^\dagger\|_2 = 2K_2(L)$.
Osserviamo infine che il buon condizionamento di L garantisce il buon condizionamento di $\begin{pmatrix} A \\ L \end{pmatrix}$. Infatti
$K_2(\begin{pmatrix} A \\ L \end{pmatrix}) = K_2(\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} \Sigma \\ M \end{pmatrix} X^{-1}) \leq K_2(\begin{pmatrix} \Sigma \\ M \end{pmatrix} X^{-1}) =$
$K_2(X^{-1}) = K_2(X) \approx 2K_2(L)$.

Osservazione 3.3.4 *Vogliamo fare alcune osservazioni relative alla scelta della matrice di regolarizzazione L. Benchè molti tipi di informazioni addizzionali relative alla soluzione x di (3.26) possano essere utilizzate l' approccio che noi adotteremo è quello di scegliere come L una matrice $p \times n$ che rappresenti l' approssimazione discreta di un operatore di derivata $(n - p)$: in tal caso L è una matrice a banda con rango massimo per righe. Definiamo "funzione di penalità del metodo di regolarizzazione di Tikhonov l' operatore $\Omega : \mathbb{R}^n \to \mathbb{R}$ definito come*

$$\Omega(x) = \|Lx\|_2. \tag{3.39}$$

In alcuni casi può risultare più appropriato scegliere come Ω la norma di Sobolev discreta in $\mathbb{R}^n$

$$\Omega^2(x) = \alpha_0^2\|x\|_2^2 + \sum_{i=1}^{q} \alpha_i^2\|L_i x\|_2^2 \tag{3.40}$$

dove L_i rappresenta l' approssimazione di un operatore di derivata i-esimo. Notiamo che Ω può essere sempre scritto nella forma (3.39) attraverso una opportuna fattorizzazione di Cholesky. Più precisamente detta

$$D = \alpha_0^2 I + \sum_{i=1}^{q} \alpha_i^2 L_i^T L_i \tag{3.41}$$

si ha D definita positiva e simmetrica: quindi applicando la fattorizzazione di Cholesky a D otteniamo che esiste una matrice G triangolare superiore con elementi diagonali positivi tale che $D = G^T G$: ovvero
$\Omega(x) = [\alpha_0^2 \|x\|_2^2 + \sum_{i=1}^{q} \alpha_i^2 \|Lx_i\|_2^2]^{1/2} = [\alpha_0^2 x^T x + \sum_{i=1}^{q} \alpha_i^2 (L_i x)^T (L_i x)]^{1/2} = [x^T(\alpha_0^2 + \sum_{i=1}^{q} \alpha_i^2 L_i^T L_i)x]^{1/2} = [x^T D x]^{1/2} = [x^T G^T G x]^{1/2} = \|Gx\|_2$.
Si noti come la nuova matrice G della funzione di penalità sia ancora strutturata: naturalmente la scelta della norma di Sobolev impone un aumento del costo computazionale del metodo di regolarizzazione dovuto alla fattorizzazione di Cholesky di D che permette di affrontare il problema utilizzando ancora la $GSVD$ di (A, G).

Esempio

In questo esempio vogliamo illustrare le differenze tra regolarizzazione in forma generale e standard mostrando che la prima può essere necessaria per assicurare il calcolo di una soluzione soddisfacente.
L' equazione che si vuole risolvere è ottenuta dalla discretizzazione dell' operatore di inversione della Trasformata di Laplace:

$$(Kf)(x) = \int_0^{+\infty} e^{-xt} f(t) dt \quad \forall x \in \mathbb{R} \tag{3.42}$$

con soluzione esatta $f(t) = 1 - exp(-t/2)$. Si osservi che (3.42) è un operatore di Fredholm di prima specie con nucleo $K(x,t) = e^{-xt}$ di classe $C^\infty(\mathbb{R}^2)$. Tale discretizzazione è ottenuta applicando le formule di quadratura su un reticolo di 16 punti con pesi tra loro uguali.
Come si vede dalla figura 3.3 la soluzione soddisfa la relazione $f(t) \to 1$ per $t \to +\infty$, e la parte asintotica nella soluzione discretizzata si evidenzia per $t > t_i \quad i \in \{8, ..., 16\}$. Usiamo come matrice di regolarizzazione la discretizzazione dell' operatore di derivata prima ($L = bidiag(-1, 1)$). Dalla figura 3.1 notiamo che nessuno dei vettori v_i della SVD di A include la parte asintotica menzionata sopra; quindi, questi vettori non sono adatti come base della soluzione regolarizzata (3.19) ottenuta dal metodo di Tikhonov standard.

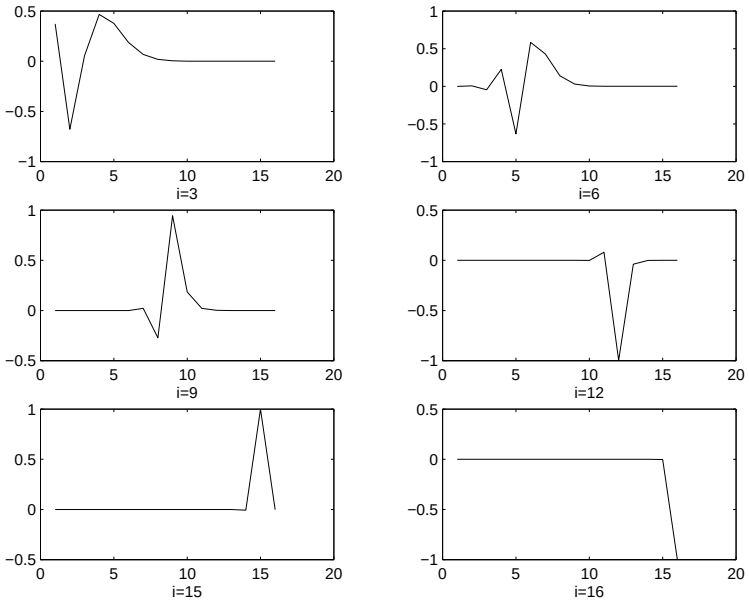

Figura 3.1. Grafico dei vettori singolari v_i che descrivono x_λ.

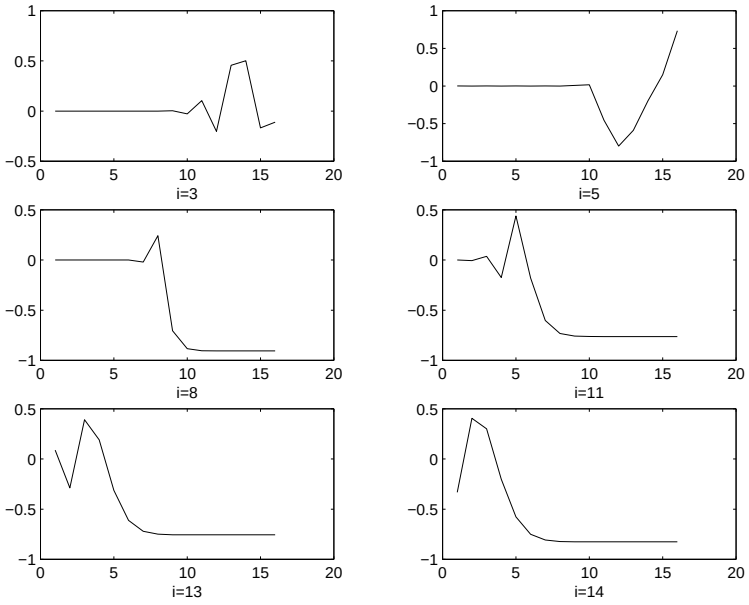

Figura 3.2. Grafico dei vettori generalizzati destri di A per alcuni i.

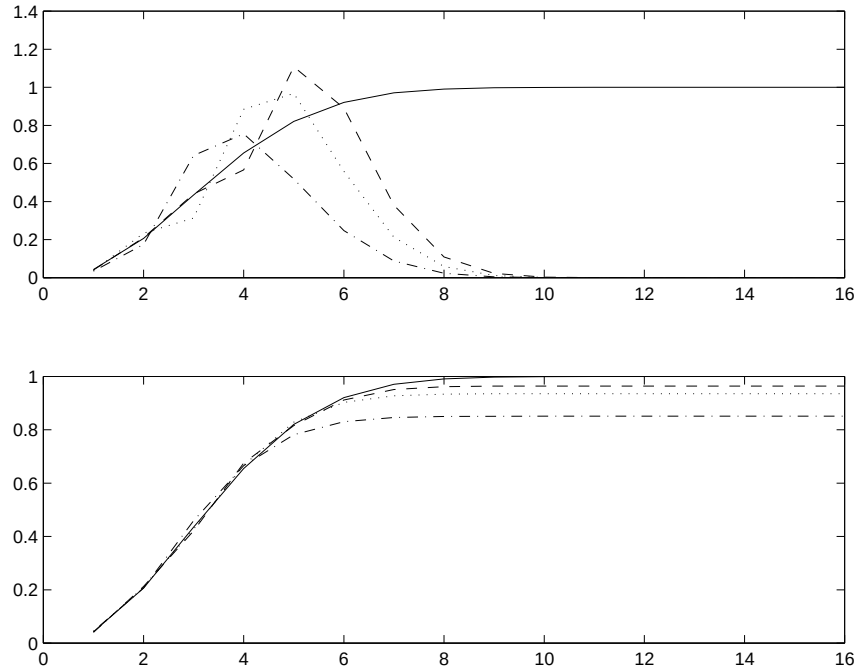

Figura 3.3. Grafico delle soluzioni standard e generale per alcuni valori di λ: $-f, --sol.\lambda = 0.001, {}^{-\cdot}sol.\lambda = 0.1, {}^{\cdot\cdot}sol.\lambda = 0.01$.

Al contrario (figura 3.2) i vettori x_i relativi alla $GSVD$ di (A, L) non sono definitivamente nulli e quindi sono più adatti a rappresentare la base rispetto alla quale sviluppare la soluzione. In figura (3.3) è presente il grafico di alcune soluzioni regolarizzate, ottenute utilizzando il metodo di regolarizzazione di Tikhonov standard e generale, corrispondenti a diversi valori del parametro λ. Per entrambi i metodi sono stati scelti gli stessi valori di λ.
Riportiamo infine alcuni valori che mostrano gli effetti della regolarizzazione generale sulla ricostruzione del dato e della soluzione:
per $\lambda = 0.1$ si ha $\|Ax_\lambda - b\|_2 = 7.7789\,10^{-3}$ e $\|f - x_\lambda\|_\infty = 1.4897\,10^{-1}$,
per $\lambda = 0.01$ si ha $\|Ax_\lambda - b\|_2 = 3.5477\,10^{-4}$ e $\|f - x_\lambda\|_\infty = 6.4514\,10^{-2}$,
per $\lambda = 0.001$ si ha $\|Ax_\lambda - b\|_2 = 2.4674\,10^{-4}$ e $\|f - x_\lambda\|_\infty = 3.5735\,10^{-2}$.
Tali risultati sono soddisfacenti soprattutto se si tiene conto del mal condizionamento di A per cui vale $K_2(A) = 1.1780\ 10^{30}$.

3.4 La soluzione ai minimi quadrati attraverso la $GSVD$

In questa sezione vogliamo rispondere alla seguente domanda: è possibile dare un' espressione di $x^\dagger$ usando la $GSVD$ di (A, L). La domanda potrebbe sembrare poco significativa in quanto abbiamo già espresso esplicitamente $x^\dagger$ usando la SVD di A e quindi non appare chiaro il ricorso alla $GSVD$. In realtà tale espressione ci permetterà
-di evidenziare l' influenza della regolarizzazione e del *noise* su x_λ,
-di introdurre, nel prossimo capitolo, il concetto di $TGSVD$.
Poichè la matrice X è invertibile l' insieme $\{x_j\}_{j=1}^n$ delle sue colonne costituisce una base di $\mathbb{R}^n$ e quindi la soluzione ai minimi quadrati di (3.26), che indicheremo con x_{LS}, può essere scritta come

$$x_{LS} = \sum_{j=1}^{n} c_j x_j \tag{3.43}$$

con $\{c_j\}_{j=1}^n$ da determinare. Estendiamo inoltre la famiglia di vettori ortonormali $\{u_j\}_{j=1}^n$ a base ortonormale di $\mathbb{R}^m$ $(m \geq n)$ aggiungendo (se $m > n$) i vettori $\{u_j\}_{j=n+1}^m$. Allora usando (3.21)

$$Ax_{LS} = \sum_{j=1}^{n} c_j Ax_j = \sum_{j=1}^{p} c_j Ax_j + \sum_{j=p+1}^{n} c_j Ax_j = \sum_{j=1}^{p} c_j \sigma_j u_j + \sum_{j=p+1}^{n} c_j u_j$$

e

$$b = \sum_{j=1}^{m} (u_j^T b) u_j$$

da cui otteniamo:

$$Ax_{LS} - b = \sum_{j=1}^{p} (c_j \sigma_j - u_j^T b) u_j + \sum_{j=p+1}^{n} (c_j - u_j^T b) u_j - \sum_{j=n+1}^{m} (u_j^T b) u_j.$$

Da questa si ha

$$K(c_1, ..., c_n) = \|Ax - b\|_2^2 = \sum_{j=1}^{p} (c_j \sigma_j - u_j^T b)^2 + \sum_{j=p+1}^{n} (c_j - u_j^T b)^2 + \sum_{j=n+1}^{m} (u_j^T b)^2$$

e quindi i punti di minimo per $\|Ax - b\|_2^2$ si deducono studiando i valori stazionari di K.
Passando alle derivate parziali otteniamo

$$\frac{\partial K}{\partial c_j} = 2(c_j \sigma_j - u_j^T b)\sigma_j \quad \forall j = 1, 2, ..., p \quad \sigma_j \neq 0$$

$$\frac{\partial K}{\partial c_j} \equiv 0 \qquad \forall j = 1, 2, ..., p \;\; \sigma_j = 0$$

e

$$\frac{\partial K}{\partial c_j} = 2(c_j - u_j^T b) \;\; \forall i = p+1, ..., n.$$

Posto $I = \{j \in \{1, 2, ..., p\} : \sigma_j \neq 0\}$ e $I^C = C_{\{1,2,...,p\}} I$ si ha che

$$x_{LS} = \sum_{j \in I} \frac{u_j^T b}{\sigma_j} x_j + \sum_{j \in I^C} c_j x_j + \sum_{j=p+1}^{n} (u_j^T b) x_j \tag{3.44}$$

,con c_j arbitrari, rappresenta la totalità delle soluzioni ai minimi quadrati di (3.26). Naturalmente x_{LS} è unica (ovvero $x_{LS} = x^\dagger$) $\iff I = \{j \in \{1, 2, ...p\} : \sigma_j = 0\} = \emptyset \iff rank(A) = n \iff N(A) = \{0\}$ in perfetto accordo con il teorema (2.1.4) applicato al caso discreto. Otteniamo allora che

$$\|x_{LS}\|_2 = \| \sum_{j \in I} \text{ e } \frac{u_j^T b}{\sigma_j} x_j + \sum_{j \in I^C} c_j x_j + \sum_{j=p+1}^{n} (u_j^T b) x_j \|_2 \tag{3.45}$$

va minimizzata rispetto alle variabili c_j con $j \in I^C$ al fine di determinare $x^\dagger$. Ricordando che

$$x_{LS} = x^\dagger + u \;\; \text{ e } \;\; u \in N(A) = span\{x_i : i \in I^C\}$$

otteniamo da (3.44) la relazione

$$x^\dagger = \sum_{j \in I} \frac{u_j^T b}{\sigma_j} x_j + \sum_{j=p+1}^{n} (u_j^T b) x_j. \tag{3.46}$$

Tale soluzione poteva essere ottenuta da (3.29) ponendo semplicemente $\lambda = 0$: in realtà avremmo dimostrato soltanto che (3.46) è soluzione ai minimi quadrati di $Ax = b$ ma non che è la soluzione generalizzata.
Si osservi che in (3.46) non intervengono i vettori $\{u_j\}_{j=n+1}^m$ utilizzati solo nella dimostrazione e che in questo caso i filtri sono dati da $f_i = 1 \;\; \forall i = 1, ..., p$ con $\sigma_i \neq 0$. Si noti infine l' analogia tra (3.46) e (3.12).
La relazione precedente può anche essere scritta come

$$x^\dagger = \sum_{j=n-rank(A)+1}^{p} \frac{u_j^T b}{\sigma_j} x_j + \sum_{j=p+1}^{n} (u_j^T b) x_j.$$

Osservazione 3.4.1 *Poichè la soluzione generalizzata $x^\dagger = A^\dagger b$ è unica, da (3.46), otteniamo*

$$A^\dagger = X\Sigma^\dagger U^T \tag{3.47}$$

avendo indicato con $\Sigma^\dagger$ la matrice $n \times m$ definita da

$$\Sigma^\dagger = \begin{pmatrix} \Sigma_p^\dagger & 0 \\ 0 & I_{n-p} \end{pmatrix}.$$

Tale relazione poteva essere anche dedotta direttamente verificando che la matrice $X\Sigma^\dagger U^T$ soddisfa le proprietà di Moore-Penrose.

