
POLITECNICO DI TORINO



CORSO DI LAUREA MAGISTRALE IN
INGEGNERIA AEROSPAZIALE

TESI DI LAUREA MAGISTRALE

METODOLOGIE PER L'IDENTIFICAZIONE E LA
REGOLARIZZAZIONE DEI CARICHI ESTERNI SU
STRUTTURE ALARI A PARTIRE DA MISURE DI
DEFORMAZIONE

RELATORI
PROF. MARCO GHERLONE
MARCO ESPOSITO

CANDIDATO
ROBERTO FERRI

DICEMBRE 2021

Indice

| | |
|---|-----------|
| Sommario | 6 |
| Introduzione | 7 |
| 1 Metodologie per la ricostruzione dei carichi esterni | 8 |
| 1.1 Approccio inverso | 8 |
| 1.2 Approccio parametrico | 10 |
| 1.2.1 Estensione tramite serie di Fourier | 13 |
| 1.3 Reti neurali artificiali | 15 |
| 1.3.1 Reti neurali multistrato | 15 |
| 1.3.2 Rete di Elman | 20 |
| 2 Ricostruzione del carico su un modello di piastra | 23 |
| 2.1 Modello | 23 |
| 2.1.1 Configurazione estensimetrica | 25 |
| 2.2 Approccio inverso | 26 |
| 2.3 Rete neurale artificiale | 28 |
| 2.3.1 Creazione dei dati di apprendimento | 28 |
| 2.3.2 Creazione della rete neurale | 29 |
| 2.3.3 Addestramento della rete neurale | 30 |
| 2.3.4 Risultati della rete neurale | 33 |
| 2.4 Confronto dei risultati | 34 |

| | | |
|----------|--|-----------|
| 3 | Regolarizzazione dei problemi inversi mal posti | 36 |
| 3.1 | Definizione del problema | 36 |
| 3.2 | Decomposizione ai Valori Singolari (SVD) | 37 |
| 3.3 | Diagramma di Picard | 39 |
| 3.4 | Curva L | 41 |
| 4 | Metodologie per la regolarizzazione della soluzione | 45 |
| 4.1 | Metodi filtranti | 46 |
| 4.1.1 | TSVD | 47 |
| 4.1.2 | Regolarizzazione di Tikhonov in forma standard | 48 |
| 4.1.3 | Regolarizzazione di Tikhonov in forma generale | 50 |
| 4.2 | Metodi iterativi | 57 |
| 4.2.1 | Metodo di Kaczmarz o ART | 57 |
| 4.2.2 | Metodo CGLS | 58 |
| 4.2.3 | Metodo CGLS preconditionato | 61 |
| 4.2.4 | Metodo LSQR | 62 |
| 4.3 | Metodo ibrido | 64 |
| 4.4 | Scelta del parametro di regolarizzazione | 66 |
| 4.4.1 | Principio di discrepanza | 68 |
| 4.4.2 | Criterio della Curva L | 69 |
| 4.4.3 | Convalida Incrociata Generalizzata (GCV) | 71 |
| 4.4.4 | Analisi NCP | 74 |
| 5 | Regolarizzazione della soluzione su un modello di piastra | 77 |
| 5.1 | Soluzione non regolarizzata | 78 |
| 5.2 | Soluzione regolarizzata | 79 |
| 5.2.1 | Metodo di Tikhonov in forma standard | 79 |
| 5.2.2 | Metodo di Tikhonov del primo ordine | 82 |

| | | |
|----------|--|------------|
| 5.2.3 | Metodo di Tikhonov del secondo ordine | 85 |
| 5.2.4 | Metodo NTR | 88 |
| 5.2.5 | Metodo MTR-1 | 90 |
| 5.2.6 | Metodo MTR-2 | 91 |
| 5.2.7 | Metodo ART | 92 |
| 5.2.8 | Metodo CGLS | 94 |
| 5.2.9 | Metodo M-CGLS | 96 |
| 5.2.10 | Metodo LSQR | 97 |
| 5.2.11 | Metodo ibrido | 98 |
| 5.2.12 | Confronto dei risultati | 101 |
| 5.3 | Analisi statistica | 102 |
| 5.3.1 | Metodo di Tikhonov in forma standard | 103 |
| 5.3.2 | Metodo di Tikhonov del primo ordine | 109 |
| 5.3.3 | Metodo ART | 114 |
| 5.3.4 | Metodo LSQR | 119 |
| 5.3.5 | Confronto dei risultati | 125 |
| 6 | Ricostruzione e regolarizzazione del carico su un cassone alare | 126 |
| 6.1 | Modello | 126 |
| 6.2 | Ricostruzione del carico | 130 |
| 6.3 | Regolarizzazione della soluzione | 131 |
| 6.3.1 | Metodo di Tikhonov in forma standard | 132 |
| 6.3.2 | Metodo di Tikhonov del primo ordine | 137 |
| 6.3.3 | Metodo ART | 143 |
| 6.3.4 | Metodo LSQR | 149 |
| 6.3.5 | Confronto dei risultati | 155 |
| | Conclusioni | 157 |

Bibliografia 158

Ringraziamenti 159

Sommario

In questa tesi verranno analizzate alcune metodologie per l'identificazione dei carichi esterni agenti su strutture alari a partire da misure discrete di deformazione e per la regolarizzazione della soluzione nel caso di misure contenenti errore. L'obiettivo dello studio è confrontare, in termini di ricostruzione del carico e di regolarizzazione della soluzione, le prestazioni di diverse metodologie applicate ad un semplice modello di piastra incastrata ad un'estremità. In particolare, le analisi verranno svolte attraverso simulazioni numeriche effettuate tramite il solutore per Analisi a Elementi Finiti MSC Nastran e la piattaforma di calcolo numerico MATLAB. Nel Capitolo 1 verrà fornita una panoramica su alcune metodologie di ricostruzione del carico partendo da misure discrete di deformazione. Nel Capitolo 2 sarà presente l'applicazione di due metodologie ad un semplice modello di piastra incastrata ad un'estremità e verranno forniti i risultati delle due analisi. Il Capitolo 3 introdurrà la natura mal condizionata dei problemi inversi e la necessità della regolarizzazione della soluzione, oltre ad alcuni strumenti utili per la comprensione del problema come la Decomposizione ai Valori Singolari (SVD) e il Diagramma di Picard. Nel Capitolo 4 verranno introdotte alcune metodologie per la regolarizzazione della soluzione e per la scelta del parametro di regolarizzazione. Nel Capitolo 5 questi metodi saranno utilizzati per la regolarizzazione della soluzione ottenuta dall'approccio inverso (il più accurato tra i due metodi testati nel Capitolo 2) nel caso di misure contenenti errore, verrà svolta un'analisi statistica sui metodi che risulteranno essere i più accurati e saranno presentati i risultati di tutte le analisi. Nel Capitolo 6, infine, sarà presente un'applicazione ad un modello di struttura alare più complesso. Nelle Conclusioni si forniranno alcune osservazioni di carattere generale sui risultati ottenuti e qualche spunto per possibili ulteriori sviluppi.

Introduzione

Il monitoraggio dello stato di salute strutturale (Structural Health Monitoring o SHM) sta assistendo ad un rapido sviluppo negli ultimi decenni e si sta stabilendo come una tecnologia chiave per la manutenzione e il progetto dei velivoli moderni e anche per la definizione delle specifiche di robustezza e rigidità. In particolare, due tecnologie si sono rivelate strumenti cruciali per la realizzazione di un effettivo sistema di monitoraggio strutturale: i metodi di identificazione dei carichi esterni ed i metodi di shape sensing. Queste due famiglie di metodi hanno l'obiettivo di calcolare i carichi esterni ed il campo di spostamenti di una struttura a partire da misure discrete di deformazione. A causa della complessità della struttura alare e delle condizioni di carico, infatti, è difficile determinare carichi alari accurati ed affidabili solo per mezzo di prove in galleria del vento e analisi teoriche e, di conseguenza, si cerca di determinarli tramite prove di volo. Tuttavia, a differenza di parametri come la quota di volo o la velocità, i carichi di volo non possono essere osservati e misurati direttamente e, quindi, nasce un problema inverso, ovvero si richiede di identificare i carichi di volo sulla base di qualche tipo di risposta strutturale dell'ala misurabile, come ad esempio le deformazioni indotte dai carichi di volo. Nonostante la relazione tra carichi e risposta strutturale dell'ala dipenda solamente dalla struttura alare stessa, a causa della complessità di quest'ultima, la relazione non può essere formulata facilmente. La soluzione del problema consiste nel trovare un mezzo per stabilire la relazione carico-deformazione che rappresenti dinamicamente le caratteristiche meccaniche della struttura alare. Il monitoraggio dei carichi esterni e lo shape sensing sono strumenti fondamentali per le operazioni di manutenzione predittiva delle strutture aerospaziali. La conoscenza dei carichi esterni in volo fornisce lo spettro dei carichi in volo, un input fondamentale per i modelli predittivi utilizzati per la stima della vita utile rimanente della struttura. Inoltre, il monitoraggio dei carichi può fornire informazioni sullo stato di salute delle strutture attraverso la determinazione dei cambiamenti dei percorsi di carico causati dalla presenza di danneggiamenti. Lo stesso approccio può essere utilizzato per lo shape sensing: è possibile determinare danneggiamenti all'interno della struttura attraverso la determinazione di anomalie nei campi di spostamento e deformazione. Il costante progresso di queste tecniche, inoltre, è stato supportato dal simultaneo miglioramento ottenuto nel campo della tecnologia della misurazione di deformazione. Lo sviluppo di sistemi di misurazione della deformazione basati su fibre ottiche ha permesso la misura facile ed affidabile di un grande numero di deformazioni discrete, non ottenibili con i classici estensimetri elettrici.

Capitolo 1

Metodologie per la ricostruzione dei carichi esterni

1.1 Approccio inverso

Il metodo di identificazione dei carichi sviluppato da Nakamura in [1] e ripreso in [2] nella definizione di un approccio integrato è in grado di ricostruire la distribuzione di pressione agente sul piano x-y a partire da misure discrete di deformazione. Il metodo è basato sull'interpolazione della distribuzione di pressione tramite elementi finiti: il dominio strutturale, dove la pressione è applicata, è discretizzato utilizzando Elementi Finiti. All'interno dell'elemento, la distribuzione di pressione è interpolata a partire dai valori nodali di pressione utilizzando le funzioni di forma. Matematicamente, ciò può essere formulato come segue:

$$p^e(x, y) = \sum_{i=1}^n N_i^e(x, y)p_i^e \quad (1.1)$$

dove $p^e(x, y)$ è la distribuzione di pressione all'interno dell'elemento, $N_i^e(x, y)$ sono le funzioni di forma dipendenti dal tipo di elemento scelto, n è il numero di nodi dell'elemento e p_i^e sono i valori nodali di pressione. Se il dominio è discretizzato con una mesh di elementi di pressione, le funzioni di forma globali $N_i(x, y)$, relative ai m_p nodi globali della mesh di pressione, possono essere calcolate utilizzando la procedura di assemblaggio adottata dal FEM. Di conseguenza, la distribuzione di pressione sull'intero dominio, $p(x, y)$ può essere espressa come:

$$p(x, y) = \sum_{i=1}^{m_p} N_i(x, y)p_i \quad (1.2)$$

dove p_i è il valore nodale dell' i -esimo nodo della mesh.

Quindi, in analogia con il classico problema strutturale FEM, i gradi di libertà del sistema, che rappresentano le incognite del sistema, sono i valori nodali di pressione p_i . Una volta

calcolati questi valori, la distribuzione di pressione completa può essere derivata tramite interpolazione.

Poiché l'obiettivo del metodo è ricostruire la distribuzione di pressione a partire da misure discrete di deformazione, è necessario trovare una relazione tra le misure discrete di deformazione e i valori nodali di pressione sconosciuti. Se si considera valida l'ipotesi di linearità elastica, la j -esima componente di deformazione misurata ε_j^m può essere espressa come la sommatoria delle ε_{ji}^m deformazioni indotte dalle m_p pressioni nodali p_i . Inoltre, nel regime lineare elastico, l' i -esima deformazione ε_{ji}^m , indotta dall' i -esima p_i , è linearmente proporzionale all' i -esima pressione nodale stessa. Di conseguenza, la j -esima componente di deformazione può essere espressa come:

$$\varepsilon_j^m = \sum_{i=1}^{m_p} \varepsilon_{ji}^m = \sum_{i=1}^{m_p} s_{ji} p_i \quad (1.3)$$

dove s_{ji} è l' i -esimo coefficiente di combinazione lineare incognito che lega la j -esima componente di deformazione all' i -esimo valore nodale di pressione. Quando sono disponibili m_s componenti di deformazione misurate, l'equazione precedente può essere espressa in forma matriciale:

$$\{\varepsilon^m\}_{m_s \times 1} = [s]_{m_s \times m_p} \{p\}_{m_p \times 1} \quad (1.4)$$

La matrice $[s]_{m_s \times m_p}$ può essere facilmente calcolata tramite la risoluzione iterativa di un problema FE diretto. L' i -esima colonna della matrice è calcolata imponendo che $p_i = 1$ e $p_k = 0$ ($k \neq i$) in $p(x, y) = \sum_{i=1}^{m_p} N_i(x, y) p_i$. Il campo di pressione risultante è quindi applicato ad un modello FE della struttura e le componenti desiderate di deformazione vengono misurate. Iterando la procedura alle m_p colonne, è possibile calcolare l'intera matrice. L'applicazione di questa procedura richiede la costruzione di un accurato modello FE della struttura. Quindi, la geometria, le proprietà dei materiali e i vincoli della struttura devono essere conosciuti. In generale, la mesh strutturale e quella di pressione condividono la discretizzazione di una porzione o di tutto il dominio strutturale e possono avere caratteristiche molto differenti. È necessario, tuttavia, essere in grado di associare la posizione dei nodi appartenenti alla mesh di dettaglio agli elementi della mesh di pressione, in modo da determinare i valori di pressione in questa posizione.

Dopo aver calcolato la matrice $[s]$, l'espressione delle pressioni nodali come funzione delle deformazioni misurate è ottenuta invertendo l'equazione 1.4. Poiché nelle applicazioni pratiche $m_s \neq m_p$, l'inversione di $[s]$ è ottenuta per mezzo della formulazione della matrice pseudo inversa di Moore-Penrose, $[s^+]$:

$$\{p\} = [s^+] \{\varepsilon^m\} \quad (1.5)$$

Sostituendo i valori di deformazione misurati in $\{\varepsilon^m\}$, l'espressione fornisce i valori nodali di pressione che meglio approssimano la distribuzione di pressione applicata che induce le deformazioni misurate. Questi valori nodali possono essere utilizzati per ricostruire l'intero campo di pressione utilizzando l'espressione 1.2.

Lo stesso principio può essere adottato per l'identificazione dei carichi concentrati. In

questo caso, la formulazione non richiede l'interpolazione dei valori nodali del carico all'interno degli elementi: i carichi ricercati, infatti, sono concentrati in posizioni specifiche; pertanto, non devono essere espansi in un dominio spaziale. Assumendo come valida l'ipotesi di linearità elastica, la j -esima componente di deformazione ε_j^m indotta dall' i -esimo carico concentrato F_i è linearmente proporzionale al carico stesso. Inoltre, se sono presenti m_f forze concentrate, la componente di deformazione può essere espressa come la sommatoria delle deformazioni indotte da ciascun carico concentrato ed il vettore delle deformazioni misurate può essere espresso come:

$$\{\varepsilon^m\}_{m_s \times 1} = [s_f]_{m_s \times m_f} [F]_{m_f \times 1} \quad (1.6)$$

dove $[s_f]_{m_s \times m_f}$ è la matrice dei coefficienti di influenza tra le deformazioni e i carichi concentrati. L' i -esima colonna della matrice è calcolata imponendo che $F_i = 1$ e $F_k = 0$ ($k \neq i$) e misurando le m_s componenti di deformazioni indotte. La pseudo inversione di $[s_f]$ permette il calcolo del vettore delle forze concentrate che induce le deformazioni misurate:

$$\{F\} = [s_f^+]\{\varepsilon^m\} \quad (1.7)$$

1.2 Approccio parametrico

Questo metodo è stato introdotto da Shkarayev in [3]. Si consideri un velivolo che svolge una manovra o è soggetto a turbolenza atmosferica. Un cambiamento delle forze interne ed esterne causa spostamenti lineari ad angolari, deformazioni e tensioni in ogni punto della struttura. I sensori sono inseriti all'interno della struttura lungo specifici percorsi, permettendo la misura delle componenti di deformazione in posizioni specifiche. I sensori di deformazione misurano i cambiamenti nelle componenti di deformazione ε^* in n posizioni ed immagazzinano queste informazioni in un computer di bordo. Questi dati vengono utilizzati come input nell'analisi inversa del modello computazionale basato sul metodo agli elementi finiti. Poiché questo approccio è basato sul metodo agli elementi finiti, è necessario sviluppare un modello agli elementi finiti del velivolo.

I carichi esterni vengono applicati in maniera incrementale al velivolo. Ogni cambiamento di forza agente sul velivolo è assunto piccolo. Ad ogni incremento, le equazioni di governo dell'approssimazione lineare agli elementi finiti ha la seguente forma:

$$[K]\{U\} = \{P\} \quad (1.8)$$

dove $[K]$ è la matrice di rigidità, $\{U\}$ è il vettore degli spostamenti nodali e $\{P\}$ è il vettore delle forze nodali equivalente all'incremento di carico corrente.

Le deformazioni nodali possono essere espresse in termini degli spostamenti nodali utilizzando la matrice deformazione-spostamento $[B]$:

$$\{\varepsilon\} = [B]\{U\} \quad (1.9)$$

Dato un carico distribuito su una superficie, $p(s)$, le forze nodali vengono espresse come

$$\{P\} = \int_s [N]^T p(s) ds \quad (1.10)$$

dove l'integrazione è effettuata sulla superficie della struttura s .

Il problema inverso è formulato come segue. Innanzitutto, è necessario stabilire un set di casi di carico possibili, m . Per l' i -esimo caso di carico, l'approssimazione del carico è espressa in forma parametrica come

$$F_i(s) = \sum_{j=1}^l a_{ij} R_{ij}(s) \quad (1.11)$$

dove $R_{ij}(s)$ sono le funzioni di distribuzione spaziale e a_{ij} sono i parametri di approssimazione sconosciuti. Questi parametri sconosciuti rappresentano alcuni parametri di volo come il fattore di carico, l'angolo di attacco e la velocità. Alcuni vincoli possono essere imposti a questi parametri:

$$\varphi(a_{ij}) \geq 0 \quad (1.12)$$

L'approccio prosegue con un'analisi agli elementi finiti diretta svolta per ogni i -esimo caso di carico. I corrispondenti spostamenti U_{ij} e deformazioni ε_{ij} sono determinati come

$$\{U_{ij}\} = [K]^{-1} \int_s [N]^T R_{ij}(s) ds \quad (1.13)$$

$$\{\varepsilon_{ij}\} = [B][K]^{-1} \int_s [N]^T R_{ij}(s) ds \quad (1.14)$$

Le tensioni vengono successivamente calcolate tramite le relazioni costitutive di Hooke

$$\{\sigma_{ij}\} = [D]\{\varepsilon_{ij}\} \quad (1.15)$$

dove $[D]$ rappresenta la matrice di elasticità. Gli spostamenti, le deformazioni e le tensioni corrispondenti all' i -esimo carico sono

$$\{U_i\} = \sum_{j=1}^m a_{ij} \{U_{ij}\} \quad (1.16)$$

$$\{\varepsilon_i\} = \sum_{j=1}^m a_{ij} \{\varepsilon_{ij}\} \quad (1.17)$$

$$\{\sigma_i\} = \sum_{j=1}^m a_{ij} \{\sigma_{ij}\} \quad (1.18)$$

I parametri di interpolazione del carico vengono calcolati utilizzando una procedura di minimizzazione ai minimi quadrati:

$$S_i = \{\varepsilon_i - \varepsilon_i^*\}^T \{\varepsilon_i - \varepsilon_i^*\} = \left\{ \sum_{j=1}^m a_{ij} \varepsilon_{ij} - \varepsilon_i^* \right\}^T \left\{ \sum_{j=1}^m a_{ij} \varepsilon_{ij} - \varepsilon_i^* \right\} \quad (1.19)$$

La minimizzazione di questa equazione rispetto ai parametri a_{ij} , tenendo conto dei vincoli, risulta in un sistema lineare algebrico di equazioni risolto per i a_{ij} coefficienti. In questo modo, un set di soluzioni di a_{ij} ottenuto per m casi di carico.

È chiaro che è necessario stabilire un adeguato criterio per selezionare l'appropriato caso di carico. Un criterio utilizza una quality function, $Q(S_i, m, n)$, in cui sono utilizzati i valori calcolati di S_i . Se i polinomi di Chebyshev vengono utilizzati per rappresentare le funzioni $R_{ij}(s)$, la quality function può essere descritta come

$$Q(S_i, m, n) = \frac{S_i}{1 - \sqrt{\frac{1}{n} \{ (m+1) [\ln(\frac{n}{m+1} + 1)] - \ln H \}}} \quad (1.20)$$

dove $1-H$ è la probabilità per cui la stima è valida. Un minimo della quality function corrisponde al k-esimo caso di carico, fornendo la miglior soluzione al problema inverso complessivo.

Per riassumere, il presente approccio segue una procedura a due fasi. La prima fase consiste nello sviluppo del modello agli elementi finiti e nell'analisi diretta di esso sottoposto al carico applicato $p(s)$. Le deformazioni superficiali ε^* vengono poi calcolate in n posizioni specifiche e queste quantità rappresentano le deformazioni "misurate" sperimentalmente. Nella seconda fase dell'analisi, il metodo inverso utilizza le deformazioni misurate ε^* e l'approssimazione parametrica del carico $F_i(s)$. Questo metodo, infatti, richiede la formulazione della distribuzione di pressione sconosciuta come combinazione lineare di distribuzioni di pressione note moltiplicate per coefficienti incogniti. Successivamente, vengono studiati i campi di deformazione ε_{ij} causati da ogni distribuzione di pressione nota $R_{ij}(s)$ e il campo di deformazione indotto dal carico complessivo viene definito come combinazione lineare dei campi di deformazione precedentemente studiati

$$\{\varepsilon_i\} = \sum_{j=1}^m a_{ij} \{\varepsilon_{ij}\} \quad (1.21)$$

Infine, tramite una procedura di minimizzazione ai minimi quadrati riguardante le deformazioni calcolate e quelle "misurate" sperimentalmente, si determinano i coefficienti

incogniti dell'approssimazione parametrica.

1.2.1 Estensione tramite serie di Fourier

Questo metodo rappresenta un'estensione del metodo precedente ed è stato introdotto in [4, 5]. Esso cerca di identificare i carichi in volo sulla base di informazioni fornite da estensimetri in tempo reale tramite un metodo di interpolazione inverso che utilizza i risultati di un modello agli elementi finiti (FEM). L'interpolazione inversa è basata sulla minimizzazione dell'errore tra i valori calcolati e misurati di deformazione. Le deformazioni ottenute tramite l'analisi agli elementi finiti sono utilizzate come deformazioni misurate "sperimentalmente". Le deformazioni indotte dai carichi di superficie rappresentati dai singoli termini di Fourier sono determinati tramite analisi agli elementi finiti. Le informazioni di deformazione indotte dal carico applicato ignoto accoppiate alle informazioni FEM di deformazione indotte dai carichi di Fourier permette la predizione dei coefficienti di Fourier del carico applicato. I coefficienti di Fourier appena calcolati vengono confrontati con gruppi di coefficienti di Fourier presenti in un database contenente dati teorici e storici e una minimizzazione ai minimi quadrati permette di determinare il set di coefficienti più probabile.

Shkarayev suggerisce di utilizzare i polinomi di Chebyshev per rappresentare le funzioni spaziali R_{ij} e una quality function basata su simulazioni numeriche e statistiche sperimentali per selezionare il caso di carico più appropriato. I polinomi di Chebyshev interpolano qualunque $f(x)$ nell'intervallo $[-1, 1]$, tuttavia un cambiamento di variabili permette l'estensione dell'interpolazione ad un generico intervallo $[a, b]$. Se si considera un caso di carico in cui sono presenti diverse funzioni in diversi intervalli, un modello di approssimazione basato sui polinomi di Chebyshev richiede diversi gruppi di polinomi e coefficienti in ogni intervallo, oltre che un diverso cambiamento di variabili su ciascun intervallo. Le serie di Fourier sono in grado di approssimare diverse funzioni continue su intervalli discreti attraverso un singolo gruppo di coefficienti e, pertanto, risultano più efficaci da un punto di vista computazionale. Questo metodo propone di esprimere l'approssimazione del carico in termini di serie di Fourier e le funzioni spaziali, di conseguenza, sono $R_{ij} \in \{1, \cos \frac{n\pi x}{L}, \cos \frac{n\pi y}{K}\}$ dove L e K sono la lunghezza e la larghezza, rispettivamente, dell'ala. Attraverso le informazioni di deformazione, questo metodo cerca di predire i coefficienti di Fourier e li compara con quelli ottenuti da un database contenente dati teorici/storici. La minimizzazione dell'errore tra il set di coefficienti misurati e quelli presenti nel database permette di identificare l'appropriato modello di carico dato un certo set di deformazioni. Questa minimizzazione può anche essere effettuata con vincoli di peso: per esempio, i tipi di carico statisticamente significanti o un tipo di carico previsto dominante sulla base dell'ambiente o posizione può avere pesi più grandi. Il carico superficiale $p(x, y)$ può essere approssimato tramite una serie cosinusoidale singola di Fourier:

$$p(x, y) \cong a_0 + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi x}{L} \quad (1.22)$$

La serie cosinusoidale è stata scelta poiché è in grado di soddisfare le probabili condizioni

finali che il carico sia piccolo all'estremità e grande alla radice. È stata selezionata la semiapertura alare poiché il carico potrebbe non essere sempre simmetrico e, di conseguenza, i set di dati potrebbero differire da una semiala all'altra.

Il database consiste nei coefficienti di Fourier a_{ij} relativi a m possibili casi di carico determinati attraverso dati teorici o storici. I coefficienti a_{ij} sono raccolti nella matrice A :

$$[A] = [\{a\}_1^* \quad \{a\}_2^* \quad \{a\}_2^* \quad \dots \quad \{a\}_m^*] \quad (1.23)$$

Sulla base della storia del carico, è possibile rappresentare i vincoli di peso tramite w_i che vengono raccolti nella matrice w :

$$[w] = [w_1 \quad w_2 \quad w_3 \quad \dots \quad w_m] \quad (1.24)$$

Se un carico incognito produce un set di deformazioni, questi valori possono essere utilizzati nell'equazione

$$S_i = \{\varepsilon_i - \varepsilon_i^*\}^T \{\varepsilon_i - \varepsilon_i^*\} = \left\{ \sum_{j=1}^m a_{ij} \{\varepsilon_{ij}\} - \varepsilon_i^* \right\}^T \left\{ \sum_{j=1}^m a_{ij} \{\varepsilon_{ij}\} - \varepsilon_i^* \right\} \quad (1.25)$$

per ottenere i coefficienti di Fourier $\{a\}$ del carico incognito. Effettuando una minimizzazione ai minimi quadrati dell'errore tra i coefficienti di Fourier calcolati e i set di coefficienti del database, si ottiene

$$S_i = [\{a\}_i^* - \{a\}]^T [\{a\}_i^* - \{a\}] \quad (1.26)$$

$$S_w = (1 - w_i) S_i \quad (1.27)$$

La minimizzazione della prima equazione accoppiata con il criterio della seconda fornisce i coefficienti a_{ij} della serie di Fourier più rappresentativa del carico.

Un'ulteriore estensione del metodo consiste nell'approssimare il carico in una serie doppia di Fourier:

$$p(x, y) \cong a_{00} + \sum_{m=1}^{\infty} a_{0m} \cos \frac{m\pi y}{K} + \sum_{n=1}^{\infty} a_{0n} \cos \frac{n\pi x}{L} + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{nm} \cos \frac{n\pi x}{L} \cos \frac{m\pi y}{K} \quad (1.28)$$

1.3 Reti neurali artificiali

Lo scopo è quello di sviluppare un approccio per l'identificazione dei carichi agenti sull'ala di un velivolo che utilizza una rete neurale artificiale per modellizzare la relazione carico-deformazione nell'analisi strutturale. In particolare, si dimostra che l'utilizzo di una rete neurale artificiale per l'identificazione dei carichi è possibile e, inoltre, una rete neurale ben addestrata rivela una convergenza estremamente veloce ed un elevato grado di accuratezza nel processo di identificazione del carico.

Le reti neurali artificiali hanno attirato una grande attenzione per la modellazione di complesse relazioni non lineari. Esse sono derivate da una modellizzazione del cervello umano e sono composte da un certo numero di unità (neuroni artificiali) interconnessi. Un grande vantaggio dell'utilizzo di un modello basato su reti neurali è l'abilità di imparare relazioni tra variabili attraverso l'esposizione ripetuta a queste variabili. Di conseguenza, invece di utilizzare una relazione analitica derivata da principi meccanici, la rete neurale artificiale impara la relazione attraverso un processo di apprendimento adattivo.

1.3.1 Reti neurali multistrato

Tra vari tipi di architetture di reti neurali artificiali, le reti neurali multistrato presentate in [6] hanno alcune caratteristiche interessanti: sono in grado di costruire una funzione di mappatura non lineare da diversi dati di input a diversi dati di output attraverso il processo di apprendimento; la rete addestrata ha una caratteristica di "generalizzazione", cioè un tipo di interpolazione tale per cui una rete ben addestrata è in grado di stimare dati di output appropriati anche per modelli non appresi; la rete addestrata opera velocemente in un processo di applicazione. Inoltre, sono in grado di considerare sia variabili discrete sia variabili continue e una delle loro caratteristiche principali è il processamento di informazioni in parallelo. Tenendo conto dei vantaggi appena descritti, questo metodo propone un approccio per identificare i carichi di volo agenti su un'ala utilizzando le reti neurali artificiali come impianto per costruire un modello della relazione tra carico e deformazione. Questo approccio è caratterizzato dai seguenti quattro sotto-processi:

1. *Test di calibrazione a terra*

Un'effettiva ala di un velivolo è utilizzata come oggetto di calibrazione prima della prova di volo e vari carichi distribuiti vengono applicati come campioni di calibrazione che vengono selezionati sulla base di analisi teoriche e dell'involuppo di volo progettato. Le risposte dell'ala in termini di deformazione vengono misurate in modo da ottenere dati che verranno usati nella fase successiva come dati di apprendimento insieme ai carichi applicati. Questa prima fase, quindi, consiste nel preparare una vasta quantità di modelli di apprendimento per le reti neurali artificiali. Anche se i set di dati di apprendimento ottenuti sono misure discrete, la caratteristica meccanica della struttura alare è contenuta nei set di dati.

2. *Addestramento della rete neurale artificiale*

Utilizzando le deformazioni come segnali di input forniti alle unità di input della rete e i carichi come segnali di insegnamento (output desiderati), l'addestramento della rete neurale artificiale viene effettuata iterativamente fino a quando l'errore tra l'output attuale e quello desiderato raggiunge un livello accettabile. Poi, la configurazione della rete per l'identificazione del carico è decisa.

3. *Test di volo*

Misura delle deformazioni indotte dai carichi aerodinamici attraverso prove di volo.

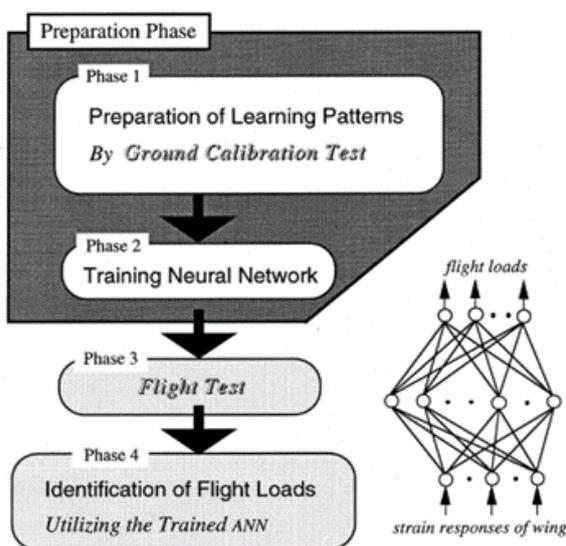


Figura 1.1: Utilizzo rete neurale [6]

4. *Identificazione dei carichi di volo*

Introdurre le deformazioni misurate in volo nella rete neurale artificiale ben addestrata e identificare i carichi di volo attraverso gli output della rete neurale.

Si può notare come una complicata analisi strutturale possa essere sostituita da un modello di rete neurale con questo approccio. Inoltre, è possibile eseguire le prove di calibrazione a terra utilizzando il velivolo in oggetto per ottenere i dati di apprendimento per le reti neurali e queste ultime possono apprendere da questi modelli in maniera efficiente ed accurata anche se i modelli di apprendimento sono costituiti da dati discreti.

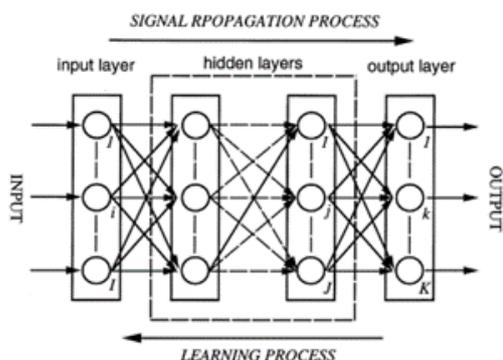


Figura 1.2: Rete neurale multistrato [6]

Poiché le strutture, in qualche caso, sono estremamente complicate e le proprietà meccaniche o le relazioni tra carichi esterni e risposte strutturali sono difficili da formulare, il modello di rete neurale artificiale diventa più adeguato evitando analisi teoriche complicate o quasi impossibili. Per costruire una rete neurale artificiale, è necessario prima decidere che tipo di rete scegliere, quante unità usare e che tipo di unità è appropriato.

A causa delle caratteristiche menzionate prima, inizialmente si considera una rete neurale

multistrato con uno strato di input, uno di output e alcuni strati nascosti.

Le reti neurali multistrato possono rappresentare qualunque funzione quando sono presenti abbastanza unità. Questo tipo di rete neurale è considerato una rete completamente connessa, in cui ogni input influenza tutti gli elementi di output. Il numero di neuroni negli strati di input e output è solitamente scelto in base al problema analizzato. Il numero degli strati nascosti e dei neuroni in ogni strato sono definiti dopo aver considerato il tipo di problema, la velocità computazionale e l'accuratezza. Ogni neurone è un elemento computazionale fondamentale. I neuroni dello strato di input operano in modo diverso rispetto a quelli degli strati nascosti e di output.

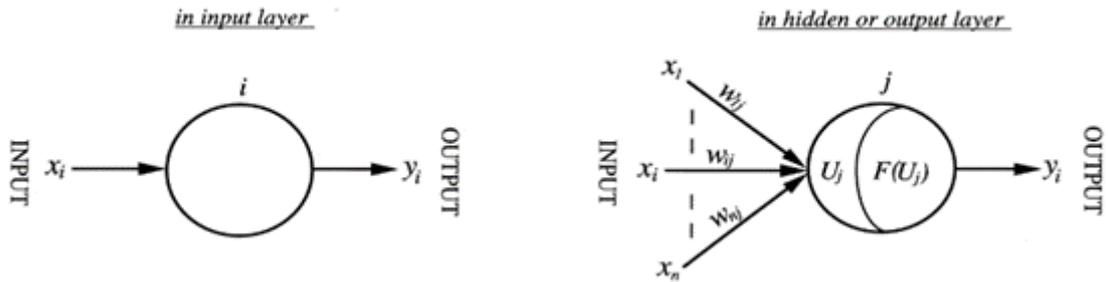


Figura 1.3: Tipologie di neuroni [6]

I neuroni appartenenti allo strato di input emettono direttamente il segnale in entrata:

$$y_i = x_i \quad (1.29)$$

dove il pedice i indica l' i -esimo neurone dello strato di input.

Ogni neurone appartenente allo strato di output o a quello nascosto, invece, forma la somma pesata

$$\sum_{i=1}^n w_{ij} x_i \quad (1.30)$$

di n input dallo strato precedente e si aggiunge un bias θ_j :

$$U_j = \sum_{i=1}^n w_{ij} x_i + \theta_j \quad i = 1, 2, \dots, n \quad (1.31)$$

Questa somma diventa il segnale di input dell'unità di processamento. Gli elementi computazionali processano e passano i risultati attraverso la funzione di attivazione F per ottenere l'output y_j :

$$y_j = F(U_j) \quad F(U_j) = F\left(\sum_{i=1}^n w_{ij} x_i + \theta_j\right) \quad i = 1, 2, \dots, n \quad (1.32)$$

Il coefficiente w_{ij} è il peso nella connessione dall'unità i dello strato precedente all'unità j dello strato attuale e il coefficiente θ_j è il bias. Gli strati nascosti non interagiscono con

l'ambiente esterno e semplicemente ricevono o emettono l'informazione da o ai neuroni nel sistema. L'output di un singolo neurone può diventare l'input per altri neuroni o può semplicemente essere uno degli output della rete. Si possono ottenere diversi modelli utilizzando diverse funzioni matematiche per F . In questo studio la funzione sigmoidea data da

$$F(U_j) = \frac{1}{1 + \exp\left(\frac{-U_j}{T}\right)} \quad (1.33)$$

è adottata come funzione di attivazione. La funzione sigmoidea ha la conveniente proprietà che la derivata $F'(U_j) = F(U_j)[1 - F(U_j)]$ è facilmente calcolabile. La funzione di attivazione è l'unica sorgente di introduzione di non linearità nella relazione input-output e, quindi, l'applicazione della funzione sigmoidea alla combinazione lineare U_j produce un output che è una funzione non lineare delle variabili di input. Si può dimostrare che, in assenza di funzioni di attivazione non lineari, una rete multistrato può essere ridotta in una rete equivalente monostrato. T è un parametro usato per modulare l'output dell'unità: nella figura sottostante 1.4 si nota come l'output della funzione sigmoidea vari con il valore di T :

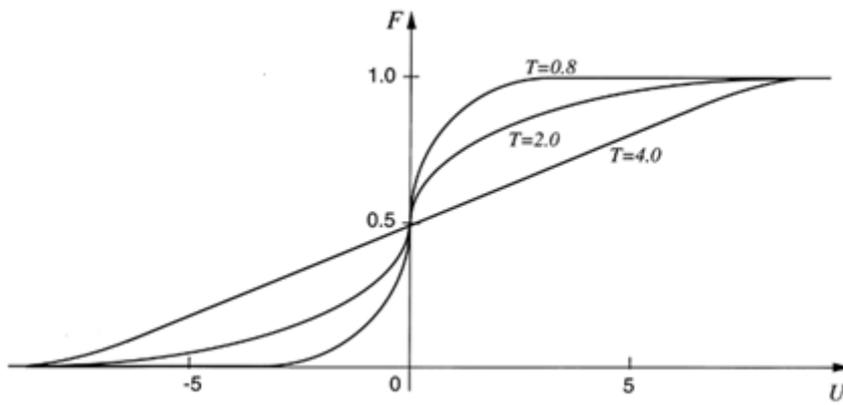


Figura 1.4: Funzione sigmoidea [6]

Il principale vantaggio di questa funzione è la sua abilità di gestire segnali di input grandi e piccoli. La pendenza di questa funzione è rappresentativa del guadagno disponibile. Per grandi valori del segnale di input (sia positivi sia negativi), il guadagno è abbastanza piccolo e per valori intermedi del segnale di input il guadagno è finito. Quindi, un appropriato livello di guadagno è ottenuto per un ampio range di segnali di input.

Un altro parametro è il bias θ_j o l'offset di input nell'unità. Sebbene questo possa essere ottenuto semplicemente aggiungendo un input costante con un peso appropriato, spesso il bias è considerato separatamente. I biases possono essere usati, per esempio, per inibire selettivamente l'attività di certi neuroni. Il bias, inoltre, influenza il grado di non linearità traslando l'input lontano dalla regione lineare della funzione sigmoidea.

Algoritmo di apprendimento

La rete neurale deve essere addestrata per riconoscere modelli noti ed estrapolare risultati esatti da questi modelli quando l'input è una nuova informazione. Per le reti neurali multistrato, l'addestramento consiste nell'aggiornamento continuo dei pesi e dei biases. Il processo di addestramento ha come obiettivo la minimizzazione dell'errore quadratico medio tra l'output attuale e quello desiderato dei neuroni nello strato di output. L'addestramento termina e la configurazione è confermata quando ogni errore delle unità nello strato di output raggiunge un livello accettabile.

Un algoritmo utilizzato per addestrare la rete è l'Improved Error Back Propagation (IEBP). Se si considera il k-esimo neurone nello strato di output, il suo errore può essere scritto come

$$e_k = [D_k - y_k] \quad (1.34)$$

dove D_k e y_k sono l'output desiderato e quello attuale rispettivamente del neurone k-esimo. L'errore quadratico medio E_k è espresso come:

$$E_k = \frac{1}{2}e_k^2 = \frac{1}{2}[D_k - y_k]^2 \quad (1.35)$$

È da notare che l'errore può essere definito solo per i neuroni dello strato di output. Per i neuroni appartenenti agli altri strati, questo confronto è impossibile. Questo è il motivo per cui i pesi e i biases devono essere aggiustati partendo dai nodi nello strato di output e tornando indietro a quelli negli strati nascosti.

La funzione di errore E_k può essere considerata una funzione quadratica multi-dimensionale dei coefficienti di connessione w_{ij} e biases θ_k nello spazio parametrico. Il suo numero dimensionale, espresso come M, è il numero totale dei pesi e biases. Quindi, la superficie errore è una superficie non lineare nello spazio M-dimensionale. Il minimo di E_k che qualche volta potrebbe non essere il minimo globale desiderato ma un minimo locale, viene determinato risolvendo le equazioni

$$\frac{\partial E_k}{\partial w_{jk}} = 0 \quad (1.36)$$

$$\frac{\partial E_k}{\partial \theta_k} = 0 \quad (1.37)$$

Il processo di apprendimento delle reti neurali è un processo di graduale adattamento dei coefficienti w_{jk} e θ_k secondo le equazioni precedenti. Definendo

$$\delta_k = [D_k - y_k]y_k[1 - y_k]\frac{1}{T} \quad (1.38)$$

come un termine di errore del neurone k -esimo nello strato di output e utilizzando una tecnica di discesa del gradiente per minimizzare la differenza quadratica media tra l'output desiderato e quello attuale, i pesi e i biases vengono aggiustati rispettivamente tramite le seguenti espressioni:

$$w_{jk}(t+1) = w_{jk}(t) + \eta \delta_k x_j \quad (1.39)$$

$$\theta_k(t+1) = \theta_k(t) + \gamma \delta_k \quad (1.40)$$

dove t indica il ciclo di apprendimento. I coefficienti η e γ rappresentano le velocità di apprendimento dei pesi e biases rispettivamente. L'algoritmo di apprendimento appena presentato è chiamato algoritmo EBP. Questo algoritmo può essere migliorato aggiungendo i termini di momento

$$\alpha \Delta w_{jk}(t-1) = \alpha [w_{jk}(t) - w_{jk}(t-1)] \quad (1.41)$$

$$\beta \Delta \theta_k(t-1) = \beta [\theta_k(t) - \theta_k(t-1)] \quad (1.42)$$

alle equazioni precedenti, per ottenere

$$w_{jk}(t+1) = w_{jk}(t) + \eta \delta_k x_j + \alpha [w_{jk}(t) - w_{jk}(t-1)] \quad (1.43)$$

$$\theta_k(t+1) = \theta_k(t) + \gamma \delta_k + \beta [\theta_k(t) - \theta_k(t-1)] \quad (1.44)$$

Questo nuovo algoritmo è chiamato Improved Error Back Propagation, che è abbreviato in algoritmo IEBP. I coefficienti α e β sono i coefficienti di momento rispettivamente per i pesi e biases. Le variazioni dei pesi e biases possono essere appianate aggiungendo i termini di momento. La velocità di convergenza e la stabilità dell'algoritmo vengono aumentate. In più, ciò aiuta l'algoritmo a non convergere ad un ottimo locale e guida la convergenza verso quello globale.

1.3.2 Rete di Elman

Un altro tipo di rete neurale utilizzata per stimare e predire i carichi aerodinamici partendo da dati di risposta strutturale è stata proposta da Carn e Trivailo in [7]. Per simulare le condizioni di carico aerodinamiche, sono stati utilizzati i dati del test di

fatica dell'impennaggio verticale di un F/A-18 International Follow-On Structural Test Project (IFOSTP) condotto dalle Royal Australian Air Forces e Canadian Forces che includono sia l'applicazione dei carichi di buffet ad alta frequenza sia i carichi di manovra a bassa frequenza. La rete neurale è stata addestrata utilizzando i dati di risposta ottenuti da diversi estensimetri come input e i carichi applicati noti come output.

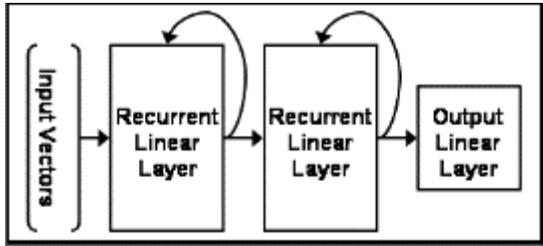


Figura 1.5: Rete di Elman [7]

Sul campione sono stati installati 19 estensimetri elettrici, 9 attuatori di carico e 2 shakers. Per questi canali di deformazione/carico, è stato selezionato un set di dati di 200,000 punti per canale di deformazione/carico che corrisponde a 5 minuti e 33 secondi di tempo di prova. Questo è stato poi diviso in un set di 160,000 punti che è stato utilizzato per addestrare la rete neurale e un set di dati di 40,000 punti che è stato utilizzato per testare la rete.

La rete utilizzata è una rete di Elman con tre strati lineari. Il primo strato di input contiene 19 neuroni (corrispondenti alle 19 variabili di input), lo strato nascosto ne contiene 8 e quello di output è a singolo neurone. Ogni neurone contiene una funzione di trasferimento lineare.

Il primo strato della rete è uno strato ricorrente di neuroni lineari che accetta input dai vettori di input della rete. L'output del primo strato è collegato all'input del secondo strato ed è anche immesso nuovamente in sé stesso con un fattore di ritardo.

Il secondo strato contiene neuroni lineari e ha come input una combinazione di output del primo strato e il suo output stesso con un fattore di ritardo temporale. L'output del secondo strato è anche immesso nello strato output che contiene un singolo neurone lineare. Lo strato di output non è ricorrente e il suo output produce il vettore di output della rete.

L'architettura degli strati ricorrenti ha la seguente struttura:

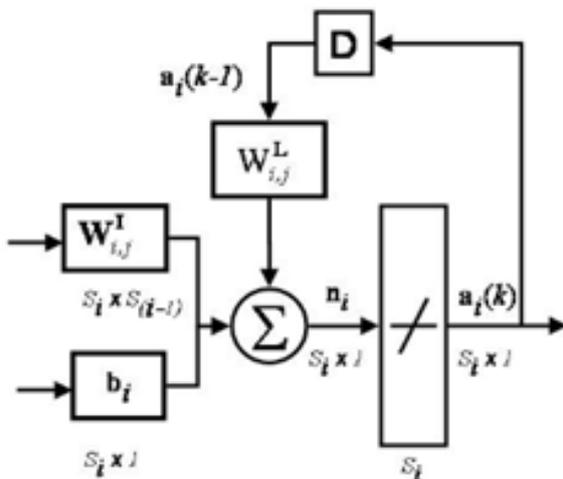


Figura 1.6: Strato ricorrente [7]

$a_i(k)$ è il vettore di output dello strato i -esimo ed è dipendente dal tempo k ;
 $W_{i,j}^I$ è il vettore dei pesi di input per l' i -esimo strato e il j -esimo input;
 $W_{i,j}^L$ è il vettore dei pesi ricorrenti per l' i -esimo strato e il j -esimo input;
 b_i è il bias per l' i -esimo strato;
 s_i è il numero di elementi dello strato considerato;
 D è il fattore di ritardo per gli input ricorrenti.

Lo strato di output, invece, non è ricorrente e ha la seguente struttura:

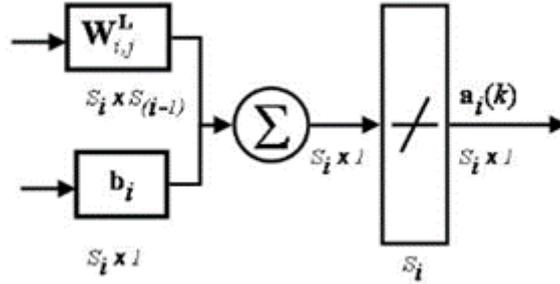


Figura 1.7: Strato non ricorrente [7]

Ogni strato lineare ricorrente ($i=1, 2$) lavora secondo l'equazione:

$$a^i(k) = f(W_{i,j}^I p_i + W_{i,j}^L a_i(k-1) + b_i) \quad (1.45)$$

dove f è la funzione di trasferimento lineare e p_i è l'input dell' i -esimo strato. Lo strato di output ($i=3$) lavora secondo l'equazione:

$$a^i(k) = f(W_{i,j}^L p_i + b_i) \quad (1.46)$$

dove f è la stessa funzione di trasferimento.

Per il primo strato della rete ($i=1$), l'input è il vettore dei valori di deformazione per un numero j di estensimetri e per un intervallo temporale:

$$p_i = \varepsilon \quad (1.47)$$

dove ε è la matrice di deformazione di input.

Per il secondo e terzo strato della rete ($i=2, 3$), l'input è preso dall'output degli strati precedenti:

$$p_i = a_{(i-1)}(k) \quad (1.48)$$

La seguente equazione rappresenta il calcolo del carico nella posizione j , L_j , come funzione della matrice di deformazione ε :

$$L_j = f(W_{3,j}^I [f(W_{2,j}^I [f(W_{1,j}^I \varepsilon + W_{1,j}^L a_1(k-1) + b_1]) + W_{2,j}^L a_2(k-1) + b_2]) + b_3) \quad (1.49)$$

La rete è stata addestrata usando la funzione MATLAB 'traingdx' con la metodologia della discesa del gradiente e l'algoritmo back-propagation con momento e velocità di apprendimento adattiva.

Capitolo 2

Ricostruzione del carico su un modello di piastra

In questo capitolo verranno applicate due metodologie introdotte nel Capitolo 1, ovvero l'approccio inverso e quello delle reti neurali artificiali, per la ricostruzione dei carichi esterni agenti su un modello di piastra. Le analisi verranno condotte attraverso simulazioni numeriche agli elementi finiti effettuate con il solutore MSC Nastran, attraverso un codice MATLAB che implementa l'approccio inverso e la Neural Net Fitting App di MATLAB per l'implementazione e l'utilizzo di una rete neurale artificiale.

2.1 Modello

Il modello analizzato consiste in una piastra rettangolare incastrata ad un'estremità e libera all'altra, su cui agisce il carico distribuito bidimensionale

$$p(x, y) = (1 - (0.1x)^2)(1 - (0.33y)^2) \quad (2.1)$$

Le dimensioni dei lati della piastra sono $a = 100mm$ e $b = 3mm$ ed, inoltre, essa è costituita da materiale isotropo con caratteristiche $E = 73GPa$ e $\nu = 0.3$. L'obiettivo è quello di ricostruire tale carico a partire da misure discrete di deformazione attraverso le due metodologie sopra citate. Le misure discrete di deformazione non sono state ottenute sperimentalmente ma tramite analisi agli elementi finiti, immaginando che gli estensimetri siano posizionati esattamente in corrispondenza dei nodi della mesh. Per l'applicazione delle due metodologie è necessario introdurre due mesh sul modello, ovvero una mesh strutturale e una mesh di pressione, con dimensioni, rispettivamente, di $0.5mm$ e $1mm$ ed entrambe costituite da elementi CQUAD4 con spessore $1mm$. Le caratteristiche del modello sono riassunte nella tabella 2.1

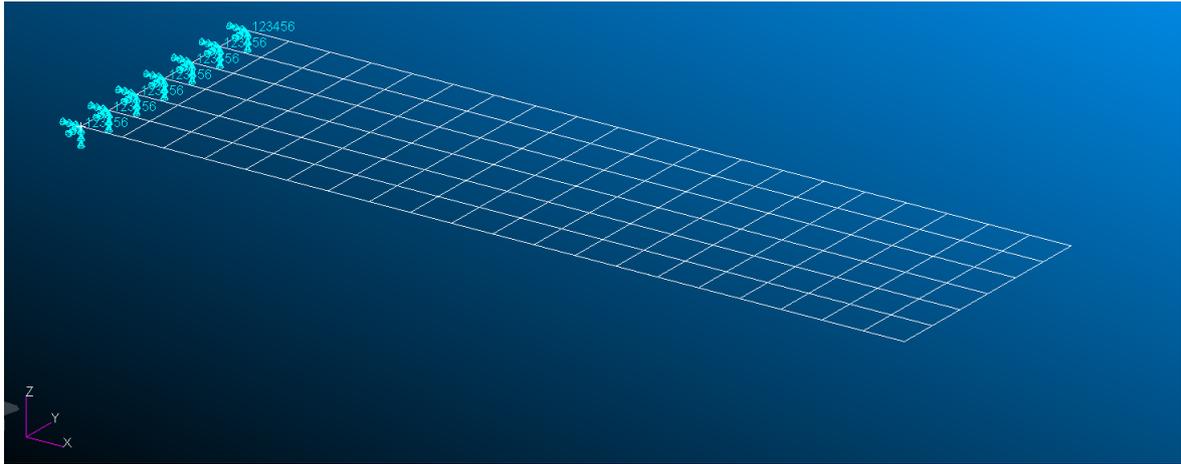


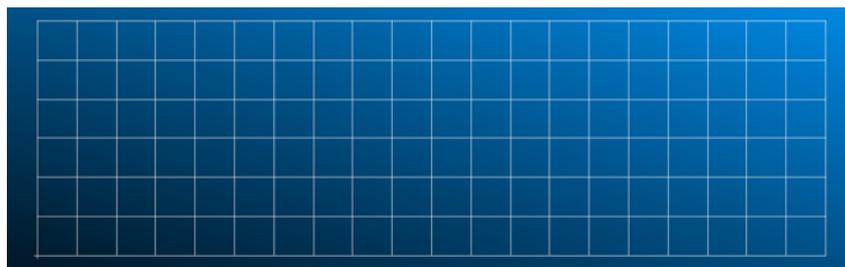
Figura 2.1: Modello

| | |
|-----------------------------------|-----------|
| Dimensioni piastra [mm] | a=10, b=3 |
| Modulo di Young (E) [GPa] | 73 |
| Coefficiente di Poisson | 0.3 |
| Dimensione mesh di pressione [mm] | 1 |
| Dimensione mesh strutturale [mm] | 0.5 |

Tabella 2.1: Caratteristiche del modello



(a)



(b)

Figura 2.2: (a) Mesh di pressione (b) Mesh strutturale

2.1.1 Configurazione estensimetrica

Per l'applicazione delle due metodologie è necessario, innanzitutto, stabilire una configurazione estensimetrica. La configurazione scelta è ispirata all'uso di sistemi basati su fibre ottiche e, di conseguenza, gli estensimetri sono disposti lungo linee che collegano le due estremità della piastra. Come accennato in precedenza, le misure discrete di deformazione sono state ottenute tramite analisi agli elementi finiti e, di conseguenza, si immagina che gli estensimetri di ciascuna linea siano posizionati in corrispondenza dei nodi della mesh strutturale. Inoltre, per semplicità è stato scelto che ogni estensimetro misuri la componente x di deformazione. La scelta delle linee e del loro numero è stata effettuata per tentativi poiché non era disponibile un Algoritmo Genetico in grado di fornire la configurazione ottimale. Inizialmente il numero di linee è stato posto pari a 3 e, tramite l'approccio inverso, è stato ricostruito il carico per ogni configurazione possibile; tuttavia, la radice dell'errore quadratico medio percentuale, definito come

$$RMSE\%_g = 100 \times \sqrt{\frac{1}{k} \sum_{i=1}^k \left(\frac{g_i - g_i^{ref}}{g_{max}^{ref}} \right)^2} \quad (2.2)$$

minimo ottenuto è stato circa del 9%. Questo errore è stato ritenuto eccessivo e, di conseguenza, il numero di linee è stato aumentato a 4. In questo caso la configurazione estensimetrica che ha fornito il miglior risultato è stata la configurazione rappresentata in figura 2.3

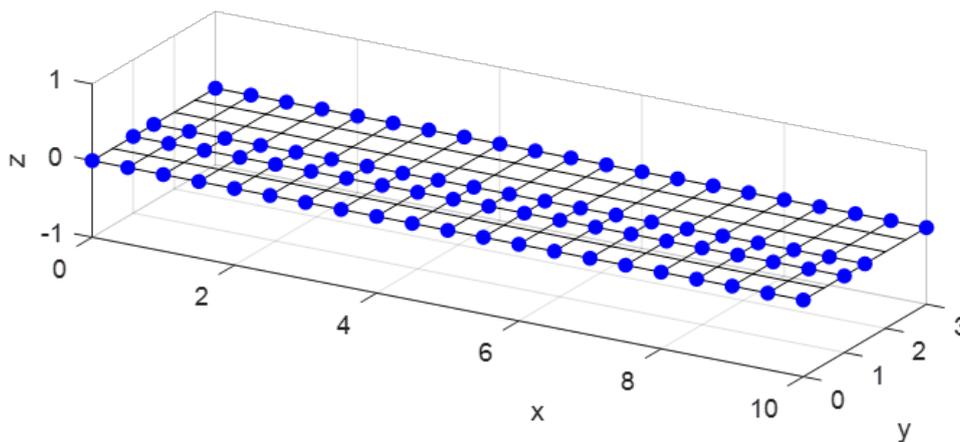


Figura 2.3: Configurazione estensimetrica

e l'errore ottenuto con essa è stato ritenuto accettabile. Pertanto, la configurazione appena rappresentata è stata utilizzata per tutte le analisi di questo studio.

