



Università degli Studi di Cagliari

DOTTORATO DI RICERCA

Ingegneria Industriale

Ciclo XXXIV

TITOLO TESI

Tecniche di ottimizzazione per la stima e il monitoraggio
di sistemi elettrici di potenza

Settore scientifico disciplinare di afferenza

ING-INF/07

Presentata da:

Antonio Vincenzo Solinas

Tutor

Prof.ssa Sara Sulis

Esame finale anno accademico 2020/2021
Tesi discussa nella sessione d'esame aprile 2022

Questa Tesi può essere utilizzata, nei limiti stabiliti dalla normativa vigente sul Diritto d'Autore (Legge 22 aprile 1941 n. 633 e succ. modificazioni e articoli da 2575 a 2583 del Codice civile) ed esclusivamente per scopi didattici e di ricerca; è vietato qualsiasi utilizzo per fini commerciali. In ogni caso tutti gli utilizzi devono riportare la corretta citazione delle fonti. La traduzione, l'adattamento totale e parziale, sono riservati per tutti i Paesi. I documenti depositati sono sottoposti alla legislazione italiana in vigore nel rispetto del Diritto di Autore, da qualunque luogo essi siano fruiti.

Indice

Introduzione	iv
1 Tecniche di ottimizzazione per problemi inversi mal posti	1
1.1 Problemi inversi mal posti	1
1.2 Norme	5
1.2.1 Norme matriciali	8
1.3 Norme ed informazione	11
1.3.1 Errore di approssimazione o residuo	11
1.3.2 Termine di regolarizzazione	16
1.3.3 Combinazione di norme ed informazione a priori	19
1.3.4 LASSO e Tikhonov, interpretazione Bayesiana, legame con MLE e MAP	22
1.4 Problemi sotto-determinati e <i>Compressive Sensing</i>	26
1.4.1 Introduzione	26
1.4.2 Sparsità e algoritmi	27
1.4.3 Caratteristiche matrice di misura	28
1.5 Regolarizzazione in ℓ_2	30
1.5.1 Decomposizione a valori singolari, SVD e GSVD	31
1.5.2 Analisi SVD	33
1.5.3 Filtraggio	35
1.5.4 Regolarizzazione di Tikhonov	35
1.5.5 Regolarizzazione di Tikhonov generalizzata	36
1.5.6 Scelta del parametro di regolarizzazione	37
1.5.7 <i>Discrepancy Principle</i> per TSVD e Tikhonov	39
1.6 Stima dell'energia dell'errore di misura	41
1.6.1 Valutazione del limite dell'energia dell'errore di misura	41
1.6.2 Sbiancamento, <i>Whitening Transformation</i>	42
1.6.3 Proposta di una particolare matrice di sbiancamento e proprietà dell'errore sbiancato	45

Appendice	47
1.A $\mathcal{P}_{e_1/r_u}, \min_{\mathbf{x}} \ \mathbf{Ax} - \mathbf{b}\ _1$ riscritto come LP	47
1.B $\mathcal{P}_{e_\infty/r_u}, \min_{\mathbf{x}} \ \mathbf{Ax} - \mathbf{b}\ _\infty$ riscritto come LP	47
1.C Esempi di matrici \mathbf{L}	48
1.D $\mathcal{P}_{r_0/e=0}, (P0) \min_{\mathbf{x}} \ \mathbf{x}\ _0$ s.t. $\mathbf{Ax} = \mathbf{b}$ riscritto come MILP	49
1.E $\mathcal{P}_{r_1/e=0}, (P1) \min_{\mathbf{x}} \ \mathbf{x}\ _1$ s.t. $\mathbf{Ax} = \mathbf{b}$, Basis Pursuit, riscritto come LP	50
1.F <i>Discrepancy Principle</i> per Tikhonov generalizzato	50
2 Identificazione delle principali sorgenti armoniche nelle reti di distribuzione	52
2.1 Introduzione	52
2.2 Formulazione dell’algoritmo di stima	55
2.2.1 Modello di stima delle sorgenti armoniche	55
2.2.2 Algoritmi per il <i>Compressive Sensing</i>	56
2.3 Valutazione dell’energia dell’errore di misura	58
2.3.1 <i>Whitening Transformation</i>	59
2.3.2 Matrice di sbiancamento proposta	62
2.3.3 Valutazione dei limiti dell’energia dell’errore	64
2.4 HSoE: prove e risultati	66
2.4.1 Configurazione di test	66
2.4.2 Risultati delle prove	69
2.5 Conclusioni	73
Appendice	76
2.A Modello della rete di test a 49 nodi	76
3 Stima dei parametri di linea e compensazione degli errori siste- matici in reti di trasmissione	78
3.1 Introduzione	78
3.2 Formulazione del modello di stima	81
3.2.1 Modello di misura e vincoli di rete	81
3.2.2 WLS aumentato inquadrate come problema di regolariz- zazione di Tikhonov	86
3.3 Gestione di casi operativi multipli	87
3.3.1 Analisi dell’incertezza della stima	87
3.3.2 Metodo proposto e regolarizzazione	90
3.4 Prove e Risultati	94
3.4.1 Stima su un insieme di rami selezionati	96
3.4.2 Stima simultanea sull’intera rete	98
3.5 Conclusioni	102

Appendice	104
3.A Derivazione della matrice di misura e della rappresentazione dell'incertezza	104
Conclusioni	106
Bibliografia	108
Lista delle figure	117
Lista delle tabelle	119
Glossario	120
Elenco dei simboli	122

Introduzione

Le moderne reti elettriche di trasmissione e di distribuzione stanno attraversando profondi cambiamenti che possono essere visti come un vero e proprio cambio di paradigma. Produzione, trasmissione, distribuzione e consumo stanno diventando sistemi altamente complessi nei quali: viene meno la centralità della produzione a causa della crescente installazione di generazione distribuita (benché il ruolo delle grandi centrali di produzione resti ad oggi preponderante), viene meno la unidirezionalità del flusso energetico e la programmabilità della produzione di energia (essenzialmente a causa della penetrazione sempre più significativa delle fonti di energia rinnovabile), si diffondono carichi di natura non lineare ed emerge la figura del *prosumer* ovvero di un soggetto che è allo stesso tempo produttore e consumatore di energia.

Questa evoluzione comporta problemi nella qualità dell'energia, comunemente identificati con l'espressione *Power Quality* (PQ), e implica un'amplificazione dell'importanza degli strumenti di gestione e controllo necessari a governare la complessità della rete. Tali strumenti basano le proprie decisioni su una conoscenza della rete che deriva dagli strumenti di stima e di monitoraggio della rete stessa. Gli strumenti di monitoraggio e di stima hanno l'obiettivo di ricostruire il cosiddetto stato, ovvero le grandezze che caratterizzano lo stato operativo del sistema, al fine di ottimizzare la gestione, prevenire un malfunzionamento o identificarne la causa mitigando o eliminando il disservizio, a partire da misure e pseudo-misure (dati storici, informazioni a priori, etc.), caratterizzate dalla propria incertezza. Si tratta quindi di un classico problema inverso: a partire dalle conseguenze (misure) si vuole ricostruire la causa. I problemi inversi sono spesso mal posti, ovvero non hanno un'unica soluzione che dipende con continuità dai dati, e necessitano di una riformulazione per essere affrontati in modo adeguato.

In questa tesi viene esplorato un approccio, basato su tecniche di ottimizzazione vincolata o con penalizzazione, ad alcuni rilevanti problemi di stima e monitoraggio di sistemi elettrici di potenza. L'obiettivo è quello di ottimizzare le prestazioni degli algoritmi di stima e monitoraggio attraverso l'integrazione

nelle formulazioni di informazioni a priori sul dominio in esame e sullo stato da ricostruire, sfruttando le conoscenze disponibili sull'errore di misura. Le formulazioni di ottimizzazione considerate sono composte da due termini, l'errore di approssimazione e il termine di regolarizzazione, attraverso i quali è possibile integrare le informazioni a priori nel modello di stima. Il termine di approssimazione o residuo valuta, in una data norma, quanto la soluzione trovata si adatti alle misure a disposizione, quindi ai dati, e viene per questo anche chiamato *fitting term*. Il termine di regolarizzazione tiene conto delle informazioni a priori sulla soluzione da ricostruire misurando, in una data norma, la soluzione stessa.

I problemi di stima relativi ai sistemi elettrici di potenza affrontati in questa tesi sono principalmente due, il primo è l'individuazione delle principali sorgenti di inquinamento armonico in una rete di distribuzione, il secondo è la stima dei parametri di linea in una rete di trasmissione e la compensazione degli errori sistematici della catena di misura.

La tesi è organizzata come segue:

- Il Capitolo 1 affronta alcuni aspetti teorici dell'approccio ai problemi inversi mal posti tramite tecniche di ottimizzazione. Dopo un richiamo alle norme, vettoriali e matriciali, e ad alcune delle loro proprietà, viene affrontato il tema della modellazione delle informazioni attraverso le norme e la loro composizione, dandone infine una interpretazione in termini statistico-probabilistici. Viene mostrato come le formulazioni composte da due termini e sfruttate nelle due applicazioni presentate nei capitoli successivi, ovvero la regolarizzazione di Tikhonov (o *Ridge regression*) e la minimizzazione in norma uno (ℓ_1) con vincolo quadratico sull'errore di approssimazione (*LASSO regression* in formulazione a norma minima), siano strettamente connesse tra di loro. Le due formulazioni sono infatti legate alla massima probabilità a posteriori (MAP) con differenti assunzioni a priori sulla soluzione da trovare. Vengono quindi introdotte due tecniche matematiche specifiche: il *Compressive Sensing* (CS), volto a ricostruire segnali sparsi da sistemi sotto-determinati, e la regolarizzazione di Tikhonov, volta a trovare soluzioni approssimate (o filtrate) per problemi mal condizionati e pertanto non affrontabili in modo diretto. Il Capitolo si conclude con la presentazione della stima dell'energia dell'errore di misura, fondamentale per sintonizzare il problema di ottimizzazione al sistema di misura.
- Il Capitolo 2 affronta il problema dell'individuazione delle sorgenti armoniche in un sistema di distribuzione. Il problema di stima viene modellato come la ricerca di generatori armonici di corrente in corrispondenza

dei principali carichi inquinanti ed è quindi inquadrato come un problema sparso ed affrontato con il CS, in particolare utilizzando algoritmi di minimizzazione in ℓ_1 . Vengono discussi gli aspetti teorici coinvolti nella valutazione delle incertezze di misura e viene presentata una matrice di sbiancamento che, consentendo il recupero di informazioni sulle distribuzioni degli errori di misura, permette la stima dei relativi limiti, fattore essenziale per la sintonizzazione dell'algoritmo al problema e al sistema di misura specifici. L'efficacia della soluzione presentata, basata sull'integrazione nella formulazione di minimizzazione ℓ_1 con vincolo quadratico dell'errore di approssimazione della stima dell'energia dell'errore di misura, è stata valutata tramite simulazioni su reti di distribuzione.

- Il Capitolo 3 affronta il problema della stima dei parametri di linea in una rete di trasmissione e la compensazione degli errori sistematici della catena di misura sfruttando un approccio a rami multipli. Il problema di stima viene modellato considerando le sorgenti di incertezza della catena di misura e, in particolare, gli errori introdotti dai trasformatori di misura e dalle unità di misura fasoriali (PMU). Si introducono quindi i vincoli derivanti dalle leggi di Kirchhoff (considerando uno o più rami contemporaneamente), che vengono linearizzati sotto ipotesi realistiche integrando misure sincrofasoriali relative a condizioni operative di rete differenti, ed infine si aggiungono le informazioni a priori (prior) sulle variabili incognite per ottenere un sistema sovra-determinato ed aumentato. La soluzione del problema è prima affrontata con i minimi quadrati pesati, o *Weighted Least Squares* (WLS), poi inquadrati come caso particolare della regolarizzazione di Tikhonov, quindi viene effettuata un'accurata valutazione dell'incertezza di stima e infine viene presentata una via per migliorare le prestazioni grazie ad una adeguata gestione delle misure corrispondenti a molteplici casi operativi. La validità della proposta è stata verificata attraverso simulazioni eseguite su una rete di trasmissione IEEE a 14 nodi.

*“Nearly everything is really interesting
if you go into it deeply enough”
Richard P. Feynman.*

Capitolo 1

Tecniche di ottimizzazione per problemi inversi mal posti

1.1 Problemi inversi mal posti

I problemi inversi sono onnipresenti in diverse aree della scienza e dell'ingegneria. Ogni volta che si vuole individuare la causa di un fenomeno osservato o ricostruire la configurazione interna di un fenomeno a partire dai suoi effetti, si ha a che fare con un problema inverso.

Nei primi anni del XX secolo, il matematico francese Jacques Hadamard [1] introdusse il concetto di problema **ben posto** (*well-posed problem*) riferendosi a un problema che ha tutte le seguenti proprietà:

- esiste una soluzione
- la soluzione è unica
- la soluzione dipende con continuità dai dati

Quando anche una sola delle tre proprietà non sussiste il problema è detto **mal posto** (*ill-posed problem*). Un problema inverso nel caso lineare (o linearizzato), discreto e reale può essere descritto dal sistema lineare:

$$\begin{aligned}\mathbf{x} : \mathbf{A}\mathbf{x} &= \mathbf{b}_{\text{true}} \\ \mathbf{A} &\in \mathbb{R}^{m \times n} \\ \mathbf{x} \in \mathbb{R}^n; \mathbf{b}_{\text{true}} &\in \mathbb{R}^m\end{aligned}\tag{1.1}$$

dove, il vettore \mathbf{x} rappresenta le n componenti dello stato o del segnale oggetto dell'indagine, il vettore \mathbf{b}_{true} rappresenta le m quantità (effetti) misurabili, e dove \mathbf{A} è la matrice associata all'applicazione lineare f (1.2) che rappresenta il

legame fisico tra lo stato incognito \mathbf{x} e le quantità \mathbf{b}_{true} .

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ \mathbf{x} &\mapsto \mathbf{b}_{\text{true}} \end{aligned} \quad (1.2)$$

Come corollario del Teorema di *Rouché-Capelli*, il problema (1.1) ha una e una sola soluzione, diversa dalla soluzione nulla¹, se e solo se $m = n = \text{rank}(\mathbf{A})$, dove $\text{rank}(\mathbf{A})$ è il rango della matrice \mathbf{A} . Inoltre, affinché il problema sia ben posto, la soluzione stimata o ricostruita deve dipendere con continuità dai dati, i dati vengono acquisiti attraverso le misure che sono affette da errori, il problema (1.1) può quindi essere riscritto come:

$$\begin{aligned} \mathbf{x} : \mathbf{A}\mathbf{x} + \mathbf{e} &= \mathbf{b} \\ \mathbf{A} &\in \mathbb{R}^{m \times n} \\ \mathbf{x} \in \mathbb{R}^n; \mathbf{b}, \mathbf{e} &\in \mathbb{R}^m \end{aligned} \quad (1.3)$$

avendo indicato con \mathbf{b} il vettore di misura e \mathbf{e} è il vettore degli errori di misura. La relazione tra la perturbazione dei dati e l'errore sulla soluzione del problema è data da [2]:

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \kappa(\mathbf{A}) \frac{\|\mathbf{b}_{\text{true}} - \mathbf{b}\|}{\|\mathbf{b}_{\text{true}}\|} \quad (1.4)$$

dove $\hat{\mathbf{x}}$ è lo stato stimato o ricostruito e $\kappa(\mathbf{A})$ è il **numero di condizionamento** della matrice \mathbf{A} che può essere enorme e/o crescere esponenzialmente con le dimensioni della matrice, come nel caso della matrice di Hilbert². Per affrontare un problema mal posto è indispensabile la riformulazione dello stesso, aggiungendo requisiti aggiuntivi e, nel caso il problema sia fortemente mal condizionato, sostituendolo con un problema prossimo a quello originario ma più stabile rispetto all'errore di misura. Se il problema è fortemente mal condizionato infatti, una piccolissima perturbazione sui dati può portare una soluzione estremamente lontana da quella originale.

In questo lavoro viene esplorato un approccio, basato su tecniche di ottimizzazione vincolata o con penalizzazione, ad alcuni problemi di stima e monitoraggio di sistemi elettrici di potenza. L'obiettivo è quello di ottimizzare le prestazioni degli algoritmi di stima e monitoraggio attraverso l'integrazione nelle formulazioni di informazioni a priori sul dominio in esame, sullo stato da

¹Nel caso di un sistema omogeneo, ovvero con $\mathbf{b}_{\text{true}} = \mathbf{0} \in \mathbb{R}^m$, il sistema ammette un'unica soluzione, in particolare la soluzione nulla $\mathbf{x} = \mathbf{0} \in \mathbb{R}^n$, anche nel caso sovra-determinato $m \geq n$

²La Matrice di Hilbert è una matrice quadrata i cui valori sono definiti da $h_{ij} = \frac{1}{i+j-1}$, $i, j = 1, \dots, n$; dove $n \times n$ è la dimensione della matrice. La matrice di Hilbert interviene nell'ambito dell'approssimazione polinomiale di funzioni, nel senso dei minimi quadrati, nell'intervallo $[0, 1]$, utilizzando la 'base canonica'. Il suo condizionamento cresce esponenzialmente con le dimensioni della matrice e supera il valore (usando la norma euclidea) di 10^{13} a partire da $n = 10$

ricostruire e sfruttando le conoscenze sull'errore di misura. Nelle formulazioni basate su ottimizzazione si possono individuare due termini:

- l'**errore di approssimazione**, altrimenti detto **residuo**, *misfit* o *data fitting term*, che può tenere conto, in una data norma indicata col pedice p , dell'errore di misura³:

$$\begin{aligned} & \|\mathbf{Ax} - \mathbf{b}\|_p \\ & p \in \{1, 2, \infty\} \end{aligned} \tag{1.5}$$

- il termine di **regolarizzazione** che può tenere conto, in una data norma indicata col pedice q , delle conoscenze a priori sul segnale o stato da ricostruire, come i limiti, la sparsità o la regolarità del segnale stesso. La matrice \mathbf{L} è detta matrice di regolarizzazione, spesso coincide con la matrice di identità ma può essere sfruttata per «modellare» al meglio le caratteristiche del segnale da ricostruire:

$$\begin{aligned} & \|\mathbf{Lx}\|_q \\ & q \in \{0, 1, 2, \infty\} \\ & \mathbf{L} \in \mathbb{R}^{d \times n}, d \leq n \end{aligned} \tag{1.6}$$

Con $\|\cdot\|_p$ ($\|\cdot\|_q$) si intende la norma, o “pseudo” norma vettoriale meglio specificata nel paragrafo 1.2. I due termini possono essere composti in diverse formulazioni: come ottimizzazione vincolata, fissandone uno come funzione obiettivo da minimizzare ed imponendo l'altro come vincolo, oppure composti in una formulazione di ottimizzazione libera con penalizzazione:

1. minimo errore e vincolo di regolarizzazione

$$\begin{aligned} \mathcal{P}_{e_p/r_q} : \quad & \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_p \text{ s.t. } \|\mathbf{Lx}\|_q \leq k_q \\ & k_q \in \mathbb{R}^+ \cup \{+\infty\} \end{aligned} \tag{1.7}$$

Quando $k_q = +\infty$ è possibile scrivere \mathcal{P}_{e_p/r_u} dove $/r_u$ indica vincolo di regolarizzazione illimitato, *unbounded*.