Il fatto che la famiglia $(R_\lambda)_\lambda$, con R_λ definito da (3.32), costituisca una famiglia di operatori di regolarizzazione per $Ax = b$ si può dedurre in modo diretto a partire dalle proprietà della $GSVD$ e della relazione appena trovata. Infatti:
(1) $(A^TA + \lambda^2 L^TL)$ è un operatore lineare, continuo e invertibile avendo supposto $N(A) \cap N(L) = \{0\}$. Inoltre si hanno le relazioni:

$$(A^TA + \lambda^2 L^TL) = X^{-T} diag(\sigma_1^2 + \lambda^2\mu_1^2, ..., \sigma_p^2 + \lambda^2\mu_p^2, 1, ..., 1)X^{-1} \tag{3.48}$$

e

$$(A^TA + \lambda^2 L^TL)^{-1} = X diag(\frac{1}{\sigma_1^2 + \lambda^2\mu_1^2}, ..., \frac{1}{\sigma_p^2 + \lambda^2\mu_p^2}, 1, ..., 1)X^T \tag{3.49}$$

(2) $\lim_{\lambda\to 0} R_\lambda b = A^\dagger b$ infatti per (3.49) si ha che
$\lim_{\lambda\to 0} R_\lambda b = \lim_{\lambda\to 0}(A^TA + \lambda^2 L^TL)^{-1}X^TA^Tb =$
$\lim_{\lambda\to 0} X(\Sigma^T\Sigma + \lambda^2 M^TM)^{-1}X^TA^Tb = X\Sigma^\dagger U^Tb = A^\dagger b.$
Da questo si deduce, che nel caso in cui il dato non sia perturbato, la soluzione regolarizzata $x_\lambda = R_\lambda b$ fornisce, per $\lambda \to 0$, una approssimazione della soluzione generalizzata.
Ci chiediamo ora cosa succede nel caso in cui si abbia un dato perturbato.

3.5 Perturbazioni sul dato

Nelle applicazioni il termine noto b di (3.26) è sempre contaminato da errori di diverso tipo: ad esempio errori di misurazione, errori di approssimazione o arrotondamento. In pratica non abbiamo a disposizione il dato esatto ma una

sua approssimazione e, talvolta, una stima dell' errore massimo con cui viene fornito. Quindi possiamo scrivere b come

$$b = \bar{b} + e \tag{3.50}$$

dove e è il vettore degli errori (*noise*) e $\bar{b}$ è il dato non perturbato.
Con queste premesse sia

$$\bar{x}^\dagger = \sum_{i \in I} \frac{u_i^T \bar{b}}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T \bar{b}) x_i \tag{3.51}$$

la soluzione generalizzata di

$$Ax = \bar{b} \tag{3.52}$$

e studiamo l' errore di approssimazione $(x_\lambda - \bar{x}^\dagger)$ relativo alla soluzione regolarizzata (ovvero l' errore commesso nella ricostruzione di $A^\dagger \bar{b}$). Dalle relazioni (3.17) e (3.51) otteniamo

$$x_\lambda - \bar{x}^\dagger = (R_\lambda b - A^\dagger \bar{b}) = [\ \sum_{i=1}^{p} f_i \frac{u_i^T e}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T e) x_i\ ] + \sum_{i=1}^{p} (f_i - 1) \frac{u_i^T \bar{b}}{\sigma_i} x_i = e_{Pert} + e_{Reg} \tag{3.53}$$

con x_λ soluzione regolarizzata generale del problema perturbato $Ax = b$. In (3.53), il termine in parentesi rappresenta l' *errore di perturbazione* indotto dall' errore di cui è affetto il dato $\bar{b}$ mentre il secondo addendo è un *errore di regolarizzazione* dovuto alla regolarizzazione della componente non perturbata $\bar{b}$ di b. Pertanto applicando un metodo di regolarizzazione con λ piccolo, essendo la maggior parte dei fattori di filtro f_i prossimi ad uno, l' errore $(x_\lambda - \bar{x}^\dagger)$ è dominato, in base alla precedente relazione, dall' errore di perturbazione

$$e_{Pert} = \sum_{i=1}^{p} f_i \frac{u_i^T e}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T e) x_i. \tag{3.54}$$

Al contrario aumentando la capacità di regolarizzazione del metodo (λ grande) la maggior parte dei fattori di filtro soddisfa la relazione $f_i << 1$ e $(x_\lambda - \bar{x}^\dagger)$ è dominato, a meno del fattore costante $\sum_{i=p+1}^{n} (u_i^T e) x_i$, dall' errore di regolarizzazione

$$e_{Reg} = \sum_{i=1}^{p} (f_i - 1) \frac{u_i^T \bar{b}}{\sigma_i} x_i. \tag{3.55}$$

Si osservi che mentre l' errore di perturbazione dipende dal parametro di regolarizzazione oltre che dal noise e l' errore di regolarizzazione dipende unicamente dal parametro λ.

Teorema 3.5.1 *Sia e la perturbazione del dato ed x_λ la corrispondente soluzione regolarizzata. Allora [11] la perturbazione relativa a x_λ è data da*

$$\frac{\|\bar{x}_\lambda - x_\lambda\|_2}{\|\bar{x}_\lambda\|_2} \leq \phi_\lambda \|A\|_2 \|X\|_2 \frac{\|e\|_2}{\|\bar{b}_\lambda\|_2} \tag{3.56}$$

dove $\bar{b}_\lambda = A\bar{x}_\lambda$ e

$$\phi_\lambda = \begin{cases} 1 & se\ \lambda > 1/\sqrt{2} \\ \frac{1}{2\lambda}(1-\lambda^2)^{-1/2} & se\ \lambda \leq 1/\sqrt{2} \end{cases} .$$

Dimostrazione.
$x_\lambda - \bar{x}_\lambda = (R_\lambda b - R_\lambda \bar{b}) = R_\lambda e \Longrightarrow \|\bar{x}_\lambda - x_\lambda\|_2 = \|R_\lambda e\|_2 = \|X \begin{pmatrix} F_p & 0 \\ 0 & I_{n-p} \end{pmatrix} \Sigma^\dagger U^T e\|_2 \leq$
$\|X \begin{pmatrix} F_p & 0 \\ 0 & I_{n-p} \end{pmatrix} \Sigma^\dagger\|_2 \|U^T e\|_2 \leq \|X\|_2 \| \begin{pmatrix} F_p & 0 \\ 0 & I_{n-p} \end{pmatrix} \Sigma^\dagger\|_2 \|e\|_2 \leq$
$\|X\|_2 \max_{i=1,2,\ldots,p}\{\frac{\gamma_i^2}{\gamma_i^2+\lambda^2}\frac{1}{\sigma_i}, 1\}\|e\|_2$. Poichè

$$\frac{{\gamma_i}^2}{\gamma_i^2+\lambda^2}\frac{1}{\sigma_i} = \frac{\gamma_i}{\gamma_i^2+\lambda^2}\frac{1}{(\gamma_i+1)^{1/2}} \leq \phi_\lambda$$

otteniamo

$$\|\bar{x}_\lambda - x_\lambda\|_2 \leq \phi_\lambda \|X\|_2 \|e\|_2.$$

Essendo

$$\|\bar{b}_\lambda\|_2 = \|A\bar{x}_\lambda\|_2 \leq \|A\|_2 \|\bar{x}_\lambda\|_2$$

si ha

$$\|\bar{x}_\lambda\|_2 \geq \frac{\|\bar{b}_\lambda\|_2}{\|A\|_2}$$

e da questa segue la relazione cercata. ∎

Definizione 3.5.2 *Il numero di condizionamento associato alla regolarizzazione generale è definito da*

$$K_\lambda = \limsup_{\|e\|_2 \to 0} \frac{\|\bar{x}_\lambda - x_\lambda\|_2}{\|x_\lambda\|_2} \tag{3.57}$$

e soddisfa la relazione [11]

$$K_\lambda \leq \phi_\lambda K_2(X)$$

con $K_2(X)$ numero di condizionamento di X.

La relazione precedente permette di evidenziare che il buon condizionamento di X garantisce la stabilità di (3.29) e del metodo di regolarizzazione rispetto a perturbazioni sul dato. Applicando la disuguaglianza triangolare a (3.53) otteniamo:

$$\|x_\lambda - \bar{x}^\dagger\|_2 = \|R_\lambda b - A^\dagger \bar{b}\|_2 \leq \|e_{Reg}\|_2 + \|e_{Pert}\|_2 \tag{3.58}$$

cioè la norma dell' errore di ricostruzione è limitata dalla somma della norma dell' errore di regolarizzazione e della norma dell' errore di perturbazione.
Attraverso la $GSVD$ di (A, L) possiamo dare una rappresentazione esplicita di tali errori mostrando come variano in funzione di λ e/o e.

-Norma dell' errore di perturbazione

$$P(\lambda) = \|e_{Pert}\|_2 = \|R_\lambda e\|_2 = \| \sum_{i=1,\sigma_i \neq 0}^{p} \frac{{\gamma_i}^2}{{\gamma_i}^2 + \lambda^2} \frac{u_i^T e}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T e) x_i \|_2 \tag{3.59}$$

$-P(0) = \| \sum_{i=1,\sigma_i \neq 0}^{p} \frac{u_i^T e}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T e) x_i \|_2$
$-\lim_{\lambda \to +\infty} P(\lambda) = \| \sum_{i=p+1}^{n} (u_i^T e) x_i \|_2$.
Poichè i fattori di filtro f_i sono decrescenti rispetto a λ la funzione $P(\lambda)$ è decrescente.

-Norma dell' errore di regolarizzazione

$$R(\lambda) = \|e_{Reg}\|_2 = \| \sum_{i=1}^{p} (f_i - 1) \frac{u_i^T \bar{b}}{\sigma_i} x_i \|_2 = \| \sum_{i=1,\sigma_i \neq 0}^{p} \frac{\lambda^2}{\lambda^2 + {\gamma_i}^2} \frac{u_i^T \bar{b}}{\sigma_i} x_i \|_2 \tag{3.60}$$

$-R(0) = 0$
$-\lim_{\lambda \to +\infty} = \| \sum_{i=1,\sigma_i \neq 0}^{p} \frac{u_i^T \bar{b}}{\sigma_i} x_i \|_2$.
Poichè i coefficienti $c_i(\lambda) = \frac{\lambda^2}{\lambda^2 + {\gamma_i}^2}$ sono crescenti rispetto a λ $R(\lambda)$ è crescente.

Troviamo pertanto che la norma dell' errore di regolarizzazione e di perturbazione hanno comportamento opposto rispetto a λ. Se desideriamo ridurre l' errore di regolarizzazione dobbiamo scegliere un piccolo valore del parametro λ e, in tal caso, l' errore di perturbazione aumenta. Dall' altro lato se vogliamo ridurre l' errore di perturbazione dobbiamo scegliere un valore di λ sufficientemente grande a cui corrisponde un errore di regolarizzazione maggiore.

Ritornando allora alla valutazione dell' errore globale si ha

$$e_{Globale} = e_{Reg} + e_{Pert} \tag{3.61}$$

e per questo si ottiene un andamento che può essere rappresentato come in figura 3.4 . Poichè l' errore globale è ottenuto come somma di una funzione

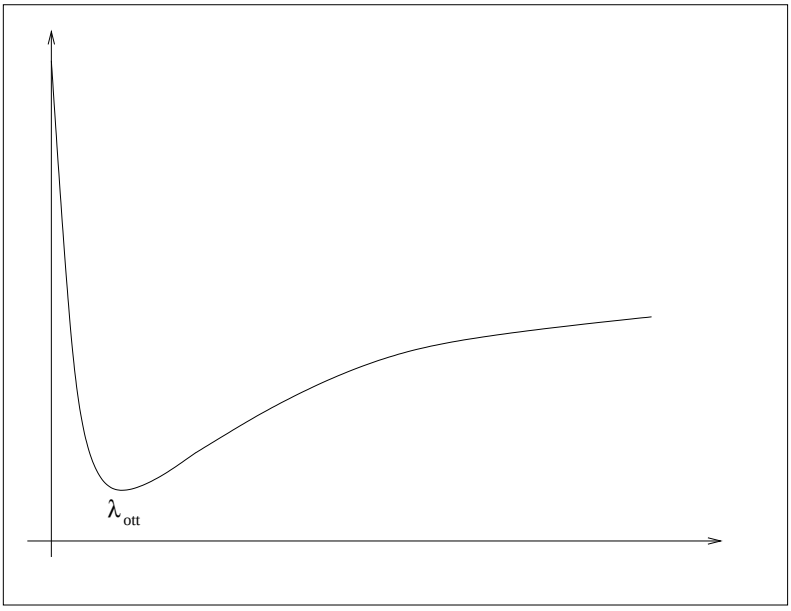

Figura 3.4. Andamento dell' errore globale.

crescente e di una decrescente esiste un parametro ottimo λ_{ott} in corrispondenza del quale l' operatore R_λ fornisce la migliore approssimazione della soluzione $\bar{x}^\dagger$ e tale parametro rappresenta il migliore compromesso tra regolarizzazione e propagazione dell' errore di perturbazione.
Questo risultato comporta che per valori di λ decrescenti le soluzioni regolarizzate x_λ inizialmente approssima la soluzione x del problema non perturbato per poi allontanarsi da questa a causa dell' influenza sempre maggiore del *noise* e. Possiamo pertanto chiamare [2], in analogia con il caso standard, questa proprietà *semiconvergenza* in modo tale da poter parlare di *convergenza* nel caso del problema non perturbato. La semiconvergenza è una proprietà addizionale che può essere richiesta per un algoritmo regolarizzante. Osserviamo esplicitamente che il valore λ_{ott} non può essere determinato nel caso di problemi concreti poichè non è noto il dato $\bar{b}$. Nonostante questo l' uso di tale

parametro è fondamentale per lo studio delle caratteristiche dei problemi inversi ottenute attraverso simulazioni numeriche in cui $\bar{b}$ è noto.
Siamo riusciti a verificare che nel caso generale ($L \neq I$) continuano a essere verificate le proprietà valide nel caso standard.

3.6 Metodi di scelta del parametro di regolarizzazione

Trattiamo ora i metodi di scelta del parametro di regolarizzazione.
Questi possono essere divisi in due famiglie a seconda che utilizzino o meno le informazioni disponibili sul noise di cui è affetto il dato $\bar{b}$. Le due classi sono caratterizzate nel seguente modo:
(1) metodi che sfruttano la conoscenza di $\|e\|_2$ o, più in generale, stime di $\|e\|_2$.
(2) metodi che non richiedono informazioni a priori relative al *noise* ma cercano di estrapolare le necessarie informazioni dal dato perturbato $b = \bar{b} + e$.
Uno dei metodi più utilizzati appartenente alla famiglia (1) è dato dal metodo della discrepanza [21] che determina il parametro di regolarizzazione λ per cui valga

$$\|Ax_\lambda - b\|_2 = \|e\|_2. \tag{3.62}$$

Diversi sono invece i metodi appartenenti alla seconda famiglia: GCV [26], L-curva [14], criterio di quasi ottimalità [16]. In questo capitolo ci occuperemo soltanto del metodo della discrepanza e della GCV fornendo una loro rappresentazione esplicita a partire dalla $GSVD$ di (A, L).