2.2 Approccio inverso

Per applicare questo approccio è necessario, innanzitutto, creare le due mesh definite in precedenza e salvarle in due file con estensione `.bdf`: `'structural_mesh_piastra.bdf'` e `'pressure_mesh_piastra.bdf'`. Inoltre, è necessario effettuare un'analisi agli elementi finiti e creare un file *report* in cui vengono salvate le componenti di deformazione x , y e xy calcolate in ogni nodo della mesh strutturale. Successivamente è possibile calcolare la matrice $[s]$: per fare ciò è stata utilizzata la *function* `pfem_matrix_v2.m` in MATLAB fornita dal docente, la quale è in grado di calcolare la matrice $[s]$ attraverso la procedura iterativa descritta nel Capitolo 1 ricevendo in input i seguenti elementi:

- **bdf_s**: file `.bdf` della mesh strutturale;
- **bdf_p**: file `.bdf` della mesh di pressione;
- **n_str**: vettore degli indici dei nodi della mesh strutturale dove vengono misurate le deformazioni;
- **Z**: vettore della stessa lunghezza di `n_str` contenente i valori 1 o 2 per indicare dove vengono misurate le deformazioni per ogni nodo;
- **nodes_s**: vettore contenente l'indice dei nodi della mesh strutturale che giacciono sulla mesh di pressione;
- **ele_s**: vettore contenente l'indice degli elementi della mesh strutturale che giacciono sulla mesh di pressione

Nel vettore `n_str`, in particolare, vengono inseriti tutti i nodi della mesh strutturale. In questo modo, la matrice $[s]$ viene, inizialmente, calcolata tenendo conto di tutti i nodi della mesh strutturale e di tutte le componenti di deformazione misurate in ciascun nodo. Successivamente, attraverso un'altra *function* denominata `p_fem.m` (sempre fornita dal docente), viene calcolato il vettore delle pressioni nodali $\{p\}$, viene ricostruita la distribuzione di pressione tramite interpolazione dei valori nodali e viene misurato il RMSE % rispetto alla distribuzione di pressione di riferimento. Per far ciò, questa *function* riceve in input i seguenti elementi:

- **bdf_s**: file `.bdf` della mesh strutturale;
- **bdf_p**: file `.bdf` della mesh di pressione;
- **n_str**: vettore degli indici dei nodi della mesh strutturale dove vengono misurate le deformazioni;

2.2. APPROCCIO INVERSO

- **Z**: vettore della stessa lunghezza di '*n_str*' contenente i valori 1 o 2 per indicare dove vengono misurate le deformazioni per ogni nodo;
- **nodes_s**: vettore contenente l'indice dei nodi della mesh strutturale che giacciono sulla mesh di pressione;
- **ele_s**: vettore contenente l'indice degli elementi della mesh strutturale che giacciono sulla mesh di pressione;
- **nodi_eps**: vettore degli indici dei nodi della mesh strutturale in cui sono posizionati gli estensimetri;
- **comp_eps**: vettore della stessa lunghezza di '*nodi_ep*'s contenente gli indici delle componenti di deformazione misurate;
- **str_m**: report contenente le tre componenti di deformazione misurate in ogni nodi della mesh strutturale;
- **s_matrix**: file .mat contenente la matrice [s];
- **func_press**: nome della funzione che definisce la distribuzione di pressione

Inserendo nei vettori '*nodi_eps*' e '*comp_eps*' solamente i nodi e le componenti di deformazione definiti dalla configurazione estensimetrica scelta, la function seleziona le righe della matrice [s] corrispondenti e calcola gli output tramite questa versione ridotta della matrice.

L' **RMSE%** ottenuto con questo approccio è pari a **2.568%** e la distribuzione di pressione ricostruita è rappresentata in figura 2.5

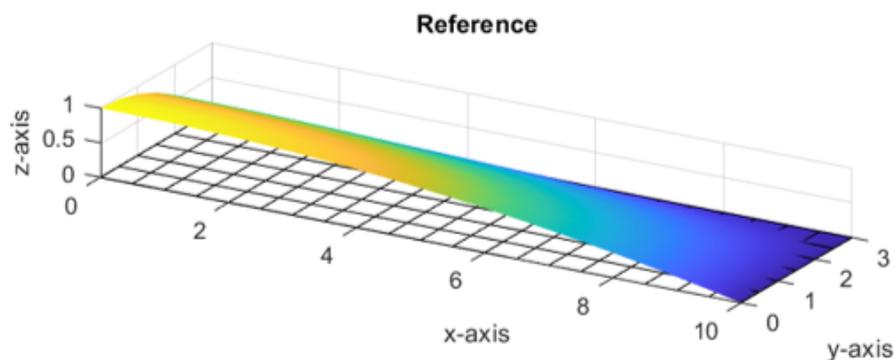


Figura 2.4: Distribuzione di pressione di riferimento

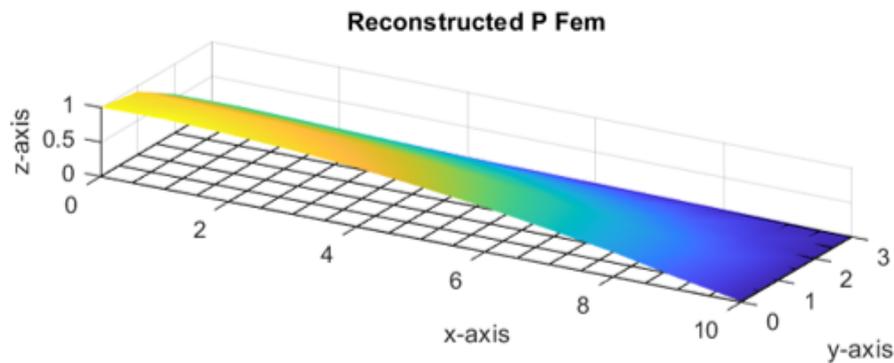


Figura 2.5: Distribuzione di pressione ricostruita tramite approccio inverso

2.3 Rete neurale artificiale

Il secondo approccio utilizzato per la ricostruzione dei carichi consiste nell'utilizzo di una rete neurale artificiale. A tale scopo è stata utilizzata la Neural Net Fitting App presente in MATLAB per la creazione e l'utilizzo della rete neurale.

L'obiettivo, in questo caso, è creare una rete neurale artificiale che, ricevendo in input gli 84 valori di deformazione misurati (la configurazione estensimetrica scelta è formata da 4 linee e su ogni linea sono definiti 21 punti di misura della deformazione), sia in grado di fornire in output i 44 valori nodali di pressione (la mesh di pressione è costituita da 44 nodi) attraverso i quali viene determinata la distribuzione di pressione complessiva tramite interpolazione.

2.3.1 Creazione dei dati di apprendimento

La prima operazione da effettuare per l'utilizzo di una rete neurale artificiale è la creazione dei dati di apprendimento. Dato che per lo studio in esame non sono state condotte misure sperimentali, i dati di apprendimento sono stati creati tramite simulazioni numeriche. In particolare, è stato selezionato un insieme costituito da 62 funzioni di carico bidimensionali di vario tipo con l'obiettivo di addestrare la rete ad una tipologia di carico

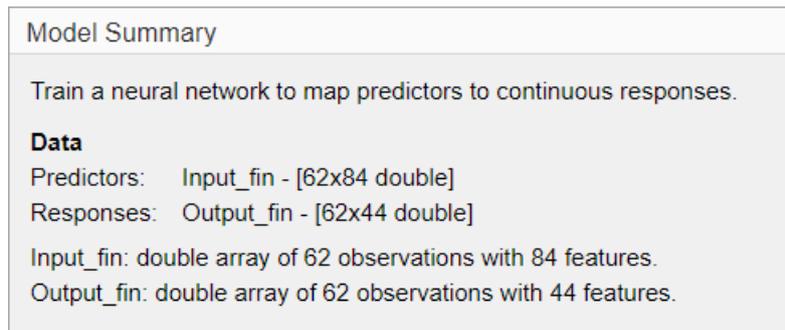


Figura 2.6: Dati di apprendimento

il più vasta possibile; per ogni funzione di carico è stata effettuata un'analisi agli elementi finiti in modo da ottenere gli 84 valori di deformazione e, successivamente, tramite l'approccio inverso descritto in precedenza, sono stati ottenuti i 44 valori nodali di pressione corrispondenti. Questi dati sono stati salvati in due file: **Input_fin.mat** è formato da 62 righe e 84 colonne e contiene le misure di deformazione, **Output_fin.mat** è formato da 62 righe e 44 colonne e contiene le misure di pressione.

2.3.2 Creazione della rete neurale

La rete neurale scelta per questo studio è una rete neurale feed-forward a tre strati (uno strato di input, uno strato nascosto e uno strato di output) con una funzione di attivazione sigmoidea nello strato nascosto e una funzione di attivazione lineare nello strato di output. Il numero di neuroni nello strato di input e nello strato di output è determinato dal tipo di problema, ovvero il numero di neuroni nello strato di input è pari a 84 in quanto la rete neurale riceve in input gli 84 valori di deformazione misurati ed il numero di neuroni nello strato di output è 44 in quanto la rete neurale fornisce in output i 44 valori nodali di pressione. L'unica variabile presente in questa architettura è il numero di neuroni nello strato nascosto. Come si può notare nelle figure 2.7 e 2.8, il numero di neuroni nello strato nascosto è stato posto pari a 10 ed il motivo verrà mostrato nel paragrafo successivo.

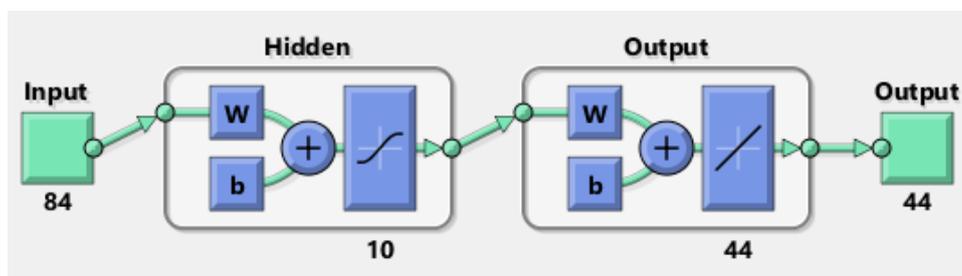


Figura 2.7

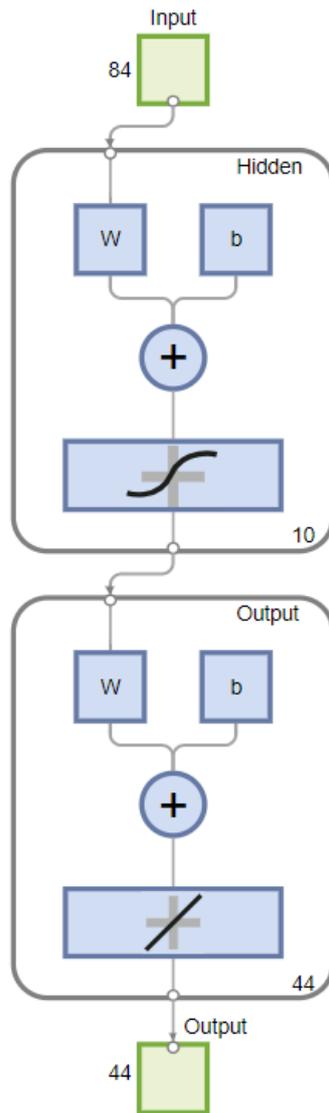


Figura 2.8: Architettura rete neurale

2.3.3 Addestramento della rete neurale

Per effettuare l'addestramento della rete neurale è necessario suddividere i dati di apprendimento in tre sottogruppi: apprendimento, validazione e verifica. In questa applicazione è stata adottata la seguente suddivisione:

- 70% dei dati per l'apprendimento;
- 15% dei dati per la validazione;
- 15% dei dati per la verifica.

Il modello viene inizialmente addestrato utilizzando il **set di dati di apprendimento**, il quale viene utilizzato per la modifica dei parametri del modello, ovvero i pesi b delle connessioni tra i neuroni e gli offsets o biases w . Il modello viene addestrato con questo set di dati utilizzando un algoritmo di apprendimento, che in questo studio coincide con l'algoritmo di Levenberg-Marquardt. In pratica, il set di dati è costituito da coppie di vettori di input e vettori di output; il modello viene lanciato con un dato vettore di input in ingresso e produce un risultato, il quale viene confrontato con il vettore di output corrispondente chiamato anche *target*. Questa operazione viene effettuata per ogni vettore di input e, sulla base del risultato del confronto e dell'algoritmo di apprendimento specifico, i parametri del modello vengono adeguati.

Successivamente, il modello ottenuto viene utilizzato per predire le risposte di un secondo set di dati chiamato **set di dati di validazione**. Questo set di dati fornisce una valutazione imparziale del modello calibrato con il set di dati di apprendimento e regola il/gli iperparametro/i del modello, che in questa applicazione è unico e corrisponde al numero di neuroni nello strato nascosto, con l'obiettivo di evitare il fenomeno dell'*overfitting* e fare in modo che il modello sia in grado di generalizzare.

Infine, il **set di dati di verifica** viene utilizzato per fornire una valutazione imparziale sul modello finale e testare la capacità di generalizzazione della rete, in quanto questi dati non vengono utilizzati per modificare i parametri del modello, ma solo per testare le capacità della configurazione finale.

Per determinare il valore ottimale del numero di neuroni nello strato nascosto, quindi, viene effettuata un'analisi al variare di questo parametro. In particolare, per ogni valore di questo numero vengono calcolati i RMSE% relativi ai set di dati di apprendimento e validazione ed, infine, viene scelto come parametro ottimale quello in corrispondenza del minimo RMSE% relativo al set di validazione. Come si può vedere in figura 2.9, il minimo si ha per $n=10$ e, di conseguenza, il numero di neuroni nello strato nascosto è stato posto pari a 10.

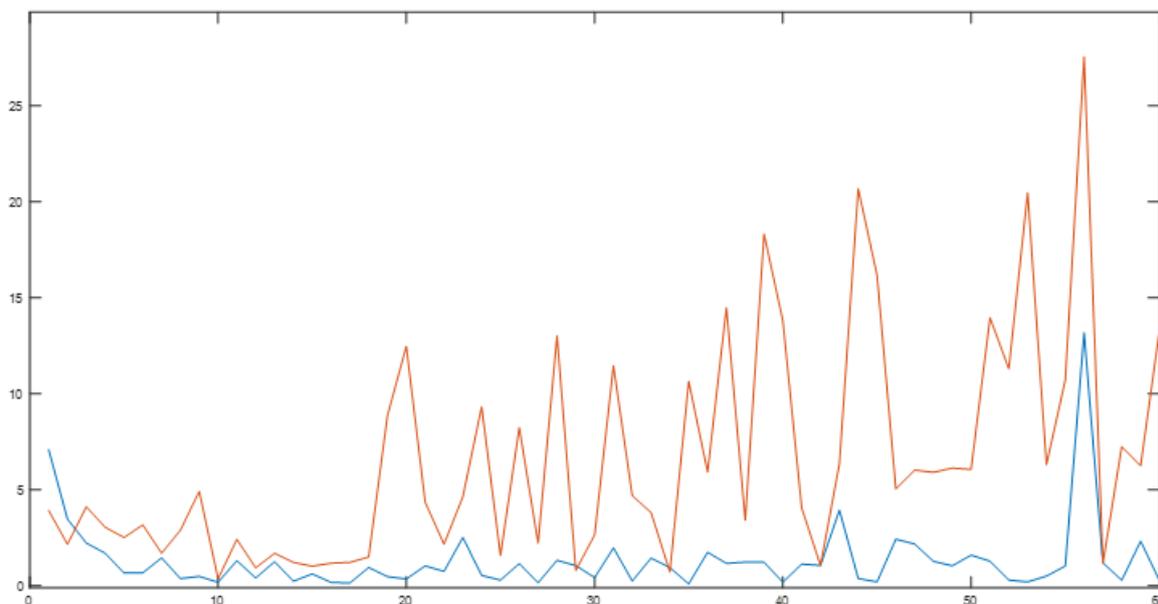


Figura 2.9: Andamento dell'RMSE% relativo al set di dati di validazione (linea arancione) e dell'RMSE% relativo al set di dati di apprendimento (linea blu) al variare del numero di neuroni nello strato nascosto

Dalla figura 2.9, si nota che quando il numero di neuroni è molto piccolo si verifica il fenomeno dell' *underfitting*, ovvero il modello è troppo semplice e non è in grado di definire accuratamente la relazione tra input e output, quindi entrambi i valori dell'RMSE% sono elevati. D'altra parte, se il numero dei neuroni è molto elevato l'RMSE% relativo al set di dati di apprendimento è molto basso, mentre quello relativo al set di dati di validazione aumenta: in questo caso si verifica il fenomeno dell' *overfitting*, ovvero il modello è molto complesso, è in grado di definire una relazione molto accurata tra input e output relativi ai dati di apprendimento ma non è in grado di generalizzare questo risultato per dei non-dati.

Dopo aver suddiviso i dati di apprendimento e aver scelto il valore ottimale del numero di neuroni nello strato nascosto, è possibile addestrare la rete neurale. In questo studio l'algoritmo di apprendimento utilizzato è quello di Levenberg-Marquardt. In figura 2.10 sono mostrati i risultati dell'addestramento.

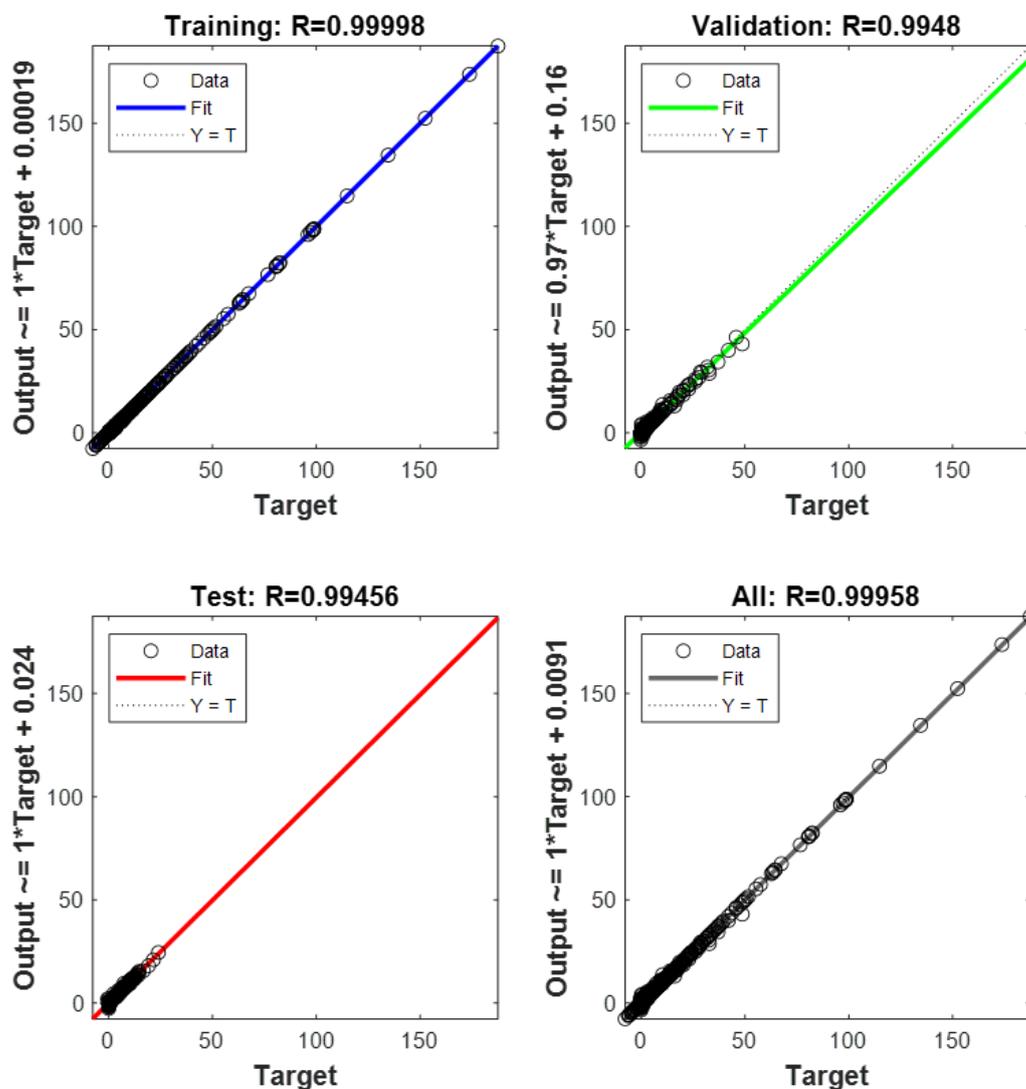


Figura 2.10: Andamenti delle curve di regressione lineare per i set di dati di validazione, apprendimento e verifica

Il grafico della regressione mostra le previsioni della rete, ovvero gli output, rispetto ai valori di target per i set di dati di apprendimento, validazione e verifica. Per un adattamento perfetto, la retta di regressione lineare dovrebbe avere un'inclinazione di 45° , in quanto vorrebbe dire che gli output della rete coincidono con i valori di target. Per questo problema si nota che l'approssimazione è molto buona per tutti i set di dati.

Un altro mezzo per verificare le prestazioni della rete è l'istogramma dell'errore. Le colonne blu rappresentano i dati di apprendimento, le colonne verdi rappresentano i dati di validazione e le colonne rosse rappresentano i dati di verifica. L'istogramma fornisce un'indicazione dei valori anomali, che sono dati in cui l'approssimazione è significativamente peggiore rispetto al resto. In questo problema, tuttavia, la maggior parte degli errori è centrata sullo 0 e ciò conferma ulteriormente la bontà dell'approssimazione fornita dalla rete neurale.

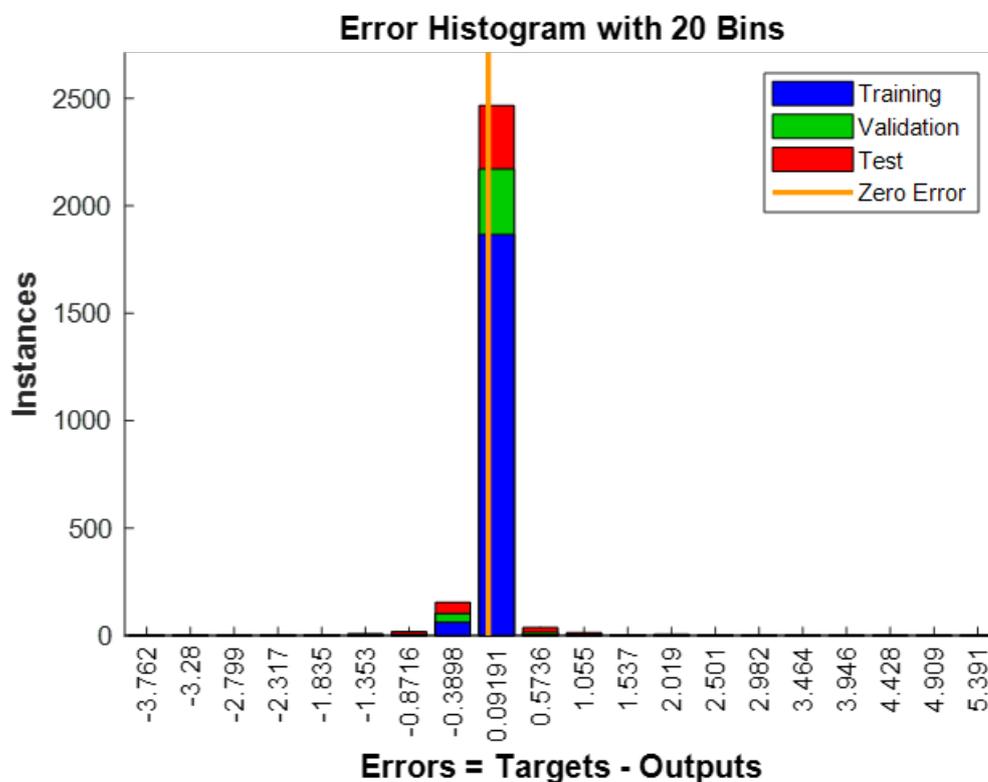


Figura 2.11: Iistogramma dell'errore

2.3.4 Risultati della rete neurale

Dopo aver creato e addestrato la rete neurale, essa è stata utilizzata per la ricostruzione del carico di nostro interesse. Per fare ciò, alla rete neurale viene fornito in input il vettore delle 84 deformazioni ottenute dalla simulazione numerica; la rete, a questo punto, fornisce in output il vettore delle pressioni nodali e, infine, tramite l'interpolazione di questi valori viene ricostruita la distribuzione di pressione completa. In figura 2.12 viene rappresentata

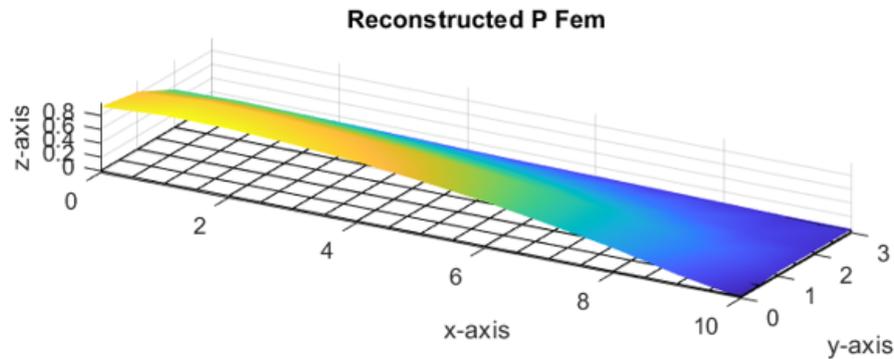


Figura 2.12: Distribuzione di pressione ricostruita tramite la rete neurale artificiale

la distribuzione di pressione ricostruita attraverso questo metodo e l'**RMSE%** ottenuto è pari a **4.245%**.

2.4 Confronto dei risultati

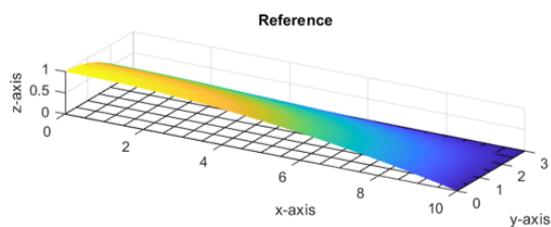
In quest'ultimo paragrafo del capitolo vengono brevemente riassunti i risultati delle analisi svolte nella tabella 2.2.

Definito il modello su cui effettuare le analisi, è stata innanzitutto definita un'adeguata configurazione estensimetrica in grado di fornire il minimo **RMSE%** ottenibile con quel numero di linee di sensori. Il **RMSE%** ottenuto tramite l'approccio inverso considerando tutte le linee di sensori, ovvero utilizzando le deformazioni misurate in tutti i nodi della mesh strutturale, è pari a **2.51%**, mentre quello ottenuto utilizzando la configurazione estensimetrica definita nel paragrafo 2.1.1 è pari a **2.56%**.

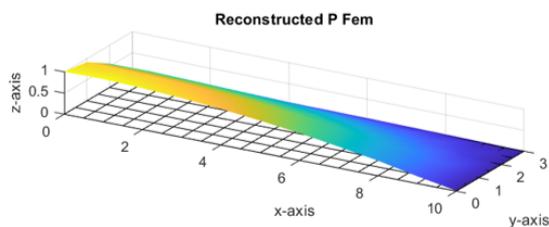
Per quanto riguarda l'approccio delle reti neurali artificiali, invece, il **RMSE%** ottenuto considerando la configurazione estensimetrica a 4 linee è pari a **4.245%**. Nonostante quest'ultimo errore sia quasi il doppio di quello ottenuto con l'approccio inverso, l'approccio delle reti neurali è da ritenersi comunque valido vista la buona ricostruzione ottenuta nonostante lo scarso numero di dati di apprendimento.

| Metodo | RMSE% |
|--|---------|
| Approccio inverso configurazione completa | 2.51 % |
| Approccio inverso configurazione a 4 linee | 2.56 % |
| Rete neurale | 4.245 % |

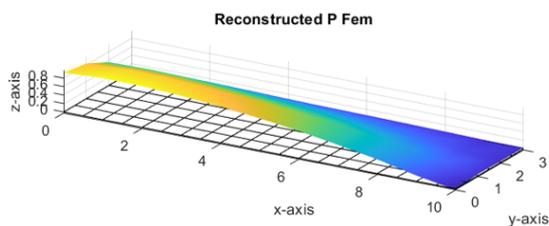
Tabella 2.2: Riassunto dei risultati



(a) Riferimento



(b) Approccio inverso



(c) Rete neurale

Figura 2.13: Distribuzioni di pressione

Capitolo 3

Regolarizzazione dei problemi inversi mal posti

3.1 Definizione del problema

Un problema inverso nasce tipicamente quando si vuole determinare un'informazione relativa a dati “interni” o nascosti del sistema utilizzando misure “esterne”. Per lo studio di questa tipologia di problema, gran parte dei contenuti di questo Capitolo sono tratti da [8]. Nel nostro caso il problema inverso consiste nella determinazione dei carichi agenti su un sistema a partire dalle misure di deformazione indotte dai carichi stessi.

I problemi inversi, tuttavia, appartengono alla classe dei problemi mal posti. Questa definizione è stata introdotta agli inizi del Novecento da Hadamard, secondo cui un problema lineare è ben-posto se soddisfa i seguenti requisiti:

- **Esistenza:** il problema deve avere una soluzione;
- **Unicità:** la soluzione deve essere unica;
- **Stabilità:** la soluzione deve dipendere continuamente dai dati.

Se il problema non rispetta uno di questi requisiti, esso è detto mal-posto.

La violazione del criterio di esistenza può spesso essere risolta tramite una riformulazione del problema, ad esempio ai minimi quadrati.

La condizione di unicità può essere più critica; tuttavia, anche in questo caso essa può spesso essere rispettata tramite una riformulazione del problema, tipicamente ottenuta aggiungendo requisiti addizionali alla soluzione, ad esempio di minima norma.

La condizione di stabilità, invece, è molto più difficile da rispettare e la sua violazione implica che piccole perturbazioni nei dati producono grandi perturbazioni nella soluzione. Il motivo di questo comportamento è il mal condizionamento della matrice dei coefficienti A . Anche in questo caso è necessario riformulare il problema in modo che la nuova soluzione sia meno sensibile alle perturbazioni, ovvero è necessario *stabilizzare* o *regolarizzare* il problema.

I problemi inversi utilizzano un modello per ricostruire un input a partire dal corrispondente output. Questi modelli matematici possono essere espressi come un sistema di equazioni lineari:

$$Ax = b \tag{3.1}$$

In questo sistema, la matrice $A \in \mathbb{R}^{m \times n}$ con $m \geq n$ descrive il modello. L'input è rappresentato dal vettore $x \in \mathbb{R}^n$ e l'output dal vettore $b \in \mathbb{R}^m$. Il problema inverso associato a questo modello calcola l'input x dato l'output b e noto il modello.

Nel nostro caso, la matrice A è solitamente rettangolare e, di conseguenza, non ammette un'inversa e il sistema è sovradeterminato, ovvero il numero di equazioni è maggiore del numero di incognite, che in termini matriciali significa un numero di righe superiore al numero di colonne $m > n$. In questo caso la soluzione potrebbe non esistere o non essere unica (esistono soluzioni se e solo se $\text{rank}(A) = \text{rank}([A \ b])$ e la soluzione è unica solo se $\text{rank}(A) = n$, altrimenti sono infinite) e, di conseguenza, il problema iniziale viene riformulato e si considera il problema di minimizzazione ai minimi quadrati associato al sistema di equazioni lineari:

$$\min_x \|Ax - b\| \tag{3.2}$$

3.2 Decomposizione ai Valori Singolari (SVD)

La decomposizione ai valori singolari, detta anche SVD, è una particolare fattorizzazione che permette di scrivere una generica matrice $A \in \mathbb{R}^{m \times n}$ come prodotto di tre matrici. Si tratta di un'interessante tecnica che permette di evidenziare alcune caratteristiche utili della matrice ed, inoltre, permette di determinare la soluzione del problema ai minimi quadrati.

Data una matrice $A \in \mathbb{R}^{m \times n}$, essa può essere scritta come

$$A = U \Sigma V^T \tag{3.3}$$

dove $U = [u_1, \dots, u_m] \in \mathbb{R}^{m \times m}$ è una matrice ortogonale le cui colonne sono i vettori singolari sinistri di A , $V = [v_1, \dots, v_n] \in \mathbb{R}^{n \times n}$ è una matrice ortogonale le cui colonne sono i vettori singolari destri di A e $\Sigma \in \mathbb{R}^{m \times n}$ è una matrice diagonale rettangolare. I termini sulla diagonale σ_j sono detti valori singolari della matrice A . Essi sono $\sigma_j \geq 0$ e sono ordinati in modo decrescente $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$. Inoltre, il rango della matrice A è pari al numero di valori singolari non nulli, ovvero

$$l = \text{rank}(A), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_l > \sigma_{l+1} = \dots = \sigma_n = 0$$

Tramite la decomposizione è anche possibile definire il numero di condizionamento di A :

$$\|A\|_2 = \sigma_1, \|A^{-1}\|_2 = \sigma_n^{-1} \quad \longrightarrow \quad \text{cond}(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$$

3.2. DECOMPOSIZIONE AI VALORI SINGOLARI (SVD)

ovvero il numero di condizionamento di A è dato dal rapporto tra il valore singolare massimo e il valore singolare minimo non nullo e ciò è valido anche se A è rettangolare ed è a rango pieno, nonostante l'espressione sia stata calcolata per una matrice quadrata. Questo parametro misura la sensibilità della soluzione alle perturbazioni nei dati: un numero di condizionamento grande implica che piccole perturbazioni nei dati provocano grandi perturbazioni nella soluzione x calcolata.

Nei problemi discreti mal posti, i valori singolari della matrice hanno un andamento del tipo:

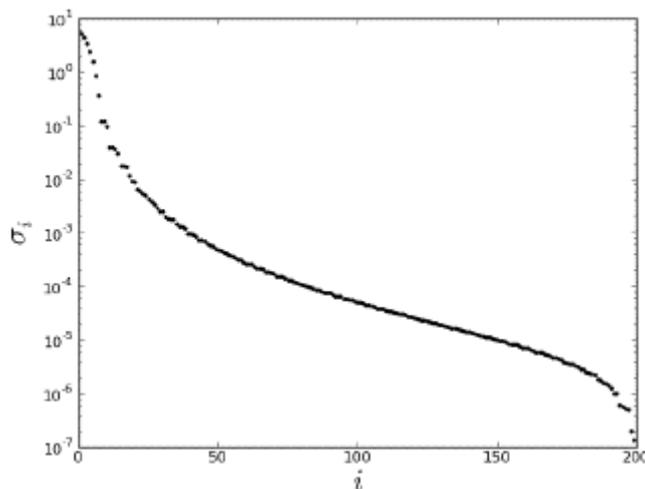


Figura 3.1: Andamento valori singolari [9]

ovvero decadono gradualmente a zero. Di conseguenza, il rapporto tra il massimo e il minimo non nullo valore singolare è elevato e ciò dimostra l'elevata sensibilità alle perturbazioni del problema discreto mal posto.

I vettori singolari, invece, hanno un andamento oscillatorio ed il numero di oscillazioni, ovvero il numero di cambiamenti di segno all'interno dei vettori, aumenta all'aumentare dell'indice i :

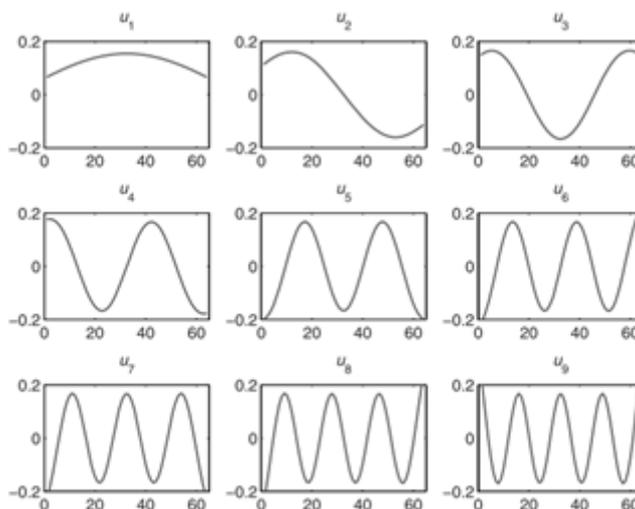


Figura 3.2: Andamento vettori singolari [9]

Ciò significa che, nel caso di problemi inversi mal posti, le alte frequenze vengono amplificate in quanto sono associate ad indici i elevati e, quindi, a piccoli valori singolari (come si vedrà più avanti, nella definizione di soluzione i valori singolari sono a denominatore). Si ha inoltre

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \dots + \sigma_l u_l v_l^T = \sum_{i=1}^l \sigma_i u_i v_i^T \quad (3.4)$$

cioè è possibile scrivere ogni matrice di rango l come somma di l matrici di rango 1. Se si definiscono $U_l \in \mathfrak{R}^{m \times l}$, $V_l \in \mathfrak{R}^{n \times l}$ e $\Sigma_l \in \mathfrak{R}^{l \times l}$ come

$$U_l = [u_1, u_2, \dots, u_l], \quad V_l = [v_1, v_2, \dots, v_l], \quad \Sigma_l = \text{diag}(\sigma_1, \dots, \sigma_l)$$

si ottiene la SVD *compatta* di A :

$$A = U_l \Sigma_l V_l^T \quad (3.5)$$

Una delle principali applicazioni della SVD è la risoluzione di problemi ai minimi quadrati e, più in generale, di sistemi lineari a matrice dei coefficienti rettangolare, compatibili o meno.

Se $A \in \mathfrak{R}^{m \times n}$ è una matrice di rango l , la SVD può essere definita come 3.5 e se $b \in \mathfrak{R}^m$ è il vettore dei termini noti, allora la soluzione di minima norma del problema 3.2 può essere scritta nella forma

$$x = A^+ b = \sum_{i=1}^l \frac{u_i^T b}{\sigma_i} v_i \quad (3.6)$$

dove $A^+ \in \mathfrak{R}^{n \times m}$ è la pseudo-inversa di Moore-Penrose, definita come

$$A^+ = V_l \Sigma_l^{-1} U_l^T \quad (3.7)$$

3.3 Diagramma di Picard

Definita la soluzione del problema generale ai minimi quadrati, è necessario comprendere la bontà di essa. Per far ciò, viene investigato il comportamento dei coefficienti SVD $u_i^T b$ (del termine di destra b) e $u_i^T b / \sigma_i$ (della soluzione). Il diagramma di questi coefficienti,

3.3. DIAGRAMMA DI PICARD

assieme ai valori singolari, è chiamato *diagramma di Picard* e viene presentato nel dettaglio in [8].

Assenza di rumore nei dati:

Un diagramma tipico per un problema discreto mal posto in assenza di rumore nei dati è il seguente:

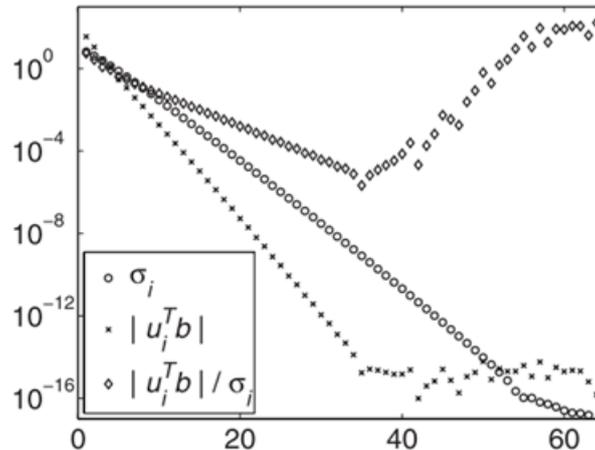


Figura 3.3: Diagramma di Picard in assenza di rumore [8]

Come si può notare, i valori singolari σ_i decadono gradualmente a 0. $|u_i^T b|$ rappresentano i coefficienti del termine di destra b in termini di SVD e decadono più velocemente dei valori singolari σ_i a causa della condizione di Picard. Ciò si verifica per $i < 35$: per $i \geq 35$, infatti, questi valori assumono un andamento all'incirca costante e ciò è dovuto agli errori inerenti (*rounding errors*) del calcolatore, ovvero errori legati alla precisione di macchina. Anche i coefficienti della soluzione $\frac{u_i^T b}{\sigma_i}$ decadono per $i < 35$, ma per $i \geq 35$ essi iniziano ad aumentare a causa della scarsa accuratezza dei valori di $u_i^T b$. Ciò mostra come anche in assenza di errori nei dati non sia possibile aspettarsi una buona soluzione del problema inverso discreto, in quanto l'influenza dei *rounding errors* nel calcolo della SVD è sufficiente a distruggere la soluzione calcolata.

Presenza di rumore nei dati:

Se si introduce un errore nel vettore di destra, quest'ultimo può essere scritto come

$$b = \bar{b} + e \quad (3.8)$$

dove \bar{b} rappresenta il vettore dei dati senza errore ed e il vettore del rumore. In questo caso, il diagramma di Picard ha la seguente forma:

I valori singolari σ_i rimangono invariati in quanto dipendono esclusivamente dalla matrice A e non dal vettore b .

I coefficienti SVD del termine di destra b , ovvero $u_i^T b$, sono dati dalla somma

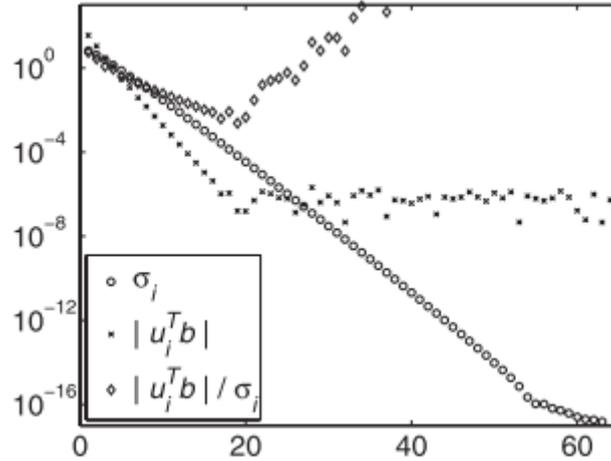


Figura 3.4: Diagramma di Picard in presenza di rumore [8]

$$u_i^T b = u_i^T (\bar{b} + e) = u_i^T \bar{b} + u_i^T e \quad (3.9)$$

Il primo termine rappresenta i coefficienti SVD del termine b senza errore e decade fino alla comparsa dei *rounding errors*. Il secondo termine, invece, rappresenta i termini SVD dell'errore. Per piccoli valori di indice prevale il primo termine e l'influenza dell'errore è nulla, per indici grandi la situazione si inverte. Complessivamente, quindi, si verifica un iniziale decadimento dei coefficienti SVD $u_i^T b$, fino a quando si appiattiscono ad un livello determinato dal livello di rumore.

I coefficienti SVD della soluzione $u_i^T b / \sigma_i$ decadono inizialmente, ma per grandi valori dell'indice i essi iniziano ad aumentare. I primi valori, quindi, non vengono influenzati dal rumore e costituiscono una buona soluzione del problema, i secondi, invece, sono dominati dal rumore. Inoltre, questi ultimi assumono un valore molto grande per alti valori dell'indice i e, pertanto, le soluzioni calcolate sono completamente dominate dai coefficienti SVD corrispondenti ai valori singolari più piccoli.

3.4 Curva L

Un altro strumento che nei capitoli successivi risulterà essere fondamentale ai fini della regolarizzazione della soluzione è la cosiddetta *curva L* e viene presentata nel dettaglio in [8]. Le norme della soluzione e del residuo assumono un ruolo fondamentale nella trattazione dei problemi discreti mal posti in quanto essi possono essere sempre calcolati indipendentemente dal metodo di regolarizzazione utilizzato.

La norma della soluzione TSVD e il residuo associato variano monotonicamente con k :

$$\|x_k\|_2^2 = \sum_{i=1}^k \left(\frac{u_i^T b}{\sigma_i} \right)^2 \leq \|x_{k+1}\|_2^2 \quad (3.10)$$

$$\|Ax_k - b\|_2^2 = \sum_{i=k+1}^n (u_i^T b)^2 + \varepsilon_\perp^2 \geq \|Ax_{k+1} - b\|_2^2 \quad (3.11)$$

dove $\varepsilon_\perp = \|(I - UU^T)b\|_2$ è la norma della componente di b che giace fuori dallo spazio delle colonne di A .

La norma della soluzione di Tikhonov e il suo residuo sono:

$$\|x_\lambda\|_2^2 = \sum_{i=1}^n \left(\varphi_i^{[\lambda]} \frac{u_i^T b}{\sigma_i} \right)^2 \quad (3.12)$$

$$\|Ax_\lambda - b\|_2^2 = \sum_{i=1}^n \left((1 - \varphi_i^{[\lambda]}) u_i^T b \right)^2 + \varepsilon_\perp^2 \quad (3.13)$$

dove $\varphi_i^{[\lambda]} = \sigma_i^2 / (\sigma_i^2 + \lambda^2)$ sono i fattori di filtro. Per vedere che queste norme variano monotonicamente con λ , viene introdotta la seguente notazione:

$$\xi = \|x_\lambda\|_2^2 \quad \text{e} \quad \rho = \|Ax_\lambda - b\|_2^2 \quad (3.14)$$

Si dimostra che

$$\xi' \equiv \frac{d\xi}{d\lambda} = -\frac{4}{\lambda} \sum_{i=1}^n (1 - \varphi_i^{[\lambda]}) (\varphi_i^{[\lambda]})^2 \frac{(u_i^T b)^2}{\sigma_i^2} \quad (3.15)$$

$$\rho' \equiv \frac{d\rho}{d\lambda} = \frac{4}{\lambda} \sum_{i=1}^n (1 - \varphi_i^{[\lambda]})^2 \varphi_i^{[\lambda]} (u_i^T b)^2 = -\lambda^2 \xi' \quad (3.16)$$

Si vede che $\xi' < 0$ e $\rho' > 0$ per tutti i λ , confermando la monotonicità delle norme rispetto a λ . Inoltre, dalla relazione $\rho' = -\lambda^2 \xi'$ si ottiene $d\xi/d\rho = -\lambda^{-2}$, mostrando che la norma della soluzione al quadrato $\|x_\lambda\|_2^2$ è una funzione monotonicamente decrescente della norma del residuo al quadrato $\|Ax_\lambda - b\|_2^2$.

Un'altra relazione utile è la seguente:

$$\rho'' \equiv \frac{d^2\rho}{d\lambda^2} = \frac{d}{d\lambda} (-\lambda^2 \xi') = -2\lambda \xi' - \lambda^2 \xi'' \quad (3.17)$$

Se si considera la curva ottenuta diagrammando ξ in funzione di ρ con parametro λ , la curvatura c_λ della curva in funzione di λ è

$$c_\lambda = \frac{\rho' \xi'' - \rho'' \xi'}{((\rho')^2 + (\xi')^2)^{3/2}} = \frac{2\lambda (\xi')^2}{((\rho')^2 + (\xi')^2)^{3/2}} \quad (3.18)$$

Quindi, $c_\lambda > 0$ per tutti i λ e la curva (ρ, ξ) è convessa.

Questa curva assume grande importanza perché mostra come la soluzione regolarizzata varia al variare del parametro di regolarizzazione λ . Inoltre, valgono le seguenti disuguaglianze:

$$0 \leq \|x_\lambda\|_2 \leq \|A^{-1}b\|_2 \quad \text{e} \quad 0 \leq \|Ax_\lambda - b\|_2 \leq \|b\|_2 \quad (3.19)$$

Poiché ogni punto (ρ, ξ) sulla curva è una soluzione del problema

$$\rho = \min \|Ax - b\|_2^2 \quad \text{soggetto a} \quad \|x\|_2^2 \leq \xi \quad (3.20)$$

allora la curva definisce un limite tra due regioni del primo quadrante: è impossibile scegliere un vettore x tale per cui il punto $(\|Ax - b\|_2^2, \|x\|_2^2)$ giaccia sotto la curva; solo i punti su o sopra la curva possono essere raggiunti. Lo stesso vale anche per la curva $(\|Ax - b\|_2, \|x\|_2)$. Tuttavia, quando ognuna di queste curve viene diagrammata in scala lineare, è difficile comprendere le sue caratteristiche a causa dei grandi intervalli di valori delle due norme.

Le caratteristiche diventano più visibili quando la curva viene disegnata in scala logaritmica doppia. La curva associata

$$\left(\frac{1}{2} \log \rho, \frac{1}{2} \log \xi \right) = (\log \|Ax_\lambda - b\|_2, \log \|x_\lambda\|_2) \quad (3.21)$$

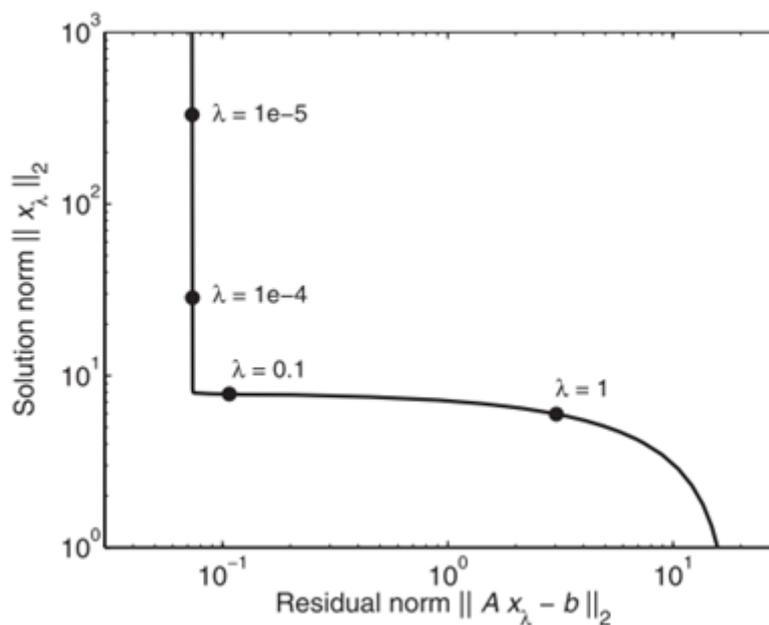


Figura 3.5: Curva L [8]

è chiamata *curva-L* per la regolarizzazione di Tikhonov e ha un ruolo chiave nell'analisi dei problemi discreti mal posti.

La soluzione di Tikhonov si comporta in maniera diversa per piccoli e grandi valori di λ . Lo stesso vale per le diverse parti della curva-L. Quando λ è grande, allora x_λ è dominata

3.4. CURVA L

dai coefficienti SVD che hanno come contributo principale il termine di destra esatto b^{exact} : la soluzione è sovra-regolarizzata. Per grandi valori di λ si ha che

$$\|x_\lambda\|_2 \approx \|x^{exact}\|_2 \quad (\text{costante}) \quad e \quad \|Ax_\lambda - b\|_2 \quad \text{aumenta con } \lambda$$

Per piccoli valori di λ la soluzione di Tikhonov è dominata dagli errori di perturbazione provenienti dal rumore invertito: la soluzione è sotto-regolarizzata, e si ha che

$$\|x_\lambda\|_2 \quad \text{aumenta con } \lambda^{-1} \quad e \quad \|Ax_\lambda - b\|_2 \approx \|e\|_2 \quad (\text{costante})$$

La conclusione è che la curva-L ha due parti distinte, ovvero una parte che è approssimativamente orizzontale e una approssimativamente verticale.

L'”angolo” che separa queste due parti si trova circa nel punto

$$(\log \|e\|_2, \log \|x^{exact}\|_2) \tag{3.22}$$

Andando verso destra, per $\lambda \rightarrow \infty$, la curva-L inizia a decrescere in quanto la quantità crescente di regolarizzazione forza la norma della soluzione verso zero.

Capitolo 4

Metodologie per la regolarizzazione della soluzione

Dall'analisi svolta, è possibile comprendere le difficoltà associate ai problemi discreti mal posti e per quale motivo la soluzione “banale” non è utile. Per questo motivo è necessario definire nuovi metodi per determinare soluzioni approssimate che sono meno sensibili alle perturbazioni rispetto alla soluzione “banale”.

Questi metodi sono chiamati *metodi di regolarizzazione* perché impongono regolarità alla soluzione calcolata (tipicamente sotto forma di requisito di regolarità della soluzione) e, imponendola, vengono eliminate alcune componenti di rumore indesiderate, ottenendo soluzioni approssimate più stabili. Tutti i contenuti di questo Capitolo sono tratti da [8–13].

Come già detto, i problemi discreti mal posti sono caratterizzati da una matrice dei coefficienti con un elevato numero di condizionamento. Ciò implica che la soluzione banale è molto sensibile ad ogni perturbazione del termine di destra, la quale rappresenta gli errori nei dati. In particolare, poiché $\text{cond}(A)$ è grande, la soluzione calcolata x può essere molto lontana da quella esatta x^{exact} . Per questo motivo è necessario introdurre alcuni metodi di regolarizzazione in grado di calcolare approssimazioni di x^{exact} meno sensibili.