2. minima norma e vincolo sull'errore di approssimazione

$$\begin{aligned} \mathcal{P}_{r_q/e_p} : \quad & \min_{\mathbf{x}} \|\mathbf{Lx}\|_q \text{ s.t. } \|\mathbf{Ax} - \mathbf{b}\|_p \leq \epsilon_p \\ & \epsilon_p \in \mathbb{R}^+ \cup \{0\} \end{aligned} \tag{1.8}$$

Quando $\epsilon_p = 0$ è possibile scrivere $\mathcal{P}_{r_q/e=0}$ dove $/e=0$ indica il vincolo di uguaglianza.

³Per semplicità, senza perdere di generalità, è stata omessa la formulazione pesata che è facilmente riconducibile alla formulazione generale: $\|\mathbf{W}(\mathbf{Ax} - \mathbf{b})\|_p = \|\mathbf{AWx} - \mathbf{bW}\|_p$, dove \mathbf{W} è la matrice dei pesi.

3. con penalizzazione,

$$\mathcal{P}_{e_p+r_q} : \min_x (\|\mathbf{Ax} - \mathbf{b}\|_p^r + \mu_{pq} \|\mathbf{Lx}\|_q^s)$$

$$\mu_{pq} \in \mathbb{R}^+$$

$$r = \begin{cases} p & p \neq \infty \\ 1 & p = \infty \end{cases} \quad s = \begin{cases} q & q \notin \{0, \infty\} \\ 1 & q \in \{0, \infty\} \end{cases} \quad (1.9)$$

Avendo indicato con $\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$ l'insieme dei reali positivi. Nella simbologia utilizzata \mathcal{P} indica “problema” ed è specificato dai pedici, nel caso di ottimizzazione vincolata il simbolo $/$ separa il termine da minimizzare a sinistra, dal vincolo a destra, questi possono essere il termine di regolarizzazione (in norma q) r_q e il residuo o errore di approssimazione (in norma q) e_p . Nel caso di vincolo di uguaglianza (*equality constraint*), ovvero errore di approssimazione nullo, viene indicato con $/e=0$, nel caso di un problema di minimizzazione dell'errore senza vincolo di regolarizzazione, viene indicato con $/r_u$. Nel Caso di ottimizzazione libera con penalizzazione, i due termini di regolarizzazione e d'errore vengono uniti dal simbolo $+$.

Alla prima formulazione può essere ricondotto il metodo dei minimi quadrati, *Least Squares* (LS), per sistemi sovra-determinati ($m > n$) con la formulazione \mathcal{P}_{e_2/r_u} , ovvero minimo errore quadratico, con $p = 2$, e vincolo di regolarizzazione illimitato (*unbounded*), con $k_q = +\infty$, trasformandolo in un problema di ottimizzazione libera.

Alla seconda formulazione può essere ricondotto il metodo dei minimi quadrati a norma minima per sistemi sotto-determinati ($m < n$) con $q = 2$, $e_p = 0$ e vincolo di uguaglianza, $\mathcal{P}_{r_2/e=0}$.

Infine alla terza formulazione si riconduce il metodo di regolarizzazione di Tikhonov con $p = q = 2$.

Il capitolo è organizzato nel seguente modo:

- Nel paragrafo 1.2 viene fatto un richiamo alle norme, vettoriali e matriciali, e ad alcune delle loro proprietà che verranno sfruttate nel lavoro di tesi.
- Il paragrafo 1.3 affronta il problema dell'uso delle diverse norme e della loro composizione all'interno delle formulazioni di ottimizzazione. In particolare, viene discusso come sfruttare le norme per integrare informazioni, sul dominio e sull'errore di misura, all'interno delle formulazioni. Viene inoltre data una interpretazione statistica della scelta delle norme nelle diverse formulazioni.

- Il paragrafo 1.4 è dedicato alle formulazioni sparse per problemi sotto-determinati. La questione viene in parte affrontata anche nel paragrafo precedente in quanto strettamente legata alla scelta della norma da utilizzare per la valutazione del termine di regolarizzazione. Il paragrafo è quindi dedicato ad inquadrare il problema all'interno del *Compressive Sensing* (CS) di cui viene data una introduzione. Le tecniche vengono poi utilizzate nel Capitolo 2 per determinare le sorgenti di inquinamento armonico in una rete elettrica di distribuzione.
- Il paragrafo 1.5 è dedicato alla regolarizzazione per problemi mal posti. Vengono presentate importanti fattorizzazioni come la decomposizione a valori singolari, *Singular Value Decomposition* SVD fondamentale per l'analisi del problema. Il paragrafo affronta la regolarizzazione di Tikhonov ed i criteri per scegliere il parametro di regolarizzazione, con un approfondimento al criterio del *Discrepancy Principle*. Le metodologie affrontate nel paragrafo sono fondamentali per l'applicazione presentata nel Capitolo 3 dedicato alla stima dei parametri di linea in una rete di trasmissione simultanea alla compensazione degli errori sistematici nella catena di misura.
- L'ultimo paragrafo del capitolo è dedicato alla stima dell'energia del vettore di errore. Il paragrafo discute in dettaglio la *Whitening Transformation*, presenta una particolare matrice di sbiancamento e mostra alcune proprietà dell'energia dell'errore. La stima dell'energia dell'errore è cruciale per limitare l'errore di approssimazione nell'algoritmo con formulazione \mathcal{P}_{r_1/e_2} , cuore del Capitolo 2. Viene inoltre utilizzata anche nel Capitolo 3 per selezionare il parametro con il criterio del *Discrepancy Principle* nella regolarizzazione di Tikhonov, formulazione $\mathcal{P}_{e_2+r_2}$.

1.2 Norme

Definizione 1.2.1. Si definisce **spazio normato** uno spazio vettoriale, reale o complesso, \mathbb{X} sul campo \mathbb{R} per il quale può essere definita una funzione

$$\begin{aligned} \|\cdot\| : \mathbb{X} &\longrightarrow \mathbb{R}^+ \cup \{0\} \\ \mathbf{x} &\mapsto \|\mathbf{x}\| \end{aligned}$$

detta **norma**, tale che per ogni $\mathbf{x}, \mathbf{y} \in \mathbb{X}$ e $\alpha \in \mathbb{R}$

1. $\|\mathbf{x}\| \geq 0 \forall \mathbf{x} \in \mathbb{X}$ e $\|\mathbf{x}\| = 0 \iff \mathbf{x} = 0$ (positività)
2. $\|\alpha\mathbf{x}\| = |\alpha| \|\mathbf{x}\|$ (omogeneità)
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (diseguaglianza triangolare)

Una funzione che gode della proprietà 3 è detta **subadditiva**. Negli spazi \mathbb{R}^n e \mathbb{C}^n possono essere definite infinite norme, tre delle norme più comunemente usate sono la norma Euclidea o norma 2, la norma 1 chiamata anche norma Manhattan o Taxicab e la norma infinito.

$$\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i| \quad (1.10)$$

$$\|\mathbf{x}\|_2 := \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (1.11)$$

$$\|\mathbf{x}\|_\infty := \max_{i=1, \dots, n} |x_i| \quad (1.12)$$

La Figura 1.1 mostra la forma delle rispettive sfere unitarie in \mathbb{R}^2

$$S := \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\} \quad (1.13)$$

Più in generale, su \mathbb{R}^n e \mathbb{C}^n si definisce la Hölder p -norm [3], indicata anche

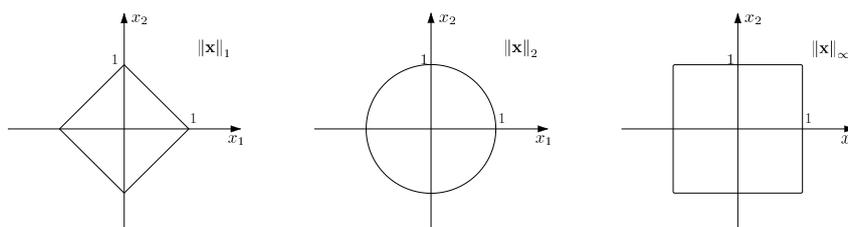


Figura 1.1: Sfere unitarie in \mathbb{R}^2 rispetto alle norme 1, 2 e ∞ .

come $\|\cdot\|_p$, ℓ_p -norm o semplicemente ℓ_p , di cui le tre norme presentate sono tre realizzazioni tra le infinite possibili

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (1.14)$$

$$\forall \mathbf{x} \in \mathbb{R}^n, \mathbb{C}^n$$

$$\forall p \in [1, +\infty)$$

È immediato verificare la consistenza della definizione di

$$\|\mathbf{x}\|_\infty := \lim_{p \rightarrow +\infty} \|\mathbf{x}\|_p = \max_{i=1, \dots, n} |x_i|$$

Teorema 1.1. *Tutte le norme di uno spazio a dimensione finita \mathcal{X} sono equivalenti, ovvero per ogni coppia di norme $\|\cdot\|_\alpha$ e $\|\cdot\|_\beta$ esistono due costanti*

positive $c, C \in \mathbb{R}^+$ tali che $\forall \mathbf{x} \in \mathcal{X}$

$$c \|\mathbf{x}\|_\beta \leq \|\mathbf{x}\|_\alpha \leq C \|\mathbf{x}\|_\beta \quad (1.15)$$

Il risultato precedente vale in particolare per gli spazi metrici \mathbb{R}^n e \mathbb{C}^n e consente di utilizzare indifferentemente una qualunque norma (ad esempio (1.14)) per studiare la convergenza di successioni di vettori.

L'estensione della (1.14) al caso $p \in (0, 1)$ è possibile ed utilizzata, anche se il risultato non è propriamente una norma, infatti non rispetta la disuguaglianza triangolare, proprietà 3 in 1.2.1.

Una menzione a parte merita la ℓ_0 -“norma” introdotta dal matematico David Donoho per indicare il numero di componenti non nulle di un vettore ovvero per indicare la cardinalità del suo **supporto**. Le virgolette introdotte da Donoho stanno a sottolineare che non si tratta propriamente di una norma in quanto non soddisfa la omogeneità, proprietà 2 in 1.2.1, tuttavia viene spesso utilizzata senza le virgolette o semplicemente indicata come ℓ_0 o $\|\cdot\|_0$:

$$\|\mathbf{x}\|_0 := \text{Card}\{i : x_i \neq 0\} \quad (1.16)$$

con $\mathbf{x} \in \mathbb{R}^n, \mathbb{C}^n$, e $\text{Card}\{\cdot\}$ indica il numero di elementi di un insieme

Va menzionato il fatto che della “norma” zero è possibile trovare anche una formulazione basata sulla p -norm (1.14)

$$\|\mathbf{x}\|_0 := \lim_{p \rightarrow 0^+} \|\mathbf{x}\|_p^p = \lim_{p \rightarrow 0^+} \sum_{i=1}^n |x_i|^p$$

definendo forzatamente la forma indeterminata $0^0 := 0$.

Introduciamo ora il concetto di convessità:

Definizione 1.2.2. Sia f una funzione da un insieme V con valori in \mathbb{R}

$$\begin{aligned} f : V \rightarrow \mathbb{R} \text{ è } \mathbf{convessa} \text{ se e solo se} \\ f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) \\ \forall \mathbf{x}, \mathbf{y} \in V, \lambda \in [0, 1] \end{aligned} \quad (1.17)$$

quando vale la disuguaglianza stretta, la funzione si definisce **strettamente convessa**.

conseguentemente

Teorema 1.2. Ogni *norma* in uno spazio normato \mathbb{X} è *convessa*.

Dimostrazione. È sufficiente applicare la disuguaglianza triangolare e l'omogeneità (Def. 1.2.1)

$$\begin{aligned} \|\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}\| \leq \|\lambda \mathbf{x}\| + \|(1 - \lambda)\mathbf{y}\| = \lambda \|\mathbf{x}\| + (1 - \lambda)\|\mathbf{y}\| \\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{X}, \lambda \in [0, 1] \end{aligned} \quad (1.18)$$

□

Il risultato precedente si applica di conseguenza anche alla $\|\cdot\|_p$ (1.14) con $p \geq 1$, inoltre per $p \neq \{1, \infty\}$ la p -norma è anche strettamente convessa. Il Teorema 1.2 non si applica invece per $p < 1$, infatti per $p \in (0, 1)$ non vale la disuguaglianza triangolare e per $p = 0$ non vale l'omogeneità, proprietà 3 e 2 in 1.2.1 rispettivamente. La Figura 1.2 mostra graficamente la non convessità per $p = \frac{1}{2}$ e la convessità per $p \in \{1, 5\}$, la figura mostra anche come la 1-norma, riportata al centro della figura, non sia strettamente convessa.

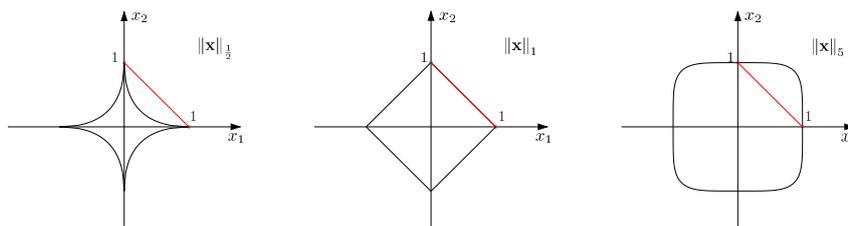


Figura 1.2: Sfere unitarie in \mathbb{R}^2 rispetto alle norme $p = \{\frac{1}{2}, 1, 5\}$ con evidenza della convessità per norme 1 e 5 e della non convessità per la $\|\cdot\|_{\frac{1}{2}}$

1.2.1 Norme matriciali

Lo spazio vettoriale delle matrici $\mathcal{M}_{m \times n}$ può essere dotato della struttura di spazio normato definendo una norma che verifichi le proprietà richieste. È possibile estendere alle matrici le norme vettoriali, ad esempio l'estensione della norma ∞ (1.12) è immediata, e corrisponde al massimo valore assoluto degli elementi della matrice. Tuttavia quando si parla di norma matriciale viene richiesta, in aggiunta alla positività, omogeneità e alla disuguaglianza triangolare (proprietà 1, 2, 3 nella Definizione 1.2.1), anche la proprietà submoltiplicativa [2].

Definizione 1.2.3. Una norma matriciale si definisce **submoltiplicativa** se per ogni coppia di matrici di dimensioni compatibili si ha:

$$4. \quad \|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\| \quad (\text{proprietà submoltiplicativa})$$

Le norme vettoriali su $\mathbb{R}^n, \mathbb{C}^n$ sono submoltiplicative, ma la semplice estensione di una norma vettoriale alle matrici non garantisce il soddisfacimento di questa proprietà, come mostra l'esempio seguente.

Esempio 1.2.1. L'estensione della $\|\cdot\|_{\infty}$ vettoriale alle matrici in $\mathbb{R}^{2 \times 2}$ porta a definire la norma

$$\|\cdot\|_{inf} = \max_{i,j=\{1,2\}} |x_{ij}|$$

nella quale si è utilizzato il pedice $_{inf}$ per distinguerla dalla norma infinito matriciale, definita in nel seguito come norma indotta. Consideriamo quindi le matrici

$$\mathbf{A} = \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{C} = \mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

abbiamo:

$$2 = \|\mathbf{AB}\|_{inf} > \|\mathbf{A}\|_{inf} \|\mathbf{B}\|_{inf} = 1$$

Vale la pena sottolineare che il segno \leq nella proprietà submoltiplicativa è dovuto al fatto che l'insieme delle matrici, considerando il prodotto righe per colonne, ammette divisori dello zero, ovvero è possibile che il prodotto sia nullo anche quando entrambi i fattori sono diversi da zero. Conseguentemente, per la positività (proprietà 1 nella Definizione 1.2.1), la norma del prodotto è nulla mentre il prodotto delle norme dei fattori è positivo, come mostra l'esempio seguente.

Esempio 1.2.2.

$$\mathbf{A} = \begin{bmatrix} 3 & -3 \\ 2 & -2 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Qualunque norma si consideri, grazie alla positività si ha $\|\mathbf{AB}\| = 0$, $\|\mathbf{A}\| > 0$, $\|\mathbf{B}\| > 0$.

Una utile proprietà che lega le norme matriciali e vettoriali è la compatibilità o consistenza:

Definizione 1.2.4. Una norma matriciale $\|\cdot\|_\gamma$ è detta **compatibile** o **consistente** con le norme vettoriali $\|\cdot\|_\alpha$ in \mathbb{C}^n e $\|\cdot\|_\beta$ in \mathbb{C}^m se:

$$\begin{aligned} \|\mathbf{Ax}\|_\beta &\leq \|\mathbf{A}\|_\gamma \|\mathbf{x}\|_\alpha \\ \forall \mathbf{A} \in \mathbb{C}^{m \times n}, \forall \mathbf{x} \in \mathbb{C}^n \end{aligned} \tag{1.19}$$

Usualmente le norme vettoriali $\|\cdot\|_\alpha$ e $\|\cdot\|_\beta$ sono uguali e il pedice si trascura. Tutte le norme matriciali, essendo definite su spazi a dimensione finita, sono equivalenti, Teorema 1.1.

Introduciamo una importante classe di norme matriciali, le norme **indotte** o **naturali**.

Definizione 1.2.5. Una norma matriciale si dice **indotta** o **naturale** dalla norma vettoriale $\|\cdot\|$ se

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \tag{1.20}$$

Una norma naturale misura, in una data norma vettoriale, il massimo allungamento relativo che un vettore può subire in seguito al prodotto per la matrice.

Teorema 1.3. *Tutte le **norme** matriciali **indotte** sono **submoltiplicative** e **compatibili** con la norma vettoriale che la induce.*

Le norme matriciali indotte più comuni, con riferimento a una matrice $\mathbf{A} \in \mathbb{C}^{m \times n}$, sono:

$$\begin{aligned}\|\mathbf{A}\|_1 &= \max_{j=1,\dots,n} \sum_{i=1}^m |a_{ij}| \\ \|\mathbf{A}\|_2 &= \sqrt{\rho(\mathbf{A}^* \mathbf{A})} \\ \|\mathbf{A}\|_\infty &= \max_{i=1,\dots,m} \sum_{j=1}^n |a_{ij}|\end{aligned}\tag{1.21}$$

Dove con $*$ si intende l'operatore trasposto coniugato e $\rho(\cdot)$ indica il raggio spettrale ovvero l'autovalore con il modulo massimo. Se \mathbf{A} è simmetrica $\|\mathbf{A}\|_2 = \rho(\mathbf{A})$. Alcune importanti proprietà che legano assieme la norma 2 matriciale con la norma 2 vettoriale che la induce sono:

$$\begin{aligned}\forall \mathbf{Q} \in \mathbb{C}^{m \times m} : \mathbf{Q}\mathbf{Q}^* &= \mathbf{Q}^* \mathbf{Q} = \mathbf{I}_m \\ \forall \mathbf{B} \in \mathbb{C}^{m \times m} \\ \forall \mathbf{x} \in \mathbb{C}^m \\ \|\mathbf{Q}\|_2 &= 1 \\ \|\mathbf{QB}\|_2 &= \|\mathbf{BQ}\|_2 = \|\mathbf{B}\|_2 \\ \|\mathbf{Qx}\|_2 &= \|\mathbf{Q}\|_2 \|\mathbf{x}\|_2 = \|\mathbf{x}\|_2\end{aligned}\tag{1.22}$$

dove \mathbf{I}_m è la matrice identità di dimensione m e \mathbf{Q} è una matrice unitaria, detta ortogonale nel caso reale. Le ultime due proprietà indicano che le matrici unitarie non alterano la norma Euclidea vettoriale o matriciale. Per verificare l'ultima proprietà è sufficiente ricordare che:

$$\|\mathbf{x}\|_2^2 = \mathbf{x}^* \mathbf{x}\tag{1.23}$$

Oltre alle norme indotte riveste una particolare importanza la norma di Frobenius:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}\tag{1.24}$$

che è l'estensione della norma 2 vettoriale. La norma di Frobenius non è una

norma matriciale indotta ma è comunque submoltiplicativa. Si ha inoltre:

$$\begin{aligned}
 \forall \mathbf{Q} \in \mathbb{C}^{m \times m} : \mathbf{Q}\mathbf{Q}^* &= \mathbf{Q}^*\mathbf{Q} = \mathbf{I}_m \\
 \forall \mathbf{B} \in \mathbb{C}^{m \times m} \\
 \|\mathbf{Q}\|_F &= \sqrt{m} \\
 \|\mathbf{QB}\|_F &= \|\mathbf{BQ}\|_F = \|\mathbf{B}\|_F
 \end{aligned} \tag{1.25}$$

1.3 Norme ed informazione

I due termini che compongono le formulazioni di ottimizzazione introdotte nel paragrafo 1.1 vengono valutati ciascuno secondo una norma, la scelta della norma può essere utilizzata per integrare nel modello informazioni a priori sul problema in esame.

