CONSIGLIO NAZIONALE DELLE RICERCHE

QUADERNI DEL GRUPPO NAZIONALE PER L'INFORMATICA MATEMATICA

CLAUDIO ESTATICO

GRADIENTE CONIUGATO
E
REGOLARIZZAZIONE DI PROBLEMI MAL POSTI

FIRENZE , 1996

Indice

Introduzione

La ricerca matematica è spinta in primo luogo dalla necessità di risolvere problemi concreti. Generalmente l'utilizzo di nuovi strumenti matematici non viene immediatamente supportato da un preciso studio teorico, ma spesso l'applicazione di nuove metodologie precede di parecchi anni la loro formulazione teorica completa.
In questo lavoro viene analizzato l'algoritmo del gradiente coniugato per il trattamento di problemi mal posti, che ha avuto uno sviluppo di questo tipo.
In prima analisi un problema matematico viene detto mal posto quando non dipende con continuità dai dati, ossia quando piccole perturbazioni del dato in ingresso danno luogo a soluzioni assai diverse.
Appare evidente che un problema di questo tipo comporta nella pratica grandi difficoltà di gestione, poichè in natura i dati a disposizione sono sempre conosciuti in maniera approssimata entro tolleranze finite.
Il trattamento di problemi mal posti si realizza per mezzo di metodologie che hanno lo scopo di rendere il problema poco sensibile a perturbazioni sui dati. Queste metodologie vengono dette "regolarizzanti".
Scopo del lavoro risulta proprio l'analisi di recenti studi che garantiscono come l'algoritmo del gradiente coniugato sia un metodo di regolarizzazione. In accordo con quanto detto, si sottolinea che il gradiente coniugato veniva già utilizzato numericamente come metodo per il trattamento di problemi mal posti, sebbene non fosse stata ancora dimostrata la sua appartenenza alla classe degli algoritmi regolarizzanti.

La monografia è articolata in modo da proporre i concetti fondamentali riguardanti la regolarizzazione e quelli più approfonditi sul gradiente coniugato.
Il contesto in cui si sviluppa lo studio è quello degli spazi di Hilbert. Più precisamente si considera una generica equazione lineare

$$Tx = y$$

dove $T : X \longrightarrow Y$ è un operatore lineare limitato tra gli spazi X, Y di Hilbert, $y \in Y$ è il dato del problema e $x \in X$ la soluzione da determinare.
Poichè come equazioni lineari mal poste si intendono anche equazioni in cui la soluzione non esiste o non è unica, la risoluzione è ricondotta alla determinazione dell'inversa generalizzata, ossia della soluzione ai minimi quadrati di norma minima.

Il primo capitolo ha lo scopo di introdurre la nozione di inversa generalizzata. L'esposizione affianca all'usuale definizione basata sull'introduzione dell'insieme delle soluzioni ai minimi quadrati, una diversa trattazione, che viene detta "operatoriale", basata su proprietà di restrizione e estensione dell'operatore T a opportuni sottospazi.
In particolare si sostituisce l'operatore T non invertibile, con un operatore $\tilde{T}$ invertibile tale che $\tilde{T}|_{N(T)^\perp} = T|_{N(T)^\perp}$ dove $N(T)$ rappresenta il nucleo dell'operatore T. L'invertibilità dell'operatore $\tilde{T}$ consente di determinare l'elemento $\tilde{x} = \tilde{T}^{-1}y \in Y$, la

cui proiezione su $N(T)^{\perp}$ corrisponde alla soluzione ai minimi quadrati di norma minima. In questo modo si ha una visione differente del concetto di inversa generalizzata che, oltre a chiarirne il significato, rappresenta un primo approccio alle difficoltà create da operatori il cui insieme immagine, o range, non sia chiuso.

Nel secondo capitolo vengono trattati i problemi mal posti e i metodi di regolarizzazione. Le nozioni acquisite nel capitolo precedente trovano adesso una naturale collocazione poichè le equazioni lineari mal poste vengono risolte proprio con la determinazione di soluzioni che approssimano la soluzione generalizzata.
Come già accennato, il capitolo ha lo scopo di fornire una breve introduzione all'argomento, che risulta vastissimo e in continuo sviluppo.
Nella prima parte del capitolo l'analisi è indirizzata all'algoritmo regolarizzante di A. Tikhonov . Questo particolare metodo consente di introdurre in maniera semplice il cosiddetto criterio della discrepanza, di fondamentale importanza per il capitolo successivo; inoltre consente di verificare la stretta relazione tra regolarizzazione e filtraggio in frequenza.
Nella parte finale del capitolo viene brevemente esaminato l'algoritmo regolarizzante di Landweber-Fridman. Nello svilupparsi del percorso che conduce alla regolarizzazione mediante gradiente coniugato, il metodo di Landweber-Fridman asssume un'importanza particolare poichè, analogamente al gradiente coniugato, è un metodo iterativo, ossia un metodo che determina in maniera sequenziale approssimazioni successive della soluzione, ognuna calcolata per mezzo delle precedenti.
Il secondo capitolo si chiude con una panoramica su alcuni metodi numerici di regolarizzazione, il cui scopo è dimostrare in che modo le proprietà individuate analiticamente si caratterizzano dal punto di vista numerico.
È importante sottolineare che, come verrà illustrato, l'equazione lineare $Tx = y$ conduce ad un problema mal posto ogniqualvolta l'operatore T non è dotato di range chiuso. Generalmente, a meno di precisazioni particolari, gli operatori considerati nel secondo capitolo soddisfano esclusivamente a questa ipotesi. Le dimostrazioni riportate fanno spesso uso della risoluzione spettrale per operatori autoaggiunti, argomento che per la sua importanza viene trattato in un'apposita appendice.

L'algoritmo del gradiente coniugato viene introdotto nel terzo capitolo.
Il gradiente coniugato è un metodo per la minimizzazione di funzioni; l'introduzione del funzionale $f : X \longrightarrow \mathbb{R}$ definito come

$$f(x) = ||Tx - y||_2^2$$

consente però di utilizzare l'algoritmo come metodo per la risoluzione ai minimi quadrati dell'equazione $Tx = y$.
La prima analisi, di tipo geometrico, viene fatta in uno spazio euclideo n-dimensionale. In questo contesto, che risulta il più semplice, si mette in luce il funzionamento dell'algoritmo. L'algoritmo viene successivamente generalizzato alla minimizzazione in spazi di Hilbert e viene confrontato con il metodo della massima discesa e con il metodo delle direzioni coniugate.
Lo studio procede con l'analisi dei teoremi di convergenza formulati negli anni 70 da Kammerer e Nashed (cfr. [21] , [22]) , i quali asseriscono che il metodo delle direzioni coniugate converge alla soluzione generalizzata.
Verrà innanzitutto dimostrata la convergenza del metodo nel caso in cui il funzionale sia definito per mezzo di operatori limitati invertibili e per mezzo di operatori limitati non invertibili con range chiuso. Successivamente verrà dimostrata la convergenza per operatori limitati con range non chiuso. In quest'ultimo caso le difficoltà nascono dal fatto che l'operatore T ristretto al sottospazio $N(T)^{\perp}$ possiede inversa non limitata e vengono superate

con l'introduzione del concetto di operatore aggiunto di operatore lineare non limitato.

La convergenza dimostrata per operatori con range non chiuso ha spinto diversi ricercatori a considerare il metodo come possibile algoritmo regolarizzante.
Nel 1986 A. Nemirovskii ha dimostrato che l'algoritmo è un metodo regolarizzante per una particolare classe di solusioni (i.e. $x \in \cup_{\mu>0} R(T^*T)^{\mu}$). Questo risultato, sebbene di notevole rilevanza, non permetteva però di concludere che l'algoritmo fosse un metodo di regolarizzazione, ossia avesse proprietà regolarizzanti che garantissero la convergenza in presenza di solusioni appartenenti a tutto l'insieme $N(T)^{\perp}$.
Nel 1990 R. Plato pubblica un articolo in cui, in maniera indiretta, dimostra che l'algoritmo del gradiente coniugato è effettivamente un algoritmo di regolarizzazione e che quindi consente la risolusione di problemi mal posti in presenza di qualsiasi solusione appartenente a $N(T)^{\perp}$ (cfr. [38]) .
La parte principale di questa monografia è l'analisi dei risultati di Nemirovskii e Plato.
La dimostrazione del teorema di Nemirovskii seguirà un percorso proposto nel 1994 da M. Hanke (cfr. [17]) . Questa dimostrazione fa uso di strumenti di analisi funzionale che possono essere ricondotti in primo luogo alla risolusione spettrale degli operatori autoaggiunti T^*T e TT^* . Il teorema di Plato, che come già detto risolve il problema dell'appartenenza del gradiente coniugato alla classe degli algoritmi regolarizzanti, viene invece illustrato nella formulazione originale dell'autore.

Una caratteristica di questo lavoro è di fornire un'analisi su "regolarizzazione e gradiente coniugato" breve ma sufficientemente completa. In quest'ottica parecchi teoremi minori, che nelle varie pubblicazioni vengono solitamente solo citati, sono stati verificati e dimostrati, mentre altri sono stati enunciati e dimostrati con formalismi differenti da quelli adottati dai rispettivi autori. Utilizzando le stesse definizioni nei diversi capitoli, l'esposizione è risultata così più omogenea.
Come già accennato sono state sviluppate inoltre appendici matematiche, le quali forniscono le necessarie nozioni di base sugli strumenti di analisi funzionale più sofisticati (in particolare la risolusione spettrale di operatori autoaggiunti non necessariamente compatti e la derivata di Fréchet di funzionali in spazi di Hilbert).

È utile precisare che l'algoritmo del gradiente coniugato è oggi applicato alla elaborazione di immagini. In questo campo il metodo, insieme a tecniche numeriche di precondizionamento, ha condotto a buoni risultati (cfr. [7] , [33]).

Un doveroso ringraziamento va rivolto alla Prof.ssa Paola Brianzi per i suggerimenti e le proposte di chiarimento di alcuni punti della trattazione.

Capitolo 1

Inversa generalizzata di operatori lineari

1.1 Introduzione

Siano X e Y spazi di Hilbert e $T : X \longrightarrow Y$ un operatore lineare limitato.
Indichiamo con $D(T)$, $N(T)$, $R(T)$ rispettivamente il dominio , lo spazio nullo e il range dell'operatore T .
Si consideri l'equazione lineare

$$Tx = y \tag{1.1}$$

dove $y \in Y$ è il dato assegnato.
Evidentemente se l'operatore T è invertibile, l'equazione (1.1) ha sempre una e una sola soluzione.
Si supponga però che $R(T) \neq Y$ e $y \notin R(T)$; in tal caso l'equazione non ammette soluzione, infatti $Tx \neq y \quad \forall x \in X$. Inoltre se l'operatore non è iniettivo ($\Leftrightarrow N(T) \neq 0$) la soluzione anche nel caso in cui esista, può non essere unica.
Queste difficoltà possono essere superate con l'introduzione della inversa generalizzata; essa consente di estendere il concetto di "unica soluzione".
Nella maggior parte della letteratura, la soluzione generalizzata viene introdotta e definita come l'elemento di norma minima nell'insieme delle soluzioni ai minimi quadrati. Questa definizione, detta variazionale, verrà presentata nel prossimo paragrafo.
Un approccio differente, che poggia sulla determinazione di un operatore invertibile $\tilde{T} : X \longrightarrow Y$ tale che $\tilde{T}|_{N(T)^\perp} = T|_{N(T)^\perp}$, conduce nel paragrafo successivo a una nuova definizione di inversa generalizzata che chiameremo operatoriale. Verrà dimostrata l'equivalenza delle due definizioni prima nel caso in cui $R(T)$ è chiuso e successivamente nel caso in cui $R(T)$ è arbitrario, ossia non necessariamente chiuso.
Nel quarto paragrafo la definizione operatoriale consentirà di verificare in maniera semplice diverse proprietà dell'inversa generalizzata; sulla base di queste proprietà verranno poi enunciate interessanti definizioni alternative.
Il capitolo si chiude con una breve analisi della inversa generalizzata per operatori compatti e con la determinazione dell'inversa generalizzata per matrici; quest'ultima risulta la versione numerica della definizione operatoriale per operatori dotati di range chiuso.

1.2 Pseudosoluzione e soluzione generalizzata

Indichiamo con P_R l'operatore di proiezione ortogonale su $\overline{R(T)}$.
Se l'operatore ha range chiuso l'equazione $Tx = P_R y$ ammette sempre un insieme di soluzioni; infatti, poichè $P_R y \in \overline{R(T)} = R(T)$, allora $\{x \in X : Tx = P_R y\} \neq \emptyset$.
Inoltre, osservando che $P_R y$ è l'elemento in $R(T)$ piu vicino a $y \in Y$, tali soluzioni sembrano essere le "migliori" ottenibili, ossia quelle la cui immagine dista meno possibile dal dato.
Generalizzando questa osservazione ad operatori a range arbitrario, introduciamo il teorema seguente che permette di definire la pseudosoluzione.
In seguito indicheremo con $\mathcal{L}(X,Y)$ l'insieme degli operatori lineari da X in Y, e con $\mathcal{B}(X,Y)$ l'insieme degli operatori lineari e limitati; l'aggiunto di un operatore $B \in \mathcal{B}(X,Y)$ verrà indicato con B^*, la chiusura di un insieme S verrà indicata con $\overline{S}$ e lo spazio ortogonale con $S^\perp$.

Teorema 1.1 *Consideriamo $T \in \mathcal{B}(X,Y)$, $u \in X$ e P_R l'operatore di proiezione ortogonale di Y su $\overline{R(T)}$.*
Le seguenti condizioni sono equivalenti:

(i) $Tu = P_R y$

(ii) $\|Tu - y\| \leq \|Tx - y\| \qquad \forall x \in X$

(iii) $T^*Tu = T^*y$

Dimostrazione.

$(i) \Rightarrow (ii)$
Si osservi innanzitutto che $y \in R(T) \oplus R(T)^\perp$ poichè, dall'equazione (i), $P_R y \in R(T)$.
Supposto $y = y_1 + y_2$ con $y_1 \in R(T)$, $y_2 \in R(T)^\perp$, si ha
$(P_R y - y) = (y_1 - (y_1 + y_2)) = y_2 \in R(T)^\perp$.
Sia $x \in X$, allora

$$\begin{aligned} \|Tx - y\|^2 &= \|Tx - P_R y\|^2 + \|P_R y - y\|^2 \\ &= \|Tx - P_R y\|^2 + \|Tu - y\|^2 \geq \|Tu - y\|^2 \end{aligned}$$

$(ii) \Rightarrow (iii)$
Poichè $\|Tu - y\| \leq \|Tx - y\| \qquad \forall x \in X$
possiamo considerare Tu come quell'elemento $\hat{w} \in R(T)$ t.c.

$$\|\hat{w} - y\| \leq \|w - y\| \ \ \forall w \in R(T)$$

Per il teorema della proiezione si ha che $\hat{w} - y \in R(T)^\perp = N(T^*)$, quindi
$T^*(\hat{w} - y) = 0 \Leftrightarrow T^*\hat{w} = T^*y$
cioè

$$T^*Tu = T^*y$$

$(iii) \Rightarrow (i)$
$Tu - y \in N(T^*) = R(T)^\perp$ quindi

$$P_R(Tu - y) = 0 \quad \Leftrightarrow \quad P_R Tu = P_R y \quad \Leftrightarrow \quad Tu = P_R y$$

Definizione 1.1 *Ogni elemento $u \in X$ che verifica le condizioni $(i) - (iii)$ del teorema 1.1 è detto* **pseudosoluzione** *o* **soluzione nel senso dei minimi quadrati.**

È facile osservare da (ii) che l'immagine di ogni pseudosoluzione approssima al meglio il dato dell'equazione (1.1), infatti minimizza la distanza $\|Tx - y\|$ su tutto lo spazio X.
Il teorema 1.1 considera un generico operatore $T \in \mathcal{B}(X,Y)$ a range arbitrario. Se $R(T)$ non è chiuso la proiezione $P_R y$ può non appartenere a $R(T)$. In tal caso, osservando (i), risulta evidente che non esistono pseudosolusioni del problema (1.1).

Proposizione 1.1 *Il problema* (1.1) *ammette pseudosoluzioni se e solo se* $y \in R(T) \oplus R(T)^\perp$.

Dimostrazione. Se $y \in R(T) \oplus R(T)^\perp$ allora $y = y' + y^\perp$, con $y' \in R(T)$ e $y^\perp \in R(T)^\perp$. L'insieme S formato dalle pseudosolusioni del problema (1.1) è la controimmagine secondo T dell'elemento y' cioè $S = \{u \in X, Tu = y'\} \neq \emptyset$.
Viceversa se $y \in \overline{R(T)} \backslash R(T)$, allora $P_R y \notin R(T)$ e quindi l'equazione (i) del teorema 1.1 non ha solusioni. ∎

Si noti che l'insieme $R(T) \oplus R(T)^\perp$ è denso in Y e coincide con esso se $R(T)$ è chiuso. Quindi se l'operatore T possiede range chiuso la pseudosolusione esiste sempre. Il passo sucessivo nella definisione di inversa generalizzata è la scelta di un elemento particolare nell' insieme delle pseudosolusioni.

Proposizione 1.2 *Se $y \in R(T) \oplus R(T)^\perp$, l'insieme $S = \{x \in X,\ Tx = P_R y\}$ delle pseudosoluzioni è non vuoto e convesso.*

Dimostrazione. Se $y \in R(T) \oplus R(T)^\perp$ allora $P_R y \in R(T)$, quindi, come già osservato nella proposisione precedente, l'insieme S è non vuoto, poichè esiste necessariamente almeno un $x \in X$ tale che $Tx = P_R y$.
Se T è iniettivo, S è costituito da un solo elemento. In questo caso generalizzando la definisione di insieme convesso, consideriamo verificata la tesi.
Supponiamo allora che S contenga almeno 2 elementi, $x', x'' \in S$.
Dimostriamo che $x''' = \alpha x' + (1-\alpha)x'' \in S$ con $\alpha \in (0,1)$.

$$\begin{aligned} \|Tx''' - y\| &= \|T(\alpha x' + (1-\alpha)x'') - y\| = \|\alpha Tx' + (1-\alpha)Tx'' - y\| \\ &= \|\alpha T(x' - x'') + Tx'' - y\| = \|Tx'' - y\| \end{aligned}$$

Pr l'ultima uguaglianza si osservi che, da $T^*Tx' = T^*y$ e $T^*Tx'' = T^*y$, si ha $T^*T(x' - x'') = 0$ quindi

$$T(x' - x'') \in N(T^*) = R(T)^\perp \ \Rightarrow \ T(x' - x'') = 0$$

Considerando nuovamente (ii) del teorema (1.1) si conclude che $x''' \in S$. ∎

Siccome l'insieme S delle pseudosolusioni è chiuso (infatti è la controimmagine di un elemento rispetto a un operatore continuo) e convesso, esso ammette un elemento di norma

minima. In accordo con quanto detto nell' introduzione, definiamo l'inversa generalizzata come la pseudosoluzione di norma minima.

Definizione 1.2 VARIAZIONALE
*L'***inversa generalizzata** $T^\dagger$ *di* T *è una applicazione* $T^\dagger : D(T^\dagger) = R(T) \oplus R(T)^\perp \longrightarrow X$ *definita come*

$$T^\dagger y = x^\dagger$$

dove

(i) $x^\dagger \in S = \{ u \in X \ : \ \|Tu - y\| \leq \|Tx - y\| \qquad \forall x \in X \} \subset X$

(ii) $\|x^\dagger\| \leq \|u\| \qquad \forall u \in S.$

La pseudosoluzione $x^\dagger$ *viene detta* **soluzione generalizzata** *del problema* (1.1).

Si noti che, se l'operatore è invertibile, $S = \{y^{-1}\}$, $x^\dagger = y^{-1}$, $T^\dagger = T^{-1}$.

1.3 Definizione operatoriale di inversa generalizzata

La definizione di inversa generalizzata adottata nel precedente paragrafo, che poggia sul teorema 1.1, viene detta variazionale.
Al fine di analizzarne le proprietà, introduciamo lo stesso concetto di inversa generalizzata in un altro modo.

Consideriamo innanzitutto il caso in cui l'operatore $T \in \mathcal{B}(X,Y)$ ha range chiuso. Quello che ci apprestiamo a fare è definire condizioni necessarie e sufficienti affinchè tale operatore possa essere scritto come $T = P_R \tilde{T}$ con P_R operatore di proiezione su $R(T)$ e $\tilde{T} \in \mathcal{B}(X,Y)$ invertibile. Definiremo così l'inversa generalizzata come $T^\dagger = P_{N(T)^\perp} \tilde{T}^{-1}$. Vediamo la cosa in dettaglio.

Teorema 1.2 *Consideriamo* $T \in \mathcal{B}(X,Y)$. *Supponiamo che*

$$T = P\tilde{T} \tag{1.2}$$

dove P *è un operatore di proiezione,* $\tilde{T} \in \mathcal{B}(X,Y)$ *invertibile con inversa anch'essa limitata. Allora :*

(i) $R(T)$ *è chiuso*

(ii) P *è la proiezione su* $R(T)$, *ossia* $P = P_R$

(iii) $\{y \in Y \ : \ y = \tilde{T}x \ , \ x \in N(T)\} = R(T)^\perp$
cioè $N(T)$ *è in corrispondenza biunivoca, rispetto a* $\tilde{T}$ *, con* $R(T)^\perp$

Viceversa se $R(T)$ *è chiuso e* $dim(\ N(T)\ ) = dim(\ R(T)^\perp\ )$ *, allora* T *ammette la rappresentazione* (1.2). *L'operatore* $\tilde{T}$ *può essere scelto in modo tale che* $\|T\| = \|\tilde{T}\|$

Il teorema precedente, con dimostrazione, enunciato nel caso più generale in cui X, Y sono spazi di Banach, si trova in [4].
L'operatore $\tilde{T}$ appare qui come una sorta di estensione invertibile dell'operatore T. Si osservi infatti che :

$$\begin{aligned} \tilde{T}x &\neq 0 \qquad && \forall\, x \in N(T)\backslash\{0\} \\ \tilde{T}x &= Tx \qquad && \forall\, x \in N(T)^\perp \end{aligned}$$

Si osservi che se vale le rappresentazione (1.2), X e Y hanno necessariamente la stessa dimensione ($\tilde{T}$ infatti è invertibile) . Il teorema può comunque essere generalizzato al caso in cui $dim(X) \neq dim(Y)$; si trattera' di immergere l'operatore T in modo opportuno, a seconda che $dim(X) > dim(Y)$ o $dim(X) < dim(Y)$. Il prossimo teorema asserisce che l'estensione di cui abbiamo bisogno esiste sempre.

Teorema 1.3 .
• *Sia* $T' \in \mathcal{B}(X', Y)$, $R(T')$ *chiuso.*
Se $dim(X') < dim(Y)$ *allora esiste uno spazio di Hilbert* X, $X \supset X'$
e un'operatore $T \in \mathcal{B}(X,Y)$ *t.c.*

(i) $Tx = T'x \quad \forall x \in X'$
$Tx = 0 \qquad \forall x \in X \ominus X'$

(ii) $R(T') = R(T)$

(iii) T *ammette rappresentazione* (1.2)

• *Sia* $T' \in \mathcal{B}(X, Y')$, $R(T')$ *chiuso.*
Se $dim(X) > dim(Y')$ *allora esiste uno spazio di Hilbert* Y, $Y \supset Y'$
e un'operatore $T \in \mathcal{B}(X,Y)$ *t.c.*

(i') $Tx = T'x \quad \forall x \in X$

(ii') $R(T') = R(T)$

(iii') T *ammette rappresentazione* (1.2)

Per la dimostrazione si consulti [4].

Definizione 1.3 OPERATORIALE
Definiamo **inversa generalizzata** *per l'operatore* $T \in \mathcal{B}(X,Y)$ *dotato di range chiuso, e la indichiamo con* $T^\dagger$ *, l'applicazione lineare* $T^\dagger : Y \longrightarrow X$ *definita come*

$$T^\dagger = P_{N(T)^\perp} \tilde{T}^{-1} \tag{1.3}$$

Riassumendo , l'inversa generalizzata $T^\dagger$ viene determinata nel modo seguente.
L'operatore T , con range chiuso, viene "esteso" con invertibilità a $\tilde{T}$ ponendo in corrispondenza biunivoca $N(T) \subset X$ con $R(T)^\perp \subset Y$. Invertendo $\tilde{T}$ e considerando come immagini solo le componenti in $N(T)^\perp$, le uniche significative rispetto a T , si ottiene $T^\dagger$.

$$\begin{aligned} T \ : \ & X \longrightarrow Y \\ & N(T)^\perp \longleftarrow R(T) \\ & N(T) \longrightarrow 0 \end{aligned} \tag{1.4}$$

$$\tilde{T} \;:\; X \longrightarrow Y \qquad (1.5)$$
$$N(T)^{\perp} \longleftrightarrow R(T)$$
$$N(T) \longleftrightarrow R(T)^{\perp}$$

$$T^{\dagger} \;:\; Y \longrightarrow X \qquad (1.6)$$
$$R(T) \longleftrightarrow N(T)^{\perp}$$
$$R(T)^{\perp} \longrightarrow 0$$

Si osservi che la trattazione adottata in questo paragrafo è indipedente da quella del paragrafo precedente, ossia la definizione 1.3 non fa uso di strumenti definiti nel paragrafo 1.2 . Occorre comunque dimostrare che le due definizioni sono equivalenti e non creano alcuna inconsistenza. Questo problema è risolto dal seguente teorema.