3.6.1 Discrepanza nel caso discreto

Si considerino le funzioni:
(1) $\lambda \in \mathbb{R}^+ \to \epsilon(\lambda, y) = \|Kx_\lambda - y\|_Y$
(2) $\lambda \in \mathbb{R}^+ \to E(\lambda, y) = \|Lx_\lambda\|_Y$.
La prima distanza, detta discrepanza, permette di studiare l' accuratezza della soluzione x_λ, ossia la precisione con cui approssima il dato: si otterranno valori piccoli di $\epsilon(\lambda, y)$ in corrispondenza di soluzioni prossime a $K^\dagger y$. La seconda misura invece la "regolarità della soluzione x_λ.
Nel caso standard ($L = I$) è possibile dare una espressione esplicita alle funzioni

$\epsilon(\lambda, y)$ $E(\lambda, y)$ attraverso il sistema singolare di K qualora questo sia compatto [5] mentre nel caso generale ($L \neq I$) non esiste una trattazione teorica esauriente. Autonomamente, vogliamo studiare le funzioni $\epsilon(\lambda, y)$ e $E(\lambda, y)$ nel caso discreto cercando di estendere alcuni risultati del caso standard-teorico.
Nel nostro caso lavoreremo sulle funzioni discrete:
(1) $\lambda \in \mathbb{R}^+ \rightarrow \epsilon(\lambda, b) = \|Ax_\lambda - b\|_2$
(2) $\lambda \in \mathbb{R}^+ \rightarrow E(\lambda, b) = \|Lx_\lambda\|_2$
attraverso la $GSVD$ di (A, L).

Proposizione 3.6.1 *Per ogni $b \in M_{m,1}(\mathbb{R})$*
(1) la funzione $\epsilon^2(\lambda, b)$ è monotona strettamente crescente a valori in $[{\delta_0}^2, {\delta_0}^2 + \sum_{i=1}^{p}(u_i^T b)^2)$ dove $\delta_0 = \|(I - UU^T)b\|_2$.
(2) la funzione $E^2(\lambda, b)$ è monotona strettamente decrescente a valori in $[0, \sum_{i=1,\sigma_i \neq 0}^{p}(\frac{u_i^T b}{\gamma_i})^2) = I_E$.

Dimostrazione.
(1) Utilizzando la $GSVD$ di (A, L) si ottiene che la funzione

$$\epsilon^2(\lambda, b) = \|Ax_\lambda - b\|_2^2 = \sum_{i=1}^{p}\left[\frac{\lambda^2}{{\gamma_i}^2 + \lambda^2} u_i^T b\right]^2 + \delta_0^2 \tag{3.63}$$

con

$$\delta_0^2 = \|(I - UU^T)b\|_2^2 = \sum_{i=1}^{m} {b_i}^2 - \sum_{i=1}^{n}(u_i^T b)^2. \tag{3.64}$$

Verifichiamo la relazione (3.64):
$\delta_0^2 = \|(I - UU^T)b\|_2^2 = b^T(I - UU^T)^T(I - UU^T)b = b^T b - b^T UU^T b =$
$\|b\|_2^2 - \|U^T b\|_2^2 = \sum_{i=1}^{m} b_i^2 - \sum_{i=1}^{n}(u_i^T b)^2$.
Ricaviamo (3.63): a tale scopo poniamo $y = X^{-1}x_\lambda$
$\|Ax_\lambda - b\|_2^2 = \|U\Sigma X^{-1}x_\lambda - b\|_2^2 = \|U\Sigma y - b\|_2^2 = \|\Sigma y - U^T b\|_2^2 =$
$\sum_{i=1}^{p}\left[\frac{\gamma_i^2}{\gamma_i^2+\lambda^2} u_i^T b\right]^2 + \sum_{i=p+1}^{n}(u_i^T b)^2 - \sum_{i=1}^{m}(u_i^T b)^2 - 2\left[\sum_{i=1}^{p}\frac{\gamma_i^2}{\gamma_i^2+\lambda^2}(u_i^T b)^2 + \sum_{i=p+1}^{n}(u_i^T b)^2\right]$
e quindi, sommando gli addenti relativi agli stessi indici e utilizzando la relazione precedentemente trovata, otteniamo la (3.63). Allora:
$-\epsilon(0, b) = \delta_0^2$
$-\lim_{\lambda \rightarrow +\infty} \epsilon^2(\lambda, b) = \sum_{i=1}^{m} {b_i}^2 - \sum_{i=p+1}^{n}(u_i^T b)^2$
$-(\epsilon^2(\lambda, b))' = \sum_{i=1}^{p} \frac{4\lambda^3 {\gamma_i}^2}{({\gamma_i}^2+\lambda^2)^3}(u_i^T b)^2 > 0 \ \ \forall \lambda > 0$ ovvero $\epsilon^2(\lambda, b)$ è strettamente crescente.

(2) Ricorrendo nuovamente alla $GSVD$ si ha:

$$E^2(\lambda, b) = \|Lx_\lambda\|_2^2 = \sum_{i=1}^{p} \left[\frac{\gamma_i}{{\gamma_i}^2 + \lambda^2} u_i^T b \right]^2 \tag{3.65}$$

da cui si ottiene:
-$E^2(0,b) = \sum_{i=0,\sigma_i \neq 0}^{p} \left(\frac{u_i^T b}{\gamma_i} \right)^2$
-$lim_{\lambda \to +\infty} E^2(\lambda, b) = 0$
-$(E^2(\lambda, b))' = \sum_{i=1}^{p} \frac{-4\lambda\gamma_i^2}{(\gamma_i^2+\lambda^2)^3}(u_i^T b)^2 < 0 \ \forall \lambda > 0$ ovvero $E^2(\lambda, b)$ è strettamente decrescente. ∎

La monotonia delle funzioni $\epsilon(\lambda, b)$ e $E(\lambda, b)$ permette di trattare in modo semplice i seguenti problemi:
(1) $\min_{x \in S}\{\|Lx\|_2\} \quad S = \{x \in M_{n,1}(\mathbb{R}) : \|Ax - b\|_2 \leq \epsilon\}$
ovvero tra le soluzioni che approssimano il dato a livello ϵ determinare la più regolare ($\|Lx\|_2$ minimo),
(2) $\min_{x \in S}\{\|Ax - b\|_2\} \quad \{x \in M_{n,1}(\mathbb{R}) : \|Lx\|_2 \leq E\}$
ovvero tra tutte le soluzioni che soddisfano opportune condizioni di regolarità, imposte a priori ($\|Lx\|_2 \leq E$), trovare quella che meglio approssima il dato ($\|Ax - b\|_2$ minimo).
Nel caso (1) supponendo ${\delta_0}^2 \leq \epsilon \leq {\delta_0}^2 + \sum_{i=1}^{p}(u_i^T b)^2$ i valori del parametro di regolarizzazione λ per cui $\|Ax_\lambda - b\|_2 \leq \epsilon$ appartengono all' intervallo $(0, \lambda_\epsilon]$ dove λ_ϵ è il valore per cui $\epsilon(\lambda_\epsilon, b) = \epsilon$. Poichè la funzione $E(\lambda, b)$ è decrescente la soluzione più regolare è ottenuta per $\lambda = \lambda_\epsilon$. Il problema è allora ricondotto alla determinazione del parametro λ_ϵ.
Nel caso (2), supponendo $E \in I_E$ i valori del parametro di regolarizzazione per cui $\|Lx_\lambda\|_2 \leq E$ appartengono all' intervallo $[\lambda_E, +\infty)$ dove λ_E è il valore tale che $E(\lambda_E, b) = E$. Poichè la funzione $\epsilon(\lambda, b)$ relativa all' accuratezza di x_λ è crescente la soluzione che meglio approssima il dato b è ottenuta per $\lambda = \lambda_E$. Il problema è allora ricondotto alla ricerca del valore λ_E.
Notiamo infine che in entrambi i casi è necessario avere a disposizione informazioni a priori (ϵ, E) per poter determinare i corrispondenti parametri λ_ϵ e λ_E.
Nelle figure 3.5 e 3.6 sono riportati i grafici delle funzioni ϵ e E nel caso dell' esempio trattato nella sezione 3.3 .
In alternativa la determinazione del parametro di regolarizzazione può essere

ottenuta mediante considerazioni di tipo statistico, che non richiedono informazioni a priori, quali la GCV.

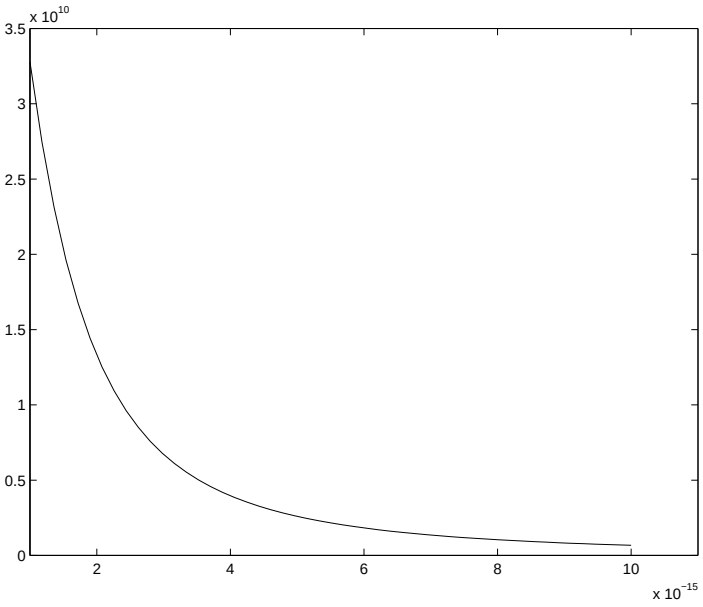

Figura 3.5. Grafico di E.

3.6.2 Generalized Cross Validation (GCV)

Per la determinazione di λ consideriamo la GCV che fornisce buoni risultati nel caso di problemi di equazioni integrali del primo tipo con nucleo non singolare. Il parametro λ è scelto in modo da minimizzare la funzione [2]

$$V(\lambda) = m\frac{\|Ax_\lambda - b\|_2^2}{[Tr(I - AR_\lambda)]^2} \tag{3.66}$$

che rappresenta una generalizzazione della funzione, introdotta nel caso standard,

$$V(\lambda) = \frac{1}{n}\sum_{k=1}^{n} |(Ax_{\lambda,k})_k - b_k|^2 \tag{3.67}$$

[2]Indichiamo con $Tr(Z)$ la traccia di una matrice quadrata Z ovvero la somma dei suoi elementi sulla diagonale principale.

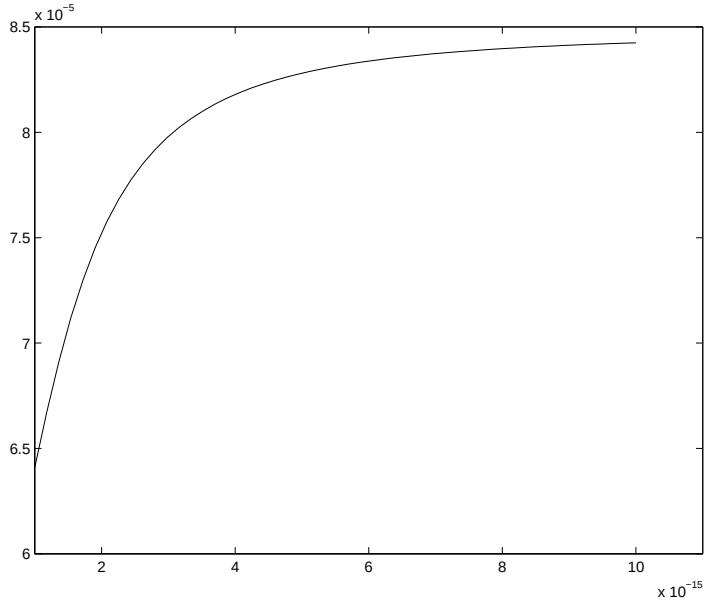

Figura 3.6. Grafico di ϵ.

, con $x_{\lambda,k}$ soluzione regolarizzata del problema (3.26) in cui la k^a componente del dato è tralasciata.
Per (3.49) abbiamo

$$I_m - AR_\lambda = I_m - Udiag(\frac{\gamma_1^2}{\gamma_1^2+\lambda^2}, ..., \frac{{\gamma_p}^2}{{\gamma_p}^2+\lambda^2}, 1, ..., 1)U^T$$

da cui otteniamo
$Tr(I_m - AR_\lambda) = m - (n-p) - \sum_{j=1}^{p} \frac{{\gamma_j}^2}{{\gamma_j}^2+\lambda^2} = m - n - \sum_{j=1}^{p}[\frac{\gamma_j^2}{\gamma_j^2+\lambda^2} - 1] = m - n + \sum_{j=1}^{p}[\frac{\lambda^2}{\gamma_j^2+\lambda^2}]$.
Usando le relazioni (3.66) e (3.63) si deduce che:

$$V(\lambda) = m \left[\frac{\sum_{j=1}^{p}[\frac{\lambda^2}{{\gamma_j}^2+\lambda^2} u_j^T b]^2 + \delta_0^2}{m - n + \sum_{j=1}^{p} \frac{\lambda^2}{\gamma_j^2+\lambda^2}} \right]^2 \tag{3.68}$$

Ad eccezione del fattore di proporzionalità $1/m$, il numeratore è uguale al quadrato del residuo, e per $L = I$ può essere espresso come

$$\sum_{j=1}^{m} \left[\frac{\lambda^2}{\lambda^2 + {\sigma_j}^2} (u_j^T b) \right]^2 . \tag{3.69}$$

Da (3.68) notiamo che quando i valori singolari γ_i sono molto piccoli , i corrispondenti termini nella somma a denominatore diventano prossimi a uno e quindi il residuo assume un andamento quasi costante per valori di λ sufficientemente grandi. Otteniamo che la funzione $V(\lambda)$ è molto schiacciata verso l' asse delle x e tale andamento rende talvolta difficile la determinazione del suo punto di minimo. Nel caso del nostro esempio il grafico della GCV è indicato in figura 3.7.

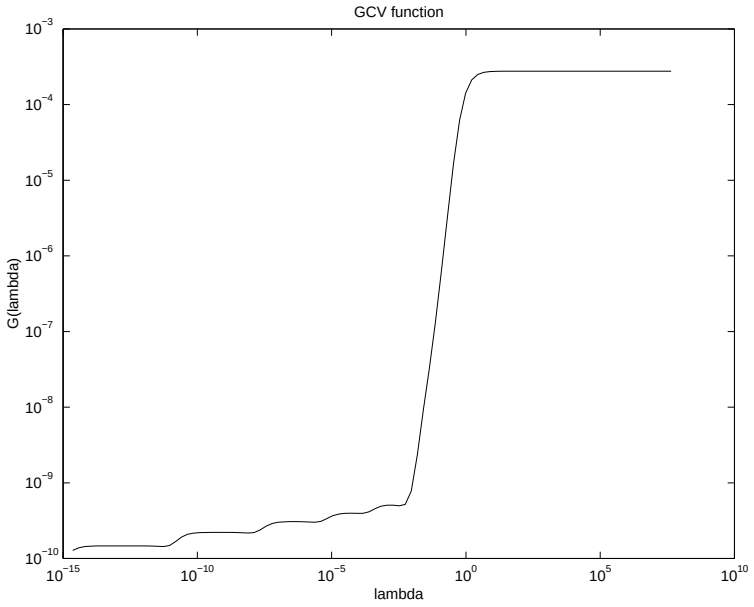

Figura 3.7. Grafico della GCV.

Rimane da chiarire come cambia il concetto di inverso L-generalizzato nel caso discreto. Ricordiamo che $x_L^\dagger$ è definito come l' unica soluzione del problema

$$\min_{x\in S}\{\|Lx\|_2\} \quad S = \{x \in M_{n,1}(\mathbb{R}) : \|Ax - b\|_2 = min\} \tag{3.70}$$

che generalizza

$$\min_{x\in S}\{\|x\|_2\} \quad S = \{x \in M_{n,1}(\mathbb{R}) : \|Ax - b\|_2 = min\}$$

a cui si riduce se $L = I_n$.
Dimostriamo che la soluzione $x_L^\dagger$ coincide con $x^\dagger$; pertanto, nel caso discreto,

le caratterizzazioni date nel capitolo 2 conducono alla definizione di soluzione generalizzata.
Basta infatti osservare che per (3.44)

$$Lx_{LS} = \sum_{i \in I} \frac{u_i^T b}{\sigma_i} Lx_i + \sum_{i \in I^C} c_i Lx_i + \sum_{i=p+1}^{n} (u_i^T b) Lx_i$$

ovvero

$$Lx_{LS} = \sum_{i \in I} \frac{u_i^T b}{\gamma_i} v_i + \sum_{i \in I^C} \mu_i c_i v_i.$$

Quindi otteniamo

$$\|Lx_{LS}\|_2^2 = \sum_{i \in I} \left(\frac{u_i^T b}{\gamma_i} \right)^2 + \sum_{i \in I^C} \mu_i^2 c_i^2$$

che è minimo per $c_i = 0 \ \forall i \in I^C$.