1.3.1 Errore di approssimazione o residuo

Il termine $\|\mathbf{Ax} - \mathbf{b}\|_p$ valuta, in una data norma, la discrepanza tra i dati misurati \mathbf{b} e la proiezione, tramite la matrice di misura \mathbf{A} , del vettore incognito \mathbf{x} , per questo viene anche detto *data fitting term*. Nelle formulazioni a norma minima (1.8) viene utilizzato come vincolo, per rilassare il vincolo di uguaglianza, nel caso di sistemi sotto-determinati o, in generale, per imporre un limite superiore (*upper bound*) in ℓ_p al residuo. Nelle formulazioni con penalizzazione (1.9), per problemi mal condizionati ad esempio, è uno dei due termini da bilanciare mentre nelle formulazioni a minimo errore, è il termine da minimizzare. La scelta della specifica norma p può essere utilizzata per tenere conto della distribuzione di probabilità dell'errore di misura $\mathbf{e} = \mathbf{b} - \mathbf{Ax}$. In generale la valutazione in ℓ_p del residuo, può modellare un errore di misura additivo specifico. In particolare, la minimizzazione in ℓ_p del residuo, senza vincolo di regolarizzazione, ovvero con vincolo di regolarizzazione illimitato:

$$\hat{\mathbf{x}}_p = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_p, p \in \{1, 2, \infty\} \tag{1.26}$$

che corrisponde al problema \mathcal{P}_{e_p/r_u} secondo la simbologia (1.7), realizza lo stimatore di massima verosimiglianza, *Maximum Likelihood Estimator* (MLE) quando i campioni dell'errore di misura $\mathbf{e} = \mathbf{b} - \mathbf{Ax}$ sono indipendenti ed identicamente distribuiti, *Independent, Identically Distributed* (IID), e seguono specifiche distribuzioni di probabilità. In particolare:

- **errore Gaussiano**, lo stimatore di massima verosimiglianza è la soluzione ottenuta con il metodo dei minimi quadrati LS, corrispondente a (\mathcal{P}_{e_2/r_u}) ,

$$\hat{\mathbf{x}}_2 = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2 \tag{1.27}$$

come verrà mostrato successivamente in 1.3.4

- **errore uniforme**, l'errore può essere modellato dalla norma infinito [4], [5, Cap. 7.1] e la soluzione della minimizzazione del residuo in ℓ_∞ , ($\mathcal{P}_{e_\infty/r_u}$) è il MLE,

$$\hat{\mathbf{x}}_\infty = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_\infty \quad (1.28)$$

- **errore laplaciano**, può essere modellato attraverso la minimizzazione del residuo in ℓ_1 (\mathcal{P}_{e_1/r_u}), nota anche come *Least Absolute Values* (LAV) o *Least Absolute Deviations* LAD, che ne realizza lo stimatore di massima verosimiglianza [6], come verrà mostrato successivamente in 1.3.4,

$$\hat{\mathbf{x}}_1 = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_1 \quad (1.29)$$

Va ricordato che tra le tre norme, solo la ℓ_2 è strettamente convessa, e solo per questa norma l'unicità della soluzione è garantita. In [3], [7], si sottolinea come la minimizzazione dell'errore di approssimazione in ℓ_1 sia meno sensibile del Least Squares (LS) rispetto a pochi errori grandi ed isolati. Questo è spiegabile considerando che (1.29) è il MLE per un errore che segue la distribuzione di **Laplace**, la cui funzione di densità di probabilità, *Probability Density Function* PDF, di media z_0 e varianza $2\tau^2$ è data da:

$$f(z) = \frac{1}{2\tau} e^{-\frac{|z-z_0|}{\tau}} \quad (1.30)$$

ovvero è una esponenziale doppia, che ha la caratteristica di essere *heavy-tailed* [8], detta anche *fat-tailed*, ovvero ha le code che decadono meno velocemente rispetto alla distribuzione Gaussiana che decade con $e^{-\frac{(z-z_0)^2}{2\sigma^2}}$.

Esempio 1.3.1. Sia $\mathbf{A} = [1, \dots, 1] \in \mathbb{R}^{m \times 1}$ la matrice di misura, e \mathbf{b} il vettore delle osservazioni di componenti $b_i, i = 1, \dots, m$, la minimizzazione in ℓ_1 dell'errore di approssimazione trova la mediana delle m osservazioni, il LS ne trova la media mentre la minimizzazione in ℓ_∞ trova il valore medio tra il minimo e il massimo valore tra le m osservazioni.

Quando si deve ricostruire lo stato a n variabili di un problema fisico, e si dispone di un modello lineare del problema $\mathbf{b}_{\text{true}} = \mathbf{Ax}$, dove il vettore delle quantità \mathbf{b}_{true} è incognito ma misurabile. Per ridurre l'impatto dell'errore di misura sulle osservazioni la scelta comune è quella di eseguire un numero di misure superiore al numero delle incognite. Il risultato è un problema inverso (1.3) sovra-determinato ($m > n$). Il problema non ha una soluzione esatta, una scelta comune è quella di ricorrere al metodo dei minimi quadrati per trovare una soluzione approssimata (1.27). Nel 1799 Laplace usa il principio della minimizzazione della somma degli errori di approssimazione in valore assoluto, Gauss e Legendre la minimizzazione dei quadrati, giustificata statisticamente da Gauss nel 1809, [3, Cap. 1.1.1] L'esempio seguente, 1.3.2, simula un problema inverso sovra-determinato a rango pieno considerando tre differenti configurazioni

d'errore: errore Gaussiano, errore uniforme, ed errore Gaussiano con l'aggiunta di una piccola percentuale di misure anomale, comunemente dette *outliers*. Il problema è stato affrontato, per tutte e tre le configurazioni, con il noto metodo dei minimi quadrati (1.27), con la minimizzazione del residuo in ℓ_1 (LAV) (1.29) ed in ℓ_∞ (1.28).

Per quanto riguarda la minimizzazione in ℓ_2 è stata affrontata con le **equazioni normali**

$$(\mathbf{A}^\top \mathbf{A}) \hat{\mathbf{x}}_2 = \mathbf{A}^\top \mathbf{b} \quad (1.31)$$

che nel caso di $m > n$ a rango pieno può essere scritta come:

$$\begin{aligned} \hat{\mathbf{x}}_2 &= (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} = \mathbf{F} \mathbf{b} \\ \mathbf{F} &\in \mathbb{R}^{n \times m} \end{aligned} \quad (1.32)$$

evidenziando come il LS sia uno stimatore lineare. Inoltre, secondo il Teorema di Gauss-Markov [3, Teorema 1.1.1.] [9, Cap. 3.2.2], nel caso di disturbo scorrelato a valore atteso nullo e con la stessa varianza su tutte le componenti, è il miglior stimatore lineare non distorto, *Best Linear Unbiased Estimator* (BLUE).

Le approssimazioni in ℓ_1 (LAV) ed in ℓ_∞ , come sottolineato in [3, Cap.4.5.4], sono complicate dal fatto che il termine $\|\mathbf{Ax} - \mathbf{b}\|_p^p$ non è differenziabile per $p = \{1, \infty\}$, ma entrambi problemi, nel caso di valori reali, possono essere riscritti come problema di programmazione lineare, *Linear Programming* (LP)⁴ e quindi affrontati col metodo del simplesso [10], [7, Cap. 6] o altro metodo dedicato ed implementato in qualunque risolutore *solver*, ad esempio GURUBI o CPLEX. Un problema di programmazione lineare è caratterizzato da una **funzione obiettivo** lineare da minimizzare o massimizzare, formalizzabile come prodotto scalare tra il vettore di variabili decisionali reali \mathbf{z} e l vettore dei pesi \mathbf{c} , e dai vincoli di uguaglianza o disuguaglianza. Un problema di **LP in Forma Canonica** si presenta come:

$$\begin{aligned} \min_{\mathbf{z}} \quad & \mathbf{c}^\top \mathbf{z} \\ \text{s.t.} \quad & \mathbf{H} \mathbf{z} \leq \mathbf{y} \\ & \mathbf{z} \geq \mathbf{0}_{n_{LP}} \end{aligned} \quad (1.33)$$

con:

$$\begin{aligned} \mathbf{H} &\in \mathbb{R}^{m_{LP} \times n_{LP}} \\ \mathbf{z}, \mathbf{c}, \mathbf{0}_{n_{LP}} &\in \mathbb{R}^{n_{LP}} \\ \mathbf{y} &\in \mathbb{R}^{m_{LP}} \end{aligned}$$

dove con $\mathbf{0}_{n_{LP}}$ si intende il vettore con n_{LP} componenti nulle. Le riscritture dei problemi della minimizzazione dei valori assoluti LAV, (\mathcal{P}_{e_1/r_u}) e del residuo

⁴La limitazione al caso reale della riscrittura dei problemi come LP è dovuta al fatto che il campo dei complessi \mathbb{C} è un **campo non ordinato**.

in ℓ_∞ ($\mathcal{P}_{e_\infty/r_u}$) come problemi di programmazione lineare sono state riportate rispettivamente in Appendice 1.A e 1.B. Entrambi i problemi sono stati quindi trasformati in un problema LP in forma canonica, implementati in MATLAB e quindi risolti attraverso il *solver* CPLEX (IBM ILOG CPLEX V12.10.0).

Esempio 1.3.2.

$$\begin{aligned} \mathbf{Ax} + \mathbf{e} &= \mathbf{b} \\ \mathbf{A} &\in \mathbb{R}^{1000 \times 100} \\ \mathbf{x} \in \mathbb{R}^{100} \quad \mathbf{b}, \mathbf{e} &\in \mathbb{R}^{1000} \end{aligned} \tag{1.34}$$

Gli elementi della matrice \mathbf{A} sono stati estratti casualmente in $[0, 1]$ (da una distribuzione di tipo uniforme), sono stati simulati 10 casi estraendo casualmente altrettante matrici. Per tutti i casi è stato utilizzato come vettore di stato incognito il vettore $\mathbf{x} = \mathbb{1}_{100} = [1, 1, \dots, 1]^\top \in \mathbb{R}^{100}$. Le quantità vere da misurare sono state ottenute proiettando il vettore di stato tramite la matrice di misura (problema diretto): $\mathbf{b}_{\text{true}} = \mathbf{Ax}$. Per simulare le misure, alle quantità vere è stato aggiunto l'errore di misura considerando la formula:

$$\begin{aligned} \mathbf{b} &= \mathbf{b}_{\text{true}} \odot (\mathbb{1}_{1000} + \epsilon) \\ \mathbf{b}, \mathbf{b}_{\text{true}}, \epsilon &\in \mathbb{R}^{1000} \end{aligned} \tag{1.35}$$

dove il simbolo \odot è il prodotto matriciale di Hadamard, ovvero il prodotto elemento per elemento. Sono state prese in considerazione tre diverse configurazioni per ϵ :

1. errore Gaussiano
2. errore Gaussiano con *outliers*
3. errore uniforme

Per tutte le configurazioni d'errore è stata considerata una incertezza estesa di misura del 1%, per la validazione statistica sono state effettuate 1000 prove Monte Carlo MC. Per ogni prova sono stati estratti i valori delle componenti di ϵ da una distribuzione Gaussiana con fattore di copertura pari a 3 (ovvero considerando una deviazione standard pari a 0.0033), per i due casi gaussiani, e da una distribuzione uniforme nell'intervallo $[-0.01, 0.01]$ nel caso uniforme⁵. Nella configurazione Gaussiana con *outliers*, questi ultimi sono stati modellati aggiungendo uno 0.5% di misure anomale in posizione casuale. L'anomalia è stata simulata estraendo i valori delle componenti di ϵ corrispondenti agli *outliers* da una Gaussiana con una deviazione standard maggiorata di 20 volte rispetto alla configurazione base. Per tutte le configurazioni il problema è stato

⁵Va sottolineato che l'errore di misura così ottenuto, pari a $\mathbf{b}_{\text{true}} \odot \epsilon$ è ancora Gaussiano o uniforme ma con diverse deviazioni standard per le varie componenti.

affrontato considerando tre diversi metodi corrispondenti alla minimizzazione dell'errore di approssimazione in ℓ_2 (1.27), ℓ_1 (1.29) e in ℓ_∞ (1.28), ottenendo le stime $\hat{\mathbf{x}}_p$ con $p \in \{2, 1, \infty\}$.

Per la valutazione delle prestazioni degli algoritmi è stato utilizzato il *Relative Root Square Error* (RRSE).

$$\text{RRSE} = \frac{\|\hat{\mathbf{x}}_p - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \quad (1.36)$$

La Tabella 1.1 riporta, per i tre metodi, i risultati ottenuti in termini di RRSE, mediati sulle 10 diverse estrazioni della matrice di misura e su 1000 prove Monte Carlo (MC) per ogni realizzazione stessa, per le tre configurazioni d'errore considerate. Nella configurazione con errore Gaussiano il metodo dei minimi

Tabella 1.1: RRSE medi, 10 casi e 1000 MC

tipo di errore	RRSE medi [%]		
	$\hat{\mathbf{x}}_2$	$\hat{\mathbf{x}}_1$	$\hat{\mathbf{x}}_\infty$
errore Gaussiano	1.91	2.37	3.90
errore Gaussiano + outliers	3.24	2.39	22.04
errore uniforme	3.31	5.24	2.51

quadrati ottiene un RRSE inferiore del 51 % rispetto alla minimizzazione in ℓ_∞ , ma quest'ultima riesce ad ottenere un miglior risultato di stima nel caso di errore uniforme, realizzando una diminuzione del RRSE medio del 24.5 % rispetto all'errore di stima ottenuto con il LS. La minimizzazione in ℓ_1 si dimostra più robusta del metodo dei minimi quadrati quando nelle misure sono presenti degli *outliers*, l'aggiunta di uno 0.5 % di *outliers* con un errore di 20 volte superiore alla configurazione d'errore Gaussiana base, porta il metodo LS ad aumentare l'errore RRSE di stima di quasi il 70 %, mentre l'errore della metodo in ℓ_1 rimane sostanzialmente invariato, pur partendo da un risultato peggiore nel caso Gaussiano base.

La Figura 1.3 riporta gli errori relativi di stima (si ricorda che $\mathbf{x} = \mathbb{1}_{100}$, componente per componente, nelle due configurazioni d'errore Gaussiana con *outliers* a sinistra e uniforme a destra, considerando per ciascuna configurazione il miglior metodo di stima testato assieme al classico LS. Per la leggibilità dei grafici, è stato visualizzato l'errore di stima relativo una sola prova MC per ognuna delle 10 realizzazioni della matrice \mathbf{A} . La figura mostra graficamente come, nel caso di errore uniforme (a destra), e di errore con *outliers*, il metodo dei minimi quadrati LS abbia un errore con varianza superiore rispettivamente ℓ_∞ e ℓ_1 .

I risultati della simulazione presentati nell'esempio mostrano come, nel com-

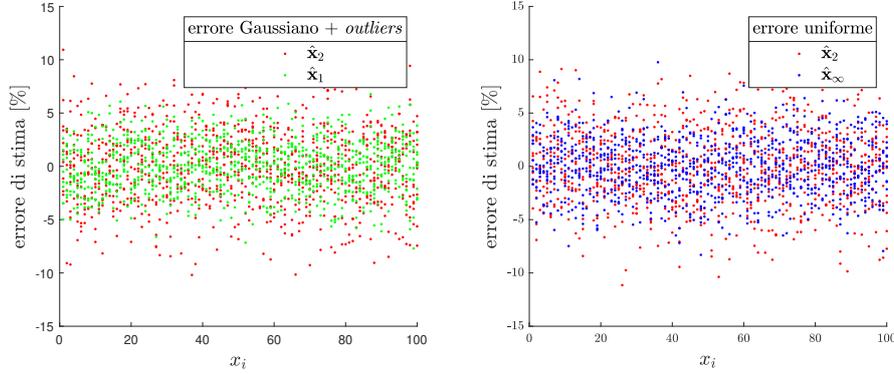


Figura 1.3: Errori relativi della stima delle componenti del vettore di stato nelle configurazioni: errore Gaussiano con outliers a sinistra (stima con minimizzazioni in ℓ_2 e in ℓ_1), errore uniforme a destra (stima con minimizzazioni in ℓ_2 e in ℓ_∞). Visualizzata una sola prova MC.

plesso, il metodo dei minimi quadrati si confermi il piú stabile ed affidabile, raggiungendo il massimo RRSE del 3.31% nelle tre configurazioni considerate, contro il 5.24% della minimizzazione in ℓ_1 ed del 22.04% di RRSE raggiunto dalla minimizzazione in ℓ_∞ nel caso di errore Gaussiano con *outliers*. Per questo motivo, in assenza di specifiche informazioni sulle caratteristiche dell'errore di misura, la valutazione del residuo viene valutata in ℓ_2 . Nel caso in cui, da informazioni a priori sul dominio d'indagine, la presenza di *outliers* tra le misure fosse probabile, la valutazione in ℓ_1 del residuo potrebbe essere una valida alternativa al LS. Qualora invece, la distribuzione dell'errore fosse di tipo uniforme, ad esempio nel caso in cui l'errore di quantizzazione fosse preponderante, come in [4], potrebbe essere preferibile un approccio che valutasse l'errore in ℓ_∞ . Va sottolineato che i risultati dell'Esempio 1.3.2 appena presentati non contraddicono il Teorema di Gauss-Markov [3, Teorema 1.1.1.] in quanto gli stimatori in ℓ_1 (1.29) e in ℓ_∞ (1.28) sono **stimatori non lineari**, ovvero per $p \in \{1, \infty\}$:

$$\begin{aligned} \nexists \mathbf{F}_p : \hat{\mathbf{x}}_p &= \mathbf{F}_p \mathbf{b} \\ \hat{\mathbf{b}} &= \mathbf{b}_{\text{true}} + \mathbf{e} \\ \mathbf{F}_p &\in \mathbb{R}^{n \times m} \end{aligned} \tag{1.37}$$

1.3.2 Termine di regolarizzazione

Il termine di regolarizzazione, detto anche di penalizzazione, $\|\mathbf{L}\mathbf{x}\|_q$ (1.6), *regularization term*, misura, in una data "norma" $q \in \{0, 1, 2, \infty\}$ lo stato del sistema che si intende ricostruire. Lo stato è pesato dalla una matrice $\mathbf{L} \in \mathbb{R}^{d \times n}$, $d \leq n$. Il termine consente di arricchire la formulazione di un problema inverso introducendo nello stesso informazioni a priori sulla soluzione ricercata. La matrice

\mathbf{L} solitamente coincide con \mathbf{I}_n , la matrice identità a n dimensioni ma, può essere utilizzata per meglio caratterizzare la soluzione. Nel caso in cui le componenti x_j del vettore di stato abbiano un differente range di variabilità, con σ_j la deviazione standard della j -esima componente, L può essere la matrice diagonale avente in posizione (j, j) il valore σ_j^{-1} per pesare ciascuna componente, secondo la propria variabilità. È anche possibile scegliere la matrice \mathbf{L} in maniera tale da non penalizzare, e quindi selezionare, particolari caratteristiche della soluzione che si vuole ricostruire, questo è possibile se vettori con tali caratteristiche sono elementi del “Nucleo”, *Kernel* di \mathbf{L} , indicato con $\mathcal{N}(\mathbf{L})$.