Teorema 1.4 *Se T ammette rappresentazione (1.2), allora $\forall y \in Y$ considerando la definizione 1.3 , si ha*

(a) $\inf_{x\in X} \|Tx - y\| = \|Tx_0 - y\|$ *con* $x_0 = \tilde{T}^{-1}y$
L'estremo inferiore è raggiunto da tutti e soli gli $x \in \Delta_0 = \{x \,:\, Tx = Tx_0\} \subset X$

(b) *(i)* $\inf_{x\in\Delta_0} \|x\| = \|x^{\dagger}\|$ *, dove* $x^{\dagger} = P_{N(T)^{\perp}}\tilde{T}^{-1}y = T^{\dagger}y$
(ii) $\|x\| > \|x^{\dagger}\| \qquad \forall\, x \in \Delta_0, \quad x \neq x^{\dagger}$

(c) L'operatore $T^{\dagger}$ che soddisfa le proprietà $(i)-(ii)$ è unico.

Dimostrazione.
Si osservi che, dal teorema 1.2 (i) , $R(T)$ risulta necessariamente chiuso.
Osservando che $Tx_0 - y$ e $T(x - x_0)$ sono ortogonali, infatti $T(x - x_0) \in R(T)$ e $Tx_0 - y = T\tilde{T}^{-1}y - y = P_R\tilde{T}\tilde{T}^{-1}y - y = (P_R - I)y = P_{R(T)^{\perp}}y \in R(T)^{\perp}$,
si ha

$$\|Tx - y\|^2 = \|Tx_0 - y\|^2 + \|T(x - x_0)\|^2$$

Quindi l'estremo inferiore in (a) assume il valore $\|Tx_0 - y\|$, che si ottiene se e solo se $Tx = Tx_0$. (a) risulta così dimostrato .
Considerando lo spazio quoziente $X\backslash N$, che si ottiene quozientando rispetto alla relazione di equivalenza

$$x' \sim x'' \;\Leftrightarrow\; x' - x'' \in N(T) \;\Leftrightarrow\; T(x' - x'') = 0$$

l'insieme Δ_0 può essere visto come la classe d'equivalenza contenente x_0.
Siccome $TP_{N(T)^{\perp}} = T$ (vedi (1.4)) , allora $x^{\dagger} \in [\Delta_0]$, infatti dalla definizione 1.3
$Tx^{\dagger} = TP_{N(T)^{\perp}}\tilde{T}^{-1}y = T\tilde{T}^{-1}y = Tx_0$. L'elemento $x^{\dagger}$ può essere preso come rappresentante della classe $[\Delta_0]$; quindi
$\forall x \in \Delta_0 \quad x = x^{\dagger} + z$, $z \in N(T)$; in più, per come è stato definito, $x^{\dagger} \in N(T)^{\perp}$ quindi

$$\|x\|^2 = \|x^{\dagger}\|^2 + \|z\|^2 \quad \forall x \in [\Delta_0]$$

Si ottiene $\|x\| > \|x^{\dagger}\| \quad \forall\, x \neq x^{\dagger}$, che dimostra (b).
Per dimostrare l'unicità dell'operatore $T^{\dagger}$, consideriamo un operatore lineare T' che soddisfi le proprietà (b). Evidentemenente $T'y = x^{\dagger}$, poichè $x^{\dagger}$ è l'unico elemento che ha la proprietà di minimo $(b)-(ii)$.
Allora

$$T^{\dagger}y = T'y \quad \forall y \in Y$$

da cui $(T^{\dagger} - T')y = 0 \qquad \forall\, y \in Y$, cioè $T^{\dagger} = T'$ che conclude la dimostrazione. ∎

Dal teorema 1.4 (a) si osserva immediatamente che l'insieme $\Delta_0 \subset X$, i cui elementi minimizzano la distanza $\|Tx - y\|$, coincide con l'insieme delle soluzioni ai minimi quadrati del paragrafo 1.1 . Come dimostrato al punto (b), l'inversa generalizzata 1.3 è l'unico elemento di norma minina in Δ_0 . L'equivalenza tra le due defizioni (variazionale e operatoriale) nel caso di operatori con range chiuso è così dimostrata.

Lo stesso percorso che conduce alla definizione 1.3 può essere seguito per operatori a range arbitrario.
Consideriamo $T \in \mathcal{B}(X, Y)$ con range non chiuso.
Si tratta di definire sotto quali condizioni una rappresentazione (1.2) può essere fatta. In questo caso vedremo che l'operatore $\tilde{T}$ ammette inversa non limitata definita su un sottospazio denso in Y. L'inversa generalizzata verrà costruita in maniera analoga, ma il dominio sarà nuovamente un insieme denso in Y. Si osservi che già nel paragrafo precedente l'inversa generalizzata è definita su $R(T) \oplus R(T)^{\perp}$, che, nel caso in cui $R(T)$ non è chiuso, risulta essere denso in Y.
Alcuni dei teoremi seguenti vengono solo enunciati. Per le dimostrazioni si può consultare [4] .

Teorema 1.5 *Sia $T \in \mathcal{B}(X, Y)$, con range non chiuso.*
Si supponga inoltre che $dim(N(T)) = dim(R(T)^{\perp})$. Allora

(i) *T ammette la rappresentazione*
$$T = P\tilde{T}$$
con P operatore di proiezione su Y, e $\tilde{T}$ dotato di inversa definita su un sottospazio denso in Y .

(ii) $P = P_{R(T)}$

(iii) $\tilde{T}^{-1}$ *è non limitato*

Definizione 1.4 *Si definisce* **inversa generalizzata** $T^{\dagger}$ *per l'operatore* $T \in \mathcal{B}(X, Y)$ *l'applicazione lineare $T^{\dagger} : D(T^{\dagger}) \to X$ che soddisfa:*

(i) $D(T^{\dagger})$ *è denso in* Y

(ii) $\forall\, y \in D(T^{\dagger}) \qquad \inf_{x\in X} \|Tx - y\| = \|Tx^{\dagger} - y\|$ *dove* $x^{\dagger} = T^{\dagger}y$

(iii) $\|x^{\dagger}\| < \|x'\| \qquad \forall\, x' \in X, \quad x' \neq x^{\dagger}$ *tale che* $Tx' = Tx^{\dagger}$

Prima di procedere si osservi che nel caso in cui $R(T)$ è chiuso, l'inversa generalizzata viene definita come $T^{\dagger} = P_{N(T)^{\perp}}\tilde{T}^{-1}$ *(eq.*(1.3)). Il teorema 1.4 dimostra che tale operatore soddisfa le proprietà di minimo enunciate.
La definizione 1.4 , che contempla anche il caso in cui $R(T)$ non è chiuso, è stata data invece basandosi sulle proprietà di minimo. Verrà dimostrato in seguito che tale operatore $T^{\dagger}$ può essere scritto in una forma corrispondente alla (1.3).
Si osservi inoltre che l'unicità dell'operatore, garantita nel caso in cui $R(T)$ è chiuso, qui non è mensionata. È possibile infatti definire diverse inverse generalizzate, considerando diversi domini densi in Y . In realtà due generiche inverse generalizzate sono coincidenti sul sottoinsieme dove sono entrambe definite.

Teorema 1.6 *Siano $T^{\dagger\prime}$, $T^{\dagger\prime\prime}$ due inverse generalizzate soddisfacenti la definizione* (1.4)
Sia $D = D(T^{\dagger\prime}) \cap D(T^{\dagger\prime\prime})$.
Allora

$$T^{\dagger\prime} y = T^{\dagger\prime\prime} y \qquad \forall\, y \in D$$

Dimostrazione. Supponiamo per assurdo che esista $\bar{y} \in D$ t.c.

$$x^{\dagger\prime} = T^{\dagger\prime}\bar{y} \neq x^{\dagger\prime\prime} = T^{\dagger\prime\prime}\bar{y}$$

Per definizione $x^{\dagger\prime}$ e $x^{\dagger\prime\prime}$ soddisfano entrambi la proprietà di minimo 1.4 – (*ii*), quindi $Tx^{\dagger\prime} = Tx^{\dagger\prime\prime}$.
Ma allora per la proprietà (iii) si dovrebbe avere $\|x^{\dagger\prime}\| < \|x^{\dagger\prime\prime}\|$ e $\|x^{\dagger\prime\prime}\| < \|x^{\dagger\prime}\|$, il che è assurdo. ∎

Corollario 1.1 *Se $T^\dagger$ è definito su tutto Y, qualsiasi altra inversa generalizzata deve necessariamente essere una restrizione di $T^\dagger$.*

Come già detto vediamo come l'inversa generalizzata, nel caso in cui $R(T)$ non sia chiuso, ammetta comunque la rappresentazione operatoriale del tipo (1.3).

Teorema 1.7 *Sia $T \in B(X,Y)$, $R(T)$ non chiuso e $dim(N(T)) = dim(R(T)^\perp)$.*
Allora, considerando $\tilde{T}$ introdotta nel teorema 1.5, l'operatore

$$T^\dagger = P_{N(T)^\perp}\tilde{T}^{-1}$$

è un'inversa generalizzata nel senso della definizione 1.4.
Inoltre considerando $x^\dagger = T^\dagger y$, $y \in D(T^\dagger)$ si ha

$$\{\, x' \in X \,:\, \|Tx' - y\| \leq \|Tx - y\| \quad \forall x \in X \,\} = \{\, x' \in X \,:\, Tx' = Tx^\dagger \,\}$$

$T^\dagger$ è non limitata.

Dimostrazione. Utilizzando la decomposizione del teorema 1.5, dimostriamo che $T^\dagger$ è effettivamente una inversa generalizzata come definita in 1.4.
$D(\tilde{T}^{-1})$ è denso in Y e $D(\tilde{T}^{-1}) = D(T^\dagger)$, quindi $T^\dagger$ soddisfa (*i*) della definizione 1.4.
Dimostriamo ora che $T^\dagger$ soddisfa (*ii*) della definizione 1.4.
Poichè $T = P_{R(T)}\tilde{T}$, allora $R(\tilde{T}) \subseteq R(T) \oplus R(T)^\perp$;
quindi $D(T^\dagger) = D(\tilde{T}^{-1}) = R(\tilde{T}) \subseteq R(T) \oplus R(T)^\perp$.
Verifichiamo innanzitutto che, $\forall y \in D(T^\dagger)$, $TT^\dagger y = P_{R(T)}y$.
Si ha
$TT^\dagger y = TP_{N(T)^\perp}\tilde{T}^{-1}y = P_{R(T)}\tilde{T}\tilde{T}^{-1} = P_{R(T)}y \qquad \forall y \in D(T^\dagger)$.
Sia $x' \in D(T)$ e $y \in D(T^\dagger) = D(\tilde{T}^{-1})$; allora
$\|Tx' - y\|^2 = \|Tx^\dagger - y\|^2 + \|Tx' - Tx^\dagger\|^2$
infatti $Tx^\dagger - y = TT^\dagger y - y = (TT^\dagger - I)y = (P_{R(T)} - I)y = -P_{R(T)^\perp}y \in R(T)^\perp$
e $Tx' - Tx^\dagger = T(x' - x^\dagger) \in R(T)$.
Quindi $\|Tx^\dagger - y\| \leq \|Tx' - y\|$, che dimostra (*ii*).
Per dimostrare infine (*iii*), si osservi che l'insieme $\Delta = \{\, x \in D(T) : Tx = Tx^\dagger\}$ contiene tutti e soli gli elementi che minimizzano la relazione al punto (ii).
Ogni elemento $\hat{x} \in \Delta$ può essere scritto nella forma $\hat{x} = x^\dagger + z$, con $z \in N(T)$, infatti
$Tx = Tx^\dagger \;\Leftrightarrow\; T(x - x^\dagger) = 0 \;\Leftrightarrow\; x - x^\dagger \in N(T)$.
Inoltre, per come è definito,

$$x^\dagger = T^\dagger y = P_{N(T)^\perp}(\tilde{T}^{-1}y) \in N(T)^\perp$$

Quindi $\|x\|^2 = \|x^\dagger\|^2 + \|z\|^2$ che verifica (*iii*).

Per completare, verifichiamo che $T^\dagger$ risulta necessariamente non limitata.
Accenniamo brevemente come si svilupperà la dimostrazione.
Se per assurdo $T^\dagger$ fosse limitata, vedremmo che l'operatore T ammetterebbe una rappresentazione $T = P_{R(T)}T_0$, con T_0 operatore limitato invertibile con inversa limitata. In questo caso il teorema 1.2 garantirebbe che T possiede range chiuso, il che contraddice l'ipotesi su T. Vediamo la cosa in dettaglio.
Supponiamo per assurdo che $T^\dagger = P_{N(T)^\perp}\tilde{T}^{-1}$ sia limitata.
Definiamo $T_0 : X \longrightarrow Y$

$$T_0 = T + UP_{N(T)}$$

dove U è un isometria che immerge $N(T) \subset X$ in $R(T)^\perp \subset Y$, che sono della stessa dimensione per ipotesi.
Poichè $P_{R(T)}UP_{N(T)} = 0$, si ha che

$$T = P_{R(T)}T_0$$

con T_0 limitato poichè somma di T e $UP_{N(T)}$ entrambi limitati.
Dimostriamo che l'inverso di T_0 è l'operatore $B : D(B) \longrightarrow X$

$$B = T^\dagger + U^{-1}P_{R(T)^\perp}$$

Osserviamo innanzitutto che, essendo $T^\dagger$ limitato, anche B è un operatore limitato e che $D(B) = D(T^\dagger)$, infatti $D(B) = D(T^\dagger) \cap D(U^{-1}P_{R(T)^\perp}) = D(T^\dagger) \cap Y = D(T^\dagger)$.
Inoltre, poichè $TU^{-1}P_{R(T)^\perp} = 0$, $UP_{N(T)}T^\dagger = 0$ e $UP_{N(T)}U^{-1}P_{R(T)^\perp} = P_{R(T)^\perp}$ (per l'ultima si osservi che $UP_{N(T)}U^{-1}|_{R(T)^\perp} = I|_{R(T)^\perp}$), si ha che $(T_0B)y = (TT^\dagger + P_{R(T)^\perp})y \qquad \forall y \in D(B)$.
Siccome, come già visto in precedenza, $TT^\dagger = P_{R(T)}$ su $D(B) = D(T^\dagger)$, allora
$T_0By = TT^\dagger y + P_{R(T)^\perp}y = (P_{R(T)} + P_{R(T)^\perp})y = y \qquad \forall y \in D(B)$,
quindi B è inversa destra di T_0.
Vediamo che B è anche inversa sinistra di T_0 e che quindi corrisponde all'inversa di T_0.
Si ha $BT_0By = By \qquad \forall y \in D(B)$.
$R(B)$ è lo spazio generato da $R(T^\dagger) = N(T)^\perp \cap D(T)$ e $N(T)$, che equivale esattamente a $D(T) = D(T_0)$. Quindi $\forall\, x \in D(T_0) \quad \exists\, y \in D(B)$ tale che $x = By$, da cui $BT_0x' = x' \quad \forall\, x' \in D(T_0)$.
B è anche inversa sinistra di T_0, quindi è l'unica inversa di T_0, ossia $B = T_0^{-1}$.
Quindi T ammette una rapppresentazione del tipo (1.2) con T_0 invertibile con inversa limitata. Per il teorema 1.2, $R(T)$ risulta essere necessariamente chiuso, in contraddizione con l'ipotesi su $R(T)$. Ne segue che $T^\dagger$ non può essere limitata. ∎

Si osservi che nel caso in cui l'operatore è invertibile la rappresentazione (1.2) si caratterizza con $P_{R(T)} = I$, $\tilde{T} = T$. Si ottiene $T^\dagger = P_{N(T)^\perp}\tilde{T}^{-1} = IT^{-1} = T^{-1}$, come ovviamente ci si aspettava.
A conclusione di questo paragrafo è utile sottolineare che tutti i risultati ottenuti valgono nel contesto degli operatori lineari tra spazi di Hilbert. Per un'analisi dettagliata dell'inversa generalizzata in spazi di Banach si può consultare [3] e [4] ; mentre nel caso in cui gli spazi siano solo topologici si può consultare [49].

1.4 Proprietà della inversa generalizzata

Enunciamo e verifichiamo alcune proprietà dell'operatore $T^\dagger$.

Teorema 1.8 *Se $T \in B(X,Y)$ è dotato di range chiuso, allora*

$$R(T^\dagger) = R(T^*) = R(T^\dagger T)$$

Dimostrazione. La dimostrazione poggia sulla nota relazione $R(T^*) = N(T)^\perp$, valida poichè $R(T)$ è chiuso.

- $R(T^\dagger) \subseteq R(T^*)$

La verifica è ovvia considerando la definizione operatoriale 1.3 .
Vediamo una dimostrazione equivalente che utilizza le relazioni esposte nella definizione variazionale 1.2 .
Dimostriamo che, fissato $y \in Y$, si ha $T^\dagger y \in N(T)^\perp = R(T^*)$.
Per il teorema di decomposizione, siccome $X = N(T) \oplus N(T)^\perp$, consideriamo
$x^\dagger = T^\dagger y = x_1 + x_2$, con $x_1 \in N(T)^\perp$, $x_2 \in N(T)$. Allora
$Tx_1 = Tx_1 + Tx_2 = TT^\dagger y = P_R y$.
Per il teorema 1.1 , x_1 è una soluzione ai minimi quadrati. Allora si deve avere necessariamente $x_2 = 0$, poichè altrimenti $\|x_1\| < \|x^\dagger\|$.
Si ottiene $x^\dagger \in N(T)^\perp = R(T^*)$.

- $R(T^\dagger) \supseteq R(T^*)$

Anche in questo caso, per mezzo della definizione 1.3 , la dimostrazione è immediata (si osservi che $R(\tilde{T}^{-1}) = X$).
Si può comunque verificare il risultato in maniera equivalente utilizzando la definizione 1.2 .
Supponiamo che $x \in R(T^*)$. Poniamo $y = Tx$.
$Tx = P_R Tx = P_R y$; x è quindi una soluzione ai minimi quadrati. Dimostriamo che è proprio quella di norma minima.
Qualsiasi altra pseudosoluzione x' verifica $Tx' = P_R y = Tx$.
Quindi $x' = x + (x' - x)$ con $x \in N(T)^\perp$ e $(x' - x) \in N(T)$ (infatti $T(x' - x) = Tx' - Tx = P_R y - P_R y = 0$).
Allora $\|x'\|^2 = \|x\|^2 + \|x' - x\|^2 \geq \|x\|^2$.
x è la pseudosoluzione di norma minima, quindi $x = T^\dagger y \in R(T^\dagger)$

- Per la seconda uguaglianza, ovviamente $R(T^\dagger) \supseteq R(T^\dagger T)$.

Inoltre, se $x \in N(T)^\perp$, cioè $x = T^\dagger y$ con $y \in Y$, si ha $x = T^\dagger P_R y \in R(T^\dagger T)$, infatti considerando il teorema 1.1 , $Tx = P_R y \Leftrightarrow Tx = P_R(P_R y)$.
Quindi $R(T^\dagger) \subseteq R(T^\dagger T)$. ∎

•

Teorema 1.9 *Sia $T \in B(X,Y)$. Allora :*

(i) $T^\dagger$ *è lineare*

(ii) $T^\dagger$ *è limitato* $\Leftrightarrow$ $R(T)$ *è chiuso*

Dimostrazione.
Verifichiamo (i) .
Siano $y_1, y_2 \in D(T^\dagger)$ con $D(T^\dagger)$ sottospazio denso in Y.
$TT^\dagger y_1 + TT^\dagger y_2 = P_R y_1 + P_R y_2 = P_R(y_1 + y_2) = TT^\dagger(y_1 + y_2)$. Allora
$T\left(T^\dagger y_1 + T^\dagger y_2 - T^\dagger(y_1 + y_2)\right) = 0$.
Quindi
(a) $T^\dagger y_1 + T^\dagger y_2 - T^\dagger(y_1 + y_2) \in N(T) \cap N(T)^\perp = \{0\}$

Inoltre $TT^\dagger(\alpha y_1) = P_R(\alpha y_1) = \alpha P_R y_1 = \alpha TT^\dagger y_1 = T(\alpha T^\dagger y_1)$, con $\alpha \in \mathbb{R}$. Allora, analogamente ad (a) , si ottiene
(b) $T^\dagger(\alpha y_1) = \alpha T^\dagger y_1$.
(a) e (b) verificano (i) .
Per (ii) abbiamo già visto e dimostrato nel teorema 1.7 che se $R(T)$ non è chiuso allora $T^\dagger$ risulta essere non limitato.
Dimostriamo ora che se $R(T)$ è chiuso, allora $T^\dagger$ è limitato.
Essendo $R(T)$ chiuso, allora

$$\exists\, m > 0 \quad \text{tale che} \quad \|Tx\| \geq m\|x\| \qquad \forall\, x \in N(T)^\perp$$

(cfr. [13] , teorema 1.2.1) , che implica $\|TT^\dagger y\| \geq m\|T^\dagger y\| \qquad \forall\, y \in Y$.
Si ottiene $\|y\| \geq \|P_R y\| = \|TT^\dagger y\| \geq m\|T^\dagger y\|$,
quindi

$$\|T^\dagger y\| \leq m^{-1}\|y\|$$

∎

Proposizione 1.3 *Se $T \in B(X,Y)$, $R(T)$ chiuso, allora*

(i) $TT^\dagger = P_{R(T)}$

(ii) $T^\dagger T = P_{N(T)^\perp}$

(iii) $T^\dagger TT^\dagger = T^\dagger$

(iv) $TT^\dagger T = T$

(v) $(TT^\dagger)^* = TT^\dagger$

(vi) $(T^\dagger T)^* = T^\dagger T$

(vii) $(T^\dagger)^\dagger = T$

Dimostrazione. Useremo sistematicamente la definizione operatoriale 1.3 . Le stesse dimostrazioni si possono fare utilizzando la definizione variazionale, che risulta però meno efficace.

(i) $TT^\dagger = T(P_{N(T)^\perp}\tilde{T}^{-1}) = (TP_{N(T)^\perp})\tilde{T}^{-1} = T\tilde{T}^{-1} = P_R\tilde{T}\tilde{T}^{-1} = P_R$ Si osservi che questa relazione è stata verificata nel caso in cui $R(T)$ non è chiuso nella dimostrazione del teorema 1.7 .

(ii) $T^\dagger T = P_{N(T)^\perp}\tilde{T}^{-1}P_R\tilde{T} = P_{N(T)^\perp}$
infatti, considerando
$x \in X$, $x = x' + x''$, $x' \in N(T)^\perp$, $x'' \in N(T)$,
$y' = \tilde{T}x'$, $y'' = \tilde{T}x''$, $y' \in R(T)$, $y'' \in R(T)^\perp$,
si ottiene

$$\begin{aligned} P_{N(T)^\perp}\tilde{T}^{-1}P_R\tilde{T}x &= P_{N(T)^\perp}\tilde{T}^{-1}P_R\tilde{T}(x' + x'') \\ &= P_{N(T)^\perp}\tilde{T}^{-1}P_R\tilde{T}x' + P_{N(T)^\perp}\tilde{T}^{-1}P_R\tilde{T}x'' \\ &= P_{N(T)^\perp}\tilde{T}^{-1}P_R y' + P_{N(T)^\perp}\tilde{T}^{-1}P_R y'' \\ &= P_{N(T)^\perp}\tilde{T}^{-1}P_R y' = P_{N(T)^\perp}\tilde{T}^{-1}y' = P_{N(T)^\perp}x' \\ &= P_{N(T)^\perp}(x' + x'') = P_{N(T)^\perp}x \end{aligned}$$

(iii) Utilizzando (ii) si ottiene
$T^\dagger T T^\dagger = P_{N(T)^\perp} T^\dagger = P_{N(T)^\perp} P_{N(T)^\perp} \tilde{T}^{-1} = P_{N(T)^\perp} \tilde{T}^{-1} = T^\dagger$.

(iv) Utilizzando (i) si ottiene
$TT^\dagger T = P_R T = T$

(v) – (vi) Per dimostrare le due relazioni , si osservi che, da (i), (ii) segue che $T^\dagger T$ e $TT^\dagger$ sono operatori di proiezione. Mostriamo che ogni operatore di proiezione è autoaggiunto; la (v), (vi) ne saranno una conseguenza diretta.
Sia P un generico operatore di proiezione su un generico sottospazio chiuso R ,
con ; $x = x' + x''$ $y = y' + y''$ $x', y' \in R$, $x'', y'' \in R^\perp$.
$(P_R x, y) = (x', y) = (x', y') = (x, y') = (x, P_R y)$
quindi $(P_R)^* = P_R$

(vii) Sapendo che $T^\dagger = P_{N(T)^\perp} \tilde{T}^{-1}$, si può notare che $T^\dagger$ ammette una rappresentazione del tipo (1.2) con $(\tilde{T}^\dagger) = \tilde{T}^{-1}$ invertibile con inversa limitata.
Si osservi inoltre che $T^\dagger x = 0$ per tutti e soli gli $x \in R(T)^\perp$, infatti $\tilde{T}^{-1}$ è invertibile e fa corrrispondere $R(T)^\perp$ con $N(T)$. Quindi $N(T^\dagger) = R(T)^\perp$.
Applichiamo allora la definizione 1.3 all'operatore $T^\dagger$. Si ha
$(T^\dagger)^\dagger = P_{N(T^\dagger)^\perp} (\tilde{T}^\dagger)^{-1} = P_{(R(T)^\perp)^\perp} (\tilde{T}^{-1})^{-1} = P_{R(T)} \tilde{T} = T$

■

Si osservi che se T è autoaggiunto, considerando $R(T) = N(T^*)^\perp = N(T)^\perp$ e le proprietà (i) – (ii) , si può concludere che $TT^\dagger = T^\dagger T$.
Altre proprietà si possono verificare facilmente facendo uso di quelle enunciate nella proposizione precedente.