In questo capitolo, attraverso l' uso della $GSVD$, siamo riusciti a dare un' espressione esplicita di x_λ, (e quindi) ai fattori di filtro nonchè a esplicitare la funzione di discrepanza e la GCV.
Tipicamente per studiare e regolarizzare il problema (2.1) si utilizzano metodi di regolarizzazione in "forma standard, dove L è scelta uguale all' identità, supportati da routine numericamente affidabili. Nell' ipotesi che K sia lineare e compatto tale specificità è confermata da una perfetta simmetria tra soluzione numerica e teorica intendendo con questo che la soluzione numerica, espressa a partire dalla SVD della matrice A di discretizzazione di K, è ricavabile dalla soluzione teorica descritta utilizzando il sistema singolare dell' operatore K. Pertanto proprio la SVD rappresenta il metodo di indagine numerica con cui svolgere la regolarizzazione in forma standard.
Attraverso l' esempio di pagina 69 abbiamo osservato che la regolarizzazione standard può non essere sufficiente a determinare una buona approssimazione di f che è invece ottenuta ricorrendo al caso generale.
Tutti i risultati di questo capitolo e del precedente forniscono la base per lo studio dei metodi di regolarizzazione in forma generale e la sua stesura è stata condotta in base alle seguenti considerazioni:
-necessità di coerenza tra approccio teorico e numerico,
-estensione al caso generale di alcuni risultati caratteristici dei metodi di regolarizzazione in forma standard,

-dualità tra $SVD - GSVD$ e regolarizzazione in forma standard-generale,
-peculiarità della $GSVD$ come generalizzazione della SVD (cap.4).
Analizzeremo nel prossimo capitolo il legame che esiste tra queste due decomposizioni per poi fornire una relazione tra soluzione standard e soluzione generale; estenderemo infine i risultati della sezione 3.2 introducendo il concetto di $TGSVD$.

Capitolo 4

Relazione tra SVD e GSVD per problemi discreti di regolarizzazione

In questo capitolo riprendiamo il problema ai minimi quadrati in forma generale:

$$min\{\|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2\} \tag{4.1}$$

con

$$A \in M_{m,n}(\mathbb{R}), L \in M_{p,n}(\mathbb{R}), m \geq n \geq p,$$
$$rank(L) = p, N(A) \cap N(L) = \{0\}. \tag{4.2}$$

dove la condizione (4.2), come già evidenziato, assicura l' unicità della soluzione del problema (4.1).
Anzichè affrontare il problema (4.1) direttamente ci proponiamo [6] di trasformarlo, mediante un opportuno cambiamento di variabile, in un problema in forma standard espresso da:

$$\min\{\|\bar{A}\bar{x} - \bar{b}\|_2^2 + \lambda^2 \|\bar{x}\|_2^2\}. \tag{4.3}$$

Il vantaggio principale di questo approccio è quello di fornire un legame tra regolarizzazione standard e generale e rendere utilizzabile metodi efficienti sia diretti che iterativi per la risoluzione numerica di (4.1) ampliando, da un lato, il contesto in cui erano stati definiti e, dall'altro, proponendo possibili alternative alla $GSVD$.
Dal punto di vista numerico questo permetterà di fornire un legame tra la $GSVD$ di (A, L) e la SVD della matrice $\bar{A}$.

4.1 Trasformazione del problema dalla forma generale a quella standard

Introduciamo in questa sezione l'algoritmo, dovuto a Elden [6], che permette di scrivere il problema generale (4.1) nella forma standard (4.3). Tale procedimento è riassumibile in quattro passi e fa sostanzialmente uso della fattorizzazione QR.

1. Calcolo la fattorizzazione QR di L^T:

$$L^T = (\, K_p, \quad K_{n-p} \,) \begin{pmatrix} R_p \\ 0 \end{pmatrix} = K_p R_p; \tag{4.4}$$

2. Calcolo la fattorizzazione QR di AK_{n-p}:

$$AK_{n-p} = (\, H_{n-p}, \quad H_{m-(n-p)} \,) \begin{pmatrix} T_{n-p} \\ 0 \end{pmatrix} = H_{n-p} T_{n-p}; \tag{4.5}$$

3. Risolvo il problema standard (4.3) con:

$$\bar{A} = H^T_{m-(n-p)} A L^\dagger \in M_{m-(n-p),p}(\mathbb{R}) \tag{4.6}$$

$$\bar{b} = H^T_{m-(n-p)} b \in M_{m-(n-p),1}(\mathbb{R}). \tag{4.7}$$

4. Calcolo la soluzione di (4.1) come

$$x = L^\dagger \bar{x} + K_{n-p} T^{-1}_{n-p} H^T_{n-p} (b - A L^\dagger \bar{x}). \tag{4.8}$$

Notazione: Per maggiore chiarezza poniamo $o = n - p$ e $q = m - (n - p)$.

Ricaviamo innanzittutto alcune relazioni fondamentali:
-Dall' ortogonalitá di (K_p, K_0) otteniamo:

$$K_p^T K_p = I_p, \quad K_0^T K_0 = I_0, \quad K_p^T K_0 = 0. \tag{4.9}$$

-$L^\dagger = K_p R_p^{-T} \in M_{n,p}(\mathbb{R})$: per l'unicità della pseudoinversa basta dimostrare che valgono le relazioni:

$$L^\dagger L, LL^\dagger \quad \text{sono simmetriche e} \quad L^\dagger L L^\dagger = L^\dagger, LL^\dagger L = L.$$

Infatti da (4.4) si ha $L = R_p^T K_p^T$ e quindi :
-$L^\dagger L = K_p R_p^{-T} R_p^T K_p^T = K_p K_p^T$ è simmetrica,
-$LL^\dagger = R_p^T K_p^T K_p R_p^{-T} = I$ è simmetrica,
-$L^\dagger L L^\dagger = L^\dagger I = L^\dagger$,
-$LL^\dagger L = IL = L$.
Si osservi come sia di particolare importanza l'ipotesi che L abbia rango pieno per poter invertire R_p e scrivere $L^\dagger$ come $K_p R_p^{-T}$.

4.2 Regolarizzazione generale e standard

Vogliamo stabilire una relazione tra la SVD di $\bar{A}$ e la $GSVD$ della matrice (A, L).
Ricordiamo che la SVD di $\bar{A}$ è data da

$$\bar{A} = \bar{U}\bar{\Sigma}\bar{V}^T \tag{4.10}$$

con

$$\bar{U} \in M_{q,p}(\mathbb{R}), \bar{\Sigma}, \bar{V} \in M_{p,p}(\mathbb{R})$$

dove $\bar{U}^T\bar{U} = I_p, \bar{V}^T\bar{V} = I_p$ e $\bar{\Sigma}$ è una matrice diagonale con elementi diagonali $(\bar{\sigma}_i)_{i=1}^{rank(\bar{A})}$ disposti in ordine non crescente.

La $GSVD$ della matrice (A, L) relativa al problema generale è data da

$$A = U\Sigma X^{-1} = (\, U_p, \quad U_0\,) \begin{pmatrix} \Sigma_p & 0 \\ 0 & I_0 \end{pmatrix} (\, W_p, \quad W_0\,)^T, \tag{4.11}$$

$$L = VMX^{-1} = V(\, M_p, \quad 0)\ (\, W_p, \quad W_0\,)^T = VM_pW_p^T, \tag{4.12}$$

$$U \in M_{m,n}(\mathbb{R}), \ \ X \in M_{n,n}(\mathbb{R}),$$

$$\Sigma_p, M_p, V \in M_{p,p}(\mathbb{R}),$$

dove $U^TU = I_n$,$V^TV = I_p$ e gli elementi diagonali $(\sigma_i)_{i=1}^p$ e $(\mu_i)_{i=1}^p$ di Σ_p e M_p, rispettivamente, sono ordinati nel modo seguente:

$$0 \leq \sigma_1 \leq \ldots \leq \sigma_p < 1, \qquad 1 \geq \mu_1 \geq \ldots \geq \mu_p > 0. \tag{4.13}$$

Ricordiamo infine che vale la relazione

$$\sigma_i^2 + \mu_i^2 = 1 \quad \forall i = 1 \ldots p \tag{4.14}$$

e da (4.11) ricaviamo inoltre

$$W^T = (\, W_p, \quad W_0 \,)^T = X^{-1} = (\, X_p, \quad X_0 \,)^{-1}. \tag{4.15}$$

Osservazione 4.2.1 *Nell' osservazione 3.3.1 abbiamo introdotto alcune relazioni tra la GSVD di (A, L) e il range e il nucleo degli operatori lineari individuati dalle matrici A e L; ci chiediamo ora se esiste un legame tra $N(\bar{A})$, $R(\bar{A})$ e $N(A)$, $R(A)$ rispettivamente. Poichè, come vedremo più avanti,*

$$\sigma_i = \frac{\bar{\sigma}_{p-i+1}}{\sqrt{\bar{\sigma}^2_{p-i+1} + 1}} \quad \forall i = 1, 2, ..., p$$

, dove $\{\bar{\sigma}_i\}_{i=1}^p$ sono i valori singolari di $\bar{A} = H_q^T A L^\dagger$, abbiamo [1] *$card(I) = rank(\bar{\Sigma}) = rank(\bar{A})$. Da queste osservazioni deduciamo che*

$$dim(N(A)) = p - rank(\bar{\Sigma}) = p - rank(\bar{A})$$

e quindi vale

$$n = dim(N(A)) + dim(R(A)) = p - rank(\bar{A}) + dim(R(A))$$

da cui otteniamo:

$$rank(\bar{A}) = p - n + dim(R(A)) = p - n + rank(A).$$

Da quest' ultima si ha:

$$p - dim(N(\bar{A})) = dim(R(\bar{A})) = rank(\bar{A}) = p - n + rank(A) = p - dim(N(A))$$

ovvero $dim(N(\bar{A})) = dim(N(A))$. In particolare $rank(\bar{\Sigma}) = rank(A)$ infatti: $rank(\Sigma) = rank(\Sigma_p) + n - p = rank(\bar{A}) + n - p = rank(A)$ e quindi A ha rango pieno se e solo se $\bar{\Sigma}$ è invertibile.
Deduciamo inoltre che se L ha rango pieno allora $\mu_p > 0$ come già evidenziato implicitamente dalla relazione (4.13): infatti supposto, per assurdo, $\mu_p = 0$ l' insieme $\{Xe_i\}_{i=p,...,n}$ di $(n-p+1)$ vettori costituirebbe un sistema di generatori per lo spazio $N(L)$ di dimensione $(n - p)$.

[1] Si ricordi che

$$I = \{i \in \{1, 2, ..., p\} : \sigma_i \neq 0\}.$$

Lemma 4.2.2 *Se L ha rango massimo per righe, allora*

$$N(L) = R(K_0) = R(X_0), \quad R(AK_0) = R(H_0) = R(U_0). \tag{4.16}$$

Dimostrazione.
-$N(L) = R(K_0)$
Poichè $dimN(L) = n - rank(L) = n - p$ e $dimR(K_0) = rank(K_0) = n - p$ (infatti le colonne di $K_0 \in M_{n,n-p}(\mathbb{R})$ costituiscono un insieme di vettori ortonormali) basta verificare una sola delle inclusioni tra $N(L)$ e $R(K_0)$. A tale scopo osserviamo che $x \in R(K_0) \iff$ (esiste y tale che $K_0 y = x$) $\implies$ $(Lx = LK_0 y = R_p^T K_p^T K_0 y = 0) \implies x \in N(L)$.
$-R(X_0) = N(L)$
Osserviamo preliminarmente che per (4.15):

$$I_n = W^T X = (\, W_p \quad W_0 \,)^T \, (\, X_p \quad X_0 \,) = \begin{pmatrix} W_p^T \\ W_0^T \end{pmatrix} (\, X_p \quad X_0 \,) = \begin{pmatrix} W_p^T X_p & W_p^T X_0 \\ W_0^T X_p & W_0^T X_0 \end{pmatrix}$$

da cui otteniamo, uguagliando i quattro blocchi,

$$W_0^T X_0 = I_0, W_p^T X_0 = 0, W_0^T X_p = 0, W_p^T X_p = I_p. \tag{4.17}$$

Osserviamo che valgono le uguaglianze: $dimN(L) = n-p$ e $rank(X_0) = n-p$: la prima relazione è stata verificata nell' osservazione 3.3.1 mentre per la seconda se, per assurdo, fosse $rank(X_0) < (n-p)$ si avrebbe $rank(X) = rank\,(\, X_p, \quad X_0 \,) < n \implies X \notin Gl_n(\mathbb{R})$ il che è assurdo.
Dimostriamo l' inclusione $R(X_0) \subseteq N(L)$: infatti $x \in R(X_0) \iff$ (esiste y tale che $x = X_0 y$) $\implies$ $(Lx = LX_0 y = V M_p W_p^T X_0 y = 0) \implies x \in N(L)$.
Da queste relazioni concludiamo, in modo analogo alla precedente verifica, che $R(X_0) = N(L)$ e quindi $N(L) = R(K_0) = R(X_0)$.
Verifichiamo innanzitutto che $R(AX_0) = R(U_0)$: infatti
$A = U\Sigma X^{-1} \iff AX = U\Sigma \iff A\,(\, X_p, \quad X_0 \,) = (\, U_p, \quad U_0 \,) \begin{pmatrix} \Sigma_p & 0 \\ 0 & I_0 \end{pmatrix} \iff$
$(\, AX_p, \quad AX_0 \,) = (\, U_p \Sigma_p, \quad U_0 \,) \implies AX_0 = U_0 \implies R(AX_0) = R(U_0)$.
Poichè $R(AK_0) = R(AX_0)$ (basta applicare la definizione di range di un operatore e le proprietà trovate) otteniamo $R(AK_0) = R(AX_0) = R(U_0)$ come volevamo dimostrare. ∎

Lemma 4.2.3 *Se L ha rango massimo per righe, allora la pseudoinversa di L è data da:*

$$L^{\dagger} = W_p(W_p^T W_p)^{-1} M_p^{-1} V^T = P X_p M_p^{-1} V^T, \tag{4.18}$$

dove P rappresenta la matrice di proiezione ortogonale su $R(L^T)$.