La figura 4.1 mostra il bisogno della regolarizzazione. Qui, la soluzione esatta x^{exact} origina l'esatto termine di destra b^{exact} , mentre il termine di destra perturbato è $b = b^{exact} + e$. A causa del mal condizionamento di A , la soluzione banale $x = A^{-1}b$ può essere distante dalla soluzione esatta x^{exact} , anche se la perturbazione è piccola, ovvero $\|e\|_2 < \|b^{exact}\|_2$. Le soluzioni regolarizzate, come ad esempio la soluzione TSVD x_k o la soluzione di Tikhonov x_λ , rappresentano una buona approssimazione di x^{exact} .

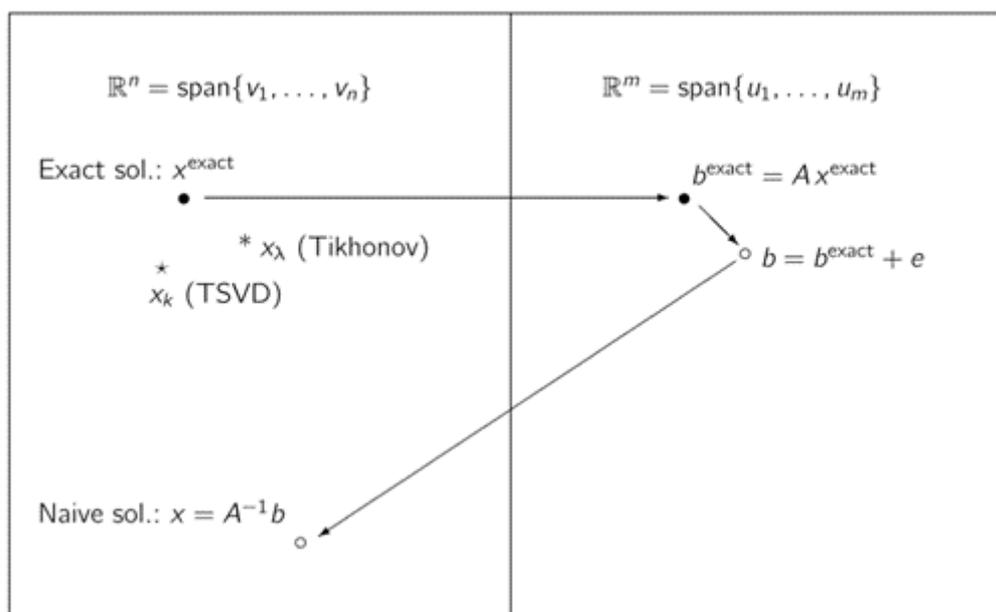


Figura 4.1: Il bisogno della regolarizzazione [8]

4.1 Metodi filtranti

Una categoria di metodi di regolarizzazione produce soluzioni che possono essere espresse come espansioni SVD filtrate:

$$x_{\text{reg}} = \sum_{i=1}^n \varphi_i \frac{u_i^T b}{\sigma_i} v_i \quad (4.1)$$

dove φ_i sono i *fattori di filtro* associati al metodo.

Nel generico problema considerato, la matrice A è o quadrata o rettangolare con più righe che colonne, ovvero $A \in \mathbb{R}^{m \times n}$ con $m \geq n$ e per $m > n$ è normale considerare il problema ai minimi quadrati 3.2 associato al problema di partenza 3.1. Quindi, quando si parla di “soluzione banale” ci si riferisce o alla soluzione $x = A^{-1}b$ (quando $m = n$) o alla soluzione ai minimi quadrati (quando $m > n$), e questa soluzione ha la stessa espansione SVD per entrambi i casi. Anche l’espressione del numero di condizionamento A è lo stesso, ovvero $\text{cond}(A) = \frac{\sigma_1}{\sigma_n}$. Pertanto, nel seguito non è necessario distinguere i due casi.

4.1.1 TSVD

Dall'analisi svolta precedentemente, è chiaro che gli errori estremamente grandi nella soluzione banale provengono dalle componenti SVD rumorose associate ai valori singolari più piccoli. In particolare, la soluzione banale è dominata dai coefficienti SVD della forma $u_i^T b / \sigma_i \approx u_i^T e / \sigma_i$ (dove e rappresenta la perturbazione di b) corrispondenti ai valori singolari più piccoli. La buona notizia è che i coefficienti SVD della forma $u_i^T b / \sigma_i \approx u_i^T b^{exact} / \sigma_i$, corrispondenti ai valori singolari più grandi, sono affidabili.

Queste considerazioni conducono ad un metodo per il calcolo di soluzioni approssimate regolarizzate che consiste nell'eliminazione delle componenti SVD dominate dal rumore. Quindi, viene definita soluzione *SVD troncata* (TSVD) x_k la soluzione ottenuta prendendo le prime k componenti della soluzione banale:

$$x_k \equiv \sum_{i=1}^k \frac{u_i^T b}{\sigma_i} v_i \quad (4.2)$$

Il *parametro di troncamento* k deve essere scelto in modo da eliminare tutti i coefficienti SVD dominati dal rumore. Un valore adeguato di k può essere trovato tramite il diagramma di Picard.

Una formulazione alternativa del metodo TSVD è basata sull'introduzione della matrice A_k , che è la matrice di rango k definita come:

$$A_k = \sum_{i=1}^k u_i \sigma_i v_i^T \quad (4.3)$$

Il numero di condizionamento di 4.3 è $cond(A_k) = \frac{\sigma_1}{\sigma_k}$, ed è solitamente molto più piccolo del numero di condizionamento $cond(A) = \frac{\sigma_1}{\sigma_n}$ della matrice di partenza A .

Pertanto, sostituire il problema originale e mal condizionato 3.1 o 3.2 con il problema ai minimi quadrati meglio condizionato

$$\min_x \|A_k x - b\|_2 \quad (4.4)$$

rappresenta una buona idea.

Tuttavia, poiché A_k non è a rango pieno (se $k < n$), non esiste un'unica soluzione a questo problema ai minimi quadrati. Per definire un'unica soluzione, è necessario aggiungere un vincolo addizionale alla soluzione x . Un tipico vincolo in molte applicazioni è quello di minima norma-2:

$$\min \|x\|_2 \quad \text{applicato a} \quad \min \|A_k x - b\|_2 \quad (4.5)$$

È facile dimostrare che la soluzione di questo problema vincolato è la soluzione TSVD x_k . Pertanto, questa formulazione è una definizione alternativa della soluzione TSVD, in cui il requisito di regolarità su x , nella forma di minimizzazione della sua norma $\|x\|_2$, è

esplicito.

Analogamente al problema 3.2 di partenza, la soluzione TSVD del problema 4.5 può essere scritta nella forma

$$x_k = A_k^+ b \quad (4.6)$$

dove $A_k^+ = \sum_{i=1}^k v_i \sigma_i^{-1} u_i^T$ è la matrice pseudo-inversa di A_k .

4.1.2 Regolarizzazione di Tikhonov in forma standard

Il metodo TSVD è intuitivo ed è facile calcolare le soluzioni x_k per diversi parametri di troncamento dopo aver calcolato la SVD. Lo svantaggio è che esso richiede esplicitamente il calcolo della SVD o, almeno, i primi k vettori e valori singolari. Questa operazione potrebbe risultare eccessiva per problemi di larga scala e, di conseguenza, sono necessari altri metodi di regolarizzazione più adatti a questi problemi. Uno di essi è il metodo di *regolarizzazione di Tikhonov*. Questo metodo incorpora esplicitamente il requisito di regolarità nella formulazione del problema. La soluzione di Tikhonov x_λ è definita come la soluzione del problema

$$\min_x \{ \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2 \} \quad (4.7)$$

Qui, il *parametro di regolarizzazione* λ è un parametro positivo che controlla il peso tra due ingredienti della funzione.

- Il primo termine $\|Ax - b\|_2^2$ misura la bontà dell'approssimazione, cioè quanto bene la soluzione x predice il dato noto (rumoroso) b . Ovviamente, se questo termine è troppo grande x non può essere considerata una buona soluzione perché non “risolve il problema”.
- Il secondo termine $\|x\|_2^2$ misura la regolarità della soluzione. L'introduzione di questo termine si basa sul fatto che la soluzione banale è dominata dalle componenti ad alta frequenza dotate di grandi ampiezze, e l'obiettivo è quello di controllare la norma di x in modo da eliminare le grandi componenti di rumore.
- L'equilibrio tra i due termini è controllato dal fattore λ^2 . Maggiore è λ , maggiore è il peso dato alla minimizzazione della norma della soluzione $\|x\|_2$ e alla sua regolarità. D'altra parte, minore è λ , maggiore è il peso dato all'approssimazione dei dati rumorosi, generando una soluzione meno regolare (si ottiene il problema originale e la soluzione banale per $\lambda = 0$).

L'obiettivo è trovare un buon equilibrio tra i due termini attraverso un adeguato valore di λ , in modo che la soluzione regolarizzata x_λ sia sufficientemente regolare e contemporaneamente approssimi bene i dati. L'obiettivo è ottenere una soluzione regolarizzata che

approssimi la soluzione esatta.

Il problema di Tikhonov può essere riformulato come

$$\min_x \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2 \quad (4.8)$$

che è chiaramente un problema ai minimi quadrati con soluzione

$$x_\lambda = (A^T A + \lambda^2 I)^{-1} A^T b \quad (4.9)$$

Inserendo la SVD di A in 4.9, si ottiene l'espressione della soluzione in termini di vettori e valori singolari:

$$x_\lambda = \sum_{i=1}^n \varphi_i^{[\lambda]} \frac{u_i^T b}{\sigma_i} v_i \quad (4.10)$$

dove sono stati introdotti i *fattori di filtro* $\varphi_i^{[\lambda]}$ per $i = 1, \dots, n$, che soddisfano

$$\varphi_i^{[\lambda]} = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \approx \begin{cases} 1 & \sigma_i \gg \lambda \\ \frac{\sigma_i^2}{\lambda^2} & \sigma_i \ll \lambda \end{cases} \quad (4.11)$$

Il comportamento di questi fattori è illustrato nella figura successiva:

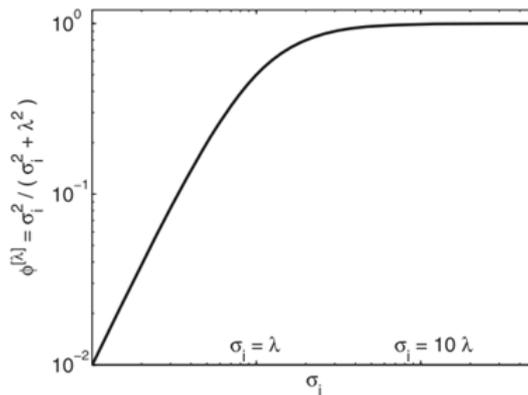


Figura 4.2: Fattori di filtro della regolarizzazione di Tikhonov [8]

Per valori singolari σ_i più grandi del parametro λ , i fattori di filtro sono vicini a 1 e le corrispondenti componenti SVD contribuiscono alla soluzione x_λ nella loro quasi totalità. D'altra parte, per valori singolari molto più piccoli di λ i fattori di filtro sono piccoli e, pertanto, queste componenti SVD sono smorzate o filtrate. In quest'ultimo caso, i fattori di filtro $\varphi_i^{[\lambda]}$ sono proporzionali a σ_i^2 e, quindi, essi decadono velocemente per eliminare i fattori crescenti $u_i^T b / \sigma_i \approx u_i^T e / \sigma_i$.

La conclusione di quest'analisi è che la soluzione di Tikhonov è una soluzione filtrata.

La transizione nei fattori di filtro $\varphi_i^{[\lambda]}$ avviene per valori singolari la cui dimensione è comparabile a λ e in questo modo è possibile usare λ per controllare il filtraggio (in modo simile all'uso del parametro di troncamento k nella TSVD). Tuttavia, qui la transizione tra le componenti SVD mantenute e filtrate è più regolare e il filtraggio può essere ottenuto senza il calcolo esplicito della SVD.; tutto ciò che è necessario è risolvere il problema ai minimi quadrati.

Il numero di condizionamento per la soluzione di Tikhonov è $k_\lambda = \frac{\sigma_1}{\lambda}$ che, in modo simile al numero di condizionamento per la TSVD, può essere molto più piccolo del numero di condizionamento di A .

4.1.3 Regolarizzazione di Tikhonov in forma generale

La regolarizzazione di Tikhonov in forma standard utilizza la norma-2 come mezzo di controllo (o soppressione) dell'errore nella soluzione regolarizzata. La norma $\|x\|_2$ è utile in molte applicazioni, ma non sempre rappresenta la scelta migliore.

E' possibile, quindi, formulare una versione generale del problema di Tikhonov sostituendo la norma $\|x\|_2$ con la discretizzazione di una *norma regolarizzante*. La forma generale della norma regolarizzante discreta è $\|Lx\|_2$, dove L è chiamata *matrice di regolarizzazione* e rappresenta o un'approssimazione discreta di un operatore di derivata o una funzione del parametro di regolarizzazione λ .

Il problema di Tikhonov in *forma generale*, quindi, ha la seguente forma:

$$\min_x \{ \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2 \} \quad (4.12)$$

La matrice L è $p \times n$ senza restrizioni sulla dimensione p .

Se L è invertibile, ovvero L^{-1} esiste, la soluzione a 4.12 può essere scritta come $x_{L,\lambda} = L^{-1}\bar{x}_\lambda$, dove \bar{x}_λ rappresenta la soluzione del problema di Tikhonov in forma standard

$$\min_{\bar{x}} \{ \|(AL^{-1})\bar{x} - b\|_2^2 + \lambda^2 \|\bar{x}\|_2^2 \} \quad (4.13)$$

La moltiplicazione per L^{-1} in $x_\lambda = L^{-1}\bar{x}_\lambda$ rappresenta un' integrazione, la quale aggiunge un'ulteriore regolarità alla soluzione di Tikhonov rispetto alla scelta $L=I$. Lo stesso discorso vale anche per matrici regolarizzanti L di forma rettangolare e non invertibili.

Come il problema in forma standard ottenuto per $L=I$, la soluzione di Tikhonov in forma generale $x_{L,\lambda}$ viene calcolata riformulando 4.12 in un problema ai minimi quadrati della forma

$$\min_x \left\| \begin{pmatrix} A \\ \lambda L \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2 \quad (4.14)$$

La soluzione di Tikhonov in forma generale $x_{L,\lambda}$ è unica quando la matrice dei coefficienti in 4.14 ha rango pieno, ovvero quando l'intersezione tra gli spazi nulli di A e L è nulla:

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \emptyset \quad (4.15)$$

Come anticipato in precedenza, in questo studio la matrice di regolarizzazione L rappresenta o un'approssimazione discreta di un operatore di derivata o una funzione del parametro di regolarizzazione λ . Nel primo caso si otterranno i metodi di Tikhonov del primo e secondo ordine, nel secondo caso si otterranno i metodi NTR, MTR-1 e MTR-2.

Regolarizzazione di Tikhonov del primo ordine

La regolarizzazione di Tikhonov del primo ordine si ottiene utilizzando come matrice di regolarizzazione L la matrice

$$L_1 = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \mathfrak{R}^{(n-1) \times n} \quad (4.16)$$

che rappresenta l'approssimazione discreta dell'operatore di derivata prima.

Regolarizzazione di Tikhonov del secondo ordine

La regolarizzazione di Tikhonov del secondo ordine si ottiene utilizzando come matrice di regolarizzazione L la matrice

$$L_2 = \begin{pmatrix} -1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix} \mathfrak{R}^{(n-2) \times n} \quad (4.17)$$

che rappresenta l'approssimazione discreta dell'operatore di derivata seconda.

Nuova Regolarizzazione di Tikhonov (NTR)

Questo metodo è stato introdotto in [9] e ha l'obiettivo di combinare le proprietà della regolarizzazione di Tikhonov in forma standard e della TSVD. Definito $\tilde{b} = U^T b$, la soluzione di Tikhonov in forma standard può essere scritta come

$$x_{STR} = \sum_{j=1}^n \frac{\sigma_j}{\sigma_j^2 + \lambda^2} \tilde{b}_j v_j \quad (4.18)$$

Da 4.18 si nota che la regolarizzazione di Tikhonov smorza ogni componente della soluzione. La TSVD, invece, non smorza alcuna componente diversa da 0. Questo nuovo metodo, quindi, cerca di smorzare esclusivamente le componenti della soluzione che contribuiscono maggiormente all'errore.

Le componenti della soluzione ad alto indice corrispondono ai vettori singolari v_j caratterizzati da un elevato andamento oscillatorio, i quali contribuiscono maggiormente all'introduzione dell'errore nella soluzione. Pertanto, queste sono le componenti della soluzione che necessitano dello smorzamento. Le componenti della soluzione a basso indice, invece, corrispondono ai vettori singolari v_j meno oscillanti che non introducono errore nella soluzione e che, quindi, non necessitano di smorzamento. Questo metodo, quindi, smorza le componenti della soluzione ad alto indice e mantiene invariate quelle a basso indice.

Definito λ il parametro di regolarizzazione, quando $\sigma_j > \lambda$ la j -esima componente non viene smorzata. Se $\sigma_j \leq \lambda$, invece, il denominatore $\sigma_j^2 + \lambda^2$ in 4.18 viene sostituito con λ^2 . Supponendo che $\sigma_k > \lambda \geq \sigma_{k+1}$, la soluzione x_{NTR} proposta da questo metodo è data da

$$x_{NTR} = \sum_{j=1}^k \sigma_j^{-1} \tilde{b}_j v_j + \sum_{j=k+1}^n \frac{\sigma_j}{\lambda^2} \tilde{b}_j v_j \quad (4.19)$$

Definita la matrice

$$P_\lambda = \text{diag}[\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \lambda^2, \dots, \lambda^2] \quad (4.20)$$

la soluzione 4.19 può essere espressa in forma matriciale come

$$x_{NTR} = V P_\lambda^{-1} \Sigma^T \tilde{b} \quad (4.21)$$

La soluzione 4.21 soddisfa

$$(A^T A + L_\lambda^T L_\lambda) x_{NTR} = A^T b \quad (4.22)$$

per una matrice di regolarizzazione $L_\lambda = D_\lambda V^T$, dove

$$D_\lambda^2 = \text{diag}[\max\{\lambda^2 - \sigma_1^2, 0\}, \max\{\lambda^2 - \sigma_2^2, 0\}, \dots, \max\{\lambda^2 - \sigma_n^2, 0\}] \quad (4.23)$$

La soluzione 4.21 può anche essere scritta nella forma

$$x_{NTR} = \sum_{j=1}^n \varphi_{\lambda,j} \frac{u_j^T b}{\sigma_j} v_j \quad (4.24)$$

con i fattori di filtro pari a

$$\varphi_{\lambda,j} = \begin{cases} 1, & j = 1, \dots, k \\ \frac{\sigma_j^2}{\lambda^2}, & j = k + 1, \dots, n \end{cases} \quad (4.25)$$

Quindi, questo metodo rappresenta una modifica della regolarizzazione di Tikhonov che preserva gli effetti delle componenti della soluzione a basso indice e smorza le componenti della soluzione ad alto indice.

Il parametro di regolarizzazione λ in questo metodo viene determinato attraverso il principio di discrepanza applicato al problema di Tikhonov in forma standard. Definita la funzione

$$\Psi(\lambda) = \|Ax_{STR} - b\|^2 \quad (4.26)$$

il parametro di regolarizzazione viene scelto in modo che sia soddisfatta la seguente relazione:

$$\Psi(\lambda) = \eta^2 \epsilon^2 \quad (4.27)$$

dove $\|e\| = \epsilon$ e $\eta > 1$ rappresenta un "safety factor" indipendente da ϵ .

Sostituendo la soluzione x_{STR} e la SVD di A in 4.26, essa può essere riscritta come

$$\Psi(\lambda) = \sum_{j=1}^n \left(\frac{\sigma_j^2}{\sigma_j^2 + \lambda^2} - 1 \right)^2 \tilde{b}_j^2 \quad (4.28)$$

Se si introduce la funzione $\Phi(\nu) = \Psi(1/\lambda^2)$ data da

$$\Phi(\nu) = \sum_{j=1}^n \frac{\tilde{b}_j^2}{\nu \sigma_j^2 + 1} \quad (4.29)$$

allora Φ decresce ed è convessa per $\nu \geq 0$. Ciò rende conveniente risolvere

$$\Phi(\nu) = \eta^2 \epsilon^2 \quad (4.30)$$

tramite il metodo di Newton, il quale converge quadraticamente e monotonicamente, a differenza della funzione $\Psi(\lambda)$ che non è garantita essere convessa.

Regolarizzazione di Tikhonov Modificata 1 (MTR-1)

Questo metodo è stato introdotto da [11] e cerca di ottenere una soluzione più accurata rispetto al metodo precedente introducendo una modifica alla matrice di regolarizzazione

e, quindi, includendo più informazioni utili nella soluzione.
Introducendo

$$\tilde{L}_\lambda = \tilde{D}_\lambda V^T \quad (4.31)$$

con

$$\tilde{D}_\lambda^2 = \text{diag}[(1 - \omega) \max\{\lambda^2 - \sigma_1^2, 0\}, \dots, (1 - \omega) \max\{\lambda^2 - \sigma_n^2, 0\}] \quad (4.32)$$

dove il valore di λ è lo stesso di quello trovato in 4.1.3 e ω ($\omega < 1$) è un parametro. Quando $\omega = 0$, questo metodo coincide con quello precedente. La soluzione, quindi, può essere scritta come

$$x_{MTR-1} = V(\Sigma^T \Sigma + \tilde{D}_\lambda^2)^{-1} \Sigma^T U^T b \quad (4.33)$$

Se $\sigma_k > \lambda \geq \sigma_{k+1}$, allora

$$\Sigma^T \Sigma + \tilde{D}_\lambda^2 = \text{diag}[\sigma_1^2, \dots, \sigma_k^2, (1 - \omega)\lambda^2 + \omega\sigma_{k+1}^2, \dots, (1 - \omega)\lambda^2 + \omega\sigma_n^2] \quad (4.34)$$

che è una matrice definita positiva. Allora 4.33 è equivalente a

$$x_{MTR-1} = \sum_{j=1}^n \tilde{\varphi}_{\lambda, \omega, j} \frac{u_j^T b}{\sigma_j} v_j, \quad \omega < 1 \quad (4.35)$$

con i fattori di filtro pari a

$$\tilde{\varphi}_{\lambda, \omega, j} = \begin{cases} 1, & 1 \leq j \leq k \\ \frac{\sigma_j^2}{(1 - \omega)\lambda^2 + \omega\sigma_j^2}, & k < j \leq n \end{cases} \quad (4.36)$$

Qui per $1 \leq j \leq k$ i fattori di filtro $\tilde{\varphi}_{\lambda, \omega, j}$ sono uguali ai fattori di filtro $\varphi_{\lambda, j}$ del metodo precedente, mentre per $k < j \leq n$ i fattori di filtro soddisfano la relazione

$$\begin{cases} \frac{\sigma_j^2}{\mu^2} < \frac{\sigma_j^2}{(1 - \omega)\mu^2 + \omega\sigma_j^2}, & 0 < \omega < 1 \\ \frac{\sigma_j^2}{\mu^2} = \frac{\sigma_j^2}{(1 - \omega)\mu^2 + \omega\sigma_j^2}, & \omega = 0 \\ \frac{\sigma_j^2}{\mu^2} > \frac{\sigma_j^2}{(1 - \omega)\mu^2 + \omega\sigma_j^2}, & \omega < 0 \end{cases} \quad (4.37)$$

L'effetto dei fattori di filtro è di eliminare le componenti SVD corrispondenti ai valori singolari più piccoli e, quindi, stabilizzare la soluzione.

In questo metodo, aggiustando il parametro ω è possibile ottenere i fattori di filtro ottimali e informazioni più utili. Pertanto la soluzione ottenuta con questo metodo potrebbe essere una migliore approssimazione della soluzione desiderata rispetto alla soluzione 4.19.

Per determinare il parametro ω si utilizza il metodo della convalida incrociata generalizzata (GCV). Definita

$$A^\# = V(\Sigma^T \Sigma + \tilde{D}_\lambda^2)^{-1} \Sigma^T U^T \quad (4.38)$$

allora il parametro ω può essere determinato minimizzando la funzione GCV

$$G(\omega) = \frac{\|(I_m - AA^\#)b\|_2^2}{(\text{trace}(I_m - AA^\#))^2} \quad (4.39)$$

nell'intervallo $(-\infty, 1)$.

Per ottenere la forma specifica di 4.39 per questo problema, è prima necessario determinare un adeguato valore del parametro di regolarizzazione λ . Quest'ultimo viene scelto in modo che la soluzione del problema di Tikhonov in forma standard x_{STR} soddisfi il principio di discrepanza.

Successivamente viene determinato il parametro k tale che $\sigma_k > \lambda \geq \sigma_{k+1}$.

Definiti i parametri λ e k , il numeratore di 4.39 può essere scritto come

$$\|(I_m - AA^\#)b\|_2^2 = \sum_{i=k+1}^n \left[\frac{\mu^2 - \sigma_i^2}{(1 - \omega)\mu^2 + \omega\sigma_i^2} u_i^T b \right]^2 \quad (4.40)$$

mentre il denominatore di 4.39 come

$$(\text{trace}(I_m - AA^\#))^2 = \left[\sum_{i=k+1}^n \frac{\mu^2 - \sigma_i^2}{(1 - \omega)\mu^2 + \omega\sigma_i^2} \right]^2 \quad (4.41)$$

Combinando 4.40 e 4.41, la funzione GCV è

$$G(\omega) = \frac{\sum_{i=k+1}^n \left[\frac{\mu^2 - \sigma_i^2}{(1 - \omega)\mu^2 + \omega\sigma_i^2} u_i^T b \right]^2}{\left[\sum_{i=k+1}^n \frac{\mu^2 - \sigma_i^2}{(1 - \omega)\mu^2 + \omega\sigma_i^2} \right]^2}, \quad \omega < 1 \quad (4.42)$$

In questo studio il minimo della funzione 4.42 viene determinato utilizzando il comando *fminbnd* di MATLAB nell'intervallo $(-\infty, 1)$.

A questo punto è possibile trovare la soluzione regolarizzata tramite la formula 4.33.

Regolarizzazione di Tikhonov Modificata 2 (MTR-2)

Questo metodo è stato presentato in [10] e cerca di sopprimere il rumore nella soluzione in maniera più efficace rispetto ai metodi precedenti attraverso l'introduzione di una nuova matrice di regolarizzazione \bar{L}_λ definita come

$$\bar{L}_\lambda = \bar{D}_\lambda V^T \quad (4.43)$$

con

$$\bar{D}_\lambda^2 = \text{diag}[\max\{\lambda^2 - \sigma_1^2, 0\}, \dots, \max\{\lambda^2 - \sigma_k^2, 0\}, \max\{\lambda^2 - \sigma_{k+1}^2 + \sigma_n^2, 0\}, \dots, \max\{\lambda^2 - \sigma_n^2 + \sigma_{k+1}^2, 0\}] \quad (4.44)$$

Anche in questo caso il parametro di regolarizzazione λ viene determinato applicando il principio di discrepanza al problema di Tikhonov in forma standard e il parametro k viene scelto in modo tale da soddisfare la relazione $\sigma_k > \lambda \geq \sigma_{k+1}$.

La soluzione, quindi, può essere scritta come

$$x_{MTR-2} = V(\Sigma^T \Sigma + \bar{D}_\lambda^2)^{-1} \Sigma^T U^T b \quad (4.45)$$

dove

$$\Sigma^T \Sigma + \bar{D}_\lambda^2 = \text{diag}[\sigma_1^2, \dots, \sigma_2^2, \lambda^2 + \sigma_n^2, \dots, \lambda^2 + \sigma_{k+1}^2] \quad (4.46)$$

Allora la soluzione 4.45 può essere scritta nella forma

$$x_{MTR-2} = \sum_{j=1}^n \bar{\varphi}_{\lambda,j} \frac{u_j^T b}{\sigma_j} v_j, \quad \omega < 1 \quad (4.47)$$

con i fattori di filtro pari a

$$\bar{\varphi}_{\lambda,j} = \begin{cases} 1, & j = 1, \dots, k \\ \frac{\sigma_j^2}{\lambda^2 + \sigma_{n-j+k+1}^2}, & j = k + 1, \dots, n \end{cases} \quad (4.48)$$

4.2 Metodi iterativi

I metodi presentati finora sono adatti a problemi in cui è possibile calcolare la SVD o le soluzioni di Tikhonov tramite la formulazione ai minimi quadrati per diversi valori del parametro di regolarizzazione. Molti problemi reali, tuttavia, sono caratterizzati da grandi matrici per cui la fattorizzazione della matrice non è possibile o non è conveniente in termini di costo computazionale. In [8] si cerca di trovare una soluzione a questo problema.

Per i metodi di larga scala, quindi, un metodo di regolarizzazione deve evitare la fattorizzazione e soddisfare i seguenti requisiti:

- i principali "elementi costitutivi" di un metodo di larga scala devono essere le moltiplicazioni matrice-vettore, in modo tale da evitare qualunque fattorizzazione della matrice;
- un metodo di larga scala deve permettere all'utente di selezionare il parametro di regolarizzazione senza, però, risolvere il problema da zero per ogni nuovo parametro.

Il primo requisito porta alla scelta dei metodi iterativi, in quanto essi richiedono solamente prodotti matrice-vettore. Una prima idea potrebbe essere quella di utilizzare in metodo iterativo ai minimi quadrati per risolvere il problema di Tikhonov. Lo svantaggio di questo approccio, tuttavia, è che per ogni parametro di regolarizzazione λ è necessario rilanciare il metodo iterativo e ciò diventa insostenibile quando il tempo computazionale è elevato.

Per soddisfare anche il secondo requisito, quindi, vengono scelti metodi iterativi che evitano un parametro di regolarizzazione predeterminato e utilizzano il numero di iterazioni per questa funzione.

I metodi iterativi necessitano sempre di un vettore di partenza $x^{[0]}$ specificato dall'utente e producono una sequenza di iterate $x^{[1]}, x^{[2]}, \dots$ che convergono a qualche soluzione. Per alcuni metodi iterativi, queste iterate tendono inizialmente verso la soluzione esatta x^{exact} , ma a partire da un certo valore del numero di iterazioni esse iniziano a divergere dalla soluzione esatta e a convergere verso la soluzione "banale" contaminata dall'errore. Questo tipo di convergenza è chiamata *semiconvergenza* e i metodi che verranno presentati successivamente saranno caratterizzati da essa.

4.2.1 Metodo di Kaczmarz o ART

Nel metodo di Kaczmarz, noto anche come tecnica di ricostruzione algebrica (ART), la generica k -esima iterazione consiste in una "spazzata" delle righe della matrice A e il vettore della soluzione $x^{[k]}$ viene aggiornato secondo il seguente algoritmo:

$$\begin{aligned}
& x^{[k^{(0)}]} = x^{[k]} \\
& \text{for } i = 1, \dots, m \\
& \quad x^{[k^{(i)}]} = x^{[k^{(i-1)}]} + \frac{b_i - a_i^T x^{[k^{(i-1)}]}}{\|a_i\|_2^2} a_i \\
& \text{end} \\
& x^{[k+1]} = x^{[k^{(m)}]}
\end{aligned} \tag{4.49}$$

dove b_i rappresenta l' i -esima componente del vettore b e a_i è l' i -esima riga della matrice A .

Per dimostrare la bontà di questo metodo è possibile fornire una spiegazione di esso da un punto di vista geometrico. Come si può vedere dall'algoritmo precedente, l'iterata $x^{[k^{(i)}]}$ viene ottenuta proiettando l'iterata precedente $x^{[k^{(i-1)}]}$ nell'iperpiano $\mathcal{H} = \{x \mid a_i^T x = b_i\}$ definito dall' i -esima riga a_i^T e il corrispondente elemento b_i . Definito θ_{ij} l'angolo tra il vettore a_i e il j -esimo vettore singolare destro v_j come

$$\cos \theta_{ij} = \frac{a_i^T v_j}{\|a_i\|_2} = \frac{e_i^T A v_j}{\|a_i\|_2} = \frac{\sigma_j e_i^T u_j}{\|a_i\|_2} = \frac{\sigma_j u_{ij}}{\|a_i\|_2} \tag{4.50}$$

si nota che la direzione dei vettori a_i è quasi coincidente con quella dei vettori v_j corrispondenti a valori singolari σ_j grandi, ovvero le componenti della soluzione non contaminate dall'errore e, quindi, migliori dal punto di vista della ricostruzione della soluzione regolarizzata. Inoltre, come detto in precedenza, le iterate del metodo sono proiettate nell'iperpiano definito da a_i e b_i e, quindi, nella direzione a_i . Pertanto, le iterate di Kaczmarz hanno grandi componenti nella direzione dei vettori singolari destri corrispondenti a grandi valori singolari e piccole in quella dei vettori singolari corrispondenti a piccoli valori singolari, ovvero le iterate tendono verso le soluzioni regolarizzate.

4.2.2 Metodo CGLS

Questo metodo rientra nella categoria dei *metodi di proiezione*; pertanto, prima dell'enunciato del metodo viene fornita una breve spiegazione riguardante questa categoria di metodi.

Metodi di proiezione

Dato che per i problemi discreti mal posti è impossibile calcolare la soluzione esatta x^{exact} , i metodi di regolarizzazione presentati finora cercano un' approssimazione di questa soluzione in un sottospazio di \mathfrak{R}^n . E' stato dimostrato, infatti, che la riduzione della dimensione dello spazio vettoriale in cui cercare la soluzione ha un effetto regolarizzante. Nel caso della TSVD, ad esempio, il sottospazio in cui viene ricercata la soluzione ha come base vettoriale i primi k vettori singolari destri.

Se i vettori della base del generico sottospazio entro cui si vuole ricercare la soluzione

vengono raccolti nelle colonne della matrice $W_k = (w_1, \dots, w_k) \in \mathfrak{R}^{n \times k}$, allora il problema ai minimi quadrati può essere riformulato come

$$\min_x \|Ax - b\|_2 \quad \text{con} \quad x \in \mathcal{W} = \text{span}\{w_1, \dots, w_k\} \quad (4.51)$$

Il problema vincolato 4.51 può essere riformulato attraverso l'introduzione del cambiamento di variabili $x = W_k y$, dove $y \in \mathfrak{R}^k$ è la nuova incognita, nel seguente modo:

$$x^{(k)} = W_k y^{(k)}, \quad y^{(k)} = \operatorname{argmin}_y \|(AW_k)y - b\|_2 \quad (4.52)$$

Il problema ai minimi quadrati 4.52 viene chiamato *problema proiettato* in quanto viene ottenuto proiettando il problema originale in un sottospazio vettoriale k -dimensionale.

Il passo successivo consiste nella scelta di un'adeguata base vettoriale del sottospazio. Poiché la base "ottimale", data dai vettori singolari della matrice A , non può essere determinata per problemi di larga scala, è necessario trovare una buona approssimazione di questa base. A tale scopo viene in aiuto il *sottospazio di Krylov* associato ad A e b , definito come la base delle potenze della matrice $A^T A$ applicato a $A^T b$:

$$K_k \equiv \text{span}\{A^T b, (A^T A)A^T b, (A^T A)^2 A^T b, \dots, (A^T A)^{k-1} A^T b\} \quad (4.53)$$

I vettori di questa base diventano sempre più ricchi nella direzione dell'autovettore principale di $A^T A$, ovvero il principale vettore singolare destro v_1 . Pertanto, questi vettori non sono immediatamente utilizzabili come vettori della base. Tuttavia, effettuando una ortonormalizzazione di essi è possibile ottenere una base di vettori orientati non solo nella direzione di v_1 ma anche degli altri vettori singolari destri. In particolare, ogni vettore w_i della base ortonormale è dominato da pochi vettori singolari v_j con indice $j \approx i$. Pertanto questa base rappresenta una buona scelta per un metodo di proiezione. Un metodo efficace per la determinazione di questa base è l'applicazione del metodo della *bidagonalizzazione di Lanczos* che verrà presentato successivamente.

L'algoritmo CGLS ha una stretta relazione con i metodi di proiezione appena descritti in quanto è possibile dimostrare che la soluzione $x^{(k)}$ ottenuta applicando k passi dell'algoritmo del Gradiente Coniugato, con vettore di partenza nullo, al sistema di equazioni normali $A^T A x = A^T b$ associato al problema ai minimi quadrati non regolarizzato $\min_x \|Ax - b\|_2$ coincide con la soluzione del problema proiettato 4.51 con $W_k = K_k$:

$$x^{(k)} = \operatorname{argmin}_x \|Ax - b\|_2 \quad \text{con} \quad x \in K_k \quad (4.54)$$

Il modo più efficace per applicare il metodo del Gradiente Coniugato al sistema di equazioni normali è detto *algoritmo CGLS* e ha la seguente forma:

$$\begin{aligned}
 x^{(0)} &= 0 \quad (\text{starting vector}) \\
 r^{(0)} &= b - Ax^{(0)} \\
 d^{(0)} &= A^T r^{(0)} \\
 &\text{for } k = 1, 2, \dots \\
 \bar{\alpha}_k &= \left\| A^T r^{(k-1)} \right\|_2^2 / \left\| Ad^{(k-1)} \right\|_2^2 \\
 x^{(k)} &= x^{(k-1)} + \bar{\alpha}_k d^{(k-1)} \\
 r^{(k)} &= r^{(k-1)} - \bar{\alpha}_k Ad^{(k-1)} \\
 \bar{\beta}_k &= \left\| A^T r^{(k)} \right\|_2^2 / \left\| A^T r^{(k-1)} \right\|_2^2 \\
 d^{(k)} &= A^T r^{(k)} + \bar{\beta}_k d^{(k-1)} \\
 &\text{end}
 \end{aligned} \tag{4.55}$$

Il vantaggio di questo algoritmo è che esso non richiede il calcolo esplicito della base ortonormale, ma solo una moltiplicazione per A e una per A^T ogni iterazione.

La norma della soluzione $\|x^{(k)}\|_2$ e del residuo $\|Ax^{(k)} - b\|_2$ sono funzioni monotone di k ,

$$\|x^{(k)}\|_2 \geq \|x^{(k-1)}\|_2, \quad \|Ax^{(k)} - b\|_2 \leq \|Ax^{(k-1)} - b\|_2, \quad k = 1, 2, \dots \tag{4.56}$$

e ciò permette l'utilizzo del metodo della curva-L per stoppare l'algoritmo prima che si verifichi il fenomeno della semiconvergenza. E', inoltre, possibile dimostrare che la soluzione appena trovata può essere espressa come una soluzione SVD filtrata. Data che la soluzione, infatti, appartiene al sottospazio di Krylov, essa può essere espressa come combinazione lineare dei vettori della sua base:

$$x^{(k)} = c_1 A^T b + c_2 (A^T A) A^T b + \dots + c_k (A^T A)^{k-1} A^T b \tag{4.57}$$

e, inserendo la SVD di A si ottiene

$$x^{(k)} = V \Phi^{(k)} \Sigma^{-1} U^T b \tag{4.58}$$

dove $\Phi^{(k)} = \text{diag}(\varphi_1^{(k)}, \dots, \varphi_n^{(k)})$ è una matrice diagonale contenente i fattori di filtro della soluzione CGLS

$$\varphi_i^{(k)} = c_1 \sigma_i^2 + c_2 \sigma_i^4 + c_3 \sigma_i^6 + \dots + c_k \sigma_i^{2k}, \quad i = 1, 2, \dots, n \tag{4.59}$$

4.2.3 Metodo CGLS preconditionato

L'algoritmo CGLS, nel caso di problemi inversi mal posti, ha una bassa velocità di convergenza e, per questo motivo, spesso ad esso viene associato un preconditionatore. Quest'ultimo, nel caso dei problemi inversi mal posti, aumenta la velocità di convergenza dell'algoritmo migliorando la posizione e la distribuzione dei valori singolari più grandi e mantenendo il resto dello spettro singolare invariato. Nel caso dei problemi inversi mal posti i vettori singolari a bassa frequenza corrispondenti ai valori singolari più grandi, infatti, hanno un ruolo molto importante nella ricostruzione della soluzione esatta, mentre i vettori singolari ad alta frequenza corrispondenti ai valori singolari più piccoli hanno un ruolo importante nell'amplificazione del rumore nella soluzione. I sottospazi generati da questi due gruppi di vettori sono detti sottospazio di segnale e sottospazio di rumore e, dato che nei problemi inversi mal posti il numero di vettori singolari grandi è piccolo rispetto alla dimensione della matrice, la dimensione dello spazio di segnale è molto piccola.

Un tipo di preconditionatore che risulta essere molto efficace per questo scopo viene ottenuto in [12] effettuando k passi dell'algoritmo di bidiagonalizzazione di Lanczos.

Definita $A = U\Sigma V^T$ la SVD di A , allora $A^T A = V\Sigma^2 V^T$ e per un adeguato valore di k si ha che

$$P = V \begin{pmatrix} \Sigma_1^{-2} & 0 \\ 0 & I_2 \end{pmatrix} V^T \quad (4.60)$$

può essere utilizzato come un preconditionatore regolarizzato inverso di $A^T A = A^T b$, dove $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_k)$, $\Sigma_2 = \text{diag}(\sigma_{k+1}, \dots, \sigma_n)$ e $I = \begin{pmatrix} I_1 & \\ & I_2 \end{pmatrix}$. In questo caso il sistema preconditionato diventa

$$PA^T A = V \begin{pmatrix} I_1 & 0 \\ 0 & \Sigma_2^2 \end{pmatrix} V^T \quad (4.61)$$

Ciò mostra che il preconditionatore P ha raggruppato i primi k vettori singolari attorno a 1 e ha mantenuto invariati i restanti, separando gli spazi di segnale e di rumore. In questo modo viene evitata la combinazione dei due sottospazi, la soluzione non viene contaminata dall'errore e ciò ha un effetto regolarizzante.

Per i problemi di larga scala, tuttavia, questo preconditionatore non può essere calcolato in maniera diretta utilizzando l'espressione 4.60 in quanto non è conveniente o possibile calcolare tutti i vettori singolari. Tuttavia, definendo $V = [V_1, V_2]$ dove $V_1 = [v_1, \dots, v_k]$ e $V_2 = [v_{k+1}, \dots, v_n]$, P può essere scritto nella forma chiusa

$$P = V_1 \Sigma_1^{-2} V_1^T + V_2 V_2^T \quad (4.62)$$

Inoltre, sfruttando l'ortogonalità di V , si ha che $V_2 V_2^T = I - V_1 V_1^T$ e P può essere scritta come

$$P = V_1 \Sigma_1^{-2} V_1^T + (I - V_1 V_1^T) \quad (4.63)$$

Ciò dimostra che è possibile ottenere P solamente utilizzando Σ_1 e V_1 . Poiché nel caso di problemi di grande scala non è possibile o non è conveniente calcolare la SVD di A , V_1 e Σ_1 vengono approssimati con X_k e $\bar{\Sigma}_k$ ottenuti effettuando k passi della bidiagonalizzazione di Lanczos:

$$\begin{aligned} P &\approx X_k \bar{\Sigma}_k^{-2} X_k^T + (I - X_k X_k^T) \\ &= \bar{V}_k Q_k \bar{\Sigma}_k^{-2} Q_k^T \bar{V}_k^T + (I - \bar{V}_k Q_k Q_k^T \bar{V}_k^T) \\ &= \bar{V}_k (B_k^T B_k)^{-1} \bar{V}_k^T + (I - \bar{V}_k \bar{V}_k^T) \end{aligned} \quad (4.64)$$

Quindi, effettuando k passi della bidiagonalizzazione di Lanczos è possibile ottenere il preconditionatore inverso regolarizzato

$$M = \bar{V}_k (B_k^T B_k)^{-1} \bar{V}_k^T + (I - \bar{V}_k \bar{V}_k^T) \quad (4.65)$$

Selezione dell'indice k

Un metodo per definire il numero di passi del processo di bidiagonalizzazione è quello di scegliere k come il più piccolo intero per cui

$$\bar{\sigma}_k < \tau \bar{\sigma}_1 \quad (4.66)$$

dove $\tau = \sqrt{\text{precisione di macchina}}$ e $\bar{\sigma}_i$ sono i valori singolari della matrice B_k ottenuta dal processo di bidiagonalizzazione.

4.2.4 Metodo LSQR

Questo metodo è stato introdotto in [13] e rientra nella categoria dei metodi di proiezione. Esso si basa sulla procedura di bidiagonalizzazione di Lanczos ed è analiticamente equivalente al metodo standard del gradiente coniugato, anche se possiede proprietà numeriche più favorevoli.

Come si vedrà anche nel paragrafo successivo 4.3, dato il generico sistema $Ax = b$, la procedura di bidiagonalizzazione associato ad esso con vettore iniziale uguale a b può essere scritta come

$$\left. \begin{aligned} \beta_1 u_1 &= b, & \alpha_1 v_1 &= A^T u_1 \\ \beta_{i+1} u_{i+1} &= A v_i - \alpha_i u_i \\ \alpha_{i+1} v_{i+1} &= A^T u_{i+1} - \beta_{i+1} v_i \end{aligned} \right\}, \quad i = 1, 2, \dots \quad (4.67)$$

dove $\alpha_i \geq 0$ e $\beta_i \geq 0$ sono scelti in modo che $\|u_i\| = \|v_i\| = 1$. Questa procedura permette di definire le matrici ortogonali U_k e V_k e la matrice bidiagonale inferiore B_k

$$U_k \equiv [u_1, u_2, \dots, u_k], \quad B_k \equiv \begin{bmatrix} \alpha_1 & & & & & \\ \beta_2 & \alpha_2 & & & & \\ & \beta_3 & \ddots & & & \\ & & \ddots & \ddots & & \\ & & & \alpha_k & & \\ & & & & \beta_{k+1} & \end{bmatrix}, \quad V_k \equiv [v_1, v_2, \dots, v_k] \quad (4.68)$$

tali che

$$U_{k+1}(\beta_1 e_1) = b \quad (4.69)$$

$$AV_k = U_{k+1}B_k \quad (4.70)$$

La matrice V_k , in particolare, ha come colonne i vettori della base del sottospazio in cui il problema iniziale viene proiettato.

Il problema proiettato ai minimi quadrati ottenuto dopo k passi della procedura di bidiagonalizzazione è definito come

$$\min \|\beta_1 e_1 - B_k y_k\| \quad (4.71)$$

e l'approssimazione della soluzione del problema iniziale è

$$x_k = V_k y_k \quad (4.72)$$

Questo metodo risolve il problema proiettato 4.71 utilizzando la fattorizzazione QR di B_k :

$$Q_k [B_k \quad \beta_1 e_1] = \left[\begin{array}{c|c} R_k & \begin{matrix} f_k \\ \bar{\phi}_{k+1} \end{matrix} \end{array} \right] \equiv \left[\begin{array}{cccccc|c} \rho_1 & \theta_2 & & & & & \phi_1 \\ & \rho_2 & \theta_3 & & & & \phi_2 \\ & & \ddots & \ddots & & & \vdots \\ & & & \rho_{k-1} & \theta_k & & \phi_{k-1} \\ & & & & \rho_k & & \phi_k \\ \hline & & & & & & \bar{\phi}_{k+1} \end{array} \right] \quad (4.73)$$

dove Q_k è un prodotto di rotazioni nel piano che hanno lo scopo di eliminare gli elementi sottodiagonali di B_k . Il vettore y_k , quindi, può essere espresso come

$$R_k y_k = f_k \quad (4.74)$$

Tuttavia, y_k in 4.74 non ha elementi in comune con y_{k-1} . Invece, $[R_k \ f_k]$ è la stessa $[R_{k-1} \ f_{k-1}]$ con una nuova riga e una nuova colonna in aggiunta. Quindi, un modo efficace di combinare 4.72 e 4.74 è

$$x_k = V_k R_k^{-1} f_k \equiv D_k f_k \quad (4.75)$$

dove le colonne di $D_k \equiv [d_1 \ d_2 \ \dots \ d_k]$ possono essere trovate dal sistema $R_k^T D_k^T = V_k^T$. Con $d_0 = x_0 = 0$, si ha che

$$d_k = \frac{1}{\rho_k} (v_k - \theta_k d_{k-1}) \quad (4.76)$$

$$x_k = x_{k-1} + \phi_k d_k \quad (4.77)$$

e solo l'ultima iterata deve essere salvata.

4.3 Metodo ibrido

Il sottospazio di Krylov in teoria rappresenta una buona base per i metodi di proiezione come l' algoritmo CGLS in quanto rappresenta una buona approssimazione della base "ottimale", ovvero quella costituita dai vettori singolari destri. Nella realtà, tuttavia, la base di Krylov potrebbe non essere una buona soluzione in quanto essa è costruita utilizzando A e b , ovvero il vettore dei termini noti perturbati dall'errore, e non b^{exact} . Per questo motivo la base di Krylov include sia vettori della base desiderati sia, a causa del rumore, vettori della base non desiderati ai fini della determinazione della soluzione regolarizzata. Di conseguenza la soluzione ottenuta tramite il metodo di proiezione potrebbe non essere accurata in quanto potrebbe contenere vettori rumorosi.

La soluzione a questo problema consiste nell'applicare un metodo di regolarizzazione al problema proiettato. Nel caso della regolarizzazione di Tikhonov, il problema proiettato regolarizzato può essere scritto come

$$x_\lambda^{(k)} = W_k y_\lambda^{(k)}, \quad y_\lambda^{(k)} = \operatorname{argmin}_y \{ \| (AW_k)y - b \|^2 + \lambda^2 \| y \|^2 \} \quad (4.78)$$

Come anticipato nel paragrafo precedente, per determinare i vettori della base ortonormale che costituiscono le colonne della matrice W_k si applica il metodo della *bidiagonalizzazione di Lanczos* alla matrice A . Se come vettore iniziale si prende b e i coefficienti α_k e β_k vengono scelti in modo che i vettori w_k e z_k siano normalizzati, l'algoritmo può essere scritto come

$$\begin{aligned}
 &w_0 = 0 \\
 &\beta_1 z_1 = b \\
 &\text{for } k = 1, 2, \dots \\
 &\quad \alpha_k w_k = A^T z_k - \beta_k w_{k-1} \\
 &\quad \beta_{k+1} z_{k+1} = A w_k - \alpha_k z_k \\
 &\text{end}
 \end{aligned} \tag{4.79}$$

Dopo k iterazioni, questo algoritmo produce due matrici con colonne ortonormali $W_k \in \mathfrak{R}^{n \times k}$ e $Z_k \in \mathfrak{R}^{m \times (k+1)}$

$$W_k = (w_1, w_2, \dots, w_k), \quad Z_{k+1} = (z_1, z_2, \dots, z_{k+1}) \tag{4.80}$$

e una matrice bidiagonale inferiore $B_k \in \mathfrak{R}^{(k+1) \times k}$

$$B_k = \begin{pmatrix} \alpha_1 & & & & \\ \beta_2 & \alpha_2 & & & \\ & \beta_3 & \ddots & & \\ & & \ddots & \alpha_k & \\ & & & & \beta_{k+1} \end{pmatrix} \tag{4.81}$$

tale che

$$A W_k = Z_{k+1} B_k \tag{4.82}$$

Le colonne della matrice ottenuta W_k rappresentano i vettori della base ortonormale ricercati.

Inserendo l'equazione 4.82 nel problema proiettato regolarizzato 4.78, si ottiene la seguente formulazione alternativa del problema:

$$x_\lambda^{(k)} = W_k y_\lambda^{(k)}, \quad y_\lambda^{(k)} = \operatorname{argmin}_y \{ \|B_k y - \beta_1 e_1\|_2^2 + \lambda^2 \|y\|_2^2 \} \tag{4.83}$$

dove $e_1 = (1, 0, \dots, 0)^T \in \mathfrak{R}^{k+1}$ e $z_1 = b/\beta_1$.

Inoltre, la norma della soluzione $x_\lambda^{(k)}$ e la norma del residuo $\|A x_\lambda^{(k)} - b\|_2$ possono essere ottenute direttamente dalla soluzione $y_\lambda^{(k)}$ del problema di Tikhonov. Infatti,

$$x_\lambda^{(k)} = \|W_k y_\lambda^{(k)}\|_2 = \|y_\lambda^{(k)}\|_2, \quad \|A x_\lambda^{(k)} - b\|_2 = \|B_k y_\lambda^{(k)} - \beta_1 e_1\|_2 \tag{4.84}$$

Pertanto, la norma della soluzione e la norma del residuo possono essere calcolate direttamente attraverso $y_\lambda^{(k)}$ e B_k . Il vantaggio di ciò è che per determinare il parametro di

regolarizzazione possono essere utilizzati metodi basati su queste norme, come il metodo della curva L o il principio di discrepanza.

4.4 Scelta del parametro di regolarizzazione

Definiti alcuni metodi di regolarizzazione basati sul filtraggio delle componenti SVD, è necessario definire una tecnica affidabile e automatica per la scelta del parametro di regolarizzazione, ovvero k (per TSVD) o λ (per Tikhonov). Nello specifico, sarebbe necessario un metodo efficace, robusto e affidabile per il calcolo del parametro di regolarizzazione a partire dai dati noti e che non richieda il calcolo della SVD (non praticabile per problemi di grande dimensione). Sfortunatamente, un metodo del genere non è stato ancora trovato; attualmente si ha a disposizione una collezione di metodi presentati nel dettaglio in [8] che, sotto certe assunzioni, lavorano bene, ma ognuno di essi occasionalmente non produce buoni risultati.

Ogni metodo per la scelta del parametro di regolarizzazione dovrebbe cercare di minimizzare gli errori nella soluzione regolarizzata e, per questo motivo, è utile analizzare questi errori. Definita la matrice diagonale $\phi^{[\lambda]}$ contenente i fattori di filtro di Tikhonov

$$\phi^{[\lambda]} = \begin{pmatrix} \varphi_1^{[\lambda]} & & \\ & \ddots & \\ & & \varphi_n^{[\lambda]} \end{pmatrix} \quad (4.85)$$

dove $\varphi_i^{[\lambda]} = \sigma_i^2 / (\sigma_i^2 + \lambda^2)$, allora è possibile scrivere la soluzione regolarizzata x_λ come

$$x_\lambda = V \phi^{[\lambda]} \Sigma^{-1} U^T b \quad (4.86)$$

Un'espressione simile è valida anche per la soluzione TSVD con $\phi^{[\lambda]}$ sostituita da $\phi^{[\lambda]} = \text{diag}(1, \dots, 1, 0, \dots, 0)$.

Il termine di destra è dato dalla somma di un "segnale" esatto e un errore $b = Ax^{exact} + e$ e l'errore nella soluzione regolarizzata di Tikhonov è, quindi, dato da

$$x^{exact} - x_\lambda = x^{exact} - V \phi^{[\lambda]} \Sigma^{-1} U^T b = V(I - \phi^{[\lambda]}) V^T x^{exact} - V \phi^{[\lambda]} \Sigma^{-1} U^T e \quad (4.87)$$

Il primo termine è detto *errore di regolarizzazione* ed è dovuto all'introduzione del filtraggio:

$$\Delta x_{bias} = V(I - \phi^{[\lambda]}) V^T x^{exact} = \sum_{i=1}^n (1 - \varphi_i^{[\lambda]}) (v_i^T x^{exact}) v_i \quad (4.88)$$

Esso esprime la deviazione del valore atteso di x_λ dalla soluzione esatta.

Il secondo termine è detto *errore di perturbazione* ed è dovuta al filtraggio della componente rumorosa nei dati:

$$\Delta x_{pert} = V\phi^{[\lambda]}\Sigma^{-1}U^T e \quad (4.89)$$

Il principale obiettivo della regolarizzazione tramite filtraggio, in presenza della matrice $\phi^{[\lambda]}$, è evitare che questo errore di perturbazione diventi troppo grande e rovini la soluzione.