A conclusione di questo paragrafo enunciamo altre definizioni di inversa generalizzata, limitandoci al caso in cui $R(T)$ è chiuso.
Queste definizioni, come già accennato, poggiano sulle proprietà appena verificate.
È interessante notare che queste proprietà sono di notevole importanza poichè caratterizzano esattamente l'inverso generalizzato, ossia esso è l'unico operatore che le soddisfa.

Definizione 1.5 Moore
Se $T \in \mathcal{B}(X, Y)$, $R(T)$ chiuso , allora $T^\dagger$ è l'unico operatore in $\mathcal{B}(Y, X)$ che soddisfi:

(i) $TT^\dagger = P_{R(T)}$

(ii) $T^\dagger T = P_{R(T^\dagger)}$

Definizione 1.6 Penrose
Se $T \in \mathcal{B}(X, Y)$, $R(T)$ chiuso , allora $T^\dagger$ è l'unico operatore in $\mathcal{B}(Y, X)$ che soddisfi:

(i) $(TT^\dagger)^* = (TT^\dagger)$

(ii) $(T^\dagger T)^* = (T^\dagger T)$

(iii) $T^\dagger T T^\dagger = T^\dagger$

(iv) $TT^\dagger T = T$

Definizione 1.7 Deoser and Whalen
Se $T \in \mathcal{B}(X, Y)$, $R(T)$ chiuso , allora $T^\dagger$ è l'unico operatore in $\mathcal{B}(Y, X)$ che soddisfi:

(i) $T^\dagger T x = x \quad \forall x \in N(T)^\perp$

(ii) $T^\dagger y = 0 \quad \forall y \in R(T)^\perp$

Nel paragrafo precedente è stato dimostrato che la definizione variazionale è equivalente a quella operatoriale. Si ha inoltre il seguente risultato:

Proposizione 1.4 *Le definizioni*
1.2 (variazionale), 1.3 (operatoriale),
1.5 (di Moore), 1.6 (di Penrose), 1.7 (di Deoser-Whalen)
sono equivalenti.

Si osservi che l'ultima proposizione è valida solo se $R(T)$ è chiuso.
Una dimostrazione di quest'ultima proposizione si può trovare in [13].
Per maggiori dettagli sulla definizione di Moore si consulti [31]; per quella di Penrose si consulti [37] ; per quella di Deoser-Whalen si consulti [10].

1.5 Inversa generalizzata di operatori compatti e matrici

Nella maggior parte delle applicazioni, l'operatore T risulta compatto.
In questo caso, utilizzando il sistema singolare $\{\mu_n; u_n, v_n\}$ di T (vedi appendice A), è possibile determinare una semplice espressione dell'operatore inverso $T^\dagger$.

Teorema 1.10 *Sia $T \in \mathcal{B}(X, Y)$ un'operatore compatto con sistema singolare $\{\mu_n; u_n, v_n\}$.*
Se $y \in R(T) \oplus R(T)^\perp$, allora

$$T^\dagger y = \sum_{n=1}^{\infty} \frac{1}{\mu_n} (y, v_n)_Y u_n \tag{1.7}$$

Dimostrazione. Sia P_R l'operatore di proiezione ortogonale su $\overline{R(T)}$.
Allora, poichè $y \in R(T) \oplus R(T)^\perp$, $P_R y \in R(T)$.
Dal teorema di Picard (vedi appendice A) applicato a $P_R y$, si ottiene

$$\sum_{n=1}^{\infty} \frac{1}{\mu_n^2} |(P_R y, v_n)|^2 < +\infty$$

Si osservi che, poichè $v_n \in \overline{R(T)}$, allora $(P_R y, v_n) = (y, P_R v_n) = (y, v_n)$.
Utilizzando nuovamente il teorema di Picard, il quale stabilisce condizioni non solo sufficienti ma anche necessarie alla convergenza, si conclude che

$$\sum_{n=1}^{\infty} \frac{1}{\mu_n} (y, v_n) u_n \qquad \text{è convergente in } X$$

Sia allora $v \in X$, $v = \sum_{n=1}^{\infty} \frac{1}{\mu_n} (y, v_n) u_n$.
Si osservi che $\{u_n\} \subset N(T)^\perp$; quindi, poichè $N(T)^\perp$ è chiuso, $v \in N(T)^\perp$.

Si ha

$$\begin{aligned} Tv &= \sum_{n=0}^{\infty} \frac{1}{\mu_n}(P_R y, v_n) T u_n \\ &= \sum_{n=0}^{\infty} \frac{1}{\mu_n}(y, v_n)\mu_n v_n \\ &= \sum_{n=0}^{\infty} (P_R y, v_n) v_n \\ &= P_R y \end{aligned}$$

Quindi, per il teorema 1.1 (i), il limite è una soluzione ai minimi quadrati ; ricordando che $v \in N(T)^{\perp}$, necessariamente $v = T^{\dagger} y$. ∎

Nel caso in cui gli spazi di Hilbert siano di dimensione finita, l'operatore T può essere rappresentato per mezzo di una matrice. Più precisamente, un generico operatore $T' \in \mathcal{L}(X,Y)$, con $dim(X) = r$, $dim(Y) = s$, viene rappresentato da una matrice $T \in M_{s,r}(\mathbb{R})$, dove $M_{s,r}(\mathbb{R})$ è lo spazio delle matrici $s \times r$ a coefficienti reali.
Questo si verifica ogni volta che un problema inverso viene discretizzato e quindi rappresentato su uno spazio di dimensione finita.

Teorema 1.11 *Sia $T \in M_{s,r}(\mathbb{R})$, $n = rank(T)$ allora esistono $M \in M_{s,n}(\mathbb{R})$ e $N \in M_{r,n}(\mathbb{R})$ con $rank(M) = rank(N)$, tali che*

$$T = MN^t$$

Per una precisa dimostrazione si consulti [40].
Diamo qui una spiegazione euristica di come si determinano M e N.
Una volta determinato il rango di T , ad esempio per mezzo dell'uso dei minori principali, si isolino n colonne linearmente indipendenti. Queste n colonne formano la matrice M.
Per ogni colonna c linearmente dipendente da quelle considerate in M, si determini la combinazione lineare di c in funzione delle colonne di M. I coefficienti trovati dalle singole combinazioni lineari formano la matrice N.
Vediamo un semplice esempio.

$$\begin{pmatrix} 1 & 3 & 5 \\ 2 & 1 & 5 \\ 1 & 0 & 2 \\ 1 & 1 & 3 \end{pmatrix} \text{ con } T \in M_{4,3}(\mathbb{R}) , \quad rank(T) = 2 .$$

Si ha $M \in M_{4,2}(\mathbb{R})$, $N_{3,2}(\mathbb{R})$.
Indicando con c_i la colonna iesima :

$$M = (c_1, c_2) = \begin{pmatrix} 1 & 3 \\ 2 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Per determinare i coefficienti della combinazione lineare che esprime c_3 in funzione di c_1 e c_2, si risolve il sistema

$$\alpha c_1 + \beta c_2 = c_3$$

Si ottiene $\alpha = 2$, $\beta = 1$. Quindi

$$N = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \alpha & \beta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix} \text{ con } \quad T = MN^t$$

La decomposizione descritta permette di calcolare in maniera diretta $T^{\dagger}$.

Definizione 1.8 *Sia $T \in M_{s,r}(\mathbb{R})$, $n = rank(T)$, $M \in M_{s,n}(\mathbb{R})$, $N \in M_{r,n}(\mathbb{R})$ con $n = rank(M) = rank(N)$, $T = MN^t$.*
Allora si definisce **pseudoinversa di T** *la matrice*

$$T^{\dagger} = N(N^tN)^{-1}(M^tM)^{-1}M^t$$

Si osservi che
(a) $T^{\dagger} \in M_{r,s}(\mathbb{R})$
(b) l'operatore corrrispondente alla matrice T è un operatore con range chiuso.

Teorema 1.12 *Se $T \in M_{s,r}(\mathbb{R})$ è la rappresentazione dell'operatore $T' \in \mathcal{B}(X,Y)$, allora la matrice pseudoinversa $T^{\dagger} \in M_{r,s}(\mathbb{R})$ è la rappresentazione matriciale dell'operatore $T'^{\dagger} \in \mathcal{B}(Y,X)$.*
Inoltre $rank(T) = rank(T^{\dagger})$

Dimostrazione. La dimostrazione è immediata considerando la definizione 1.6 di Penrose e le usuali proprietà del calcolo matriciale. Occorre verificare le relazioni $(i) - (iv)$ della definizone 1.6 . Mostriamo la verifica di (iii); le altre si verificano in modo analogo.
(iii) $T^{\dagger}TT^{\dagger} = T^{\dagger}$

$$\begin{aligned} T^{\dagger}TT^{\dagger} &= N(N^tN)^{-1}(M^tM)^{-1}M^tMN^tN(N^tN)^{-1}(M^tM)^{-1}M \\ &= N(N^tN)^{-1}(M^tM)^{-1}M^{-1} = T^{\dagger} \end{aligned}$$

Per dimostrare che $rank(T) = rank(T^{\dagger})$, si osservi che , per il teorema 1.8 $R(T^{\dagger}) = R(T^*)$.
La matrice che rappresenta T^* è la matrice trasposta.
Quindi $rank(T^{\dagger}) = rank(T^t) = rank(T)$ ∎

Il metodo descritto è utile per la risoluzione di problemi semplici di piccola dimensione. Vediamo adesso un altro algoritmo per calcolare $T^{\dagger}$; esso è l'applicazione numerica del procedimento che ha condotto alla definizione operatoriale 1.3 e risulta particolarmente efficace nel caso in cui la matrice T è di dimensione molto grande.
In moltissimi problemi numerici, come ad esempio la elaborazione della immagini, la matrice, oltre ad avere una precisa struttura, è simmetrica. Consideriamo quindi una generica matrice hermitiana.
Per non appesantire troppo le notazioni, d'ora in poi chiameremo con lo stesso nome l'operatore lineare e la matrice che lo rappresenta.
Sappiamo dall'algebra lineare che se $H \in M_{n,n}(\mathbb{R})$ hermitiana, allora $\exists$ $C, F \in M_{n,n}(\mathbb{R})$ con C matrice unitaria, le cui colonne formano una base ortogonale di autovettori di H e $F = diag(\lambda_1, \ldots, \lambda_2)$, dove λ_i rappresenta l'autovalore corrispondente all'autovettore dell'iesima colonna, t.c.

$$H = C^*FC$$

Si può supporre che, posto $k = rank(H)$, con $k \leq n$, allora
$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_k| > 0$ e $\lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_n = 0$.
Un ragionamento analogo a quello fatto all'inizio del paragrafo per la determinazione delle matrici M e N, ci permette di osservare che, posto

$$G = diag(g_1, \ldots, g_n) \quad \text{con} \quad g_i = \begin{cases} 1 & i = 1, \ldots, k \\ 0 & altrimenti \end{cases} ,$$

si ha che

$$P_R = C^*GC$$

dove R è lo spazio generato dalle colonne di H.
Poichè H è hermitiana $R(H) = R(H^*) = N(H)^{\perp}$.
Consideriamo ora la definizione 1.3 di inversa generalizzata.

Posto $\tilde{H} = C^*VC$ dove $V = diag(v_1, \dots v_n)$ con $v_i = \begin{cases} \lambda_i & i = 1, \dots, k \\ 1 & altrimenti \end{cases}$

verifichiamo che

$$H = P_R \tilde{H}$$

Si ha che $\tilde{H}$ è invertibile e
$P_{R^\perp}\tilde{H} = (C^*GC)(C^*VC) = C^*FC = H$, infatti $GV = F$.
Inoltre $\tilde{H}^{-1} = C^*WC$ con $W = diag(w_1, \dots w_n)$, $w_i = \begin{cases} \frac{1}{\lambda_i} & i = 1, \dots, k \\ 1 & altrimenti \end{cases}$

Sappiamo infine che

$$H^\dagger = P_{N^\perp}\tilde{H}^{-1}$$

quindi

$$H^\dagger = (C^*GC)(C^*WC) = C^*GWC$$

Posto $Z = GW$, si ottiene $Z = diag(z_1, \dots z_n)$ $z_i = \begin{cases} \frac{1}{\lambda_i} & i = 1, \dots, k \\ 0 & altrimenti \end{cases}$.

La matrice $H^\dagger$ può essere scritta come

$$H^\dagger = C^*ZC$$

Quest'ultimo procedimento è utile se implementato. Accennniamo a tal proposito che gli autovettori e gli autovalori di H possono essere calcolati per mezzo di metodi di approssimazione come, ad esempio, il metodo delle potenze e il metodo basato sulla decomposizione QR di H, dove Q è una matrice unitaria e R triangolare superiore (si consulti [8]) .

Capitolo 2

Metodi di regolarizzazione per problemi mal posti

2.1 Introduzione

Ogni volta che si affronta un problema concreto, i dati a disposisione sono sempre affetti da errore; è chiaro infatti che a causa della tolleranza della apparecchiatura che effettua le rilevazioni, i dati acquisiti da fenomeni naturali sono noti entro un margine fissato. Nel caso in cui una piccola variazione del dato comporti una grande variazione della solusione, non si hanno garansie sulla affidabilità del risultato.
Scopo di questo paragrafo è analizzare metodologie che permettano di affrontare questo problema che, come vedremo, rientra nella categoria dei problemi mal posti.
In quest'ottica verrà sviluppata la teoria generale dei metodi di regolarizzazione per la risolusione di equasioni lineari in spasi di Hilbert. Verrà illustrato il noto algoritmo regolarizzante di A.N. Tikhonov e l'algoritmo iterativo di Landweber-Fridman. Questi, per la loro semplicità, consentono una rapida analisi delle proprietà caratteristiche dei metodi di regolarizzazione.

2.2 Problemi mal posti e problemi inversi

Consideriamo una generica equazione

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

dove K è un'applicazione tra gli spazi di topologici X e Y , e $y \in Y$ è il dato assegnato.
Agli inizi di questo secolo J. Hadamard classificò la categoria dei problemi mal posti.

Definizione 2.1 *Il problema* (2.1) *è detto* **ben posto** *se sono soddisfatte le condizioni seguenti*

(a) $\forall y \in Y$ *la soluzione* $x \in X$ *esiste*

(b) la soluzione è unica

(c) la soluzione dipende con continuità da y *, cioè l'applicazione* $y \longrightarrow x$ *è continua.*

Ogni problema non ben posto è detto **mal posto.**

Risulta evidente che l'essere ben posto è una proprietà legata non solo all'operatore K, ma dipende dagli spazi X, Y e dalle loro topologie.
La condizione (a) equivale alla surgettività dell'operatore K, ed è quindi soddisfatta se $Y = K(X)$, mentre la condizione (b) equivale all'iniettività. Affinchè entrambe le condizioni $(a), (b)$ siano soddisfatte è necessario e sufficiente che l'applicazione sia invertibile.
Come accennato nell'introduzione, la condizione (c) è quella di maggior interesse e sta alla base dello studio dei problemi non ben posti. Si osservi che questa condizione dipende dalle topologie introdotte in X e in Y, le quali possono rendere l'applicazione $y \longrightarrow x$ continua o meno.
Nasce quindi l'idea di rendere ben posto qualsiasi problema modificando la terna (K,X,Y). Ad esempio nel caso in cui K sia iniettiva e continua in un generico insieme compatto $\tilde{X} \subset X$, si può considerare il problema rispetto alla terna $(K, \tilde{X}, K(\tilde{X}))$. In questo caso l'applicazione risulta ovviamente surgettiva e l'inversa K^{-1}, ristretta all'insieme chiuso $K(\tilde{X})$, risulta continua (cfr. [47]).
Il problema così ridefinito soddisfa le condizioni (a) e (c) ed è quindi ben posto.
Si osservi che nel caso in cui l'applicazione K è lineare e gli spazi X, Y sono di Hilbert, è sempre possibile attivare una strategia del genere, qui enunciata nel caso in cui K è iniettivo e continuo su un sottospazio compatto $\tilde{X}$ (si consulti [2] pag. 19).
Una riformulazione di questo tipo è però priva di interesse pratico poichè modifica l'insieme dei dati Y. Si supponga che i dati in ingresso abbiano una componente che non appartiene a $K(\tilde{X})$. In questo caso il problema ristretto alla terna $(K, \tilde{X}, K(\tilde{X}))$ non ammette soluzioni.
Un esempio tipico è quello in cui l'operatore K è un operatore di convoluzione. È noto che se il nucleo ha certe proprietà di regolarità, l'insieme $K(X)$ risulta essere formato da funzioni "liscie" (addirittura $\mathcal{C}^\infty$). Spesso i dati, affetti da errore, non appartengono alla stessa classe di funzioni $\mathcal{C}^\infty$; non è possibile quindi affrontare il problema rispetto alla terna $(K, X, K(X))$.

Consideriamo nuovamente le condizioni (a) e (b).
L'invertibilità di K è una richiesta molto restrittiva che in svariati problemi non viene rispettata. Alla luce dei risultati enunciati nel capitolo precedente, è possibile considerare la definizione 2.1 nel senso della soluzione generalizzata e dare così un significato più ampio e concreto al concetto di problema ben posto.
Al fine di operare con la pseudoinversa, d'ora in poi considereremo X, Y spazi di Hilbert e K operatore lineare e continuo; per utilizzare le stesse notazioni del capitolo 1, indicheremo l'operatore con il simbolo T.
La soluzione generalizzata $x^\dagger$ se esiste è unica; l'inversa generalizzata $T^\dagger$ garantisce quindi l'unicità, ma non l'esistenza della soluzione. Sappiamo infatti che, nel caso in cui R(T) non è chiuso, si ha $D(T^\dagger) = R(T) \oplus R(T)^\perp \neq Y$ (cfr. proposizione 1.1).
Viceversa, nel caso in cui $R(T)$ è chiuso, oltre a fornire una unica soluzione $\forall y \in Y$, l'operatore lineare $T^\dagger$ è anche limitato; l'applicazione $y \longrightarrow x^\dagger = T^\dagger y$ è continua e soddisfa il punto (c) della definizione di problema ben posto . Abbiamo così dimostrato la seguente proposizione.

Proposizione 2.1 *Se T è un operatore lineare limitato tra spazi di Hilbert, il problema* (2.1) *, nel senso della soluzione generalizzata, è ben posto se e solo se $R(T)$ è chiuso.*

Questo risultato è di notevole importanza per la classificazione dei problemi ben posti. Si consideri l'operatore di **Fredholm di prima specie** con range di dimensione non finita, nella forma

$$\int_a^b k(r,s)\, u(s)\ ds = g(r) \qquad c \leq r \leq d \tag{2.2}$$

dove il nucleo $k : L^2([c,d] \times [a,b]) \longrightarrow \mathbb{R}$ e $g : L^2([a,b]) \longrightarrow \mathbb{R}$ sono funzioni assegnate, e $u : L^2([c,d]) \longrightarrow \mathbb{R}$ è la soluzione da determinare.
Si osservi che tale operatore, essendo compatto, ha range non chiuso (infatti ogni operatore compatto possiede range chiuso se e solo se questo ha dimensione finita; si consulti [2] , pag. 17 osservazione 3.1) e dà luogo a un problema mal posto.
Problemi legati all'esistensa della solusione si hanno ad esempio nel caso in cui il nucleo assume la forma $k(r,s) = m(r)n(s)$ con $g(r) \neq Cm(r)$ $C \in \mathbb{R}$
Inoltre, nel caso in cui la solusione esiste, questa non è necessariamente unica; nell'esempio precedente, se u è una solusione, qualunque altra funsione

$$u' = u + \tilde{u} \qquad \text{con} \qquad \int_a^b n(s)\tilde{u}(s)ds = 0$$

risolve l'equasione (2.2).
Più interessante è lo studio della dipendenza continua dai dati.
Il lemma di Riemman-Lebesgue (cfr. [6] pag. 159) asserisce che

$$\int_a^b k(r,s)\sin ns\ ds \longrightarrow 0 \qquad \text{per } n \longrightarrow +\infty.$$

Consideriamo quindi il dato perturbato

$$\tilde{g}_n(r) = g(r) + A\int_a^b k(r,s)\sin ns\ ds \qquad \text{con } A \in \mathbb{R} \text{ costante arbitraria.}$$

Ovviamente al dato $\tilde{g}_n$ corrisponde la solusione $\tilde{u}_n(s) = u(s) + A\sin ns$.
Quindi, sebbene $\tilde{g}_n \longrightarrow g$, la differensa tra le soluzioni corrispondenti può essere molto grande.
L'operatore di Fredholm di prima specie (2.2) rientra nella categoria degli operatori il cui range è non chiuso e, come ci aspettavamo, dà luogo a un problema mal posto.

Lo studio dei problemi mal posti ha acquistato notevole importansa con la necessità di affrontare i cosiddetti **problemi inversi.**
Il concetto di problema inverso non ammette una caratterizzazione precisa. Supponiamo di aver fissato per "problema diretto" la determinazione di una certa solusione avendo a disposizione un modello matematico e dati in ingresso. Uno scambio di ruolo tra ciò che intendiamo per dato in ingresso e solusione in uscita fa nascere un nuovo problema che chiameremo "problema inverso". Questo esempio chiarirà i termini della questione.
Supponiamo di avere a disposizione un trasduttore che, posto a contatto con una sorgente, restituisca un opportuno valore numerico α. Considerando quest'ultimo come problema diretto, il problema inverso si pone nel modo seguente: " se α è il valore restituito dal trasduttore, in quale configurazione si trova la sorgente ? ".
Generalmente in tutti i casi in cui si ha una sorgente che dà luogo a un campo (particelle che emettono onde elettromagnetiche, una corda che vibrando produce onde sonore , ecc.), per problema inverso si intende la determinazione di informazioni sulle sorgenti avendo a disposizione informazioni sul campo generato.
Generalmente i problemi inversi sono mal posti; vediamo perchè.
Consideriamo un modello matematico riassunto da un operatore lineare $T : X \longrightarrow Y$ con X, Y spazi di Hilbert.
Per problema diretto intendiamo la determinazione, a partire dal dato $x \in X$, del risultato $y = T(x)$.
Supponiamo ulteriormente che l'operatore T sia compatto, con sistema singolare $\{\mu_n; u_n, v_n\}$ (cfr. appendice A), ossia

$$Tx = \sum_{n=1}^{\infty} \mu_n (x, u_n)_X v_n \quad \forall x \in X$$

Il problema inverso sarà la determinazione del valore x la cui immagine secondo T sia y. Utilizziamo l'espressione dell'inversa generalizzata calcolata nel capitolo precedente:

$$T^\dagger y = \sum_{n=1}^{\infty} \frac{1}{\mu_n}(y, v_n)_Y u_n$$

Si può osservare che, mentre nel problema diretto i coefficienti $\mu_n \longrightarrow 0$ per $n \to +\infty$, nel problema inverso i corrispondenti coefficienti ${\mu_n}^{-1}$ divergono.
Inoltre, rispetto al sistema ortonormale v_n , l'errore che perturba il dato y ha componenti principalmente localizzate in corrispondenza dei valori grandi di n (questo è causato dal fatto che l'errore è solitamente una funzione altamente oscillante e le funzioni singolari più oscillanti sono quelle corrispondenti a valori singolari piccoli; si pensi all'analogia con le armoniche di Fourier ; per maggiori dettagli si consulti [2]). Si ottiene quindi che in un problema inverso le componenti dovute all'errore vengono moltiplicate per i coefficienti ${\mu_n}^{-1}$ divergenti, mentre in un problema diretto le eventuali componenti di errore vengono moltiplicate per valori sempre più piccoli.

È utile osservare che si incontrano problemi mal posti in tutte le branchie della fisica e dell'ingegneria. Una serie di esempi di problemi inversi può essere trovata in [2].

Una generica metodologia che permetta di affrontare un problema non ben posto viene detta "di regolarizzazione".
Come suggerisce il nome, il suo scopo è quello di approssimare la soluzione cercando di fare in modo che essa soddisfi a certe proprietà di regolarità che hanno in genere le soluzioni corrispondenti a dati non perturbati. Una prima analisi di tali metodi viene fatta nel paragrafo seguente.

2.3 Metodi di regolarizzazione

Si consideri nuovamente l'equazione

$$Tx = y \tag{2.3}$$

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

Supponiamo inoltre $R(T)$ non chiuso; abbiamo già visto che l'equazione (2.3) dà così luogo ad un problema mal posto. Cerchiamo di risolvere l'equazione nel senso della soluzione generalizzata. Il teorema 1.9 - (ii) asserisce che $T^\dagger$ è un operatore non limitato. Questa caratteristica comporta la mancata dipendenza continua dai dati.
L'idea che sta alla base dei metodi di regolarizzazione è di definire una famiglia di operatori R_α limitati che approssimino con continuità $T^\dagger$.

Definizione 2.2 *Una famiglia di operatori $\{R_\alpha\}$ dipendenti dal parametro reale $\alpha > 0$, dove $R_\alpha : Y \longrightarrow X \quad \forall \alpha > 0$ è detto* **algoritmo regolarizzante** *per il problema (2.3) nel senso della soluzione generalizzata se sono soddisfatte le condizioni seguenti:*

(i) R_α è lineare e continuo $\quad \forall \alpha > 0$

(ii) $\forall y \in R(T) \oplus R(T)^\perp \qquad \lim_{\alpha \to 0} R_\alpha y = x^\dagger \qquad$ *dove* $x^\dagger = T^\dagger y$

La definizione precedente considera esclusivamente operatori R_α lineari. Questa condizione semplifica parecchio lo studio successivo, ma non riveste carattere di necessità (esistono infatti metodi di regolarizzazione basati su famiglie di operatori non limitati). A tal proposito si osservi che nel capitolo successivo verrà enunciata una definizione di caratterepiù generale, che non richiede l'ipotesi di linearità (cfr. definizione 3.6).