Definizione 1.3.1. Il “Nucleo”, *Kernel* o *Null-Space* di una generica matrice $\mathbf{A} \in \mathbb{R}^{m \times n}$, indicato con $\mathcal{N}(\mathbf{A})$, è il sotto-spazio di \mathbb{R}^n :

$$\mathcal{N}(\mathbf{A}) := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{0} \in \mathbb{R}^m\} \subseteq \mathbb{R}^n \quad (1.38)$$

In Appendice 1.C vengono proposti degli esempi per non penalizzare soluzioni “costanti” o “lisce” scegliendo opportunamente \mathbf{L} .

La scelta della specifica norma con la quale viene valutato il termine di regolarizzazione impatta sulle caratteristiche della soluzione che viene ricostruita. Storicamente, il primo problema inverso nel quale compare il termine di regolarizzazione è la soluzione di un sistema lineare (1.3) **sotto-determinato** ($m < n$). Un tale sistema ammette $\infty^{(n-m)}$ soluzioni (se a rango pieno, con $\text{rank}(\mathbf{A}) = m$), secondo il già citato Teorema di Rouché-Capelli. Questo, da un lato impone la necessità di impostare dei vincoli, dall’altro offre la possibilità di sfruttare i vincoli stessi per caratterizzare la soluzione da ricostruire. In assenza di specifiche motivazioni, il vincolo solitamente scelto è quello della norma (euclidea) minima ℓ_2 , pervenendo alla formulazione :

$$\hat{\mathbf{x}}_{r_2} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \quad (1.39)$$

nota come *minimum-norm least squares*, minimi quadrati a norma minima, equivalente al problema a $\mathcal{P}_{r_2/e=0}$ secondo la simbologia proposta nel paragrafo 1.1. Il problema è risolubile con le **equazioni normali di seconda specie**

$$\hat{\mathbf{x}}_{r_2} = \mathbf{x} : \begin{cases} (\mathbf{A}\mathbf{A}^\top)\mathbf{y} = \mathbf{b} \\ \mathbf{x} = \mathbf{A}^\top\mathbf{y} \end{cases} \quad (1.40)$$

ricavabili con il metodo dei moltiplicatori di Lagrange. Nel caso in cui $\text{rank}(\mathbf{A}) = m$, la matrice $\mathbf{A}\mathbf{A}^\top \in \mathbb{R}^{m \times m}$ è invertibile, e quindi si ha:

$$\hat{\mathbf{x}}_{r_2} = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{b} \quad (1.41)$$

La scelta della norma euclidea non è unica, è possibile utilizzare una qualunque norma vettoriale, ma l’unicità della soluzione è garantita solo quando la norma è strettamente convessa. Le scelte della norma con la quale valutare e minimiz-

zare lo stato da ricostruire, determina le caratteristiche della soluzione trovata, avendo informazioni a priori sulle caratteristiche della soluzione, è possibile scegliere la norma più opportuna. La norma 2 penalizza le componenti grandi del vettore di stato, conseguentemente seleziona una soluzione $\hat{\mathbf{x}}_2$ con componenti piccole. Con la norma ∞ viene selezionata una soluzione con componenti omogenee in modulo, mentre con la norma 0 e 1 viene selezionata una soluzione con poche componenti diverse da zero, ovvero una soluzione “sparsa”.

L'esempio seguente propone un sistema sotto-determinato 1×2 affrontato come problema a norma minima con vincolo di uguaglianza, $\mathcal{P}_{r_q/e=0}$:

$$\begin{aligned} \hat{\mathbf{x}}_{r_q} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{x}\|_q \text{ s.t. } \mathbf{Ax} = \mathbf{b} \\ q = \{0, 1, 2, \infty\} \end{aligned} \quad (1.42)$$

Esempio 1.3.3. Il sistema $\mathbf{Ax} = \mathbf{b}$ sotto-determinato con

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} 2 \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix};$$

ammette infinite soluzioni corrispondenti a tutti i punti della retta di equazione $x_1 + 2x_2 - 2 = 0$. Imponendo il vincolo di norma minima alla soluzione, utilizzando diverse norme (1.42), otteniamo le soluzioni:

$$\hat{\mathbf{x}}_{r_0} = \hat{\mathbf{x}}_{r_1} = (0, 1); \quad \hat{\mathbf{x}}_{r_2} = (0.4, 0.8); \quad \hat{\mathbf{x}}_{r_\infty} = \left(\frac{2}{3}, \frac{2}{3}\right)$$

La Figura 1.4 mostra la soluzione grafica del problema.

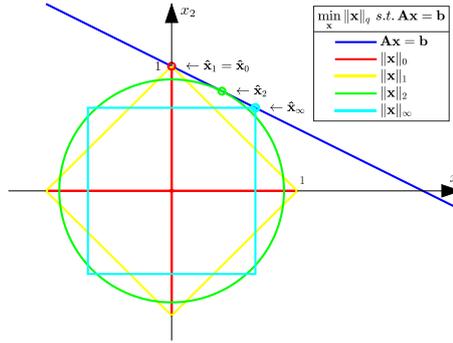


Figura 1.4: Soluzione grafica del problema $\mathcal{P}_{r_q/e=0}$ a norma minima con vincolo di uguaglianza, $\hat{\mathbf{x}}_{r_q} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{x}\|_q \text{ s.t. } \mathbf{Ax} = \mathbf{b}$, nel caso $\mathbf{A} \in \mathbb{R}^{1 \times 2}$ e $q \in \{0, 1, 2, \infty\}$.

L'esempio evidenzia come la soluzione in ℓ_2 ha componenti piccole, la soluzione con norma minima in ℓ_∞ ha le componenti omogenee $\hat{\mathbf{x}}_{r_\infty} = (\frac{2}{3}, \frac{2}{3})$, mentre le soluzioni a norma minima in ℓ_0 e ℓ_1 siano identiche e “sparse”, ovvero

hanno un solo elemento diverso da zero, come evidenziato nella soluzione grafica del problema riportata in Figura 1.4. Le formulazioni a minima norma per $q \in \{0, 1\}$ (1.42), sono chiamate rispettivamente problema (P0) e (P1), quest'ultimo è anche conosciuto come *Basis Pursuit* (BP)[11], verranno riprese nel paragrafo 1.4 ed il (P1) sperimentato nel Capitolo 2.

1.3.3 Combinazione di norme ed informazione a priori

Le più comuni combinazioni dei due termini consistono nell'aggiunta, alla valutazione dell'errore di approssimazione (1.5) in ℓ_2 , che in assenza di termine di regolarizzazione coincide con il metodo dei minimi quadrati LS, della valutazione del termine di regolarizzazione (1.6) in ℓ_2 o in ℓ_1 .

- Termine di regolarizzazione in ℓ_2 Per quanto riguarda l'aggiunta del termine di **regolarizzazione** in ℓ_2 , nel 1963 **Tikhonov** [12] introduce l'omonima regolarizzazione, nell'ambito della soluzione di problemi inversi mal posti, in formulazione Lagrangiana $\mathcal{P}_{e_2+r_2}$, che nella forma standard, con $\mathbf{L} = \mathbf{I}_n$, assume la forma:

$$\hat{\mathbf{x}}_{e_2+r_2} = \mathbf{x} : \min_x (\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \mu_{2,2} \|\mathbf{x}\|_2^2), \mu_{2,2} \in \mathbb{R}^+ \quad (1.43)$$

Nel 1970 Hoerl e Kennard [13, 14] introducono, in ambito statistico, la *Ridge Regression*, o regressione di cresta, per affrontare problemi implicanti matrici quasi-singolari con l'aggiunta del termine di regolarizzazione in ℓ_2 che produce uno slittamento degli autovalori rendendo invertibile la matrice. Il metodo di Tikhonov verrà presentato in modo analitico ed esteso nelParagrafo 1.5 e sperimentato nel Capitolo 3.

- Termine di regolarizzazione in ℓ_1 L'aggiunta del termine di regolarizzazione in ℓ_1 è stata introdotta da Tibshirani in [15] in ambito statistico, come tecnica di "restringimento", *shrinking*, portando al metodo di regressione o **regolarizzazione** denominato **LASSO**, *Least Absolute Shrinkage and Selection*, sia in formulazione Lagrangiana $\mathcal{P}_{e_2+r_1}$:

$$\hat{\mathbf{x}}_{e_2+r_1} = \mathbf{x} : \min_{\mathbf{x}} (\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \mu_{2,1} \|\mathbf{x}\|_1), \mu_{2,1} \in \mathbb{R}^+ \quad (1.44)$$

sia nell'equivalente formulazione con vincolo di regolarizzazione o a minimo errore \mathcal{P}_{e_2/r_1} :

$$\hat{\mathbf{x}}_{e_2/r_1} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \text{ s.t. } \|\mathbf{x}\|_1 \leq \epsilon_{2,1} \in \mathbb{R}^+ \quad (1.45)$$

Nell'ambito del *Signal Processing* (SP), ed in particolare nell'ambito della rappresentazione "sparsa" di un segnale, Chen, Donoho e Saunders in [16] introducono il metodo *Basis Pursuit De-Noising* BPDN che coincide, a meno di costanti, al metodo LASSO in formulazione (1.44) anche se, a

differenza di quest'ultimo, è indirizzato a sistemi sotto-determinati. L'utilizzo del termine di regolarizzazione in ℓ_1 , nella formulazione a minima norma \mathcal{P}_{r_1/e_2} , sempre per sistemi sotto-determinati, viene anche chiamato problema (P2) [11] e verrà presentato nel paragrafo 1.4 e sperimentato nel Capitolo 3.

Aggiungere il termine di regolarizzazione, o di *shrinking* all'impostazione dei minimi quadrati di un problema mal posto, può consentire di ottenere un vettore stimato più stabile. Il teorema di Gauss-Markov assicura che con il LS si ottiene lo stimatore lineare non distorto (non polarizzato o *unbiased*) a varianza minima. La regolarizzazione in ℓ_2 produce ancora uno stimatore lineare, questo non accade per la regolarizzazione mediante termini in ℓ_1 . Sia la regolarizzazione di Tikhonov sia la LASSO possono essere in grado di ridurre la varianza del LS, introducendo però una distorsione. Si tratta quindi di stimatori polarizzati.

L'esempio seguente mostra la soluzione grafica di un problema in \mathbb{R}^2 affrontato con le due regolarizzazioni presentate, e mostra come una soluzione $\hat{\mathbf{x}}_2$, inizialmente ottenuta con il metodo dei minimi quadrati (1.27) venga ristretta in una soluzione: più "sparsa" $\hat{\mathbf{x}}_{e_2/r_1}$ con la regolarizzazione di LASSO, e più piccola $\hat{\mathbf{x}}_{e_2/r_2}$ con la regolarizzazione di Tikhonov.

Esempio 1.3.4. Sistema lineare sovra-determinato $m \times n$ con $n = 2$

$$\begin{aligned} \mathbf{x} : \mathbf{Ax} &= \mathbf{b} \\ \mathbf{A} &\in \mathbb{R}^{m \times 2} \\ \mathbf{x} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad \mathbf{b} \in \mathbb{R}^m \end{aligned}$$

La Figura 1.5 mostra graficamente la soluzione del problema a minimo errore quadratico con vincolo di regolarizzazione in ℓ_1 a sinistra,

$$\hat{\mathbf{x}}_{e_2/r_1} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 \text{ s.t. } \|\mathbf{x}\|_1 \leq t; \quad (1.46)$$

e in ℓ_2 , a destra,

$$\hat{\mathbf{x}}_{e_2/r_2} = \mathbf{x} : \min_x \|\mathbf{Ax} - \mathbf{b}\|_2^2 \text{ s.t. } \|\mathbf{x}\|_2^2 \leq t^2; \quad (1.47)$$

Il termine $\|\mathbf{Ax} - \mathbf{b}\|_2^2 = \text{const}^2$ è una conica nelle variabili x_1, x_2 , al variare della costante il termine descrive un fascio di ellissi, tracciate in rosso, centrate sulla soluzione dei minimi quadrati $\hat{\mathbf{x}}_2$ (1.27), identica a destra e sinistra. La figura mostra in blu a sinistra e in rosso a destra, le sfere di raggio t in ℓ_1 e ℓ_2 rispettivamente. Le due nuove stime $\hat{\mathbf{x}}_{e_2/r_1}$ e $\hat{\mathbf{x}}_{e_2/r_2}$ sono ottenute come intersezione tra il fascio di ellissi e le sfere di raggio t in ℓ_1 e ℓ_2 rispettivamente e si trovano in posizione traslata rispetto al centro del fascio di ellissi evidenziando il fatto che siano stimatori *biased*. Le due metodologie sono utilizzate nel *Machine Learning* (ML) o *Data Mining* per evitare il rischio di *overfitting*, ovvero

per evitare che la soluzione del problema sia poco generalizzabile. Il risultato è raggiunto rilassando il vincolo di minimo errore di approssimazione, che si chiama anche *data fitting term*, bilanciandolo con il termine di regolarizzazione.

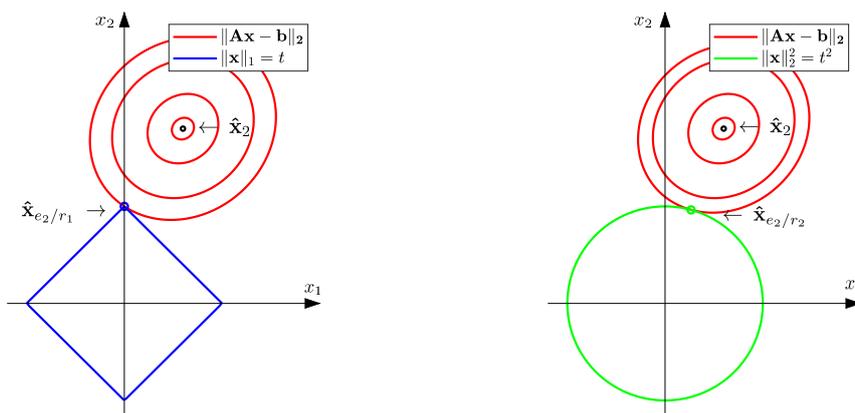


Figura 1.5: Soluzione grafica del problema a minimo errore quadratico $\min \|\mathbf{Ax} - \mathbf{b}\|_2^2$ con vincolo di regolarizzazione $\|\mathbf{x}\|_1 \leq t$, e $\|\mathbf{x}\|_2^2 \leq t^2$ a destra sinistra.

Nell'estensione da $n = 2$ a più dimensioni, le ellissi diventano ellissoidi e le sfere diventano ipersfere, in particolare, l'ipersfera in ℓ_1 è un romboide caratterizzato dall'aver tanti angoli, bordi lineari e facce, il che offre opportunità per trovare soluzioni di stima con elementi nulli, quindi "sparse". Questa caratteristica viene sfruttata nel *Machine Learning* (ML) o *Data Mining* per ridurre il numero di variabili, *Features Selection*, e nel *Signal Processing* per cercare soluzioni sparse in problemi sotto-determinati. Va sottolineato che, nel caso sotto-determinato, il termine $\|\mathbf{Ax} - \mathbf{b}\|_2^2 = \text{const}^2$ descrive sempre una conica, ma essendo $m < n$ la conica è degenere, e corrisponde a due rette parallele nel caso bidimensionale, come mostra l'esempio 1.3.5, e a due piani o iperpiani in più dimensioni.

Esempio 1.3.5. Si prenda il sistema dell'esempio 1.3.3, e lo si affronti con la minimizzazione del termine di regolarizzazione in ℓ_1 e ℓ_2 , sostituendo, in entrambi i casi, il vincolo di uguaglianza con un vincolo quadratico sull'errore di approssimazione $\|\mathbf{Ax} - \mathbf{b}\|_2^2 \leq \epsilon_2^2$. Tale vincolo, trasformato in uguaglianza ed applicato al problema dell'esempio 1.3.3, conduce all'equazione:

$$(x_1 + 2x_2 - 2)^2 = \epsilon_2^2$$

che è una conica degenera, coincidente con una coppia di parallele:

$$x_1 + 2x_2 - 2 + \epsilon_2 = 0$$

$$x_1 + 2x_2 - 2 - \epsilon_2 = 0$$

La figura 1.6 propone la soluzione grafica, individuando la soluzione sparsa $\hat{\mathbf{x}}_{r_1/\epsilon_2}$ in ℓ_1 e la soluzione con componenti piccole $\hat{\mathbf{x}}_{r_2/\epsilon_2}$ in ℓ_2

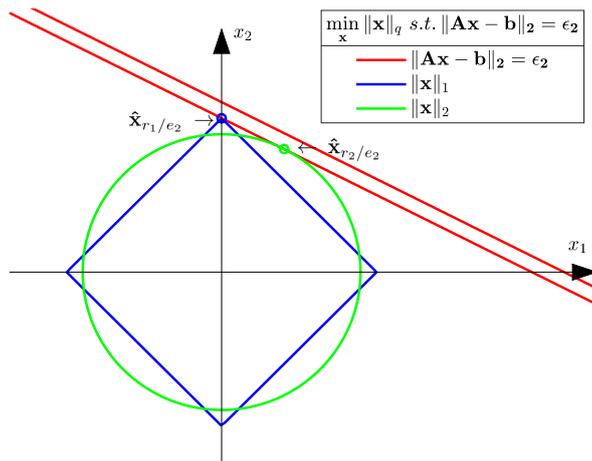


Figura 1.6: Minima norma in ℓ_1 e in ℓ_2 con vincolo quadratico sul residuo, $\mathcal{P}_{r_q/\epsilon_2} : \min_{\mathbf{x}} \|\mathbf{x}\|_q \text{ s.t. } \|\mathbf{Ax} - \mathbf{b}\|_2 \leq \epsilon_2$ per $q \in \{1, 2\}$. $\hat{\mathbf{x}}_{r_1/\epsilon_2}$ e $\hat{\mathbf{x}}_{r_2/\epsilon_2}$ sono le soluzioni per $q = 1$ e $q = 2$ rispettivamente.

1.3.4 LASSO e Tikhonov, interpretazione Bayesiana, legame con MLE e MAP

Come è stato accennato all'inizio del paragrafo 1.3.1, i metodi a minimo errore di approssimazione solo legati alla massima verosimiglianza del vettore incognito \mathbf{x} date le misure \mathbf{b} , e prendono in considerazione solo i dati facendo assunzioni circa la distribuzione dell'errore di misura. I metodi di regolarizzazione invece, oltre ai dati e alle assunzioni sugli stessi, sfruttano anche informazioni o assunzioni a priori circa la variabilità del vettore incognito.