Per le soluzioni TSVD, gli errori di regolarizzazione e perturbazione hanno la seguente forma:

$$\Delta x_{bias} = \sum_{i=k+1}^n (v_i^T x^{exact})v_i, \quad \Delta x_{pert} = \sum_{i=1}^k \frac{u_i^T e}{\sigma_i} v_i \quad (4.90)$$

È possibile utilizzare il parametro di troncamento k per evitare un errore di perturbazione troppo grande, a causa della divisione per piccoli valori singolari, al costo di introdurre un errore di regolarizzazione nella soluzione regolarizzata.

Entrambi gli errori sono presenti nella soluzione regolarizzata e la loro dimensione dipende dal parametro di regolarizzazione. Se k è vicino a n , allora la maggior parte delle componenti SVD sono incluse e l'errore di regolarizzazione è piccolo, ma l'errore di perturbazione è grande in quanto viene applicato un filtraggio molto piccolo. D'altra parte, se k è piccolo, molte componenti SVD sono escluse e l'errore di perturbazione è piccolo, ma l'errore di regolarizzazione è grande. L'obiettivo della scelta del parametro k , quindi, è bilanciare la dimensione dei due errori.

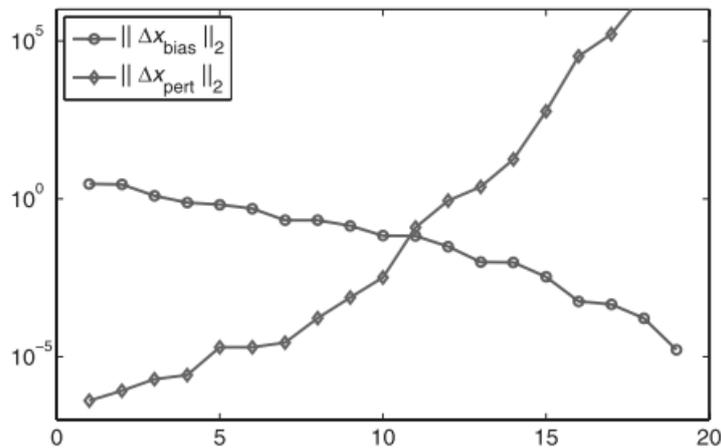


Figura 4.3: Errori di perturbazione e regolarizzazione [8]

Lo stesso discorso vale, ovviamente, anche per la scelta del parametro di regolarizzazione di Tikhonov. Se λ è piccolo, tutti i fattori di filtro $\varphi_i^{[\lambda]}$ sono vicini a 1 e l'errore di regolarizzazione è piccolo (mentre l'errore di perturbazione è grande). Se λ è grande,

molti fattori di filtro sono piccoli e, quindi, anche l'errore di perturbazione è piccolo, mentre l'errore di regolarizzazione è grande.

4.4.1 Principio di discrepanza

Se per semplicità si considera un errore bianco gaussiano, cioè tutti gli elementi di $e \in \mathfrak{R}^m$ provengono dalla stessa distribuzione gaussiana con valor medio nullo e deviazione standard η il valore atteso della norma della perturbazione può essere definito come $\|e\|_2 = n^{\frac{1}{2}}\eta$. Un approccio ancora più conservativo si ha includendo un "safety factor" ν_{dp} .

Secondo questo principio, il parametro di troncamento k dev'essere scelto in modo che la norma del residuo sia uguale alla "discrepanza" nei dati, misurata da $\nu_{dp}\|e\|_2$. Poiché raramente questa relazione può essere soddisfatta esattamente, viene scelto il *più grande* k tale che $\|Ax_k - b\|_2 \geq \nu_{dp}\|e\|_2$.

TSVD:

$$k = k_{DP} \quad \text{tale che} \quad \|Ax_k - b\|_2 \geq \nu_{dp}\|e\|_2 > \|Ax_{k+1} - b\|_2 \quad (4.91)$$

Lo stesso discorso vale per la regolarizzazione di Tikhonov con x_k sostituita da x_λ e, in questo caso, l'uguaglianza può essere ottenuta.

Regolarizzazione di Tikhonov:

$$\lambda = \lambda_{DP} \quad \text{tale che} \quad \|Ax_\lambda - b\|_2 = \nu_{dp}\|e\|_2 \quad (4.92)$$

In entrambi i casi esiste un'unica soluzione in quanto la norma del residuo varia monotonicamente con k e λ .

Il principale svantaggio del principio di discrepanza è che spesso $\|e\|_2$ (o η) non è noto esattamente e, di conseguenza, è necessario utilizzare una stima della norma (o della deviazione standard). Sfortunatamente la qualità del parametro di regolarizzazione calcolato k_{DP} o λ_{DP} è molto sensibile all'accuratezza della stima di $\|e\|_2$. In particolare, una stima troppo bassa porta ad una soluzione sotto-regolarizzata (in quanto k è troppo grande e λ è troppo piccolo).

In termini della curva-L, il parametro di regolarizzazione λ_{DP} scelto con il principio di discrepanza corrisponde al punto di intersezione tra la curva-L e una linea verticale posizionata a $\|Ax_\lambda - b\|_2 \approx \nu_{dp}\|e\|_2$. Inoltre, la parte quasi verticale della curva-L si trova a $\|Ax_\lambda - b\|_2 \approx \|e\|_2$; di conseguenza una sottostima della norma dell'errore nel principio di discrepanza porta all'intersezione della curva-L con la linea verticale in un punto in cui λ è molto piccolo e la norma della soluzione è grande (soluzione fortemente sotto-

regolarizzata). Anche una piccola sottostima della norma dell'errore può avere un grande impatto sul parametro di regolarizzazione, rendendo il metodo rischioso.

4.4.2 Criterio della Curva L

Il comportamento della curva L è fortemente legato alla dimensione dell'errore di regolarizzazione Δx_{bias} e dell'errore di perturbazione Δx_{pert} . quindi, è possibile studiare la dipendenza della soluzione dall'errore di regolarizzazione tramite questa curva. Per comprendere ciò, per semplicità si considera il metodo TSVD (tutto ciò può essere esteso a Tikhonov); definito k_η l'indice di transizione tra l'andamento decrescente e quello piatto dei coefficienti $|u_i^T b|$, si verifica che:

- Quando k è più piccolo di k_η , l'errore di regolarizzazione domina in x_k e il comportamento della sua norma è largamente determinato da quante componenti SVD vengono incluse. Minore è il numero di componenti, minore è la norma della soluzione $\|x_k\|_2$ e maggiore è la norma del residuo $\|Ax_k - b\|_2$. In particolare,
 - La norma della soluzione $\|x_k\|_2$ è quasi una costante data da $\|x^{exact}\|_2$ eccetto per k molto piccoli dove $\|x_k\|_2$ diventa più piccola quando $k \rightarrow 0$
 - La norma del residuo $\|Ax_k - b\|_2$ aumenta quando $k \rightarrow 0$ fino a quando raggiunge il suo valore massimo $\|b\|_2$ per $k = 0$
- Quando k è più grande di k_η , l'errore di perturbazione domina x_k e il comportamento della sua norma è dominato da questa componente. In particolare,
 - La norma della soluzione $\|x_k\|_2$ aumenta all'aumentare di k , in quanto vengono incluse componenti di errore sempre maggiori
 - La norma del residuo $\|Ax_k - b\|_2$, d'altra parte, rimane quasi costante al valore approssimato di $\|e\|_2$ (eccetto quando $k \approx \eta$)

Da quest'analisi segue che la curva-L è costituita da due parti distinte, ovvero una parte piatta (quando l'errore di regolarizzazione domina) e una parte verticale (quando l'errore di perturbazione domina). La scala log-log per la curva-L evidenzia le diverse caratteristiche di queste due parti.

Un'importante osservazione è che la curva-L deve avere una parte di transizione tra la parte orizzontale e verticale. Molto spesso questa transizione avviene in una piccola regione e la curva-L presenta un angolo tra le due parti. Inoltre, questa regione di transizione è associata ai valori di λ o k per cui la componente di errore dominante cambia.

L'idea chiave del criterio della curva-L è scegliere un valore di λ in corrispondenza dell'angolo della curva-L, nella speranza che questo valore fornisca un buon equilibrio tra gli errori di regolarizzazione e perturbazione.

Per la regolarizzazione di Tikhonov, una definizione naturale dell'angolo è il punto della curva-L con massima curvatura. La curvatura dev'essere misurata nel sistema log-log,

quindi è necessario calcolare la curvatura \hat{c} della curva $(\log \|Ax_\lambda - b\|_2, \log \|x_\lambda\|_2)$.
Definiti

$$\xi = \|x_\lambda\|_2^2 \tag{4.93}$$

$$\rho = \|Ax_\lambda - b\|_2^2 \tag{4.94}$$

tramite alcuni passaggi matematici è possibile trovare un'espressione della curvatura della curva-L in scala log-log in termini delle tre quantità ξ , ρ e ξ' :

$$\hat{c}_\lambda = 2 \frac{\xi \rho \lambda^2 \xi' \rho + 2 \lambda \xi \rho + \lambda^4 \xi \xi'}{\xi'} \frac{1}{(\lambda^2 \xi^2 + \rho^2)^{3/2}} \tag{4.95}$$

La conclusione di quest'analisi è che per ogni valore di λ è possibile calcolare la curvatura \hat{c}_λ , come mostrato nella figura

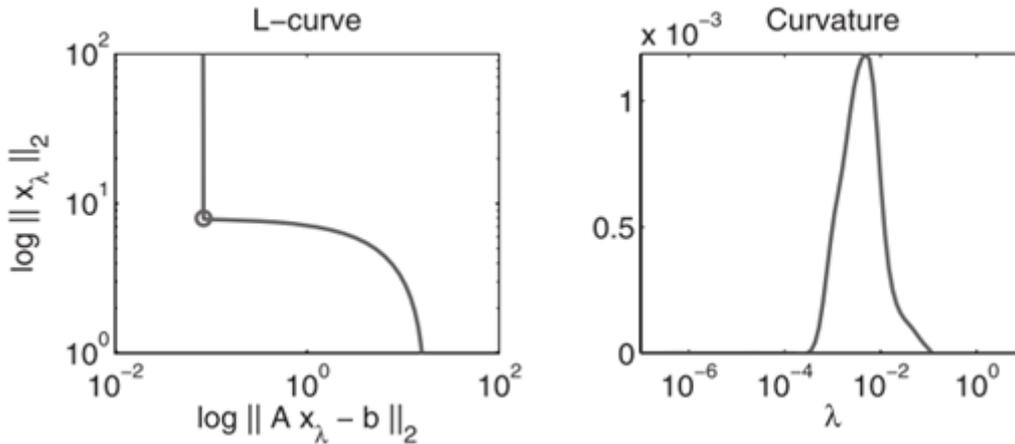


Figura 4.4: Andamento della curva-L e della sua curvatura [8]

Regolarizzazione di Tikhonov:

$\lambda = \lambda_L$ tale che la curvatura \hat{c}_λ sia massima

Quest'analisi non è valida immediatamente per il metodo TSVD in quanto questo metodo produce un set finito di soluzione per $k=1,2,\dots$ e, di conseguenza, la curva-L corrispondente consiste in un set finito di punti. Tuttavia, è possibile estendere l'analisi anche a questo caso ed è bene porre la soluzione TSVD in corrispondenza dell'angolo globale della curva-L discreta:

TSVD:

$k = k_L$ in corrispondenza dell'angolo globale della curva-L discreta

Questo criterio, tuttavia, non sempre garantisce una buona scelta del parametro di regolarizzazione. Un esempio in cui il criterio della curva-L fallisce si ha quando le componenti SVD $v_i^T x^{exact}$ della soluzione esatta decadono velocemente a 0 (in questo caso x^{exact} appare molto regolare in quanto è dominata dai primi vettori singolari v_i , i quali sono regolari). L'angolo della curva-L è posizionato nel punto dove la norma della soluzione diventa dominata dalle componenti SVD rumorose. Spesso ciò è dove le componenti iniziano a crescere, ma per soluzioni molto regolari ciò avviene per k più grandi dopo aver incluso troppe componenti rumorose e, quindi, λ_L produce una soluzione sottoregolarizzata.

Un altro esempio in cui il criterio della curva-L fallisce è quando il cambiamento delle norme del residuo e della soluzione è piccolo per due valori consecutivi di k . Anche in questo caso è necessario introdurre troppi coefficienti SVD rumorosi nella soluzione prima che la sua norma aumenti e l'angolo nella curva-L sia visibile. Sfortunatamente questo comportamento dei coefficienti SVD è spesso osservato nei problemi di larga scala.

Un'ultima criticità di questo criterio è che se il rumore va a zero, il parametro di regolarizzazione λ_L diverge da quello ottimale e, quindi, x_λ potrebbe non convergere a x^{exact} quando $\|e\|_2 \rightarrow 0$. Fortunatamente, nella pratica molto raramente il rumore tende a zero.

4.4.3 Convalida Incrociata Generalizzata (GCV)

Questo metodo rappresenta un approccio diverso per la scelta del parametro di regolarizzazione in quanto l'obiettivo in questo caso è quello di trovare il valore di λ o k tale per cui Ax_λ o Ax_k preveda il dato *esatto* b^{exact} nel miglior modo possibile.

Per la derivazione del metodo è più semplice considerare il caso della TSVD, in cui la matrice di filtro è $\varphi_i^{[\lambda]} = \text{diag}(1, \dots, 1, 0, \dots, 0)$. La differenza tra b^{exact} e il predittore Ax_k è data da

$$Ax_k - b^{exact} = AV\phi^{[k]}\Sigma^{-1}U^T(b^{exact} + e) - b^{exact} \quad (4.96)$$

Utilizzando l'analisi SVD, la norma dell'errore di previsione è:

$$\|Ax_k - b^{exact}\|_2^2 = \sum_{i=1}^k (u_i^T e)^2 + \sum_{i=k+1}^n (u_i^T b^{exact})^2 \approx k\eta^2 + \sum_{i=k+1}^n (u_i^T b^{exact})^2 \quad (4.97)$$

Ricordando che $|u_i^T b^{exact}| < \eta$ per $i > k_\eta$,

$$k < k_\eta : \quad \|Ax_k - b^{exact}\|_2^2 \approx k\eta^2 + \sum_{i=k+1}^n (u_i^T b^{exact})^2 \quad (4.98)$$

$$k > k_\eta : \quad \|Ax_k - b^{exact}\|_2^2 \approx k\eta^2 \quad (4.99)$$

L'analisi mostra che per $k < k_\eta$ la norma dell'errore di previsione decresce all'aumentare di k (perché il termine $u_k^T b^{exact}$ viene sostituito con η , che è più piccolo, quando si passa da k a $k+1$), mentre per $k > k_\eta$ la norma aumenta con k . Il minimo si ha vicino alla transizione, ovvero per $k \approx k_\eta$. Quindi è una buona idea cercare il parametro di regolarizzazione che minimizza l'errore di previsione.

A tale scopo, la convalida incrociata (cross validation) è una tecnica statistica utile per questo obiettivo. Nella convalida incrociata, i dati (nel nostro caso gli elementi del termine di destra) vengono separati in due gruppi: un gruppo viene utilizzato per calcolare una soluzione, la quale viene successivamente utilizzata per predire gli elementi nell'altro gruppo. Ad esempio, è possibile togliere b_i , l' i -esimo elemento di b , e calcolare la soluzione di Tikhonov basata sul problema ridotto:

$$x_\lambda^{(i)} = ((A^{(i)})^T A^{(i)} + \lambda^2 I_{n-1})^{-1} (A^{(i)})^T b^{(i)} \quad (4.100)$$

dove $A^{(i)}$ e $b^{(i)}$ sono le versioni ridotte di A e b senza, rispettivamente, l' i -esima riga e l' i -esimo elemento. A questo punto è possibile utilizzare $x_\lambda^{(i)}$ per predire l'elemento $b^{(i)}$ (che era stato eliminato) tramite la riga "mancante" di A , ovvero $A(i, :)x_\lambda^{(i)}$. L'obiettivo è, quindi, scegliere il parametro di regolarizzazione che minimizza l'errore di previsione per tutti gli elementi di b :

$$\min_\lambda \frac{1}{m} \sum_{i=1}^m (A(i, :)x_\lambda^{(i)} - b_i)^2 \quad (4.101)$$

In questo modo, però, sono coinvolti m problemi di Tikhonov differenti. Tuttavia, è possibile dimostrare che è possibile sostituire questo problema di minimizzazione con

$$\min_\lambda \frac{1}{m} \sum_{i=1}^m \left(\frac{A(i, :)x_\lambda - b_i}{1 - h_{ii}} \right)^2 \quad (4.102)$$

dove x_λ è la soluzione di Tikhonov e h_{ii} sono gli elementi diagonali della matrice $A(A^T A + \lambda^2 I)^{-1} A^T$.

Questo problema di minimizzazione è più facile da risolvere in quanto è presente un solo problema di Tikhonov. Sfortunatamente, gli elementi diagonali h_{ii} cambiano se le righe di A vengono scambiate e, quindi, la soluzione dipende dal particolare ordine dei dati, ovvero gli elementi di b .

Il *metodo della convalida incrociata generalizzata* (GCV) è stato introdotto per rimediare a questo inconveniente e consiste nella sostituzione di ogni elemento diagonale h_{ii} con la media degli elementi diagonali. Quindi, il problema di minimizzazione semplificato ha la forma

$$\min_{\lambda} \frac{1}{m} \sum_{i=1}^m \left(\frac{A(i, :)x_{\lambda} - b_i}{1 - \text{trace}(A(A^T A + \lambda^2 I)^{-1} A^T)/m} \right)^2 \quad (4.103)$$

in cui la somma a numeratore è la norma del residuo al quadrato $\|Ax_{\lambda} - b\|_2^2$, mentre la traccia della matrice, inserendo la SVD di A, può essere scritta come

$$\text{trace}(A(A^T A + \lambda^2 I)^{-1} A^T) = \text{trace}(U\phi^{[\lambda]}U^T) = \text{trace}(\phi^{[\lambda]}) = \sum_{i=1}^n \varphi_i^{[\lambda]} \quad (4.104)$$

Regolarizzazione di Tikhonov

$$\lambda = \lambda_{GCV} \text{ che minimizza } G(\lambda) = \frac{\|Ax_{\lambda} - b\|_2^2}{(m - \sum_{i=1}^n \varphi_i^{[\lambda]})^2} \quad (4.105)$$

Per il metodo TSVD, la traccia è molto più semplice, in quanto $\text{trace}(\phi^{[K]}) = k$.

TSVD

$$k = k_{GCV} \text{ che minimizza } G(K) = \frac{\|Ax_k - b\|_2^2}{(m - k)^2} \quad (4.106)$$

L'esperienza mostra che questo metodo è robusto e accurato se il rumore è bianco (in quanto questa assunzione è alla base della derivazione della funzione GCV).

Nella figura 4.5 sono rappresentate due funzioni $G(\lambda)$ per la regolarizzazione di Tikhonov.

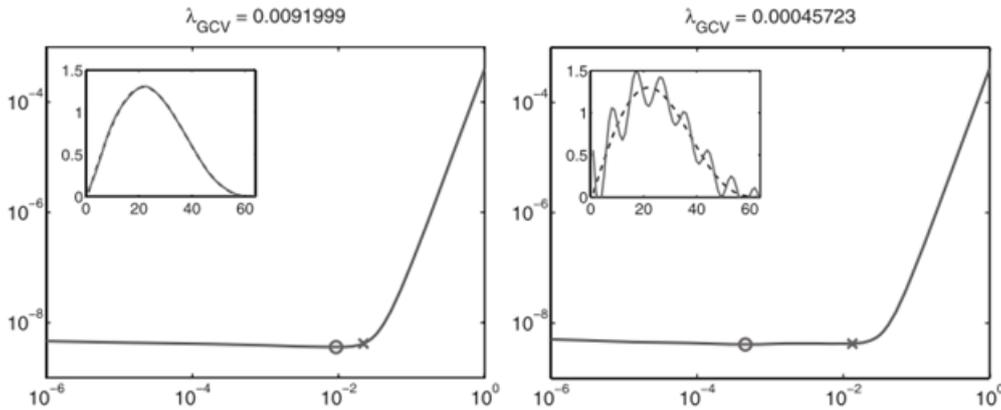


Figura 4.5: Due esempi di soluzione regolarizzata [8]

Il diagramma di sinistra mostra il tipico comportamento della GCV quando il metodo funziona bene e produce un buon parametro di regolarizzazione λ_{GCV} vicino a quello ottimale (che è vicino al punto di transizione tra la parte “piatta” e “ripida” della funzione). Il diagramma a destra mostra un esempio in cui il metodo non produce un buon risultato e ciò accade tutte le volte in cui il minimo si trova sulla parte “piatta” della curva, generando un λ_{GCV} troppo piccolo.

4.4.4 Analisi NCP

Per comprendere al meglio questo criterio è necessario ricordare che quando la soluzione è sovraregolarizzata (poche componenti SVD sono incluse) allora il vettore residuo è dominato dalle componenti SVD dei dati esatti b^{exact} e quando la soluzione è sotto-regolarizzata (troppe componenti SVD sono incluse) allora il vettore residuo è dominato dalle componenti SVD del rumore e.

Idealmente, per scegliere il parametro di troncamento nella TSVD bisognerebbe monitorare il diagramma di Picard e selezionare il parametro di regolarizzazione tale per cui tutti i coefficienti $|u_i^T b|$ sopra il livello di rumore η vengano inclusi nella soluzione regolarizzata. In pratica, bisognerebbe scegliere il parametro k in corrispondenza del quale i coefficienti rumorosi $u_i^T b$ cambiano comportamento e si stabilizzano ad un livello invece che decrescere.

L'obiettivo è quello di implementare ciò senza calcolare la SVD o utilizzare il grafico di Picard. L'approccio consiste nel vedere il vettore residuo come una serie temporale e considerare il termine di destra esatto b^{exact} come un "segnale" che appare molto differente dal vettore rumore e. L'obiettivo è quello di trovare il parametro di regolarizzazione per cui il residuo passa dall'essere simile ad un segnale (dominato dalle componenti di b^{exact}) all'essere simile ad un rumore (dominato dalle componenti di e).

Nello specifico, nel metodo TSVD il vettore residuo è dato da $r_k = \sum_{i=k+1}^n u_i^T b u_i$. Quando k è troppo piccolo, il vettore residuo r_k include sia componenti di segnale $u_i^T b \approx u_i^T b^{exact}$ per i nell'intervallo $k < i < k_\eta$ sia componenti di rumore $u_i^T b \approx \pm \eta$ (per $k_\eta < i \leq n$). Quando k è troppo grande, r_k include solo componenti rumorose $u_i^T b \approx \pm \eta$ (per tutti i valori di i nell'intervallo $k < i \leq n$). Quindi l'obiettivo è scegliere k_η come il più piccolo k per cui r_{k+1} appare statisticamente come rumore.

Lo strumento con cui viene effettuato ciò è detto *periodogramma cumulativo normalizzato* (*normalized cumulative periodogram, NCP*).

Si definisce $p_\lambda \in \mathfrak{R}^{m/2}$ lo spettro di potenza del vettore residuo di Tikhonov r_λ , con elementi

$$(p_\lambda)_k = |\text{dft}(Ax_\lambda - b)_k|^2 \quad k = 1, 2, \dots, m/2 \quad (4.107)$$

dove con dft si intende la trasformata di Fourier discreta. Gli elementi dello spettro di potenza rappresentano la potenza nel segnale associata ad ogni componente spettrale di base e disposti in funzione della loro frequenza, ovvero la prima componente $(p_\lambda)_1$ rappresenta il segnale costante e l'ultima componente rappresenta la frequenza più alta. Allora, il vettore $c(r_\lambda) \in \mathfrak{R}^{\frac{m}{2}-1}$ i cui elementi contengono le somme cumulative dello spettro di potenza

$$c(r_\lambda)_k = \frac{\|p_\lambda(2 : k + 1)\|_1}{\|p_\lambda(2 : m)\|_1} \quad k = 1, \dots, \frac{m}{2} - 1 \quad (4.108)$$

è detto *periodogramma cumulativo normalizzato (NCP)* per il vettore residuo.

La normalizzazione garantisce che il più grande elemento è 1 e la prima componente di p_λ non è inclusa nel calcolo del NCP. La stessa definizione è valida per il NCP $c(r_\lambda)$ associato al vettore residuo per il metodo TSVD, con parametro di troncamento k .

Se il vettore r_λ è costituito da rumore bianco, lo spettro di potenza atteso è piatto e i punti della curva NCP giacciono su una linea retta da $(0, 0)$ a $(m/2-1, 1)$. Il rumore reale non ha uno spettro ideale piatto, ma la curva NCP ha comunque un andamento circa lineare. In termini statistici, la curva NCP deve trovarsi all'interno dei limiti di Kolmogorov-Smirnoff $\pm 1.35q^{-1/2}$ della linea retta, dove $q=m/2-1$, con un livello di significatività del 5%.

Se il residuo, invece, non è costituito da rumore bianco allora la curva NCP non giace all'interno di questi limiti. In particolare, se il residuo è dominato da rumore a bassa o ad alta frequenza allora le curve NCP hanno il seguente aspetto:

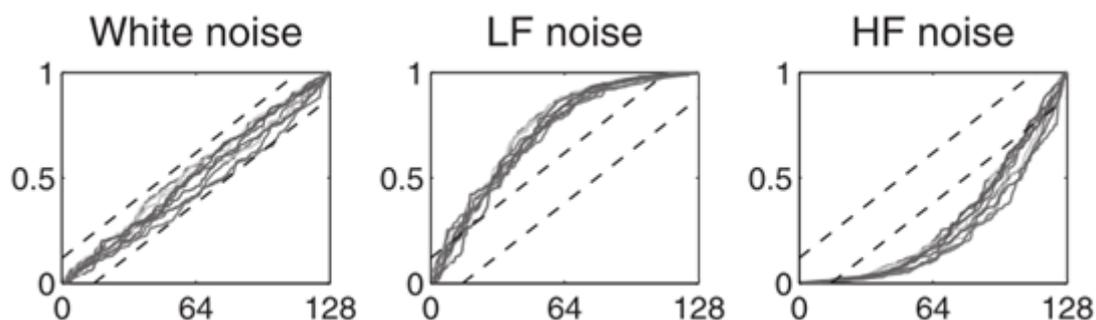


Figura 4.6: Andamenti delle curve NCP [8]

L'obiettivo del metodo è determinare quel valore del parametro di regolarizzazione per cui il vettore residuo cambia comportamento e passa dall'essere un "segnale" all'essere "rumore". Quindi, se si assume che ci sia rumore bianco nei dati, l'obiettivo è quello di determinare il valore del parametro per cui la curva NCP ha un andamento circa lineare ed è compresa nei limiti di Kolmogorov-Smirnoff.

Nella figura 4.7 è raffigurato un esempio di evoluzione della curva NCP per il metodo di Tikhonov al variare del parametro di regolarizzazione.

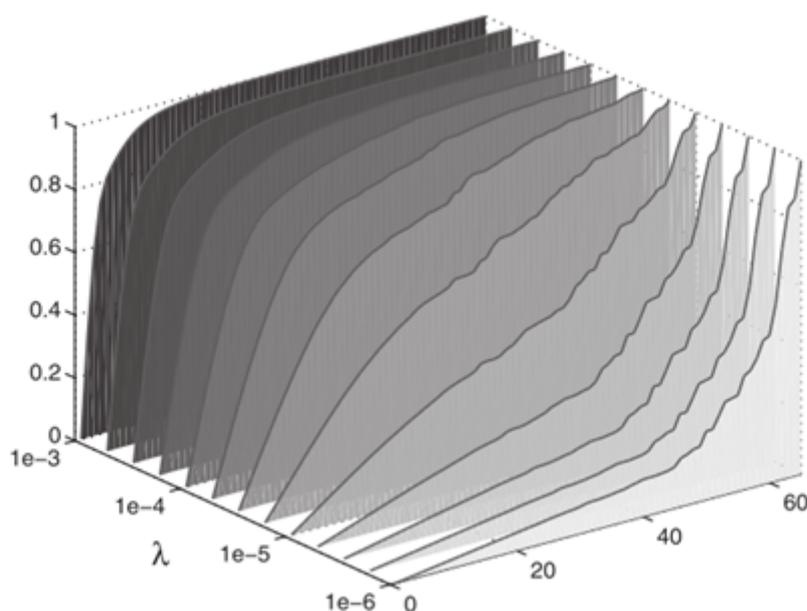


Figura 4.7: Evoluzione della curva al variare del parametro di regolarizzazione [8]

Il parametro di regolarizzazione varia da $\lambda = 10^{-3}$ a $\lambda = 10^{-6}$, ovvero si passa dalla sovra-regolarizzazione alla sotto-regolarizzazione. Tra questi valori, c'è un λ per cui il vettore residuo corrispondente x_λ è costituito il più possibile da rumore bianco e corrisponde alla curva NCP quasi lineare.

Il criterio NCP, quindi, consiste nella scelta del parametro di regolarizzazione in corrispondenza del quale la curva NCP ha un andamento il più simile possibile alla linea retta $c_{white} = \left(\frac{1}{q}, \frac{2}{q}, \dots, 1\right)^T$ e ciò viene misurato tramite la norma-2.

Regolarizzazione di Tikhonov

$$\lambda = \lambda_{NCP} \quad \text{che minimizza} \quad d(\lambda) = \|c(r_\lambda) - c_{white}\|_2 \quad (4.109)$$

TSVD

$$k = k_{NCP} \quad \text{che minimizza} \quad d(k) = \|c(r_k) - c_{white}\|_2 \quad (4.110)$$

Capitolo 5

Regolarizzazione della soluzione su un modello di piastra

Dopo aver introdotto una serie di metodologie per la regolarizzazione della soluzione, in questo Capitolo viene presentata una loro applicazione pratica al problema introdotto nel Capitolo 2. In particolare, poiché il risultato migliore in termini di ricostruzione del carico è stato ottenuto tramite l'approccio inverso e poiché l'accuratezza dell'approccio delle reti neurali è fortemente condizionato dal numero di dati di apprendimento, che in questo studio è limitato per questioni pratiche, i metodi di regolarizzazione sono applicati alla soluzione ottenuta tramite l'approccio inverso.

In questo Capitolo viene mantenuta la configurazione estensimetrica ottimale descritta nel Capitolo 2 e il problema inverso mal posto che si vuole risolvere è descritto dal sistema

$$[S]\{p\} = \{\varepsilon\}, \quad (5.1)$$

dove $[S] \in \mathfrak{R}^{84 \times 84}$ è la matrice ottenuta dall'approccio inverso e che definisce la relazione pressione-deformazione, $\{p\} \in \mathfrak{R}^{44 \times 1}$ è il vettore delle pressioni nodali e costituisce l'incognita del problema e $\{\varepsilon\} \in \mathfrak{R}^{84 \times 1}$ è il vettore delle deformazioni misurate.

Quest'ultimo, in particolare, è caratterizzato dalla presenza di un errore, il quale rappresenta l'errore di misura delle deformazioni che si verifica sperimentalmente.

Il vettore perturbato viene ottenuto a partire da quello ideale attraverso la seguente relazione:

$$\{\varepsilon\} = \overline{\{\varepsilon\}}(1 + \varphi e) \quad (5.2)$$

dove:

- $\{\varepsilon\}$ è il vettore delle deformazioni misurate sperimentalmente e contenente l'errore;
- $\overline{\{\varepsilon\}}$ è il vettore delle deformazioni ideali, ovvero prive di errore;
- $\varphi = 0.05$ è il parametro che descrive l'intensità dell'errore;

- e è un vettore di valori random compresi tra -1 e 1 appartenenti ad una distribuzione normale con valor medio nullo $\mu = 0$ e deviazione standard $\eta = 1/3$, in modo che il 99.7% dei valori estratti da questa distribuzione appartenga al range di valori desiderato.

5.1 Soluzione non regolarizzata

Dopo aver introdotto il vettore delle deformazioni perturbato, sono stati riapplicati i due metodi di ricostruzione del carico con i seguenti risultati:

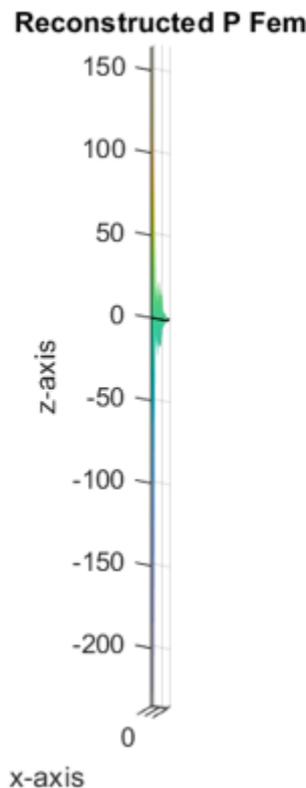


Figura 5.1: Soluzione non regolarizzata ottenuta tramite l'approccio inverso

In particolare, tramite l'approccio inverso è stato ottenuto un **RMSE%** pari a **4863.32%**, mentre con l'approccio delle reti neurali l'errore risulta essere del **125.87%**. Si può notare, quindi, come le reti neurali rappresentino un potente strumento per la risoluzione del problema e come mostrino una maggiore robustezza all'errore rispetto al metodo inverso, nonostante il set di dati di apprendimento sia lo stesso utilizzato nel Capitolo 2 e che, quindi, non contenga dati perturbati. Nel seguito, tuttavia, la regolarizzazione verrà unicamente applicata alla soluzione ottenuta tramite l'approccio inverso a causa dell'elevata dispendiosità nella creazione dei dati di apprendimento contenenti errore necessari alla rete neurale.

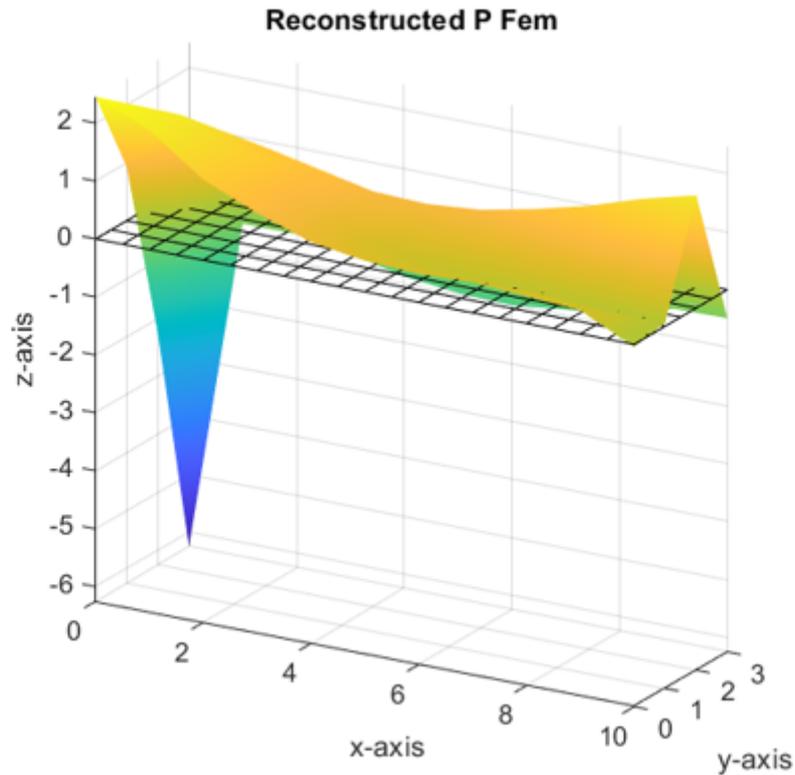


Figura 5.2: Soluzione non regolarizzata ottenuta tramite rete neurale

| Metodo | RMSE% |
|-------------------|----------|
| Approccio inverso | 4863.32% |
| Rete neurale | 125.87% |

Tabella 5.1: Confronto risultati

5.2 Soluzione regolarizzata

In questo paragrafo vengono applicati i metodi di regolarizzazione presentati nel Capitolo 4 alla soluzione ottenuta tramite l'approccio inverso con l'obiettivo di ottenere una migliore approssimazione della soluzione.

5.2.1 Metodo di Tikhonov in forma standard

La prima operazione da compiere per l'applicazione di questo metodo è la scelta del parametro di regolarizzazione. A tale scopo sono stati utilizzati il principio di discrepanza, il metodo della curva L, il metodo della funzione GCV e l'analisi NCP descritti nel Ca-

pitolo 4. Inoltre, grazie al fatto che la distribuzione di riferimento è nota, viene calcolato il parametro di regolarizzazione λ ottimale che minimizza l' RMSE % e in questo modo è possibile determinare il criterio più accurato. I parametri di regolarizzazione ottenuti sono riassunti in tabella 5.2:

| Criterio | Parametro λ |
|--------------------------|---------------------|
| (Valore ottimale) | (3.245E-05) |
| Principio di discrepanza | 4.567E-05 |
| Curva L | 1.881E-05 |
| GCV | 1.509E-05 |
| NCP | 1.339E-06 |

Tabella 5.2: Parametri di regolarizzazione

Successivamente, il metodo di regolarizzazione viene applicato utilizzando i parametri di regolarizzazione appena trovati e le distribuzioni ottenute sono rappresentate nelle figure 5.3, 5.4, 5.5 e 5.6.

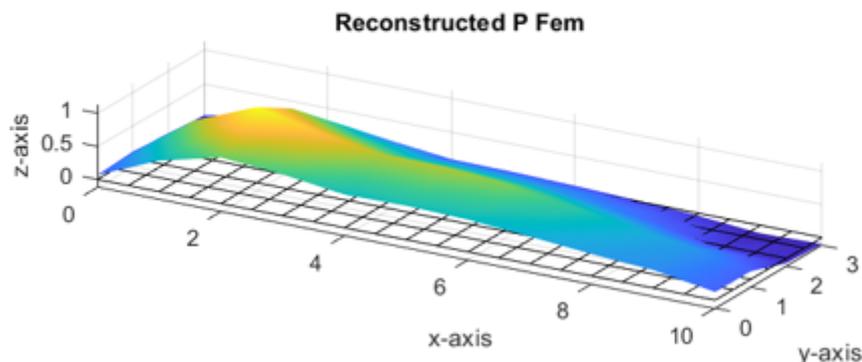


Figura 5.3: Distribuzione ottenuta con λ calcolato tramite principio di discrepanza

Il principio di discrepanza, il metodo della funzione GCV e il metodo della curva L permettono di determinare un parametro di regolarizzazione λ dello stesso ordine di grandezza di quello ottimale, mentre il λ ottenuto tramite l'analisi NCP è di un ordine di grandezza inferiore. Pertanto, come si può notare dalla figura 5.6, la soluzione ottenuta è fortemente sotto-regolarizzata ed è caratterizzata da molte componenti rumorose.

E', inoltre, possibile calcolare i valori dei RMSE % relativi alle soluzioni ottenute tramite i vari criteri e che vengono riassunti in tabella 5.3. Il **minimo** RMSE % è ottenuto prendendo il valore ottimale di λ ed è pari a **25.32** %. Gli errori ottenuti tramite il principio di discrepanza, il metodo della curva L e il metodo GCV sono leggermente superiori ma vicini al valore minimo, mentre quello ottenuto tramite l' analisi NCP è,

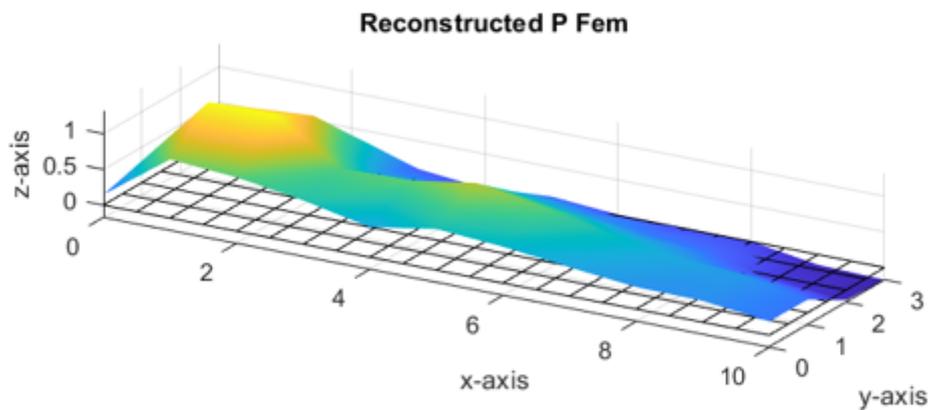


Figura 5.4: Distribuzione ottenuta con λ calcolato tramite metodo della curva L

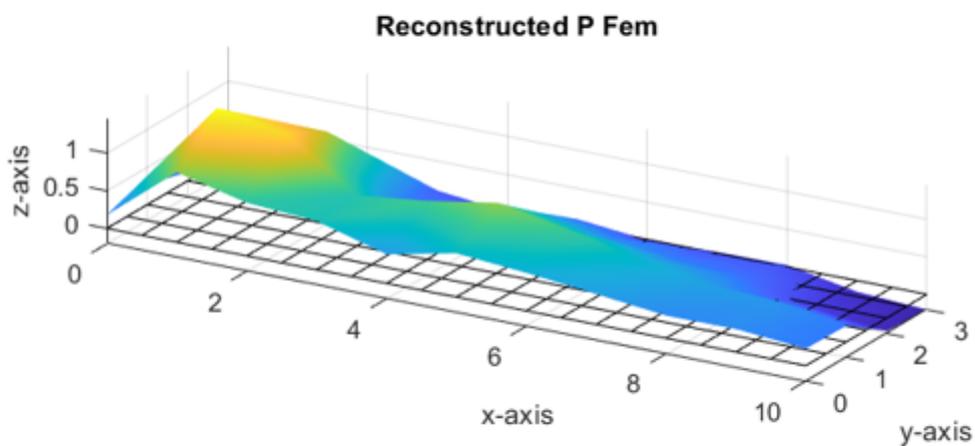


Figura 5.5: Distribuzione ottenuta con λ calcolato tramite metodo GCV

come anticipato, molto più elevato. In particolare, il principio di discrepanza risulta essere il miglior criterio di scelta del parametro di regolarizzazione per questo metodo.

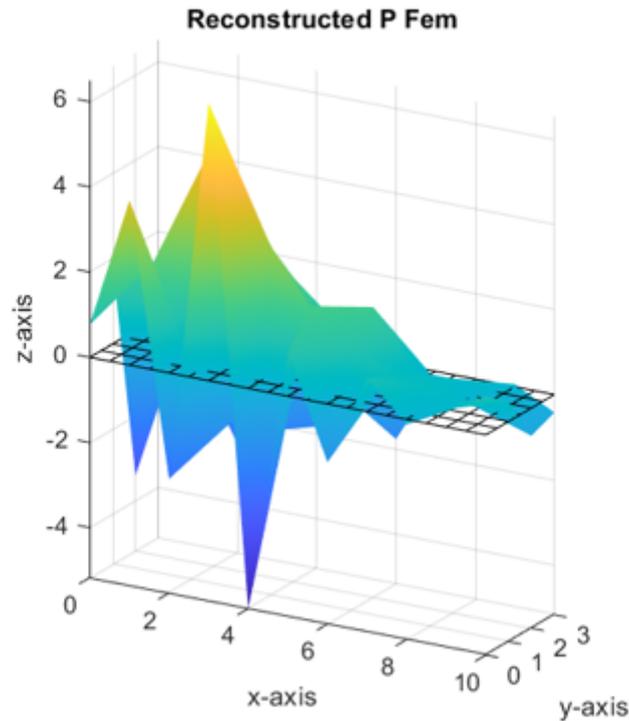


Figura 5.6: Distribuzione ottenuta con λ calcolato tramite analisi NCP

| Criterio | RMSE % |
|--------------------------|-----------|
| (Valore ottimale) | (25.32 %) |
| Principio di discrepanza | 25.68 % |
| Curva L | 26.29 % |
| GCV | 27.29 % |
| NCP | 197.68 % |

Tabella 5.3: RMSE % ottenuti tramite i vari criteri

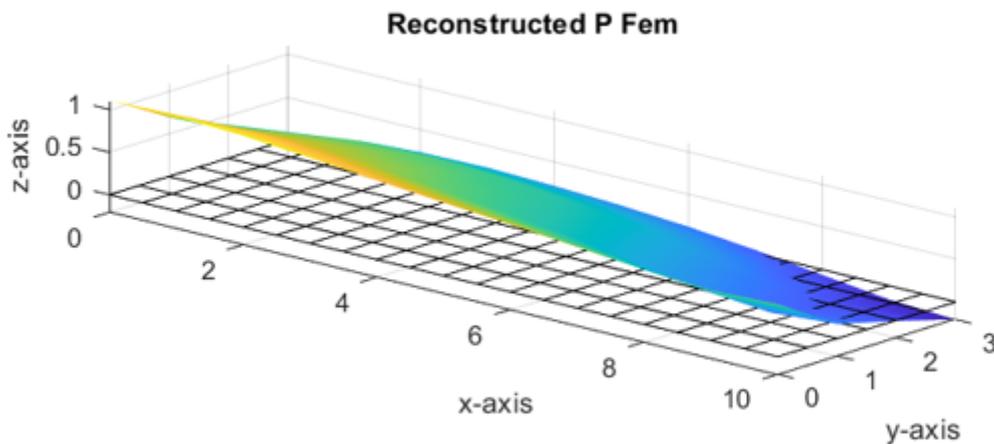
5.2.2 Metodo di Tikhonov del primo ordine

Anche per questo metodo il parametro di regolarizzazione viene calcolato applicando il principio di discrepanza, il metodo della curva L, il metodo GCV e l'analisi NCP; questi valori, inoltre, sono confrontati con il valore ottimale di λ che minimizza il RMSE %.

| Criterio | Parametro λ |
|--------------------------|---------------------|
| (Valore ottimale) | (2.58E-04) |
| Principio di discrepanza | 1.39E-04 |
| Curva L | 1.72E-04 |
| GCV | 2.37E-05 |
| NCP | 2.09E-03 |

Tabella 5.4: Parametri di regolarizzazione

Le distribuzioni di pressioni ottenute con questi valori di λ sono rappresentate nelle figure 5.7, 5.8, 5.9 e 5.10.

**Figura 5.7:** Distribuzione ottenuta con λ calcolato tramite principio di discrepanza

Il principio di discrepanza e il metodo della curva L permettono di ottenere un parametro di regolarizzazione dello stesso ordine di grandezza di quello ottimale, mentre il metodo GCV e l'analisi NCP ottengono un parametro di regolarizzazione di un ordine di grandezza, rispettivamente, più piccolo e più grande.

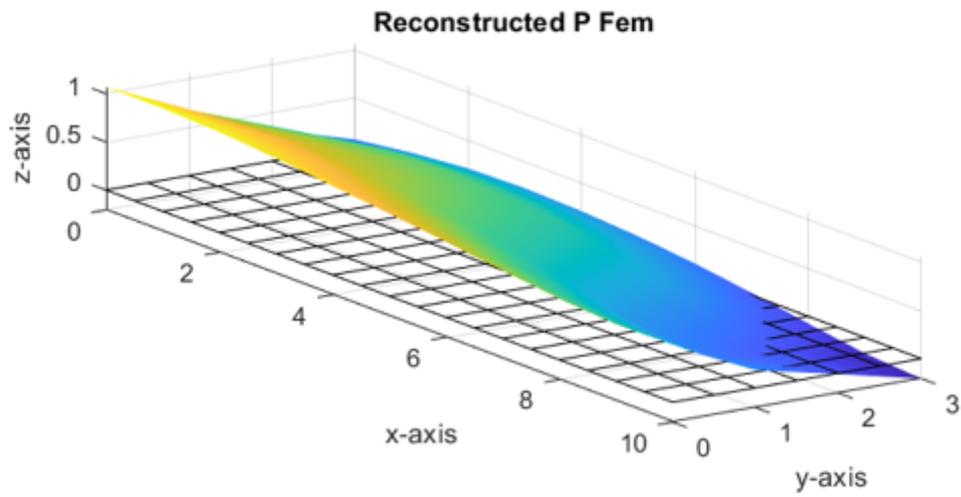


Figura 5.8: Distribuzione ottenuta con λ calcolato tramite metodo della curva L

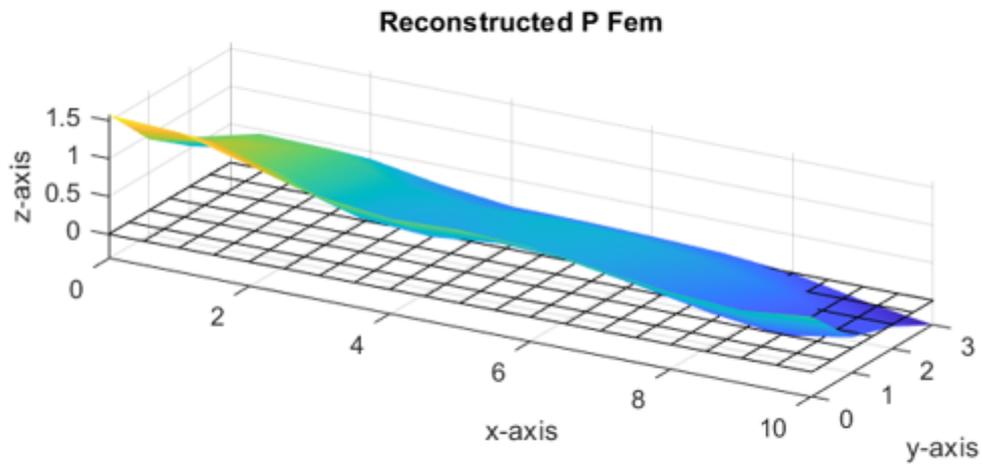


Figura 5.9: Distribuzione ottenuta con λ calcolato tramite metodo GCV

I primi due metodi, quindi, permettono di ottenere una migliore approssimazione della soluzione e ciò è confermato dai valori dei RMSE % relativi ai vari criteri. Il **minimo**

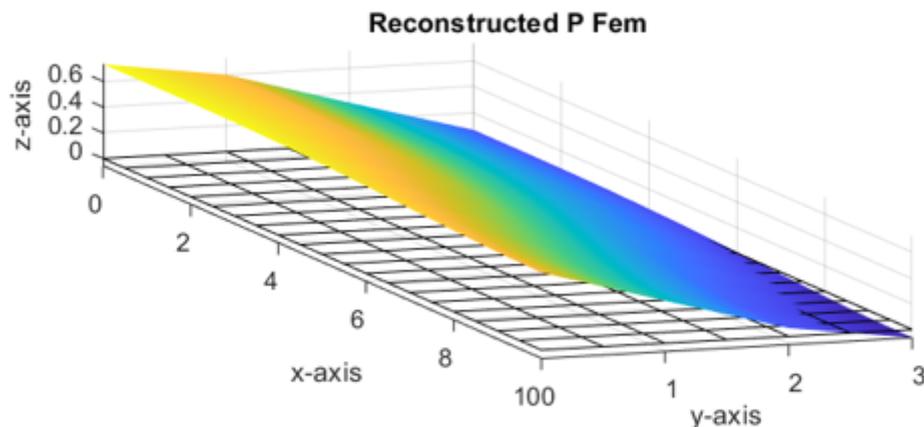


Figura 5.10: Distribuzione ottenuta con λ calcolato tramite analisi NCP

RMSE % ottenuto con il valore ottimale di λ è pari a **16.34 %**, quelli relativi al principio di discrepanza e al metodo della curva L sono molto vicini al valore minimo, mentre gli errori ottenuti con il metodo GCV e l'analisi NCP sono più alti.

| Criterio | RMSE % |
|--------------------------|---------|
| (Valore ottimale) | 16.34 % |
| Principio di discrepanza | 17.17 % |
| Curva L | 16.69 % |
| GCV | 26.65 % |
| NCP | 19.73 % |

Tabella 5.5: RMSE % ottenuti tramite i vari criteri

5.2.3 Metodo di Tikhonov del secondo ordine

In questo metodo la matrice di regolarizzazione L_2 è rappresentata dall'approssimazione discreta dell'operatore di derivata seconda. In analogia con i metodi precedenti, il parametro di regolarizzazione viene calcolato applicando il principio di discrepanza, il metodo della curva L, il metodo GCV e l'analisi NCP e questi valori sono confrontati con quello ottimale che ottimizza il RMSE %. I parametri di regolarizzazione ottenuti sono riassunti in tabella 5.6.

| Criterio | Parametro λ |
|--------------------------|---------------------|
| (Valore ottimale) | (4.63E-03) |
| Principio di discrepanza | 5.04E-04 |
| Curva L | 3.72E-04 |
| GCV | 2.17E-05 |
| NCP | 1.05E-03 |

Tabella 5.6: Parametri di regolarizzazione

Le distribuzioni di pressione ottenute con questi parametri di regolarizzazione sono rappresentate nelle figure 5.11, 5.12, 5.13 e 5.14.

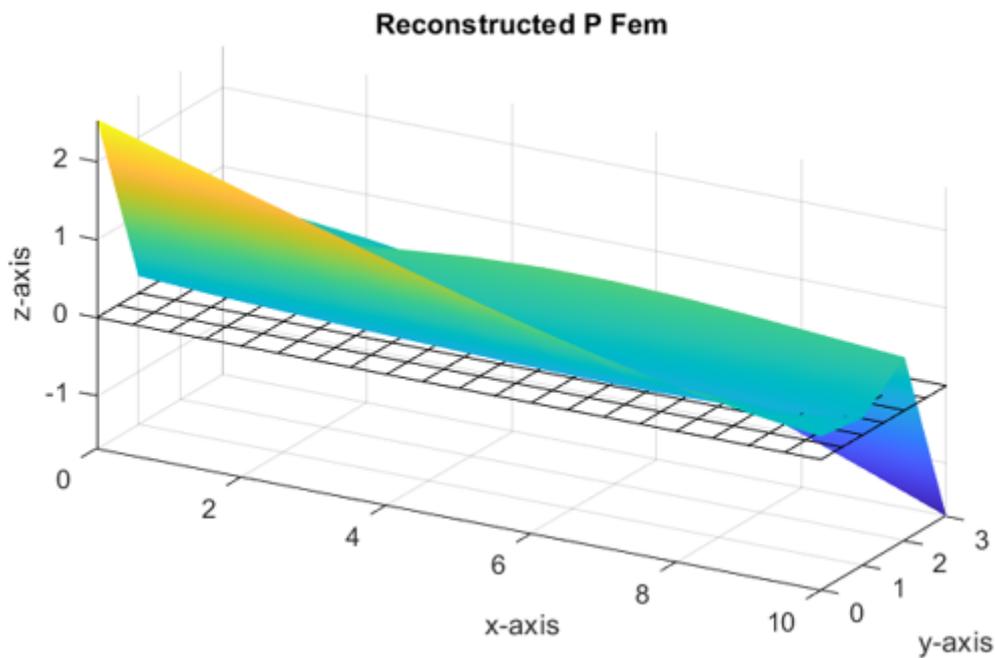


Figura 5.11: Distribuzione ottenuta con λ calcolato tramite principio di discrepanza

Per questo metodo, l'unico criterio in grado di fornire un parametro di regolarizzazione dello stesso ordine di quello ottimale è dato dall'analisi NCP; il principio di discrepanza, il metodo della curva L e il metodo GCV forniscono, invece, parametri di uno o due ordini di grandezza più piccoli.