Come si può vedere dalla definizione precedente, ogni algoritmo regolarizzante converge alla soluzione generalizzata qualora questa ésista, ossia nel caso in cui $y \in D(T^\dagger)$. È importante però sottolineare che il dominio di ogni operatore R_α è tutto l'insieme Y. Nel caso in cui il dato sia affetto da errore può succedere che y non appartenga a $D(T^\dagger)$. In questo caso l'inversa generalizzata non ha alcuna utilità, mentre l'uso di un algoritmo regolarizzante permette di ottenere una soluzione approssimata del dato non perturbato.

L'idea centrale nella realizzazione pratica di un metodo di regolarizzazione, cioè nella definizione di come deve essere fatta la famiglia $\{R_\alpha\}$, è quella di utilizzare informazioni aggiuntive che permettano di gestire il problema mal posto in una maniera più "regolare". Queste informazioni vanno dedotte dalla formulazione del problema. Ad esempio si può fare in modo che le soluzioni approssimate soddisfino a certe proprietà tipiche delle soluzioni corrispondenti a dati "esatti" (si supponga di avere a disposizione l'ampiezza massima E delle soluzioni corrispondenti a dati non perturbati, $\|x^\dagger\| \leq E$).
Nei paragrafi successivi verranno studiate alcune particolari famiglie regolarizzanti $\{R_\alpha\}$; qui procediamo ancora in maniera del tutto generale.
Utilizzando la rappresentazione sprettrale dell'operatore T^*T (cfr. Appendice B) , ogni algoritmo regolarizzante ammette una rappresentazione semplice.
Sappiamo che $x^\dagger$ è una soluzione nel senso dei minimi quadrati, quindi $T^*Tx^\dagger = T^*y$.
Nel caso in cui T^*T è invertibile si ha chiaramente $x^\dagger = (T^*T)^{-1}T^*y$.
Sappiamo che un algoritmo regolarizzante R_α "approssima" la soluzione $x^\dagger$. Scriviamo allora

$$x^\dagger{}_\alpha = \tilde{R}_\alpha(T^*T)T^*y \tag{2.4}$$

Qui $\tilde{R}_\alpha$ è una funzione reale definita su $\sigma(T^*T) \subseteq [0, \|T\|^2]$, continua e $\tilde{R}_\alpha(T^*T)$ va letta nel contesto della risoluzione spettrale.
La famiglia di funzioni $\{\tilde{R}_\alpha\}$ ha il ruolo di approssimare, per $\alpha \longrightarrow 0$, la funzione $f(t) = t^{-1}$. Il teorema seguente aiuta a chiarire il significato dell'operatore $\tilde{R}_\alpha(T^*T)$ e stabilisce condizioni sufficienti affinchè $x^\dagger{}_\alpha \longrightarrow x^\dagger$ per $\alpha \longrightarrow 0$.
Il teorema è enunciato nel caso in cui l'operatore è compatto; per una formulazione generale si consulti [13] - teorema (3.2.2) .

Teorema 2.1 *Si consideri il problema (2.3) , con T compatto. Sia $\{\tilde{R}_\alpha\}_{\alpha>0}$ una famiglia di funzioni continue di variabile reale, definite su $[0, \|T\|^2]$.*
Si supponga inoltre che

(i) $\tilde{R}_\alpha(t) \longrightarrow t^{-1} \quad$ *per* $\alpha \longrightarrow 0$, $\forall t \in [0, \|T\|^2]$

(ii) $(t\tilde{R}_\alpha(t))$ *sia uniformemente limitato*

Allora, se $y \in D(T^\dagger)$, si ha

$$x^\dagger{}_\alpha = \tilde{R}_\alpha(T^*T)T^*y \longrightarrow x^\dagger = T^\dagger y \qquad \text{per } \alpha \longrightarrow 0 \tag{2.5}$$

Dimostrazione. Si osservi che, se P è un polinomio reale, si ha $P(T^*T)T^* = T^*P(TT^*)$; la verifica può essere fatta per sostituzione diretta. Una generalizzazione di questo risultato, dovuta a Weierstrass, assicura che questo è valido per ogni funzione reale e continua definita su $\sigma(T^*T) = \sigma(TT^*)$. Si ottiene

$$x^\dagger{}_\alpha = \tilde{R}_\alpha(T^*T)T^*y = T^*\tilde{R}_\alpha(TT^*)y \in R(T^*) \tag{2.6}$$

Possiamo allora sviluppare $x^\dagger{}_\alpha$ rispetto al sistema singolare $\{\mu_n; u_n, v_n\}$ dell'operatore T.

$$x^\dagger{}_\alpha = \sum_{n=1}^{\infty} \tilde{R}_\alpha({\mu_n}^2)(T^*y, u_n)u_n = \sum_{n=1}^{\infty} \tilde{R}_\alpha({\mu_n}^2)(y, Tu_n)u_n$$

$$= \sum_{n=1}^{\infty} \tilde{R}_\alpha(\mu_n{}^2)\mu_n(y,v_n)u_n = \sum_{n=1}^{\infty} \mu_n{}^2\tilde{R}_\alpha(\mu_n{}^2)\mu_n^{-1}(y,v_n)u_n$$

Le ipotesi $(i),(ii)$ garantiscono che $\mu_n{}^2\tilde{R}_\alpha(\mu_n{}^2) \longrightarrow 1$, quindi, ricordando la rappresentazione di inversa generalizzata per operatori compatti vista nel capitolo precedente, si ottiene

$$x^\dagger{}_\alpha \longrightarrow \sum_{n=1}^{\infty} \mu^{-1}(y,v_n)u_n = x^\dagger$$

■

Il teorema 2.1 garantisce la convergenza all'inversa generalizzata nel caso in cui $y \in D(T^\dagger)$. Inoltre, per la continuità della funzione $\tilde{R}_\alpha$, l'operatore $\tilde{R}_\alpha(T^*T)T^*$ è lineare e continuo (cfr. appendice B). Si ha così che questo metodo, basato sulla famiglia di funzioni continue $\{\tilde{R}_\alpha\}_{\alpha>0}$ rientra nella categoria degli algoritmi regolarizzanti lineari. Si osservi che gli algoritmi che considereremo nei prossimi paragrafi ammettono una rappresentazione di questo tipo, pur di scegliere un'opportuna famiglia $\{\tilde{R}_\alpha\}$.
Alla luce di quest'ultima considerazione, spesso indicheremo con la stessa notazione R_α , sia l'algoritmo regolarizzante della definizione 2.2 , che l'operatore (2.4); risulterà dal contesto a cosa ci si riferisce.
Il significato della condizione (i) è stato accennato. La condizione (ii) è una condizione di regolarità sulla famiglia di funzioni. Vedremo nel prossimo paragrafo che la famiglia di funzioni che definisce la regolarizzazione di Tikhonov verifica queste due condizioni.

Abbiamo già detto che nella maggior parte dei casi di interesse pratico y non appartiene a $D(T^\dagger)$. In questo caso la soluzione ottenuta mediante l'algoritmo regolarizzante diverge.

Teorema 2.2 *Nelle ipotesi del teorema precedente, se y non appartiene a $D(T^\dagger)$, allora, scelta qualsiasi successione $\{\alpha_n\}_{n\in\mathbb{N}}$, con $\alpha_n \longrightarrow 0$ per $n \longrightarrow +\infty$, si ha che $R_{\alpha_n}(T^*T)T^*y$ non converge (puntualmente).*

Dimostrazione. Sia P l'operatore di proiezione su $\overline{R(T)} = N(T^*)^\perp$.
Supponiamo per assurdo che si abbia la convergenza, cioè $\exists z \in X$ t.c.
$R_{\alpha_n}(T^*T)T^*Py = R_{\alpha_n}(T^*T)T^*y \longrightarrow z$ per $n \longrightarrow +\infty$.
Allora
$TR_{\alpha_n}(T^*T)T^*y = TT^*R_{\alpha_n}(TT^*)y = TT^*R_{\alpha_n}(TT^*)Py.$
Ma per le ipotesi (i) e(ii) del teorema precedente si ha che
$TT^*R_{\alpha_n}(TT^*)Py \longrightarrow Py,$ per $n \longrightarrow +\infty$.

Quindi $Py = Tz$, e necessariamente $y \in D(T^\dagger)$,
che è in contraddizione con l'ipotesi.

■

Corollario 2.1 *Se y non appartiene a $D(T^\dagger)$, allora*

$$\lim_{\alpha\to 0}\|R_\alpha(T^*T)T^*y\| = +\infty$$

Il corollario segue dal fatto che ogni successione non divergente (quindi limitata) in uno spasio di Hilbert ammette sottosuccessione convergente, mentre il teorema precedente assicura la divergenza ad ogni successione.
Gli ultimi due teoremi enunciati mostramo che la condizione $y \in D(T^\dagger)$ è necessaria e sufficiente affinchè si abbia la convergenza di un algoritmo di tipo (2.4). Si intuisce così il comportamento e l'uso di un algoritmo regolarizzante nel caso di un dato perturbato $y = y' + y''$ con $y' \in D(T^\dagger)$, $y'' \in Y\backslash D(T^\dagger)$.
Utilizzare l'operatore R_α, con α piccolo, produce una approssimazione di $T^\dagger$ troppo "fine", la quale, anche se riproduce bene la soluzione corrispondente al dato y' , comporta un grosso errore a causa della divergenza della componente dovuta a y''.
Occorre quindi scegliere un opportuno parametro $\alpha > 0$, tale che l'operatore corrispondente approssimi in maniera soddisfacente $T^\dagger$, ma nello stesso tempo non risenta dell'errore sul dato. Questo discorso, ora accennato, verrà chiarito nel resto del capitolo e riassume il significato della regolarizzazione dei problemi mal posti.

2.4 Regolarizzazione di Tikhonov

Nella risoluzione di problemi mal posti, le strategie comunemente usate possono essere raggruppate in due categorie.
La prima consiste nel considerare unicamente soluzioni appartenenti a una particolare classe (ad esempio una sfera di raggio E). Tra queste soluzioni si cercherà quella che meglio approssima i dati, ossia quella che rende minima la distanza $\|Tx - y\|$.
La seconda metodologia è in un certo senso duale della precedente. Si tratta di cercare tra la soluzioni che approssimano il dato entro un margine di errore ϵ prefissato, la più "regolare" (ad esempio quella di norma minima, oppure , nel caso di spazi di funzioni, quella meno oscillante). Una particolare famiglia di operatori regolarizzanti proposta da Tikhonov (cfr. [42]) permette di gestire in maniera precisa entrambi i metodi.
Tikhonov cerca una soluzione che soddisfi entrambi i vincoli

(1) $\|Tx - y\| \le \epsilon$

(2) $\|x\| \le E$

minimizzando il funzionale, dipendente dal parametro α ,

$$\Phi_\alpha[x,y] = \|Tx - y\|_Y^2 + \alpha\|x\|_X^2 \tag{2.7}$$

Si osservi innanzitutto che non è sempre possibile ottenere soluzioni che soddisfino i due vincoli; si consideri ad esempio $X = Y$, $T = I$, $\|y\| = 3E$, $\epsilon < E$.
Lo studio del funzionale $\Phi_\alpha[x,y]$ permette di stabilire l'esistenza di una soluzione e, nel caso affermativo, di determinarla.

Teorema 2.3 *Per ogni $y \in Y$, $\alpha > 0$ il funzionale $\Phi_\alpha[x,y]$ ammette uno e un solo punto di minimo $x_\alpha \in N(T)^\perp$, soluzione dell'equazione di Eulero*

$$(T^*T + \alpha I)x_\alpha = T^*y \tag{2.8}$$

Dimostrazione. Supponiamo che x_α sia un punto di minimo di $\Phi_\alpha[x,y]$.
Allora $\forall w \in X$, ponendo $f(t) = \Phi_\alpha[x_\alpha + tw, y]$, si ha $f'(0) = 0$.
Si osservi che per la continuità e la linearità dell'operatore T, $f(t)$ risulta continua e derivabile. Inoltre

$$\begin{aligned} f(t) &= \Phi_\alpha[x_\alpha + tw, y] = \|Tx_\alpha - y + tTw\|_Y^2 + \alpha\|x_\alpha + tw\|_X^2 \\ &= \|Tx_\alpha - y\|_Y^2 + t^2\|Tw\|_Y^2 + 2t(Tx_\alpha - y, Tw)_Y + \alpha\|x_\alpha\|_X^2 + t^2\alpha\|w\|_X^2 + 2t\alpha(x_\alpha, w)_X \\ &= f(0) + 2t(T^*Tx_\alpha - T^*y + \alpha x_\alpha, w)_X + t^2(\|Tw\|_Y^2 + \alpha\|w\|_X^2). \end{aligned}$$

Si ha $f'(0) = 0 \Leftrightarrow (T^*Tx_\alpha - T^*y + \alpha x_\alpha, w)_X = 0$.
L'unico elemento ortogonale ad ogni $w \in X$ è l'elemento nullo, quindi se x_α è punto di

minimo , allora
$T^*Tx_\alpha - T^*y + \alpha x_\alpha = 0$ o, equivalentemente,
$(T^*T + \alpha I)x_\alpha = T^*y$.
Dimostriamo adesso che tale punto di minimo esiste sempre ed è unico.
Ciò segue dal fatto che l'operatore $T^*T + \alpha I$ è definito positivo (i suoi autovalori sono contenuti nell'insieme $[\alpha, \|T\|^2 + \alpha]$) e quindi ha inverso continuo. Ne segue l'esistenza e l'unicità di $x_\alpha = (T^*T + \alpha I)^{-1}T^*y$.
Concludiamo dimostrando che tale soluzione x_α appartiene a $N(T)^\perp$.
Come già visto nella dimostrazione del teorema 2.1 , per ogni funzione reale e continua $h(t)$, $h(T^*T)T^* = T^*h(TT^*)$.
Ponendo $h(s) = (s+\alpha)^{-1}$, $s \in \mathbb{R}, s > 0$, si ottiene
$(T^*T + \alpha I)^{-1}T^* = T^*(TT^* + \alpha I)^{-1}$.
Abbiamo così mostrato che $x_\alpha \in R(T^*) \subseteq N(T)^\perp$. ∎

Nel prossimo paragrafo vedremo come viene trattato il problema dei vincoli esposto all'inizio, utilizzando il funzionale $\Phi_\alpha[x, y]$ (e quindi l'equazione di Eulero (2.8)) .
Dimostriamo adesso che l'equazione di Eulero , al variare di $\alpha > 0$, definisce un algoritmo regolarizzante e studiamone la convergenza.
Utilizzando le notazioni introdotte nel paragrafo precedente, consideriamo la funzione reale

$$\tilde{R}_\alpha(t) = (t+\alpha)^{-1} \tag{2.9}$$

Si osservi che

(i) fissato $\alpha > 0$
$\tilde{R}_\alpha$ è continua su $[0, +\infty)$

(ii) $\tilde{R}_\alpha(t) \longrightarrow t^{-1}$ $(\alpha \longrightarrow 0)$

(iii) $|t\tilde{R}_\alpha| = |\frac{t}{t+\alpha}|$ è uniformemente limitata, infatti

$|\frac{t}{t+\alpha}| = \frac{1}{1+\frac{\alpha}{t}} \leq 1$, $\quad t > 0$.

Allora la famiglia di operatori $\{R_\alpha\}_{\alpha>0}$, con

$$R_\alpha = \tilde{R}_\alpha(T^*T)T^* = (T^*T + \alpha I)^{-1}T^* \tag{2.10}$$

definisce un algoritmo regolarizzante lineare detto **algoritmo regolarizzante di Tikhonov del primo ordine.**
I risultati del paragrafo precedente sulla convergenza alla soluzione generalizzata garantiscono che

(i) $x^\dagger{}_\alpha = R_\alpha y \longrightarrow x^\dagger = T^\dagger y$, se $y \in R(T) \oplus R(T)^\perp$

(ii) $\{x^\dagger{}_\alpha\}$ diverge se y non appartiene a $R(T) \oplus R(T)^\perp$.

È possibile anche ottenere una maggiorazione sull'ordine di convergenza dell'algoritmo (2.10). A questo scopo si osservi che nel teorema 2.1 , condizione sufficiente alla convergenza è che $y \in R(T) \oplus R(T)^\perp$, ossia $Py \in R(T)$, dove P è l'operatore di proiezione di Y su $\overline{R(T)}$. È ragionevole pensare che più il dato soddisfi a certe condizioni di regolarità, più la successione $\{x^\dagger{}_\alpha\}$ converga rapidamente alla soluzione.
Vediamo infatti che la condizione $Py \in R(T(T^*T)^\gamma)$ $0 < \gamma \leq 1$, permette di stimare il raggio di convergenza. Si osservi che gli insiemi $R(T(T^*T)^\gamma)$, al crescere di γ , formano una famiglia di insiemi decrescenti, ognuno costituito da elementi sempre più regolari.

Teorema 2.4 *Se* $\exists w \in X$ *tale che* $Py = T(T^*T)^\gamma w$, $0 < \gamma \leq 1$, *allora*

$$\|T^\dagger y - x^\dagger{}_\alpha\| \leq \alpha^\gamma \|w\|$$

dove $\{x^\dagger{}_\alpha\}$ *è la successione determinata con l'algoritmo di Tikhonov* (2.10).

Prima di procedere, enunciamo due risultati (lemma 2.1 e 2.2) che saranno utilizzati nel corso della dimostrazione del teorema.

Lemma 2.1 $R((T^*T)^\gamma) \subseteq N(T)^\perp$

Dimostrazione. Per semplicità, verifichiamo la relazione solo nel caso in cui l'operatore T è compatto.
Utilizziamo il sistema singolare $\{\mu_n; u_n, v_n\}$ per l'operatore T; sia $x \in X$.
Allora
$(T^*T)^\gamma x = \sum_{n=1}^{\infty} \mu_n{}^{2\gamma}(x, u_n)_X u_n$.
Inoltre $\{u_n\}$ è un sistema ortonormale completo per $\overline{R(T^*)}$.
Quind $(T^*T)^\gamma x \in \overline{R(T^*)} = N(T)^\perp$. ∎

Lemma 2.2 *Si consideri la famiglia di funzioni reali* $\{\tilde{R}_\alpha\}$ (2.9), $\tilde{R}_\alpha(t) = (t+\alpha)^{-1}$, $\alpha > 0$.
Allora, fissato $\gamma \in (0, 1]$,

$$t^\gamma(1 - t\tilde{R}_\alpha(t)) < \alpha^\gamma$$

per ogni $t \in [0, +\infty)$.

Dimostrazione. Si osservi che, essendo $t \geq 0$ e $\alpha > 0$, segue

$$\begin{aligned} t^\gamma|1 - tR_\alpha(t)| &= t^\gamma\left|1 - \frac{t}{t+\alpha}\right| \\ &= \frac{t^\gamma\alpha}{t+\alpha} \\ &= \frac{1}{t^{1-\gamma}\alpha^{-1} + t^{-\gamma}} \end{aligned}$$

Quindi :

(i) Se $0 < t \leq \alpha$

$$\frac{1}{t^{1-\gamma}\alpha^{-1} + t^{-\gamma}} \leq \frac{1}{t^{-\gamma}} = t^\gamma \leq \alpha^\gamma$$

(ii) Se $t > \alpha$

$$\frac{1}{t^{1-\gamma}\alpha^{-1} + t^{-\gamma}} \leq \frac{1}{t^{1-\gamma}\alpha^{-1}} = t^{\gamma-1}\alpha \leq \alpha^{\gamma-1}\alpha = \alpha^\gamma$$

∎

Possiamo ora dimostrare il teorema 2.4 .

Dimostrazione. **(teorema 2.4)**

$$\begin{aligned} x^\dagger{}_\alpha &= \tilde{R}_\alpha(T^*T)T^*y = \tilde{R}_\alpha(T^*T)T^*(Py + y - Py) \\ &= \tilde{R}_\alpha(T^*T)T^*Py = \tilde{R}_\alpha(T^*T)(T^*T)^{\gamma+1}w \end{aligned}$$

Si osservi che $TT^\dagger y = Py = T(T^*T)^\gamma w$, quindi
$T(T^\dagger y - (T^*T)^\gamma w) = 0$ cioè $T^\dagger y - (T^*T)^\gamma w \in N(T)$.
Dal lemma 2.1 , sappiamo che $(T^*T)^\gamma w \in N(T)^\perp$.
Ricordando che anche $T^\dagger y \in N(T)^\perp$, si ha $T^\dagger y = (T^*T)^\gamma w$.
Quindi

$$\begin{aligned} \|T^\dagger y - x^\dagger{}_\alpha\| &= \|(T^*T)^\gamma w - \tilde{R}_\alpha(T^*T)(T^*T)^{\gamma+1}w\| \\ &= \|(T^*T)^\gamma(I - (T^*T)\tilde{R}_\alpha(T^*T))w\| \\ &\leq \|(T^*T)^\gamma(I - (T^*T)\tilde{R}_\alpha(T^*T))\|\,\|w\| \\ &\leq \alpha^\gamma\|w\|. \end{aligned}$$

Per l'ultima maggiorazione si è utilizzato il lemma 2.2 (si consideri come al solito la risoluzione spettrale dell'operatore T^*T, i cui autovalori sono contenuti in $[0, \|T\|^2] \subset [0, +\infty)$) .

Il teorema ora dimostrato asserisce che l'algoritmo regolarizzante di Tikhonov, nel caso migliore contemplato, converge almeno con ordine α . Questo si ottiene se $\gamma = 1$, cioè se $T^\dagger y \in R(T^*T)$.
Si può dimostrare che in realtà $O(\alpha)$ è il miglior raggio di convergenza ottenibile dall'algoritmo di Tikhonov e si ottiene necessariamente nel caso in cui $T^\dagger y \in R(T^*T)$.

La convergenza dell'algoritmo di Tikhonov non è la proprietà più importante per la risoluzione di problemi inversi. Più che sapere se e come un algoritmo converge nel caso di un dato y, è importante conoscere le qualità dell'algoritmo nel caso in cui si abbiano a disposizione dati affetti da errore e quindi non conosciuti esattamente. Vedremo nel prossimo paragrafo che l'algoritmo di Tikhonov garantisce soluzioni affidabili (stabili e sufficientemente approssimate) anche in questo caso.

2.5 Soluzioni vincolate , approssimate regolarizzate e metodo della discrepanza

Si considerino le due funzioni introdotte nel paragrafo precedente

(1) $\alpha \longrightarrow \epsilon(\alpha, y) = \|Tx_\alpha - y\|_Y$

(2) $\alpha \longrightarrow E(\alpha, y) = \|x_\alpha\|_X$

La prima distanza, detta **discrepanza**, permette di stimare l'accuratezza della soluzione x_α , ossia la precisione con cui approssima il dato; qualora l'inversa generalizzata esista, si ottengono valori piccoli di $\epsilon(\alpha, y)$ in corrispondenza di soluzioni prossime a $T^\dagger y$.
La seconda in un certo senso misura la "regolarità" della soluzione x_α ; se ad esempio $X = C^1(I)$ e $\|x\|_X{}^2 = \int_I (x'(t))^2 dt$, soluzioni poco oscillanti corrispondono a valori piccoli di $E(\alpha, y)$.
Nel caso dell'algoritmo di Tikhonov del primo ordine si possono avere utili indicazioni sull'andamento di queste due funzioni.

Teorema 2.5 *Per ogni $y \in Y$, considerando l'algoritmo regolarizzante di Tikhonov 2.9 ,*

(i) la funzione $(\epsilon(\alpha, y))^2$ è monotona non decrescente a valori in $(\|y^\perp\|^2, \|y\|^2)$, dove $y^\perp = (I - P)y$ con P operatore di proiezione ortogonale di Y su $R(T)$

(ii) la funzione $(E(\alpha, y))^2$ è monotona non crescente, a valori in $(0, \|T^\dagger y\|^2)$ se $y \in R(T) \oplus R(T)^\perp = D(T^\dagger)$, o a valori in $(0, +\infty)$ se $y \notin R(T) \oplus R(T)^\perp$

Dimostrazione. Dalla definizione si ha

$$x_\alpha = R_\alpha y = (T^*T + \alpha I)^{-1}T^*y .$$

Si osservi che $R_\alpha y = R_\alpha P_{N(T^*)^\perp} y = R_\alpha P_{\overline{R(T)}} y$. Inoltre, come si è visto nel teorema 2.3 , l'algoritmo di Tikhonov si può scrivere nella forma equivalente

$$x_\alpha = T^*(TT^* + \alpha I)^{-1}y$$

D'ora in poi spesso indicheremo con lo stesso simbolo sia il prodotto scalare in X che quello in Y, lo stesso dicasi per le norme. Possiamo allora scrivere:

$$\begin{aligned} E^2(\alpha, y) &= \|x_\alpha\|_X{}^2 = \|(T^*T + \alpha I)^{-1}T^*y\|^2 \\ &= \|(T^*T + \alpha I)^{-1}T^*Py\|^2 = \|T^*(TT^* + \alpha I)^{-1}Py\|^2 \\ &= (TT^*(TT^* + \alpha I)^{-2}Py, Py) \end{aligned}$$

$$\begin{aligned} \epsilon^2(\alpha, y) &= \|Tx_\alpha - y\|_Y{}^2 = \|TR_\alpha y - y\|^2 \\ &= \|TR_\alpha Py - Py - (I - P)y\|^2 \\ &= \|(TR_\alpha - I)Py\|^2 + \|(I - P)y\|^2 \\ &= \|(TT^* - (TT^* + \alpha I))(TT^* + \alpha I)^{-1}Py\|^2 + \|y^\perp\|^2 \\ &= \alpha^2\|(TT^* + \alpha I)^{-1}Py\|^2 + \|y^\perp\|^2 \end{aligned}$$

Indicando con $\{E_\lambda\}_{\lambda \geq 0}$ la famiglia spettrale associata all'operatore TT^* (cfr. appendice B) e ricordando che $\sigma(T^*T) \subset [\,0, \|T^*T\|\,] = [\,0, \|T\|^2\,]$ (cfr. [15] , teorema 3.3.1) , si ottengono le relazioni seguenti

$$E^2(\alpha, y) = \int_0^{\|T\|^2} \frac{\lambda}{(\lambda + \alpha)^2} d\|E_\lambda Py\|_Y^2$$

$$\epsilon^2(\alpha, y) = \alpha^2 \int_0^{\|T\|^2} \frac{1}{(\lambda + \alpha)^2} d\|E_\lambda Py\|_Y^2 \;+\; \|y^\perp\|^2$$

Consideriamo le funzioni

$$\xi(\alpha, \lambda) = \frac{\lambda}{(\lambda + \alpha)^2} \leq \lambda^{-1}$$

e

$$\psi(\alpha, \lambda) = \frac{\alpha^2}{(\lambda + \alpha)^2} \leq 1$$

definite su $\{(0, +\infty) \times (0, \|T\|^2)\}$.