Dimostrazione.
Poichè $L \in M_{p,n}(\mathbb{R})$ ha rango massimo per righe e $p \leq n$ si ha $rank(L) = p$ in accordo con (4.2): quest' ultima condizione, insieme a (4.12), ci permette di scrivere

$$L^{\dagger} = L^T(LL^T)^{-1} = W_p M_p V^T (V M_p W_p^T W_p M_p V^T)^{-1}$$

ed essendo

$$(V M_p W_p^T W_p M_p V^T)^{-1} = V M_p^{-1} (W_p^T W_p)^{-1} M_p^{-1} V^T$$

(infatti per verifica diretta, ricordando che per ortogonalità $V^T V = V V^T = I_p$, abbiamo

$$\begin{gathered}(V M_p W_p^T W_p M_p V^T) V M_p^{-1} (W_p^T W_p)^{-1} M_p^{-1} V^T \\ = V M_p W_p^T W_p M_p (V^T V) M_p^{-1} (W_p^T W_p)^{-1} M_p^{-1} V^T \\ = I\end{gathered}$$

e analogamente

$$(V M_p^{-1} (W_p^T W_p)^{-1} M_p^{-1} V^T)(V M_p W_p^T W_p M_p V^T) = I)$$

otteniamo:

$$L^{\dagger} = W_p M_p V^T V M_p^{-1} (W_p^T W_p)^{-1} M_p^{-1} V^T = W_p (W_p^T W_p)^{-1} M_p^{-1} V^T$$

dove nell' ultimo passaggio si è utilizzata l' uguaglianza $M_p V^T V M_p^{-1} = I$.
Poichè $XW^T = I$ si ha

$$P = PXW^T = PX_pW_p^T;$$

mentre dall'uguaglianza $L = V M_p W_p^T$ deduciamo $R(L^T) = R(W_p)$: infatti $dim R(L^T) = dim R(W_p)$ e $x \in R(L^T) \iff$ (esiste y tale che$L^T y = x) \iff W_p M_p V^T y = x \iff x \in R(W_p)$. Ricordando che $XW^T = I$, $X_p W_p^T = I_p$ ovvero

$$W_p = PW_p = (PX_pW_p^T)W_p = PX_p(W_p^T W_p)$$

da cui si deduce

$$PX_p = W_p(W_p^T W_p)^{-1}$$

e, quindi, vale la relazione cercata

$$L^\dagger = W_p(W_p^T W_p)^{-1} M_p^{-1} V^T = PX_p M_p^{-1} V^T.$$

∎

Osservazione 4.2.4 *Verifichiamo che vale la relazione* $W_p^T K_0 = 0$. *Infatti dal lemma (4.2.3) si ottiene:*

$$W_p = L^\dagger V M_p(W_p^T W_p) \Longleftrightarrow W_p^T = (W_p^T W_p)^T M_p^T V^T (L^\dagger)^T = (W_p^T W_p) M_p^T V^T R_p^{-1} K_p^T$$

da cui

$$W_p^T K_0 = (W_p^T W_p)^T M_p^T V^T R_p^{-1} K_p^T K_0 = 0$$

Lemma 4.2.5 *La sottomatrice* X_0 *in* $(\, X_p, \quad X_0\,)$ *è data da*

$$X_0 = K_0 T_0^{-1} H_0^T U_0.$$

Dimostrazione.
Osserviamo innanzittutto che $K_0 K_0^T$ è la matrice di proiezione ortogonale su $R(K_0) = N(L)$: per la verifica basta dimostrare che

$$(K_0 K_0^T) \text{ è idempotente e simmetrica.}$$

Abbiamo infatti:

$$(K_0 K_0^T)(K_0 K_0^T) = K_0(K_0^T K_0) K_0^T = K_0 K_0^T$$

mentre la simmetria si ottiene passando alla trasposta.
Poichè $\{x_i\}_{i=p+1}^n \subset N(L)$ abbiamo $K_0 K_0^T X_0 = X_0$ e quindi vale:

$$(W_0^T K_0)(K_0^T X_0) = W_0{}^T((K_0 K_0^T) X_0) = W_0^T X_0 = I_0 \tag{4.19}$$

da cui segue

$$(W_0^T K_0)^{-1} = K_0^T X_0. \tag{4.20}$$

Per la dimostrazione utilizzeremo le identità:

$$(U_0 W_0^T K_0)^\dagger = (W_0^T K_0)^{-1} U_0^T \quad \text{e} \quad (AK_0)^\dagger = T_0^{-1} H_0^T \tag{4.21}$$

che, come fatto precedentemente, vanno verificate in modo diretto.
Ricordando che, per l'osservazione precedente, $W_p^T K_0 = 0$, otteniamo per (4.11)

$$AK_0 = (U_p \Sigma_p W_p^T + U_0 W_0^T) K_0 = U_p \Sigma_p (W_p^T K_0) + U_0 W_0^T K_0 = U_0 W_0^T K_0, \tag{4.22}$$

e quindi per (4.20), (4.21), (4.22) si ha:

$$\begin{aligned} K_0 T_0^{-1} H_0^T = K_0 (AK_0)^\dagger = K_0 (U_0 W_0^T K_0)^\dagger = K_0 (W_0^T K_0)^{-1} U_0^T = K_0 (K_0^T X_0) U_0^T \\ = (K_0 K_0^T) X_0 U_0^T = X_0 U_0^T \end{aligned}$$

il che implica

$$K_0 T_0^{-1} H_0^T = X_0 U_0^T$$

ovvero, per l'ortogonalità di U_0, $X_0 = K_0 T_0^{-1} H_0^T U_0$ come volevamo dimostrare.

■

Il teorema seguente esprime la relazione tra la $GSVD$ di (A, L) e la SVD di $\bar{A}$ ottenuta da A e L attraverso il procedimento descritto nelle precedenti pagine.

Teorema 4.2.6 *Siano A, L e $\bar{A}$ definite mediante le equazioni(4.1),(4.3),(4.6) e supponiamo che la SVD di $\bar{A}$ e la GSVD di (A,L) siano espresse da (4.10),(4.11), (4.12). Allora:*

$$\bar{U} = H_q^T U_p \Pi, \quad \bar{\Sigma} = \Pi \Sigma_p M_p^{-1} \Pi, \quad \bar{V} = V \Pi, \tag{4.23}$$

$$U_p = H_q \bar{U} \Pi, U_0 = H_0, \quad X_0 = K_0 T_0^{-1}, \tag{4.24}$$

dove $H = (\, H_0, \quad H_q\,)$, K_0 e T_0 sono definite dalle equazioni (4.4),(4.5) e $\Pi = antidiag(1, ..., 1)$ è la matrice antidiagonale di ordine p.

Dimostrazione.
Dimostriamo innanzitutto la relazione (4.23). Usando (4.6), (4.11) e il Lemma (4.2.3) otteniamo

$$\begin{aligned} \bar{A} = H_q^T A L^\dagger = H_q^T (U_p \Sigma_p W_p^T + U_0 W_0^T) W_p (W_p^T W_p)^{-1} M_p^{-1} V^T = H_q^T U_p \Sigma_p M_p^{-1} V^T \\ = (H_q^T U_p \Pi)(\Pi \Sigma_p M_p^{-1} \Pi)(V \Pi)^T \end{aligned}.$$

Poichè H_0 e U_0 sono matrici con rango massimo (infatti sono sottomatrici di matrici ortogonali) otteniamo, applicando il lemma (4.2.2) e $H_0^T H_q = 0$, che $H_0^T U_p = 0$. Allora,

$$H^T U_p = (\, H_0, \quad H_q \,)^T U_p = \begin{pmatrix} H_0^T U_p \\ H_q^T U_p \end{pmatrix} = \begin{pmatrix} 0 \\ H_q^T U_p \end{pmatrix}, \tag{4.25}$$

da cui segue che

$$(H_q^T U_p)^T (H_q^T U_p) = (H^T U_p)^T (H^T U_p) = U_p^T H H^T U_p = U_p^T U_p = I_p,$$

cioè le colonne della matrice $H_q^T U_p$ sono ortonormali.
In particolare, essendo Π ortonormale, otteniamo che $H_q^T U_p \Pi$ è ortonormale. Inoltre, essendo M_p diagonale e invertibile, $\Pi \Sigma_p M_p^{-1}$ è una matrice diagonale con entrate positive ordinate in modo non decrescente mentre $V\Pi$ è ortonormale.
Allora (4.23) rappresenta la SVD della matrice $\bar{A}$.
Dimostriamo ora (4.24). Da (4.25) abbiamo che

$$U_p = H \begin{pmatrix} 0 \\ H_q^T U_p \end{pmatrix} = H_q H_q^T U_p = H_q (H_q^T U_p \Pi)\Pi = H_q \bar{U} \Pi.$$

Poichè la matrice U_0 non è unica possiamo scegliere $U_0 = H_0$. Dal lemma (4.2.5) otteniamo allora che:

$$X_0 = K_0 T_0^{-1} H_0^T U_0 = K_0 T_0^{-1} H_0^T H_0 = K_0 T_0^{-1}.$$

∎

4.3 Algoritmo $gsvd - stdform$

Vogliamo ora ricavare esplicitamente l'algoritmo per la $GSVD$ di (A, L), eq.ni (4.11), (4.12), facendo uso di alcune relazioni prima definite, che riportiamo qui di seguito per maggiore chiarezza:
-Fattorizzazione QR di L^T:

$$L^T = (\, K_p, \quad K_0 \,) \begin{pmatrix} R_p \\ 0 \end{pmatrix} = K_p R_p$$

-Fattorizzazione QR di AK_{n-p}:

$$AK_o = (\, H_0, \quad H_q) \begin{pmatrix} T_0 \\ 0 \end{pmatrix} = H_0 T_0$$

Pongo:

$$\bar{A} = H_q^T A L^\dagger \in M_{q,p}(\mathbb{R})$$

$$\bar{b} = H_q^T b \in M_{q,1}(\mathbb{R}).$$

Ci riferiremo a questo algoritmo con il termine $gsvd-stdform$ per distinguerlo dall' algoritmo $gsvd$ basato su [24] e trattato nel capitolo 1.
Una parte molto delicata dell' algoritmo di Elden è la costruzione della matrice $\bar{A}$ su cui si basa sostanzialmente il calcolo della $GSVD$ di (A, L): tenuto conto che $\bar{A} = H_q^T A L^\dagger$ si nota che il mal condizionamento di A si rifletterà, in generale, su $\bar{A}$ (vedi (4.34)). Vogliamo pertanto analizzare quale sia il modo migliore di calcolare $\bar{A}$ localizzando la nostra attenzione sul calcolo della pseudoinversa di L. In generale sono possibili due alternative:
-$L^\dagger = K_p R_p^{-T}$
-$L^\dagger = pinv(L)$ dove con $pinv$ intendiamo la pseduoinversa di L calcolata atraverso la SVD di L.
Per meglio capire quale scelta sia più opportuna osserviamo preliminarmente che $K_2(L) = K_2(L^T) = K_2(K_p R_p) = K_2(R_p)$: tale relazione evidenzia come il condizionamento di L si ripercuota interamente su R_p. Rimanendo nell' ipotesi che L rappresenti la discretizzazione di un operatore differenziale calcoliamo pertanto $L^\dagger$ come $K_p R_p^{-T}$ in modo numericamente stabile e con un più basso costo computazionale al pari della scelta $L^\dagger = pinv(L)$. Il calcolo di $L^\dagger$ come $pinv(L)$ sarebbe necessario qualora la matrice L fosse mal condizionata; tuttavia tale condizione non va presa in considerazione nel caso di problemi di regolarizzazione in quanto L è sempre scelta ben condizionata dovendo garantire il buon condizionamento di X..

Detta Π=antidiag(1,1,...,1) $\in GL_p(\mathbb{R})$ da (4.23) e (4.24) abbiamo $U = (\, U_p \quad U_0\,) = (\, H_q \bar{U}\Pi, \quad H_0\,)$, $V = \bar{V}\Pi$ mentre rimangono da determinare $\Sigma\,, M\,, W^T = X^{-1}\,$.
Le matrici Σ_p e M_p , a partire dalle quali si costruiscono $\Sigma = \begin{pmatrix} \Sigma_p & 0 \\ 0 & I_0 \end{pmatrix}$ e $M = (\, M_p \quad 0\,)$, si ottengono facilmente dalle relazioni $\bar{\Sigma} = \Pi \Sigma_p M_p^{-1} \Pi$ e

$\sigma_i^2 + \mu_i^2 = 1 \ \ \forall i = 1, 2, ..., p$. Infatti abbiamo:

$$\bar{\Sigma} = \Pi \Sigma_p M_p^{-1} \Pi \Longleftrightarrow \Sigma_p M_p^{-1} = \Pi \bar{\Sigma} \Pi \Longleftrightarrow \frac{\sigma_i}{\mu_i} = \bar{\sigma}_{p-i+1} \forall i = 1, 2, ..., p :$$

e quindi $\forall i = 1, 2, ..., p$

$$\begin{cases} \frac{\sigma_i}{\mu_i} = \bar{\sigma}_{p-i+1} \\ \sigma_i^2 + \mu_i^2 = 1 \end{cases} \Longleftrightarrow \begin{cases} \sigma_i = \bar{\sigma}_{p-i+1} \mu_i \\ (\bar{\sigma}_{p-i+1}^2 + 1)\mu_i^2 = 1 \end{cases} \Longleftrightarrow$$

$$\begin{cases} \mu_i = \frac{1}{\sqrt{\bar{\sigma}_{p-i+1}^2 + 1}} \\ \sigma_i = \frac{\bar{\sigma}_{p-i+1}}{\sqrt{\bar{\sigma}_{p-i+1}^2 + 1}} \end{cases}.$$

Ricordando che la matrice $\bar{A}$ è, in generale, mal condizionata esisterà $k \in \{1, 2, ..., p\}$ tale che $\bar{\sigma}_j \approx 0 \ \ \forall j = k+1, ..., p$ ovvero si avrà $\bar{\sigma}_{p-j+1} \approx 0 \ \ \forall j = 1, 2, ..., p-k$. Questo ci permette di scrivere:

$$\mu_i = \frac{1}{\sqrt{\bar{\sigma}_{p-j+1}^2 + 1}} \approx 1 - \frac{1}{2}\bar{\sigma}_{p-j+1}^2 \approx 1, \quad \sigma_i = \frac{\bar{\sigma}_{p-j+1}}{\sqrt{\bar{\sigma}_{p-j+1}^2 + 1}} \approx \bar{\sigma}_{p-j+1} \approx 0 \tag{4.26}$$

e attraverso di essa individuiamo una analogia tra mal condizionamento di A, Σ e buon condizionamento di L, M. Osserviamo inoltre che i valori $(\sigma_i)_{i=1}^p$ e $(\mu_i)_{i=1}^p$ di (A, L) sono stati calcolati nel modo migliore possibile ovvero utilizzando esclusivamente i valori singolari di A: per tale ragione risulteranno poco sensibili a eventuali perturbazioni di $\bar{A}$.

Ricaviamo ora $W^T = X^{-1}$. Il calcolo di W, o equivalentemente di X, può essere eseguito in più modi utilizzando diverse relazioni: va però osservato che l' espressione ottenuta non è sempre numericamente stabile. Calcoliamo la matrice $X = (X_p, X_0)$: poichè $X_0 = K_0 T_0^{-1}$ rimane da determinare soltanto la sottomatrice X_p.

Ricordando che $P = L^\dagger L$ possiamo ricavare $L^\dagger$ in due diversi modi:

-utilizzando la relazione $L^\dagger = K_p R_p^{-T}$ ottenuta a partire dalla fattorizzazione QR di L^T;

$$P = L^\dagger L = K_p R_p^{-T} R_p K_p^T = K_p K_p^T \tag{4.27}$$

-utilizzando la SVD di L, o meglio la $SVDS$ [7] che sfrutta la sparsità di L $(L = QDZ^T)$:

$$P = L^\dagger L = Z D^\dagger Q^T Q D Z^T = Z_p Z_p^T \tag{4.28}$$

con Z_p sottomatrice di Z ottenuta considerando le sue prime p colonne.
Nonostante le due espressione di P siano semplici e ottenute mediante il calcolo di fattorizzazioni di uso comune il condizionamento di P risulta essere, in generale, troppo alto per invertire P usando (4.27) o (4.28) e dedurne X_p attraverso (4.26). Dobbiamo pertanto trovare un' espressione alternativa che risulti numericamente praticabile. A tale scopo, anzichè calcolare la matrice X, calcoliamo $W = (\, W_p \quad W_0 \,)$. Dalla relazione (4.12) otteniamo $W_p = L^T V M_p^{-1}$ e quindi utilizzando la (4.11) si ha:

$A = U_p \Sigma_p W_p^T + U_0 W_0^T \leftrightarrow U_0 W_0^T = A - U_p \Sigma_p W_p^T \leftrightarrow W_0^T = U_0^T (A - U_p \Sigma_p W_p^T) =$

$U_0^T A - U_0^T U_p \Sigma_p W_p^T = U_0^T A$ cioè $W_0 = A^T H_0$ e pertanto

$$(\, W_p, W_0 \,) = (\, L^T V M_p^{-1}, A^T H_0 \,) \tag{4.29}$$

Osserviamo che l' espressione $W_p = L^T V M_p^{-1}$ è numericamente significativa se L è ben condizionata.
È possibile infine una seconda alternativa per il calcolo di W ricavabile formalmente dalla $GSVD$ di (A,L) sfruttando la relazione
$\Sigma^T \Sigma + M^T M = I$:

$$W = A^T U \Sigma + L^T V M.$$

Si osservi come quest' ultima relazione non richieda l' inversione di alcuna matrice. Queste osservazioni, oltre che evidenziare alcune scelte fatte nella formulazione dell' algoritmo $gsvd - stdform$, fanno da un lato intuire che la stabilità dell' algoritmo dipenderà da $K_2(L)$ (cfr (4.35)) e dall' altro che la sua efficacia è legata alla scelta di una matrice L ben condizionata. Per il listato dell'algoritmo $gsvd - stdform$ si consulti l' appendice in cui è anche presente un commento per i passaggi fondamentali.