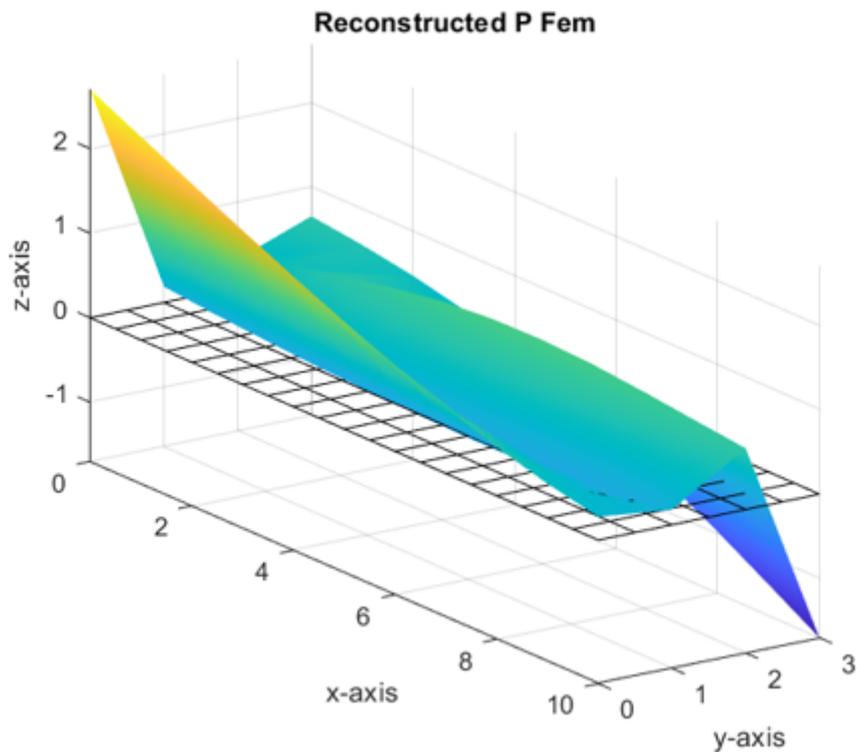


Figura 5.12: Distribuzione ottenuta con λ calcolato tramite metodo della curva L

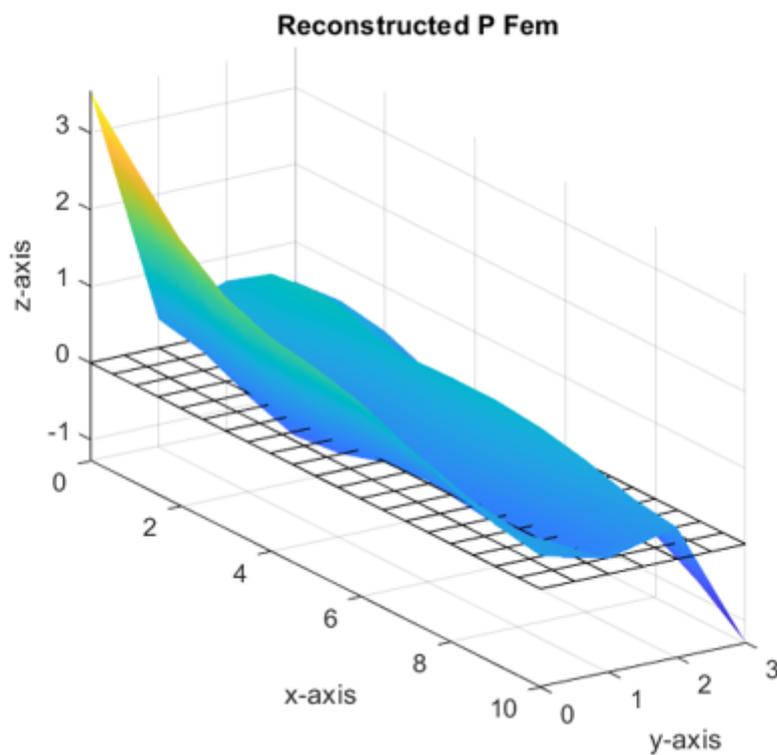


Figura 5.13: Distribuzione ottenuta con λ calcolato tramite metodo GCV

In generale, tuttavia, questo metodo non risulta essere un buon metodo di regolarizzazione per il problema studiato e ciò è evidenziato dai valori dei RMSE %: il **minimo RMSE**

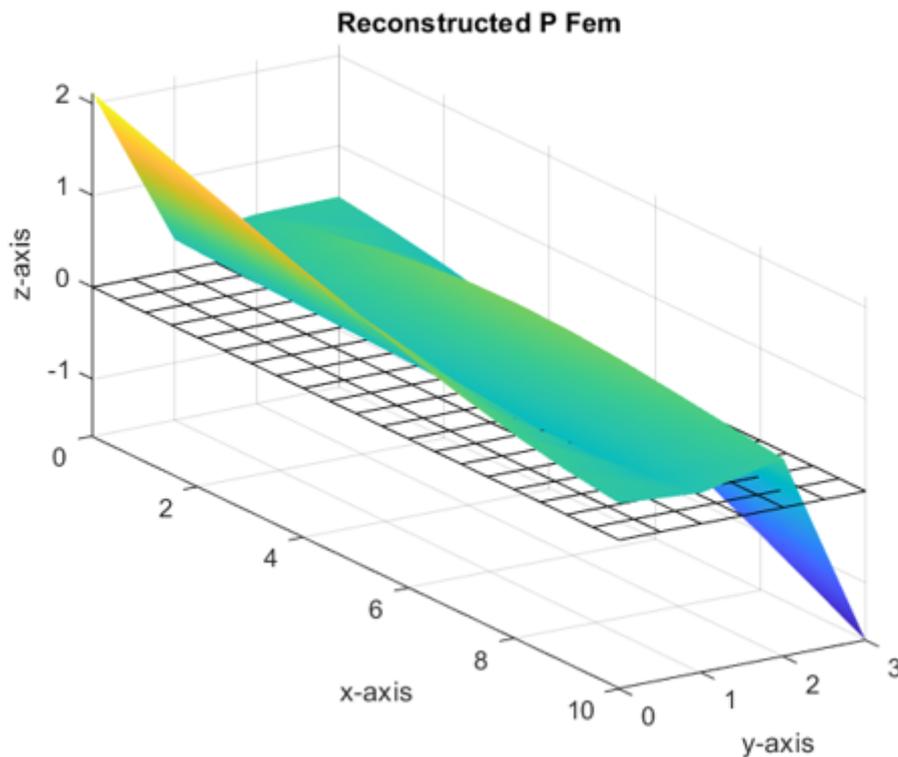


Figura 5.14: Distribuzione ottenuta con λ calcolato tramite analisi NCP

% ottenuto con il valore ottimale di λ è pari a **39.18 %**, decisamente più alto di quelli ottenuti nelle altre analisi. Inoltre, i criteri di scelta del parametro di regolarizzazione non forniscono valori vicini a quello ottimale e, di conseguenza, i RMSE % ottenuti sono ancora superiori, come mostrato nella tabella 5.7.

| Criterio | RMSE % |
|--------------------------|-----------|
| (Valore ottimale) | (39.18 %) |
| Principio di discrepanza | 67.07 % |
| Curva L | 71.02 % |
| GCV | 74.92 % |
| NCP | 56.48 % |

Tabella 5.7: RMSE % ottenuti tramite i vari criteri

5.2.4 Metodo NTR

Per poter applicare i quattro criteri di scelta del parametro di regolarizzazione descritti nel Capitolo 4 e utilizzati nei metodi precedenti è necessario conoscere, oltre alla matrice $[S]$ che descrive il modello e il vettore delle deformazioni misurate $\{\varepsilon\}$, la matrice di regolarizzazione. In questo metodo, tuttavia, la matrice di regolarizzazione è una funzione del parametro di regolarizzazione che deve essere determinato, di conseguenza questi

criteri non possono essere applicati in maniera diretta.

L'autore di questo metodo, in [9], ha, quindi, applicato il principio di discrepanza al problema di Tikhonov in forma standard e ha utilizzato il parametro ottenuto come parametro di regolarizzazione di questo metodo. In questo studio viene seguita la stessa strada e, inoltre, viene calcolato il valore di λ ottimale che minimizza il RMSE %.

| Criterio | Parametro λ |
|--------------------------|---------------------|
| (Valore ottimale) | (5.43E-05) |
| Principio di discrepanza | 4.57E-05 |

Tabella 5.8: Parametro di regolarizzazione

La distribuzione di pressione ottenuta utilizzando il parametro descritto precedentemente è rappresentata in figura 5.15.

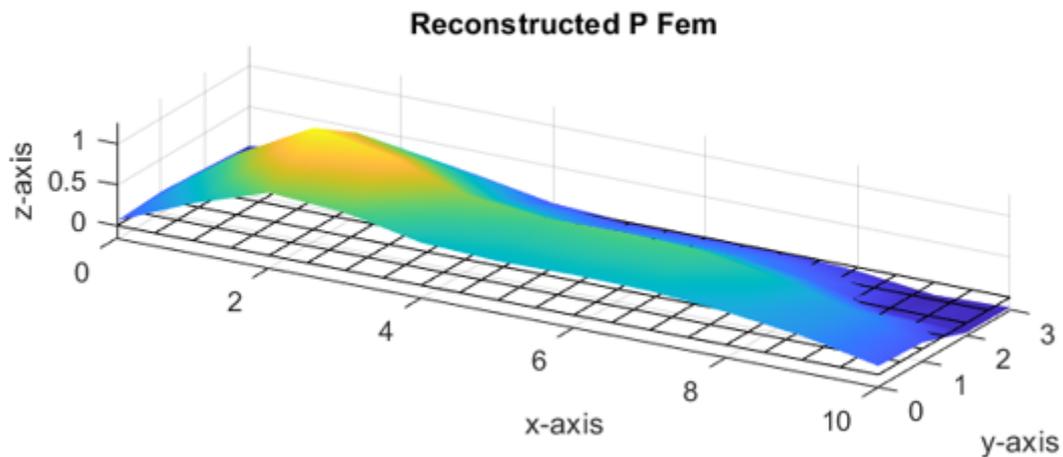


Figura 5.15: Distribuzione ottenuta tramite metodo NTR

Dalla tabella 5.8 si nota che il parametro ottenuto tramite la procedura descritta precedentemente è vicino al valore ottimale e, di conseguenza, il RMSE % relativo a questa distribuzione è molto vicino al valore minimo ottenibile con questo metodo.

| Criterio | RMSE % |
|--------------------------|------------|
| (Valore ottimale) | (26.499 %) |
| Principio di discrepanza | 26.83 % |

Tabella 5.9: Confronto RMSE %

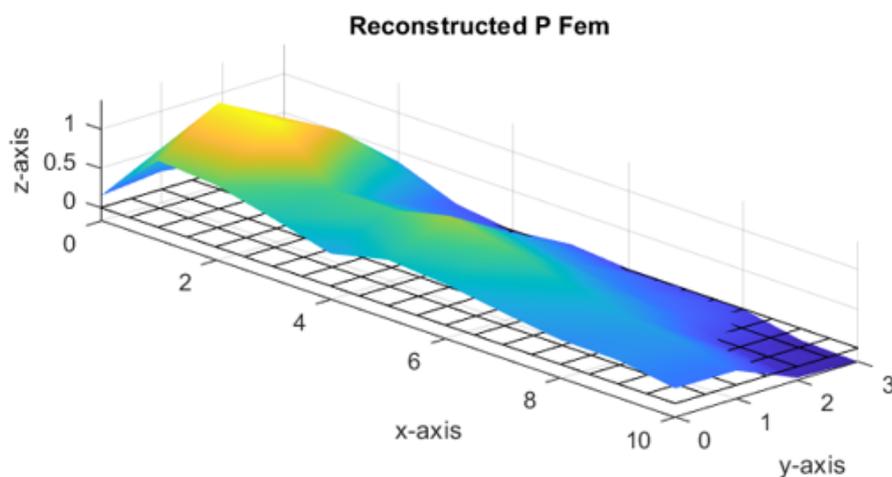
5.2.5 Metodo MTR-1

Anche per questo metodo il parametro di regolarizzazione viene calcolato applicando il principio di discrepanza al problema di Tikhonov in forma standard, mentre il parametro ω viene ottenuto minimizzando la funzione GCV descritta nel paragrafo 4.1.3.

| Criterio | Parametro |
|--------------------------|---------------------------|
| (Valore ottimale) | ($\lambda = 1.256e-04$) |
| Principio di discrepanza | $\lambda = 4.57e-05$ |
| GCV | $\omega = 0.917$ |

Tabella 5.10: Parametri

La distribuzione di pressione ottenuta con i parametri in tabella 5.10 è rappresentata in figura 5.16.

**Figura 5.16:** Distribuzione ottenuta tramite metodo MTR-1

In questo caso il parametro ottenuto tramite il principio di discrepanza è di un ordine di grandezza inferiore rispetto al valore ottimale e, di conseguenza, non risulta essere un buon criterio di scelta. Il **minimo RMSE %** ottenibile con il valore di λ ottimale è, infatti, pari a **25.52 %**, mentre quello ottenuto con questo metodo è più alto:

| Criterio | RMSE % |
|--------------------------|-----------|
| (Valore ottimale) | (25.52 %) |
| Principio di discrepanza | 26.89 % |

Tabella 5.11: Confronto RMSE %

5.2.6 Metodo MTR-2

Analogamente ai due metodi precedenti, anche in questo caso il parametro di regolarizzazione viene ottenuto applicando il principio di discrepanza al problema di Tikhonov in forma standard in quanto la matrice di regolarizzazione è una funzione del parametro e, di conseguenza, i criteri di scelta non sono direttamente applicabili a questo metodo. Inoltre, viene calcolato il valore ottimale del parametro e dal confronto in tabella 5.12 si nota che i due valori sono molto vicini .

| Criterio | Parametro λ |
|--------------------------|---------------------|
| (Valore ottimale) | (5.43E-05) |
| Principio di discrepanza | 4.57E-05 |

Tabella 5.12: Parametro di regolarizzazione

La distribuzione ottenuta con questo parametro è rappresentata in figura 5.17.

In tabella 5.13, invece, vengono confrontati i risultati in termini di RMSE % tra il caso ottimale e quello ottenuto tramite il principio di discrepanza.

In particolare, dalle tabelle 5.12 e 5.13 si nota che i risultati ottenuti sono praticamente identici a quelli del metodo NTR.

| Criterio | RMSE % |
|--------------------------|------------|
| (Valore ottimale) | (26.499 %) |
| Principio di discrepanza | 26.83 % |

Tabella 5.13: Confronto RMSE %

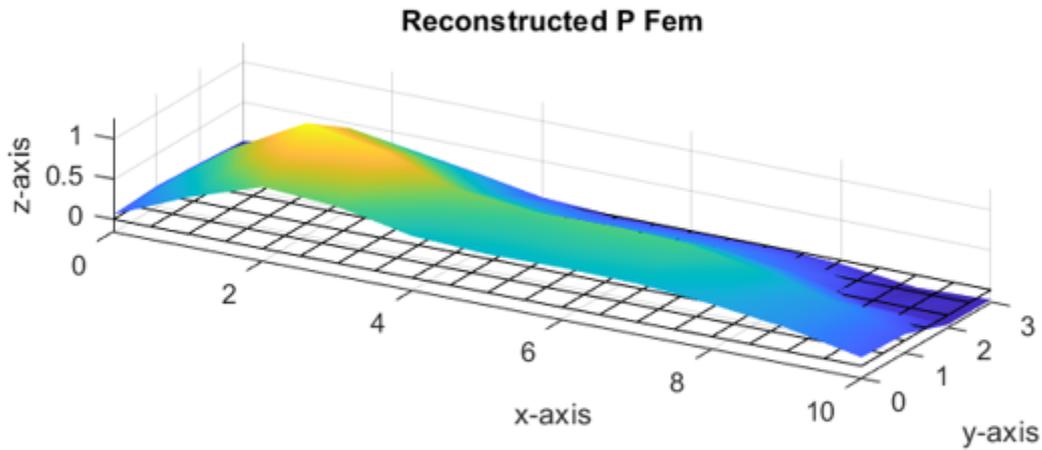


Figura 5.17: Distribuzione ottenuta tramite metodo MTR-2

5.2.7 Metodo ART

Per questo metodo iterativo il parametro di regolarizzazione, ovvero il numero di iterazioni del ciclo, viene determinato tramite il metodo della curva L e il principio di discrepanza; inoltre, esso viene confrontato in tabella 5.14 con il valore ottimale che minimizza il RMSE %.

| Criterio | Parametro k |
|--------------------------|-------------|
| (Valore ottimale) | (218) |
| Principio di discrepanza | 176 |
| Curva L | 155 |

Tabella 5.14: Parametro di regolarizzazione

Le distribuzioni di pressione ottenute con i parametri ricavati dai due metodi sono rappresentate nelle figure 5.18 e 5.19.

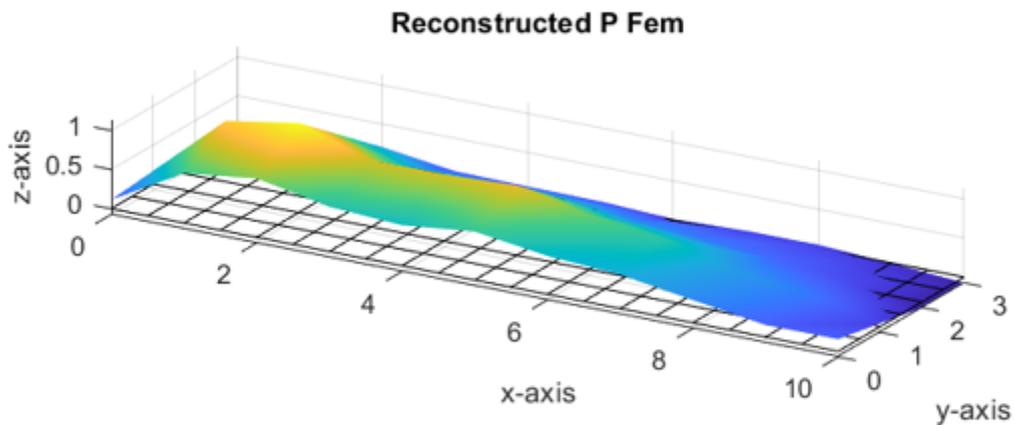


Figura 5.18: Distribuzione ottenuta con k calcolato tramite principio di discrepanza

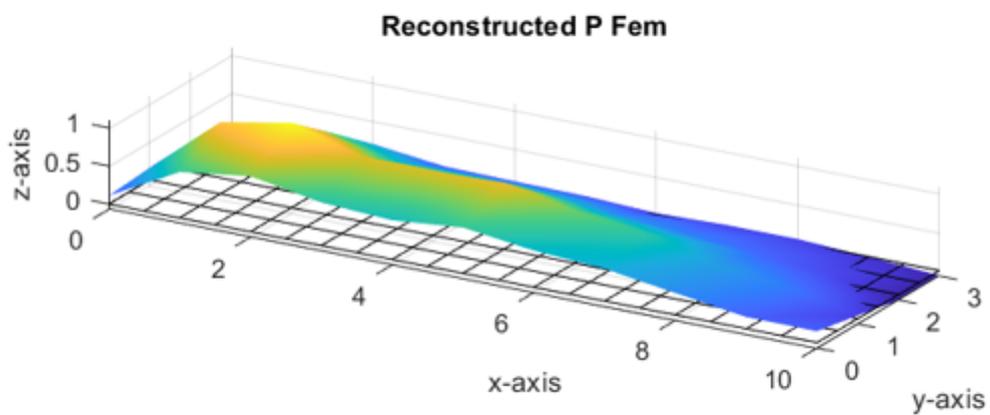


Figura 5.19: Distribuzione ottenuta con k calcolato tramite curva L

Per verificare la bontà delle distribuzioni di pressione appena ricostruite e l' accuratezza

dei criteri di scelta del parametro di regolarizzazione, vengono calcolati i relativi RMSE % e confrontati con il valore minimo ottenuto con il parametro ottimale.

| Criterio | RMSE % |
|--------------------------|-----------|
| (Valore ottimale) | (22.76 %) |
| Principio di discrepanza | 22.89 % |
| Curva L | 23.07 % |

Tabella 5.15: Confronto RMSE %

Come si può notare dalla tabella 5.15, nonostante i valori dei parametri di regolarizzazione ottenuti tramite il principio di discrepanza e il metodo della curva L non siano molto vicini al valore ottimale, i RMSE % ottenuti sono relativamente bassi, in particolare quello ottenuto tramite il principio di discrepanza, e vicini al **minimo RMSE %** pari a **22.76 %**.

5.2.8 Metodo CGLS

Analogamente a ciò che è stato svolto nel metodo ART, anche per questo metodo il parametro di regolarizzazione viene calcolato tramite il principio di discrepanza e il metodo della curva L e, successivamente, confrontato in tabella 5.16 con il valore ottimale che minimizza il RMSE %.

| Criterio | Parametro k |
|--------------------------|-------------|
| (Valore ottimale) | (5) |
| Principio di discrepanza | 4 |
| Curva L | 6 |

Tabella 5.16: Parametro di regolarizzazione

Le distribuzioni di pressione ottenute con i parametri ricavati dai due metodi sono rappresentate nelle figure 5.20 e 5.21.

Per verificare la bontà delle distribuzioni di pressione appena ricostruite e l'accuratezza dei criteri di scelta del parametro di regolarizzazione, vengono calcolati i relativi RMSE % e confrontati con il valore minimo ottenuto con il parametro ottimale.

| Criterio | RMSE % |
|--------------------------|-----------|
| (Valore ottimale) | (27.09 %) |
| Principio di discrepanza | 27.18 % |
| Curva L | 28.52 % |

Tabella 5.17: Confronto RMSE %

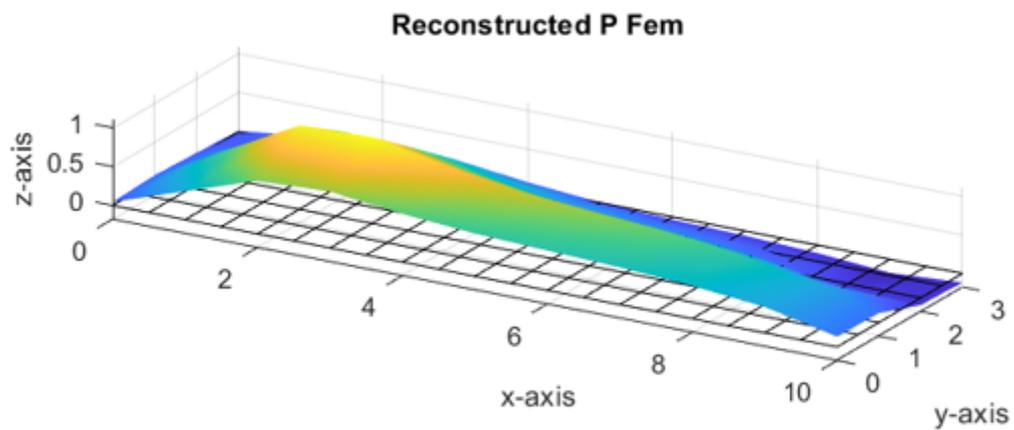


Figura 5.20: Distribuzione ottenuta con k calcolato tramite principio di discrepanza

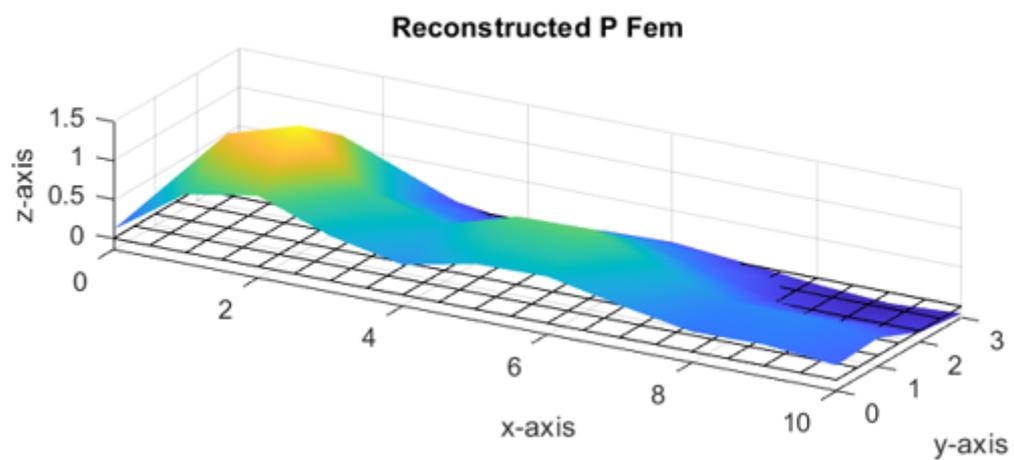


Figura 5.21: Distribuzione ottenuta con k calcolato tramite curva L

Come si può notare dalla tabella 5.17, il parametro di regolarizzazione ottenuto tramite il principio di discrepanza è molto vicino a quello ottimale e il RMSE % corrispondente si avvicina al **minimo RMSE %** pari a **27.09 %**. Nonostante anche il parametro di regolarizzazione determinato dal metodo della curva L sia molto vicino a quello ottimale, il corrispondente RMSE % è più alto.

5.2.9 Metodo M-CGLS

La prima operazione da compiere per l' applicazione di questo metodo è la scelta del numero di passi del processo di bidiagonalizzazione di Lanczos. Per fare ciò si utilizza il metodo introdotto nel paragrafo 4.2.3, ovvero k è il più piccolo intero per cui

$$\bar{\sigma}_k < \tau \bar{\sigma}_1 \quad (5.3)$$

dove $\tau = \sqrt{\text{precisione di macchina}}$ e $\bar{\sigma}_i$ sono i valori singolari della matrice B_k ottenuta dal processo di bidiagonalizzazione.

Dopo aver fatto ciò viene scelto il parametro di regolarizzazione, ovvero il numero di iterazioni del ciclo, attraverso il metodo della curva L e, successivamente, confrontato in tabella 5.18 con il parametro ottimale che minimizza il RMSE %.

| Criterio | Parametro k |
|-------------------|-------------|
| (Valore ottimale) | (3) |
| Curva L | 3 |

Tabella 5.18: Parametro di regolarizzazione

Come si può notare, per questo metodo il criterio della curva L fornisce il parametro ottimale e, di conseguenza, si è deciso di non utilizzare altri criteri per la scelta del parametro.

La distribuzione di pressione ottenuta con i parametri appena definiti è rappresentata in figura 5.22.

In questo caso, quindi, il RMSE % ottenuto tramite questo metodo coincide con il **minimo RMSE %** pari a **28.18 %**.

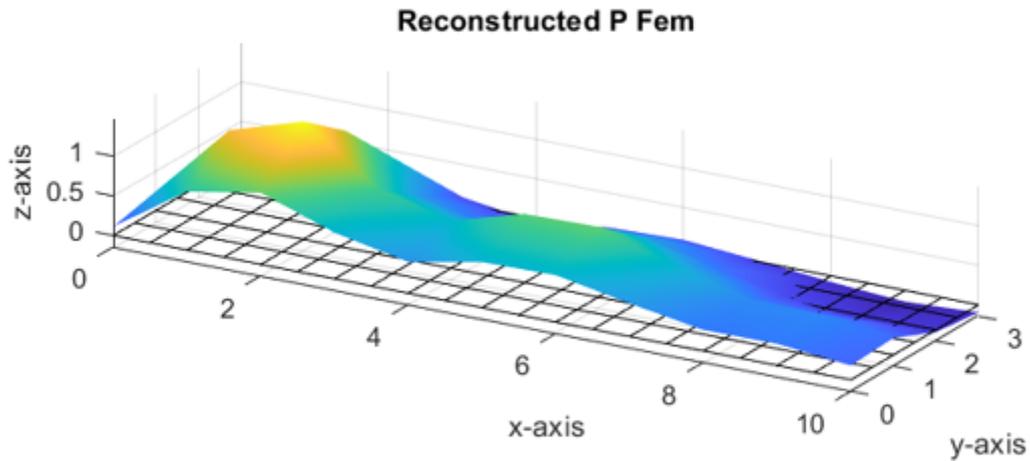


Figura 5.22: Distribuzione ottenuta tramite metodo M-CGLS

5.2.10 Metodo LSQR

Questo metodo risolve il problema ai minimi quadrati tramite il processo di bidiagonalizzazione di Lanczos.

Per determinare il numero k di passi dell'algoritmo di bidiagonalizzazione viene utilizzato il criterio della curva L e, successivamente, confrontato in tabella 5.19 con il numero ottimale che minimizza il RMSE %.

| Criterio | Parametro k |
|-------------------|-------------|
| (Valore ottimale) | (6) |
| Curva L | 7 |

Tabella 5.19: Parametro di regolarizzazione

La distribuzione di pressione ottenuta con il parametro trovato è rappresentata in figura 5.23.

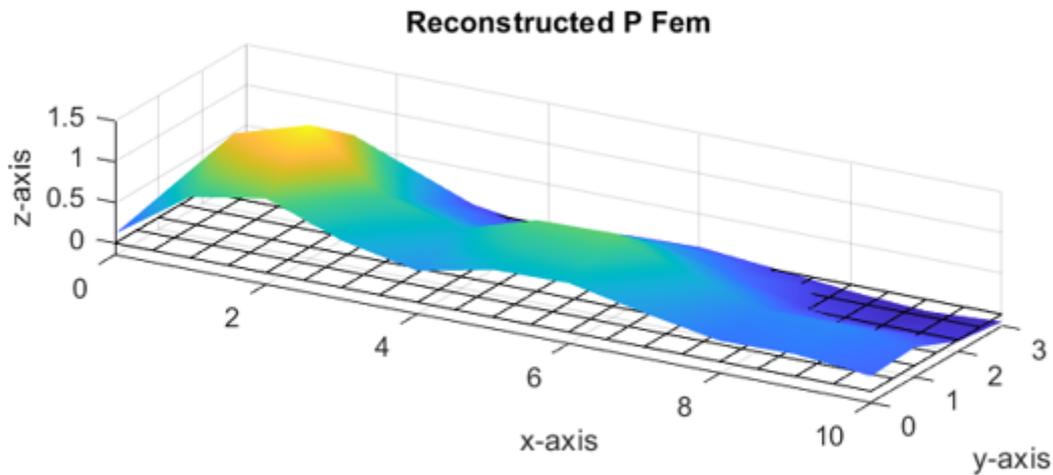


Figura 5.23: Distribuzione ottenuta tramite metodo LSQR

Per verificare la bontà della soluzione appena definita e l'accuratezza del metodo della curva L nella scelta del parametro di regolarizzazione, viene calcolato il RMSE % relativo alla soluzione e confrontato con il valore minimo ottenuto con il parametro ottimale.

| Criterio | RMSE % |
|-------------------|-----------|
| (Valore ottimale) | (26.82 %) |
| Curva L | 28.52 % |

Tabella 5.20: Confronto RMSE %

Come si può notare in tabella 5.20, nonostante il parametro di regolarizzazione trovato tramite il metodo della curva L sia vicino a quello ottimale, il RMSE % della soluzione è più alto del **minimo RMSE %** pari a **26.82 %** di circa il 2 %.

5.2.11 Metodo ibrido

Il metodo ibrido proposto in questo paragrafo consiste nell'applicazione della regolarizzazione di Tikhonov in forma standard al metodo proiettato ottenuto tramite il processo di bidiagonalizzazione di Lanczos. La formulazione teorica è descritta nel paragrafo 4.3, qui l'algoritmo di bidiagonalizzazione viene lanciato per $k = 1, 2, 3, \dots$ e vengono salvate

le matrici B_k e W_k . Per ogni quinta iterazione, viene formato il problema proiettato e viene applicata la regolarizzazione di Tikhonov con un criterio di scelta del parametro per arrivare a $\lambda^{(k)}$. Quando il parametro di regolarizzazione tende a stabilizzarsi e a rimanere costante per diversi valori di k significa che la dimensione del sottospazio di Krylov è abbastanza grande da comprendere tutte le informazioni necessarie e che l'opportuna regolarizzazione è stata applicata. Infine, è possibile calcolare la soluzione tramite la moltiplicazione per W_k .

Questa operazione è stata svolta per i quattro criteri di scelta del parametro di regolarizzazione presentati nel Capitolo 4 con i risultati riassunti in tabella 5.21.

| Criterio | Parametro λ |
|--------------------------|---------------------|
| (Valore ottimale) | (3.245E-05) |
| Principio di discrepanza | 4.567E-05 |
| Curva L | 1.881E-05 |
| GCV | 4.539E-05 |
| NCP | 3.329E-03 |

Tabella 5.21: Parametri di regolarizzazione

Il parametro di regolarizzazione ottimale e quelli ottenuti tramite il metodo della curva L e il principio di discrepanza coincidono con i valori trovati per il metodo di Tikhonov in forma standard, mentre quelli ottenuti tramite i metodo GCV e NCP sono differenti. Le distribuzioni di pressione corrispondenti sono rappresentate nelle figure 5.24, 5.25, 5.26 e 5.27.

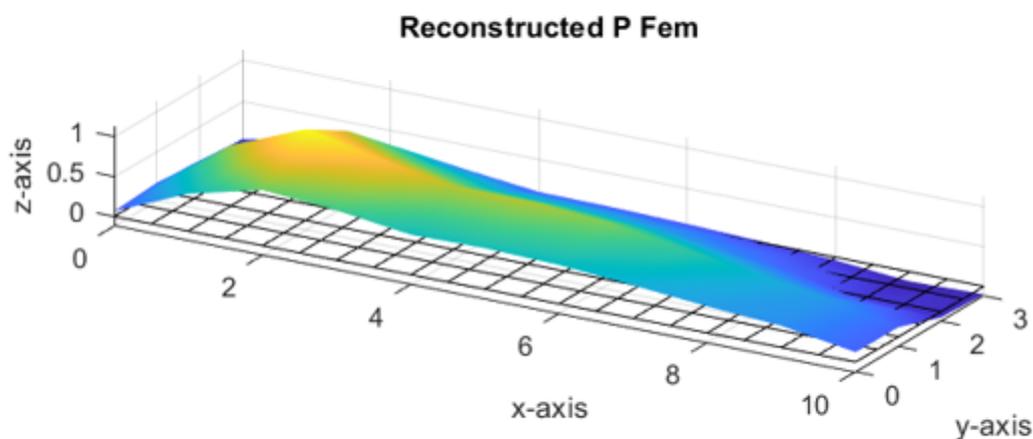


Figura 5.24: Distribuzione ottenuta con λ calcolato tramite principio di discrepanza

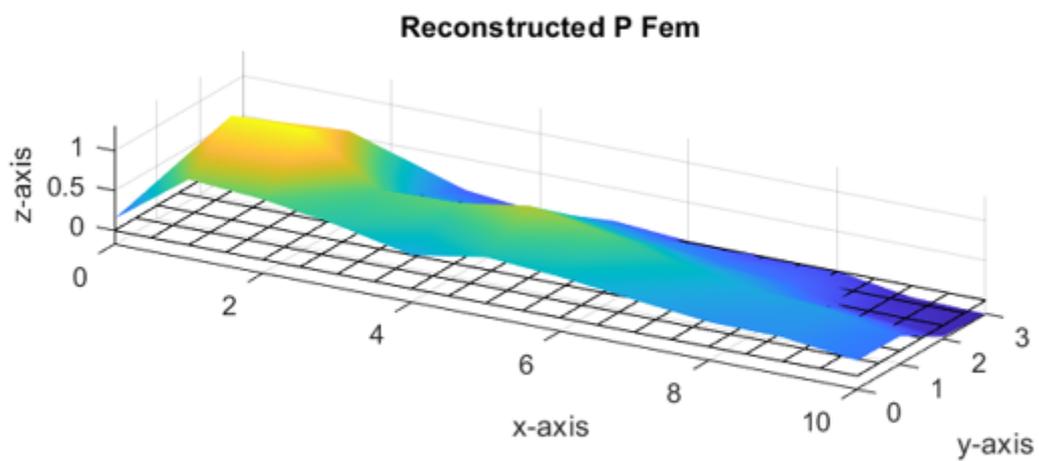


Figura 5.25: Distribuzione ottenuta con λ calcolato tramite metodo della curva L

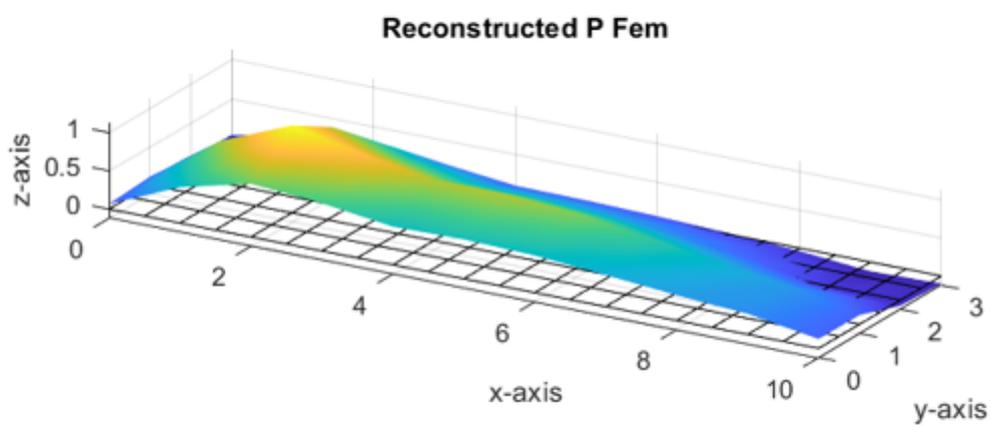


Figura 5.26: Distribuzione ottenuta con λ calcolato tramite metodo GCV

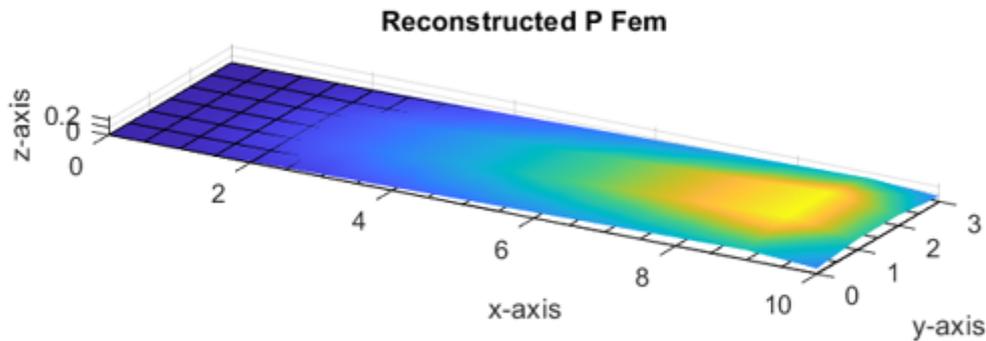


Figura 5.27: Distribuzione ottenuta con λ calcolato tramite analisi NCP

I rispettivi RMSE % vengono riassunti in tabella 5.22 e si può notare che, scegliendo il corretto numero k di passi dell'algoritmo di bidiagonalizzazione di Lanczos, i risultati tendono verso quelli ottenuti con il metodo di Tikhonov in forma standard in quanto quest'ultimo è stato utilizzato come metodo di regolarizzazione del problema proiettato.

| Criterio | RMSE % |
|--------------------------|-----------|
| (Valore ottimale) | (25.32 %) |
| Principio di discrepanza | 25.68 % |
| Curva L | 26.29 % |
| GCV | 25.67 % |
| NCP | 49.04 % |

Tabella 5.22: Confronto RMSE %

5.2.12 Confronto dei risultati

In tabella 5.23 sono, infine, riassunti i risultati ottenuti con questi metodi. Come si può vedere, il miglior risultato è stato ottenuto tramite il Metodo di Tikhonov del primo ordine e il metodo della Curva L come criterio di scelta del parametro di regolarizzazione, con un RMSE % pari a 16.69 %. Altri metodi che hanno fornito ottimi risultati sono il Metodo di Tikhonov in forma standard, il Metodo ART e il Metodo LSQR.

| | Princ. Discrep. | | Curva L | | GCV | | NCP | | OPT | |
|---------------|-------------------|--------|-------------------|--------|-------------------|--------|-------------------|---------|-------------------|---------|
| | $\lambda \circ K$ | RMSE % | $\lambda \circ K$ | RMSE % |
| Tikh | 4.57E - 05 | 25.68% | 1.88E - 05 | 26.29% | 1.51E - 05 | 27.29% | 1.34E - 06 | 197.68% | 3.245E - 05 | 25.32% |
| Tikh-1 | 1.39E - 04 | 17.17% | 1.72E - 04 | 16.69% | 2.37E - 05 | 26.65% | 2.09E - 03 | 19.73% | 2.58E - 04 | 16.34% |
| Tikh-2 | 5.04E - 04 | 67.07% | 3.72E - 04 | 71.02% | 2.17E - 05 | 74.92% | 1.05E - 03 | 56.48% | 4.63E - 03 | 39.18% |
| NTR | 4.57E - 05 | 26.83% | | | | | | | 5.43E - 05 | 26.499% |
| MTR-1 | 4.57E - 05 | 26.89% | | | | | | | 1.256E - 04 | 25.52% |
| MTR-2 | 4.57E - 05 | 26.83% | | | | | | | 5.438E - 05 | 26.499% |
| ART | 176 | 22.89% | 155 | 23.07% | | | | | 218 | 22.76% |
| CGLS | 4 | 27.18% | 6 | 28.52% | | | | | 5 | 27.09% |
| M-CGLS | | | 3 | 28.18% | | | | | 3 | 28.18% |
| LSQR | | | 7 | 28.52% | | | | | 6 | 26.82% |
| Ibrido | 4.567E - 05 | 25.68% | 1.88E - 05 | 26.29% | 4.539E - 05 | 25.67% | 3.329E - 03 | 49.04% | 3.2E - 05 | 25.32% |

Tabella 5.23: Riassunto dei risultati

5.3 Analisi statistica

Le analisi presenti nel paragrafo precedente sono state svolte considerando un unico vettore di perturbazione e uguale per tutti i casi. In questo paragrafo l'obiettivo è verificare che i risultati ottenuti siano rappresentativi dell'effettiva bontà di questi metodi e non dipendenti da quel dato vettore di perturbazione e , quindi, studiare la loro robustezza. Sulla base dei risultati precedenti, vengono selezionati i migliori quattro metodi (due filtranti e due iterativi), ovvero il metodo di Tikhonov in forma standard, il metodo di Tikhonov del primo ordine, il metodo ART e, infine, il metodo LSQR. Per ogni metodo viene svolta un'analisi statistica, cioè la singola operazione di regolarizzazione viene effettuata un numero elevato di volte (2000 in tre casi e 200 in uno) con vettori di perturbazione random e vengono analizzati i valori dei RMSE % e dei parametri di regolarizzazione ottenuti. Per la visualizzazione dei dati raccolti e la valutazione del loro comportamento statistico vengono utilizzati i seguenti strumenti:

- Visualizzazione dei dati:
 - **Istogramma**: rappresentazione grafica in cui il range dei possibili valori viene diviso in intervalli, viene creato un grafico a barre e l'altezza di ogni colonna corrisponde alla frequenza assoluta con cui i valori in quell'intervallo appaiono nei dati;
 - **Box Plot**: rappresentazione grafica di una distribuzione di dati in cui è presente un rettangolo diviso in due parti, da cui escono due segmenti tratteggiati. Il rettangolo è delimitato dal primo e dal terzo quartile ed è diviso al suo interno dalla mediana (secondo quartile). I segmenti delimitano l'intervallo in cui è presente circa il 99 % dei dati. I dati anomali sono rappresentati tramite crocette rosse.

- Misura della centralità:
 - **Valor medio:** media aritmetica dei dati appartenenti alla popolazione studiata

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

- **Mediana:** valore che bipartisce la distribuzione in due sotto-distribuzioni di uguale numerosità.
- Misura della dispersione:

- **Deviazione standard:** radice quadrata della media aritmetica degli scarti assoluti tra i valori della distribuzione e un valore medio

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

- **Scarto interquartile (IQR):** differenza tra il terzo e il primo quartile, ovvero l'ampiezza della regione che contiene la metà "centrale" dei valori osservati.

L'istogramma fornisce un'indicazione qualitativa della forma della distribuzione dei dati. La forma esatta della distribuzione, invece, è data dalla sua funzione di densità di probabilità (PDF). La PDF è una funzione matematica $f(x)$ tale che l'area sottesa dalla curva compresa tra due valori di x è la probabilità che un valore random x si trovi tra i due valori che delimitano l'area. La PDF di una distribuzione discreta coincide con l'istogramma normalizzato e l'area di una colonna rappresenta la probabilità che un dato scelto in maniera random si trovi in quell'intervallo.

Per comprendere al meglio, quindi, la forma della distribuzione dei dati, viene creato un grafico in cui è rappresentata la PDF della distribuzione discreta ottenuta normalizzando l'istogramma corrispondente, a cui si sovrappone la PDF della distribuzione normale. Dal confronto tra i due andamenti è, infatti, possibile avere un'idea più chiara sulla forma della distribuzione di dati studiata.

5.3.1 Metodo di Tikhonov in forma standard

Per lo studio statistico di questo metodo vengono svolte 2000 analisi, ognuna di esse con un diverso vettore di perturbazione (il quale viene creato in maniera random), e il criterio di scelta del parametro di regolarizzazione considerato è il principio di discrepanza in quanto nel paragrafo precedente è stato il metodo che ha fornito il valore più accurato tra i quattro criteri utilizzati. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regolarizzazione λ .

RMSE %

I valori del RMSE % ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

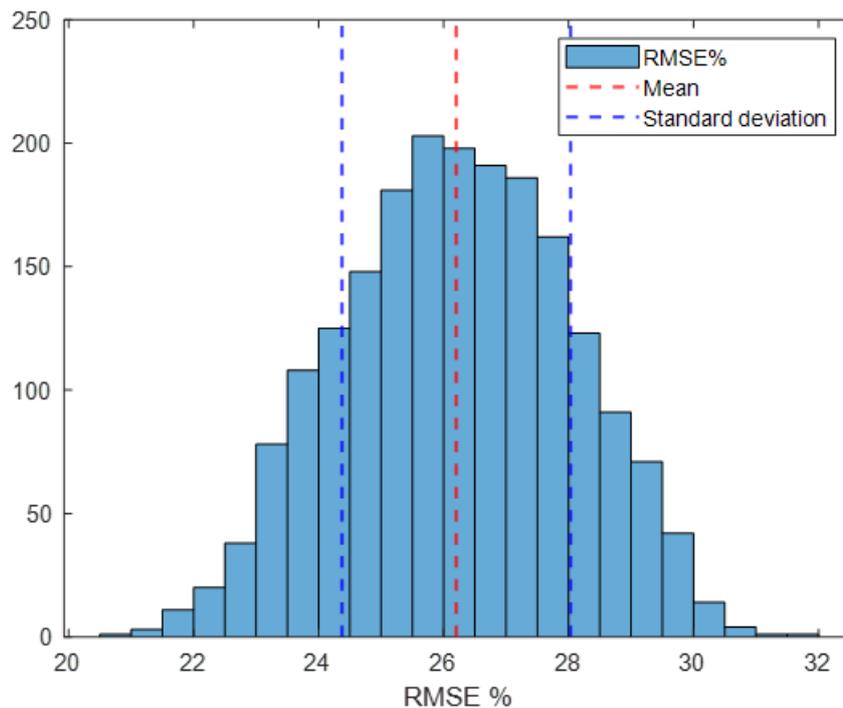


Figura 5.28: Istogramma con valor medio e deviazione standard in evidenza

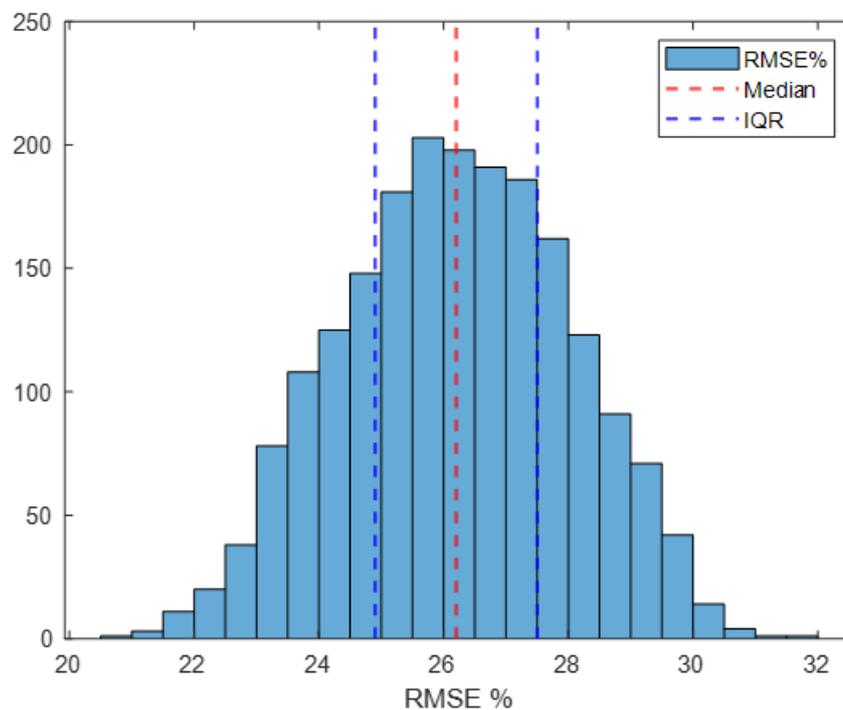


Figura 5.29: Istogramma con mediana e IQR in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.24. Come si può vedere, la media e la mediana hanno valori molto vicini e, di conseguenza, la distribuzione può essere considerata simmetrica. Gli indici di dispersione, invece, sono relativamente piccoli e ciò dimostra la robustezza del metodo alla variabilità della perturbazione. Inol-

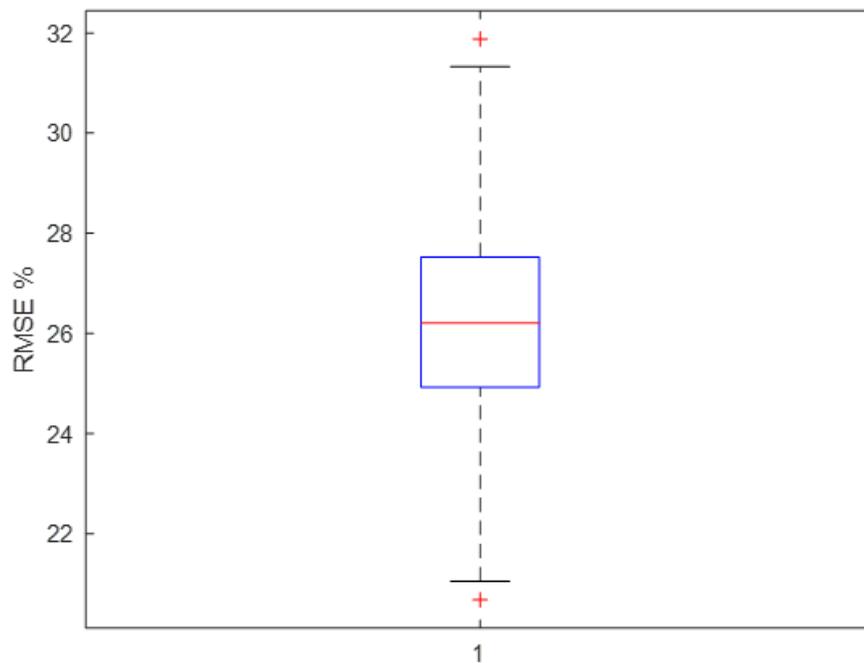


Figura 5.30: Box plot RMSE %

tre, il RMSE % ottenuto nel paragrafo precedente è molto vicino al valor medio e ciò indica che è stato analizzato un caso statisticamente probabile.

| RMSE % | |
|---------------------|----------|
| Valor medio | 26.203 % |
| Mediana | 26.205 % |
| Deviazione standard | 1.830 % |
| IQR | 2.602 % |

Tabella 5.24: Dati statistici del RMSE %

Nella figura 5.31, infine, è rappresentata la funzione di densità di probabilità (PDF) della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

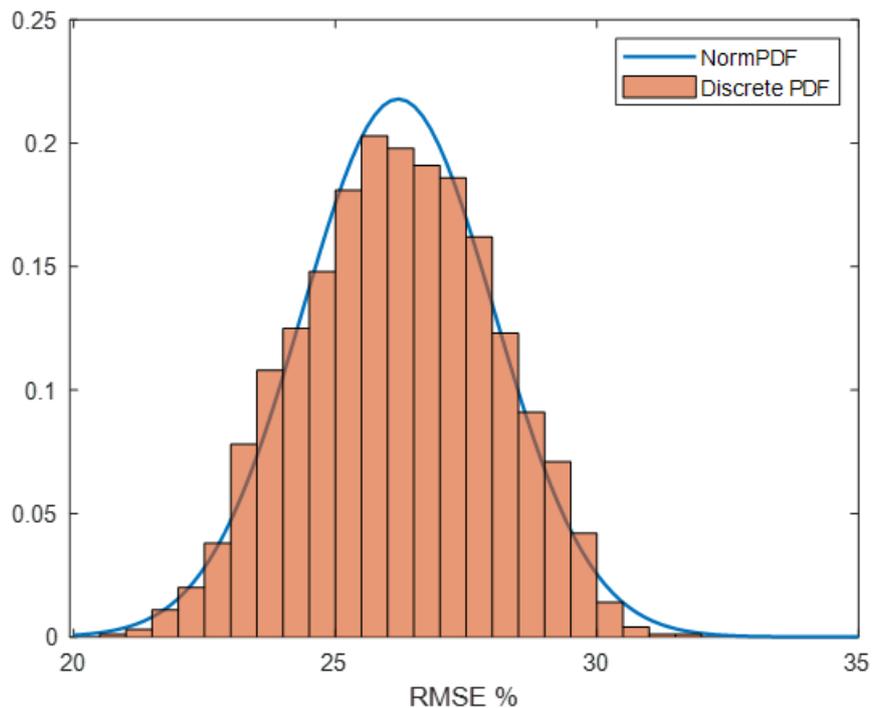


Figura 5.31: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, come la PDF della distribuzione discreta abbia un andamento praticamente coincidente con la PDF della distribuzione normale.