Fissato $\bar{\lambda} \in (0, \|T\|^2)$,
$\xi(\alpha, \bar{\lambda})$ è monotona decrescente dal valore $\bar{\lambda}^{-1}$ per $\alpha = 0$, al valore 0 per $\alpha \longrightarrow +\infty$,
$\psi(\alpha, \bar{\lambda})$ è monotona crescente dal valore 0 per $\alpha = 0$, al valore 1 per $\alpha \longrightarrow +\infty$.
Inoltre

$$(a) \quad \int_0^{\|T\|^2} \lambda^{-1} d\|E_\lambda P y\|_Y{}^2 = \|T^\dagger y\|^2$$

$$(b) \quad \int_0^{\|T\|^2} 1 d\|E_\lambda P y\|_Y{}^2 = \|P y\|^2$$

dove nel caso in cui $y \in R(T) \oplus R(T)^\perp$, $T^\dagger y$ rappresenta l'inversa generalizzata, mentre nel caso in cui y non appartiene a $R(T) \oplus R(T)^\perp$, con abuso di linguaggio indichiamo $\|T^\dagger y\| = +\infty$ (si confronti il corollario 2.1).

Utilizzando il teorema di convergenza dominata per le famiglie di funzioni reali $\Xi_\alpha(\lambda) = \xi(\alpha, \lambda)$ e $\Psi_\alpha(\lambda) = \psi(\alpha, \lambda)$, la monotonia rispetto a α delle funzioni $\xi(\alpha, \lambda)$ e $\psi(\alpha, \lambda)$ e i limiti $(a), (b)$ si ottiene la tesi. ∎

Rappresentiamo graficamente i risultati del teorema.

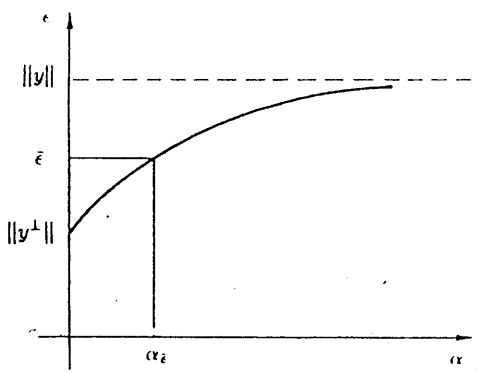
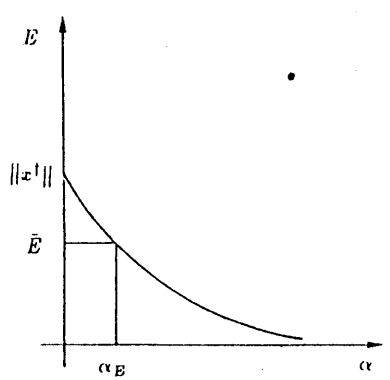
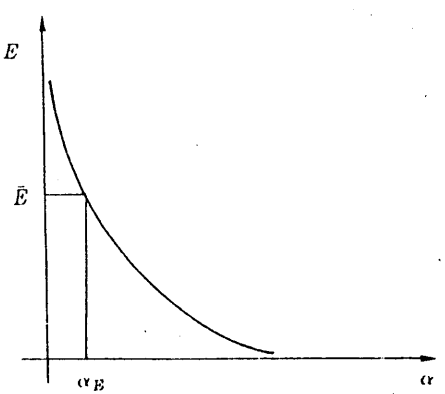


Figura A



$y \in R(T) \oplus R(T)^\perp$



$y \notin R(T) \oplus R(T)^\perp$

La monotonia delle funzioni ϵ e E rende particolarmente semplice il trattamento dei due problemi introdotti all'inizio del paragrafo precedente:

(1) Tra le soluzioni x che approssimano il dato a livello ϵ (ossia $\|Tx - y\| \leq \epsilon$), determinare la più regolare ($\|x\|$ minimo)

(2) Tra le soluzioni x soddisfacenti a certe condizioni di regolarità (ossia $\|x\| \leq E$), determinare quella che meglio approssima il dato ($\|Tx - y\|$ minimo).

Nel caso (1) , supponendo $\|y^\perp\| \leq \epsilon \leq \|y\|$, i valori del parametro di regolarizzazione α per cui $\|Tx_\alpha - y\| \leq \epsilon$, appartengono all'insieme $I_1 = (0, \alpha_\epsilon)$ dove α_ϵ è il valore tale che $\|Tx_{\alpha_\epsilon} - y\| = \epsilon$, (vedi fig. A) .
Nell'insieme I_1 , la soluzione più regolare si ottiene per $\alpha = \alpha_\epsilon$; (vedi fig. B) .

Nel caso (2) , supponendo $E < \|x^\dagger\|$, i valori del parametro di regolarizzazione α per cui $\|x_\alpha\| \leq E$ appartengono all'insieme $I_2 = (\alpha_E, \infty)$, dove α_E è il valore tale che $\|x_{\alpha_E}\| = E$ (vedi fig. B) .
Nell'insieme I_2 , la soluzione che meglio approssima il dato si ottiene per $\alpha = \alpha_E$, (vedi fig. A) .

Considerando la osservazioni fin qui svolte, la risoluzione del problema (1) può essere affrontata nel seguente modo.
Per ogni valore del parametro di regolarizzazione $\alpha > 0$, utilizzando l'operatore R_α (2.10), si calcoli il minimo x_α del funzionale $\Phi_\alpha[x, y]$. Si determini quindi il valore $\bar{\alpha}$ tale che

$$\|Tx_{\bar{\alpha}} - y\| = \epsilon$$

Questo valore esiste ed è unico se $\|y^\perp\| < \epsilon < \|y\|$.
La soluzione $x_{\bar{\alpha}}$ ottenuta è quella di norma minima tra tutte quelle ammesse.

Per quanto riguarda il problema (2) si procede nel seguente modo. Per ogni valore del parametro $\alpha > 0$, si calcoli come nel caso precedente la soluzione regolarizzata $R_\alpha y$. Si determini poi il valore $\bar{\alpha}$ tale che $\|x_{\bar{\alpha}}\| = E$. Questo valore esiste ed è unico se $E < \|x^\dagger\|$.
La soluzione $x_{\bar{\alpha}}$ ottenuta è quella che meglio approssima il dato tra quelle ammesse.

La soluzione del problema (1) è detta soluzione approssimata regolarizzata ed è la soluzione del problema variazionale

$$\|x_\alpha\|_X = \inf\{\|x\|_X \ : \ \|Tx - y\|_Y \leq \epsilon\}$$

La soluzione del problema (2) è detta soluzione vincolata a livello E ed è la soluzione del problema variazionale

$$\|Tx_\alpha - y\|_Y = \inf\{\|Tx - y\|_Y \ : \ \|x\|_X \leq E\}$$

Nel caso in cui vi siano indicazioni sia sull'errore che sulla norma della soluzione, cioè nel caso in cui esistano e siano ammesse soluzioni tali che $\|Tx - y\| \leq \epsilon$ e $\|x\| \leq E$, si può scegliere direttamente $\bar{\alpha} = \left(\frac{\epsilon}{E}\right)^2$. Questa scelta, che va sotto il nome di metodo di Miller, ha valide basi teoriche. La soluzione che si ottiene è un buon compromesso tra approssimazione del dato e regolarità. Ovviamente tale soluzione ha una norma maggiore

di quella del problema (1) e una discrepanza maggiore di quella del problema (2) . Per una ampia trattazione si può consultare [2] o [29] .

È già stato più volte ricordato che nei problemi concreti il dato a disposizione è contaminato da una componente di errore detto noise dovuto a cause sistematiche (tra le quali la precisione con cui l'elaboratore maneggia i dati) e accidentali. In pratica non abbiamo mai a disposizione il dato $y \in Y$ ma una sua approssimazione e, nei casi migliori, una stima dell'errore massimo con cui viene fornito il dato. Nella parte restante di questo paragrafo vedremo come i risultati ottenuti per i problemi (1) e (2) , i quali sono stati formulati nella ipotesi di conoscere "esattamente" il dato $y \in Y$, siano la base su cui lavorare per affrontare il problema concreto.

Indichiamo con y^δ una approssimazione del dato "esatto" $y \in R(T)$, e con y^δ_{err} la componente in $CR(T)$ di errore sul dato.

$$y^\delta = y + y^\delta_{err}$$
$$\|y - y^\delta\| = \|y^\delta_{err}\| \leq \delta$$

Si tenga presente che considereremo come dati non affetti da errore solo quelli in $R(T)$. Questa scelta viene fatta al fine di rendere più semplice i calcoli (non si dovrà fare continuamente ricorso all'operatore di proiezione su $R(T)$, altrimenti necessario), ma ha una motivazione più forte. Sappiamo che l'operatore T rappresenta una generica macchina che, ricevendo in ingresso un dato x , fornisce in uscita il corrispondente risultato y. Possiamo allora considerare un dato y^δ non appartenente a $R(T)$ come affetto da errore poichè non è tra quelli che la macchina, con un funzionamento corretto, è in grado di restituire.

Vediamo come agisce un generico algoritmo regolarizzante lineare in presenza di un dato perturbato y^δ .
Indicando con $x^\dagger$ la soluzione corrispondente al dato $y \in R(T)$, si ottiene

$$\begin{aligned} x_\alpha = R_\alpha y^\delta &= R_\alpha y + R_\alpha y^\delta_{err} \\ &= R_\alpha T^\dagger x^\dagger + R_\alpha y^\delta_{err} \end{aligned}$$

Valutiamo allora l'andamento della soluzione x_α in funzione del parametro α , confrontandola con la soluzione corrispondente al dato esatto $x^\dagger$.

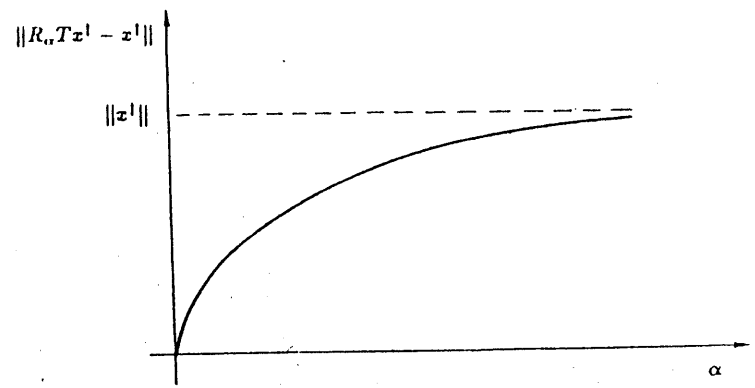
$$\begin{aligned} x_\alpha - x^\dagger &= R_\alpha y^\delta - x^\dagger = R_\alpha T x^\dagger + R_\alpha y^\delta_{err} - x^\dagger \\ &= {}^{*}(R_\alpha T x^\dagger - x^\dagger) + R_\alpha y^\delta_{err} \end{aligned}$$

L'errore che si commette è la somma dei due termini $(R_\alpha T x^\dagger - x^\dagger)$ e $R_\alpha y^\delta_{err}$.
Il primo rappresenta l'errore dovuto all'algoritmo regolarizzante R_α il quale approssima l'operatore $T^\dagger$ (si osservi che questo errore è indipendente dall'errore sul dato). Chiamiamo questa componente **errore di approssimazione** .
Il secondo termine è dovuto alla perturbazione y^δ_{err} e rappresenta come l'errore sul dato si propaga sulla soluzione.

Studiamo, al variare di α , il comportamento dei due addendi. Analogamente a quanto è stato fatto nella dimostrazione precedente, indicando con $\{F_\lambda\}_{\lambda \geq 0}$ la famiglia spettrale associata all'operatore T^*T, si ottiene

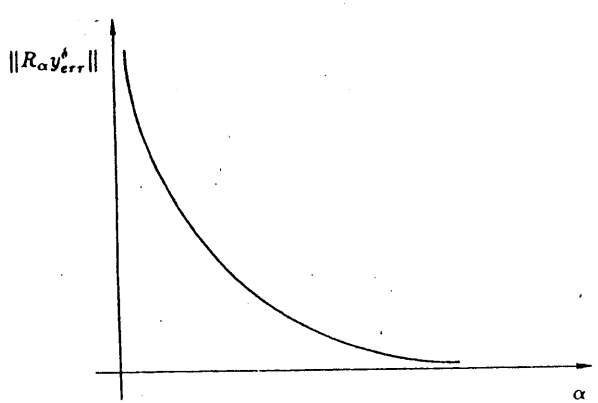
$$\begin{aligned} \|R_\alpha T x^\dagger - x^\dagger\|^2 &= \|(T^*T + \alpha I)^{-1} T^* T x^\dagger - x^\dagger\|^2 = \int_0^{\|T\|^2} \left(\frac{\lambda}{\lambda+\alpha} - 1\right)^2 d\|F_\lambda x^\dagger\|^2 \\ &= \int_0^{\|T\|^2} \frac{\alpha^2}{(\lambda+\alpha)^2} d\|F_\lambda x^\dagger\|^2 \end{aligned}$$

Considerando nuovamente l'andamento della funzione reale $\frac{\alpha^2}{(\lambda+\alpha)^2}$ vista nel teorema precedente, si ha che l'errore di approssimazione è una funzione crescente, dal valore 0 per $\alpha = 0$ al valore $\|x^\dagger\|$ per $\alpha \longrightarrow +\infty$



Errore di approssimazione

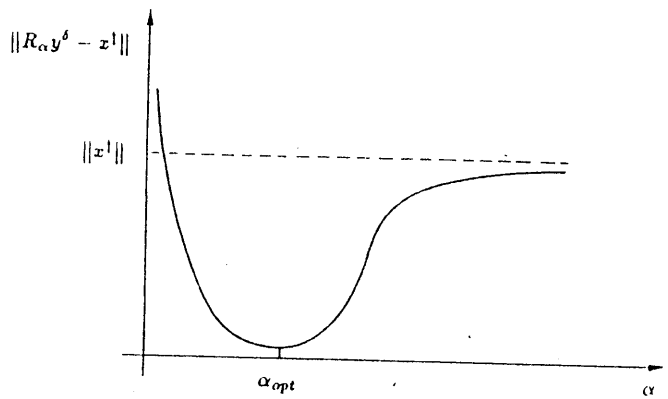
Per quanto riguarda l'errore generato dall'errore sul dato $R_\alpha y^\delta_{err}$, questo può essere visto come se fosse la soluzione ottenuta dal dato "non perturbato" y^δ_{err}. Si ottiene quindi un andamento di $\|R_\alpha y^\delta_{err}\|^2$ analogo a quello di $E = \|x_\alpha\|^2$ visto nel teorema 2.5 , cioè decrescente dal valore $\|T^\dagger y^\delta_{err}\|^2$ (nella maggior parte dei casi $+\infty$ poichè solitamente y^δ_{err} non appartiene a $R(T) \oplus R(T)^\perp$) per $\alpha \longrightarrow 0$, al valore 0 per $\alpha \longrightarrow +\infty$.



Possiamo allora ritornare alla valutazione dell'errore globale

$$\|x_\alpha - x^\dagger\|^2 \leq \|R_\alpha T x^\dagger - x^\dagger\|^2 + \|R_\alpha y^\delta_{err}\|^2$$

Si ottiene un andamento che può essere rappresentato nel seguente modo:



Esiste un parametro di regolarizzazione ottimale in corrispondenza del quale l'operatore R_α fornisce la migliore approssimazione della soluzione $x^\dagger$.
Il funzionamento di un algoritmo regolarizzante si può riassumere nel seguente modo:
" valori piccoli del parametro α riducono l'errore di approssimazione, mentre l'errore prodotto dalla perturbazione sul dato viene controllato da valori grandi di α ; il parametro ottimale , detto α_{opt} rappresenta il miglior compromesso tra approssimazione e propagazione degli errori " .

Dal punto di vista qualitativo, la situazione esposta nella figura precedente si verifica in svariati problemi numerici. Ad esempio nelle approssimazioni per discretizzazioni di problemi ben posti (equazioni differenziali, integrazione numerica , ecc.), esiste un valore ottimale del passo di discretizzazione; discretizzazioni più fini , sebbene possano far supporre una risoluzione più precisa del problema, in realtà comportano un errore maggiore. Le cause di questo fenomeno sono gli errori di calcolo dovuti alla precisione limitata del calcolatore (numero di macchina). Ogni volta che la discretizzazione diventa più fine, la mole di calcoli aumenta come anche l'errore ad essi associato. Se si sceglie un passo di discretizzazione minore di quello ottimale, il miglioramento ottenuto dalla maggiore precisione con cui si maneggia il problema non compensa l'aumento dell'errore dovuto ai calcoli e la soluzione risulta peggiore.
Vi è però una sostanziale differenza tra il valore ottimale del parametro di regolarizzazione e il valore ottimale del passo di discretizzazione di un problema ben posto, infatti l'esistenza di α_{opt} non è dovuta a cause numeriche, ma unicamente alla natura del problema (che è mal posto) e al dato y^δ (che è perturbato). Si osservi che una risoluzione simbolica di un problema ben posto per mezzo di discretizzazione comporta invece risultati migliori per passi sempre più piccoli.

Si intuisce che la risoluzione del problema è ricondotta alla determinazione del parametro di regolarizzazione ottimale.

Analizziamo un importante criterio per la scelta del parametro α proposto da Morozov, che va sotto il nome di metodo della discrepanza (cfr. [32]).
Il criterio di Morozov nasce dall'osservazione intuitiva che, conoscendo il dato a meno di δ, non è ragionevole cercare una soluzione che approssimi il dato con una precisione maggiore di δ .
Il metodo della discrepanza consiste appunto nella scelta del parametro $\overline{\alpha}$ tale che

$$\|Tx_{\overline{\alpha}} - y^\delta\| = \delta \qquad (2.11)$$

Utilizzando i risultati del problema (1) considerato precedentemente, si possono trarre indicazioni sulla soluzione $x_{\overline{\alpha}}$ che rafforzano la validità del criterio.
Riconsideriamo il problema nella forma seguente :
" avendo a disposizione il dato approssimato y^δ , si determini una soluzione approssimata di norma minima ".
In virtù di quanto detto circa la motivazione intuitiva, appare naturale cercare la soluzione di norma minima nell'insieme

$$C_\delta = \{\, x \in X : \|Tx - y^\delta\| \leq \delta \,\}$$

Il problema ora formulato è equivalente al problema (1) (si osservi cha la tolleranza qui indicata con δ , nel problema (1) è indicata con ϵ) . Dall'analisi svolta sul problema (1), si conclude che la soluzione di norma minima si ottiene in corrispondenza del valore del parametro di regolarizzazione $\overline{\alpha}$ tale che

$$\|Tx_{\overline{\alpha}} - y^\delta\|; = \delta$$

La soluzione così ottenuta è quella di norma minima, ossia più regolare, tra quelle che consideriamo affidabili, cioè tra quelle che approssimano il dato con la stessa precisione con cui esso è fornito.
La proprietà di minimo della soluzione $x_{\overline{\alpha}}$, può essere anche verificata direttamente. Si ha il seguente risultato.

Lemma 2.3 *Si consideri l'insieme chiuso e convesso*

$$C_\delta = \{x \in X : \|Tx - y^\delta\| \leq \delta\}$$

e sia $\bar{x}$ l'unico elemento di norma minima in C_δ.
Allora

$$\|T\bar{x} - y^\delta\| = \delta$$

cioè $\bar{x}$ corrisponde alla soluzione $x_{\overline{\alpha}}$ ottenuta col metodo della discrepanza.

Dimostrazione. Supponiamo per assurdo che l'elemento di norma minima $\bar{x}$ non appartenga al bordo di C_δ , cioè

$$\|T\bar{x} - y^\delta\| < \delta$$

Allora esiste $t \in (0,1)$ tale che , per t sufficientemente prossimo a 1 , si ha ancora $\|T(t\bar{x}) - y\| < \delta$, quindi $t\bar{x} \in C_\delta$.
Ma $\|t\bar{x}\| = t\|x\| < \|x\|$, in contraddizione con l'ipotesi.
Quindi necessariamente $\|T\bar{x} - y^\delta\| = \delta$. ∎

La scelta del parametro $\overline{\alpha}$ basata sul criterio di Morozov nella pratica si realizza determinando la soluzione per una successione di valori positivi $\{\alpha_n\}_{n=1}^{N}$. Tra questi valori si sceglie quello tale che $\|Tx_{\alpha_n} - y^\delta\|$ sia più vicino possibile a δ .
La scelta è detta "a posteriori" poichè necessita della determinazione delle varie soluzioni x_{α_n} , ossia dipende dal particolare dato y^δ e non solo dal livello di errore δ.
Si osservi che il metodo della discrepanza risulta particolarmente efficace nel caso di algoritmi regolarizzanti iterativi, poichè la soluzione regolarizzata ad ogni passo è immediatamente disponibile (questo discorso verrà chiarito nel prossimo capitolo).

Nel caso dell'algoritmo di Tikhonov è possibile avere una maggiorazione "a priori" , cioè indipendente dal particolare y^δ, del parametro $\overline{\alpha}$. Questa stima è dovuta a Vinokurov (si consulti [48]).

Teorema 2.6 *Si consideri l'algoritmo reglarizzante di Tikhonov* (2.10) *e sia* $\overline{\alpha} > 0$ *il parametro di regolarizzazione corrispondente al metodo della discrepanza. Si supponga inoltre* $\delta < \|y^\delta\|$.
Allora

$$\overline{\alpha} \leq \frac{\delta \|T\|^2}{\|y^\delta\| - \delta} \tag{2.12}$$

Dimostrazione.

$$\begin{aligned}
\|y^\delta\| - \delta &= \|y^\delta\| - \|Tx_{\overline{\alpha}} - y^\delta\| \leq \|Tx_{\overline{\alpha}}\| \\
&= \|T(T^*T + \overline{\alpha}I)^{-1}T^*y^\delta\| \\
&= \|T(\frac{1}{\overline{\alpha}} - \frac{1}{\overline{\alpha}}\frac{T^*T}{T^*T + \overline{\alpha}I})T^*y^\delta\| \\
&= \frac{1}{\overline{\alpha}}\|T[1 - T^*T(T^*T + \overline{\alpha}I)^{-1}]T^*y^\delta\| \\
&= \frac{1}{\overline{\alpha}}\|TT^*y^\delta - TT^*Tx_{\overline{\alpha}}\| \\
&= \frac{1}{\overline{\alpha}}\|TT^*(y^\delta - Tx_{\overline{\alpha}})\| \\
&\leq \frac{1}{\overline{\alpha}}\|T\|^2\|y^\delta - Tx_{\overline{\alpha}}\| \\
&\leq \frac{1}{\overline{\alpha}}\|T\|^2\delta
\end{aligned}$$

Da cui

$$\overline{\alpha} \leq \frac{\delta \|T\|^2}{\|y^\delta\| - \delta}$$

■

Corollario 2.2 *Si consideri la funzione* $\alpha = \alpha(\delta)$ *dove* $\alpha(\delta)$ *è il valore del parametro di regolarizzazione corrispondente al metodo della discrepanza, cioè*

$$\alpha(\delta) = \{\alpha > 0 \ : \ \|Tx_\alpha - y^\delta\| = \delta\}$$

Allora

$$\alpha(\delta) \longrightarrow 0^+ \quad , per \quad \delta \longrightarrow 0^+$$

di ordine ≥ 1.

La scelta di Morozov non è l'unica che offra buoni risultati. Un metodo alternativo è stato proposto da Arcangeli e consiste nella scelta del parametro α tale che

$$\sqrt{\alpha}\|Tx_\alpha - y^\delta\| = \delta$$

Anche il criterio di Arcangeli, come quello di Morozov, necessita della conoscenza del livello di errore δ. Per un'analisi delle proprietà di questa scelta si consulti [1] .