Partiamo dal generico problema inverso (1.3), che riportiamo per facilità di esposizione:

$$\begin{aligned} \mathbf{x} : \mathbf{Ax} + \mathbf{e} &= \mathbf{b} \\ \mathbf{A} &\in \mathbb{R}^{m \times n} \\ \mathbf{x} \in \mathbb{R}^n; \mathbf{b}, \mathbf{e} &\in \mathbb{R}^m \end{aligned} \tag{1.48}$$

dove

$$\begin{aligned} \mathbf{A} &= \{a_{i,j}\}; & \mathbf{x} &= \{x_j\}; & \mathbf{b} &= \{b_i\}; & \mathbf{e} &= \{e_i\} \\ i &= 1, \dots, m; & j &= 1, \dots, n; \end{aligned}$$

se si assume che le componenti dell'errore e_i siano IID Gaussiane, a media nulla e varianza σ^2 , ovvero:

$$e_i \sim N(0, \sigma^2), \quad i = 1, \dots, m \quad (1.49)$$

Chiamando

$$\mathbf{A}_i^\top = [a_{i,1}, \dots, a_{i,n}], \quad i = 1, \dots, m;$$

il vettore corrispondente alla i -esima riga della matrice \mathbf{A} , si ha che anche le singole misure sono Gaussiane, con media: $\mathbf{A}_i^\top \mathbf{x}$ e varianza σ^2

$$b_i \sim N(\mathbf{A}_i^\top \mathbf{x}, \sigma^2), \quad j = 1, \dots, m; \quad (1.50)$$

La *Likelihood Function* $\mathcal{L}(\cdot)$, ovvero la **funzione di verosimiglianza** è quindi data da:

$$\begin{aligned} \mathcal{L}(\mathbf{x}) &= \mathcal{L}(\mathbf{x} \mid b_1, \dots, b_m) = \mathcal{L}(\mathbf{x} \mid \mathbf{b}) \\ &:= P(\mathbf{b} \mid \mathbf{x}) \\ &= \prod_{i=1}^m \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(b_i - \mathbf{A}_i^\top \mathbf{x})^2}{2\sigma^2}} \end{aligned} \quad (1.51)$$

Dove con $P(\cdot)$ si indica la densità di probabilità, e in particolare con $P(\mathbf{b} \mid \mathbf{x})$ ci si riferisce alla probabilità che si realizzino le misure \mathbf{b} dato lo stato \mathbf{x} . Massimizzando la verosimiglianza rispetto allo stato incognito \mathbf{x} , e passando ai logaritmi⁶, si ha:

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MLE}} &= \arg \max_{\mathbf{x}} P(\mathbf{b} \mid \mathbf{x}) \\ &= \arg \max_{\mathbf{x}} \log P(\mathbf{b} \mid \mathbf{x}) \\ &= \arg \max_{\mathbf{x}} \sum_{i=1}^m \left[-\frac{1}{2\sigma^2} (b_i - \mathbf{A}_i^\top \mathbf{x})^2 \right] \\ &= \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \\ &= \hat{\mathbf{x}}_2 \end{aligned} \quad (1.52)$$

e quindi ritroviamo il metodo dei minimi quadrati LS (1.27), ovvero la minimizzazione del residuo in ℓ_2 senza limitazioni sul termine di regolarizzazione, \mathcal{P}_{e_2/r_u} . Il termine da minimizzare prende anche il nome di, *Residual Sum of Squares* (RSS):

$$\text{RSS} = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \quad (1.53)$$

⁶Il passaggio è consentito dal fatto che il logaritmo naturale, come tutti i logaritmi in una qualunque base maggiore di 1, è una funzione monotona crescente.

Applicando lo stesso procedimento nell'ipotesi di errore IID Laplaciano (1.30), si verifica che, in questo caso, lo stimatore di massima verosimiglianza lo si ottiene minimizzando la somma dei valori assoluti del residuo, LAV, ovvero minimizzando il residui in ℓ_1 senza vincoli sul termine regolarizzazione, \mathcal{P}_{e_1/r_u} , (1.29). Dopo avere verificato che i metodi fondati sulla minimizzazione del residuo, senza ulteriori vincoli, sono legati alla verosimiglianza, equivalente a $P(\mathbf{b}|\mathbf{x})$, affrontiamo il problema della determinazione della massimizzazione della probabilità a posteriori $P(\mathbf{x}|\mathbf{b})$ introducendo assunzioni a priori sul vettore incognito \mathbf{x} . Parlando di probabilità condizionata è necessario ricordare il **Teorema di Bayes** che nel caso del problema (1.48) può essere scritto come:

$$P(\mathbf{x} | \mathbf{b}) = \frac{P(\mathbf{b} | \mathbf{x}) P(\mathbf{x})}{P(\mathbf{b})} \quad (1.54)$$

che può essere interpretato come

$$Posterior = \frac{Likelihood \cdot Prior}{Evidence} \quad (1.55)$$

Il problema della massima probabilità a posteriori, **Maximum A Posteriori Probability** (MAP), nel caso del problema di misura (1.48) può essere formalizzato come:

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MAP}} &= \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{b}) \\ &= \arg \max_{\mathbf{x}} \frac{P(\mathbf{b} | \mathbf{x}) P(\mathbf{x})}{P(\mathbf{b})} \\ &= \arg \max_{\mathbf{x}} P(\mathbf{b} | \mathbf{x}) P(\mathbf{x}) \\ &= \arg \max_{\mathbf{x}} \log(P(\mathbf{b} | \mathbf{x}) P(\mathbf{x})) \\ &= \arg \max_{\mathbf{x}} [\log P(\mathbf{b} | \mathbf{x}) + \log P(\mathbf{x})] \end{aligned} \quad (1.56)$$

Assumiamo come in precedenza che le componenti dell'errore e_i siano IID Gaussiane e che seguano la (1.49), e di conseguenza le misure b_i siano anch'esse Gaussiane e seguano la (1.50), la *Likelihood* in (1.56) è data dalla (1.51). Per quanto riguarda la *Prior*, ovvero $P(\mathbf{x})$, assumiamo che tutte le componenti del vettore \mathbf{x} , ovvero x_j , $j = 1, \dots, n$, siano IID e consideriamo due diversi casi:

- *Prior* Gaussiane con media nulla e varianza γ^2

$$\begin{aligned} x_j \sim N(0, \gamma^2) &= \frac{1}{\gamma\sqrt{2\pi}} e^{-\frac{x_j^2}{2\gamma^2}}, \\ &j = 1, \dots, n \end{aligned} \quad (1.57)$$

e quindi

$$P(\mathbf{x}) = \prod_{j=1}^n \frac{1}{\gamma\sqrt{2\pi}} e^{-\frac{x_j^2}{2\gamma^2}} \quad (1.58)$$

- *Prior* Laplaciana con media nulla varianza $2\tau^2$

$$x_j \sim \text{Laplace}(0, \tau) = \frac{1}{2\tau} e^{-\frac{|x_j|}{\tau}}, \quad (1.59)$$

$$j = 1, \dots, n$$

e quindi

$$P(\mathbf{x}) = \prod_{j=1}^n \frac{1}{2\tau} e^{-\frac{|x_j|}{\tau}} \quad (1.60)$$

1. Dall'applicazione della MAP (1.56) al caso con **errori di misura** e_i e **Prior** x_j **entrambi Gaussiani** si ha:

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MAP}} &= \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{b}) \\ &\vdots \\ &= \arg \max_{\mathbf{x}} [\log P(\mathbf{b} | \mathbf{x}) + \log P(\mathbf{x})] \\ &= \arg \max_{\mathbf{x}} \left(\log \prod_{i=1}^m \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(b_i - \mathbf{A}_i^T \mathbf{x})^2}{2\sigma^2}} + \log \prod_{j=1}^n \frac{1}{\gamma\sqrt{2\pi}} e^{-\frac{x_j^2}{2\gamma^2}} \right) \\ &= \arg \max_{\mathbf{x}} \left\{ \sum_{i=1}^m \left[-\frac{(b_i - \mathbf{A}_i^T \mathbf{x})^2}{2\sigma^2} \right] + \sum_{j=1}^n \left(-\frac{x_j^2}{2\gamma^2} \right) \right\} \\ &= \arg \min_{\mathbf{x}} \frac{1}{2\sigma^2} \left[\sum_{i=1}^m (b_i - \mathbf{A}_i^T \mathbf{x})^2 + \frac{\sigma^2}{\gamma^2} \sum_{j=1}^n x_j^2 \right] \\ &= \arg \min_{\mathbf{x}} \left(\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \mu_{2,2} \|\mathbf{x}\|_2^2 \right) \\ &= \hat{\mathbf{x}}_{e_2+r_2} \end{aligned} \quad (1.61)$$

ritroviamo quindi la regolarizzazione di Tikhonov (1.43) con $\mu_{2,2} = \frac{\sigma^2}{\gamma^2}$

2. Dall'applicazione della MAP (1.56) al caso con **errori di misura** e_i e

Gaussiani e *Prior* x_j **Laplaciane** si ha:

$$\begin{aligned}
 \hat{\mathbf{x}}_{\text{MAP}} &= \arg \max_{\mathbf{x}} P(\mathbf{x} \mid \mathbf{b}) \\
 &\quad \vdots \\
 &= \arg \max_{\mathbf{x}} [\log P(\mathbf{b} \mid \mathbf{x}) + \log P(\mathbf{x})] \\
 &= \arg \max_{\mathbf{x}} \left(\log \prod_{i=1}^m \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(b_i - \mathbf{A}_i^T \mathbf{x})^2}{2\sigma^2}} + \log \prod_{j=1}^n \frac{1}{2\tau} e^{-\frac{|x_j|}{\tau}} \right) \\
 &= \arg \max_{\mathbf{x}} \left\{ \sum_{i=1}^m \left[-\frac{(b_i - \mathbf{A}_i^T \mathbf{x})^2}{2\sigma^2} \right] + \sum_{j=1}^n \left(-\frac{|x_j|}{\tau} \right) \right\} \\
 &= \arg \min_{\mathbf{x}} \frac{1}{2\sigma^2} \left[\sum_{i=1}^m (b_i - \mathbf{A}_i^T \mathbf{x})^2 + \frac{2\sigma^2}{\tau} \sum_{j=1}^n |x_j| \right] \\
 &= \arg \min_{\mathbf{x}} \left(\|\mathbf{Ax} - \mathbf{b}\|_2^2 + \mu_{2,1} \|\mathbf{x}\|_1 \right)
 \end{aligned} \tag{1.62}$$

ritroviamo quindi la regolarizzazione LASSO (1.44) con $\mu_{2,1} = \frac{2\sigma^2}{\tau}$.

Per concludere quindi, sia la regolarizzazione di Tikhonov sia quella di LASSO sono legati alla MAP, ed assumono errori di misura Gaussiani. Le due regolarizzazioni differiscono per le assunzioni sulle *Prior*, Gaussiane per Tikhonov e quindi valutate in ℓ_2 , Laplaciane e quindi valutate in ℓ_1 per la regolarizzazione di LASSO nelle varie formulazioni. Valutando il termine di regolarizzazione in ℓ_1 , come accennato in precedenza, vengono selezionati vettori *sparsi*, argomento del prossimo paragrafo. La regolarizzazione in ℓ_2 , chiamata semplicemente regolarizzazione, sarà affrontata in modo analitico in 1.5

1.4 Problemi sotto-determinati e *Compressive Sensing*

1.4.1 Introduzione

Quando il problema inverso (1.3) è sotto-determinato, ovvero con un numero di misure inferiore al numero di incognite $m < n$, non è possibile in generale ricostruire lo stato. Il sistema ammette infatti ∞^{n-m} soluzioni, anche se è possibile sceglierne una aggiungendo dei vincoli, come mostrato nel paragrafo precedente. La teoria del *Compressive Sensing* (CS) [17–19], detta anche *Compressive Sampling*, sviluppa un approccio che consente, in determinate condizioni, di ricostruire in modo esatto, o con buona approssimazione, alcune tipologie di segnale, anche in presenza di un numero di misurazioni ridotte. Il punto di partenza è l'osservazione del fatto che alcune tipologie di segnale sono bene

approssimate da una espansione sparsa in una opportuna base.⁷ Affinché sia possibile ricostruire un segnale a partire da un numero ridotto di misure, sono necessarie delle condizioni sul vettore da ricostruire e sulla matrice di misura; è necessario inoltre disporre di adeguati strumenti matematici per ricercare in modo adeguato la soluzione.

1.4.2 Sparsità e algoritmi

Il segnale da ricostruire deve essere *sparsa*.

Definizione 1.4.1. Un vettore $\mathbf{x} \in \mathbb{R}^n$ è detto *S-sparso* con $S \in \mathbb{N}$, $S < n$ se:

$$\|\mathbf{x}\|_0 \leq S \quad (1.63)$$

Indichiamo con

$$\Sigma_S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_0 \leq S\} \quad (1.64)$$

Il sottoinsieme dei vettori *S-sparsi* di \mathbb{R}^n . Se $S \ll n$ il vettore si dice semplicemente *sparsa*.

Il problema del CS è la ricostruzione di un vettore sparso a partire da un numero ridotto di misure, quindi con $m < n$, formalmente questo porta ad una formulazione in termini di minimizzazione in ℓ_0 della soluzione con vincolo di uguaglianza, quindi alla formulazione $\mathcal{P}_{r_0/e=0}$ nota anche come problema (P0) [11]:

$$(P0) \quad \hat{\mathbf{x}}_{r_0} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \quad (1.65)$$

Il problema è di tipo combinatorio di minimizzazione, è cioè NP-hard, ovvero non è risolvibile con un algoritmo di complessità polinomiale, risulta quindi numericamente intrattabile all'aumentare delle dimensioni del problema. In ogni caso, per problemi di piccole medie dimensione è gestibile riscrivendolo in termini di *Mixed Integer Programming* (MIP) con vincoli e funzione obiettivo lineari, *Mixed Integer Linear Programming* (Mixed Integer Linear Programming (MILP)), ovvero come problema di programmazione lineare intera, più precisamente binaria, e risolvibile ad esempio con l'algoritmo *branch and bound* implementato in un *solver* come CPLEX o GURUBI. In Appendice 1.D viene presentata la riscrittura del problema (P0) in termini di MILP. La complessità computazione della minimizzazione in ℓ_0 non è l'unico motivo per cercare soluzioni alternative, infatti non sempre i segnali sono esattamente sparsi, più

⁷Questo fatto è noto da tempo ed è utilizzato ad esempio nei comuni dispositivi digitali, è infatti alla base della compressione *lossy*, ovvero con perdita, della voce (G723, G729, GSM), audio (MP3, AAC ed altri), immagini (JPEG, JPEG-2000) o video (MPEG2, MPEG4, H263, H264, H265) etc. In tutti questi algoritmi, il segnale viene rappresentato in una base nella quale è sparso, ovvero è possibile rappresentarlo efficientemente con un numero ridotto di coefficienti.

spesso hanno un certo numero di componenti, in modulo, piccole ma non nulle. Un vettore così fatto è bene approssimabile da un vettore *S-sparso*.

Definizione 1.4.2. \mathbf{x}_S è il vettore *S-sparso* che meglio approssima $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x}_S := \arg \min_{\mathbf{x}^*: \|\mathbf{x}^*\|_0 \leq S} \|\mathbf{x}^* - \mathbf{x}\|_p, \quad \forall p \in \mathbb{R}^+ \cup \{+\infty\} \quad (1.66)$$

\mathbf{x}_S può essere ottenuto a partire dal vettore \mathbf{x} ponendo a zero tutte le $n - S$ componenti più piccole in modulo. Nel paragrafo precedente è stato osservato come la minimizzazione del termine di regolarizzazione in ℓ_1 seleziona una soluzione sparsa (Esempi 1.3.3, 1.3.4, 1.3.5). Si passa quindi all’algoritmo conosciuto come *Basis Pursuit* (BP), [16], riportato anche come problema (P1) [11]:

$$(P1) \quad \hat{\mathbf{x}}_{r_1} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b} \quad (1.67)$$

equivalente al problema $\mathcal{P}_{r_1/e=0}$. È possibile riscrivere il problema (P1) come problema LP in **Forma Standard**, come mostrato in Appendice 1.E.

Tuttavia, sia il problema (P0) sia il problema (P1) prendono in considerazione il vincolo di uguaglianza, e sono quindi rivolti ad un problema inverso esatto (1.1). Nella realtà tutti i problemi reali hanno a che fare con le misure, e quindi l’errore di misura. Il problema reale che quindi si deve affrontare è un problema affetto da errore (1.3). Per affrontare il problema è necessario rilassare il vincolo di uguaglianza e sostituirlo con un vincolo sull’errore di approssimazione valutato in ℓ_2 pervenendo al problema (P2) in [11], noto anche come *Basis Pursuit De-Noising* (BPDN) in [16], o LASSO in [9], (1.3.4, 1.3.5).

$$(P2) \quad \hat{\mathbf{x}}_{r_1/e_2} = \mathbf{x} : \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \epsilon_2 \quad (1.68)$$

L’algoritmo (P2) è definito a meno del parametro ϵ_2 e, qualora fosse possibile determinare una stima dell’energia del vettore di errore \mathbf{e} in (1.3), è possibile sfruttare il “Principio di Discrepanza”, *Discrepancy Principle* (DP), di Morozov [20]. Il principio della discrepanza di Morozov suggerisce di scegliere ϵ_2 in modo tale che valga la relazione:

$$\|\mathbf{e}\|_2^2 \leq \epsilon_2^2 \quad (1.69)$$

Ovvero il parametro al quadrato deve essere dello stesso ordine di grandezza del quadrato della norma del vettore dell’errore di misura, ovvero della sua energia.

1.4.3 Caratteristiche matrici di misura

Introduciamo ora alcuni concetti fondamentali del CS che consentono di caratterizzare la matrice di misura e, in funzione di essa, dare indicazioni sulla capacità degli algoritmi di ricostruire una soluzione sparsa.

Definizione 1.4.3. Data una matrice $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \leq n$, e siano $\mathbf{a}_j \in \mathbb{R}^n$, $j = 1, \dots, n$ le sue colonne, si definisce coerenza della matrice \mathbf{A} , indicata con $\mu(\mathbf{A})$, il massimo tra i prodotti scalari delle sue colonne, normalizzate:

$$\mu(\mathbf{A}) = \max_{1 \leq i \neq j \leq n} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2} \quad (1.70)$$

Si ha che:

$$\mu(\mathbf{A}) \in \left[\sqrt{\frac{n-m}{m(n-1)}}, 1 \right] \quad (1.71)$$

Affinché sia possibile ricostruire una soluzione sparsa la matrice deve essere **incoerente**. La coerenza, o meglio l'incoerenza, è legata al concetto di ortogonalità: se due colonne sono parallele, il prodotto scalare normalizzato è pari a 1 e la coerenza è massima. Se $m = n$ e \mathbf{A} è una matrice ortogonale, la coerenza è nulla. Il valore minimo di coerenza è ottenuto da una matrice le cui colonne costituiscono un telaio stretto equiangolare, *Equiangular Tight Frame*, ETF [18].