Parametro di regolarizzazione λ

I valori di λ ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

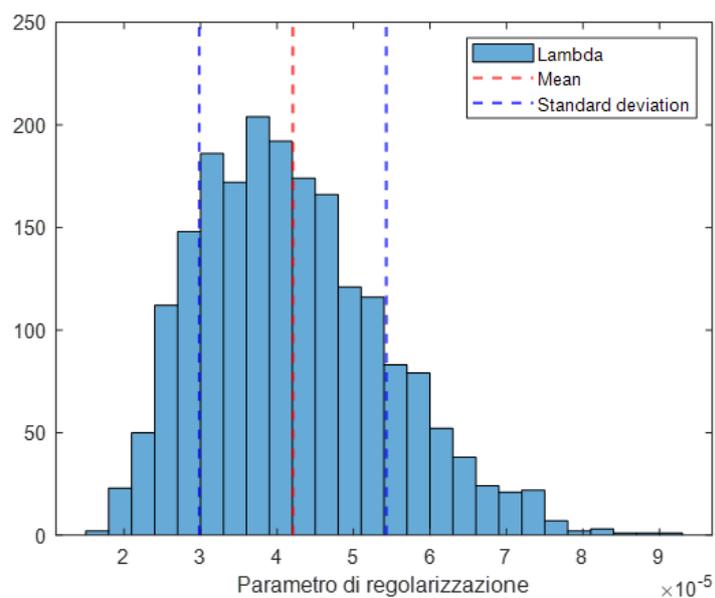


Figura 5.32: Istogramma con valor medio e deviazione standard in evidenza

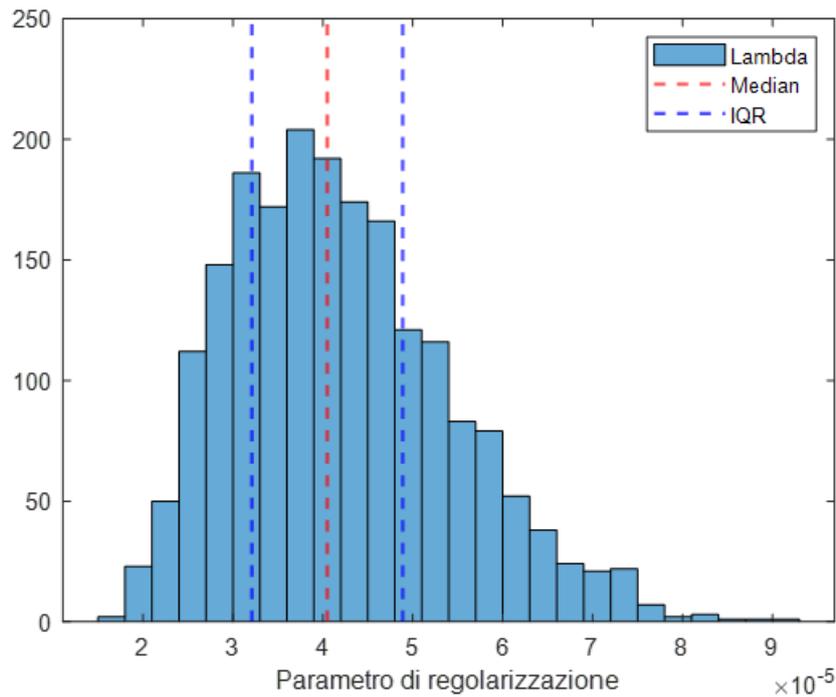


Figura 5.33: Istogramma con mediana e IQR in evidenza

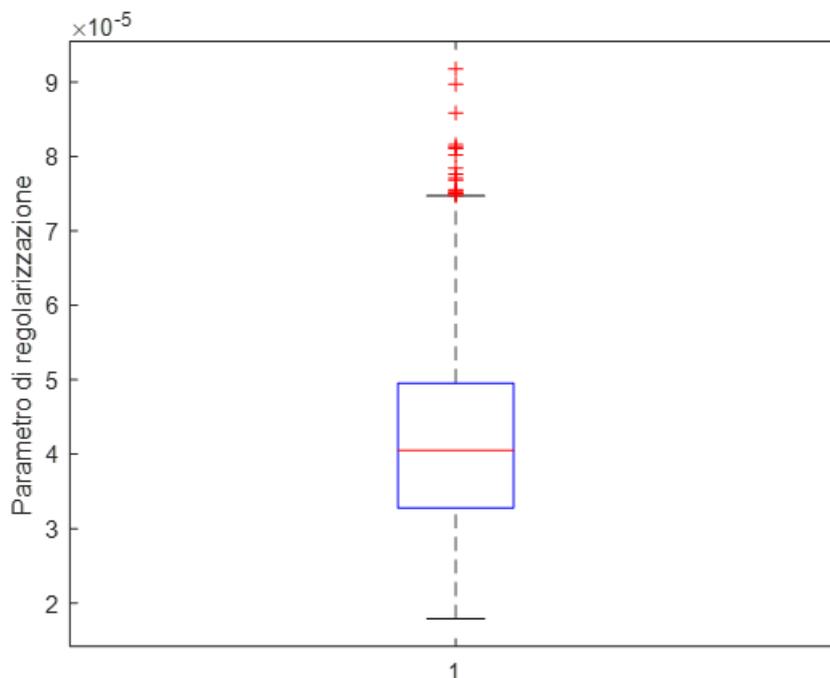


Figura 5.34: Box plot λ

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.25. Anche in questo caso la media e la mediana hanno valori simili tra loro e la distribuzione è quasi simmetrica.

Gli indici di dispersione, inoltre, sono relativamente piccoli e il metodo risulta essere robusto alla variabilità della perturbazione. Come per il RMSE %, il λ calcolato nel

paragrafo precedente è vicino al valor medio, a conferma del fatto che è stato analizzato un caso statisticamente probabile.

| Parametro di regolarizzazione λ | |
|---|----------|
| Valor medio | 4.21E-05 |
| Mediana | 4.05E-05 |
| Deviazione standard | 1.22E-05 |
| IQR | 1.68E-05 |

Tabella 5.25: Dati statistici di λ

Nella figura 5.35, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

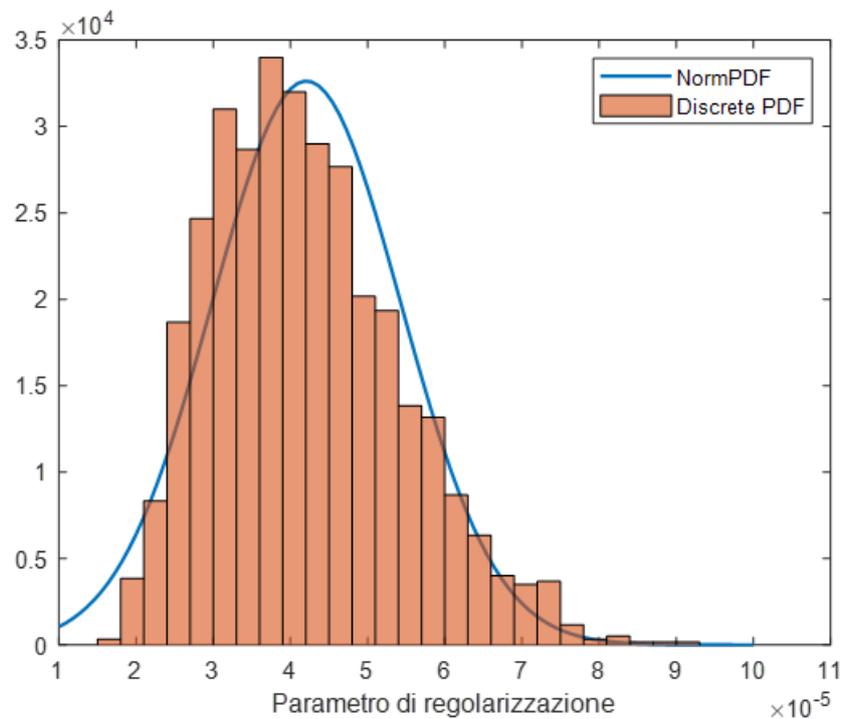


Figura 5.35: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che l'andamento della PDF della distribuzione discreta è circa uguale a quello della PDF della distribuzione normale.

5.3.2 Metodo di Tikhonov del primo ordine

Per lo studio statistico di questo metodo vengono svolte 2000 analisi, ognuna di esse con un diverso vettore di perturbazione e il criterio di scelta del parametro di regolarizzazione considerato è il metodo della curva L in quanto nel paragrafo precedente è stato il metodo che ha fornito il valore più accurato tra i quattro criteri. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regolarizzazione λ .

RMSE %

I valori del RMSE % ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

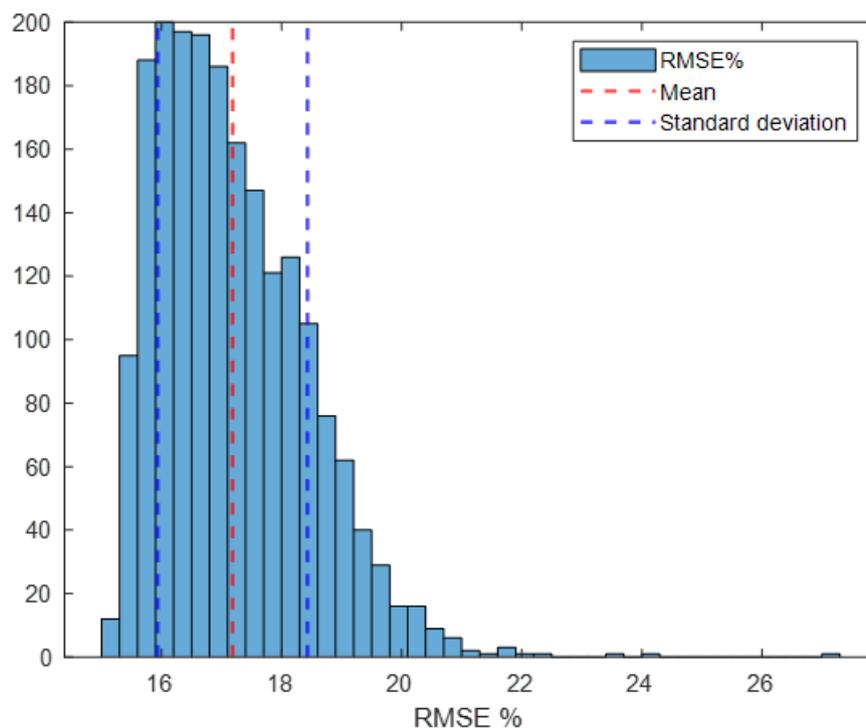


Figura 5.36: Istogramma con valor medio e deviazione standard in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.26. Come si può notare, la media e la mediana hanno valori molto simili e, di conseguenza, la distribuzione è circa simmetrica.

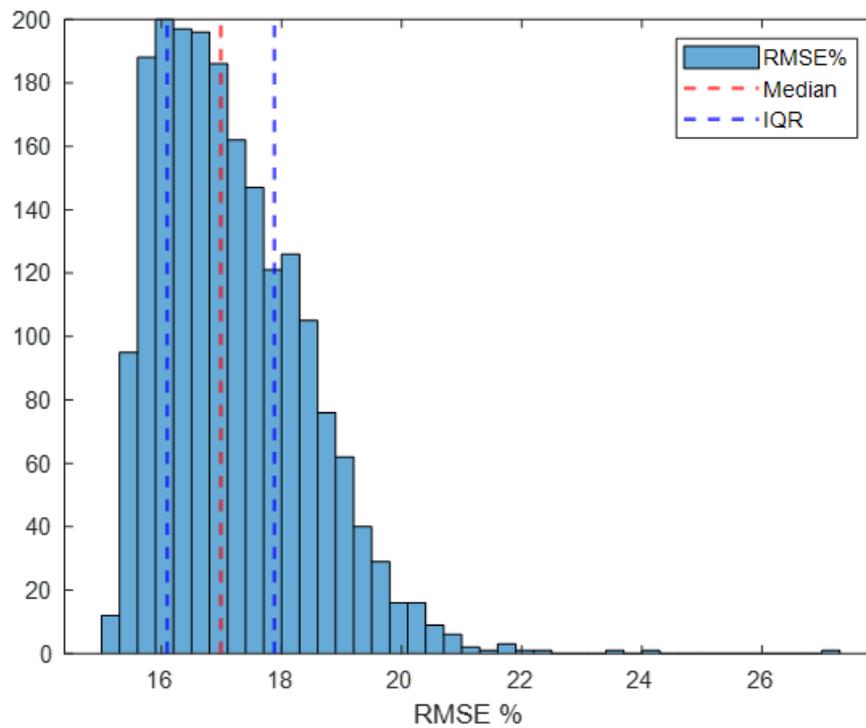


Figura 5.37: Istogramma con mediana e IQR in evidenza

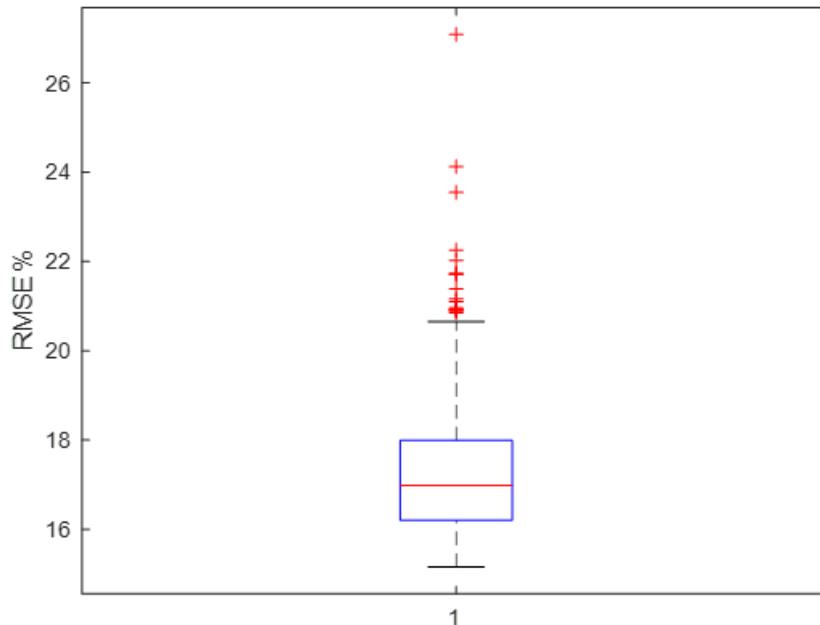


Figura 5.38: Box plot RMSE %

Gli indici di dispersione, invece, sono relativamente piccoli e ciò dimostra la robustezza di questo metodo alla variabilità della perturbazione. Inoltre, il RMSE % ottenuto

nel paragrafo precedente è leggermente inferiore al valor medio e ciò indica che è stato analizzato un caso leggermente "fortunato" nonostante i valori siano simili.

| RMSE % | |
|---------------------|----------|
| Valor medio | 17.183 % |
| Mediana | 16.987 % |
| Deviazione standard | 1.248 % |
| IQR | 1.784 % |

Tabella 5.26: Dati statistici del RMSE %

Nella figura 5.39, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

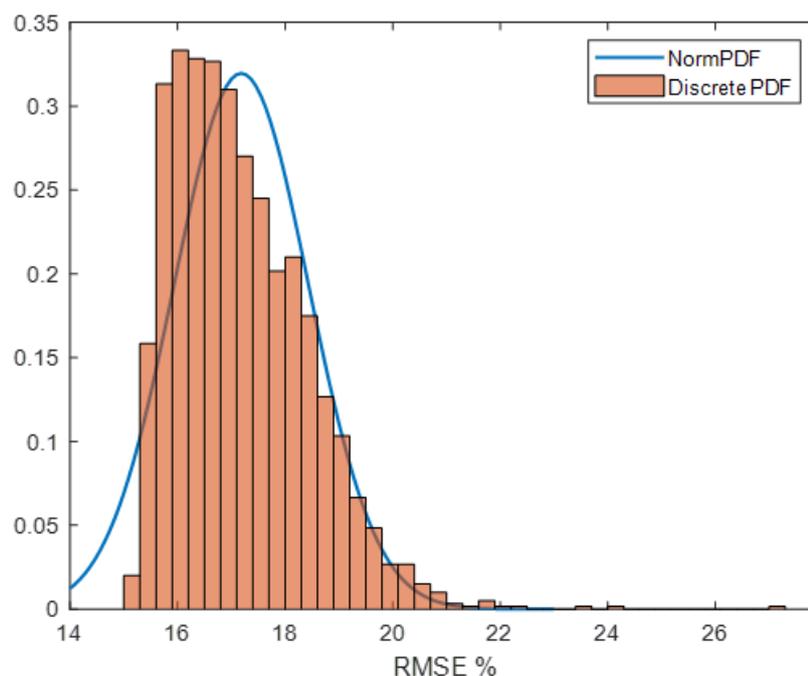


Figura 5.39: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che la PDF della distribuzione discreta ha un andamento simile a quello della distribuzione normale.

Parametro di regolarizzazione λ

I valori di λ ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

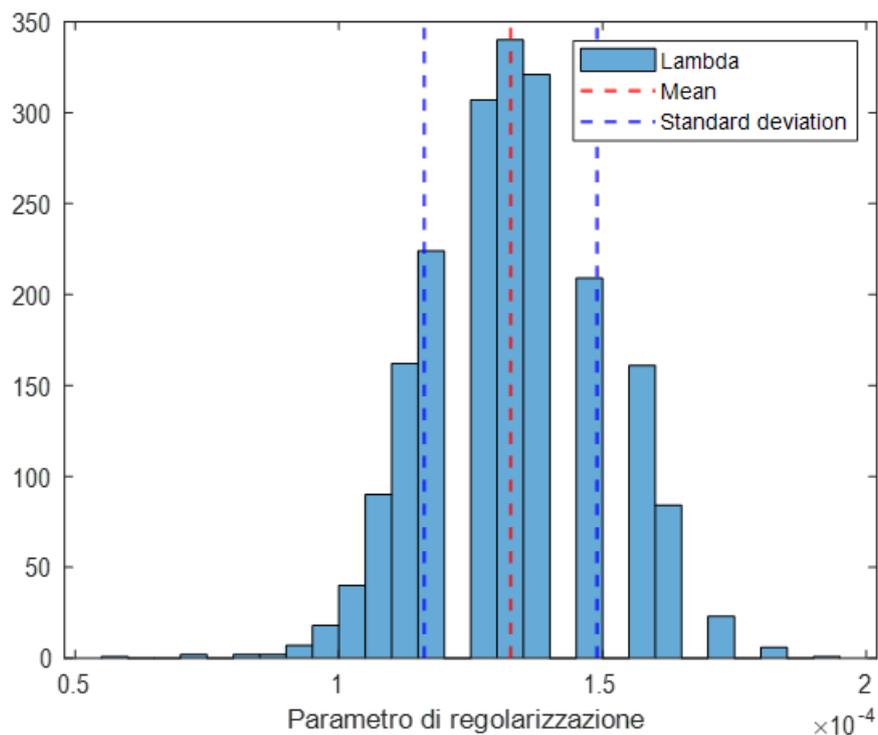


Figura 5.40: Istogramma con valor medio e deviazione standard in evidenza

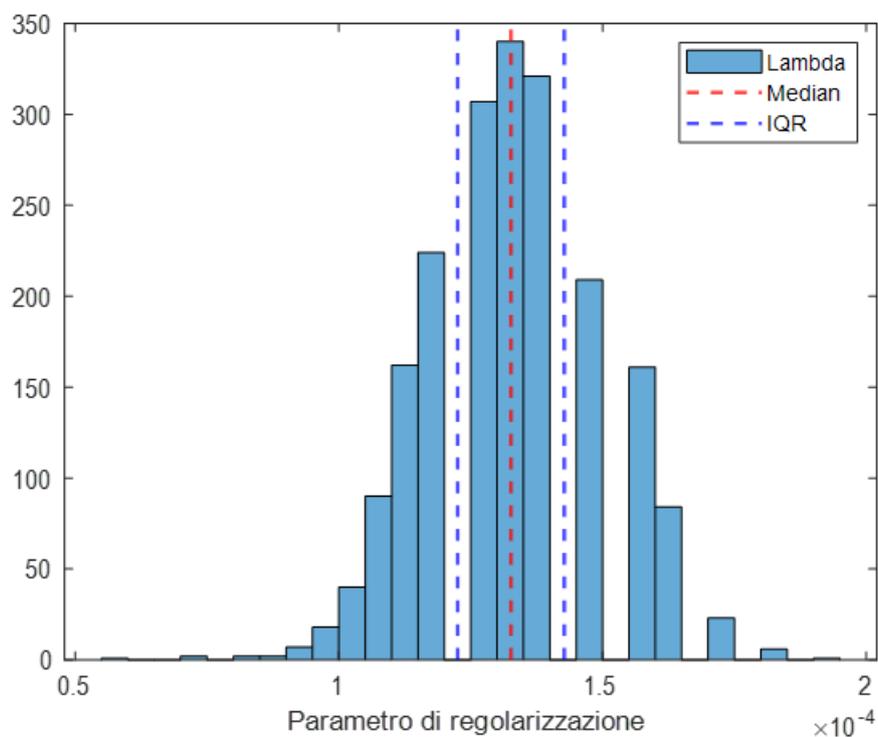


Figura 5.41: Istogramma con mediana e IQR in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.27. In questo caso la media e la mediana hanno lo stesso valore e, quindi, la distribuzione è simmetrica. Gli indici di dispersione sono piccoli e il metodo risulta essere robusto alla variabilità della

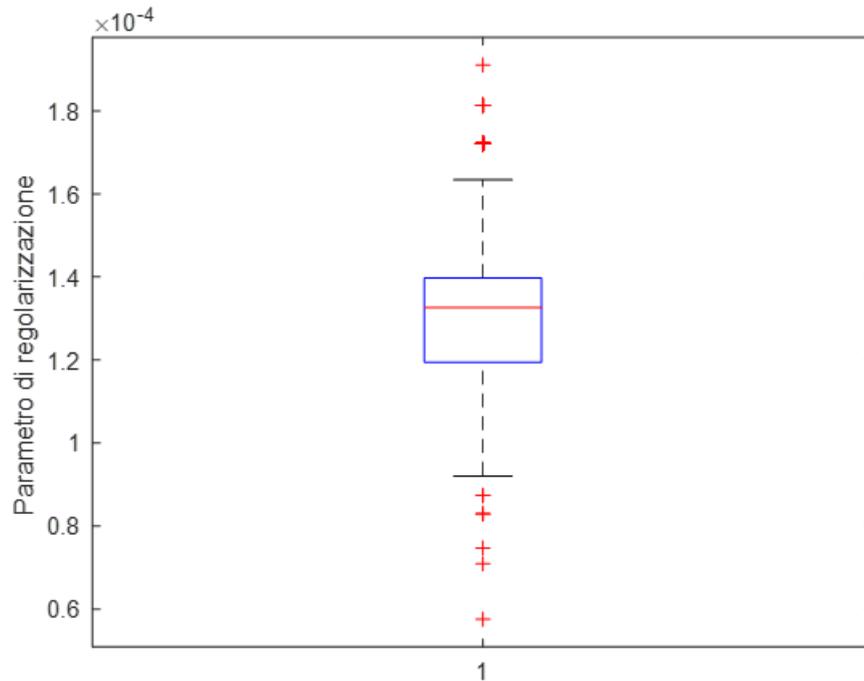


Figura 5.42: Box plot λ

perturbazione. Il λ calcolato nel paragrafo precedente è leggermente superiore al valor medio, ma comunque vicino e, pertanto, è possibile dire che complessivamente è stato analizzato un problema statisticamente probabile.

| Parametro di regolarizzazione λ | |
|---|----------|
| Valor medio | 1.33E-04 |
| Mediana | 1.33E-04 |
| Deviazione standard | 1.64E-05 |
| IQR | 2.02E-05 |

Tabella 5.27: Dati statistici di λ

Nella figura 5.43, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

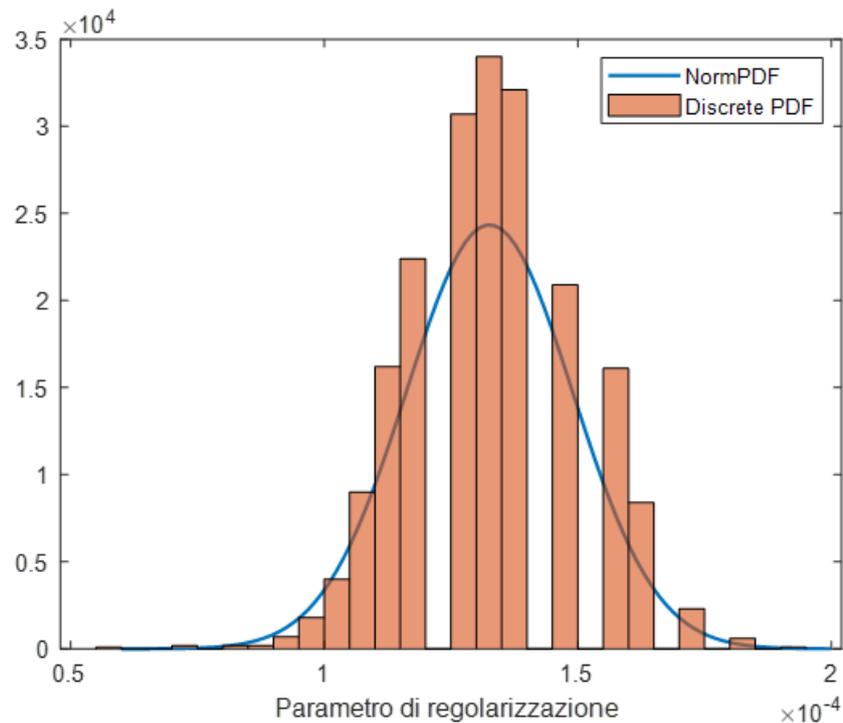


Figura 5.43: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che l'andamento della PDF della distribuzione discreta è simile a quello della PDF della distribuzione normale, nonostante la probabilità nell'intorno del valor medio sia superiore.

5.3.3 Metodo ART

Per lo studio statistico di questo metodo vengono svolte 200 analisi e non 2000 come nelle altre analisi statistiche a causa dell'elevato tempo computazionale che questo metodo richiede. Ogni analisi viene svolta con un diverso vettore di perturbazione creato in maniera random e il criterio di scelta del parametro di regolarizzazione considerato è il principio di discrepanza in quanto nel paragrafo precedente è stato il metodo che ha fornito il valore più accurato. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regolarizzazione k .

RMSE %

I valori del RMSE % ottenuti nelle 200 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

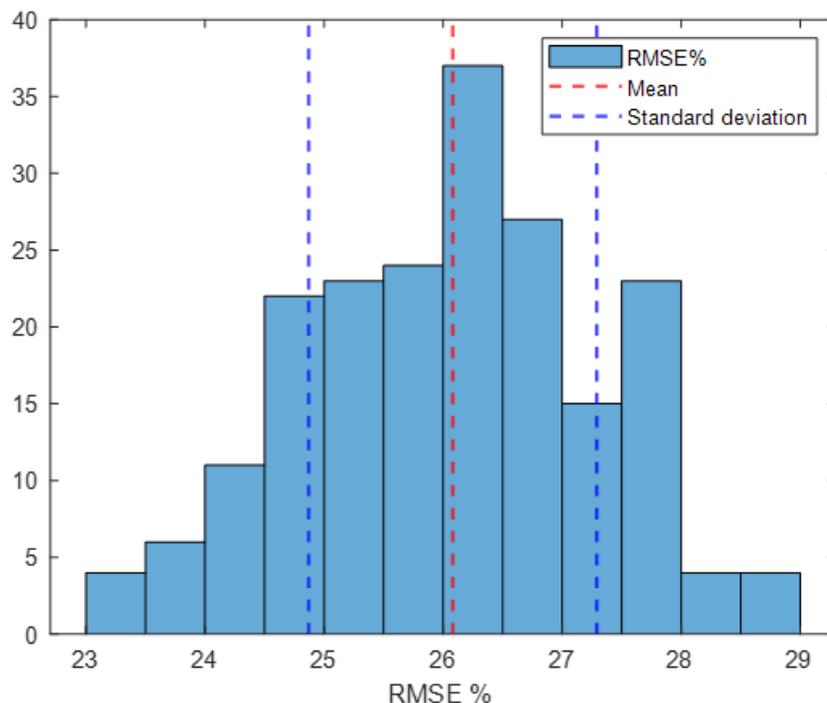


Figura 5.44: Istogramma con valor medio e deviazione standard in evidenza

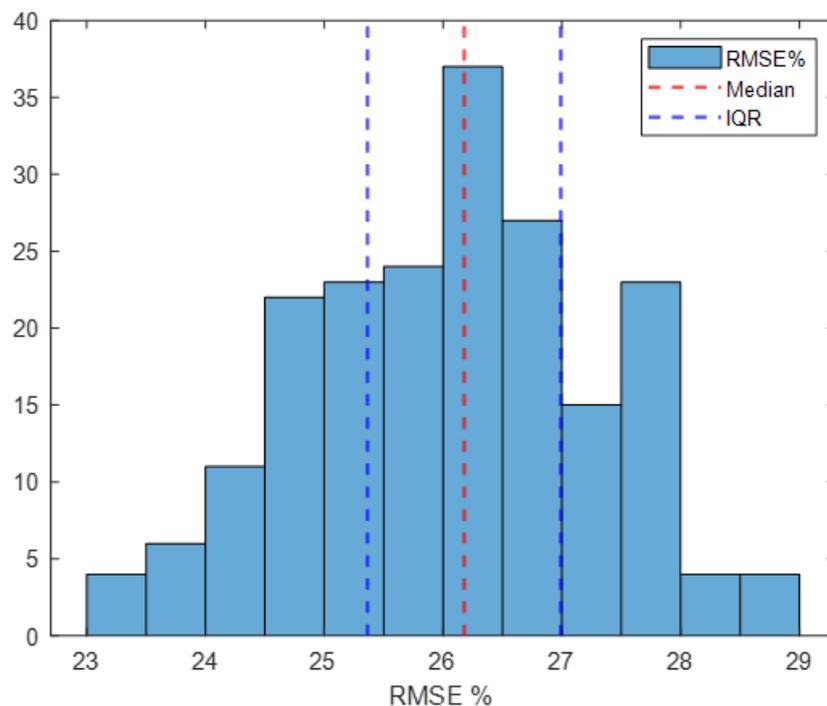


Figura 5.45: Istogramma con mediana e IQR in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.28. La media e la mediana hanno valori molto simili e, di conseguenza, la distribuzione può essere considerata simmetrica. Gli indici di dispersione, inoltre, sono relativamente bassi e il metodo si dimostra essere robusto alla variabilità della perturbazione. A differenza

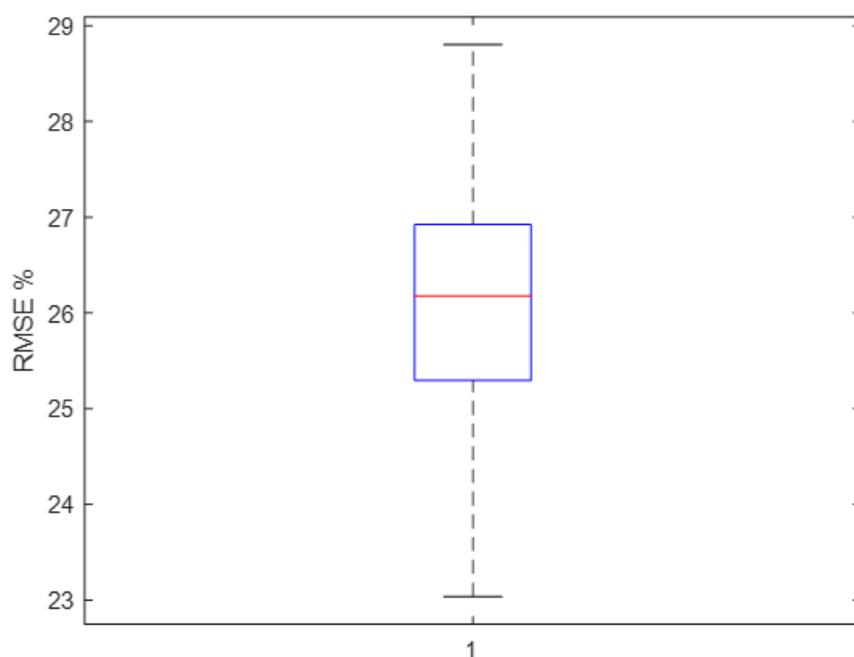


Figura 5.46: Box plot RMSE %

dei due metodi precedenti analizzati dal punto di vista statistico, tuttavia, il RMSE % ottenuto nel paragrafo precedente è decisamente più basso del valor medio e ciò indica che è stato analizzato un caso "fortunato" e statisticamente poco probabile.

| RMSE % | |
|---------------------|----------|
| Valor medio | 26.08 % |
| Mediana | 26.178 % |
| Deviazione standard | 1.209 % |
| IQR | 1.627 % |

Tabella 5.28: Dati statistici del RMSE %

Nella figura 5.47, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

Si può vedere, quindi, come la PDF della distribuzione discreta abbia un andamento praticamente coincidente con la PDF della distribuzione normale.

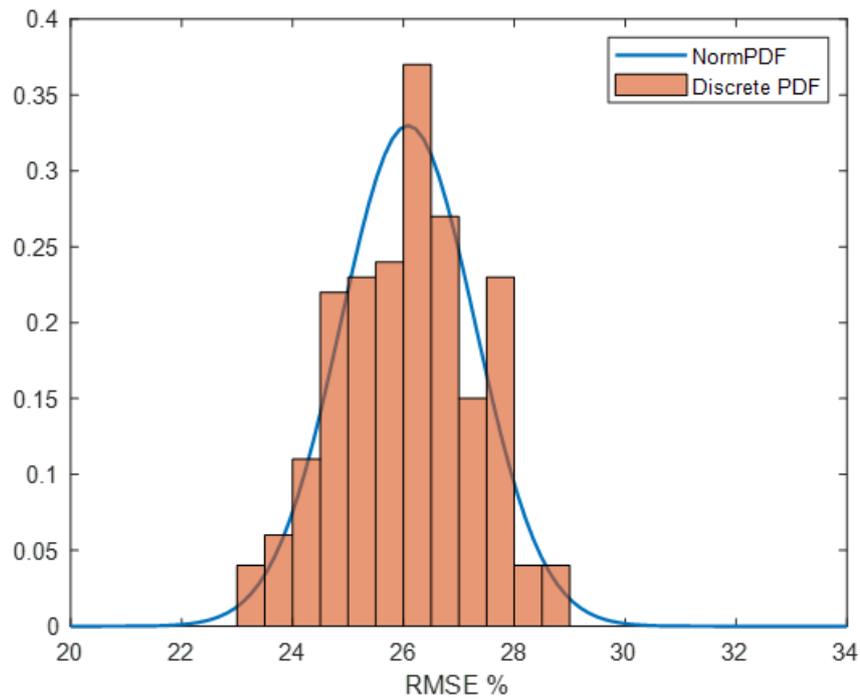


Figura 5.47: PDF della distribuzione discreta e PDF della distribuzione normale

Parametro di regolarizzazione k

I valori di k ottenuti nelle 200 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

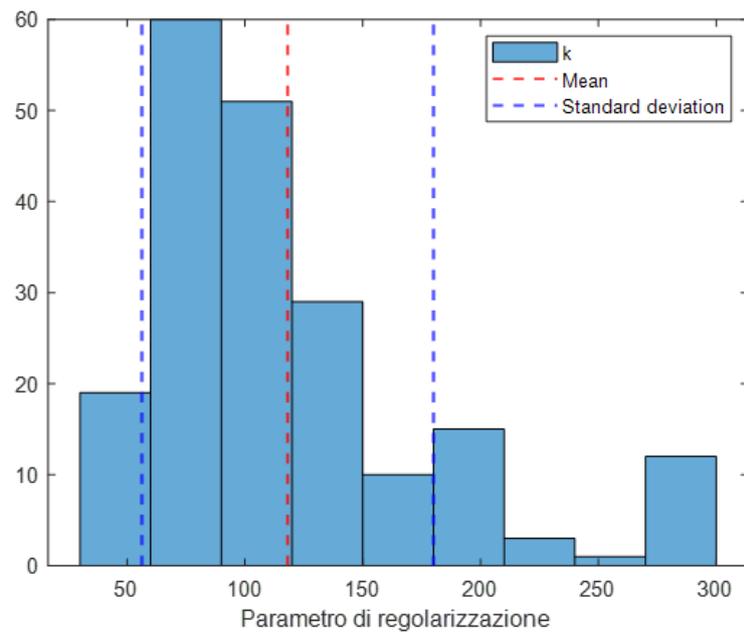


Figura 5.48: Istogramma con valor medio e deviazione standard in evidenza

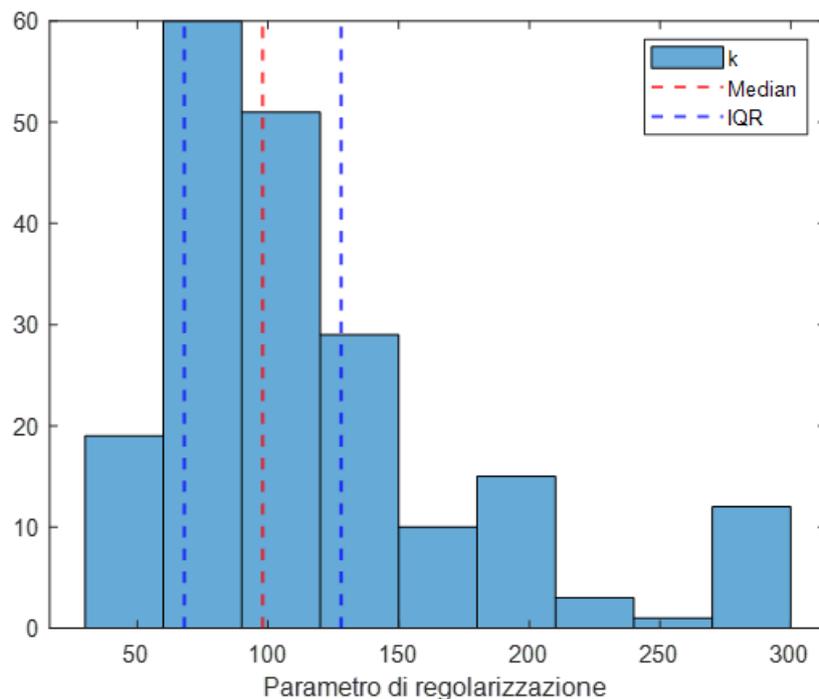


Figura 5.49: Istogramma con mediana e IQR in evidenza

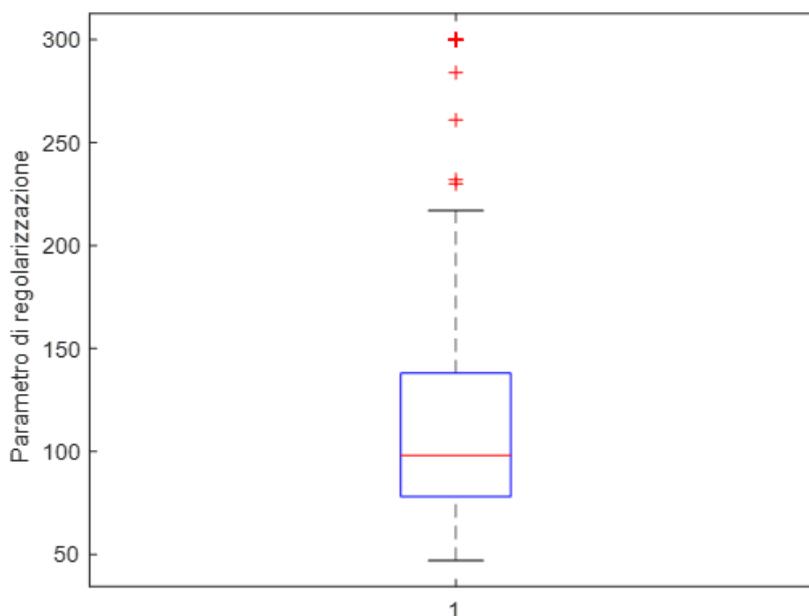


Figura 5.50: Box plot k

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.29. In questo caso la media e la mediana hanno valori abbastanza differenti e, come si può vedere dall'andamento dell'istogramma, la distribuzione non è simmetrica.

Inoltre, così come nel caso del RMSE %, il k calcolato nel paragrafo precedente è distante dal valor medio, a dimostrazione del fatto che il caso analizzato è statisticamente poco probabile.

| Parametro di regolarizzazione k | |
|---------------------------------|-----|
| Valor medio | 118 |
| Mediana | 98 |
| Deviazione standard | 62 |
| IQR | 60 |

Tabella 5.29: Dati statistici di k

Nella figura 5.51, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

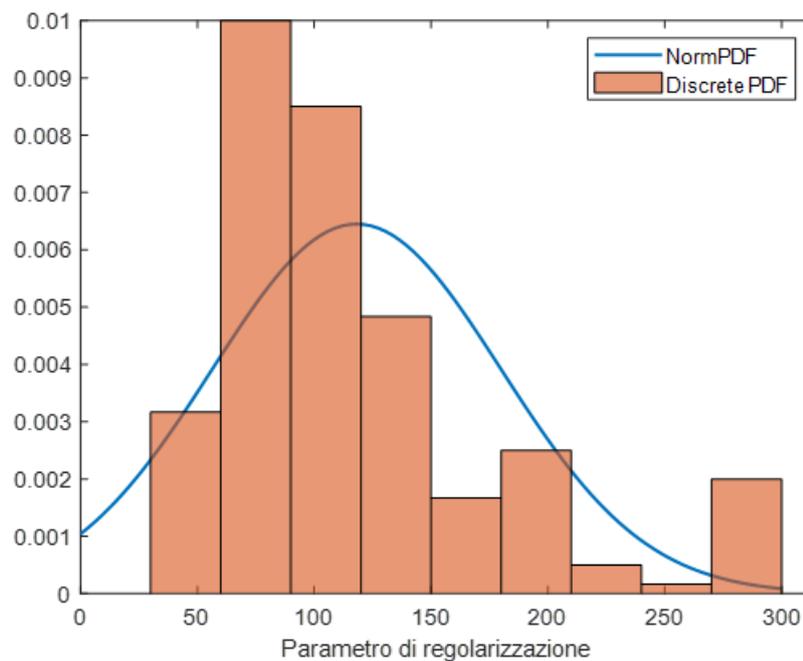


Figura 5.51: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che la PDF della distribuzione discreta non è simmetrica e che, quindi, non coincide con quella della distribuzione normale.

5.3.4 Metodo LSQR

Per lo studio statistico di questo metodo vengono svolte 2000 analisi, ognuna di esse con un diverso vettore di perturbazione (il quale viene creato in maniera random), e il criterio di scelta del parametro di regolarizzazione considerato è il metodo della curva L in quanto è stato l'unico criterio adottato nel paragrafo precedente. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regolarizzazione k.

RMSE %

I valori del RMSE % ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

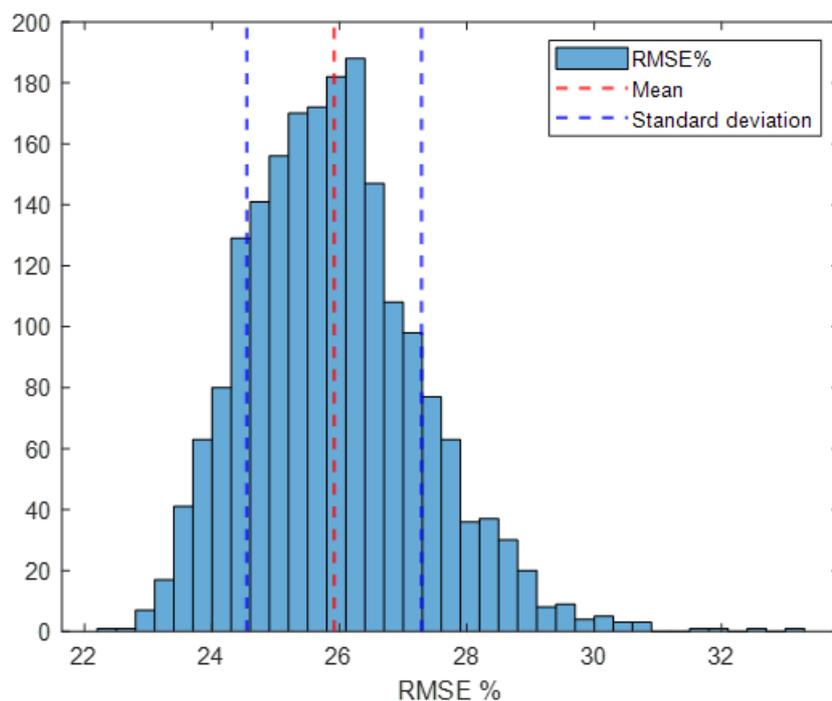


Figura 5.52: Istogramma con valor medio e deviazione standard in evidenza

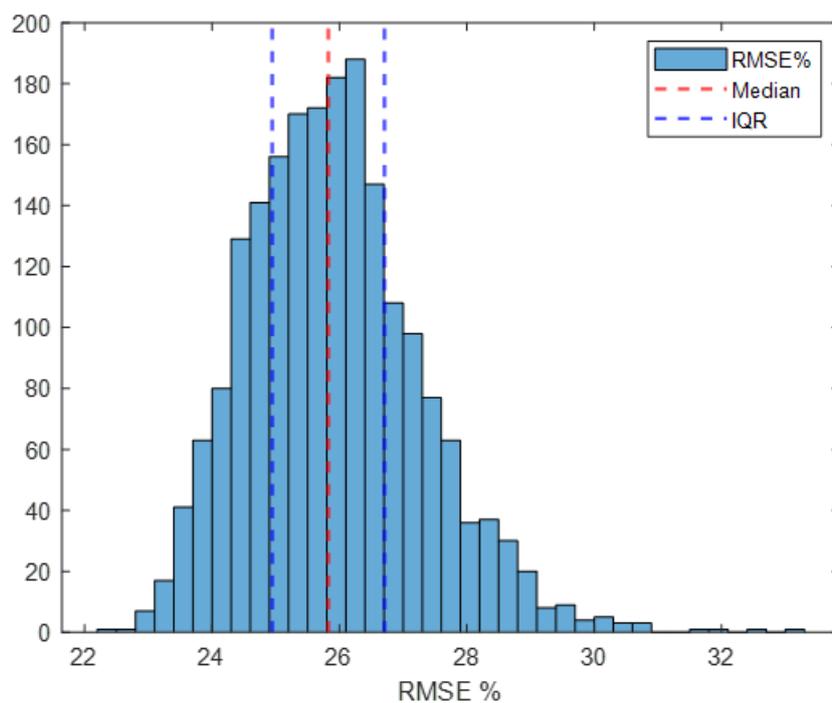


Figura 5.53: Istogramma con mediana e IQR in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.30. Come si può vedere, la media e la mediana hanno valori molto vicini e, di conseguenza, la distribuzione

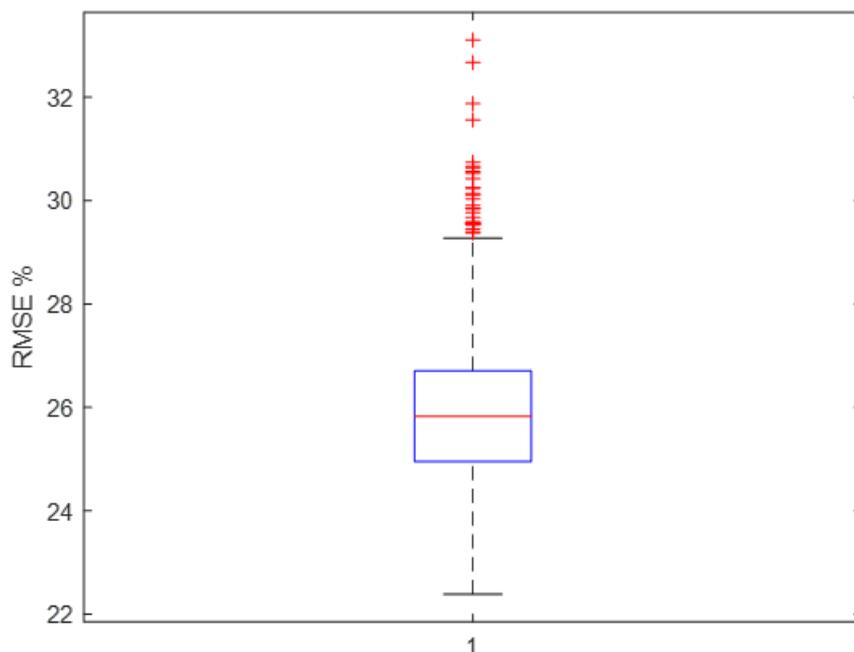


Figura 5.54: Box plot RMSE %

può essere considerata simmetrica.

Gli indici di dispersione, invece, sono relativamente piccoli e ciò dimostra la robustezza del metodo alla variabilità della perturbazione. Il RMSE % calcolato nel paragrafo precedente, tuttavia, è abbastanza superiore al valore medio e ciò indica che è stato analizzato un caso "sfortunato" e statisticamente poco probabile.

| RMSE % | |
|---------------------|----------|
| Valor medio | 25.918 % |
| Mediana | 25.826 % |
| Deviazione standard | 1.372 % |
| IQR | 1.761 % |

Tabella 5.30: Dati statistici del RMSE %

Nella figura 5.55, infine, è rappresentata la funzione di densità di probabilità (PDF) della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

Si può vedere, quindi, come la PDF della distribuzione discreta abbia un andamento praticamente coincidente con la PDF della distribuzione normale.

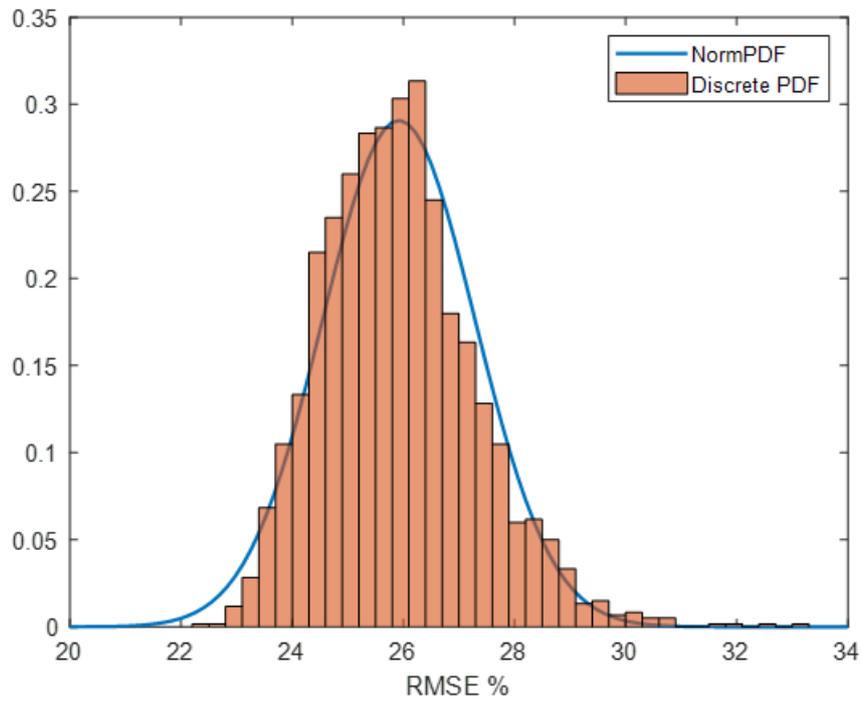


Figura 5.55: PDF della distribuzione discreta e PDF della distribuzione normale

Parametro di regolarizzazione k

I valori di k ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

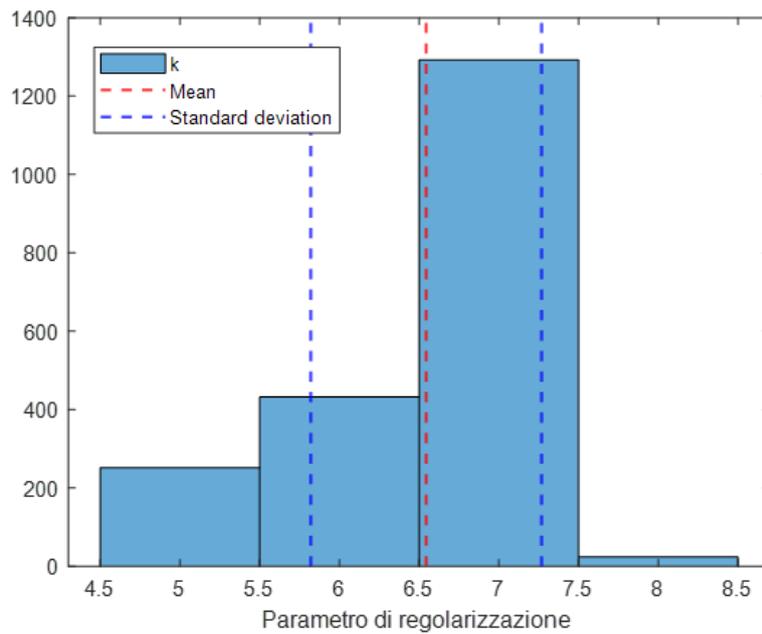


Figura 5.56: Istogramma con valor medio e deviazione standard in evidenza

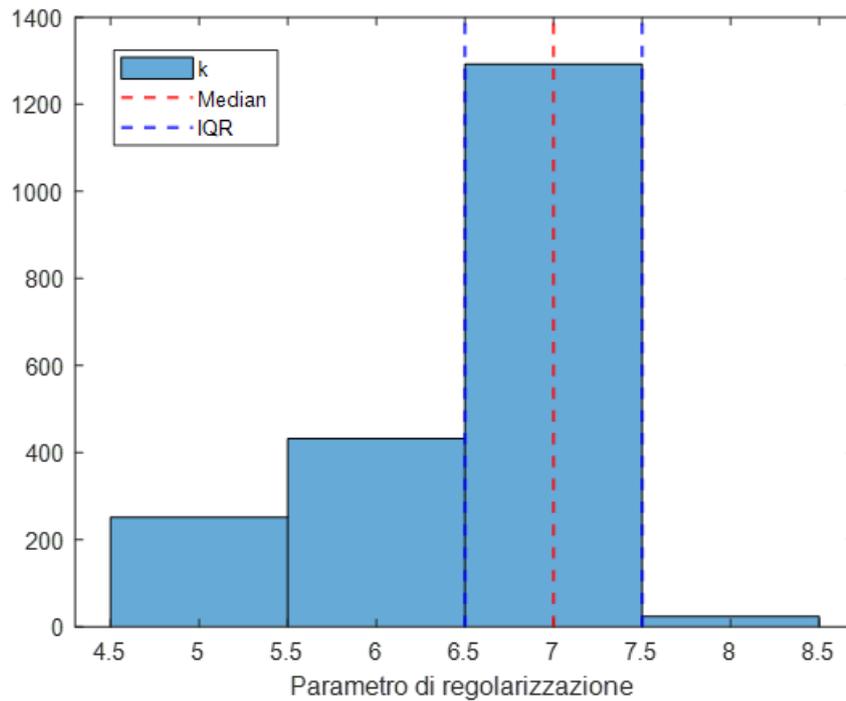


Figura 5.57: Istogramma con mediana e IQR in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 5.31. In questo caso la media e la mediana hanno valori leggermente differenti e, come si può vedere dall'andamento dell'istogramma, la distribuzione non è esattamente simmetrica.