2.6 Finestre spettrali e regolarizzazione di ordine superiore

L'algoritmo regolarizzante di Tikhonov del primo ordine (2.10) è definito come

$$R_\alpha = (T^*T + \alpha I)^{-1}T^*$$

Supponiamo per il momento che l'operatore T , oltre ad essere lineare e limitato, sia autoaggiunto. Indichiamo con $\{E_\lambda\}_{\lambda \geq 0}$ la famiglia spettrale associata all'operatore T. L'operatore R_α ammette la seguente rappresentazione :

$$\begin{aligned}
R_\alpha &= \int_0^{\|T\|} \frac{\lambda}{\lambda^2 + \alpha} dE_\lambda \\
&= \int_0^{\|T\|} (\frac{\lambda^2}{\lambda^2 + \alpha})\frac{1}{\lambda} dE_\lambda \\
&= \int_0^{\|T\|} W_\alpha(\lambda)\frac{1}{\lambda} dE_\lambda
\end{aligned}$$

dove

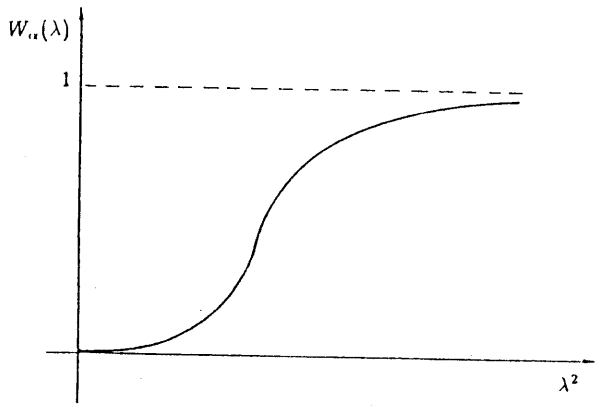
$$W_\alpha(\lambda) = \frac{\lambda^2}{\lambda^2 + \alpha}$$

Ricordando che ,qualora l'operatore autoaggiunto T sia invertibile, l'inversa è data da

$$T^{-1} = \int_0^{\|T\|} \frac{1}{\lambda} dE_\lambda$$

è facile notare che l'operatore di Tikhonov è una versione filtrata dell'operatore inverso T^{-1}. La famiglia di funzioni reali $\{W_\alpha\}_{\alpha > 0}$ definisce una famiglia di finestre spettrali con le seguenti proprietà

- $W_\alpha(\lambda) \simeq 1 \quad$ per $\lambda^2 \gg \alpha$
- $W_\alpha(\lambda) \simeq \frac{\lambda^2}{\alpha} \quad$ per $\alpha \gg \lambda^2$
- $W_\alpha(\lambda) \longrightarrow 0 \quad$ per $\alpha \longrightarrow +\infty$



Essa riduce gli effetti delle componenti corrispondenti agli autovalori più piccoli. Queste componenti sono quelle che risentono maggiormente dell'errore sui dati (si confronti nel paragrafo 2.2 la proprietà analoga per i valori singolari μ_n corrispondenti a n grandi). Proprio in questo senso l'operatore R_α offre soluzioni stabilizzate , meno suscettibili cioè, da perturbazioni sul dato.
Si osservi che la richiesta di considerare T autoaggiunto è sorta al fine di poter utilizzare la rappresentazione spettrale, ma non riveste carattere di necessità.
Ritorniamo al caso di generico operatore lineare limitato.

Si consideri l'operatore di Fredholm di prima specie a nucleo convolutivo

$$T : L^2(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n) \ , \quad k \in L^1(\mathbb{R}^n)$$
$$(Tx)(t) = \int_{\mathbb{R}^n} k(t-s)x(s)ds$$

e l'equazione

$$Tx = y \tag{2.13}$$

con $y \in L^2(\mathbb{R}^n)$.
In questo caso, utilizzando la teoria delle trasformate di Fourier, è possibile caratterizzare le finestre spettrali in una maniera più semplice.
Sia $x \in L^2(\mathbb{R}^n)$. Possiamo scrivere

$$\begin{aligned}(\hat{Tx})(\xi) &= (k \hat{*} x) = \hat{k}(\xi)\hat{x}(\xi) \\ &= \int_{\mathbb{R}^n} k(t)e^{-2\pi i\xi t}dt \int_{\mathbb{R}^n} x(t)e^{-2\pi i\xi t}dt \\ (Tx)(t) &= \int_{\mathcal{A}} (\hat{Tx})(\xi)e^{+2\pi i\xi t}d\xi = \int_{\mathcal{A}} \hat{k}(\xi)\hat{x}(\xi)e^{+2\pi i\xi t}d\xi\end{aligned}$$

dove $\mathcal{A} \subseteq \mathbb{R}^n$ è la banda, che spesso considereremo limitata, della funzione $k(t)$.
Inoltre, ricordando che $\hat{\bar{k}} = \overline{\hat{k}}$,

$$\begin{aligned}(T^*x)(t) &= \int_{\mathcal{A}} \overline{\hat{k}(\xi)}\hat{x}(\xi)e^{+2\pi i\xi t}d\xi \\ (T^*T)(t) &= (T^*(Tx))(t)\end{aligned}$$

$$\begin{aligned}&= \int_{\mathcal{A}} \overline{\hat{k}(\xi)}\hat{k}(\xi)\hat{x}(\xi)e^{+2\pi i\xi t}d\xi \\ &= \int_{\mathcal{A}} |\hat{k}(\xi)|^2\hat{x}(\xi)e^{+2\pi i\xi t}d\xi\end{aligned}$$

Utilizzando la linearità dell'integrale, l'equazione di Eulero (2.8) assume la forma seguente

$$\int_{\mathcal{A}} (|\hat{k}(\xi)|^2 + \alpha)\hat{x_\alpha}(\xi)e^{+2\pi i\xi t}d\xi + \alpha \int_{C\mathcal{A}} \hat{x_\alpha}(\xi)e^{+2\pi i\xi t}d\xi = \int_{\mathcal{A}} \overline{\hat{k}(\xi)}\hat{y}(\xi)e^{+2\pi i\xi t}d\xi \tag{2.14}$$

dove $C\mathcal{A}$ è l'insieme di frequenze fuori banda.
Si ottiene

$$\hat{x_\alpha}(\xi) = \begin{cases} \frac{\overline{\hat{k}(\xi)}}{|\hat{k}(\xi)|^2+\alpha}\hat{y}(\xi) & \xi \in \mathcal{A} \\ 0 & \xi \in C\mathcal{A} \end{cases}$$

da cui

$$\begin{aligned}x_\alpha(t) &= \int_{\mathcal{A}} \frac{\overline{\hat{k}(\xi)}}{|\hat{k}(\xi)|^2 + \alpha}\hat{y}(\xi)e^{+2\pi i\xi t}d\xi \\ &= \int_{\mathcal{A}} \frac{|\hat{k}(\xi)|^2}{|\hat{k}(\xi)|^2 + \alpha}\frac{\hat{y}(\xi)}{\hat{k}(\xi)}e^{+2\pi i\xi t}d\xi \\ &= \int_{\mathcal{A}} \tilde{W_\alpha}(\xi)\frac{\hat{y}(\xi)}{\hat{k}(\xi)}e^{+2\pi i\xi t}d\xi\end{aligned}$$

dove si è posto

$$\tilde{W_\alpha}(\xi) = \frac{|\hat{k}(\xi)|^2}{|\hat{k}(\xi)|^2 + \alpha} \tag{2.15}$$

Si osservi che, qualora esista la soluzione generalizzata, questa assume la forma

$$x^\dagger(t) = \int_{\mathcal{A}} \frac{\hat{y}(\xi)}{\hat{k}(\xi)}e^{+2\pi i\xi t}d\xi \tag{2.16}$$

Alla finestra spettrale $W_\alpha(\lambda)$ ottenuta utilizzando la risolusione spettrale dell' operatore R_α, corrisponde il filtro in frequenza $\tilde{W_\alpha}(\xi)$.
$\tilde{W_\alpha}(\xi)$ ha le proprietà di un filtro "passa bassso" ed elimina le armoniche di frequenza elevata, contaminate maggiormente da errore. La famiglia $\{\tilde{W_\alpha}(\xi)\}_{\alpha>0}$ caratterizza completamente l'algoritmo regolarizzante di Tikhonov per l'operatore di Fredholm.
Si possono costruire algoritmi regolarizzanti definendo opportunamente famiglie di filtri. Più precisamente:

Teorema 2.7 *Si consideri il problema* (2.13) .
Se la famiglia di funzioni $\{W_\alpha(\xi)\}_{\alpha>0}$ *soddisfa alle seguenti proprietà*

(a) $\forall\alpha > 0$, W_α *è continua q.o.*

(b) $\forall\alpha > 0$, $0 \le W_\alpha(\xi) \le 1$

(c) $\lim_{\alpha\to 0} W_\alpha(\xi) = 1 \quad \forall\xi \in \mathcal{A}$

(d) $\forall \alpha > 0\,, \quad \frac{W_\alpha(\xi)}{\hat{k}(\xi)} \in L^\infty(\mathbb{R}^n)$

allora la famiglia di operatori definita da

$$(R_\alpha)(y) = \int_{\mathcal{A}} W_\alpha(\xi) \frac{\hat{y}(\xi)}{\hat{k}(\xi)} e^{+2\pi i \xi t} d\xi$$

è un algoritmo regolarizzante per il problema (2.13).

Dimostrazione. Verifichiamo che la famiglia $\{W_\alpha(\xi)\}$ soddisfa le condizioni $(i) - (ii)$ della definizione 2.2 di algoritmo regolarizzante.
Per (i) , si osservi che R_α è lineare per le proprietà di linearità della trasformata di Fourier; inoltre è continuo :
infatti, utilizzando l'uguaglianza di Parceval,

$$\begin{aligned} \|R_\alpha y\|_Y^2 &= \int_{\mathcal{A}} \left| \frac{W_\alpha(\xi)}{\hat{k}(\xi)} \right|^2 |\hat{y}(\xi)|^2 d\xi \\ &\leq C(\alpha) \int_{\mathcal{A}} |\hat{y}(\xi)|^2 d\xi \\ &= C(\alpha) \|y\|_X^2 \end{aligned}$$

dove $|\frac{W_\alpha(\xi)}{\hat{k}(\xi)}|^2 \leq C(\alpha) < +\infty$ per (d).
Per quanto riguarda (ii)

$$\begin{aligned} \|R_\alpha y - x^\dagger\|^2 &= \|R_\alpha T x - x^\dagger\|^2 \\ &= \int_{\mathcal{A}} \left| \frac{W_\alpha(\xi)}{\hat{k}(\xi)} \hat{k}(\xi) \hat{x}(\xi) - \hat{x}(\xi) \right|^2 d\xi \\ &= \int_{\mathcal{A}} |W_\alpha(\xi) - 1|^2 |\hat{x}(\xi)|^2 d\xi \end{aligned}$$

Per (b) si ha

$$(W_\alpha(\xi) - 1)\hat{x}(\xi) \leq \hat{x}(\xi) \in L^2(\mathbb{R}^n)$$

allora, utilizzando il teorema di convergenza dominata e (c) si ottiene

$$\|R_\alpha y - x^\dagger\|^2 \longrightarrow 0\,, \quad \alpha \longrightarrow 0\,, \quad \forall y \in L^2(\mathbb{R}^n).$$

■

La proprietà (d) è verificata ogni volta che $W_\alpha(\xi)$ tende a zero almeno come $\hat{k}(\xi)$ o $Supp(W_\alpha(\xi))$ è contenuto in un intervallo limitato di $\mathbb{R}^n$, cioè ogni volta che la funzione antitrasformata di $W_\alpha(\xi)$ è a banda limitata. Ad esempio, nel caso in cui $\hat{k}(\xi)$ non si annulli mai, la finestra rettangolare di semiampiezza $\frac{1}{\alpha}$ è il filtro più semplice che soddisfi (d) (si osservi però che tale filtro non offre buoni risultati a causa della forte discontinuità nei punti $\{-\frac{1}{\alpha}, +\frac{1}{\alpha}\}$) .
In spazi di funzioni i metodi di regolarizzazione si caratterizzano quindi come particolari metodi di filtraggio in frequenza; in relazione alla particolare classe di dati in ingresso, la teoria del filtraggio suggerisce la scelta dell'algoritmo più opportuno.

Nel caso in cui l'operatore T sia compatto, ricorrendo al sistema singolare $\{\mu_n; u_n, v_n\}$, la rappresentazione spettrale di $R_\alpha y$ assume una forma più semplice. Con un procedimento analogo a quello sviluppato nel cap. 1.5 per ottenere l'inversa generalizzata

$$T^\dagger y = \sum_{n=1}^{\infty} \frac{1}{\mu_n} (y, v_n)_Y u_n\,,$$

applicato a $R_\alpha = (T^*T + \alpha I)^{-1} T^*$, si ottiene

$$\begin{aligned} R_\alpha y &= \sum_{n=1}^{\infty} \frac{\mu_n}{\mu_n^2 + \alpha} (y, v_n)_Y u_n \\ &= \sum_{n=1}^{\infty} \left(\frac{\mu_n^2}{\mu_n^2 + \alpha} \right) \frac{1}{\mu_n} (y, v_n)_Y u_n \\ &= \sum_{n=1}^{\infty} W_{\alpha;n} \frac{1}{\mu_n} (y, v_n)_Y u_n \end{aligned}$$

Analogamente all'equazione di Fredholm, si determina la famiglia di successioni filtro $\{W_{\alpha;n}\}_{\alpha>0}$. Ogni successione ha un andamento qualitativo illustrato in figura (b).

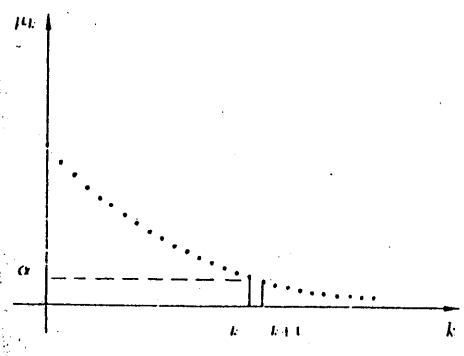


Figura (a)

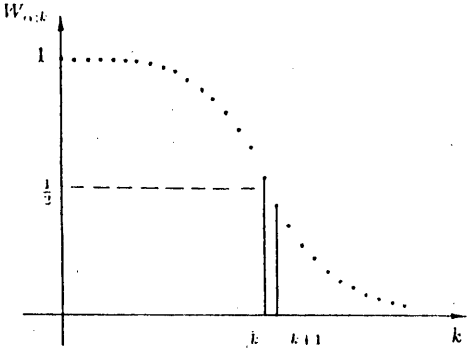


Figura (b)

Per ogni fissato $\alpha > 0$, il filtro $W_{\alpha;n}$ ha la funzione di attenuare le componenti corrispondenti agli n grandi, le quali, come è già stato detto, risentono maggiormente delle perturbazioni sul dato (infatti queste componenti sono moltiplicate per $\mu_n^{-1} \longrightarrow +\infty$).
Ovviamente anche in questo caso è possibile definire successioni di filtri alternative a quella ottenuta con l'algoritmo di Tikhonov.
Le proprietà enunciate nel teoremia 2.7 si caratterizzano nel modo seguente

(i) $\forall \alpha > 0 \quad 0 \leq W_{\alpha;n} \leq 1$

(ii) $\lim_{\alpha \to 0} W_{\alpha;n} = 1 \quad \forall n \in \mathbb{N}^*$

(iii) $\lim_{n \to +\infty} W_{\alpha;n} = 0 \quad \forall \alpha > 0$

Ad esempio , il filtro

$$\bar{W}_{\alpha;n} = \begin{cases} 1 & n \leq \frac{1}{\alpha} \\ 0 & n > \frac{1}{\alpha} \end{cases},$$

corrisponde nel caso continuo al filtraggio con finestra rettangolare e conduce alla nota decomposizione ai valori singolari troncata (Truncated S.V.D.) che verrà analizzata numericamente nel paragrafo 2.8 .

L'algoritmo regolarizzante di Tikhonov è stato introdotto per mezzo del funzionale $\Phi_\alpha[x,y]$ (2.7), il quale stima la regolarità della soluzione per mezzo della norma $\|\bullet\|_X^2$.
Consideriamo adesso un operatore lineare $L : X \longrightarrow Y$ che permetta di controllare la regolarità della soluzione secondo criteri diversi. Ad esempio, nel caso in cui X sia uno spazio di funzioni, l'operatore L può rappresentare la derivata prima, ossia $Lx = x'$, in modo da controllare le oscillazioni della soluzione più che la sua ampiezza.
Per mezzo di tale operatore è possibile definire il seguente funzionale

$$\Phi_\alpha[x,y] = \|Tx - y\|_Y^2 + \alpha\|Lx\|_X^2$$

dove la stima della regolarità di x è demandata all'operatore L.
Un teorema del tutto analogo al 2.3 asserisce che, fissato α, il minimo del funzionale esiste e corrisponde alla soluzione dell'**equazione di Eulero generalizzata**

$$(T^*T + \alpha L^*L)x_\alpha = T^*y \tag{2.17}$$

Va precisato che l'unicità della soluzione è garantita solo nel caso in cui $N(T) \cap N(L) = \{0\}$, relazione che è banalmente verificata nel caso dell'equazione (2.8) (dove $L = I$).
Nel caso in cui $N(T) \cap N(L) = \{0\}$, esiste un elemento $u \in X$, $u \neq 0$, tale che $Tu = Lu = 0$; allora se x_α è soluzione dell'equazione (2.17), ogni altro elemento $\tilde{x_\alpha} = x_\alpha + \beta u$, con $\beta \in \mathbb{R}$, risulta anch'esso soluzione.

Nel caso già considerato in cui $X = L^2(\mathbb{R}^n)$, oltre alla derivata prima vengono spesso utilizzate derivate successive o combinazioni lineari di esse. L'operatore L assume la forma

$$L = \sum_{i=0}^{N} a_i x^{(i)}$$

con $a_i \in \mathbb{R}$, $a_i \geq 0$ e il funzionale

$$\Phi_\alpha[x,y] = \|Tx - y\|_Y{}^2 + \sum_{i=0}^{N} a_i \|x^{(i)}\|^2.$$

L'equazione di Eulero viene risolta nello spazio dove le derivate N-esime sono calcolabili; più precisamente
$U \subset L^2(\mathbb{R}^n)$
$U = \{u : u^{(i)} \in C(\mathbb{R}^n) \cap L^2(\mathbb{R}^n) \quad i = 0, \ldots, N-2, \quad u^{N-1}$ assolutamente continua$\}$

I coefficienti a_i, che hanno la funzione di dare peso differente alle varie derivate, possono essere ulteriormente generalizzati considerando funzioni peso $a_i \in C^i(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$.
L'algoritmo così costruito viene detto **algoritmo regolarizzante di Tikhonov di ordine N**.
Con semplici calcoli ([2]), si può dimostrare che all'algoritmo di Tikhonov di ordine N, corrisponde il filtro in frequenza seguente

$$W_\alpha(\xi) = \frac{|\hat{k}(\xi)|^2}{|\hat{k}(\xi)|^2 + \alpha \sum_{i=0}^{N} a_i |\xi|^{2i}}$$

2.7 Metodi iterativi di regolarizzazione

In alternativa alla regolarizzazione di Tikhonov, l'equazione (2.3) può essere risolta con metodi iterativi.
Per metodo iterativo intendiamo un algoritmo che determini approssimazioni successive della soluzione, ognuna delle quali calcolata per mezzo delle precedenti.
Semplificando le notazioni, considerando un arbitrario punto iniziale $x_0 \in X$, un algoritmo iterativo A, che possiamo immaginare come una mappa $X \longrightarrow X$, genera una sequenza $\{x_k\}_{k\in\mathbb{N}} \subset X$ definita dalla relazione seguente

$$x_{k+1} = A(x_k)$$

Si osservi che x_{k+1} dipende in realtà da $x_0, x_1, \ldots, x_k$.
La proprietà principale che deve avere un metodo iterativo è la convergenza alla soluzione esatta x^*, cioè $x_k \longrightarrow x^*$ per $k \longrightarrow +\infty$. In alcuni casi la soluzione esatta viene raggiunta necessariamente in un numero finito di passi; il metodo iterativo diventa così diretto, intendendo per diretto ogni metodo che permetta di ottenere la soluzione in un numero finito di operazioni.
Un'altra caratteristica che un metodo iterativo deve avere è la stazionarietà, cioè, supponendo che l'iterata n-esima risolva l'equazione (ossia $x_n = x^*$), si deve verificare che $x_m = x_n \quad \forall m > n$, o equivalentemente $A(x^*) = x^*$.
Importantissima è la velocità di convergenza. Un metodo iterativo che, sebbene converga alla soluzione, richieda una mole enorme di calcoli non è in pratica utilizzabile.
I metodi iterativi hanno acquistato notevole interesse con l'introduzione dei calcolatori. L'implementazione di tali metodi risulta particolarmente semplice e la velocità di calcolo in continua crescita offerta dalla tecnologia li rende sempre più efficienti. Si osservi inoltre che, dal punto di vista numerico, sono generalmente più stabili dei metodi diretti.
Per una breve introduzione sugli algoritmi iterativi e sulla varie definizioni riguardanti la velocità di convergenza (raggio di convergenza, convergenza lineare, superlineare, ecc.) si consulti [25]. Per un'analisi e un confronto di diversi metodi può essere utile consultare [36]

I metodi iterativi possiedono proprietà regolarizzanti; essi risultano così un valido strumento per il trattamento di problemi inversi. Vedremo come il numero di iterazioni gioca il ruolo del parametro di regolarizzazione α della definizione 2.2 di algoritmo regolarizzante.
Più precisamente possiamo immaginare di definire la famiglia di operatori regolarizzanti $\{R_\alpha\}_{\alpha>0}$ nel seguente modo

$$R_\alpha y = S_{n(\alpha,y)} y$$

dove S_n è l'operatore ottenuto considerando le prime n iterazioni del metodo e l'applicazione $n_{(\alpha,y)}$ rappresenta la relazione che determina il numero di iterazioni necessarie in funzione del dato y e del "livello di accuratezza" α. Si osservi che l'applicazione n dipende dal valore $y \in Y$; ciò significa che in genere non è possibile considerare una semplice funzione $n(\alpha)$, ossia il numero di iterazioni necessarie per ottenere un operatore R_α prossimo all'operatore inverso, dipende fortemente dal dato del problema.

Un metodo iterativo che per la sua semplicità permette una verifica delle proprietà di regolarizzazione è il metodo di Landweber-Fridman. L'analisi di tale algoritmo aiuta a capire in che modo un metodo iterativo regolarizza un problema mal posto.
Il metodo viene costruito iterando l'equazione ai minimi quadrati (*iii*) del teorema 1.1.
In generale la ricerca delle radici di una generica equazione $f(x) = 0$ può essere ricondotta allo studio dei punti fissi dell'equazione $g(x) = x - f(x)$, cioè alla ricerca dei punti $\bar{x}$ tali

che $\bar{x} = g(\bar{x})$.
La ricerca di tali punti può essere fatta per mezzo del seguente procedimento iterativo $x_{k+1} = g(x_k)$, che, se convergente, risulta essere banalmente consistente con l'equazione iniziale $f(x) = 0$ (si consulti [8]) .
Applicando questo risultato all'equazione (iii) del teorema 1.1 , si ottiene l'**algoritmo di Landweber-Fridman**:

$$x_{n+1} = x_n + \tau(T^*y - T^*Tx_n)$$

dove τ è un valore reale, detto **parametro di rilassamento** , che permette di controllare la convergenza.
Con un semplice calcolo per induzione si verifica che il metodo assume la forma seguente

$$x_{n+1} = \tau[I + (I - \tau T^*T) + (I - \tau T^*T)^2 + \ldots + (I - \tau T^*T)^n]T^*y + (I - \tau T^*T)^{n+1}x_0 \quad (2.18)$$

È immediato così constatare che l'algoritmo di Landweber-Fridman risulta lineare.

Proposizione 2.2 *Il metodo iterativo lineare di Landweber-Fridman è un algoritmo lineare di regolarizzazione*

Dimostrazione. Consideriamo per semplicità $x_0 = 0$. Si osservi innanzitutto che, in accordo con il teorema 2.1 , opportunamente generalizzato al caso iterativo, la famiglia di funzioni reali continue $\{\tilde{R}_n\}_{n\in\mathbb{N}^*}$, definite da

$$\tilde{R}_n(t) = \tau\sum_{j=1}^{n}(1-\tau t)^j \quad ,$$

genera il metodo (il parametro $\alpha \longrightarrow 0$ ora va interpretato come $n \longrightarrow +\infty$).
Una verifica delle condizioni $(i) - (ii)$ del teorema 2.1 consente di affermare che il metodo è effettivamente un algoritmo di regolarizzazione.
Per quanto riguarda la condizione (i), si ha

$$\lambda\tau\sum_{n=1}^{\infty}(1-\tau t)^n = \tau\tfrac{1}{1-(1-\tau t)} = \tfrac{1}{t} \quad \forall t \in (0, \|T\|^2) \quad \text{se } \tau \in (0, \tfrac{2}{\|T\|^2})$$

Inoltre

$$|t\tau\sum_{n=1}^{\infty}(1-\tau t)^n| = |\tfrac{t\tau(1-(1-\tau t)^n)}{1-(1-\tau t)}|$$

$$= |\tfrac{t\tau(1-(1-t\tau)^n)}{t\tau}| \leq 1 \quad \forall t \in (0, \|T\|^2) \quad \text{se } \tau \in (0, \|T\|^{-2})$$

che verifica la condizione (ii) e che conclude la dimostrazione. ∎

Come importante esempio, consideriamo l'equazione integrale a nucleo convolutivo (2.2). È interesssante verificare che l'iterata n-esima ha l'effetto di un filtro "passa basso" (cfr. paragrafo precedente). La relazione (2.18) diventa

$$\hat{x_{n+1}}(\xi) = (1 - \tau|\hat{k}(\xi)|^2)^{n+1}\hat{x_0}(\xi) + \tau\left[1 + (1 - \tau|\hat{k}(\xi)|^2) + (1 - \tau|\hat{k}(\xi)|^2)^2 + \ldots + (1 - \tau|\hat{k}(\xi)|^2)^n\right]\overline{\hat{k}(\xi)}\hat{y}(\xi)$$

Osservando che

$$1 + (1 - \tau|\hat{k}(\xi)|^2) + (1 - \tau|\hat{k}(\xi)|^2)^2 + \ldots + (1 - \tau|\hat{k}(\xi)|^2)^n = \frac{1 - (1 - \tau|\hat{k}(\xi)|^2)^{n+1}}{1 - (1 - \tau|\hat{k}(\xi)|^2)} = \frac{1 - (1 - \tau|\hat{k}(\xi)|^2)^{n+1}}{\tau|\hat{k}(\xi)|^2}$$

si ha:

$$\hat{x_{n+1}}(\xi) = (1 - \tau|\hat{k}(\xi)|^2)^{n+1}\hat{x_0}(\xi) + \left[1 - (1 - \tau|\hat{k}(\xi)|^2)^{n+1}\right]\frac{\hat{y}(\xi)}{\hat{k}(\xi)}$$

Possiamo così scrivere l'espressione dell'iterata n-esima :

$$\begin{aligned} x_{n+1}(t) = & \int_{\mathcal{A}} \hat{x_0}(\xi)e^{+2\pi i\xi t}d\xi + \\ & \int_{\mathcal{A}} (1 - \tau|\hat{k}(\xi)|^2)^{n+1}\hat{x_0}(\xi)e^{+2\pi i\xi t}d\xi + \\ & \int_{\mathcal{A}} \left[1 - (1 - \tau|\hat{k}(\xi)|^2)^{n+1}\right]\frac{\hat{y}(\xi)}{\hat{k}(\xi)}e^{+2\pi i\xi t}d\xi \end{aligned}$$

A parte i primi due integrali, che indichiamo con $U_n(x_0(\tau))$, nulli se $x_0(\tau) \equiv 0$, l'iterata (n+1)-esima può essere scritta nel modo seguente

$$(S_{n+1}y)(t) = U_n(x_0(r))(t) + \int_{\mathcal{A}} W_{n+1}(\xi)\frac{\hat{y}(\xi)}{\hat{k}(\xi)}e^{+2\pi i\xi t}d\xi$$

dove

$$W_{n+1}(\xi) = 1 - (1 - \tau|\hat{k}(\xi)|^2)^{n+1}$$

è un filtro in frequenza dipendente da n, cioè dal numero di iterate effettuate.
Il filtro soddisfa alle condizioni del teorema 2.7 , infatti :

(a) $\forall n \in \mathbb{N}$ $W_n(\xi)$ è continua q.o.