Definizione 1.4.4. Data una matrice $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \leq n$, e un numero $S < n$, è detta “costante di isometria ristretta” δ_S il più piccolo numero tale che:

$$(1 - \delta_S) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_S) \|\mathbf{x}\|_2^2 \quad (1.72)$$

$$\forall \mathbf{x} : \|\mathbf{x}\|_0 \leq S$$

Se $\delta_S \in (0, 1)$, La matrice \mathbf{A} soddisfa la “proprietà di isometria ristretta”, *Restricted Isometry Properties*, (RIP) di ordine S e costante δ_S .

Il nome deriva dal fatto che se \mathbf{A} fosse una isometria, fosse cioè una matrice ortogonale, allora $\delta_S = 0$, come è immediato verificare dalle equazioni (1.22).

Teorema 1.4. *Se la matrice \mathbf{A} soddisfa RIP di ordine $2S$ con $\delta_{2S} < \sqrt{2} - 1$, siano $\hat{\mathbf{x}}_{r_1}$ la soluzione di (1.67), $\mathbf{x} \in \mathbb{R}^n$ e \mathbf{x}_S è il vettore S -sparso $\in \mathbb{R}^n$ con tutti gli elementi di \mathbf{x} posti a 0, tranne gli S elementi con modulo maggiore, allora $\exists C_{1,0} \in \mathbb{R}^+$ per cui valgono:*

$$\|\hat{\mathbf{x}}_{r_1} - \mathbf{x}\|_2 \leq C_{1,0} \cdot \frac{\|\mathbf{x} - \mathbf{x}_S\|_1}{\sqrt{S}} \quad (1.73)$$

$$\|\hat{\mathbf{x}}_{r_1} - \mathbf{x}\|_1 \leq C_{1,0} \cdot \|\mathbf{x} - \mathbf{x}_S\|_1$$

Il Teorema 1.4 assicura che se \mathbf{x} è S -sparso allora (1.67) trova la soluzione esatta. Se invece \mathbf{x} non è S -sparso, il teorema assicura che la soluzione ricostruita $\hat{\mathbf{x}}_{r_1}$ è di qualità simile a quella ricostruibile con la miglior approssimazione S -sparso di \mathbf{x} .

Teorema 1.5. *Se la matrice \mathbf{A} soddisfa la RIP di ordine $2S$ con $\delta_{2S} < \sqrt{2} - 1$, siano $\hat{\mathbf{x}}_{r_1/e_2}$ la soluzione di (1.68) con il vincolo sulla norma della discrepanza, $\mathbf{x} \in \mathbb{R}^n$ e \mathbf{x}_S è il vettore S -sparso $\in \mathbb{R}^n$ con tutti gli elementi di \mathbf{x} posti a 0,*

tranne gli S elementi con modulo maggiore, allora $\exists C_{2,0}, C_{2,1} \in \mathbb{R}^+$ per cui vale:

$$\|\hat{\mathbf{x}}_{r_1/\epsilon_2} - \mathbf{x}\|_2 \leq C_{2,0} \cdot \frac{\|\mathbf{x} - \mathbf{x}_S\|_1}{\sqrt{S}} + C_{2,1} \cdot \epsilon_2 \quad (1.74)$$

In questo caso l'errore dato dalla soluzione di (1.68) è la somma di due termini, il primo è equivalente all'errore che si avrebbe in assenza di errore di misura, il secondo è proporzionale al livello del disturbo dell'errore di misura. Se la matrice \mathbf{A} soddisfa la RIP di ordine $2S$ con costante $\delta_{2S} = 1/4$ le due costanti sono $C_{2,0} \leq 5.5$ e $C_{2,1} \leq 6$ [19]. L'algoritmo (1.68), grazie al Teorema 1.5 garantisce robustezza al rumore. Le matrici con le caratteristiche migliori per il CS sono le matrici casuali, ad esempio le matrici Gaussiane o Bernoulliane [18, 19], le quali con alta probabilità soddisfano la RIP, posto che il numero di misure m soddisfi la:

$$m \geq C \cdot S \cdot \log(n/S) \quad (1.75)$$

con C una costante positiva che dipende dal caso specifico. Si noti che il numero delle misure necessario cresce in modo logaritmico rispetto al numero delle incognite n . L'equazione (1.75) dà quindi una indicazione sul numero minimo di misure per approssimare, con buone speranze di successo, la ricostruzione di un problema sparso attraverso gli algoritmi basati sulla minimizzazione del termine di regolarizzazione in ℓ_1 , ovvero con gli algoritmi (P1), e (P2), che rilassa il vincolo di uguaglianza del primo, considerando un errore di approssimazione legato all'energia dell'errore di misura. Il tema della stima dell'errore di misura sarà affrontato nel Paragrafo 1.6, mentre nel Capitolo 2 verrà affrontato il problema dell'individuazione delle sorgenti di inquinamento armonico in una rete di distribuzione elettrica tramite il CS. In particolare verranno utilizzati algoritmi derivanti dall'applicazione dei problemi (P1) e (P2); in quest'ultimo caso le prestazioni dipendono dalla accurata valutazione dell'energia dell'errore di misura.

1.5 Regolarizzazione in ℓ_2

In questo paragrafo verrà affrontato il tema della regolarizzazione di un problema di misura (1.3) che riportiamo di seguito per facilità di esposizione e di lettura:

$$\begin{aligned} \mathbf{x} : \mathbf{Ax} + \mathbf{e} &= \mathbf{b} \\ \mathbf{A} &\in \mathbb{R}^{m \times n} \\ \mathbf{x} \in \mathbb{R}^n; \mathbf{b}, \mathbf{e} &\in \mathbb{R}^m \end{aligned} \quad (1.76)$$

concentrandoci sul caso $m \geq n$. Anche in questo paragrafo, per semplicità di trattazione, affrontiamo il problema nel campo reale ma i discorsi sono analoghi anche nel campo complesso.

1.5.1 Decomposizione a valori singolari, SVD e GSVD

Per approssciare il problema introduciamo due importanti fattorizzazioni. La prima fattorizzazione è la decomposizione a valori singolari *Singular Value Decomposition* (SVD) ([3], [21, pag. 209 e seguenti], [22])

$$\begin{aligned}
 & \forall \mathbf{A} \in \mathbb{R}^{m \times n} \\
 & \exists \\
 & \mathbf{U} \in \mathbb{R}^{m \times m} : \mathbf{U}^\top \mathbf{U} = \mathbf{U} \mathbf{U}^\top = \mathbf{I}_m \in \mathbb{R}^{m \times m} \\
 & \mathbf{V} \in \mathbb{R}^{n \times n} : \mathbf{V}^\top \mathbf{V} = \mathbf{V} \mathbf{V}^\top = \mathbf{I}_n \in \mathbb{R}^{n \times n} \\
 & \mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p) \in \mathbb{R}^{m \times n} \\
 & \text{with } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0, p = \min(m, n) \\
 & \implies \mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top
 \end{aligned} \tag{1.77}$$

Si ha inoltre

$$\begin{aligned}
 & \text{rank}(\mathbf{A}) = r \Leftrightarrow \sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0; \\
 & \mathcal{N}(\mathbf{A}) = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}; \\
 & \mathcal{R}(\mathbf{A}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}; \\
 & \mathcal{N}(\mathbf{A}^\top) = \text{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}; \\
 & \mathcal{R}(\mathbf{A}^\top) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}; \\
 & \|\mathbf{A}\|_2 = \sigma_1; \\
 & \|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2} \\
 & \kappa_2(\mathbf{A}) = \frac{\sigma_1}{\sigma_r}.
 \end{aligned} \tag{1.78}$$

Dove, $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ e $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ sono i vettori colonna di \mathbf{U} e \mathbf{V} rispettivamente, l'operatore $\text{span}\{\cdot\}$ indica il sottospazio generato da un insieme di vettori colonna, quindi $\mathcal{R}(\mathbf{A})$ è il sottospazio generato dai vettori colonna di \mathbf{A} , κ_2 è il numero di condizionamento in ℓ_2 . La matrice \mathbf{A} può anche essere espressa come somma di r matrici di rango 1 costruite dal prodotto esterno tra i vettori colonna u_j e v_j delle matrici \mathbf{U} e \mathbf{V} :

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^\top \tag{1.79}$$

definendo quindi la *economy-size* SVD:

$$\begin{aligned}
 \mathbf{A}_r &= \mathbf{U}_r \Sigma_r \mathbf{V}_r^\top \\
 \mathbf{A}_r &\in \mathbb{R}^{m \times n} : \text{rank}(\mathbf{A}) = r \\
 \mathbf{U}_r &= [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r] \in \mathbb{R}^{m \times r} \\
 \mathbf{V}_r &= [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r] \in \mathbb{R}^{n \times r} \\
 \Sigma_r &= \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \in \mathbb{R}^{r \times r}
 \end{aligned} \tag{1.80}$$

Nel caso in cui $r \ll \max(m, n)$ la *economy-size* SVD offre enormi vantaggi computazionali, questo fatto verrà sfruttato nell'applicazione del Capitolo 3 nella quale la matrice da fattorizzare è fortemente *slim*, ovvero $m \gg n$.

La matrice definita come:

$$\mathbf{A}_k := \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^\top, \quad k \in [1, r] \tag{1.81}$$

è detta anche SVD troncata, *Truncated Singular Value Decomposition* (TSVD) ed il Teorema di Eckart afferma costituisce la migliore approssimazione di \mathbf{A} di rango k :

$$\|\mathbf{A} - \mathbf{A}_k\|_2 = \min_{\text{rank} \mathbf{B} \leq k} \|\mathbf{A} - \mathbf{B}\|_2 = \sigma_{k+1} \tag{1.82}$$

Il risultato precedente resta valido anche sostituendo la norma 2 matriciale con la norma di Frobenius.

Se la matrice \mathbf{A} è molto mal condizionata, dalla (1.78) si ha che i valori singolari decrescono molto velocemente ed è quindi sufficiente un $k \ll \min(m, n)$ per avere una eccellente approssimazione, ed una notevole riduzione sia del carico computazionale sia dell'uso della memoria. La TSVD, si vedrà in seguito, offre anche dei vantaggi in termini di regolarizzazione.

La decomposizione a valori singolari riveste un ruolo importante anche in statistica e nel ML essendo legata alla Analisi delle Componenti Principali, *Principal Component Analysis* (PCA).

Nel 1976 Van Loan [23] introduce la ***Generalized Singular Value Decomposition*** (GSVD), generalizzazione della decomposizione SVD, rivolgendosi ad

una coppia di matrici $\{\mathbf{A}, \mathbf{L}\}$:

$$\begin{aligned}
 & \mathbf{A} \in \mathbb{R}^{m \times n}; \mathbf{L} \in \mathbb{R}^{d \times n}; d \leq n; \mathcal{N}(\mathbf{A}) \cap \mathcal{N}(\mathbf{L}) = \{\emptyset\} \\
 & \exists \\
 & \mathbf{U} \in \mathbb{R}^{m \times m} : \mathbf{U}^\top \mathbf{U} = \mathbf{U} \mathbf{U}^\top = \mathbf{I}_m \in \mathbb{R}^{m \times m} \\
 & \mathbf{V} \in \mathbb{R}^{d \times d} : \mathbf{V}^\top \mathbf{V} = \mathbf{V} \mathbf{V}^\top = \mathbf{I}_d \in \mathbb{R}^{d \times d} \\
 & \mathbf{X} \in \mathbb{R}^{n \times n} : \text{rank}(\mathbf{X}) = n, \Rightarrow \text{non singolare} \\
 & \boldsymbol{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_d, 1, \dots, 1) \in \mathbb{R}^{m \times n} \\
 & \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^{d \times n} \\
 & 0 \leq \sigma_1 \leq \dots \leq \sigma_d \leq 1 \quad \text{and} \quad 1 \geq \lambda_1 \dots \geq \lambda_d \geq 0 \\
 & \boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} + \mathbf{L}^\top \mathbf{L} = \mathbf{I}_n \in \mathbb{R}^{n \times n} \Rightarrow \sigma_i^2 + \lambda_i^2 = 1, i = 1, \dots, n \\
 & \implies \mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{X}^\top, \quad \mathbf{L} = \mathbf{V} \boldsymbol{\Lambda} \mathbf{X}^\top
 \end{aligned} \tag{1.83}$$

La decomposizione GSVD ha in generale una complessità doppia rispetto alla decomposizione SVD della matrice \mathbf{A} .

1.5.2 Analisi SVD

Il problema di misura (1.76) può essere affrontato con il LS, ovvero con minimizzazione dell'errore di approssimazione in ℓ_2 , \mathcal{P}_{e_2/r_u} (1.27) che può essere risolto tramite le equazioni normali (1.31) e per mezzo della decomposizione a valori singolari SVD (1.77) diventa:

$$\hat{\mathbf{x}}_2 = \mathbf{V} \boldsymbol{\Sigma}^\dagger \mathbf{U}^\top \mathbf{b} = \mathbf{A}^\dagger \mathbf{b} \tag{1.84}$$

dove \mathbf{A}^\dagger è la **pseudo inversa di Monroe-Penrose** data da:

$$\begin{aligned}
 \mathbf{A}^\dagger & := \mathbf{V} \boldsymbol{\Sigma}^\dagger \mathbf{U}^\top \\
 \boldsymbol{\Sigma}^\dagger & = \text{diag} \left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_r}, 0, \dots, 0 \right) \\
 r & = \text{rank}(\mathbf{A}) \leq n \leq m
 \end{aligned} \tag{1.85}$$

E ora possibile scrivere la **soluzione dei minimi quadrati $\hat{\mathbf{x}}_2$ in forma esplicita**:

$$\hat{\mathbf{x}}_2 = \sum_{j=1}^r \frac{\mathbf{u}_j^\top \mathbf{b}}{\sigma_j} \mathbf{v}_j \tag{1.86}$$

che esprime la soluzione dei minimi quadrati nella base $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, e i coefficienti $\mathbf{u}_j^\top \mathbf{b} \in \mathbb{R}$ sono anche detti coefficienti di Fourier associati allo spettro $\{\sigma_1, \dots, \sigma_n\}$. La propagazione dell'errore di misura introdotta nel Paragrafo 1.1 con l'equazione (1.4) in termini di una norma generica, può ora essere espressa

in termini di ℓ_2

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}_2\|_2}{\|\mathbf{x}\|_2} \leq \kappa_2(\mathbf{A}) \frac{\|\mathbf{b}_{\text{true}} - \mathbf{b}\|_2}{\|\mathbf{b}_{\text{true}}\|_2} \quad (1.87)$$

Introducendo esplicitamente il vettore d'errore \mathbf{e}^8 nella soluzione dei minimi quadrati (1.86) si ha:

$$\begin{aligned} \hat{\mathbf{x}}_2 &= \sum_{j=1}^r \frac{\mathbf{u}_j^\top \mathbf{b}}{\sigma_j} \mathbf{v}_j = \sum_{j=1}^r \frac{\mathbf{u}_j^\top \mathbf{b}_{\text{true}}}{\sigma_j} \mathbf{v}_j + \sum_{j=1}^r \frac{\mathbf{u}_j^\top \mathbf{e}}{\sigma_j} \mathbf{v}_j \\ &= \hat{\mathbf{x}}_{2,\text{exact}} + \sum_{j=1}^r \frac{\mathbf{u}_j^\top \mathbf{e}}{\sigma_j} \mathbf{v}_j \end{aligned} \quad (1.88)$$

dove $\hat{\mathbf{x}}_{2,\text{exact}}$ è la soluzione ai minimi quadrati in assenza di errore di misura, ovvero è la soluzione di (1.1).

Dalla (1.77) discende che gli scalari σ_j decrescono all'aumentare dell'indice j e decrescono molto velocemente se \mathbf{A} è mal condizionata⁹. Il Sistema (1.1) è un modello matematico di un problema fisico che lega lo stato \mathbf{x} alle grandezze fisiche \mathbf{b}_{true} attraverso la matrice di misura \mathbf{A}^{10} ; questo matematicamente significa che il vettore \mathbf{b}_{true} può essere espresso come combinazione lineare delle colonne di \mathbf{A} , ovvero significa che:

$$\mathbf{b}_{\text{true}} \in \mathcal{R}(\mathbf{A}) \quad (1.89)$$

il che implica che i coefficienti $\mathbf{u}_j^\top \mathbf{b}_{\text{true}} \in \mathbb{R}$ decrescano almeno quando σ_j .

Viceversa si può assumere che:

$$\mathbf{e} \notin \mathcal{R}(\mathbf{A}) \quad (1.90)$$

e di conseguenza $\mathbf{u}_j^\top \mathbf{e}$ non decresce come σ_j , se il sistema è molto mal condizionato questo comporta che:

$$\left\| \sum_{\sigma_j > 0} \frac{\mathbf{u}_j^\top \mathbf{e}}{\sigma_j} \mathbf{v}_j \right\|_2 \gg \|\hat{\mathbf{x}}_{2,\text{exact}}\|_2 \quad (1.91)$$

rendendo necessario un processo di regolarizzazione che consenta di tagliare i contributi delle alte frequenze, ovvero i contributi dei σ_j piccoli (j grandi) di $\sum_{\sigma_j > 0} \frac{\mathbf{u}_j^\top \mathbf{e}}{\sigma_j} \mathbf{v}_j$. La **Picard Condition** discreta, asserisce che una condizione necessaria per una buona regolarizzazione è che i coefficienti di Fourier decadano più velocemente rispetto ai valori singolari [24].

⁸il segno del vettore di errore \mathbf{e} è stato cambiato per semplicità, senza perdere di generalità.

⁹Dalla (1.78), se $\kappa_2(\mathbf{A})$ è molto grande, $\sigma_1 \gg \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$.

¹⁰Solitamente l'equazione (1.1) è il risultato di una linearizzazione, in questo caso il vettore \mathbf{b}_{true} non rappresenta esattamente le grandezze fisiche ma le rappresenta a meno dell'approssimazione del primo ordine.

1.5.3 Filtraggio

Introducendo un filtro Φ nella (1.86), formulazione esplicita della soluzione ai minimi quadrati del problema (1.76), si ha:

$$\hat{\mathbf{x}}_{2,\text{filt}} = \mathbf{V}\Psi\Sigma^\dagger\mathbf{U}^\top\mathbf{b} = \mathbf{A}_{\text{filt}}^\dagger\mathbf{b} = \sum_{\sigma_j>0} \phi_j \frac{\mathbf{u}_j^\top\mathbf{b}}{\sigma_j} \mathbf{v}_j \quad (1.92)$$

$$\phi_j = \begin{cases} \sim 0 & \sigma_j \text{ piccoli} \\ \sim 1 & \sigma_j \text{ grandi} \end{cases} \quad (1.93)$$

$$\Phi = \text{diag}(\phi_j)$$

e i termini $\phi_j \mathbf{u}_j^\top \mathbf{b}$ sono i nuovi coefficienti di Fourier. Il Teorema di Eckart (1.82) suggerisce la soluzione basata sulla TSVD:

$$\hat{\mathbf{x}}_{\text{TSVD}} = \sum_{j=1}^k \frac{\mathbf{u}_j^\top\mathbf{b}}{\sigma_j} \mathbf{v}_j = \sum_{\sigma_j>0} \phi_j \frac{\mathbf{u}_j^\top\mathbf{b}}{\sigma_j} \mathbf{v}_j \quad (1.94)$$

con

$$\phi_{\text{TSVD},j} = \begin{cases} 1 & j \leq k \\ 0 & j > k \end{cases} \quad (1.95)$$

$$\Phi_{\text{TSVD}} = \text{diag}(\phi_{\text{TSVD},j})$$

nella quale il parametro k è il parametro di regolarizzazione da determinare in maniera opportuna.