4.4 Stabilità

A ulteriore conferma della validità dell' algoritmo di Elden per la regolarizzazione di problemi mal posti vogliamo verificare le relazioni:

$$\|Ax - b\|_2 = \|\bar{A}\bar{x} - \bar{b}\|_2, \quad \|Lx\|_2 = \|\bar{x}\|_2. \tag{4.30}$$

Queste uguaglianze sono particolarmente importanti in connessione con i metodi di individuazione del parametro di regolarizzazione in quanto evidenziano che ogni strategia di scelta del parametro , basato su queste norme e applicato al problema in forma generale e standard, porterà allo stesso parametro di regolarizzazione.

Veniamo ora alla verifica di (4.30): per (4.8) si ha
$x = L^\dagger\bar{x}+K_0T_0^{-1}H_0^T(b-AL^\dagger\bar{x})$ e quindi $Ax = AL^\dagger\bar{x}+AK_0T_0^{-1}H_0^T(b-AL^\dagger\bar{x}) = AL^\dagger\bar{x}+H_0T_0T_0^{-1}H_0^T(b-AL^\dagger)\bar{x} = AL^\dagger\bar{x}+H_0H_0^T(b-AL^\dagger\bar{x}) \Longrightarrow H_q^T(Ax-b) = H_q^TAx-H_q^Tb = H_q^T[AL^\dagger\bar{x}+H_0H_0^T(b-AL^\dagger\bar{x})]-H_q^Tb = H_q^TAL^\dagger\bar{x}-H_q^Tb = \bar{A}\bar{x}-\bar{b}$.
Allora

$$\|Ax-b\|_2 = \|H_q^T(Ax-b)\| = \|\bar{A}\bar{x}-\bar{b}\|_2$$

come volevamo dimostrare.

Analoga è la verifica della seconda relazione:
$Lx = LL^\dagger\bar{x}+LK_0T_0^{-1}H_0^T(b-AL^\dagger\bar{x}) = LL^\dagger\bar{x}+R_p^TK_p^TK_0T_0^{-1}H_0^T(b-AL^\dagger\bar{x}) = LL^\dagger\bar{x} = \bar{x}$ da cui $\|Lx\|_2 = \|\bar{x}\|_2$.

Vogliamo ora analizzare la stabilità dell' algoritmo di Elden: per semplificare la trattazione supporremo A con rango massimo.

Teorema 4.4.1 *Se A è una matrice a rango pieno, allora la pseudoinversa di $\bar{A} = H_q^TAL^\dagger$ è:*

$$\bar{A}^\dagger = LA^\dagger H_q. \tag{4.31}$$

Dimostrazione.

Poichè, per ipotesi, la matrice A ha rango pieno possiamo esprimere $A^\dagger$ a partire dalla $GSVD$ di (A,L) espressa da (4.11): ovvero

$$A^\dagger = X\Sigma^{-1}U^T.$$

(si ricordi che Σ è invertibile se e solo se A ha rango pieno).

Utilizzando la precedente espressione di $A^\dagger$, il teorema 4.2.6 e (4.12) otteniamo

$$LA^\dagger H_q = VMX^{-1}X\Sigma^{-1}U^TH_q = VM\Sigma^{-1}U^TH_q$$

$$= V\begin{pmatrix} M_p, & 0\end{pmatrix}\begin{pmatrix}\Sigma_p^{-1} & 0\\ 0 & I_0\end{pmatrix}\begin{pmatrix}U_p^T\\ U_0^T\end{pmatrix}H_q$$

$$= V\begin{pmatrix} M_p\Sigma_p^{-1}, & 0\end{pmatrix}\begin{pmatrix}U_p^TH_q\\ U_0^TH_q\end{pmatrix} = VM_p\Sigma_p^{-1}U_p^TH_q$$

$$= (V\Pi)(\Pi\Sigma_pM_p^{-1}\Pi)^{-1}(H_q^TU\Pi)^T = \bar{V}\bar{\Sigma}^{-1}\bar{U}^T = \bar{A}^\dagger$$

∎

Sia $\tilde{U}\tilde{\Sigma}\tilde{X}^{-1} = A + E$ la $GSVD$ di A calcolata attraverso l' algoritmo di Van Loan dove la norma di E soddisfa la condizione $\|E\|_2 = c\varepsilon\|A\|_2$ in cui c è una funzione crescente in m e n e ε rappresenta la precisione di macchina. Allora:

$$\begin{gathered} E(\tilde{\Sigma}\tilde{X}^{-1})^{-1} = (\tilde{U}\tilde{\Sigma}\tilde{X}^{-1} - A)(\tilde{\Sigma}\tilde{X}^{-1})^{-1} \\ = \tilde{U} - A(\tilde{\Sigma}\tilde{X}^{-1})^{-1} = \tilde{U} - U\Sigma X^{-1}(\tilde{\Sigma}\tilde{X}^{-1})^{-1} \\ = \tilde{U} - U\Sigma X^{-1}\tilde{X}\tilde{\Sigma}^{-1} \\ \cong \tilde{U} - U \end{gathered}$$

e quindi

$$\begin{gathered} \tilde{U} - U\|_2 = \|E(\tilde{\Sigma}\tilde{X}^{-1})^{-1}\|_2 \leq \|E\|_2\|(\tilde{\Sigma}\tilde{X}^{-1})^{-1}\|_2 \\ \cong \|E\|_2\|(\Sigma X^{-1})^{-1}\|_2 \\ = \|E\|_2\|(\Sigma X^{-1})^{-1}U^T\|_2 = \|E\|_2\|A^\dagger\|_2 \end{gathered}$$

e ricordando che $K_2(A) = \|A\|_2\|A^\dagger\|_2$ otteniamo:

$$\|\tilde{U} - U\|_2 <\sim \frac{\|E\|_2}{\|A\|_2}K_2(A). \quad (4.32)$$

Calcoliamo ora una maggiorazione per $\|\tilde{\bar{U}} - \bar{U}\|_2$ dove $\tilde{\bar{U}}$ è la versione numerica di $\bar{U}$ calcolata mediante l'algoritmo proposto. Seguendo un procedimento analogo al precedente si ha: $\tilde{\bar{U}}\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T = \bar{A} + \bar{E}$ (SVD calcolata per $\bar{A}$) $\Longleftrightarrow \bar{E} = \tilde{\bar{U}}\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T - \bar{A}$ e quindi:

$$\begin{gathered} \bar{E}(\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T)^{-1} = (\tilde{\bar{U}}\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T - \bar{A})(\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T)^{-1} \\ = \tilde{\bar{U}}\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T(\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T)^{-1} - \bar{A}(\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T)^{-1} \\ = \tilde{\bar{U}} - \bar{U}\bar{\Sigma}\bar{V}^T(\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T)^{-1} \cong \tilde{\bar{U}} - \bar{U}. \end{gathered}$$

Dalla precedente relazione otteniamo:

$$\begin{gathered} \|\tilde{\bar{U}} - \bar{U}\|_2 \cong \|\bar{E}(\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T)^{-1}\|_2 \leq \|\bar{E}\|_2\|(\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T)^{-1}\|_2 \\ \cong \|\bar{E}\|_2\|(\bar{\Sigma}\bar{V}^T)^{-1}\|_2 = \|\bar{E}\|_2\|(\bar{\Sigma}\bar{V}^T)^{-1}\bar{U}^T\|_2 \\ = \|\bar{E}\|_2\|\|\bar{A}^\dagger\|_2 \end{gathered}$$

da cui si ha

$$\|\tilde{\bar{U}} - \bar{U}\|_2 <\sim \|\bar{E}\|_2\|\bar{A}^\dagger\|_2 = \|\bar{E}\|_2\frac{K_2(\bar{A})}{\|\bar{A}\|_2}. \quad (4.33)$$

La maggiorazione:

$$\begin{gathered} K_2(\bar{A}) = \|\bar{A}\|_2\|\bar{A}^\dagger\|_2 = \|\bar{A}\|_2\|LA^\dagger H_q\|_2 \\ = \|\bar{A}\|_2\|LA^\dagger\|_2 \leq \|\bar{A}\|_2\|L\|_2\|A^\dagger\|_2 \\ = \|H_q^T AL^\dagger\|_2\|L\|_2\|A^\dagger\|_2 \\ \leq \|A\|_2\|L^\dagger\|_2\|L\|_2\|A^\dagger\|_2 = K_2(A)K_2(L) \end{gathered} \quad (4.34)$$

sostituita in (4.33) permette di concludere:

$$\|\tilde{\bar{U}} - \bar{U}\|_2 \leq \frac{\|\bar{E}\|_2}{\|\bar{A}\|_2} K_2(A)K_2(L). \tag{4.35}$$

Pertanto:

$$\|U_p - \tilde{U}_p\|_2 = \|H_q(\bar{U} - \tilde{\bar{U}})\Pi\|_2 = \|\bar{U} - \tilde{\bar{U}}\|_2 \leq \frac{\|\bar{E}\|_2}{\|\bar{A}\|_2} K_2(A)K_2(L).$$

Sia $L + E = \tilde{V}\tilde{M}\tilde{X}^{-1}$ la ricostruzione di L ottenuta mediante l'algoritmo di Elden. Ricordando che $\tilde{\bar{U}}\tilde{\bar{\Sigma}}\tilde{\bar{V}}^T = \bar{A} + \bar{E}$ rappresenta la SVD di $\bar{A}$ da (4.23) otteniamo $\tilde{\bar{V}} = \tilde{V}\Pi$ e quindi

$$\|\tilde{V} - V\|_2 = \|\tilde{\bar{V}}\Pi - \bar{V}\Pi\|_2 = \|\tilde{\bar{V}} - \bar{V}\|_2 \approx \alpha\bar{E}$$

con α costante di proporzionalità.
Stimiamo infine $\|X^{-1} - \tilde{X}^{-1}\|_2$.
Poichè $\bar{X}^{-1} = \begin{pmatrix} M_p^{-1}\bar{V}\Pi L \\ H_0^T A \end{pmatrix}$ abbiamo:

$$\|X^{-1} - \tilde{X}^{-1}\|_2 = \| \begin{pmatrix} (M_p^{-1}\bar{V} - \tilde{M_p}^{-1}\tilde{\bar{V}})\Pi L \\ (H_0 - \tilde{H_0})^T A \end{pmatrix} \|_2 \approx$$

$$\| \begin{pmatrix} 0 \\ E^T A \end{pmatrix} \|_2 = \|E^T A\|_2 \leq \|E\|_2 \|A\|_2$$

con E errore relativo alla matrice H_0 ottenuta dalla fattorizzazione QR di AK_0. Concludiamo quindi che la soluzione calcolata attraverso il metodo di Elden è stabile quando L è ben condizionata: tale condizione è come vedremo (Teorema (4.6.1)) soddisfatta per opportune scelte di L. Si osservi che L deve comunque essere ben condizionata al fine di assicurare la "regolarità della soluzione del problema (4.1).

4.5 Implementazione Numerica

Abbiamo dimostrato, attraverso i precedenti teoremi, che un problema di regolarizzazione discreta nella forma generale è riconducibile a un problema nella forma standard individuando così un legame tra la SVD associata al problema (4.3) e la $GSVD$ del problema (4.1). In questo ordine di idee si è verificato che la stabilità della $GSVD$ calcolata, mediante l'algoritmo di Elden, dipende

dal numero di condizionamento della matrice di regolarizzazione L: quindi la $GSVD$ calcolata è affidabile se L è ben condizionata. Vogliamo ora analizzare due importanti situazioni, comunemente implementate nei metodi di regolarizzazione, nelle quali la matrice di regolarizzazione L rappresenta l' approssimazione discreta di un operatore di derivata prima o seconda.

Teorema 4.5.1 *Sia* $e = [1, 1, \ldots, 1]^T \in M_{n,1}(\mathbb{R})$, $f = [1, 2, \ldots, n]^T \in M_{n,1}(\mathbb{R})$, *e definiamo* $\hat{a} = Ae/n$ *e* $\tilde{a} = Af/n$. *Se* $L = L_1$ *rappresenta la matrice di approssimazione discreta per un operatore di derivata prima,*

$$L_1 = \begin{pmatrix} 1 & -1 & & & 0 \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ 0 & & & 1 & -1 \end{pmatrix} \in M_{n-1,n}(\mathbb{R}) \tag{4.36}$$

allora

$$K_0 = n^{-1/2}e \quad , H_0 = \frac{\hat{a}}{\|\hat{a}\|_2} \quad , T_0 = \sqrt{n}\|\hat{a}\|_2 \quad , K_0 T_0^{-1} = \frac{e}{n\|\hat{a}\|_2}. \tag{4.37}$$

Se $L = L_2$ *rappresenta la matrice di approssimazione discreta di un operatore di derivata seconda,*

$$L_2 = \begin{pmatrix} -1 & 2 & -1 & & & 0 \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & -1 & 2 & -1 \end{pmatrix} \in M_{n-2,n}(\mathbb{R}) \tag{4.38}$$

allora

$$K_0 = \left(n^{-1/2}e, \quad \alpha_n f - \beta_n e \right), \tag{4.39}$$

$$H_0 = \left(\tfrac{\hat{a}}{\|\hat{a}\|_2}, \quad \tfrac{\bar{a}-\gamma\hat{a}}{\|\bar{a}-\gamma\hat{a}\|_2} \right), \tag{4.40}$$

$$T_0 = \begin{pmatrix} \sqrt{n}\|\hat{a}\|_2, & n\|\hat{a}\|_2(\alpha_n\gamma - \beta_n) \\ 0, & n\alpha_n\|\tilde{a} - \gamma\hat{a}\|_2 \end{pmatrix}, \tag{4.41}$$

$$K_0 {T_0}^{-1} = n^{-1} \left(\tfrac{e}{\|\hat{a}\|_2}, \quad \tfrac{f-\gamma e}{\|\tilde{a}-\gamma\hat{a}\|_2} \right), \tag{4.42}$$

dove

$$\alpha_n = \left(\frac{12}{(n+1)n(n-1)}\right)^{1/2}, \beta_n = \left(\frac{3(n+1)}{n(n-1)}\right)^{1/2}, \gamma = \frac{\hat{a}^T\tilde{a}}{\|\hat{a}\|_2}. \tag{4.43}$$

Osserviamo esplicitamente che valgono le stime:

$$K_2(L_1) \approx 0.64n \quad \text{e} \quad K_2(L_2) \approx 0.41n^2$$

che forniscono un legame tra condizionamento di L e ordine della matrice.

Esempio

Riprendiamo l' esempio di pag. 69 su cui vogliamo applicare la $gsvd$ e la $gsvd - stdform$.