(b) $\forall n \in \mathbb{N}$ $0 \leq W_n(\xi) \leq 1$ $\forall \xi \in \mathcal{A}$,
che è verificata se e solo se
$0 \leq (1 - \tau|\hat{k}(\xi)|^2) \leq 1 \Leftrightarrow 0 \leq \tau|\hat{k}(\xi)|^2 \leq 1 \Leftrightarrow 0 \leq \tau \leq \frac{1}{|\hat{k}(\xi)|^2}$ $\forall \xi \in \mathcal{A}$

(c) $\lim_{n\to+\infty} W_n(\xi) = 1$ $\forall \xi \in \mathcal{A}$, condizione che è verificata se e solo se
$|1 - \tau|\hat{k}(\xi)|^2| < 1$
che dà luogo, ripercorrendo passaggi analoghi al punto (ii), a

$$0 < \tau < \frac{2}{|\hat{k}(\xi)|^2} \quad \forall \xi \in \mathcal{A}$$

(d) $\frac{W_n(\xi)}{\hat{k}(\xi)} \in L^\infty(\mathbb{R}^n)$
Il numeratore, per (b), è limitato; verifichiamo che il rapporto è limitato nei punti in cui $\hat{k}(\xi) \longrightarrow 0$.
Sviluppando secondo Taylor, possiamo scrivere che $W_n(\xi) \simeq \tau n|\hat{k}(\xi)|^2$ per $\xi \in \mathcal{A}$ dove $|\hat{k}(\xi)|$ è piccolo.
Quindi $\frac{W_n(\xi)}{\hat{k}(\xi)} \simeq \tau n\hat{k}(\xi) \in L^\infty(\mathbb{R}^n)$.

La famiglia di funzioni $\{W_n(\xi)\}_{n\in\mathbb{N}}\}$ definisce così un algoritmo regolarizzante dove il parametro di regolarizzazione è il numero di iterazioni.
Le condizioni (b)–(c) imposte sul parametro di rilassamento τ forniscono $0 < \tau \leq |\hat{k}(\xi)|^{-2}$.
La convergenza richiesta in (c) è assicurata anche per $\tau \in \left(|\hat{k}(\xi)|^{-2}, \frac{2}{|\hat{k}(\xi)|^2}\right)$; per questi valori la condizione (b) non è soddisfatta poichè $W_n(\xi) \in (0,2)$. Si osservi che anche per questi $\tau \in \left(|\hat{k}(\xi)|^{-2}, \frac{2}{|\hat{k}(\xi)|^2}\right)$ si ottiene un filtro che, sebbene raddoppi l'ampiezza di alcune armoniche, converge alla funzione $f(t) \equiv 1$.
Consideriamo quindi

$$\tau \in (0, \frac{2}{\|K\|_\infty^2})$$

dove $\|K\|_\infty = sup_{\xi\in\mathbb{R}^n}|\hat{k}(\xi)|$. Si ottiene

$$\lim_{n\to+\infty} x_n(t) = x^\dagger(t) + x_0{}^{(0)}(t)$$

dove $x_0{}^{(0)}(t) = \int_{\mathcal{A}} \hat{x}(\xi)e^{+2\pi i\xi t}d\xi$ è la componente di x_0 parallela a $N(T)$.
Il limite è la soluzione ai minimi quadrati che ha la distanza minima da x_0.

I risultati del paragrafo 2.5 sulla regolarizzazione in presenza di dati perturbati , ottenuti in corrispondenza dell'algoritmo di Tikhonov, possono essere facilmente generalizzati al caso iterativo.
Indichiamo nuovamente con y^δ una approssimazione a livello δ del dato esatto $y \in R(T)$ e con y^δ_{err} la componente di errore sul dato.

Allora :

$$x_n - x^\dagger = S_n y^\delta - x^\dagger = (S_n T x^\dagger - x^\dagger) + S_n y^\delta_{err}$$

Analogamente a quanto già visto, la differenza tra la soluzione esatta e l'iterata n-esima è la somma dell'errore di approssimazione, dovuto esclusivamente al numero di iterazioni effettuate, e dell'errore dovuto alla perturbazione y^δ_{err}. Utilizzando l'uguaglianza di Parceval, nel caso in cui $x_0 = 0$, si ottiene

$$\begin{aligned}
\|S_n T x^\dagger - x^\dagger\|^2 &= \|\tau[I + (I - \tau T^*T) + (I - \tau T^*T)^2 + \ldots + (I - \tau T^*T)^{n-1}]T^*Tx^\dagger - x^\dagger\| \\
&= \int_{\mathcal{A}} |1 - (1 - \tau|\hat{k}(\xi)|^2)^n - 1|^2 \frac{|\hat{k}(\xi)|^2|\hat{x^\dagger}(\xi)|^2}{|\hat{k}(\xi)|^2} d\xi \\
&= \int_{\mathcal{A}} |1 - \tau|\hat{k}(\xi)|^2|^{2n}|\hat{x^\dagger}(\xi)|^2 d\xi
\end{aligned}$$

L'errore di approssimazione è una funzione decrescente di n dal valore $\|x^\dagger\|^2$ per $n = 0$, al valore 0 per $n \longrightarrow +\infty$ (cfr. figura A seguente).
L'errore dovuto alla perturbazione sul dato assume la forma seguente

$$\|S_n y^\delta_{err}\|^2 = \int_{\mathcal{A}} |1 - (1 - \tau|\hat{k}(\xi)|^2)^n|^2 \frac{|\hat{y^\delta_{err}}(\xi)|^2}{|\hat{k}(\xi)|^2} d\xi$$

ed è una funzione di n crescente dal valore 0 per $n = 0$ ad un valore grande per $n \longrightarrow +\infty$ (spesso $+\infty$ poichè y^δ_{err} solitamente non appartiene a $R(T) \oplus R(T)^\perp$, cfr. figura B seguente).

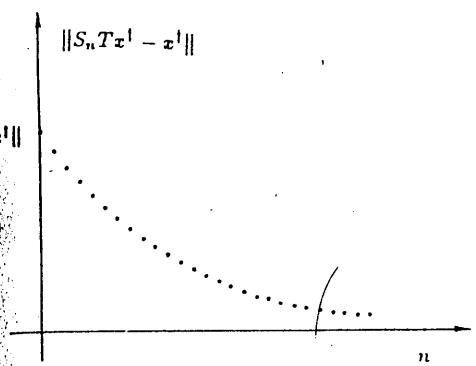


Figura A

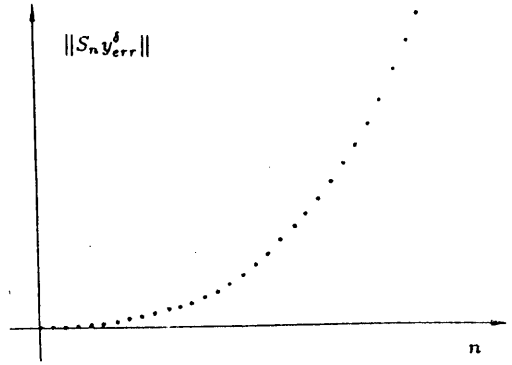
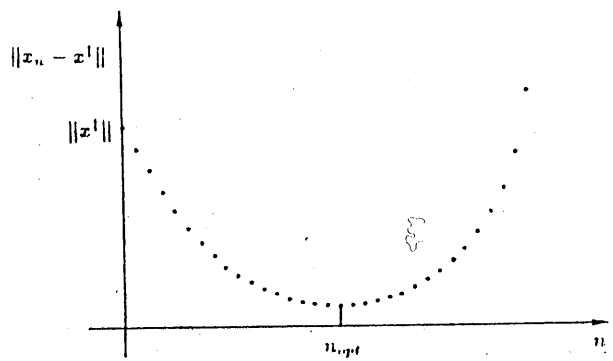


Figura B

L'errore globale $\|x_n - x^\dagger\|^2$ ha le caratteristiche rappresentate in figura seguente:



Esiste un valore ottimale n_{opt} in corrispondenza del quale si ottengono i risultati migliori, e che rappresenta il miglior compromesso tra approssimazione dell'operatore inverso e propagazione degli errori.

Generalizziamo infine il criterio della discrepanza al caso iterativo.
Si supponga di decomporre il dato perturbato y^δ nel seguente modo

$$y^\delta = y' + y^\perp$$

con $y' \in \overline{R(T)}$ e $y^perp \in R(T)^\perp$.
Poichè consideriamo come dati esatti solo quelli appartenenti a $R(T)$, allora si ha

$$\|y - y^\delta\| \leq \|y - y'\| + \|y^\perp\| \leq \delta$$

che implica

$$\|y^\perp\| \leq \delta$$

Valutiamo la discrepanza in presenza del dato perturbato.

$$\begin{aligned}\epsilon^2(n, y^\delta) &= \|Tx_n - y^\delta\|^2 \\ &= \int_{\mathcal{A}} |\hat{k}(\xi)[1-(1-\tau|\hat{k}(\xi)|^2)^n]\frac{\hat{y^\delta}(\xi)}{\hat{k}(\xi)} - \hat{y^\delta}(\xi)|^2 d\xi + \int_{C\mathcal{A}} |\hat{y^\delta}(\xi)|^2 d\xi \\ &= \int_{\mathcal{A}} |1-\tau\hat{k}(\xi)|^{2n}|\hat{y^\delta}(\xi)|^2 d\xi + \int_{C\mathcal{A}} |\hat{y^\delta}(\xi)|^2 d\xi\end{aligned}$$

La discrepanza $\epsilon(n, y^\delta)$ una funzione decrescente di n dal valore $\|y^\delta\|$ per $n = 0$, al valore $\|y^\perp\|$ per $n \longrightarrow +\infty$.
Con analogia a quanto visto per l'algoritmo di Tikhonov nel paragrafo 2.5 , è possibile definire una regola d'arresto delle iterazioni che chiamiamo ancora **criterio della discrepanza** .
Essa si caratterizza nel seguente modo :
in presenza di dati perturbati a livello δ , accettiamo come soluzione quella ottenuta dalla prima iterata tale che

$$\|Tx_n - y\| \leq \delta$$

Poichè, come abbiamo visto, $\|x_n\|$ è una funzione crescente di n , ne segue che la soluzione determinata è quella di norma minima tra quelle che consideriamo affidabili, cioè tra quelle la cui discrepanza è minore o uguale alla precisione δ con cui si ha a disposizione il dato.
Si osservi che il criterio, nel caso dell'algoritmo di Landweber-Fridman, è ben definito poichè la discrepanza è monotonicamente decrescente a $\|y^\perp\| \leq \delta$.
Spesso il criterio formulato conduce ad un numero di iterazioni troppo elevato. Nelle applicazioni pratiche si utilizza allora una costante positiva $C > 1$, la quale permette di fissare l'accuratezza della soluzione.
Il criterio della discrepanza risulta una efficace regola di arresto anche per algoritmi iterativi di regolarizzazione non lineari. Nel prossimo capitolo verrà ampliamente trattato in relazione al metodo del gradiente coniugato.

2.8 Trattamento numerico dei metodi di regolarizzazione

Nei paragrafi precedenti sono stati introdotti metodi analitici per problemi mal posti. La risoluzione numerica di tali problemi, che solitamente sono formulati in spazi continui di dimensione infinita, necessita di una discretizzazione, ossia di una approssimazione in spazi di dimensione finita. Il problema (2.3) si caratterizza quindi come sistema lineare

$$\tilde{T}\underline{x} = \underline{y} \qquad (2.19)$$

dove la matrice $\tilde{T}$ è l'approsssimazione numerica dell'operatore T, la cui dimensione dipende dal tipo di discretizzazione adottata.
Ovviamente il problema (2.19) è ben posto, poichè formulato in spazi di dimensione finita (cfr. proposizione 2.1). Da un punto di vista numerico però, l'essere ben posto non è sufficiente a garantire stabilità e affidabilità della soluzione. La proprietà che invece caratterizza ogni problema del tipo (2.19) è il condizionamento della matrice $\tilde{T}$:

un'equazione lineare (2.3) mal posta dà luogo a una matrice mal condizionata.

Si è tentati di ovviare al cattivo condizionamento della matrice, semplicemente utilizzando discretizzazioni più precise, ossia di dimensione maggiore. Questa strategia, nel caso

di problemi mal posti, non è adatta, poichè discretizzazioni più fini danno luogo a matrici inverse generalizzate sempre peggio condizionate.
Nel caso in cui T è compatto, si può fornire una semplice spiegazione di questo fenomeno.
Sia $\{\mu_n; u_n, v_n\}$ il sistema singolare dell'operatore T (vedi appendice A).
Supponiamo di aver discretizzato T per mezzo di una matrice $\tilde{T} \in M_p(\mathbb{R})$ invertibile; inoltre si consideri la decomposizione in valori singolari della matrice $\tilde{T}$, ossia

$$\tilde{T}\underline{x} = \sum_{i=1}^{p} \tilde{\mu}_i(\underline{x}, \underline{\tilde{u}}_i)\underline{\tilde{v}}_i$$

con $\tilde{\mu}_1 \geq \tilde{\mu}_2 \geq \ldots \geq \tilde{\mu}_p > 0$.
Si verifica che i p valori singolari di $\tilde{T}$ risultano essere una approssimazione dei primi p valori singolari dell'operatore T. Una spiegazione euristica di questo fatto è la seguente: ricordando che i primi p valori singolari sono quelli corrispondenti alle componenti più significative e meno oscillanti, si può concludere che queste sono le più facili da approssimare, ossia sono quelle rispetto alle quali la discretizzazione adottata risulta più sensibile.
Si supponga ora di realizzare una discretizzazione più fine, che dia luogo a una matrice $\bar{\tilde{T}} \in M_{p'}(\mathbb{R})$, $p' > p$. In questo caso i p' valori singolari di $\bar{\tilde{T}}$ sono una approssimazione dei primi p' valori singolari di T.
Consideriamo allora le rispettive inverse generalizzate. I valori singolari di $\tilde{T}^\dagger$ sono
$\tilde{\mu}_p^{-1} \geq \tilde{\mu}_{p-1}^{-1} \geq \ldots \geq \tilde{\mu}_1^{-1}$
mentre quelli di $\bar{\tilde{T}}^\dagger$ sono
$\bar{\tilde{\mu}}_{p'}^{-1} \geq \bar{\tilde{\mu}}_{p'-1}^{-1} \geq \ldots \geq \bar{\tilde{\mu}}_1^{-1}$
Si ottiene

$$cond(\tilde{T}^\dagger) = \frac{\frac{1}{\tilde{\mu}_p}}{\frac{1}{\tilde{\mu}_1}} = \frac{\tilde{\mu}_1}{\tilde{\mu}_p}$$

e

$$cond(\bar{\tilde{T}}^\dagger) = \frac{\frac{1}{\bar{\tilde{\mu}}_{p'}}}{\frac{1}{\bar{\tilde{\mu}}_1}} = \frac{\bar{\tilde{\mu}}_1}{\bar{\tilde{\mu}}_{p'}}$$

Considerando nuovamente le proprietà delle componenti associate ai primi valori singolari, si osservi che le due approssimazioni di μ_1 non differiscono di molto (la approssimazione ottenuta per mezzo della prima discretizzazione è gia buona), mentre, per quanto riguarda i valori singolari più piccoli, in genere $\tilde{\mu}_p \gg \bar{\tilde{\mu}}_{p'}$. Si è così ottenuto che

$$cond(\bar{\tilde{T}}^\dagger) \gg cond(\tilde{T}^\dagger)$$

Si osservi inoltre che $cond(\tilde{T}) = cond(\tilde{T}^\dagger)$; l'instabilità in presenza di perturbazione sui dati si manifesta però maggiormente nel problema inverso per le ragioni già evidenziate nel paragrafo 2.2 .

Prima di procedere con l'analisi numerica dei metodi di regolarizzazione, vediamo come discretizzare l'equazione di Fredholm di prima specie (2.2)

$$\int_a^b k(r,s)x(r)dr = y(s) \qquad c \leq s \leq d$$

con $x \in L^2([a,b])$, $y \in L^2([c,d])$, $k \in L^2([a,b] \times [c,d])$.

L'idea più semplice, che va sotto il nome di **metodo di collocazione**, è richiedere che l'equazione sia verificata su un numero finito di punti $s_1, s_2, \ldots, s_M \in [c,d]$, detti punti di collocazione. Si ottengono così M equazioni integrali

$$\int_a^b k(r,s_i)x(r)dr = y(s_i) \qquad i = 1,\ldots,M \tag{2.20}$$

Esaminiamo due diversi metodi per la risoluzione del sistema ottenuto.

(1) Le equazioni possono essere risolte numericamente per mezzo di qualsiasi formula di quadratura (cfr. [8]) con nodi $r_1, r_2, \ldots, r_N \in [a,b]$ e pesi $\beta_1, \beta_2, \ldots, \beta_N$, ossia

$$\int_a^b k(r,s_i)x(r)dr \simeq \sum_{j=1}^{N} \beta_j k(r_j,s_i)x(r_j).$$

Si ottiene il seguente sistema lineare

$$\sum_{j=1}^{N} \beta_j k(r_j,s_i)x(r_j) = y(s_i) \qquad i = 1,\ldots,M$$

nell'incognita $(x(r_1),\ldots,x(r_N))$, o equivalentemente,

$$\tilde{T}\underline{x} = \underline{y} \quad \text{dove} \quad \tilde{T} \in M_{M,N}(\mathbb{R}) \tag{2.21}$$

$\tilde{T}_{i,j} = \beta_j k(r_j,s_i)$
$x_j = x(r_j) \quad j = 1,\ldots,N$
$y_i = y(s_i) \quad i = 1,\ldots,M.$
Il sistema (2.21) , risolto nel senso della soluzione generalizzata con metodi di regolarizzazione numerici (che vedremo più avanti), dà la soluzione cercata. Si osservi che nel caso in cui $[a,b] = [c,d]$, si possono scegliere i punti di collocazione uguali ai nodi di integrazione; in questo caso si ottiene una matrice quadrata.

(2) Più efficiente è il metodo basato sulle funzioni di rappresentazione $\{K_i\}_{i=1}^{M}$, definite come

$$K_i \in L^2([a,b]) \qquad K_i(r) = k(r,s_i)$$

Indichiamo con $< \bullet, \bullet >_b$ $\quad L^2([a,b]) \times L^2([a,b]) \longrightarrow \mathbb{R}$ il prodotto scalare in $L^2([a,b])$ definito nel modo seguente

$$< a,b >= \int_a^b a(r)b(r)dr$$

e con $L : \; L^2([a,b]) \longrightarrow \mathbb{R}^M$ l'operatore lineare definito da

$$Lf = \begin{pmatrix} < K_1, f > \\ < K_2, f > \\ \cdots \\ \cdots \\ < K_M, f > \end{pmatrix}$$

L'insieme di equazioni (2.20) può così essere espresso nella forma seguente

$$L\underline{x} = \underline{y} \tag{2.22}$$

dove $\underline{y}$ è il vettore come definito in (1).
Dimostriamo che la soluzione generalizzata di (2.22) , che si osservi appartiene a $L^2([a,b])$, è combinazione lineare delle funzioni K_i (da cui il nome "funzioni di rappresentazione").
Sappiamo che $x^\dagger \in N(L)^\perp = \overline{R(L^*)} = R(L^*)$, infatti L^* è un operatore discreto.
Valutiamo allora l'espressione dell'operatore L^*. Occorre determinare l'operatore L^* tale che

$$(Lf,\underline{\phi})_{\mathbb{R}^M} = < f, L^*\underline{\phi} > \qquad \forall f \in L^2([a,b]) \quad \forall \underline{\phi} \in \mathbb{R}^M$$

Si ha $(Lf,\underline{\phi})_{\mathbb{R}^M} = \sum_{i=1}^{M} \phi_i < K_i, f > = < \sum_{i=1}^{M} \phi_i K_i, f >$,
quindi

$$L^*\underline{\phi} = \sum_{i=1}^{M} \phi_i K_i \qquad \forall \underline{\phi} \in \mathbb{R}^M$$

Ne segue che $x^\dagger \in L^2[a,b]$ ammette la rappresentazione

$$x^\dagger = \sum_{i=1}^{M} \alpha_i K_i = L^*\overline{\alpha}$$

per $\overline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)^t$ opportuno.
Il problema è così ricondotto alla determinazione del vettore dei coefficienti $\overline{\alpha}$.
Verifichiamo che questo è la soluzione del sistema lineare

$$G\overline{\alpha} = \underline{y} \tag{2.23}$$

dove $G \in M_{M,M}(\mathbb{R})$, detta matrice di Gram, è simmetrica semidefinita positiva con $(G)_{i,j} = < K_i, K_j >$.

Sappiamo che $L^*\underline{\phi} = \sum_{i=1}^{M} \phi_i K_i \quad \forall \underline{\phi} \in \mathbb{R}^M$, pertanto

$$LL^*\underline{\phi} = L(\sum_{i=1}^{M} \phi_i K_i) = \begin{bmatrix} < \sum_{i=1}^{M} \phi_i K_i, K_1 > \\ < \sum_{i=1}^{M} \phi_i K_i, K_2 > \\ \cdots \\ \cdots \\ < \sum_{i=1}^{M} \phi_i K_i, K_M > \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{M} \phi_i < K_i, K_1 > \\ \sum_{i=1}^{M} \phi_i < K_i, K_2 > \\ \cdots \\ \cdots \\ \sum_{i=1}^{M} \phi_i < K_i, K_M > \end{bmatrix} = G\underline{\phi}$$

$$\Leftrightarrow \qquad G = LL^*$$

La matrice di Gram risulta autoaggiunta semidefinita positiva.
Inoltre

$$Lx^\dagger = \underline{y} \quad \Leftrightarrow \quad LL^*\overline{\alpha} = \underline{y} \quad \Leftrightarrow \quad G\overline{\alpha} = \underline{y}$$

quindi effettivamente $\overline{\alpha}$ è soluzione del sistema (2.23).

La matrice G può essere decomposta nel prodotto

$$G = UDU^t$$

con $U \in M_{M,M}(\mathbb{R})$ matrice ortonormale le cui colonne sono gli autovettori $\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_M$ associati rispettivamente agli autovalori $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M \geq 0$ e $D \in M_{M,M}(\mathbb{R})$ matrice diagonale contenente gli autovalori, $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_M)$.
Si ottiene (cfr. cap 1.7)

$$\overline{\alpha}^\dagger = UD^\dagger U^t \underline{y}$$

con $D^\dagger = \begin{cases} \frac{1}{\lambda_i} & \text{se } \lambda_i > 0 \\ 0 & \text{se } \lambda_i = 0 \end{cases}$
Esplicitando le componenti, la soluzione assume la forma seguente

$$\begin{aligned} x^\dagger &= \sum_{i=1}^{M} \alpha_i K_i \\ &= \sum_{i=1}^{M} \left(\sum_{j=1, \lambda_j \neq 0}^{M} \frac{(\underline{u}_j, \underline{y})}{\lambda_j} u_{i,j} \right) K_i \\ &= \sum_{j=1, \lambda_j \neq 0}^{M} \left(\frac{(\underline{u}_j, \underline{y})}{\lambda_j} \right) \eta_j \end{aligned}$$

con

$$\eta_j = \sum_{i=1}^{M} u_{i,j} K_i$$

Si osservi che anche in questo caso per la determinazione della matrice di Gram, occorre utilizzare una formula di quadratura.
La soluzione ottenuta in funzione delle K_i può essere valutata nei punti $\{x_l\}_{l=1}^N$; si determina la soluzione numerica seguente

$$x_l = \sum_{j=1, \lambda_j \neq 0}^{M} \left(\frac{(\underline{u}_j, \underline{y})}{\lambda_j} \right) \sum_{i=1}^{M} u_{i,j} k(r_l, s_i) \qquad l = 1, \ldots, N$$

Procediamo come accennato con l'illustrazione di alcuni metodi numerici per il trattamento di problemi mal posti, ossia metodi per la regolarizzazione del sistema lineare derivante dalla discretizzazione. È oppurtuno precisare che le considerazioni e i risultati della restante parte del paragrafo possono essere applicati semplicemente al problema algebrico di risoluzione di un sistema mal condizionato, senza necessariamente fare riferimento al problema continuo da cui esso deriva.
Indichiamo d'ora in poi la matrice $\tilde{T}$ semplicemente con T. Supponiamo T quadrata (eventualmente la si può rendere quadrata aggiungendo gli opportuni zeri). Riformuliamo il problema nel modo seguente:
" si risolva il sistema lineare $T\underline{x} = \underline{y}$, dove la matrice $T \in M_n(\mathbb{R})$ è mal condizionata ".
Una soluzione numerica diretta di tale problema darebbe luogo a soluzioni inaffidabili, spesso molto oscillanti rispetto alla soluzione analitica. Si ricordi inoltre che abbiamo a disposizione dati comunque affetti da errore (nel migliore dei casi va comunque contemplato il numero di macchina dell'elaboratore).