1.5.4 Regolarizzazione di Tikhonov

Un metodo di filtraggio alternativo e più flessibile viene fornito dalla regolarizzazione di Tikhonov (che è una formulazione $\mathcal{P}_{e_2+r_2}$), già introdotta nei paragrafi 1.3.3 e 1.3.4. Riscriviamo (1.43) omettendo qualche pedice:

$$\hat{\mathbf{x}}_{\text{Tikhonov}} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \mu \|\mathbf{x}\|_2^2, \mu \in \mathbb{R}^+ \quad (1.96)$$

che è equivalente al problema ai minimi quadrati:

$$\hat{\mathbf{x}}_{\text{Tikhonov}} = \arg \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A} \\ \sqrt{\mu}\mathbf{I}_n \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0}_{n \times 1} \end{bmatrix} \right\|_2^2, \mu \in \mathbb{R}^+ \quad (1.97)$$

infatti, applicando le equazioni normali all'ultima formulazione si ha:

$$(\mathbf{A}^\top\mathbf{A} + \mu\mathbf{I}_n) \hat{\mathbf{x}}_{\text{Tikhonov}} = \mathbf{A}^\top\mathbf{b} \quad (1.98)$$

che è facile ricondurre all'equazione precedente. Appliciamo ora la decomposizione SVD alla matrice \mathbf{A} abbiamo:

$$\begin{aligned} \hat{\mathbf{x}}_{\text{Tikhonov}} &= \mathbf{V}(\Sigma^\top\Sigma + \mu\mathbf{I}_n)^{-1}\Sigma^\top\mathbf{U}^\top\mathbf{b} \\ &= \sum_{\sigma_j>0} \phi_{\text{Tikhonov},j} \frac{\mathbf{u}_j^\top\mathbf{b}_w}{\sigma_j} \mathbf{v}_j \end{aligned} \quad (1.99)$$

dove

$$\phi_{\text{Tikhonov},j} = \frac{\sigma_j^2}{\sigma_j^2 + \mu} \quad (1.100)$$

$$\Phi_{\text{Tikhonov}} = \text{diag}(\phi_{\text{Tikhonov},j})$$

che si comporta come un filtro (1.93) per opportune scelte di μ , parametro di regolarizzazione.

1.5.5 Regolarizzazione di Tikhonov generalizzata

Nel Paragrafo 1.3.2, si è accennato all'uso di una matrice $\mathbf{L} \in \mathbb{R}^{d \times n}$, $d \leq n$, in Appendice 1.C è possibile trovare qualche esempio, diversa dall'identità, per selezione caratteristiche particolari nel vettore \mathbf{x} stimato. Quando si introduce tale matrice all'interno della formulazione di Tikhonov si parla di Tikhonov generalizzato (si tratta sempre di una formulazione $\mathcal{P}_{e_2+r_2}$) che ha la seguente forma:

$$\hat{\mathbf{x}}_{\text{TGen}} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \mu_g \|\mathbf{L}\mathbf{x}\|_2^2 \quad (1.101)$$

$$\mu_g \in \mathbb{R}^+$$

$$\mathbf{L} \in \mathbb{R}^{d \times n}, \quad d \leq n \quad (1.102)$$

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

equivalente al problema ai minimi quadrati:

$$\hat{\mathbf{x}}_{\text{TGen}} = \arg \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A} \\ \sqrt{\mu_g} \mathbf{L} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0}_{n \times 1} \end{bmatrix} \right\|_2^2 \quad (1.103)$$

e quindi a:

$$(\mathbf{A}^\top \mathbf{A} + \mu_g \mathbf{L}^\top \mathbf{L}) \hat{\mathbf{x}}_{\text{TGen}} = \mathbf{A}^\top \mathbf{b} \quad (1.104)$$

Applicando ora la GSVD (1.83) alla coppia di matrici $\{\mathbf{A}, \mathbf{L}\}$, tenendo conto delle condizioni (1.102), si ha:

$$\begin{aligned} (\mathbf{A}^\top \mathbf{A} + \mu_g \mathbf{L}^\top \mathbf{L}) \hat{\mathbf{x}}_{\text{TGen}} &= \mathbf{A}^\top \mathbf{b} \\ (\mathbf{X}\Sigma^\top \mathbf{U}^\top \mathbf{U}\Sigma \mathbf{X}^\top + \mu_g \mathbf{X}\Lambda^\top \mathbf{V}^\top \mathbf{V}\Lambda \mathbf{X}^\top) \hat{\mathbf{x}}_{\text{TGen}} &= \mathbf{X}\Sigma^\top \mathbf{U}^\top \mathbf{b} \\ \mathbf{X}(\Sigma^\top \Sigma + \mu_g \Lambda^\top \Lambda) \mathbf{X}^\top \hat{\mathbf{x}}_{\text{TGen}} &= \mathbf{X}\Sigma^\top \mathbf{U}^\top \mathbf{b} \end{aligned} \quad (1.105)$$

posto che $\mathcal{N}(\mathbf{A}) \cap \mathcal{N}(\mathbf{L}) = \{\emptyset\}$, e che \mathbf{X} è non singolare, è possibile invertire le matrici $(\Sigma^\top \Sigma + \mu_g \Lambda^\top \Lambda)$ e \mathbf{X} , tuttavia è numericamente preferibile risolvere il sistema:

$$\begin{aligned} \mathbf{X}^\top \hat{\mathbf{x}}_{\text{TGen}} &= (\Sigma^\top \Sigma + \mu_g \Lambda^\top \Lambda)^{-1} \Sigma^\top \mathbf{U}^\top \mathbf{b} \\ &= \sum_{\sigma_j > 0} \phi_{\text{TGen},j} (\mathbf{u}_j^\top \mathbf{b}) \mathbf{c}_j \end{aligned} \quad (1.106)$$

con

$$\phi_{\text{TGen},j} = \frac{\sigma_j^2}{\sigma_j^2 + \mu_g \lambda_j^2} \quad (1.107)$$

$$\Phi_{\text{TGen}} = \text{diag}(\phi_{\text{TGen},j})$$

avendo indicato con \mathbf{c}_j il j -esimo versore della base canonica in \mathbb{R}^n . Ricordando che: $\sigma_i^2 + \lambda_i^2 = 1$, $i = 1, \dots, d$, si nota che il metodo Tikhonov generalizzato opera un filtraggio con un parametro di regolarizzazione che varia per ogni valore singolare.

Nel caso in cui la matrice \mathbf{L} fosse quadrata e a rango pieno, quindi invertibile, ponendo:

$$\begin{aligned} \mathbf{y} &= \mathbf{L}\mathbf{x} \\ \mathbf{B} &= \mathbf{A}\mathbf{L}^{-1} \end{aligned} \quad (1.108)$$

e sostituendoli nell'espressione della regolarizzazione di Tikhonov generalizzata (1.101) si ottiene l'espressione standard:

$$\hat{\mathbf{y}}_{\text{Tikhonov}} = \arg \min_{\mathbf{y}} \|\mathbf{B}\mathbf{y} - \mathbf{b}\|_2^2 + \mu_g \|\mathbf{y}\|_2^2, \quad \mu_g \in \mathbb{R}^+ \quad (1.109)$$

$$\hat{\mathbf{x}}_{\text{TGen}} = \mathbf{L}^{-1} \hat{\mathbf{y}}_{\text{Tikhonov}} \quad (1.110)$$

Questo approccio, quando possibile, è preferibile perché la decomposizione GSVD ha una complessità superiore alla già di per sé esigente SVD. Nell'applicazione del Capitolo 3 è stato utilizzato questo approccio essendo la \mathbf{L} diagonale e a rango pieno.

Per problemi di grandi dimensioni, tali da rendere la decomposizione SVD, ed ancor meno la decomposizione GSVD, computazionalmente proibitive, è possibile ricorrere a metodi basati su proiezioni in sotto spazi di Krylov, [25–27].

1.5.6 Scelta del parametro di regolarizzazione

La regolarizzazione di Tikhonov è una formulazione $\mathcal{P}_{e_2+r_2}$, ovvero è un metodo di minimizzazione non vincolata che coinvolge l'errore di approssimazione e il termine di regolarizzazione, entrambi valutati in ℓ_2 al quadrato, in una formulazione con penalizzazione a meno del parametro di regolarizzazione μ , formulazione detta anche Lagrangiana. Inoltre, la regolarizzazione di Tikhonov agisce introducendo un filtro (1.100) nella soluzione ai minimi quadrati. Sia guardando alla formulazione Lagrangiana del problema (1.43), sia guardando alla soluzione esplicita, quindi al filtro, appare chiaro che all'aumentare di μ aumenta l'effetto di regolarizzazione e di filtraggio¹¹, concedendo qualcosa in termini di errore di approssimazione o residuo. Viceversa, più piccolo è il pa-

¹¹Per quanto riguarda la regolarizzazione con la TSVD vale l'opposto: l'effetto filtrante e di regolarizzazione diminuisce all'aumentare di k , si veda (1.95).

parametro di regolarizzazione μ , minore è l'errore di approssimazione e minore è l'effetto filtrante. I metodi per selezionare il parametro di regolarizzazione, possono essere suddivisi in due categorie [28]:

- metodi che prescindono dalle informazioni sull'errore di misura.
- metodi basati sulla conoscenza o sulla disponibilità di una buona stima dell'energia dell'errore di misura $\|\mathbf{e}\|_2^2$.

Alla prima categoria appartengono il metodo basato sulla *L-curve* e il metodo chiamato *Generalized Cross Validation* GCV.

La **L-curve** [29] è uno strumento grafico di analisi di un problema inverso mal posto. La curva è costruita visualizzando i punti $(\|\mathbf{A}\hat{\mathbf{x}}_{2,\text{filt}} - \mathbf{b}\|_2, \|\mathbf{L}\hat{\mathbf{x}}_{2,\text{filt}}\|_2)$ in doppia scala logaritmica, al variare del parametro di regolarizzazione, $\hat{\mathbf{x}}_{2,\text{filt}}$ è la soluzione filtrata (regolarizzata o stimata). Il nome deriva dal fatto che la curva assume una tipica forma ad L. La Figura 1.7 mostra una generica *L-curve*. Tutti i metodi per determinare il parametro di regolarizzazione definiscono sulla curva un punto nella regione del gomito, area che rappresenta un buon compromesso tra regolarizzazione ed errore di approssimazione. In particolare, il criterio della *L-curve*, chiamato anche criterio **L-corner** individua il punto a massima curvatura, con il miglior compromesso tra errore di approssimazione e norma della soluzione. Il parametro μ corrispondente al *corner* viene selezionato come parametro di regolarizzazione. Va sottolineato che il tracciamento della

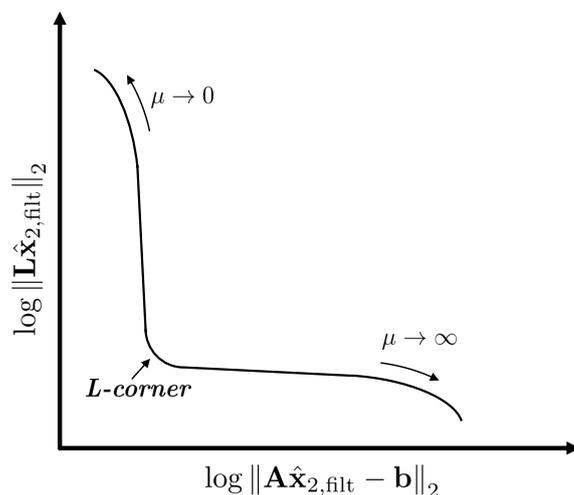


Figura 1.7: Forma generica della *L-curve*.

curva è reso possibile dal fatto che, per tutti i metodi di regolarizzazione presentati, esiste la forma esplicita della soluzione, grazie alla decomposizione SVD

(GSVD), per cui è possibile determinare i punti $(\|\mathbf{A}\hat{\mathbf{x}}_{2,\text{filt}} - \mathbf{b}\|_2, \|\mathbf{L}\hat{\mathbf{x}}_{2,\text{filt}}\|_2)$ al variare del parametro di regolarizzazione.

Il metodo della **GCV** è un metodo statistico, e si basa sulla considerazione che se un dato elemento del termine noto, ovvero una misura b_i fosse esclusa, (tecnica statica chiamata *Leave one out*), la corrispondente soluzione regolarizzata (o filtrata) dovrebbe essere in grado di darne una buona stima, inoltre, il parametro di regolarizzazione dovrebbe essere indipendente da trasformazioni ortogonali del vettore \mathbf{b} . Questo porta alla scelta del parametro di regolarizzazione che minimizza

$$G := \frac{\|\mathbf{A}\hat{\mathbf{x}}_{2,\text{filt}} - \mathbf{b}\|_2^2}{[\text{trace}(\mathbf{I}_m - \mathbf{A}\mathbf{A}_{2,\text{filt}})]^2} \quad (1.111)$$

dove $\hat{\mathbf{x}}_{2,\text{filt}}$ e $\mathbf{A}_{2,\text{filt}}$ sono, la soluzione regolarizzata e la matrice che, moltiplicata per \mathbf{b} , la produce:

$$\hat{\mathbf{x}}_{2,\text{filt}} = \mathbf{A}_{2,\text{filt}}\mathbf{b} \quad (1.112)$$

1.5.7 *Discrepancy Principle* per TSVD e Tikhonov

Alla seconda categoria appartiene solo il “metodo della discrepanza”, *Discrepancy Principle* (DP) di Morozov [20], già menzionato come criterio per fissare il parametro ϵ_2 (1.69) nell’algoritmo (P2) (1.68). Si noti come, anche nel (P2), l’errore di approssimazione è valutato in ℓ_2 , benché la regolarizzazione sia in ℓ_1 per favorire la sparsità della soluzione trovata. Il tema della stima della norma dell’errore di misura sarà affrontato nel Paragrafo 1.6. Il criterio del DP può essere applicato a tutti i metodi di filtro presentati, per cui, con riferimento alla formulazione generale: (1.92), può essere formulato come:

$$\|\mathbf{A}\hat{\mathbf{x}}_{2,\text{filt}} - \mathbf{b}\|_2^2 \simeq \|\mathbf{e}\|_2^2 \leq \delta^2 \quad (1.113)$$

- **TSVD e DP**

si vuole determinare k in (1.94) in modo che

$$k : \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{TSVD}}\|_2 \leq \delta \quad (1.114)$$

considerato che

$$\left\| \mathbf{b} - \mathbf{A} \sum_{j=1}^k \frac{\mathbf{u}_j^\top \mathbf{b}}{\sigma_j} \mathbf{v}_j \right\|_2 \geq \left\| \mathbf{b} - \mathbf{A} \sum_{j=1}^{k+1} \frac{\mathbf{u}_j^\top \mathbf{b}}{\sigma_j} \mathbf{v}_j \right\|_2 \quad (1.115)$$

è sufficiente provare valori incrementali di k con criterio d’arresto

$$\left\| \mathbf{b} - \mathbf{A} \sum_{j=1}^k \frac{\mathbf{u}_j^\top \mathbf{b}}{\sigma_j} \mathbf{v}_j \right\|_2 \leq \delta \quad (1.116)$$

- **Tikhonov e DP**

si vuole determinare μ in (1.99) in modo che

$$\mu : \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{Tikhonov}}\|_2 \leq \delta \quad (1.117)$$

Sviluppando i calcoli utilizzando la fattorizzazione SVD

$$\begin{aligned} \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{Tikhonov}}\|_2^2 &= \left\| \mathbf{b} - \mathbf{A}(\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I}_n)^{-1} \mathbf{A}^\top \mathbf{b} \right\|_2^2 \\ &= \left\| \left[\mathbf{I}_m - \mathbf{\Sigma}(\mathbf{\Sigma}^\top \mathbf{\Sigma} + \mu \mathbf{I}_n)^{-1} \mathbf{\Sigma}^\top \right] \mathbf{U}^\top \mathbf{b} \right\|_2^2 \\ &= \sum_{\sigma_j > 0} \left[\left(1 - \frac{\sigma_j^2}{\sigma_j^2 + \mu} \right) \mathbf{u}_j^\top \mathbf{b} \right]^2 \\ &= \sum_{\sigma_j > 0} \left[\left(\frac{1}{\beta \sigma_j^2 + 1} \right) \mathbf{u}_j^\top \mathbf{b} \right]^2 \end{aligned} \quad (1.118)$$

$$\text{dove } \beta = \frac{1}{\mu}$$

Definiamo la funzione $\Psi_{\text{Tikhonov}}(\beta)$:

$$\Psi_{\text{Tikhonov}}(\beta) := \sum_{\sigma_j > 0} \left(\frac{\mathbf{u}_j^\top \mathbf{b}}{\beta \sigma_j^2 + 1} \right)^2 - \delta^2 \quad (1.119)$$

In [30] si mostra come Ψ_{Tikhonov} sia convessa e monotona decrescente, inoltre, per ragionevoli valori di δ (norma dell'errore di misura), si ha¹²:

$$\begin{aligned} \Psi_{\text{Tikhonov}}(0) &= \sum_{\sigma_j > 0} (\mathbf{u}_j^\top \mathbf{b})^2 - \delta^2 > 0 \\ \lim_{\beta \rightarrow \infty} \Psi_{\text{Tikhonov}}(\beta) &= -\delta^2 < 0 \\ \implies \exists \beta^* : \Psi_{\text{Tikhonov}}(\beta^*) &= 0 \end{aligned} \quad (1.120)$$

che possiamo trovare col metodo di Newton utilizzando la derivata rispetto a β data da:

$$\Psi'_{\text{Tikhonov}}(\beta) = -2 \sum_{\sigma_j > 0} \frac{\sigma_j^2 (\mathbf{u}_j^\top \mathbf{b})^2}{(\beta \sigma_j^2 + 1)^3} \quad (1.121)$$

e poi scegliere $\mu^* = \frac{1}{\beta^*}$

- **Tikhonov generalizzato e DP**

si vuole determinare μ in (1.106) in modo che

$$\mu : \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{TGen}}\|_2 \leq \delta \quad (1.122)$$

Si procede in modo simile a quanto fatto per il metodo di Tikhonov standard, questa volta ricorrendo alla fattorizzazione GSVD (1.83). In Appendice 1.F i dettagli.

¹²Se \mathbf{A} è quadrata e a rango pieno si ha $\Psi_{\text{Tikhonov}}(0) = \|\mathbf{b}\|_2^2 - \delta^2$

1.6 Stima dell'energia dell'errore di misura

1.6.1 Valutazione del limite dell'energia dell'errore di misura

Nel Paragrafo 1.3 è stato evidenziato come la valutazione dell'errore di approssimazione in ℓ_2 sia la più robusta e quindi quella comunemente utilizzata¹³, in particolare nelle regolarizzazioni LASSO e Tikhonov, che differiscono per la scelta della norma utilizzata per la valutazione del termine di regolarizzazione: in ℓ_1 per LASSO, per selezionare soluzioni sparse ed ℓ_2 per Tikhonov. Il metodo LASSO, nella formulazione \mathcal{P}_{r_1/e_2} a minima norma con vincolo quadratico prende il nome di problema (P2) ed è un metodo utilizzato nel CS, come visto nel Paragrafo 1.4 dedicato alle formulazioni sparse per problemi sotto-determinati. La scelta del parametro di regolarizzazione del (P2) è basata sul *Discrepancy Principle*, e sempre su questo principio è possibile scegliere il parametro di regolarizzazione in ℓ_2 , come visto nel paragrafo precedente 1.5.7.