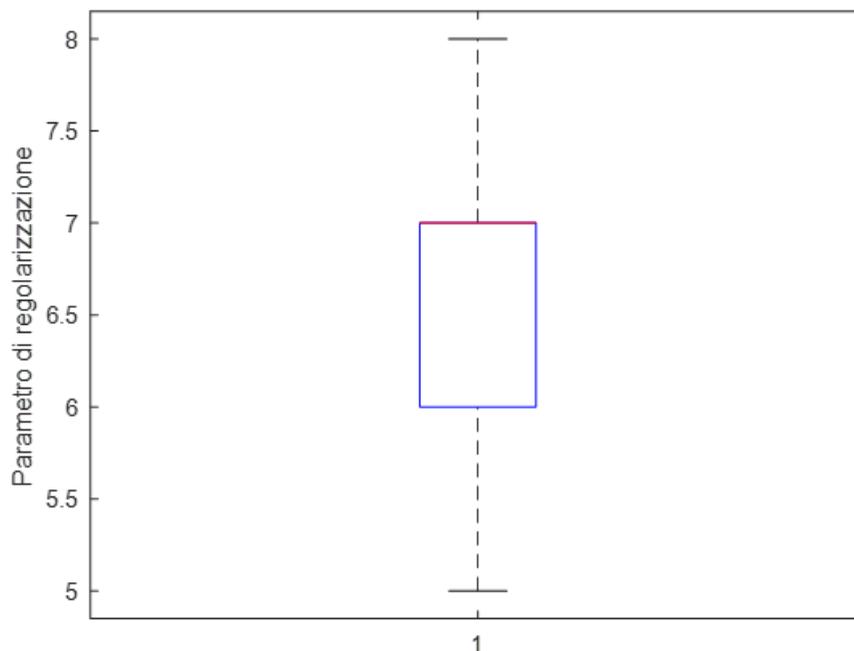


Figura 5.58: Box plot k

Inoltre, si può notare come la gran parte dei valori di k ottenuti coincide con il valore

calcolato nel paragrafo precedente.

| Parametro di regolarizzazione k | |
|---------------------------------|-------|
| Valor medio | 6.544 |
| Mediana | 7 |
| Deviazione standard | 0.724 |
| IQR | 1 |

Tabella 5.31: Dati statistici di k

Nella figura 5.59, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

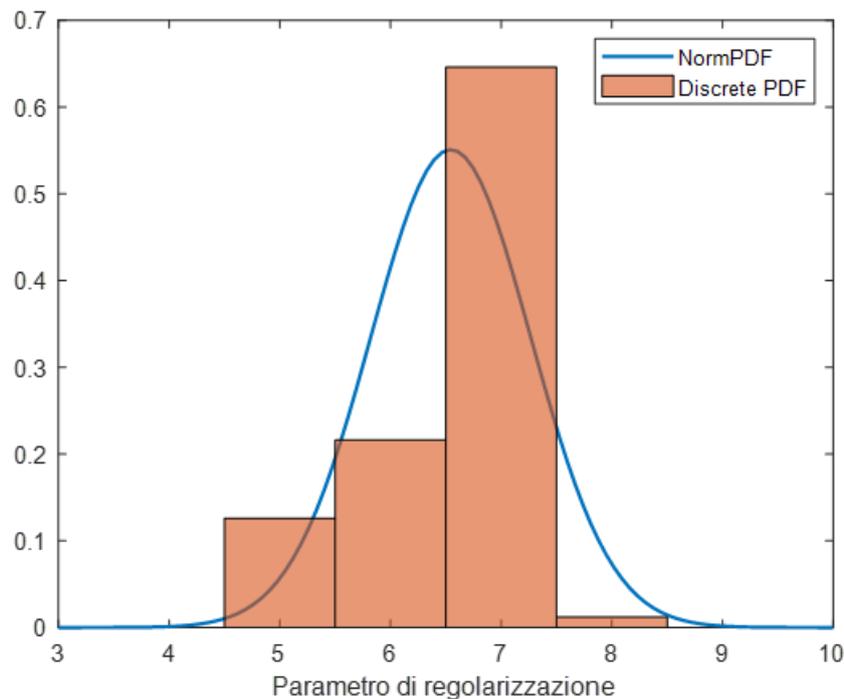


Figura 5.59: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che la PDF della distribuzione discreta non è simmetrica e che, quindi, non coincide con quella della distribuzione normale.

5.3.5 Confronto dei risultati

In tabella 5.32, infine, sono riassunti i risultati ottenuti dall'analisi statistica relativi al RMSE % e al parametro di regolarizzazione in termini di media, mediana, deviazione standard e scarto interquartile.

| | RMSE % | | | | $\lambda o k$ | | | |
|---------------|------------|----------|-------------|---------|---------------|----------|------------|----------|
| | Val. Medio | Mediana | Dev. Stand. | IQR | Val. Medio | Mediana | Dev. Stand | IQR |
| Tikh | 26.203 % | 26.205 % | 1.830 % | 2.602 % | 4.21E-05 | 4.05E-05 | 1.22E-05 | 1.68E-05 |
| Tikh-1 | 17.183 % | 16.987 % | 1.248 % | 1.784 % | 1.33E-04 | 1.33E-04 | 1.64E-05 | 2.02E-05 |
| ART | 26.08 % | 26.178 % | 1.209 % | 1.627 % | 118 | 98 | 61.85 | 60 |
| LSQR | 25.918 % | 25.826 % | 1.372 % | 1.761 % | 6.544 | 7 | 0.724 | 1 |

Tabella 5.32: Riassunto dei risultati

Come si può notare, i dati statistici confermano i risultati ottenuti nel paragrafo 5.2 con un unico vettore di perturbazione, ad eccezione del Metodo ART in cui gli indici di centralità relativi al RMSE % si sono rivelati essere maggiori rispetto al valore ottenuto in precedenza.

Capitolo 6

Ricostruzione e regolarizzazione del carico su un cassone alare

In quest'ultimo Capitolo viene presentata un'applicazione delle metodologie di ricostruzione del carico e di regolarizzazione della soluzione al cassone alare introdotto in [2]. Poiché per questo modello non sono disponibili dei dati di apprendimento, l'unico metodo utilizzato per la ricostruzione del carico è l'approccio inverso. Il metodo utilizzato in [2] per la regolarizzazione della soluzione è quello di Tikhonov in forma standard; nel Capitolo 5, tuttavia, è stato mostrato tramite l'analisi statistica sul modello di piastra che il metodo di Tikhonov del primo ordine e il metodo iterativo LSQR forniscono valori medi del RMSE % più bassi. In questo studio, quindi, si è deciso di confrontare in termini statistici il metodo di Tikhonov in forma standard, il metodo di Tikhonov del primo ordine, il metodo LSQR e il metodo ART.

6.1 Modello

Geometria e materiali

Il cassone alare è costituito da due pannelli (superiore e inferiore) connessi a due longheroni tramite quattro irrigidimenti angolari e rinforzati tramite due correnti. L'intero cassone alare è diviso in sette scomparti per mezzo di sei centine rettangolari e la sezione d'estremità. La geometria del cassone e di ogni componente è rappresentata in figura 6.1. Il cassone alare presenta due differenti leghe di Alluminio. I pannelli, i longheroni e le centine sono costituiti dalla lega 7075, mentre i correnti e gli irrigidimenti sono costituiti dalla lega 6060. Le proprietà meccaniche delle due leghe sono presentate in tabella 6.1.

| Lega | E [GPa] | ν | G [GPa] |
|------|---------|-------|---------|
| 6060 | 66 | 0.33 | 24.8 |
| 7075 | 72 | 0.32 | 27.2 |

Tabella 6.1: Proprietà meccaniche delle leghe di Alluminio

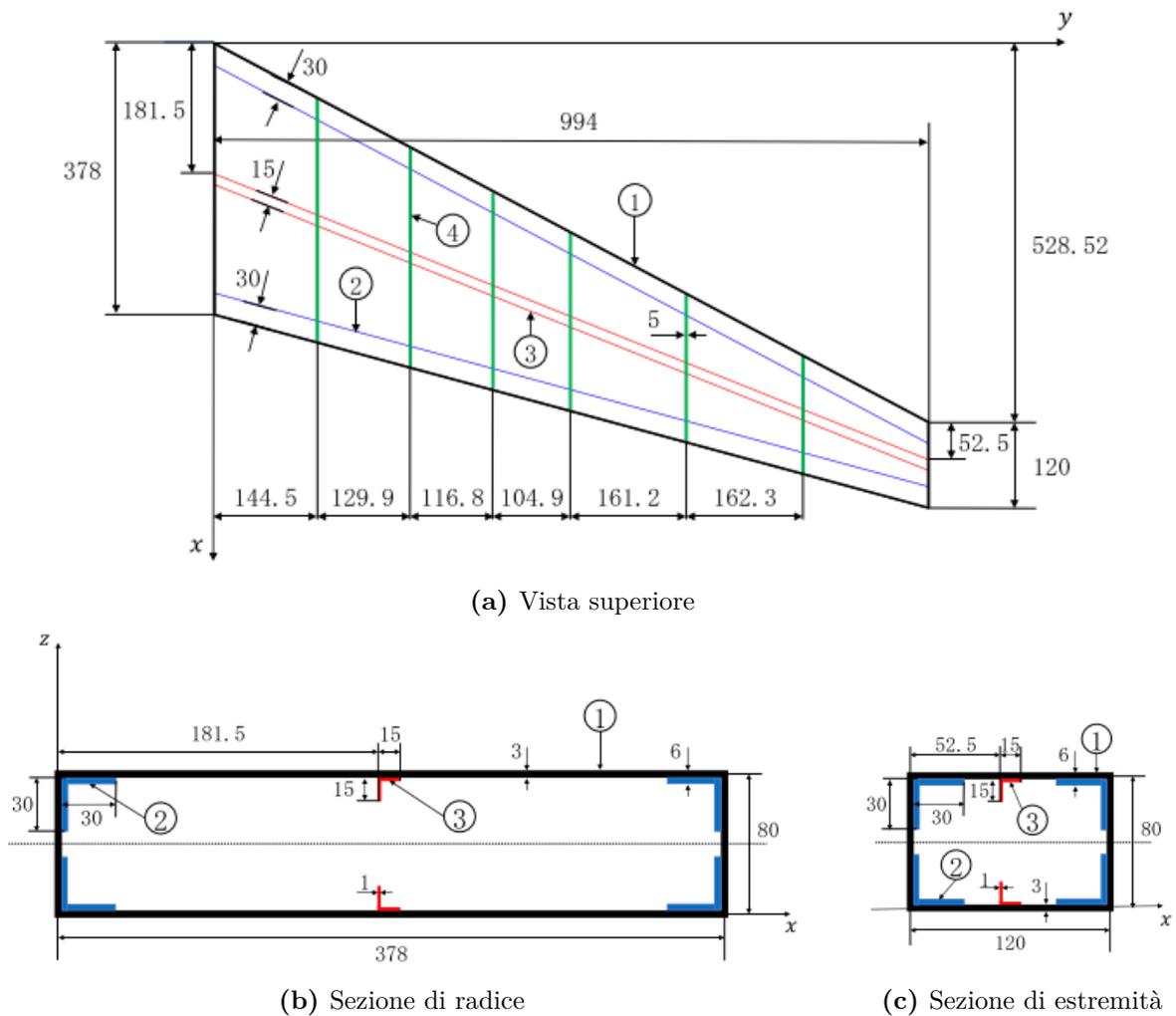


Figura 6.1: Geometria del cassone alare [2]

Carico

Il carico consiste in una distribuzione di pressione che simula una distribuzione di pressione aerodinamica. Essa è applicata sul pannello superiore ed è rappresentata in figura 6.2. In termini matematici, è dato dal prodotto tra un polinomio cubico funzione di x e uno quadratico funzione di y :

$$p(x, y) = p_1(x) \cdot p_2(y) = (a_x x^3 + b_x x^2 + c_x x + d_x) \cdot (a_y y^2 + b_y y + c_y) \quad (6.1)$$

Questa distribuzione, lungo l'apertura alare, ha un massimo in corrispondenza della sezione di radice ed è nulla in corrispondenza della sezione di estremità. Lungo la direzione della corda, invece, è nulla in corrispondenza dei due bordi ed è massima in corrispondenza del terzo di corda. Matematicamente, il polinomio funzione di y è costruito imponendo che la funzione sia nulla in corrispondenza della sezione di estremità e abbia un massimo pari a 1 in corrispondenza della sezione di radice. Il polinomio funzione di x , invece, è

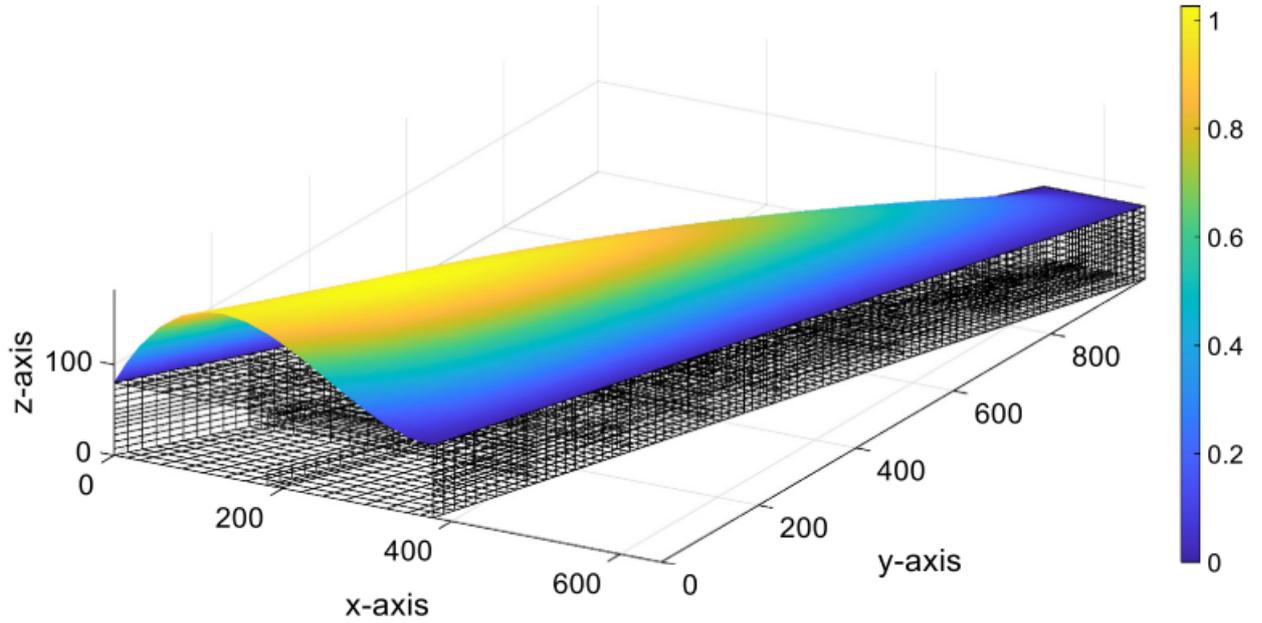


Figura 6.2: Carico [2]

costruito imponendo che sia nullo in corrispondenza dei due bordi e che abbia un massimo pari a 1 in corrispondenza del terzo di corda. Quindi, le tre condizioni necessarie a trovare i coefficienti di $p_2(y)$ sono:

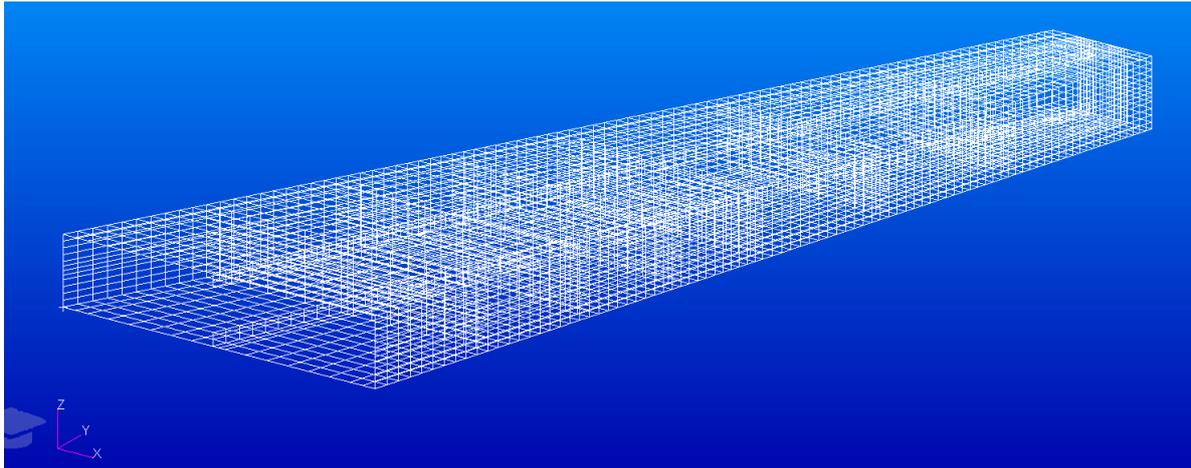
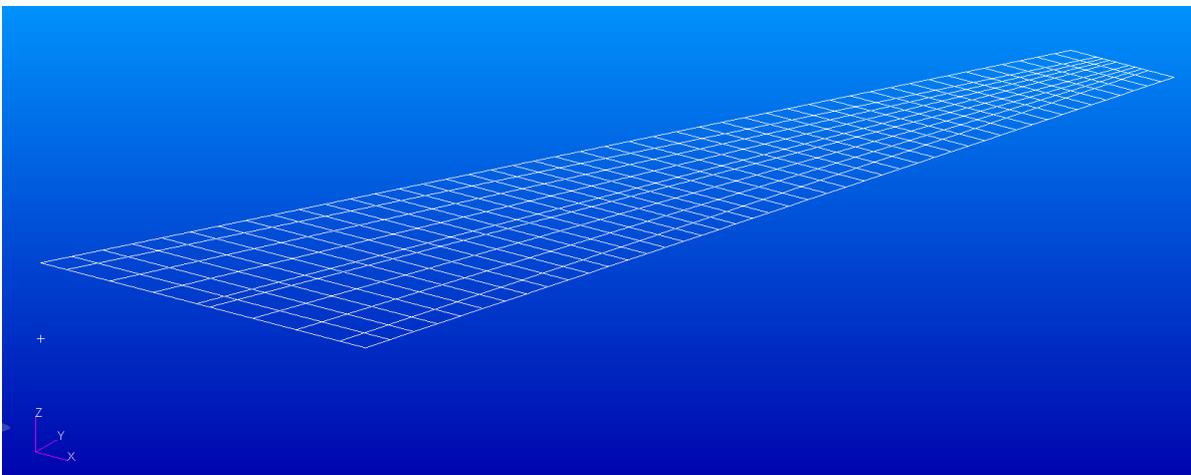
$$\begin{cases} p_2(y = 994) = 0 & p_2 \text{ uguale a } 0 \text{ nella sezione di estremità} \\ p_2(y = 0) = 1 & p_2 \text{ è uguale a } 1 \text{ nella sezione di radice} \\ p_{2,y}(y = 0) = 0 & p_2 \text{ ha un massimo nella sezione di radice} \end{cases} \quad (6.2)$$

Le quattro condizioni necessarie a trovare i coefficienti di $p_1(x)$, invece, sono:

$$\begin{cases} p_1\left(x = \frac{528.52}{994}y\right) = 0 & p_1 \text{ uguale a } 0 \text{ nel bordo d'attacco} \\ p_1\left(x = \frac{270.52}{994}y + 378\right) = 0 & p_1 \text{ uguale a } 0 \text{ nel bordo d'uscita} \\ p_1\left(x = \frac{618.69}{994}y + 126\right) = 1 & p_1 \text{ uguale a } 1 \text{ ad un terzo della corda} \\ p_{1,x}\left(x = \frac{618.69}{994}y + 126\right) = 0 & p_1 \text{ ha un massimo ad un terzo della corda} \end{cases} \quad (6.3)$$

Modello

L'applicazione dell'approccio inverso per la ricostruzione del carico richiede la definizione di un modello dettagliato agli elementi finiti della struttura. Inoltre, poiché la natura di questo studio è numerica, il modello viene anche utilizzato per simulare il reale comportamento del cassone alare e, di conseguenza, utilizzato per generare i dati di input in termini di deformazione. Il modello di dettaglio, caratterizzato dalla mesh strutturale, è costituito da 9792 elementi CQUAD4 NASTRAN® e 7129 nodi e rappresentato in figura 6.3.

**Figura 6.3:** Mesh strutturale**Figura 6.4:** Mesh di pressione

E', inoltre, necessaria anche una mesh di pressione per l'interpolazione dei valori nodali calcolati tramite l'approccio inverso e per ottenere, quindi, la distribuzione completa. Essa è costituita da 405 elementi CQUAD4 NASTRAN® e 460 nodi ed è rappresentata in figura 6.4.

Configurazione estensimetrica

La configurazione estensimetrica scelta per questa analisi è la stessa di quella introdotta in [2] ed è rappresentata in figura 6.5. Essa è caratterizzata dalla presenza di 7 linee di sensori e le loro posizioni sono definite attraverso un Algoritmo Genetico. Quest'ultimo ottimizza la configurazione in modo che l'accuratezza dell'identificazione del carico esterno sia la migliore ottenibile con quel numero di linee, ovvero il RMSE % misurato tra i valori nodali di pressione calcolati e quelli di riferimento è il minimo ottenibile.

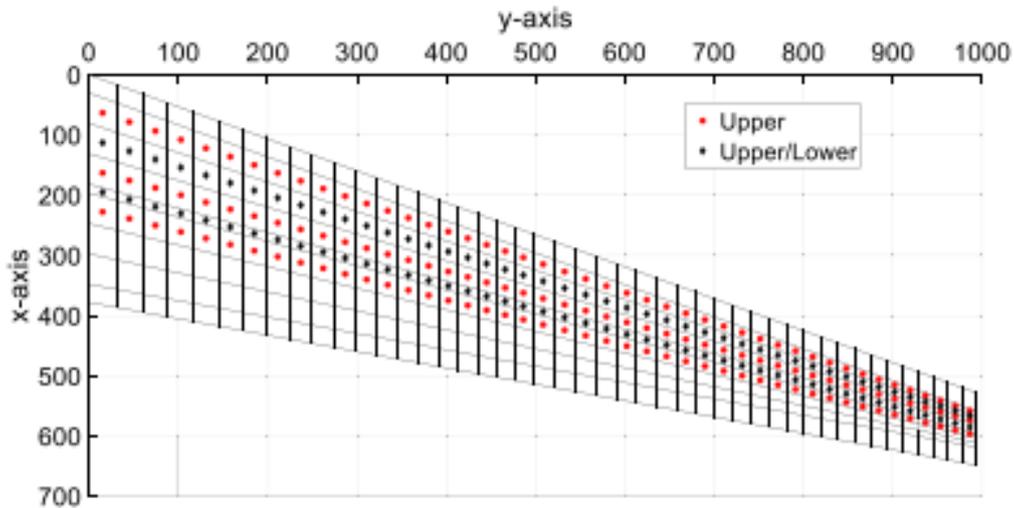


Figura 6.5: Configurazione estensimetrica. Upper indica i sensori posizionati sul pannello superiore, Upper/Lower indica i sensori posizionati sul pannello superiore e quello inferiore. [2]

6.2 Ricostruzione del carico

Per svolgere questo compito, poiché non sono disponibili dei dati di apprendimento per l'addestramento di una rete neurale, l'unico metodo utilizzato è quello dell'approccio inverso.

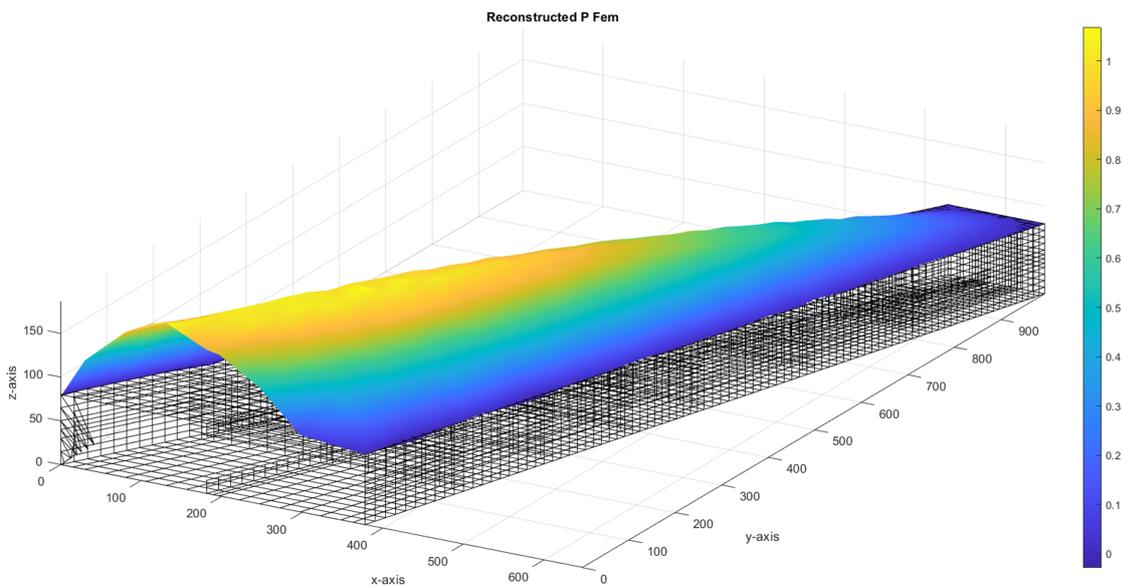


Figura 6.6: Ricostruzione del carico tramite approccio inverso

Come si vede in figura 6.6, la distribuzione di pressione ricostruita risulta essere molto accurata e il **RMSE %** ottenuto è pari a **1.65 %**.

A causa della natura mal posta dei problemi inversi, tuttavia, se si introduce un errore (anche molto piccolo) nelle misure di deformazione si ottiene una soluzione caratterizzata da un errore molto elevata e non rappresentativa del problema. A dimostrazione di ciò, introducendo un errore dello stesso tipo di quello introdotto nel Capitolo 5, si ottiene un **RMSE %** pari a **8.34E04 %** e la sua distribuzione sulla struttura è rappresentata in figura 6.7.

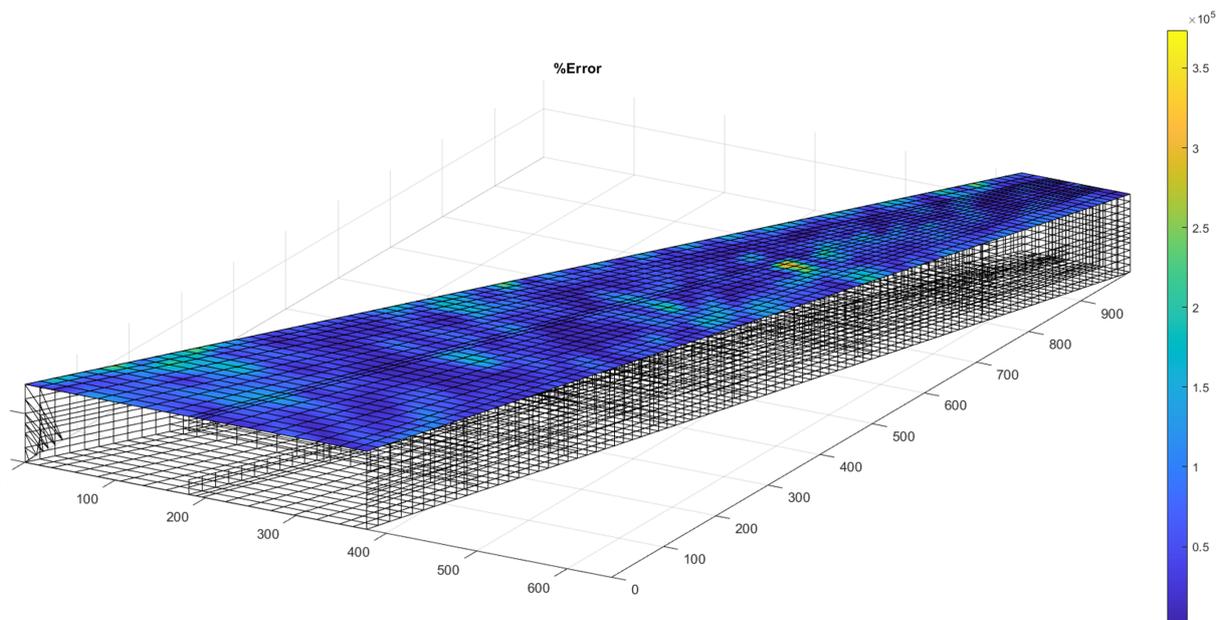


Figura 6.7: Distribuzione dell'errore associato alla soluzione non regolarizzata

6.3 Regolarizzazione della soluzione

Per il motivo spiegato nel paragrafo precedente, quindi, è necessario regolarizzare la soluzione e ridurre l'effetto dell'errore. Per fare ciò, viene sfruttato il lavoro che è stato svolto sul modello di piastra, ovvero si utilizzano le metodologie di regolarizzazione e di scelta del parametro che hanno fornito i migliori risultati su di esso. In particolare, si utilizzano il metodo di Tikhonov in forma standard (e principio di discrepanza), il metodo di Tikhonov del primo ordine (e principio di discrepanza), il metodo ART (e metodo della curva L) e, infine, il metodo LSQR (e metodo della curva L). Per ciascun metodo viene effettuata un'analisi statistica con l'obiettivo di ottenere dei risultati rappresentativi della bontà generale dei vari metodi e non legati a un particolare vettore di perturbazione.

6.3.1 Metodo di Tikhonov in forma standard

Per lo studio statistico di questo metodo vengono svolte 2000 analisi, ognuna di esse con un diverso vettore di perturbazione (il quale viene creato in maniera random), e il criterio di scelta del parametro di regolarizzazione considerato è il principio di discrepanza. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regolarizzazione λ .

RMSE %

I valori del RMSE % ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

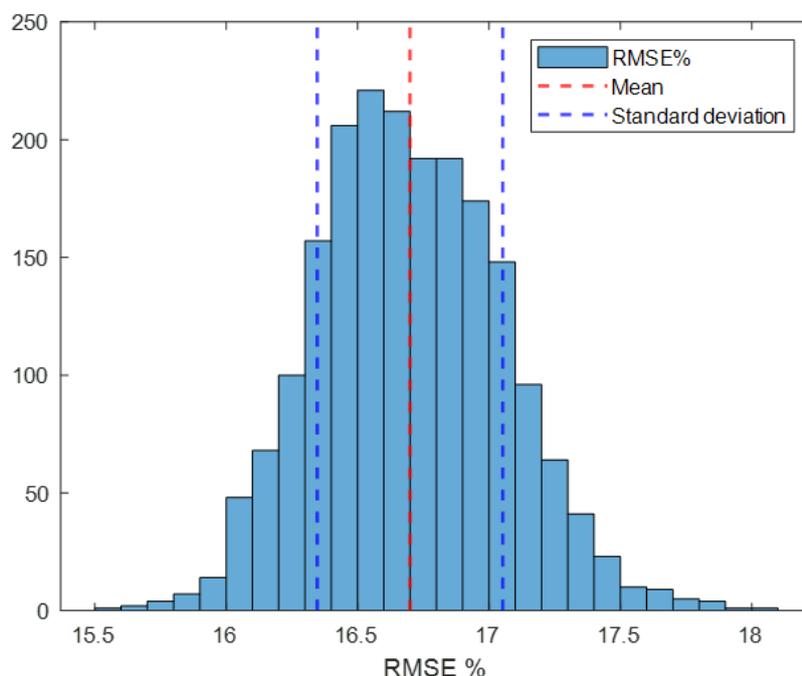


Figura 6.8: Istogramma con valor medio e deviazione standard in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.2. Come si può vedere, la media e la mediana hanno valori molto vicini e, di conseguenza, la distribuzione può essere considerata simmetrica. Gli indici di dispersione, invece, sono relativamente piccoli e ciò dimostra la robustezza del metodo alla variabilità della perturbazione.

| RMSE % | |
|---------------------|---------|
| Valor medio | 16.69 % |
| Mediana | 16.68 % |
| Deviazione standard | 0.353 % |
| IQR | 0.498 % |

Tabella 6.2: Dati statistici del RMSE %

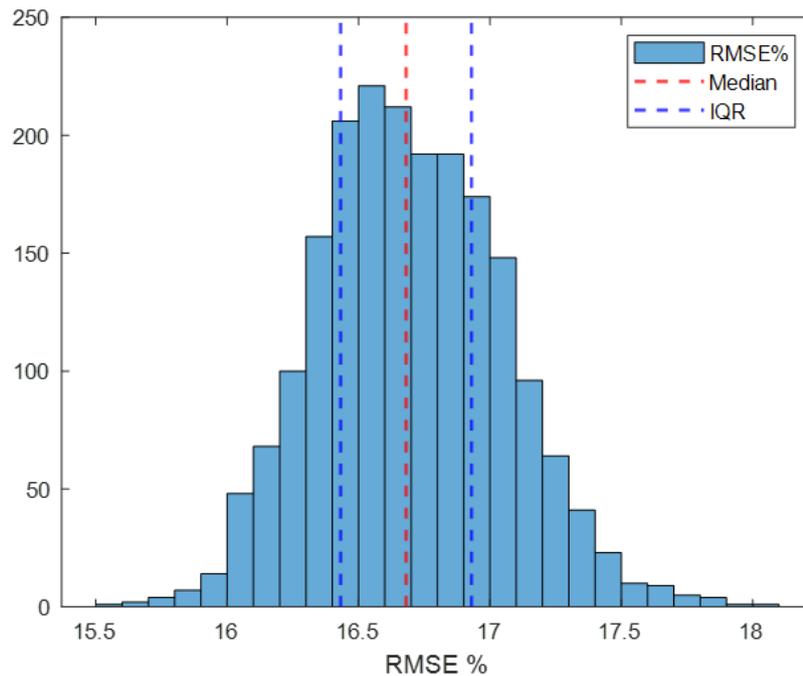


Figura 6.9: Istogramma con mediana e IQR in evidenza

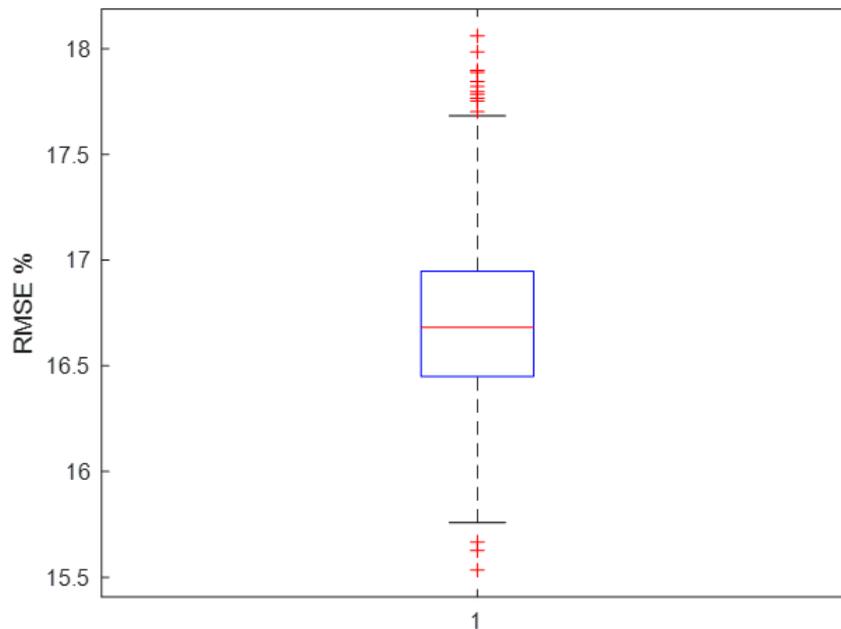


Figura 6.10: Box plot RMSE %

Nella figura 6.11, infine, è rappresentata la funzione di densità di probabilità (PDF) della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

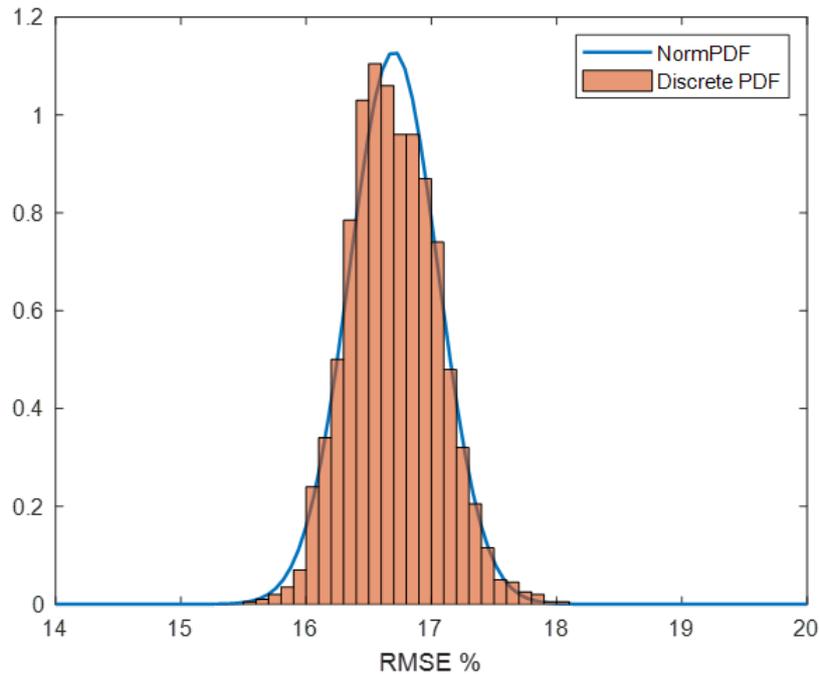


Figura 6.11: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, come la PDF della distribuzione discreta abbia un andamento praticamente coincidente con la PDF della distribuzione normale.

Parametro di regolarizzazione λ

I valori di λ ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

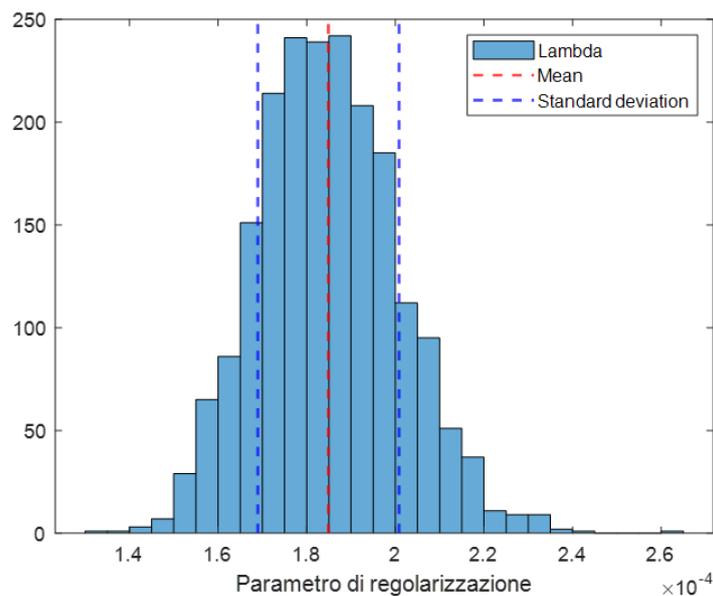


Figura 6.12: Istogramma con valor medio e deviazione standard in evidenza

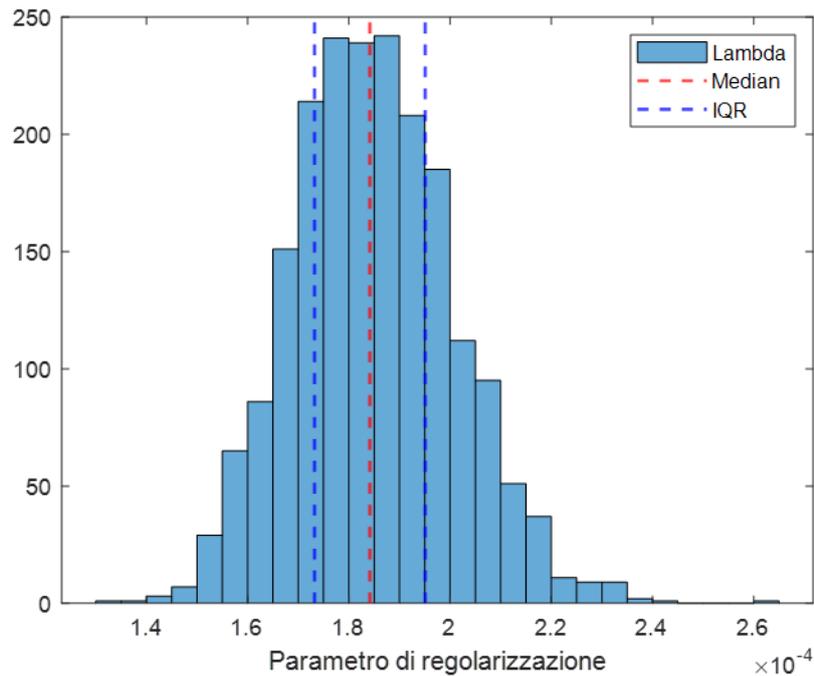


Figura 6.13: Istogramma con mediana e IQR in evidenza

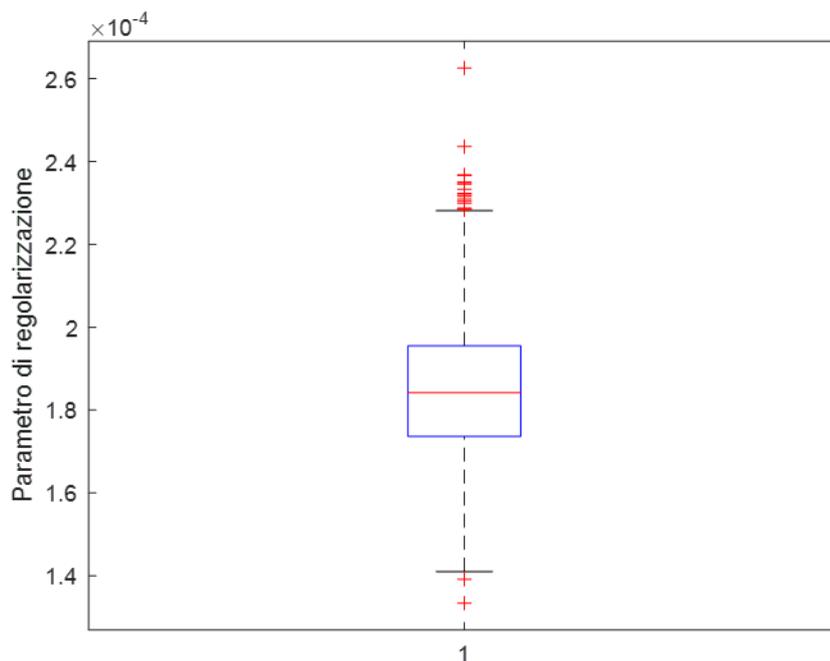


Figura 6.14: Box plot λ

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.3. Anche in questo caso la media e la mediana hanno valori simili tra loro e la distribuzione è quasi simmetrica.

Gli indici di dispersione, inoltre, sono relativamente piccoli e il metodo risulta essere robusto alla variabilità della perturbazione.

| Parametro di regolarizzazione λ | |
|---|----------|
| Valor medio | 1.85E-04 |
| Mediana | 1.84E-04 |
| Deviazione standard | 1.59E-05 |
| IQR | 2.19E-05 |

Tabella 6.3: Dati statistici di λ

Nella figura 6.15, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

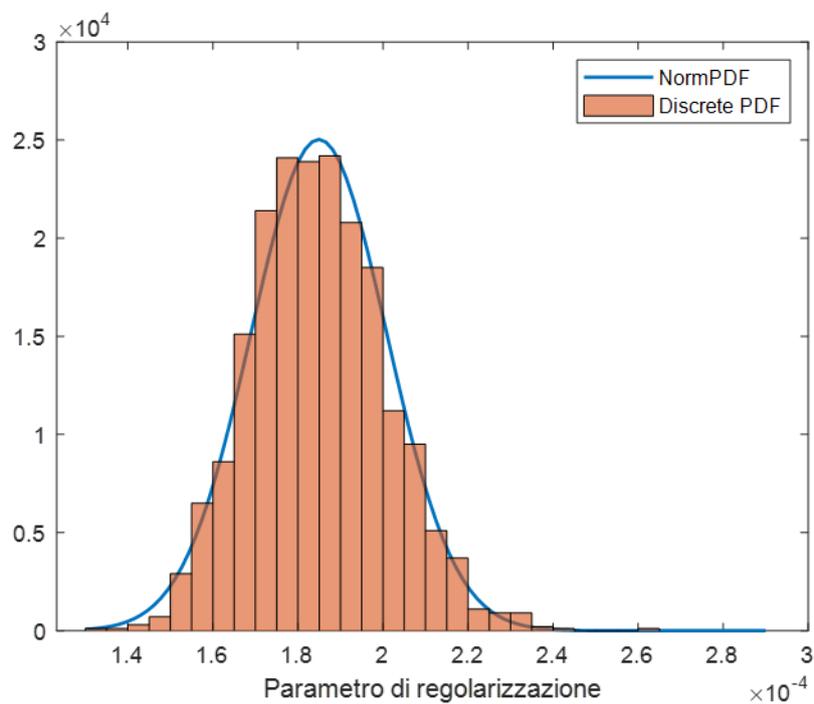


Figura 6.15: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che l'andamento della PDF della distribuzione discreta è circa uguale a quello della PDF della distribuzione normale.

In figura 6.16, infine, viene rappresentato un esempio di soluzione regolarizzata per un caso statisticamente probabile in cui il RMSE % è pari al 18.83 %.

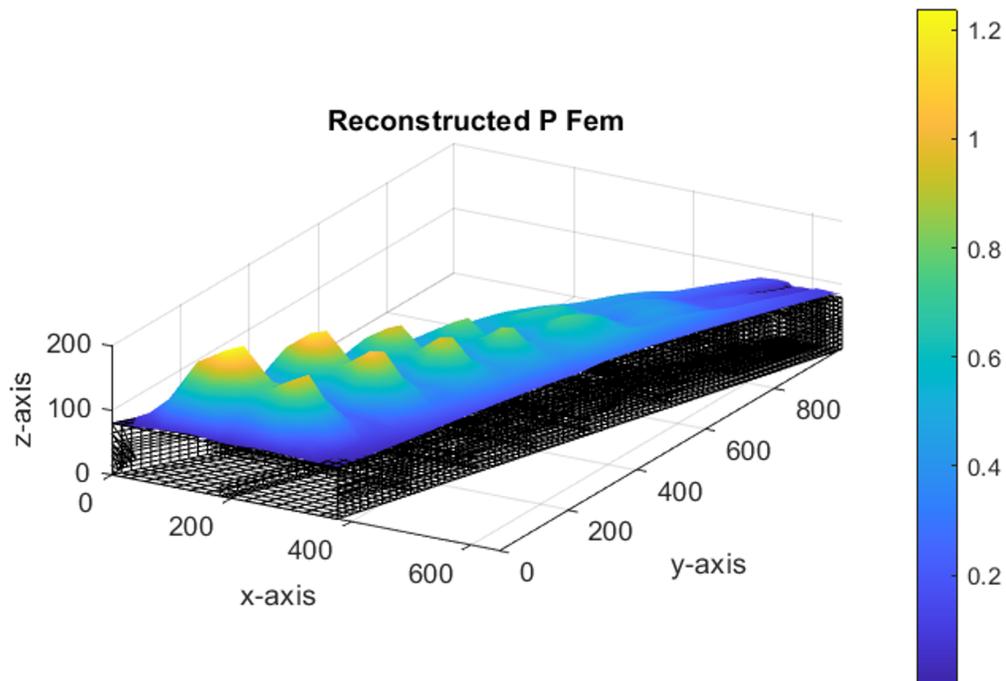


Figura 6.16: Soluzione regolarizzata tramite metodo di Tikhonov in forma standard

6.3.2 Metodo di Tikhonov del primo ordine

Per lo studio statistico di questo metodo vengono svolte 2000 analisi, ognuna di esse con un diverso vettore di perturbazione e il criterio di scelta del parametro di regolarizzazione considerato è il principio di discrepanza. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regolarizzazione λ .

RMSE %

I valori del RMSE % ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.4. Come si può notare, la media e la mediana hanno valori molto simili e, di conseguenza, la distribuzione è circa simmetrica.

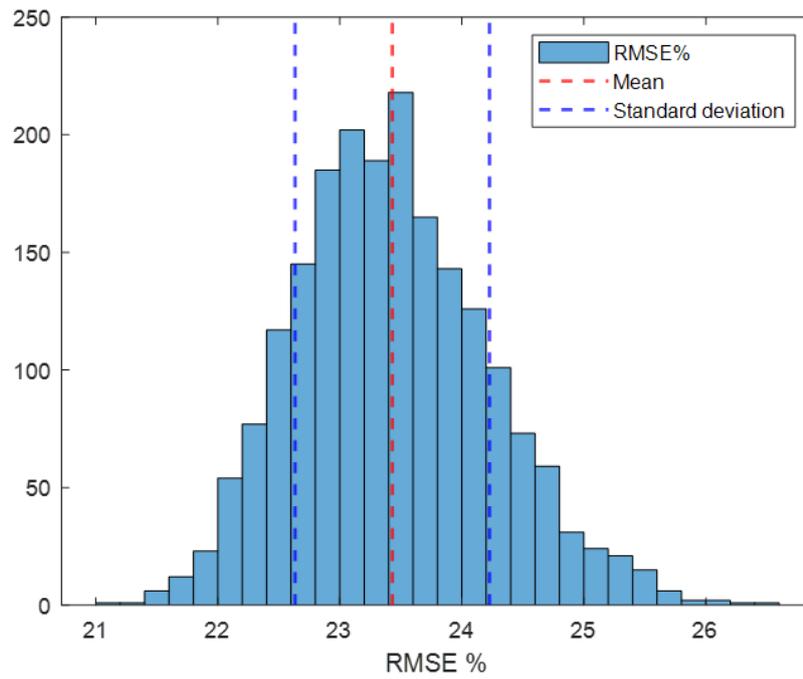


Figura 6.17: Istogramma con valor medio e deviazione standard in evidenza

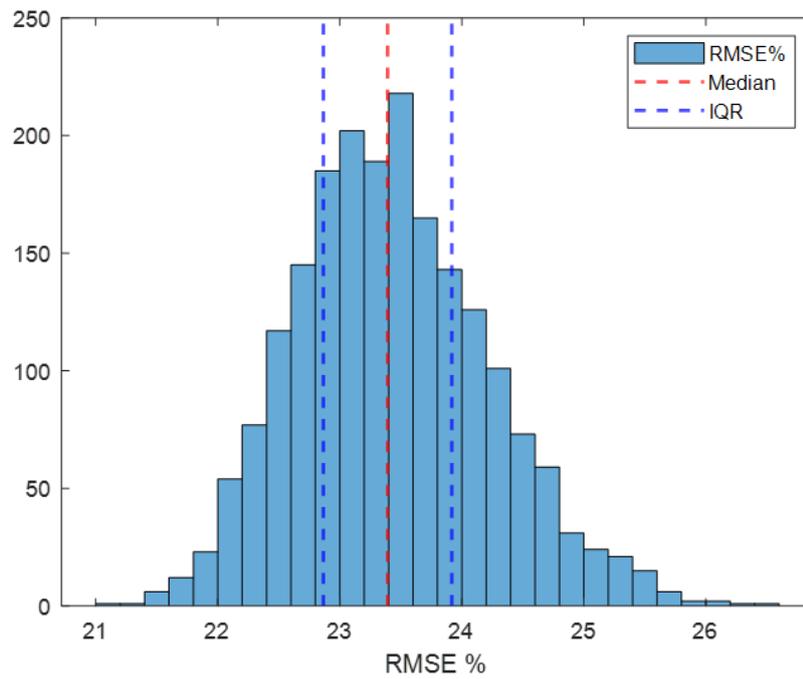


Figura 6.18: Istogramma con mediana e IQR in evidenza

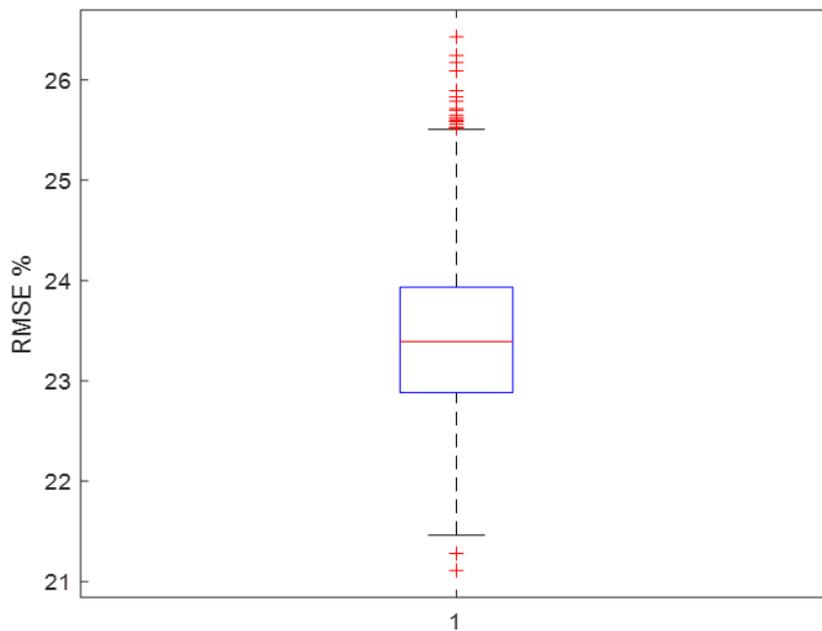


Figura 6.19: Box plot RMSE %

Gli indici di dispersione, invece, sono relativamente piccoli e ciò dimostra la robustezza di questo metodo alla variabilità della perturbazione.

| RMSE % | |
|---------------------|---------|
| Valor medio | 23.43 % |
| Mediana | 23.39 % |
| Deviazione standard | 0.797 % |
| IQR | 1.052 % |

Tabella 6.4: Dati statistici del RMSE %

Nella figura 6.20, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

Si può vedere, quindi, che la PDF della distribuzione discreta ha un andamento simile a quello della distribuzione normale.