Lo scopo è ottenere una soluzione

$$\underline{x} = \sum_{i=1}^{k} c_i \underline{x}_i$$

dove il valore $k \leq n$ e il sistema di vettori $\{\underline{x}_i\}_{i=1}^k$ va scelto in relazione al problema considerato e al metodo utilizzato.
Si osservi che, mentre il sistema ha dimensione n (generalmente molto grande), la soluzione che cerchiamo è combinazione lineare di $k \leq n$ vettori; in questo modo risolviamo il sistema in uno spazio di dimensione minore che garantisce però maggiore stabilità. Nei nostri esempi il sistema $\{\underline{x}_i\}_{i=1}^k$ sarà sempre formato da vettori ortonormali.
L'efficienza del metodo dipende dalla scelta della base $\{\underline{x}_i\}_{i=1}^k$; la base canonica $\underline{x}_i = \underline{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, con 1 all'i-esimo posto, comunemente utilizzata, spesso non offre i risultati migliori. Inoltre, come già osservato, una base può risultare adatta per la risoluzione di un sistema $K\underline{x} = \underline{y}$, mentre può non esserlo per un altro sistema $K'\underline{x} = \underline{y}$.
L'opportunità di utilizzare una particolare base può essere "misurata" per mezzo del numero di condizionamento di T rispetto ai vettori $\{\underline{x}_i\}_{i=1}^k$ che indicheremo con $\mathcal{K}$, e che è definito nel seguente modo

$$\mathcal{K} = \frac{\mu_1(T)}{\mu_k(TX_k)} \tag{2.24}$$

dove X_k è la matrice $n \times k$ formata dai vettori colonna $\{\underline{x}_i\}_{i=1}^k$ e $\mu_i(T)$ è l'i-esimo dei valori singolari, ordinati in senso decrescente, della matrice T. Più il numero di condizionamento è piccolo, minore è la sensibilità della soluzione a perturbazioni sui dati, quindi migliore è la stabilità.
Fissata la base $\{\underline{x}_i\}_{i=1}^k$, il numero di condizionamento in effetti dipende dal più piccolo valore singolare della matrice TX_k. $\mathcal{K}$ può essere reso piccolo considerando un numero minore di vettori, cioè considerando la base $\{\underline{x}_i\}_{i=1}^{k'}$ $k' < k$. Anche qui si presenta il compromesso tra approssimazione del dato e stabilità : se una base di cardinalità minore permette maggiore regolarità, occorre però controllare che la distanza $\|T\underline{x} - \underline{y}\|$ sia entro la tolleranza richiesta.
Elenchiamo ora alcuni semplici metodi numerici; in seguito verrà brevemente affrontato il problema della stabilità in presenza di dati perturbati.

1 Fattorizzazione QR troncata (TQR)

La fattorizzazione QR, ossia decomposizione della matrice T nel prodotto di una matrice unitaria Q per una triangolare superiore R, è un procedimento comunemente usato per la risoluzione di sistemi lineari (cfr. [8]). Esso ha buone proprietà di stabilità e di costo. Nel caso in cui T sia mal condizionata, il metodo si caratterizzaa nel modo seguente.
Indicata nuovamente con X la matrice la cui i-esima colonna corrisponde al vettore i-esimo della base $\{\underline{x}_i\}_{i=1}^n$, si determini la decomposizione QR di TX_m, dove X_m è la matrice $(n \times m)$ formata dalle prime m colonne di X, con $m \leq n$. In formule

$$TX_m = QR$$

con $Q \in M_{n,m}$ e $R \in M_{m,m}$.
Il parametro m corrisponde al numero massimo di vettori rispetto ai quali vogliamo sia espressa la soluzione $\underline{x} = \sum_{i=1}^{m} c_i \underline{x}_i$ e va scelto sulla base di considerazioni tra costo e affidabilità.

Si risolvano poi, con "sostituzione all'indietro", le prime $k \leq m$ equazioni (da cui il nome di fattorizzazione "troncata") del sistema

$$R\underline{c} = Q^t y$$

Otteniamo così la soluzione approssimata

$$\underline{x}^{(k)} = \sum_{i=1}^{k} c_i \underline{x}_i \tag{2.25}$$

Il valore k gioca il ruolo di parametro di regolarizzazione. Come si intuisce, esso deve essère il più piccolo tra quelli che permettono una buona approssimazione del dato, in modo da ottenere un numero di condizionamento K minore possibile.
Il residuo $\|T\underline{x}^{(k)} - y\|_2$, che misura l'accuratezza della soluzione, può essere facilmente stimato calcolando la norma l^2 della ultime $(n-k)$ componenti di $Q^t y$. Un calcolo iterativo di tale norma per diversi valori di k, dà indicazioni sulla scelta del valore ottimale.

2 Regolarizzazione di Tikhonov (del primo ordine)

Analogamente a quanto visto nei paragrafi precedenti, la soluzione del sistema $T\underline{x} = \underline{y}$ viene determinata risolvendo il problema di minimo

$$\min_{\underline{x}} (\|T\underline{x} - \underline{y}\|_2^2 + \alpha \|\underline{x}\|_2^2)$$

dove $\| \bullet \|_2$ indica la norma l^2, e il problema di minimo è equivalente all'equazione di Eulero

$$(T^t T + \alpha I)\underline{x} = T^t \underline{y}$$

Si osservi che la soluzione dell'equazione di Eulero è la soluzione ai minimi quadrati del sistema sovradeterminato $(2n \times n)$

$$\begin{pmatrix} T \\ \alpha \end{pmatrix} \underline{x} = \begin{pmatrix} \underline{y} \\ 0 \end{pmatrix}$$

Per la soluzione dell'equazione di Eulero può essere utilizzata la decomposizione ai valori singolari (S.V.D.) della matrice T (vedi [30]). Si determina così la seguente decomposizione

$$T = VDU^t$$

dove D è matrice diagonale contenente i valori singolari di T ,ossia $d_{i,i} = \mu_i \geq O$, e U , V sono ortogonali le cui colonne sono i corrispondenti vettori singolari $\underline{u}_i$, $\underline{v}_i$ (si osservi che $\underline{v}_i = {\mu_i}^{-1} T \underline{u}_i$ e $\underline{u}_i = {\mu_i}^{-1} T^t \underline{v}_i$ come da definizione di sistema singolare - vedi appendice A).
Occorre quindi esprimere il vettore dei dati y come combinazione lineare dei vettori singolari $\underline{v}_i$

$$\underline{y} = \sum_{i=1}^{n} \beta_i \underline{v}_i$$

il vettore dei coefficienti $\underline{b} = (\beta_1, \beta_2 \ldots, \beta_n)^t$ può essere determinato immediatamente facendo uso della matrice ortogonale V

$$\underline{b} = V^t \underline{y}$$

Si ottiene, posto $\underline{x}_\alpha = \sum_{i=1}^{n} c_i \underline{u}_i$,

$$\begin{aligned} (T^t T + \alpha I)\underline{x}_\alpha = T^t y \quad &\Leftrightarrow \quad \sum_{i=1}^{n} c_i (T^t T + \alpha I)\underline{u}_i = \sum_{i=1}^{n} \beta_i T^t \underline{v}_i \\ &\Leftrightarrow \quad \sum_{i=1}^{n} c_i ({\mu_i}^2 + \alpha)\underline{u}_i = \sum_{i=1}^{n} \beta_i \mu_i \underline{u}_i \\ &\Leftrightarrow \quad c_i = \frac{\beta_i \mu_i}{{\mu_i}^2 + \alpha} \end{aligned}$$

Si ottiene la soluzione

$$\underline{x}_\alpha = \sum_{i=1}^{n} \left(\frac{\beta_i}{\mu_i + \frac{\alpha}{\mu_i}} \right) \underline{u}_i \tag{2.26}$$

o equivalentemente, posto $\underline{c}_\alpha = (c_1, c_2, \ldots, c_n)^t \quad c_i = \frac{\beta_i}{\mu_i + \frac{\alpha}{\mu_i}}$,

$$\underline{x}_\alpha = U \underline{c}_\alpha$$

Secondo la terminologia adottata precedentemente, il sistema $\{\underline{x}_i\}$ in questo caso corrisponde al sistema di vettori singolari $\{\underline{u}_i\}$, ossia $X = U$.
È importante sottolineare che i valori β_i , μ_i e i vettori $\underline{u}_i$ e $\underline{v}_i \quad i = 1 \ldots n$, sono indipendenti dal parametro di regolarizzazione α e quindi vengono calcolati una sola volta.

3 Decomposizione ai valori singolari troncata (T.S.V.D.)

Questo metodo è la versione numerica dell'omonimo sviluppato nel paragrafo 2.6 .
Si decomponga la matrice T come già visto al punto 2 :

$$T = VDU^t$$

Esprimendo nuovamente il vettore $\underline{y}$ in funzione del sistema $\{\underline{v}_i\}_{i=1}^n$, $\underline{y} = \sum_{i=1}^{n} \beta_i \underline{v}_i$,
e calcolando i coefficienti $\underline{b} = (\beta_1, \beta_2 \ldots, \beta_n)^t$, con $\underline{b} = V^t \underline{y}$,
la soluzione del sistema $T\underline{x} = \underline{y}$ assume la forma $DU^t \underline{x} = V^t \underline{y} \quad \Leftrightarrow \quad U^t \underline{x} = D^{-1} V^t \underline{y}$,
ossia

$$\underline{x} = \sum_{i=1}^{n} \left(\frac{\beta_i}{\mu_i} \right) \underline{u}_i$$

La regolarizzazione della soluzione si ottiene considerando uno sviluppo di $\underline{x}$ troncato al k-esimo termine

$$\underline{x}^{(k)} = \sum_{i=1}^{k} \left(\frac{\beta_i}{\mu_i} \right) \underline{u}_i \tag{2.27}$$

Come per il metodo TQR, k assume il ruolo di parametro di regolarizzaazione; se uno sviluppo con pochi termini offre buoni risultati, allora il metodo risulta particolarmente utile.
Questo dipende, oltre che dalla matrice T, anche dal dato $\underline{y}$:
se i coefficienti β_i tendono a zero più velocemente dei μ_i , allora la differenza
$\|\underline{x} - \underline{x}^{(k)}\|_2^2 = \sum_{i=k+1}^{n} (\frac{\beta_i}{\mu_i})^2$ risulta piccola e il metodo efficace.

Si osservi che, nel caso in cui questa situazione non si verifichi, la T.S.V.D. non deve essere utilizzata, ma occorre cercare una soluzione rispetto a una base diversa da $\{\underline{u}_i\}$.
Come per il metodo TQR , la ricerca del valore k ottimale deve essere fatta iterativamente, ossia attraverso prove successive (utilizzando strategie di tipo dicotomico, per esempio).

4 Regolarizzazione di Tikhonov di ordine superiore

Come evidenziato nel paragrafo 2.6 , si tratta di risolvere il seguente problema di minimo

$$\min_{\underline{x}}(\|T\underline{x}-\underline{y}\|_2^2+\alpha\|L\underline{x}\|_2^2)$$

dove la matrice L rappresenta l'operatore che stima la regolarità della soluzione.
Nel caso in cui la matrice T sia la discretizzazione dell'operatore di Fredholm di prima specie, L generalmente è la discretizzazione di un opportuno operatore differenziale, solitamente alle differenze finite. Ad esempio, si consideri il metodo di collocazione con reticolo formato da nodi equidistanti $r_j = a + j\delta \quad j = 0,\dots,N-1, \quad \delta = \frac{b-a}{N-1}$.
L'approssimazione dell'operatore $\tilde{L} = x'$ con differenze finite assume la forma

$$x'(r_j) = \frac{1}{\delta}[x(r_{j+1}) - x(r_j)] \qquad j = 0,\dots,N-2$$

e l'approssimazione dell'operatore $\tilde{\tilde{L}} = x''$

$$\begin{aligned} x''(r_j) &= \frac{1}{\delta}[x'(r_{j+1}) - x'(r_j)] \\ &= \frac{1}{\delta}\left\{\frac{1}{\delta}[x(r_{j+1}) - x(r_j)] - \frac{1}{\delta}[x(r_j) - x(r_{j-1})]\right\} \\ &= \frac{1}{\delta^2}\{x(r_{j+1}) - 2x(r_j) + x(r_{j-1})\} \qquad j = 1,\dots,N-2 \end{aligned}$$

Questi, a meno di fattori di scala, assumono rispettivamente la forma

$$\tilde{L}x = x' \longrightarrow L = \begin{pmatrix} 1 & -1 & & & & \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \dots & \dots & \\ & & & & 1 & -1 \end{pmatrix}_{(n-1)\times n}$$

$$\tilde{\tilde{L}}x = x'' \longrightarrow L = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & 1 & -2 & 1 & \\ & & & \dots & \dots & \dots \\ & & & & 1 & -2 & 1 \end{pmatrix}_{(n-2)\times n}$$

In generale è possibile utilizzare ordini di derivazione superiori, e combinazioni lineari delle rispettive matrici, ma queste non sempre comportano miglioramenti apprezzabili. Si osservi che, nel caso in cui L sia l'approssimazione della derivata p-esima, la matrice corrispondente è di dimensione $(n-p)\times n$ e rango $(n-p)$ (ad esesmpio la derivata 4^a ha una approssimazione alle differenze finite di dimensione $(n-4)\times n$ la cui prima riga è $(1,-4,6,-4,1,0,\dots,0)$).

Ritornando al problema di minimo enunciato all'inizio, analogamente a quanto fatto al punto 2, questo è equivalente all'equazione di Eulero modificata

$$(T^tT + \alpha L^tL)\underline{x} = T^t\underline{y}$$

la quale a sua volta è equivalenta alla risolusione ai minimi quadrati del sistema sovradeterminato $(2n \times n)$

$$\begin{pmatrix} T \\ \alpha L \end{pmatrix}\underline{x} = \begin{pmatrix} \underline{y} \\ 0 \end{pmatrix}$$

Indichiamo con K la matrice le cui colonne sono gli autovettori generalizzati di (T, L) , cioè i vettori $\underline{k}_i$ soluzione del sistema

$$\lambda_i(T^tT)\underline{k}_i = (L^tL)\underline{k}_i \qquad i = 1,\dots,n$$

(si osservi che L^tL ha rango $(n-p)$, quindi necessariamente esistono $(n-p)$ autovalori generalizzati λ_i non nulli).
Per mezzo della matrice K è possibile determinare una particolare decomposizione per T e per L detta decomposizione ai valori singolari generalizzati (Generalized S.V.D.) (si consulti [45]). Si ha così

$$T = VD_aK^{-1} \qquad L = UD_bK^{-1}$$

dove $V(n\times n)$ e $U((n-p)\times(n-p))$ sono ortogonali,

$$D_a = diag(a_1, a_2, \dots, a_n) \ ,$$

$$D_b = \left(\begin{array}{cccc|c} b_1 & & & & \\ & b_2 & & & \\ & & \dots & & O \\ & & & b_{n-p} & \end{array}\right)_{(n-p)\times n}$$

Inoltre, posto $b_i = 0 \quad i = n-p+1,\dots,n$, si può dimostrare che $\lambda_i = \frac{b_i}{a_i} \quad i = 1,\dots,n$. In questo modo le ultime p colonne di K formano una base per il nucleo di L.

Con l'ausilio della G.S.V.D. , l'equazione di Eulero viene risolta nel seguente modo.

$$\begin{aligned} (T^tT + \alpha L^tL)\underline{x}_\alpha = T^t\underline{y} &\Leftrightarrow (K^{-t}D_a^2K^{-1} + \alpha K^{-t}D_b^tD_bK^{-1})\underline{x}_\alpha = K^{-t}D_aV^t\underline{y} \\ &\Leftrightarrow (D_a^2 + \alpha D_b^tD_b)K^{-1}\underline{x}_\alpha = D_aV^t\underline{y} \end{aligned}$$

Posto $\underline{y} = \sum_{i=1}^{n}\beta_i\underline{v}_i$ con $\underline{b} = (\beta_1,\dots,\beta_n)$ e $\underline{b} = V^t\underline{y}$, si ottiene la soluzione

$$\underline{x}_\alpha = \sum_{i=1}^{n}\left(\frac{a_i\beta_i}{a_i^2 + \alpha b_i^2}\right)\underline{k}_i \qquad (2.28)$$

Il problema è così ricondotto alla determinazione del parametro di regolarizzazione α più opportuno. Anche in questo caso, analogamente al punto 2, si osservi che

$a_i, b_i, \beta_i, \underline{k}_i, \underline{u}_i, \underline{v}_i \qquad i = 1, \ldots, n$ sono indipendenti dal parametro α e vengono calcolati una sola volta.
Si osservi inoltre che questo metodo rientra nella categoria vista all'inizio del paragrafo, ponendo $c_i = \frac{a_i \beta_i}{a_i^2 + \alpha b_i^2}$ e $\underline{x}_i = \underline{k}_i$.
È importante sottolineare che il calcolo numerico della G.S.V.D. non è affatto semplice e, a tutt'oggi, non esistono pacchetti software appositamente dedicati; in pratica il metodo è difficilmente utilizzabile .

Sappiamo che l'obiettivo principale dei metodi esposti è rendere la soluzione poco sensibile a pertubazioni del dato (regolarizzazione).
Si possono trarre importanti indicazioni analizzando il comportamento di tali metodi in presenza di dati affetti da errore.
A tal proposito indichiamo con $\underline{y}$ il dato non perturbato, con $\underline{y}_\delta$ il dato a disposizione, e consideriamo $\|\underline{y} - \underline{y}_\delta\|_2 = \delta$.

1 T.Q.R.

Sia $\underline{x} = \sum_{i=1}^{n} d_i \underline{x}_i$ la soluzione corrispondente al dato esatto e $\underline{x}_\delta = \sum_{i=1}^{k} \tilde{d}_i \underline{x}_i$ la soluzione corrispondente al dato perturbato, troncata al k-esimo termine.
Poichè la base $\{\underline{x}_i\}_{i=1}^n$ è scelta ortogonale, si ha

$$\|\underline{x} - \underline{x}_\delta\|_2^2 = \sum_{i=1}^{k} (d_i - \tilde{d}_i)^2 + \sum_{i=k+1}^{n} d_i^2$$

Si osservi che, anche in assenza di perturbazioni, la soluzione troncata non è "buona" nel caso in cui i primi k vettori non riescano ad approssimare sufficientemente la soluzione (cioè, per $i > k$, i coefficienti d_i non siano piccoli).
Si può ulteriormente maggiorare la prima sommatoria.
Si consideri la matrice R della decomposizione $TX = QR$, partizionata nel seguente modo
$R = \begin{pmatrix} R_1 & S \\ 0 & R_2 \end{pmatrix}$ con R_1 di dimensione $(k \times k)$ e R_2 di dimensione $((n-k) \times (n-k))$.
Posto $\underline{d} = (d_1, \ldots, d_n)$, $\underline{\tilde{d}} = (\tilde{d}_1, \ldots, \tilde{d}_k)$, si indichi
con $\underline{d}_1$ il vettore formato dalle prime k componenti di $\underline{d}$,
con $\underline{d}_2$ il vettore formato dalle seguenti $(n-k)$ componenti ,
con Q_1 la matrice formata dalle prime k colonne di Q ,
e con Q_2 la matrice formata dalle ultime $(n-k)$ colonne.

Si ha così
$$Q \begin{pmatrix} R_1 & S \\ 0 & R_2 \end{pmatrix} \begin{pmatrix} \underline{d}_1 \\ \underline{d}_2 \end{pmatrix} = \underline{y} \quad \Leftrightarrow \quad \begin{pmatrix} R_1 & S \\ 0 & R_2 \end{pmatrix} \begin{pmatrix} \underline{d}_1 \\ \underline{d}_2 \end{pmatrix} = \begin{pmatrix} Q_1^t \underline{y} \\ Q_2^t \underline{y} \end{pmatrix}$$
e
$$Q_1 R_1 \underline{\tilde{d}} = \underline{y}_\delta \quad \Leftrightarrow \quad R_1 \underline{\tilde{d}} = Q_1^t \underline{y}_\delta$$
Si ottiene

$$R_1(\underline{d}_1 - \underline{\tilde{d}}) = Q_1^t \underline{y} - Q_1^t \underline{y}_\delta - S\underline{d}_2$$

$$\|\underline{d}_1 - \underline{\tilde{d}}\|_2 \leq \|R_1^{-1}\|_2 (\|Q^t\|_2 \|\underline{y} - \underline{y}_\delta\|_2 + \|S\|_2 \|\underline{d}_2\|_2)$$

$$= \|R_1^{-1}\|_2 (\delta + \|S\|_2 \|\underline{d}_2\|_2)$$

Dall'ultima espressione si può notare che, sebbene k crescenti riducano $\|\underline{d}_2\|_2$, la differenza $\|\underline{d}_1 - \underline{\tilde{d}}\|_2$ può crescere molto; infatti, a causa del peggior condizionamento di $(KX)_k$, il termine $\|R_1^{-1}\|_2$ può diventare molto grande. Questo risultato è coerente con quanto detto nella prima parte del paragrafo circa il significato di k.

2 Regolarizzazione di Tikhonov

Consideriamo l'algoritmo sviluppato con l'ausilio della S.V.D. ; in questo caso la soluzione corrispondente al dato non perturbato (senza regolarizzazione) è data da
$\underline{x} = \sum_{i=1}^{n} (\frac{\beta_i}{\mu_i}) \underline{u}_i$ dove $\underline{y} = \sum_{i=1}^{n} \beta_i \underline{v}_i$.
La soluzione regolarizzata in presenza del dato perturbato è invece data da
$\underline{x}_\delta = \sum_{i=1}^{n} (\frac{\bar{\beta}_i}{\mu_i + \frac{\alpha}{\mu_i}})$ dove $\underline{y}_\delta = \sum_{i=1}^{n} \bar{\beta}_i \underline{v}_i$
Possiamo così scrivere

$$\underline{x} - \underline{x}_\delta = \sum_{i=1}^{n} \left(\frac{(\beta_i - \bar{\beta}_i)\mu_i + \alpha \frac{\beta_i}{\mu_i}}{\mu_i^2 + \alpha} \right) \underline{u}_i$$

Anche in questo caso, per $\alpha \longrightarrow 0$, l'errore cresce a causa delle ultime componenti $\underline{u}_i$ che hanno i valori singolari μ_i sempre più piccoli (si osservi che l'unica informazione di cui disponiamo è $|\beta_i - \bar{\beta}_i| < \delta$), mentre per α grande l'errore cresce a causa del termine $\frac{\beta_i}{\mu_i}$ al numeratore. La determinazione del valore di α che rende minima la distanza $\|\underline{x} - \underline{x}_\delta\|_2$ non è semplice.

3 T.S.V.D.

Utilizzando le notazioni già introdotte, la soluzione corrispondente al dato perturbato troncata al k-esimo termine, è data da $\underline{x}_\delta = \sum_{i=1}^{k} \frac{\bar{\beta}_i}{\bar{\mu}_i} \underline{v}_i$, e l'errore

$$\underline{x} - \underline{x}_\delta = \sum_{i=1}^{k} \left(\frac{\beta_i - \bar{\beta}_i}{\mu_i} \right) \underline{u}_i + \sum_{i=k+1}^{n} \frac{\beta_i}{\mu_i} \underline{u}_i$$

da cui

$$\|\underline{x} - \underline{x}_\delta\|_2^2 \leq \delta^2 \sum_{i=1}^{k} \left(\frac{1}{\mu_i^2} \right) + \sum_{i=k+1}^{n} \left(\frac{\beta_i}{\mu_i} \right)^2$$
$$= \delta^2 M(k) + N(k)$$

Si constata nuovanente la presenza di un termine ($M(k)$) crescente al crescere di k e di un termine ($N(k)$) decrescente, che rappresenta il compromesso tra approssimazione del dato e regolarità della soluzione.

Accennniamo brevemente che, nel caso della regolarizzazione di Tikhonov, esistono algoritmi per la determinazione del parametro α che non necessitano della conoscenza del livello di errore δ sui dati, informazione che spesso non è conosciuta.
Il più famoso è il metodo della "generalized cross-validation".
L'idea che sta alla base di tale metodo è di scegliere il parametro in modo che, escludendo un valore y_j dal vettore $\underline{y}$ dei dati e considerando solo i rimanenti, l'operatore ottenuto riesca ad approssimare al meglio il dato y_j non utilizzato. Questo procedimento di tipo statistico è ricorrente in calcolo numerico. Si rimanda a [50] per maggiori dettagli.