Illustriamo nella seguente tabella la "capacità di ricostruzione, relativa alle matrici (A, L), degll' algoritmo $gsvd$ (risp. $gsvd - stdform$) mediante il calcolo della norma infinito delle matrici $(A - \hat{A})$ e $(L - \hat{L})$ (risp. $(A - \tilde{A})$ e $(L - \tilde{L})$) avendo indicato con $\hat{A} = \hat{U}\hat{\Sigma}\hat{W}^T$ e $\hat{L} = \hat{V}\hat{M}\hat{W}^T$ (risp. $\tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{W}^T$ e $\tilde{L} = \tilde{V}\tilde{M}\tilde{W}^T$) le matrici effettivamente calcolate dall' algoritmo.

n	$K_2(A)$	$K_2(L)$	$\|A - \hat{A}\|_\infty$	$\|L - \hat{L}\|_\infty$	$\|A - \tilde{A}\|_\infty$	$\|L - \tilde{L}\|_\infty$
4	$1.6\,10^{16}$	2.4	$3\,10^{-16}$	$1\,10^{-15}$	$1\,10^{-16}$	$1\,10^{-16}$
8	$8.8\,10^{27}$	5.0	$9\,10^{-16}$	$2\,10^{-15}$	$5\,10^{-16}$	$1\,10^{-15}$
12	$1.0\,10^{29}$	7.5	$1\,10^{-15}$	$3\,10^{-15}$	$3\,10^{-16}$	$2\,10^{-15}$
16	$1.0\,10^{30}$	10	$1\,10^{-15}$	$4\,10^{-15}$	$6\,10^{-16}$	$5\,10^{-15}$
20	$6.2\,10^{29}$	12	$3\,10^{-15}$	$5\,10^{-15}$	$1\,10^{-15}$	$5\,10^{-15}$
24	$2.0\,10^{31}$	15	$3\,10^{-15}$	$6\,10^{-15}$	$1\,10^{-15}$	$7\,10^{-15}$
28	$1.0\,10^{31}$	17	$3\,10^{-15}$	$8\,10^{-15}$	$4\,10^{-15}$	$8\,10^{-15}$
32	$1.6\,10^{31}$	20	$3\,10^{-15}$	$8\,10^{-15}$	$2\,10^{-15}$	$8\,10^{-15}$
36	$5.3\,10^{30}$	22	$3\,10^{-15}$	$8\,10^{-15}$	$2\,10^{-15}$	$8\,10^{-15}$
40	$2.0\,10^{31}$	25	$3\,10^{-15}$	$1\,10^{-14}$	$4\,10^{-15}$	$9\,10^{-15}$

Per quello che riguarda il costo computazionale dei due algoritmi, indicato nella tabella qui di seguito, possiamo mettere in evidenza che questo è inferiore nel caso dell' algoritmo $gsvd - stdform$ anche se l'ordine di grandezza è uguale a quello ottenuto implementando l' algoritmo $gsvd$. Praticamente uguale è invece la "capacità di ricostruzione delle matrici A e L, attraverso entrambi i

metodi, risultando sempre prossima alla precisione di macchina.

n	$gsvd$	$gsvd - stdform$
4	$2.0980\,10^3$	$1.3070\,10^3$
8	$1.5625\,10^4$	$1.0773\,10^4$
12	$5.2733\,10^4$	$3.8100\,10^4$
16	$1.2212\,10^5$	$9.0146\,10^4$
20	$2.3917\,10^5$	$1.8307\,10^5$
24	$4.1353\,10^5$	$3.1529\,10^5$
28	$6.7444\,10^5$	$5.0030\,10^5$
32	$9.9910\,10^5$	$7.6483\,10^5$
36	$1.4702\,10^6$	$1.0862\,10^6$
40	$1.9384\,10^6$	$1.5055\,10^6$

Questo esempio permette di verificare l' accuratezza degli algoritmi $gsvd$ e $gsvd - stdform$ e segnare un punto a favore della $GSVD$ come metodo di indagine per la regolarizzazione.

4.6 $GSVD$ Troncata: $TGSVD$

Un metodo alternativo alla regolarizzazione di Tikhonov standard

$$min\{\|Ax - b\|_2^2 + \lambda^2\|x\|_2^2\}$$

per la risoluzione del problema discreto $Ax = b$ è dato dalla SVD troncata ($TSVD$) in cui vengono tralasciati i valori singolari più piccoli di A: la matrice

$$A = \sum_{i=1}^{rank(A)} u_i\sigma_i v_i^T = U\Sigma V^T \in M_{m,n}(\mathbb{R})$$

viene pertanto sostituita da

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T = U\Sigma_k V^T \in M_{m,n}(\mathbb{R})$$

con $k = rank(A_k) < rank(A)$ e $\Sigma_k = diag(\sigma_1, ..., \sigma_k, 0, ..., 0)$.
In termini del problema ai minimi quadrati associato, questo corrisponde a approssimare la pseudosoluzione

$$x^\dagger = A^\dagger b = \sum_{i=1}^{rank(A)} \frac{u_i^T b}{\sigma_i} v_i$$

con

$$x_k^\dagger = A_k^\dagger b = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i \tag{4.44}$$

commettendo un errore $\|x^\dagger - x_k^\dagger\|_2^2 = \sum_{i=k+1}^{rank(A)} (\frac{u_i^T b}{\sigma_i})^2$ sulla soluzione $x^\dagger$ e $\|A - A_k\|_2 = \sigma_{k+1}$ su A: il vettore $x_k^\dagger$ è allora la pseudosoluzione del nuovo problema

$$A_k x = b \quad \forall k < rank(A).$$

Supposto che il numero di oscillazioni nei vettori singolari destri e sinistri, u_i e v_i, aumenti al crescere di i la soluzione $x_k^\dagger$ risulta più regolare di $x^\dagger$ avendo tralasciato i valori singolari $\{\sigma_i\}_{i=k+1}^{rank(A)}$ più piccoli e corrispondenti ai vettori $\{v_i\}_{i=k+1}^{rank(A)}$ più oscillanti.
Per motivare il ricorso alla $TSVD$ è necessario supporre la presenza di oscillazioni nei vettori singolari destri e sinistri di A che è stata dimostrata per le matrici totalmente positive [11,16] e che riccorre spesso come caratteristica dei problemi mal posti. Ci chiediamo ora se è possibile trovare un analoga proprietà per i vettori singolari generalizzati $\{x_i\}_{i=1}^n$.

Teorema 4.6.1 *Siano $v_k(A) \quad \forall i = 1, 2, ..., n$ i vettori singolari destri di A relativi agli autovalori $\sigma_k(A)$ della SVD di A e esprimiamo x_i in termini di questi vettori:*

$$x_i = \sum_{k=1}^{n} \xi_{k,i} v_k(A) \quad \forall i = 1, 2, ..., n.$$

Se $\sigma_i \neq 0$ allora

$$|\xi_{k,i}| = \begin{cases} \min\{\sigma_i/\sigma_k(A), {\Pi_p}^{-1}\} & se\ \sigma_k(A) \neq 0 \\ {\Pi_p}^{-1} & se\ \sigma_k(A) = 0 \end{cases}$$

con Π_p definita nel teorema 3.3.3 e σ_i in (3.22).

Dimostrazione.
Per il teorema 3.3.3
-${\Pi_p}^{-1} \geq \|X\|_2 \geq \|x_i\|_2 = \|\sum_{k=1}^n \xi_{k,i} v_k(A)\|_2 = \sqrt{\sum_{k=1}^n \xi_{k,i}^2} \geq |\xi_{k,i}|$
e inoltre
-$\sigma_i^2 = \|\sigma_i v_i\|_2^2 = \|Ax_i\|_2^2 = \|A(\sum_{j=1}^n \xi_{j,i} v_k(A))\|_2^2 = \|\sum_{j=1}^n \xi_{j,i}\sigma_k(A) u_k(A)\|_2^2 = \sum_{j=1}^n \xi_{j,i}^2 \sigma_k(A)^2 \Longrightarrow (se\ \ \sigma_k(A) \neq 0)\ \ |\xi_{k,i}| \leq \frac{\sigma_i^2 - \sum_{j\neq k} \xi_{j,i}^2 \sigma_k(A)^2}{\sigma_k(A)^2} \leq (\frac{\sigma_i}{\sigma_k(A)})^2.$ ∎

Questo teorema mostra che x_i è dominato da quei vettori $v_k(A)$ per cui $\sigma_i/\sigma_k(A) >$

1 ovvero $\sigma_k(A) < \sigma_i$. Pertanto x_1 è dominato da $v_n(A)$, x_2 è dominato da $v_{n-1}(A)$, $v_n(A)$ e così via: possiamo allora concludere che le oscillazioni nei vettori $v_i(A)$ si ripercuotono sui vettori x_i in ordine opposto ovvero le oscillazioni nei vettori generalizzati aumentano al diminuire di i.

Richiamate nel capitolo 3 alcune caratteristiche della $TSVD$, per la cui trattazione dettagliata si può fare riferimento a [6,11], e evidenziate le problematiche legate ai vettori $\{x_i\}_{i=1}^n$ vogliamo estendere tale decomposizione alla regolarizzazione generale.

Definizione 4.6.2 *Sia* $\Sigma_k^\dagger = diag(0, ..., 0, \sigma_{p-k+1}^{-1}, ..., \sigma_p^{-1})$ *e* $1 \leq k \leq p$.
Definiamo soluzione GSVD troncata (TGSVD) di

$$\min\{\|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2\}$$

il vettore x_k^* *ottenuto tralasciando le componenti di* $(X_p \Sigma_p^\dagger U^T b)$ *corrispondenti ai* $(p-k)$ *valori singolari* σ_i *più piccoli: cioè*

$$x_k^* = R_k b \tag{4.45}$$

con

$$R_k = X \begin{pmatrix} \Sigma_k^\dagger & 0 \\ 0 & I_0 \end{pmatrix} U^T = X_p \Sigma_k^\dagger U_p^T + X_0 U_0^T.$$

L' introduzione della $TGSVD$, dovuta ad Hansen [11], è basata sull' algoritmo di Elden attraverso cui è possibile evidenziare alcune analogie con la $TSVD$. Autonomamente dall' approccio di Hansen, che riprenderemo più avanti, diamo una caratterizzazione della $TGSVD$ a cui siamo arrivati studiando lo sviluppo della pseudosoluzione

$$x^\dagger = \sum_{i=1, \sigma_i \neq 0}^{p} \frac{u_i^T b}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T b) x_i \tag{4.46}$$

ottenuta dalla $GSVD$ di (A, L).
Nella sezione 3.4 la necessità di esprimere $x^\dagger$ usando il sistema singolare generalizzato $\{\gamma_i, u_i, x_i\}_i$, in luogo di quello standard $\{\sigma_i, u_i, v_i\}_i$, era imposta per fornire una rappresentazione esplicita dell' errore globale (3.53). Proprio attraverso l' espressione (4.46), che in 3.4 aveva carattere più teorico che numerico, possiamo definire la $TGSVD$: tralasciando infatti in (4.46) i primi $(p-k)$ valori generalizzati (più piccoli o nulli)

$$\sigma_1 \leq .... \leq \sigma_{p-k}$$

, a cui corrispondono i vettori $\{x_i\}_{i=1}^{p-k}$ più oscillanti, otteniamo

$$x_k^* = \sum_{i=p-k+1}^{p} \frac{u_i^T b}{\sigma_i} x_i + \sum_{i=p+1}^{n} (u_i^T b) x_i \tag{4.47}$$

che coincide con l' espressione (4.45).
La necessità di utilizzare la $GSVD$ al posto della SVD per il calcolo della pseudosoluzione di $Ax = b$ è dovuta al fatto che i vettori generalizzati destri x_i sono, in alcuni casi, più adatti dei vettori ordinari v_i come base di $x^\dagger$: questo accade ogni qualvolta che è necessario introdurre un metodo di regolarizzazione generale per assicurare una maggiore stabilità della soluzione. A conferma di questo, riprendendo l' esempio trattato nel capitolo 3, confrontiamo i grafici della soluzione $x_k^\dagger$ e x_k^*: da questi si osserva (figura 4.1) che per nessun valore di k si ottiene una soluzione $x_k^\dagger$ che approssima la soluzione esatta in modo valido essendo i vettori v_i (figura 3.1) asintoticamente nulli. Al contrario per opportuni valori di k (figura 4.2), il vettore x_k^* fornisce una buona approssimazione per $f = \{f(t_i)\}_{i=1}^{16}$ sul reticolo $\{t_i\}_{i=1}^{16}$.

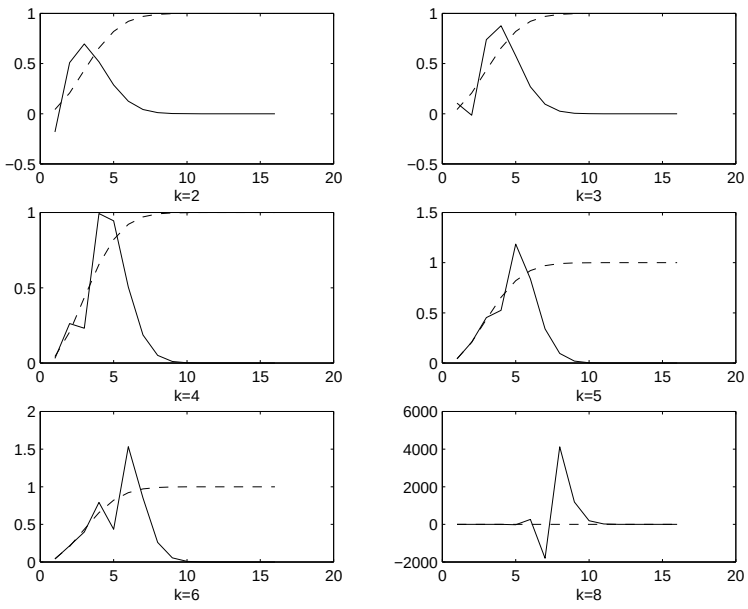

Figura 4.1. Grafico della soluzione $x_k^\dagger$ per alcuni k

Ad esempio:
per $k = 2$ si ha $\|Ax_k^* - b\|_2 = 2.4170\,10^{-3}$ e $\|f - x_k^*\|_\infty = 1.1690\,10^{-1}$,

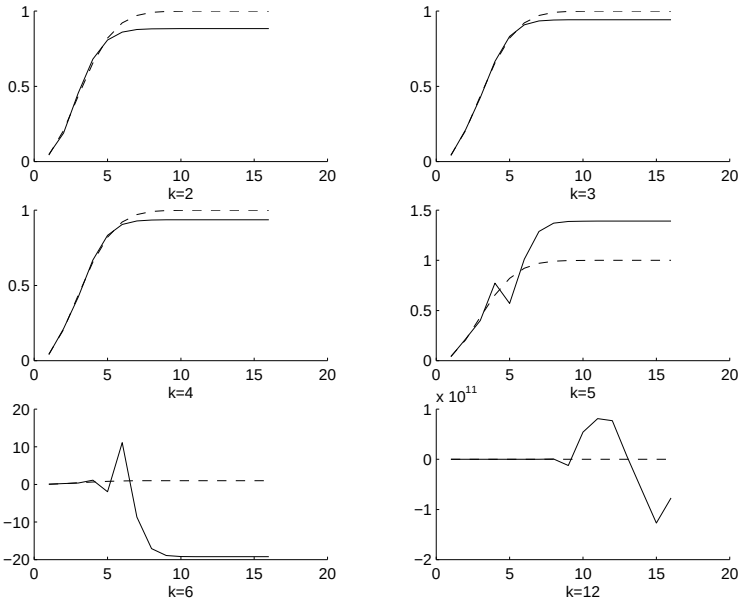

Figura 4.2. Grafico della soluzione x_k per alcuni k

per $k = 3$ si ha $\|Ax_k^* - b\|_2 = 2.5475\,10^{-3}$ e $\|f - x_k^*\|_\infty = 5.7721\,10^{-2}$,
per $k = 4$ si ha $\|Ax_k^* - b\|_2 = 2.5243\,10^{-4}$ e $\|f - x_k^*\|_\infty = 6.3814\,10^{-2}$.

Osserviamo esplicitamente che $x_k^\dagger$ e x_k^* non rappresentano lo stesso vettore (a meno che non individuino $x^\dagger$) in quanto appartengono a sottospazi diversi:

$$x_k^\dagger \in span\{v_1, ..., v_k\} \quad \text{e} \quad x_k^* \in span\{x_i\}_{i \in I \cup \{p+1,...,n\}}$$

con $I = \{i \in \{1, 2, ...p\} : \sigma_i \neq 0\}$.
Analizziamo ora la funzione di discrepanza ϵ_k al variare del parametro k. Da (4.47) e (3.21) si ha
$Ax_k^* = \sum_{i=p-k+1}^{p} \frac{u_i^T b}{\sigma_i} Ax_i + \sum_{i=p+1}^{n} (u_i^T b) Ax_i = \sum_{i=p-k+1}^{p} (u_i^T b) u_i + \sum_{i=p+1}^{n} (u_i^T b) u_i$
da cui, completando $\{u_i\}_{i=1}^n$ a base ortonormale di $\mathbb{R}^m$ (se $m > n$) mediante i vettori $\{u_i\}_{i=n+1}^m$, otteniamo

$$\epsilon_k^2 = \|Ax_k^* - b\|_2^2 = \sum_{i=1}^{p-k} (u_i^T b)^2 + \sum_{i=n+1}^{m} (u_i^T b)^2$$

che è decrescente rispetto a k. Nell' ipotesi di dati perturbati

$$b = \bar{b} + e$$

ci chiediamo come vada scelto il parametro di troncamento k affinchè la soluzione sia significativa per il problema $Ax = b$: infatti, come si osserva dalla figura 4.2, scelte sbagliate di k fornicono soluzioni x_k^* non buone. Supposto e un vettore *random* che appare come $white - noise$, intendento con questo che le componenti $u_i^T \bar{b}$ sono costanti e indipendenti da i, ricaviamo che:
-al diminuire di k, per (4.47), la soluzione x_k^* è caratterizzata dal segnale b e l' effetto del *noise* è poco evidente; a fronte di questo, per la monotonia ϵ_k aumenta la discrepanza.
-al crescere di k cresce l' influenza del noise e sulla soluzione x_k^* poichè aumenta il numero di addendi $\frac{u_i^T b}{\sigma_i}$ relativi ai σ_i più piccoli. Parallelamente si riscontra anche un maggior numero di oscillazioni della soluzione x_k^* e un aumento del termine di penalità

$$\Omega(x_k^*) = \|Lx_k^*\|_2^2 = \sum_{i=p-k+1}^{p} \frac{(u_i^T b)^2}{\gamma_i^2}$$

che ne controlla la regolarità. Il parametro di troncamento k deve pertanto rappresentare il giusto compromesso tra stabilità della soluzione x_k^*, compromessa dalla presenza in (4.47) di valori σ_i troppo piccoli, e buona approssimazione di $x^\dagger$. Forniamo allora una stima dell' errore di perturbazione [11] sulla soluzione troncata x_k^* del problema $Ax = b = \bar{b} + e$ supponendo di lasciare inalterata la matrice dei coefficienti.