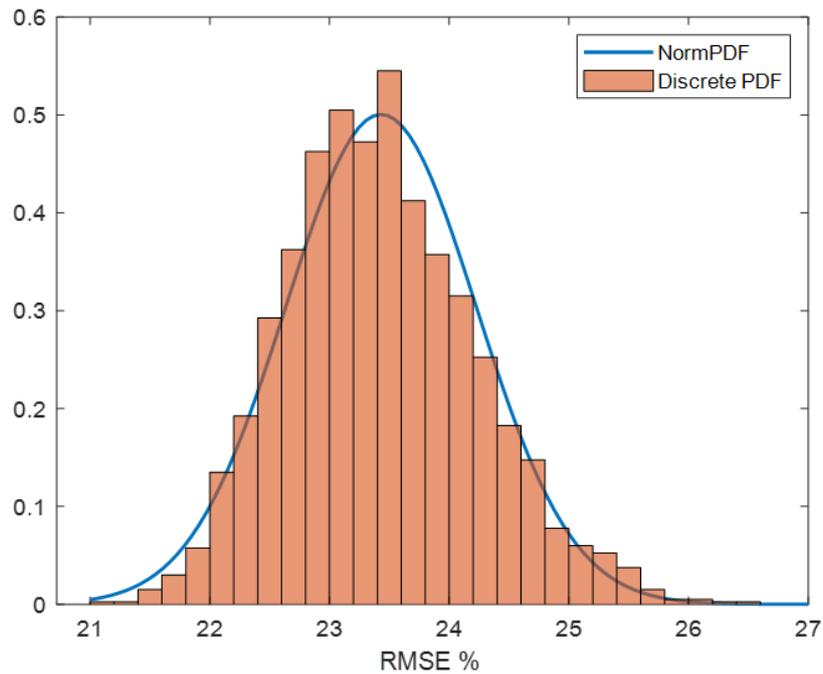


Figura 6.20: PDF della distribuzione discreta e PDF della distribuzione normale

Parametro di regolarizzazione λ

I valori di λ ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

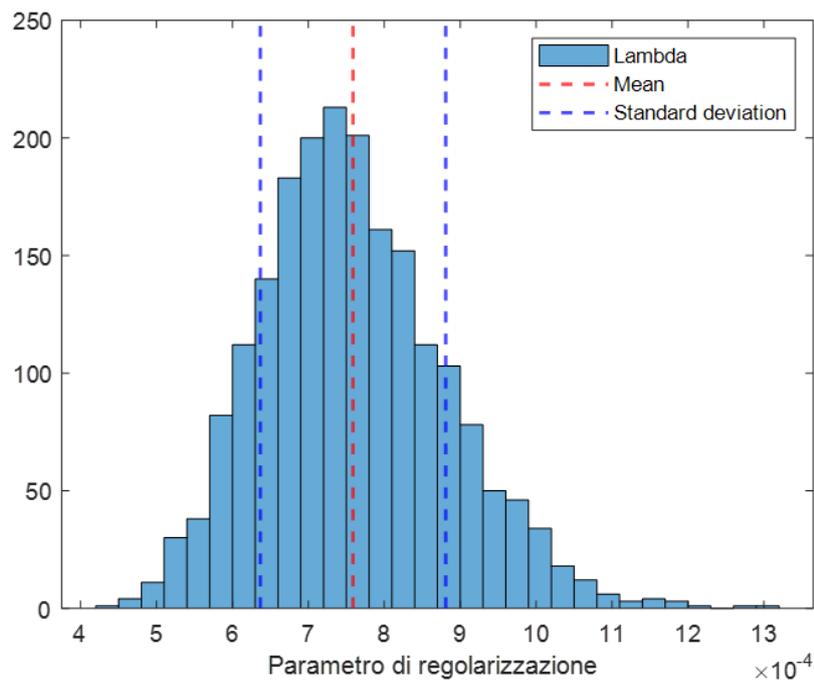


Figura 6.21: Istogramma con valor medio e deviazione standard in evidenza

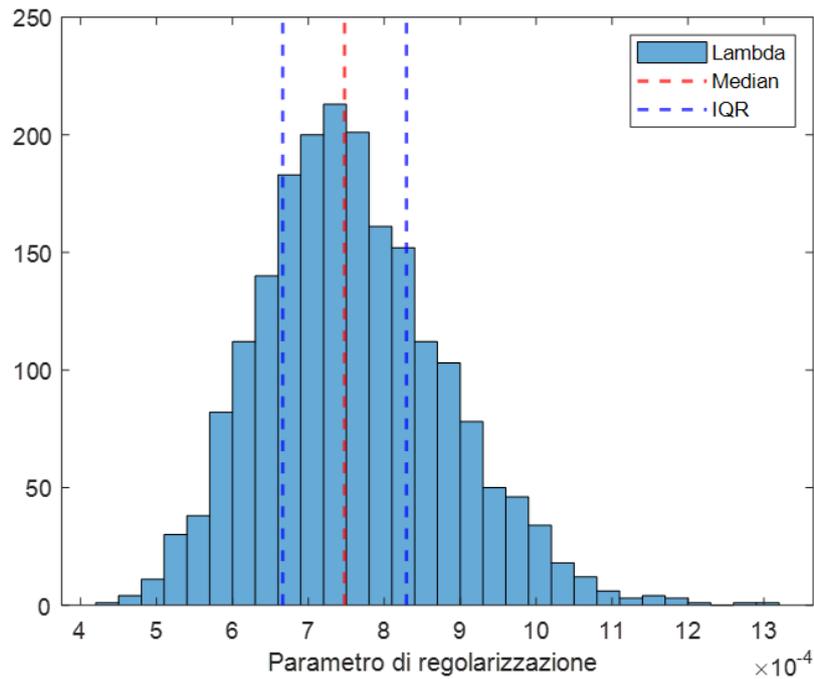


Figura 6.22: Istogramma con mediana e IQR in evidenza

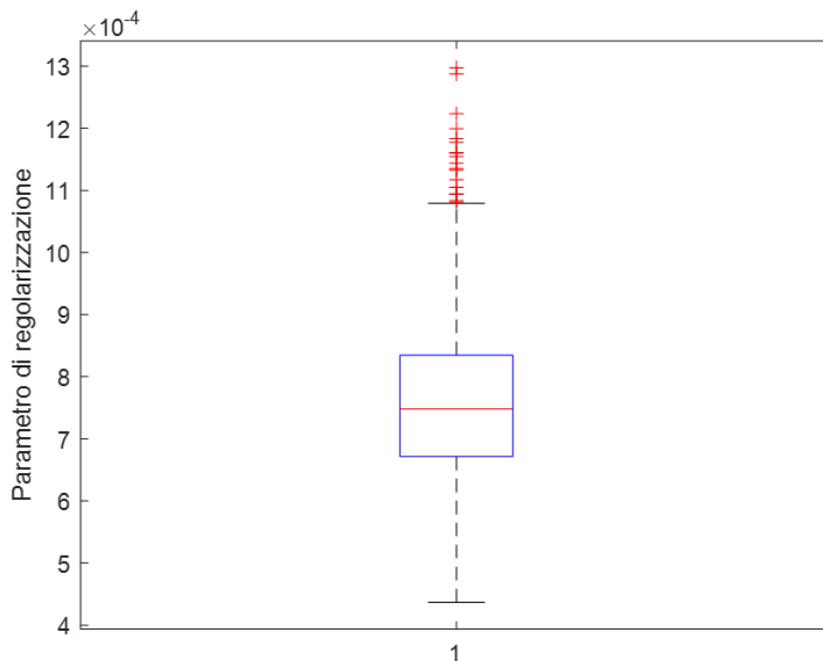


Figura 6.23: Box plot λ

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.5. Anche in questo caso la media e la mediana hanno valori vicini e, quindi, la distribuzione è quasi simmetrica. Gli indici di dispersione sono piccoli e il metodo risulta essere robusto alla variabilità della perturbazione.

| Parametro di regolarizzazione λ | |
|---|----------|
| Valor medio | 7.59E-04 |
| Mediana | 7.48E-04 |
| Deviazione standard | 1.22E-04 |
| IQR | 1.63E-04 |

Tabella 6.5: Dati statistici di λ

Nella figura 6.24, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

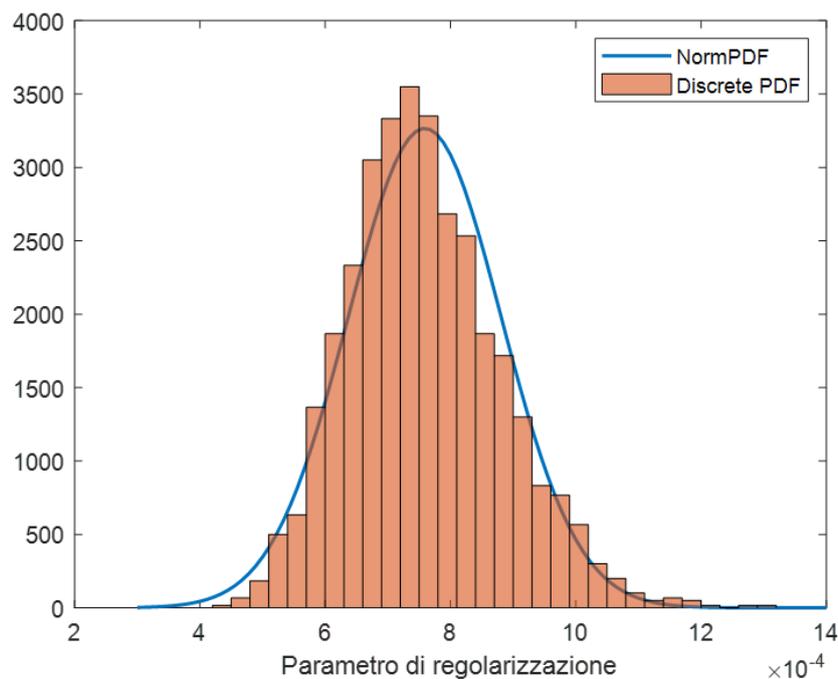


Figura 6.24: PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che l'andamento della PDF della distribuzione discreta è simile a quello della PDF della distribuzione normale, nonostante la probabilità nell'intorno del valor medio sia superiore.

In figura 6.25, infine, viene rappresentato un esempio di regolarizzazione della soluzione tramite questo metodo per un caso statisticamente probabile in cui il RMSE % è pari a 25.49 %.

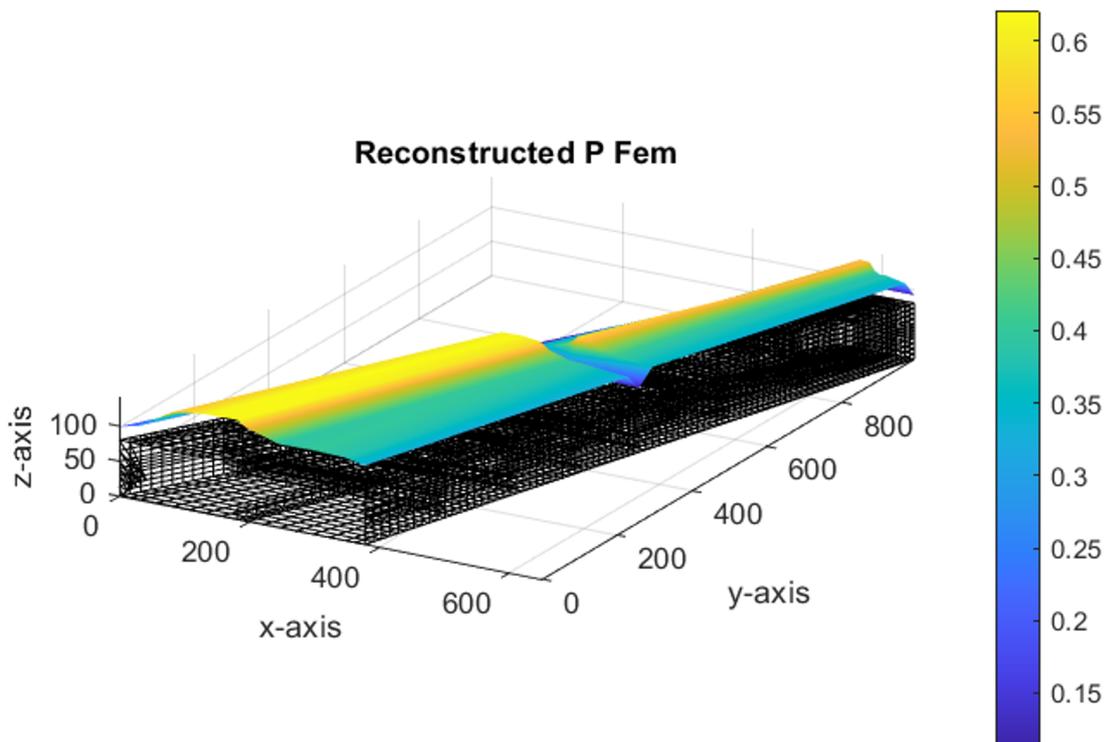


Figura 6.25: Regularizzazione della soluzione tramite metodo di Tikhonov del primo ordine

6.3.3 Metodo ART

Per lo studio statistico di questo metodo vengono svolte 200 analisi e non 2000 come nelle altre analisi statistiche a causa dell'elevato tempo computazionale che questo metodo richiede. Ogni analisi viene svolta con un diverso vettore di perturbazione creato in maniera random e il criterio di scelta del parametro di regularizzazione considerato è il metodo della curva L. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regularizzazione k .

RMSE %

I valori del RMSE % ottenuti nelle 200 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

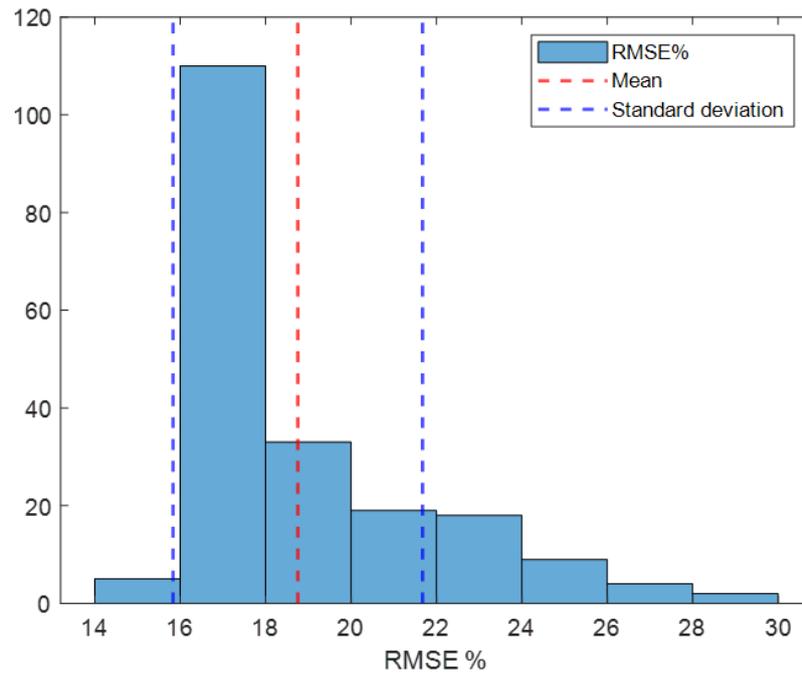


Figura 6.26: Istogramma con valor medio e deviazione standard in evidenza

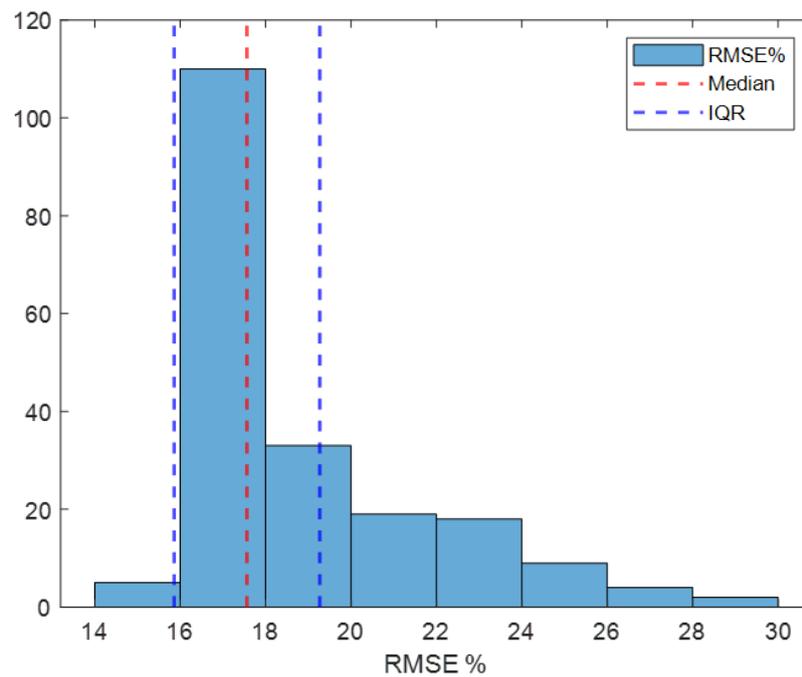


Figura 6.27: Istogramma con mediana e IQR in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.6. La media e la mediana hanno valori leggermente diversi e, di conseguenza, la distribuzione non è esattamente simmetrica. Gli indici di dispersione, inoltre, sono relativamente bassi e il

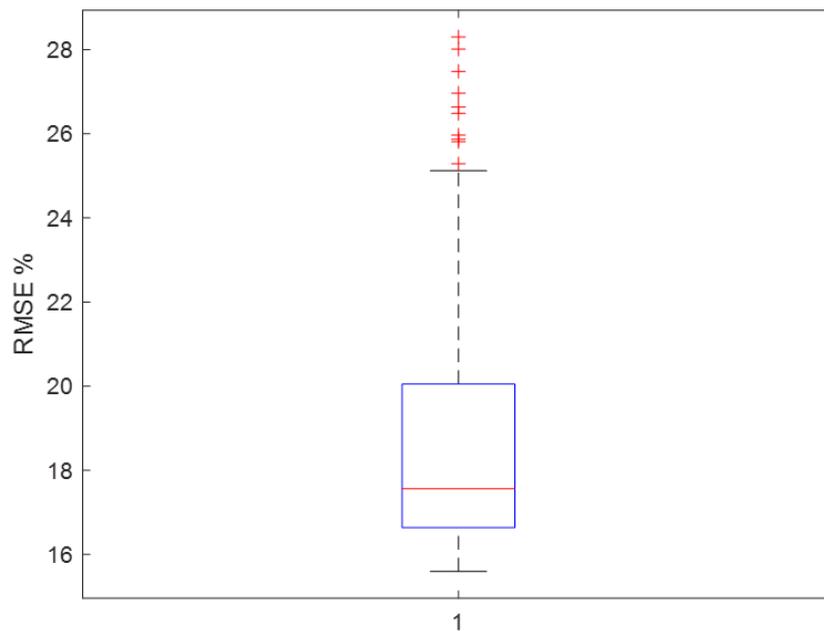


Figura 6.28: Box plot RMSE %

metodo si dimostra essere robusto alla variabilità della perturbazione.

| RMSE % | |
|---------------------|----------|
| Valor medio | 18.755 % |
| Mediana | 17.563 % |
| Deviazione standard | 2.920 % |
| IQR | 3.410 % |

Tabella 6.6: Dati statistici del RMSE %

Nella figura 6.29, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

Si può vedere, quindi, come la PDF della distribuzione discreta non abbia un andamento simmetrico, come detto in precedenza, e come non coincida con la PDF della distribuzione normale. In particolare, in questa distribuzione la probabilità che il RMSE % abbia un valore compreso tra il 16 % e il 18% è molto più elevata della probabilità che caratterizza la distribuzione normale.

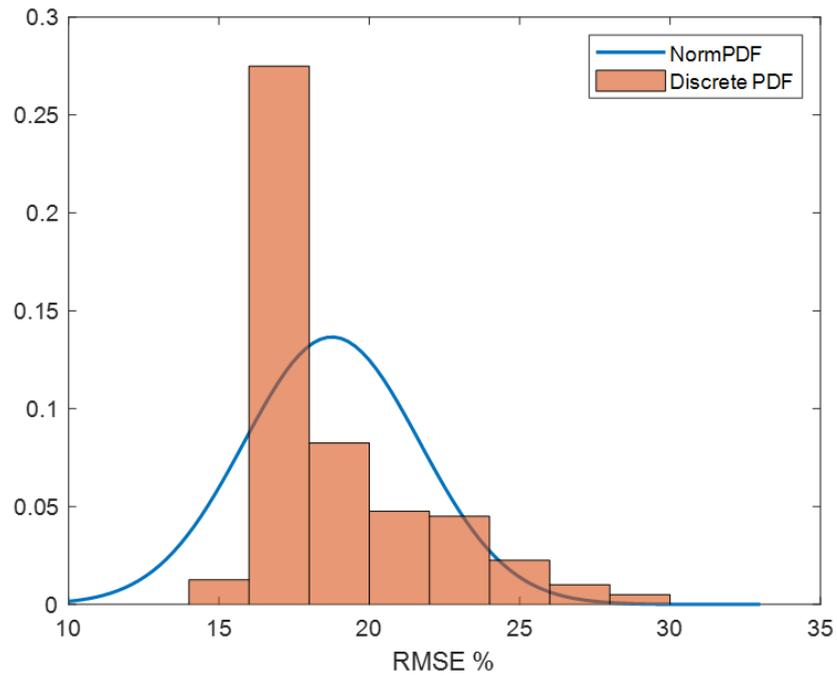


Figura 6.29: PDF della distribuzione discreta e PDF della distribuzione normale

Parametro di regolarizzazione k

I valori di k ottenuti nelle 200 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

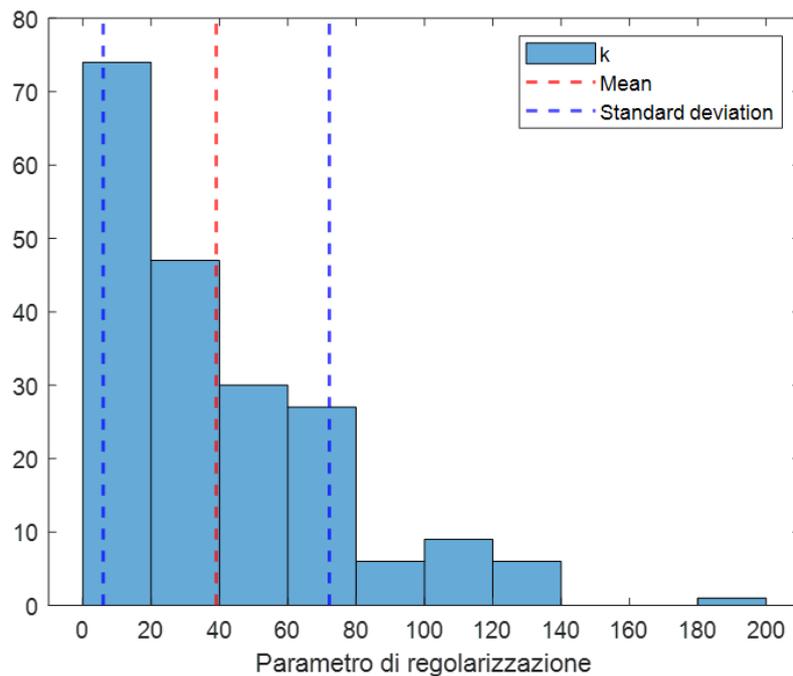


Figura 6.30: Istogramma con valor medio e deviazione standard in evidenza

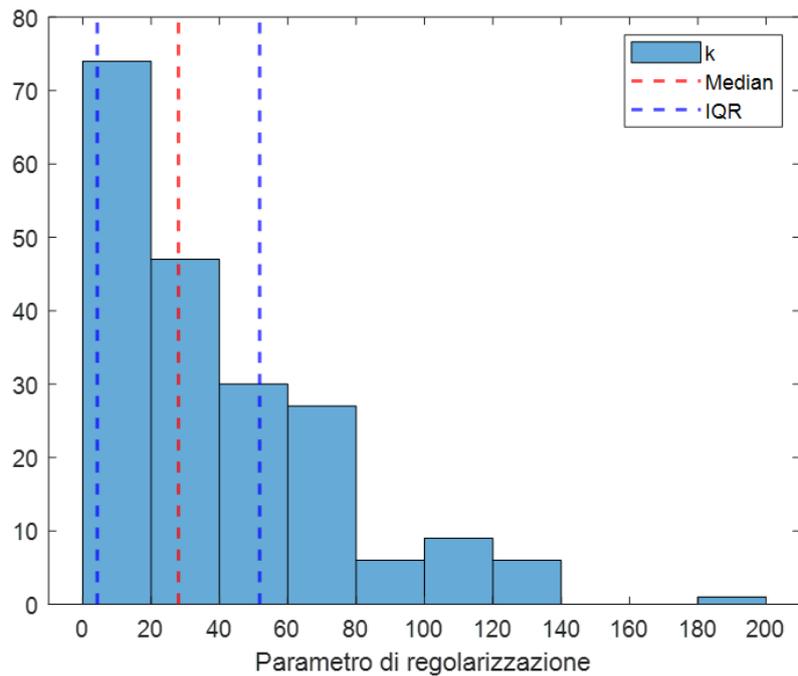


Figura 6.31: Istogramma con mediana e IQR in evidenza

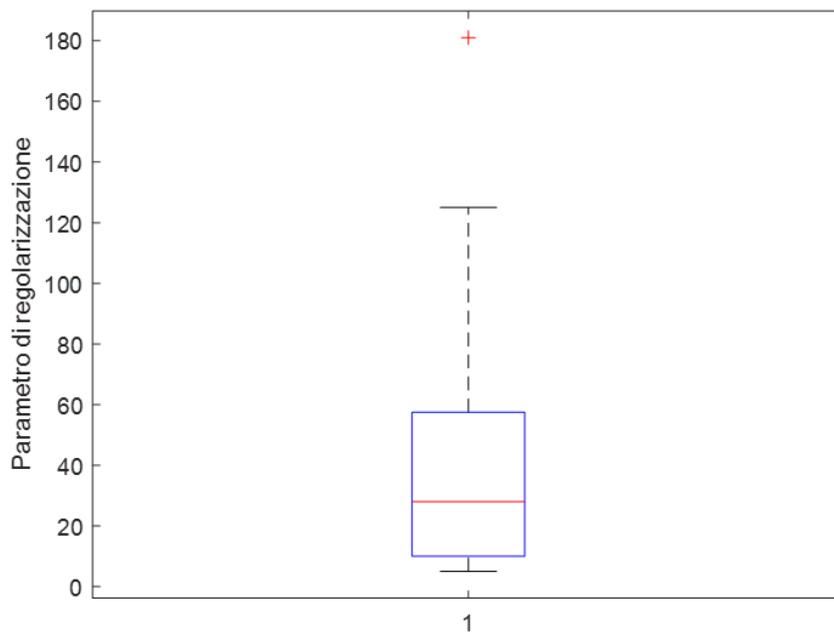


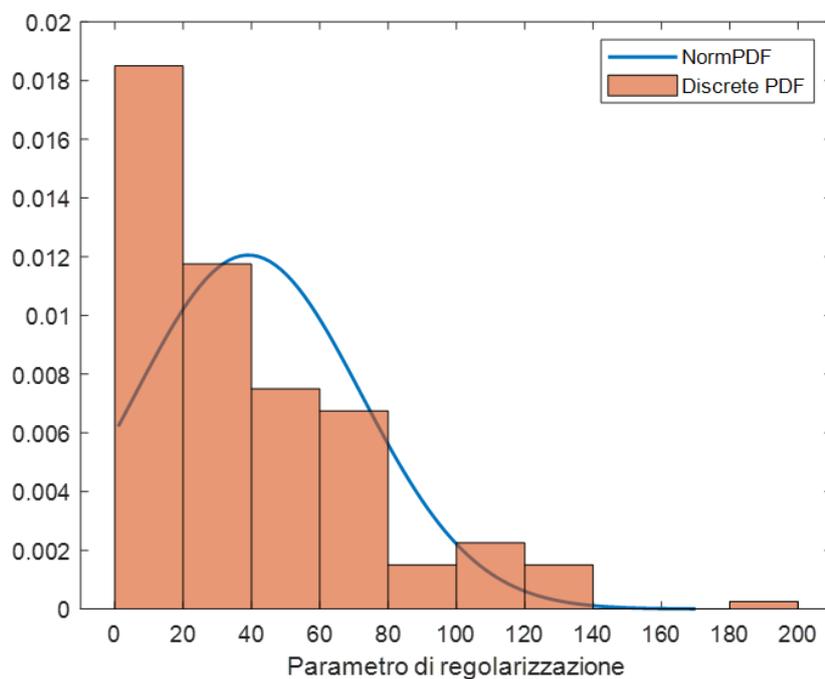
Figura 6.32: Box plot k

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.7. In questo caso la media e la mediana hanno valori abbastanza differenti e, come si può vedere dall'andamento dell'istogramma, la distribuzione non è simmetrica.

| Parametro di regolarizzazione k | |
|---------------------------------|-------|
| Valor medio | 39 |
| Mediana | 28 |
| Deviazione standard | 33.09 |
| IQR | 47.5 |

Tabella 6.7: Dati statistici di k

Nella figura 6.33, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

**Figura 6.33:** PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che la PDF della distribuzione discreta non è simmetrica e che, quindi, non coincide con quella della distribuzione normale.

In figura 6.34, infine, viene rappresentato un esempio di regolarizzazione della soluzione tramite questo metodo per un caso statisticamente probabile in cui il RMSE % è pari a 16.75 %.

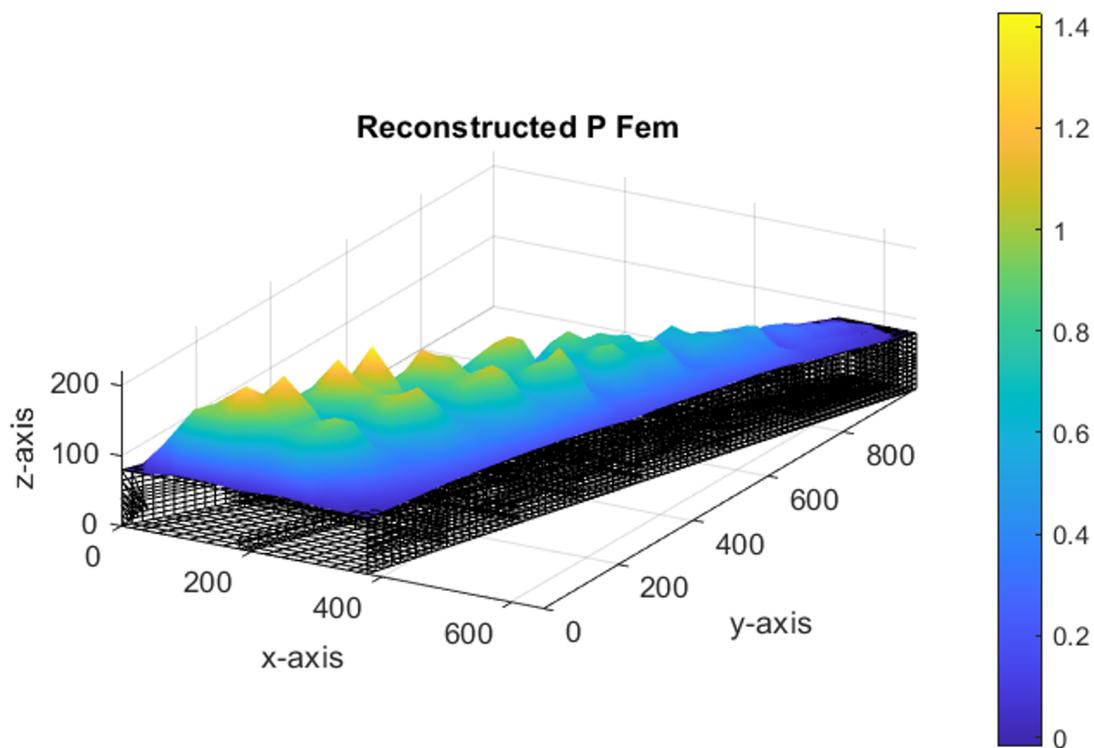


Figura 6.34: Regolarizzazione della soluzione tramite metodo ART

6.3.4 Metodo LSQR

Per lo studio statistico di questo metodo vengono svolte 2000 analisi, ognuna di esse con un diverso vettore di perturbazione e il criterio di scelta del parametro di regolarizzazione considerato è il metodo della curva L. Viene effettuato uno studio statistico sia del RMSE % e sia del parametro di regolarizzazione k .

RMSE %

I valori del RMSE % ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

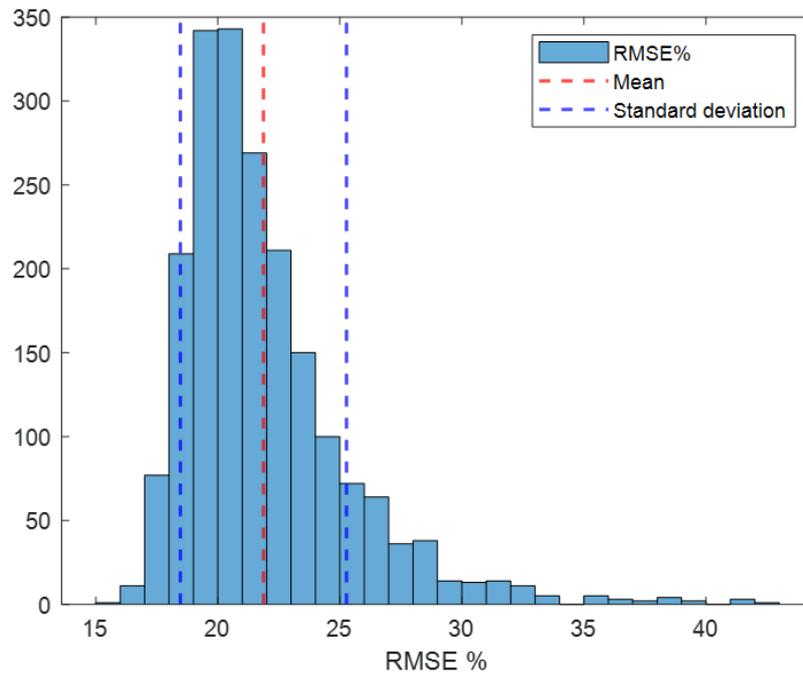


Figura 6.35: Istogramma con valor medio e deviazione standard in evidenza

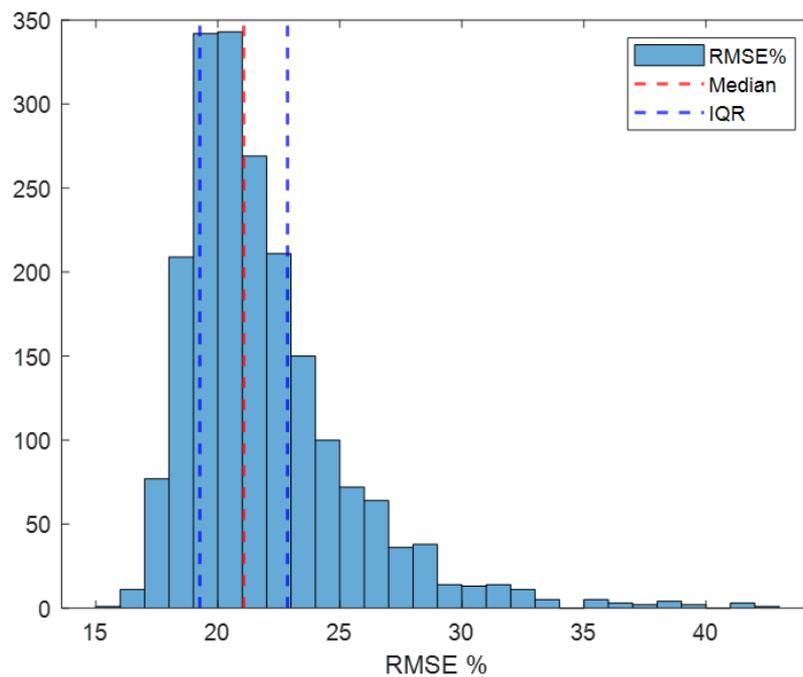


Figura 6.36: Istogramma con mediana e IQR in evidenza

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.8. Come si può vedere, la media e la mediana hanno valori leggermente diversi e, di conseguenza, la distribuzione non è esattamente simmetrica. Gli indici di dispersione, invece, sono relativamente

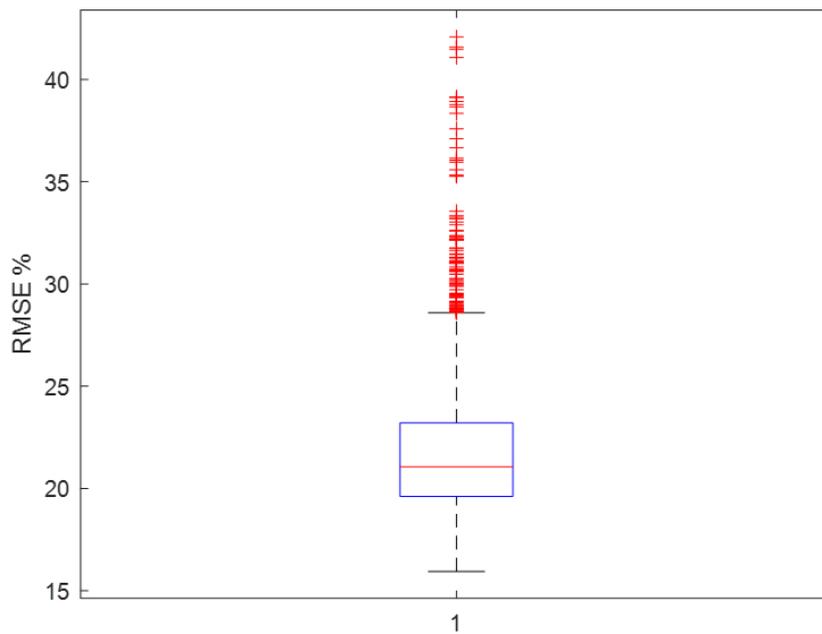


Figura 6.37: Box plot RMSE %

piccoli e ciò dimostra la robustezza del metodo alla variabilità della perturbazione.

| RMSE % | |
|---------------------|----------|
| Valor medio | 21.870 % |
| Mediana | 21.060 % |
| Deviazione standard | 3.403 % |
| IQR | 3.592 % |

Tabella 6.8: Dati statistici del RMSE %

Nella figura 6.38, infine, è rappresentata la funzione di densità di probabilità (PDF) della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

Si può vedere, quindi, come la PDF della distribuzione discreta abbia un andamento non coincidente con la PDF della distribuzione normale a causa della sua non simmetria. Inoltre, nella PDF della distribuzione discreta la probabilità associata ai valori vicini alla media e alla mediana è più elevata di quella che caratterizza la PDF della distribuzione normale.

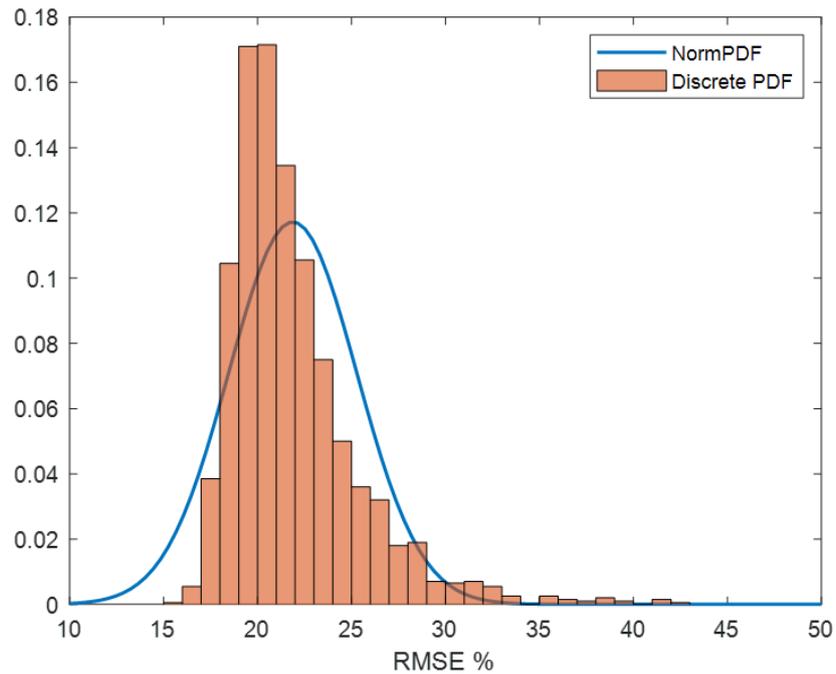


Figura 6.38: PDF della distribuzione discreta e PDF della distribuzione normale

Parametro di regolarizzazione k

I valori di k ottenuti nelle 2000 analisi sono raccolti e visualizzati in un istogramma e in un box plot.

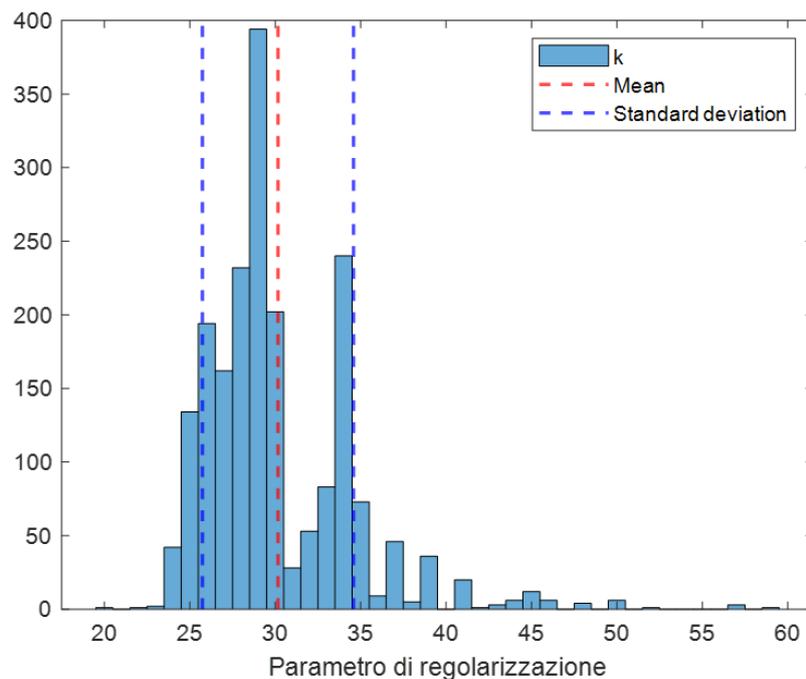


Figura 6.39: Istogramma con valor medio e deviazione standard in evidenza

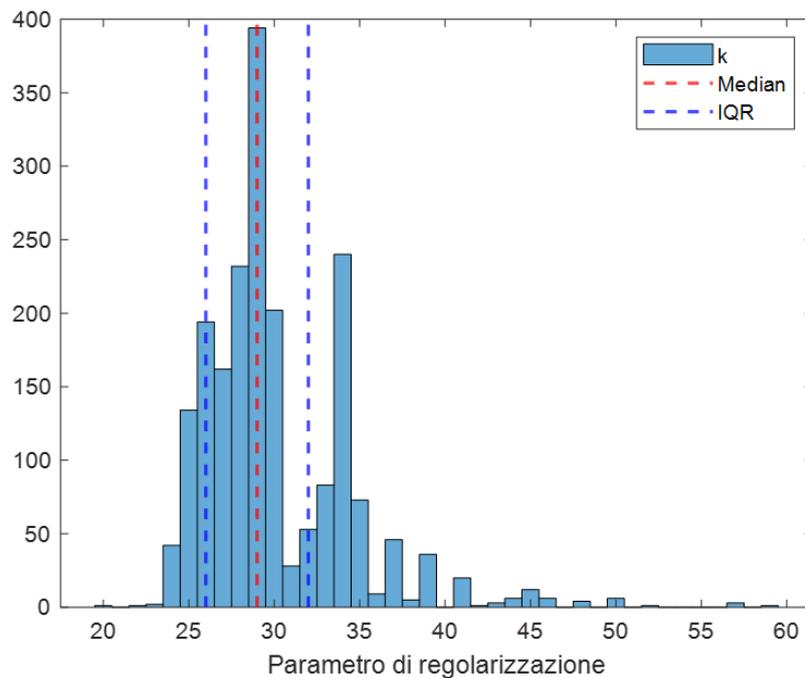


Figura 6.40: Istogramma con mediana e IQR in evidenza

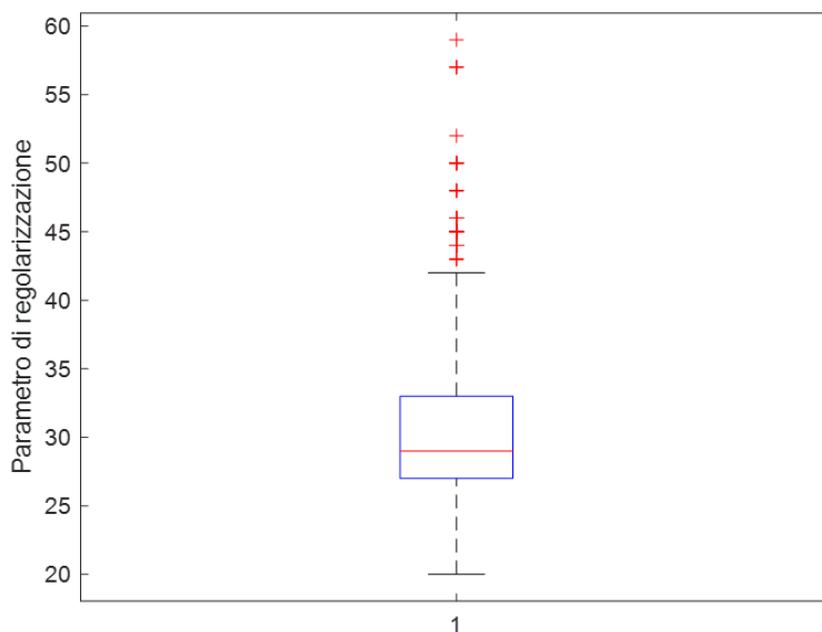


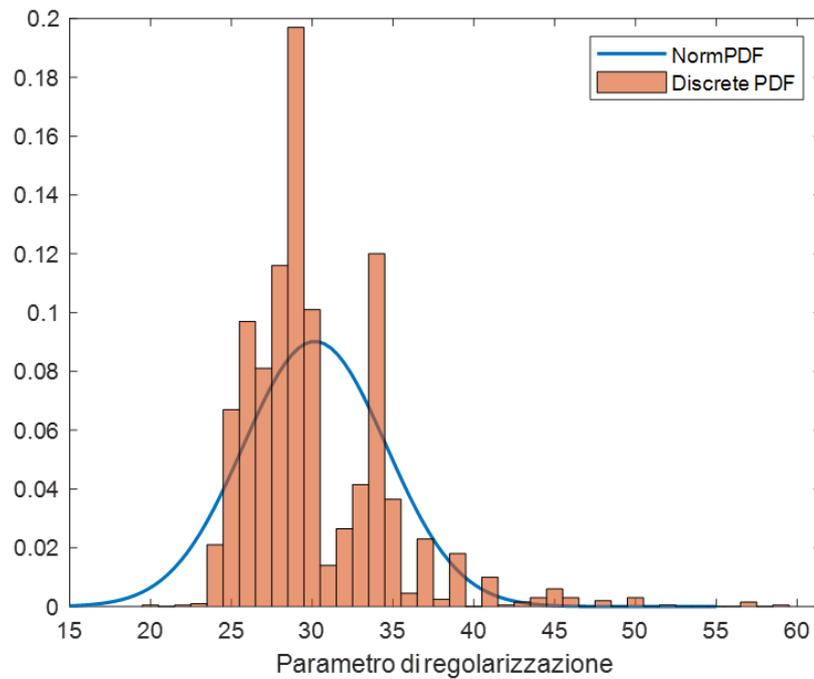
Figura 6.41: Box plot k

I valori degli indici di centralità e dispersione sono raccolti in tabella 6.9. In questo caso la media e la mediana hanno valori leggermente differenti e, come si può vedere dall'andamento dell'istogramma, la distribuzione non è esattamente simmetrica. Inoltre, si nota che la probabilità associata ai valori vicini a quelli rappresentanti gli indici di centralità è molto elevata.

| Parametro di regolarizzazione k | |
|-----------------------------------|-------|
| Valor medio | 30 |
| Mediana | 29 |
| Deviazione standard | 4.423 |
| IQR | 6 |

Tabella 6.9: Dati statistici di k

Nella figura 6.42, infine, è rappresentata la funzione di densità di probabilità della distribuzione discreta ottenuta normalizzando l'istogramma precedente e la funzione di densità di probabilità della distribuzione normale.

**Figura 6.42:** PDF della distribuzione discreta e PDF della distribuzione normale

Si può vedere, quindi, che la PDF della distribuzione discreta non è simmetrica e, in generale, non coincide con quella della distribuzione normale.

In figura 6.43 viene rappresentato un esempio di regolarizzazione della soluzione tramite questo metodo per un caso statisticamente probabile in cui il RMSE % è pari a 22.6 %.

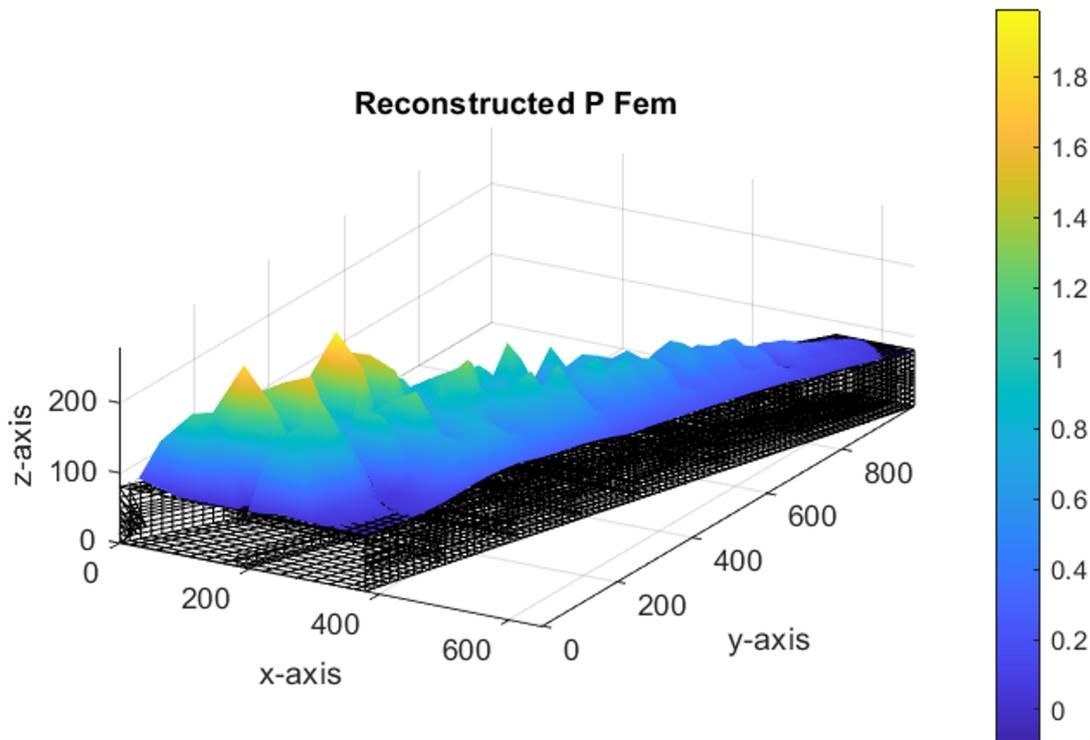


Figura 6.43: Regolarizzazione della soluzione tramite metodo LSQR

6.3.5 Confronto dei risultati

In tabella 6.10, infine, sono riassunti i risultati ottenuti dall'analisi statistica relativi al RMSE % e al parametro di regolarizzazione in termini di media, mediana, deviazione standard e scarto interquartile.

| | RMSE % | | | | $\lambda \circ k$ | | | |
|---------------|------------|---------|-------------|---------|-------------------|----------|------------|----------|
| | Val. Medio | Mediana | Dev. Stand. | IQR | Val. Medio | Mediana | Dev. Stand | IQR |
| Tikh | 16.69 % | 16.68 % | 0.353 % | 0.498 % | 1.85E-04 | 1.84E-04 | 1.59E-05 | 2.19E-05 |
| Tikh-1 | 23.43 % | 23.39 % | 0.797 % | 1.053 % | 7.59E-04 | 7.48E-04 | 1.22E-04 | 1.63E-04 |
| ART | 18.75 % | 17.56 % | 2.92 % | 3.41 % | 39 | 28 | 33.09 | 47.5 |
| LSQR | 21.87 % | 21.06 % | 3.40 % | 3.59 % | 30 | 29 | 4.4233 | 6 |

Tabella 6.10: Riassunto dei risultati

Come si può vedere, in questa applicazione finale il metodo di regolarizzazione che risulta essere il più accurato è quello di Tikhonov in forma standard e non quello del primo ordine come nel caso della piastra. Tra i metodi iterativi, invece, il Metodo ART risulta essere più accurato del Metodo LSQR, mentre nel caso della piastra quest'ultimo risulta essere leggermente migliore.

Conclusioni

La presente tesi ha riguardato lo studio e il confronto di alcune metodologie per la ricostruzione del carico agente su una struttura a partire da misure di deformazione e per la regolarizzazione della soluzione nel caso di misure di deformazione contenenti errore. Dopo aver introdotto alcune metodologie per la ricostruzione del carico, sono stati utilizzati l'approccio inverso e l'approccio delle reti neurali per ricostruire il carico agente su un semplice modello di piastra. Successivamente, sono state riportate alcune informazioni generali sulla natura mal posta dei problemi inversi e sul problema della regolarizzazione e sono state richiamate alcune metodologie per la regolarizzazione della soluzione e per la scelta del parametro di regolarizzazione. Anche in questo caso esse sono state applicate al semplice modello di piastra e, in particolare, all'approccio inverso. Infine, sulla base dei risultati ottenuti dalle analisi svolte sul modello di piastra, sono state selezionate le migliori metodologie per la ricostruzione e per la regolarizzazione e sono state applicate ad un modello di cassone alare. Le prestazioni delle varie metodologie sono state confrontate tramite la valutazione della radice dell'errore quadratico medio percentuale.

Dalla serie di analisi e confronti effettuati, risulta che il miglior metodo per la ricostruzione del carico è l'approccio inverso. Per quanto riguarda la regolarizzazione, invece, i risultati migliori sono stati forniti dal metodo di Tikhonov in forma standard e il metodo di Tikhonov del primo ordine. In tutte le analisi relative alla regolarizzazione, inoltre, i valori più accurati dei parametri di regolarizzazione sono stati ottenuti tramite il principio di discrepanza e il metodo della curva L. Ciò che è risultato dalle analisi, inoltre, è l'elevata sensibilità dell'approccio inverso all'errore nelle misure di deformazione; tuttavia, è stato mostrato che le varie metodologie di regolarizzazione sono in grado di sopperire a questo problema.

D'altro canto, le prestazioni delle reti neurali dipendono fortemente dalla quantità e dalla tipologia dei dati di apprendimento. Ciò che non è stato effettuato in questo studio e che potrebbe costituire un approfondimento è addestrare la rete neurale aggiungendo dati di apprendimento che includono errori di misura: in questo modo si otterrebbe una rete neurale artificiale robusta all'errore di deformazione in grado di fornire una maggiore accuratezza nella ricostruzione del carico.

Bibliografia

- [1] Toshiya Nakamura, Hirotaka Igawa e Atsushi Kanda. “Inverse identification of continuously distributed loads using strain data”. In: *Aerospace Science and Technology* 23.1 (2012), pp. 75–84.
- [2] Marco Esposito, Marco Gherlone e Pier Marzocca. “External loads identification and shape sensing on an aluminum wing box: An integrated approach”. In: *Aerospace Science and Technology* 114 (2021), p. 106743.
- [3] S Shkarayev, R Krashantisa e A Tessler. “An inverse interpolation method utilizing in-flight strain measurements for determining loads and structural response of aerospace vehicles”. In: (2004).
- [4] Cameron Coates, Priya Thamburaj e Cassandra Kim. “An inverse method for selection of Fourier coefficients for flight load identification”. In: *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. 2005, p. 2183.
- [5] Cameron W Coates e Priya Thamburaj. “Inverse method using finite strain measurements to determine flight load distribution functions”. In: *Journal of aircraft* 45.2 (2008), pp. 366–370.
- [6] X Cao, Y Sugiyama e Y Mitsui. “Application of artificial neural networks to load identification”. In: *Computers & structures* 69.1 (1998), pp. 63–78.
- [7] PM Trivailo e CL Carn. “The inverse determination of aerodynamic loading from structural response data using neural networks”. In: *Inverse Problems in Science and Engineering* 14.4 (2006), pp. 379–395.
- [8] Per Christian Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.
- [9] Martin Fuhry e Lothar Reichel. “A new Tikhonov regularization method”. In: *Numerical Algorithms* 59.3 (2012), pp. 433–445.
- [10] Hong-Xia Dou et al. “Signal Restoration Combining Modified Tikhonov Regularization and Preconditioning Technology”. In: *IEEE Access* 5 (2017), pp. 24275–24283.
- [11] Xiao-Juan Yang e Li Wang. “A modified Tikhonov regularization method”. In: *Journal of Computational and Applied Mathematics* 288 (2015), pp. 180–192.
- [12] Mansoor Rezghi e Seyed Mohammad Hosseini. “Lanczos based preconditioner for discrete ill-posed problems”. In: *Computing* 88.1 (2010), pp. 79–96.
- [13] Christopher C Paige e Michael A Saunders. “LSQR: An algorithm for sparse linear equations and sparse least squares”. In: *ACM Transactions on Mathematical Software (TOMS)* 8.1 (1982), pp. 43–71.

Ringraziamenti

Vorrei dedicare quest'ultima parte della tesi a chi ha contribuito alla realizzazione di essa e, in generale, mi ha permesso di raggiungere questo traguardo.

Un ringraziamento particolare va al mio relatore prof. Marco Gherlone per avermi seguito con grande disponibilità durante l'intera stesura della tesi a partire dalla scelta dell'argomento e, più in generale, per tutti gli insegnamenti ricevuti in questi anni.

Grazie anche al mio correlatore Marco Esposito per la grande disponibilità mostrata nell'aiutarmi durante l'utilizzo dei codici necessari ad effettuare le analisi.

Ringrazio infinitamente i miei genitori, mia sorella, i miei nonni e i miei zii per i loro insegnamenti e il loro infinito supporto durante l'elaborazione della tesi e, più in generale, durante gli anni di studi.

Grazie, infine, ai miei amici di sempre per il loro continuo supporto e per, con alcuni di loro, aver affrontato insieme il percorso universitario.