Teorema 4.6.3 *Sia e l' errore di perturbazione del termine noto e x_k^* la soluzione di $Ax = b$ ottenuta attraverso la $TGSVD$. Allora la perturbazione relativa alla soluzione $\bar{x}_k^*$ del problema non perturbato soddisfa la relazione:*

$$\frac{\|x_k^* - \bar{x}_k^*\|_2}{\|\bar{x}_k^*\|_2} \leq \frac{\|A\|_2}{\sigma_{p-k+1}} \frac{\|e\|_2}{\|\bar{b}_k\|_2} \tag{4.48}$$

con $\bar{b}_k = A\bar{x}_k^*$.

Dimostrazione.
$x_k^* - \bar{x}_k^* = R_k b - R_k \bar{b} = R_k e \Longrightarrow \|x_k^* - \bar{x}_k^*\|_2 = \|R_k e\|_2 = \|X \begin{pmatrix} \Sigma_k^\dagger & 0 \\ 0 & I_0 \end{pmatrix} U^T e\|_2 \leq$
$\|X\|_2 \| \begin{pmatrix} \Sigma_k^\dagger & 0 \\ 0 & I_0 \end{pmatrix} \|_2 \|U^T e\|_2 =$
$\|X\|_2 \max_{i=p-k+1,\ldots,p}\{\frac{1}{\sigma_i}, 1\}\|e\|_2 = \|X\|_2 \frac{1}{\sigma_{p-k+1}}\|e\|_2.$
Poichè $\bar{b}_k = A\bar{x}_k^*$ si ha

$$\|\bar{b}_k\|_2 = \|A\bar{x}_k^*\|_2 \leq \|A\|_2 \|\bar{x}_k^*\|_2$$

e utilizzando la relazione precedentemente trovata otteniamo la tesi. ∎

Definizione 4.6.4 *Il numero di condizionamento [11] associato alla TGSVD è definito da*

$$K_k = \lim sup_{\|e\|_2 \to 0} \frac{\|x_k^* - \bar{x}_k^*\|_2}{\|\bar{x}_k^*\|_2} \tag{4.49}$$

e soddisfa la relazione

$$K_\lambda \leq \frac{K_2(X)}{\sigma_{p-k+1}}$$

dove $K_2(X)$ è il condizionamento della matrice X rispetto alla norma 2.

In (4.49) si osserva, oltre l' analogia con (3.57), che al diminuire di k la stima di K_k diventa via via migliore in quanto va aumentando il valore σ_{p-k+1}. Come già accennato l' approccio di Hansen [11] alla $TGSVD$ è basato sull' algoritmo di Elden: l' idea è quella di calcolare $\bar{x}_k^\dagger$, pseudosoluzione di $\bar{A}_k x = b$, e definire x_k^* mediante la relazione

$$x_k^* = L^\dagger \bar{x}_k^\dagger + K_0 T_0^{-1} H_0 (b - A L^\dagger \bar{x}_k^\dagger) \tag{4.50}$$

come si è fatto per la soluzione regolarizzata (4.8). L' unica cosa da verificare è che (4.47) coincide proprio con la soluzione definita in (4.50).

Teorema 4.6.5 *Sia*

$$\bar{A} = \bar{U}\bar{\Sigma}\bar{V}^T$$

la SVD di $\bar{A}$ e (3.21) la $GSVD$ di (A, L).
Allora

$$x_k^\dagger = \bar{A}_k^\dagger b$$

con $\bar{A}_k^\dagger = V diag(\gamma_i^{-1}, ..., \gamma_k^{-1}, 0, ..., 0)\bar{U}^T$ ovvero la soluzione ottenuta sostituendo $\bar{x}_k$ in (4.8) è proprio la $TGSVD$ definita in (4.45).

Dimostrazione. Per la dimostrazione si consulti [11]. ∎

Ritornando alla soluzione x_k^* concludiamo che tralasciando i contributi $(u_i^T b/\sigma_i)$, corrispondenti ai σ_i più piccoli, si ottiene semplicemente un altro modo di ottenere una soluzione regolare che possiede proprietà simili a x_λ. Naturalmente in x_k^* la dipendenza da L non è esplicita ma va letta, oltre che nei valori σ_i, nei

vettori $\{u_i, x_i\}_i$ che cambiano per diverse scelte di L. Dalle pagine precedenti dovrebbe essere chiaro che l' algoritmo di Elden ha un duplice ruolo; infatti, dal punto di vista numerico, fornisce nell' ipotesi (4.2), un algoritmo stabile e computazionalmente valido per la risoluzione di (4.1) e il calcolo della $GSVD$ di (A, L) e, dal punto di vista teorico, lega la regolarizzazione generale a quella standard.

All' inizio di questo lavoro la routine *gsvd* contenuta in [14] forniva buoni risultati solo nel caso di matrici A con condizionamento arbitrario e L ben condizionata: pertanto la routine $gsvd - stdform$ da noi implementata ne rappresentava una valida alternativa in quanto aveva, nell' ambito delle simulazioni numeriche da noi fatte, le stesse caratteristiche di stabilità e un costo computazionale dello stesso ordine.

Entrambe risultano tuttavia di scarsa applicabilità nel caso in cui sia A che L sono mal condizionate: le motivazioni di questa instabilità sono dovute a fattori molto differenti che ci proponiamo di analizzare qui di seguito.

Per quello che riguarda la routine $gsvd - stdform$ sottolineiamo subito che è applicabile soltanto se L ha rango pieno in quanto tale ipotesi è di fondamentale importanza per l' inversione di R_p e il calcolo di $L^\dagger$. Pertanto supposta valida l' ipotesi $rank(L) = p$, si deduce, in virtù della relazione $K_2(L) = K_2(R_p)$, che se L è mal condizionata l' inversione di R_p è numericamente instabile e costituisce una prima possibile causa di instabilità per l' intero algoritmo. In realtà questa possibile instabilità nel calcolo di R_p^{-1} si ripercuote in più punti dell' algoritmo come, ad esempio, nel calcolo di $Y_p = L^T V M_p^{-1}$ e $\bar{A} = H_q^T A K_p R_p^{-T}$. Osserviamo però che il mal condizionamento di L non ha effetti negativi sui valori singolari di $M_p = diag(\mu_1, ..., \mu_p)$ per cui vale

$$\mu_i = \frac{1}{\sqrt{1 + \sigma_{p-i+1}^2}} \approx 1 \quad se \quad \sigma_{p-i+1} \approx 0.$$

Il mal condizionamento di L preclude l' affidabilità dell' algoritmo $gsvd - stdform$ anche a causa di alcune problematiche legate alla stabilità di routine utilizzate dall' algoritmo di Elden. Infatti se L non è ben condizionata il ricorso alla fattorizzazione QR in (4.5) può generare matrici H e T anch' esse instabili e quindi alcune relazioni di ortogonalità usate nei lemmi 4.3.2 e 4.3.3 e nel teorema 4.3.6 potrebbero non essere soddisfatte numericamente.

Per quello che riguarda la routine *gsvd* in [15], basata su [24], ricordiamo che

questa calcola le matrici U, Σ, V, X, M legate a (A, L) dalle relazioni

$$\begin{cases} A = U\Sigma X^{-1} \\ L = VMX^{-1} \end{cases} . \tag{4.51}$$

La sua instabilità non è da ricercare in più punti dell' algoritmo ma è localizzata nel calcolo di X: infatti se A e L sono mal condizionate anche X sarà tale e quindi il ricorso al calcolo di X^{-1} fornisce una matrice fortemente instabile che compromette la possibiltà di ricostruire A e L in modo efficiente attraverso le relazioni (4.51). L' idea, che è alla base della routine del Matlab 5.2 è quella di modificare l' algoritmo [22] in maniera tale da non costruire X^{-1} ma X^T sostituendo quindi le relazioni (4.51) con

$$\begin{cases} A = U\Sigma X^T \\ L = VMX^T \end{cases} . \tag{4.52}$$

La routine ottenuta risulta così stabile anche per matrici A e L mal condizionate non richiedendo più l' inversione di X.
Va comunque osservato che nell' ambito della regolarizzazione generale dei problemi mal-posti la *gsvd* del Matlab non permette niente di più rispetto alla routine $gsvd - stdform$ o a quella contenuta in [15] in quanto tutte le relazioni dedotte nel capitolo 3 fanno riferimento alla decomposizione (4.51) e non a (4.52). Naturalmente anche se si prendessero le relazioni (4.52) come punto di partenza tutte le formule del capitolo 3 sarebbero ancora valide pur di indicare con x_j la j-esima colonna di X^{-T} anzichè di X e questo richiederebbe ugualmente l' inversione di X. La possibilità di invertire X in modo stabile è subordinata al suo buon condizionamento che è garantito, per la dimostrazione data nel capitolo 3, solo se L è ben condizionata.

Appendice

In questa appendice è riportato l' algoritmo $gsvd - stdform$ (cfr.[6]) trattato nel capitolo 4.

```
function[U,Sigma,Y,V,M]=gsvd_stdform(A,L)
[m,n1]=size(A);
[p,n2]=size(L);
if (n1==n2)&(m>=n1)&(n1>=n2)
   n=n1;
   %Fattorizzazione QR di L';
   [K,R]=qr(L');
   %Partizionamento di K=(K_p,K_o)
   K_p=K(:,1:p);
   K_o=K(:,p+1:n);
   %Calcolo la matrice R_p
   R_p=R(1:p,:);
   %Calcolo la fattorizzazione QR di AK_o
   [H,T]=qr(A*K_o);
   %Partizionamento di H=(H_o,H_q)
   H_q=H(:,(n-p)+1:m);
   H_o=H(:,1:n-p);
   %Partizionamento di T
   T_o=T(1:n-p,:);
   %Calcolo la matrice A_s
   A_s=(H_q)'*A*K_p*inv(R_p)';
   %Calcolo la SVD di A_s
   [U_s,C,W]=svd(A_s);
```

```
   %Calcolo U_p
   U_p=H_q*U_s*antid(m-(n-p));
   %Costruisco la matrice M=(M_p,0)
   M=zeros(p,n);
   for i=1:p
      M(i,i)=1/(pythag(C(p-i+1,p-i+1),1));
   end
   %Costruisco M_p
   M_p=M(:,1:p);
   %Costruisco la matrice Sigma
   Sigma=zeros(m,n);
   for i=1:p
      Sigma(i,i)=C(p-i+1,p-i+1)/(pythag(C(p-i+1,p-i+1),1));
   end
   for i=p+1:n
      Sigma(i,i)=1;
   end
   %Costruisco U=(U_p,U_o)
   U(:,1:p)=U_p(:,1:p);
   U(:,p+1:n)=H_o(:,1:n-p);
   %Calcolo V
   V=W*antid(p);
   Y_o=A'*H_o;
   Y_p=L'*V*inv(M_p);
  %Costruisco la matrice Y
  Y(:,1:p)=Y_p(:,1:p);
  Y(:,p+1:n)=Y_o(:,1:n-p);
  %Y=A'*U*Sigma+L'*V*M;
else
   error('Le dimensioni delle matrici non sono compatibili')
end
```

La routine precedente fa uso della funzione *antidiag*, qui di seguito riportata, che costruisce la matrice con il valore 1 sulla antidiagonale.

```
function[a]=antid(n)
a=zeros(n,n);
for i=1:n
   a(i,n-i+1)=1;
end
```

Bibliografia

[1] M. Bertero, P. Boccacci, *Intoduction to inverse problems in imaging*, Institute of Physics Publications (1998).

[2] M. Bertero, *Problemi lineari non ben posti e metodi di regolarizzazione*, C.N.R. (Pubblicazioni Dell' Istituto Di Analisi Globale E Applicazioni), Serie: Problemi non ben posti ed inversi n.4 (1982).

[3] D. Bini, *Metodi numerci per l' algebra linare*, Zanichelli (1988).

[4] A. Bjorck, *Numerical Methods for Least Square Problems*, SIAM, Philadelphia (1996).

[5] C. Estatico, *Gradiente coniugato e regolarizzazione di problemi mal posti* , C.N.R Quaderni Del Gruppo Nazionale Per L' Informatica Matematica (1996).

[6] L. Elden *Algorithms for regularization of ill-conditioned least-square problems* BIT (1977), 134-145.

[7] G. Golub, C. Van Loan *Matrix Computation* 3Ed, Johns Hopkins, Baltimore (1996).

[8] C.W. Groetsch *Elements of applicable functional analysis* M.Dekker, Inc (1980).

[9] M. Hanke, *Regularization with differential operators. An iterative approach*, J. Numer. Funct. Anal. Optim. 13 (1992), 523-540.

[10] M. Hanke, *Regularization of inverse problems*, Kluwer (1996).

[11] P.C. Hansen, *Regularization, GSVD and truncated GSVD*, BIT 29 (1989), 491-504.

[12] P.C. Hansen, *Truncated SVD solutions to discrete ill-posed problems with ill-determined numerical rank*, SIAM J. Sci. Stat. Comput. 11 (1990), 503-518.

[13] P.C. Hansen, *Relation between SVD and GSVD of discrete regularization problems in standard and general form*, Lin. Alg. Appl. 141 (1990), 165-176.

[14] P.C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Review 34 (1992), 561-580.

[15] P.C. Hansen, *Regularization Tools, A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems -Version 2.0 for Matlab 4.0*, Numerical Algorithms 6 (1994), 1-35.

[16] P.C. Hansen, *Regularization Tools, A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems -Version 3.0 for Matlab 5.2*,(1998). Netlib: netlib@research.att.com (file: numeralgo/na4).

[17] P.C. Hansen, T. Sekii, H. Shibahashi, *The modified truncated SVD method for regularization in general form*, SIAM J. Sci. Stat. Comput. 13 (1992), 1142-1150.

[18] J. Locker, P.M Prenter, *Regularization with differential operators I: General Theory*, J. Math. Anal. Appl. 74 (1980),504-529.

[19] J. Locker, P.M Prenter, *Regularization with differential operators II: Weak least square finite element solution to first kind integral equation*, SIAM J. Num. Anal. 17 (1980), 247-267.

[20] *Matlab Reference Guide*, The MathWorks, Mass. (1998).

[21] V.A. Morozov, *Methods for Solving Incorrectely Posed Problems*, Springer Verlag, New York (1984).

[22] W. Rudin, *Functional Analysis*, MC.Graw-Hill (1991).

[23] A.N. Tikhonov, A.V. Goncharsky *Solutions of Ill-Posed Problems in the Natural Sciences*, MIR Publishers, Moscow, 1987.

[24] C. Van Loan, *Computing the CS and the Generalized Singular Value Decomposition*, Numer. Math.46 (1985), 479-491.

[25] J.M. Varah, *Pitfalls in the numerical solution of linear ill posed problems*, SIAM J. Sci. Stat. Comput. 4 (1983),164-176.

[26] G. Wahba, *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mahematics, Vol.59, SIAM Philadelphia, (1990).