Il DP suggerisce di scegliere il parametro di regolarizzazione in maniera tale che l'errore di approssimazione sia circa uguale, o non superiori, la norma dell'errore di misura:

$$\|\mathbf{A}\hat{\mathbf{x}}_{reg} - \mathbf{b}\|_2^2 \simeq \|\mathbf{e}\|_2^2 \leq \delta^2 \quad (1.123)$$

dove con $\hat{\mathbf{x}}_{reg}$ indichiamo il vettore stimato (regolarizzato). Affinché possa essere applicato, è essenziale disporre di una stima dell'estremo superiore dell'energia dell'errore di misura:

$$\xi_{\mathbf{e}} = \|\mathbf{e}\|_2^2 \leq \delta^2 \quad (1.124)$$

dove $\xi_{\mathbf{e}}$ è una variabile aleatoria pari all'energia dell'errore. Conoscendo la media e la deviazione standard di $\xi_{\mathbf{e}}$, che indichiamo con $\mu_{\xi_{\mathbf{e}}}$ e $\sigma_{\xi_{\mathbf{e}}}$ rispettivamente, è possibile scegliere δ come:

$$\delta^2 = \mu_{\xi_{\mathbf{e}}} + \lambda\sigma_{\xi_{\mathbf{e}}} \quad (1.125)$$

dove la scelta di λ varia il livello di fiducia della stima associata alla disegualianza:

$$\xi_{\mathbf{e}} = \|\mathbf{e}\|_2^2 \leq \mu_{\xi_{\mathbf{e}}} + \lambda\sigma_{\xi_{\mathbf{e}}} \quad (1.126)$$

Tipicamente la scelta ricade su $\lambda \geq 2$ che permette di garantire un alto livello di fiducia.

¹³In assenza di esigenze specifiche come la gestione degli *outliers* oppure informazioni a priori che portino ad ipotizzare un errore di tipo uniforme. In questi due casi potrebbe essere preferibile valutare l'errore di approssimazione in ℓ_1 o ℓ_∞ rispettivamente.

In [17] vengono analizzati due casi nei quali le componenti e_i , $i = 1, \dots, m$ del vettore d'errore sono IID: in un caso sono Gaussiane, nell'altro sono uniformi. Dalla (1.126) discende che:

$$\xi_e = \sum_{i=1}^m e_i^2 \quad (1.127)$$

- Nel caso le e_i siano **Gaussiane**

$$e_i \sim N(0, \sigma^2), \quad i = 1, \dots, m \quad (1.128)$$

la (1.127) è la somma di Gaussiane al quadrato, che si comporta come una distribuzione χ^2 , in particolare:

$$\xi_e = \|\mathbf{e}\|^2 \sim \sigma^2 \chi^2(m) \quad (1.129)$$

di media $\sigma^2 m$ e deviazione standard $\sigma^2 \sqrt{2m}$, per cui è possibile utilizzare come stima:

$$\delta^2 = \sigma^2 (m + \lambda \sqrt{2m}) \quad (1.130)$$

- Nel caso le e_i siano **uniformi**

$$e_i \sim \mathcal{U}\left(-\frac{q}{2}, \frac{q}{2}\right), \quad i = 1, \dots, m \quad (1.131)$$

la ξ_e in (1.127) segue una somma di uniformi al quadrato di media $\frac{mq^2}{12}$ e deviazione standard $\frac{\sqrt{mq^2}}{6\sqrt{5}}$ per cui è possibile utilizzare come stima:

$$\delta^2 = \frac{mq^2}{12} + \lambda \frac{\sqrt{mq^2}}{6\sqrt{5}} \quad (1.132)$$

Le considerazioni precedenti sono valide quando le componenti dell'errore di misura seguono tutte la stessa distribuzione di probabilità, con gli stessi parametri. Queste condizioni sono difficilmente riscontrabili in casi pratici ed in particolare nei sistemi elettrici di potenza:

- le misure sono effettuate su quantità di diversa natura (tensioni, correnti, potenze con valori anche molto diversi) e parametri di diversa natura (modulo e fase ad esempio); conseguentemente le componenti hanno diversa accuratezza;
- nella formulazione dei problemi di stima l'errore dato dagli strumenti di misura subisce spesso delle trasformazioni per cui la distribuzione delle componenti del vettore d'errore e non è più quella originale.

1.6.2 Sbiancamento, *Whitening Transformation*

Per quanto concerne il primo punto, si può utilizzare la tecnica dello sbiancamento. Con riferimento al generico problema di misura (1.3), si supponga che il vet-

tore \mathbf{b} sia affetto da errore con matrice di covarianza ricavata dall'applicazione della legge di propagazione dell'incertezza [31] alla matrice

$$\Sigma_{\text{meas}} = \begin{bmatrix} \sigma_{\text{meas},1} & & & 0 \\ & \sigma_{\text{meas},2} & & \\ & & \ddots & \\ 0 & & & \sigma_{\text{meas},m} \end{bmatrix} \quad (1.133)$$

che include le incertezze standard $\sigma_{,i}$, $i = 1, \dots, m$ delle misure assunte decorrelate¹⁴. Si ha quindi:

$$\Sigma_{\mathbf{b}} = \mathbf{J}\Sigma_{\text{meas}}\mathbf{J}^T \quad (1.134)$$

dove \mathbf{J} è la matrice Jacobiana delle trasformazioni effettuate a partire dalle misure originali per ottenere \mathbf{b} .

Il vettore \mathbf{b} affetto da errore può essere visto come un vettore casuale, nel caso di valore atteso nullo, si ha:

$$\Sigma_{\mathbf{b}} = E[\mathbf{b}\mathbf{b}^T] \quad (1.135)$$

dove il simbolo $E[\cdot]$ è l'operatore aspettazione (*expectation*), o valore atteso. Operare lo sbiancamento, *Whitening Transformation*, di un vettore casuale significa applicare su di esso una trasformazione lineare in maniera tale che il vettore risultante abbia la matrice di identità come matrice di covarianza.

Definizione 1.6.1. La matrice $\mathbf{W} \in \mathbb{R}^{m \times m}$ è una matrice di sbiancamento per il vettore $\mathbf{b} \in \mathbb{R}^m$ se, applicata al vettore, lo trasforma in un vettore casuale $\mathbf{b}_W = \mathbf{W}\mathbf{b}$ avente la matrice di identità come matrice di covarianza:

$$\Sigma_{\mathbf{b}_W} = \mathbf{I}_m \in \mathbb{R}^{m \times m} \quad (1.136)$$

Teorema 1.6. La matrice \mathbf{W} può essere ottenuta per mezzo della fattorizzazione dell'inversa della matrice di covarianza di \mathbf{b} :

$$\begin{aligned} \forall \mathbf{W} \in \mathbb{R}^{m \times m} : \mathbf{W}^T \mathbf{W} &= \Sigma_{\mathbf{b}}^{-1} \\ \mathbf{b}_W \triangleq \mathbf{W}\mathbf{b} &\implies \Sigma_{\mathbf{b}_W} = \mathbf{I}_m \end{aligned} \quad (1.137)$$

Dimostrazione.

$$\begin{aligned} \Sigma_{\mathbf{b}_W} &= E[\mathbf{b}_W \mathbf{b}_W^T] \\ &= E[\mathbf{W}\mathbf{b}\mathbf{b}^T \mathbf{W}^T] = \mathbf{W} E[\mathbf{b}\mathbf{b}^T] \mathbf{W}^T \\ &= \mathbf{W}\Sigma_{\mathbf{b}} \mathbf{W}^T = \mathbf{W}\mathbf{W}^{-1} \mathbf{W}^{-T} \mathbf{W}^T \\ &= \mathbf{I}_m \end{aligned} \quad (1.138)$$

¹⁴Se le misure sono correlate, la matrice Σ_{meas} non è diagonale, ma il procedimento può essere eseguito comunque.

□

La matrice di covarianza $\Sigma_{\mathbf{b}}$ è simmetrica e definita positiva, è quindi possibile sfruttare la fattorizzazione di Cholesky:

Teorema 1.7. *Ogni matrice $\mathbf{A} \in \mathbb{R}^{m \times m}$ simmetrica e definita positiva è fattorizzabile in modo unico come:*

$$\mathbf{A} = \mathbf{R}^T \mathbf{R} \quad (1.139)$$

con $\mathbf{R} \in \mathbb{R}^{m \times m}$ matrice triangolare superiore.

È possibile operare in due modi:

- metodo 1, si inverte $\Sigma_{\mathbf{b}}$, poi si applica la fattorizzazione di Cholesky

$$\begin{aligned} \mathbf{W}_{\text{Ch}_1} &= \mathbf{R}_{\text{Ch}_1} \\ \text{con} & \\ \mathbf{R}_{\text{Ch}_1}^T \mathbf{R}_{\text{Ch}_1} &= \Sigma_{\mathbf{b}}^{-1} \end{aligned} \quad (1.140)$$

- metodo 2, si fattorizza la $\Sigma_{\mathbf{b}}$ con Cholesky, poi si inverte la matrice triangolare

$$\begin{aligned} \mathbf{W}_{\text{Ch}_2} &= \mathbf{R}_{\text{Ch}_2}^{-T} \\ \text{con} & \\ \mathbf{R}_{\text{Ch}_2}^T \mathbf{R}_{\text{Ch}_2} &= \Sigma_{\mathbf{b}} \end{aligned} \quad (1.141)$$

Sia \mathbf{W}_{Ch_1} sia \mathbf{W}_{Ch_2} soddisfano la (1.137) ma il metodo 2 è più efficiente perché la matrice invertita è triangolare. Da qui in poi $\mathbf{W}_C = \mathbf{W}_{\text{Ch}_2}$ indicherà la matrice di sbiancamento ottenuta attraverso la fattorizzazione di Cholesky con il metodo 2 (1.141). Moltiplicando la (1.3) per la matrice di sbiancamento si ottiene:

$$\begin{aligned} \mathbf{b}_{w_C} &= \mathbf{W}_C \mathbf{b} = \mathbf{W}_C \mathbf{A} \mathbf{x} + \mathbf{W}_C \mathbf{e} \\ &= \mathbf{A}_{w_C} \mathbf{x} + \mathbf{e}_{w_C} \end{aligned} \quad (1.142)$$

dove \mathbf{e}_{w_C} è il vettore dell'errore sbiancato e si ha $\Sigma_{\mathbf{b}_{w_C}} = \Sigma_{\mathbf{e}_{w_C}} = \mathbf{I}_m$

In [32], lavoro ripreso in modo esteso nel Capitolo 2, è stato mostrato come \mathbf{W}_C non sia in grado di risolvere il secondo problema per la determinazione dell'energia dell'errore. Il motivo è legato alla trasformazione che subisce il vettore $\mathbf{e}_{\text{meas}} = [e_{\text{meas},1}, \dots, e_{\text{meas},m}]^T$ infatti:

$$\mathbf{e} \simeq \mathbf{J} \mathbf{e}_{\text{meas}} \quad (1.143)$$

conseguentemente le e_i , componenti del vettore \mathbf{e} , non seguono la stessa distribuzione di probabilità delle $e_{,i}$ componenti del vettore \mathbf{e}_{meas} .

1.6.3 Proposta di una particolare matrice di sbiancamento e proprietà dell'errore sbiancato

\mathbf{W}_C non è l'unica matrice di sbiancamento possibile, anzi, le matrici di sbiancamento sono infinite, in particolare, nel caso in cui \mathbf{J} sia quadrata ed invertibile si ha:

$$\begin{aligned}\Sigma_{\mathbf{b}}^{-1} &= \mathbf{J}^{-\top} \Sigma_{\text{meas}}^{-\frac{1}{2}} \Sigma_{\text{meas}}^{-\frac{1}{2}} \mathbf{J}^{-1} = \\ &= (\mathbf{J}^{-\top} \Sigma_{\text{meas}}^{-\frac{1}{2}} \mathbf{Q}^\top) (\mathbf{Q} \Sigma_{\text{meas}}^{-\frac{1}{2}} \mathbf{J}^{-1}) \\ \forall \mathbf{Q} : \mathbf{Q} \mathbf{Q}^\top &= \mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_m\end{aligned}\tag{1.144}$$

e le infinite matrici

$$\mathbf{W}_Q \triangleq \mathbf{Q} \Sigma_{\text{meas}}^{-\frac{1}{2}} \mathbf{J}^{-1}\tag{1.145}$$

soddisfano la (1.137).

In [32], lavoro presentato in dettaglio nel Capitolo 2, viene proposta una nuova matrice di sbiancamento per il caso specifico in cui la \mathbf{J} è relativa alla trasformazione da coordinate polari a rettangolari. La matrice proposta in [32] può essere generalizzata per una qualunque \mathbf{J} quadrata ed invertibile. La matrice di sbiancamento definita come:

$$\mathbf{W}_R \triangleq \Sigma_{\text{meas}}^{-\frac{1}{2}} \mathbf{J}^{-1}\tag{1.146}$$

consente di invertire la trasformazione in (1.143), infatti:

$$\mathbf{e}_{w_R} \simeq \Sigma_{\text{meas}}^{-\frac{1}{2}} \mathbf{J}^{-1} \mathbf{J} \mathbf{e}_{\text{meas}} = \Sigma_{\text{meas}}^{-\frac{1}{2}} \mathbf{e}_{\text{meas}}\tag{1.147}$$

Se matrice $\Sigma_{\text{meas}}^{-\frac{1}{2}}$ è diagonale, \mathbf{e}_{w_R} è la versione riscalata di $\mathbf{e}_{w_{\text{meas}}}$ e le rispettive componenti seguono la stessa distribuzione di probabilità¹⁵, ed hanno tutte varianza unitaria $\Sigma_{\mathbf{b}_{w_R}} = \Sigma_{\mathbf{e}_{w_R}} = \mathbf{I}_m$

Applicando la trasformazione di sbiancamento \mathbf{W}_R al generico problema inverso (1.76) si

$$\begin{aligned}\mathbf{b}_{w_R} &= \mathbf{W}_R \mathbf{b} = \mathbf{W}_R \mathbf{A} \mathbf{x} + \mathbf{W}_R \mathbf{e} \\ &= \mathbf{A}_{w_R} \mathbf{x} + \mathbf{e}_{w_R}\end{aligned}\tag{1.148}$$

Nell'ipotesi di componenti e_i di errore Gaussiano (uniforme) si ha che le componenti $e_{w_R,i}$ sono ancora Gaussiane (uniformi) e tutte con la stessa varianza (unitaria); è quindi possibile utilizzare la procedura descritta nel sotto paragrafo precedente 1.6.1 per stimare δ^2 sia nel caso Gaussiano (1.130), sia nel caso uniforme (1.125).

Il *Discrepancy Principle* per il problema (1.148) diventa:

$$\|\mathbf{A}_{w_R} \hat{\mathbf{x}}_{reg} - \mathbf{b}_{w_R}\|_2^2 \simeq \|\mathbf{e}_{w_R}\|_2^2 \leq \delta^2\tag{1.149}$$

¹⁵Nel caso in cui le misure siano correlate, Σ_{meas} non è diagonale, è comunque possibile definire $\Sigma_{\text{meas}}^{-\frac{1}{2}}$ che a sua volta non è diagonale, quindi \mathbf{e}_{w_R} non è la versione riscalata di $\mathbf{e}_{w_{\text{meas}}}$ e le rispettive componenti possono non seguire la stessa distribuzione di probabilità

Infine si dimostra che l'energia del vettore di errore sbiancato \mathbf{e}_W è indipendente dalla specifica matrice di trasformazione utilizzata per lo sbiancamento, (1.137).

Teorema 1.8. *L'energia del vettore di errore sbiancato \mathbf{e}_W è indipendente dalla specifica matrice di trasformazione utilizzata per lo sbiancamento:*

$$\begin{aligned} \forall \mathbf{W} \in \mathbb{R}^{m \times m} : \mathbf{W}^\top \mathbf{W} &= \Sigma_{\mathbf{b}}^{-1} \\ \|\mathbf{e}_W\|_2^2 &:= \|\mathbf{W}\mathbf{e}\|_2^2 \\ &= \|\mathbf{R}_{\text{Ch}_1}\mathbf{e}\|_2^2 \end{aligned} \quad (1.150)$$

dove $\mathbf{R}_{\text{Ch}_1} : \mathbf{R}_{\text{Ch}_1}^\top \mathbf{R}_{\text{Ch}_1} = \Sigma_{\mathbf{b}}^{-1}$ è la matrice triangolare superiore ottenuta dalla fattorizzazione di Cholesky dell'inversa della matrice di covarianza, equazione (1.140).

Dimostrazione. Qualunque matrice può essere decomposta come prodotto di una matrice ortogonale e di una triangolare superiore:

$$\mathbf{W} = \mathbf{Q}_W \mathbf{R}_W \quad (1.151)$$

dove $\mathbf{Q}_W : \mathbf{Q}_W \mathbf{Q}_W^\top = \mathbf{Q}_W^\top \mathbf{Q}_W = \mathbf{I}_m$ è una matrice ortogonale e $\mathbf{R}_W \in \mathbb{R}^{m \times m}$ è una matrice triangolare superiore. Se la matrice \mathbf{W} è di sbiancamento, soddisfa la (1.137) quindi:

$$\begin{aligned} \Sigma_{\mathbf{b}}^{-1} &= \mathbf{W}^\top \mathbf{W} = \\ &= \mathbf{R}_W^\top \mathbf{Q}_W^\top \mathbf{Q}_W \mathbf{R}_W \\ &= \mathbf{R}_W^\top \mathbf{R}_W \\ &= \mathbf{R}_{\text{Ch}_1}^\top \mathbf{R}_{\text{Ch}_1} \\ \implies \mathbf{R}_W &= \mathbf{R}_{\text{Ch}_1} \end{aligned} \quad (1.152)$$

L'ultimo passaggio è giustificato dall'unicità della decomposizione di Cholesky (Teorema 1.7). Per concludere, sfruttando il fatto che le trasformazioni ortogonali non alterano la norma Euclidea di un vettore o di una matrice, proprietà (1.22), si ha:

$$\begin{aligned} \forall \mathbf{W} \in \mathbb{R}^{m \times m} : \mathbf{W}^\top \mathbf{W} &= \Sigma_{\mathbf{b}}^{-1} \\ \|\mathbf{W}\mathbf{e}\|_2^2 &= \|\mathbf{Q}_W \mathbf{R}_{\text{Ch}_1} \mathbf{e}\|_2^2 = \|\mathbf{Q}_W\|_2^2 \|\mathbf{R}_{\text{Ch}_1} \mathbf{e}\|_2^2 \\ \implies \|\mathbf{e}_W\|_2^2 &= \|\mathbf{R}_{\text{Ch}_1} \mathbf{e}\|_2^2 \end{aligned} \quad (1.153)$$

□

Il risultato precedente generalizza quanto riportato in [32, Equazione 23], lavoro presentato in dettaglio nel capitolo successivo.

Appendice

1.A \mathcal{P}_{e_1/r_u} , $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_1$ riscritto come LP

La minimizzazione dei valori assoluti LAV (1.29), corrispondente al problema \mathcal{P}_{e_1/r_u} , può essere riscritta per mezzo di un vettore di m variabili decisionali ausiliarie $\mathbf{w} = [w_1, \dots, w_i, \dots, w_m]^\top \in \mathbb{R}^m$ in termini di problema lineare

$$\begin{aligned} \min \quad & \sum_{i=1}^m w_i \\ \text{s.t.} \quad & -\mathbf{w} \leq \mathbf{Ax} - \mathbf{b} \leq \mathbf{w} \\ & \mathbf{w} \geq \mathbf{0}_m \in \mathbb{R}^m \end{aligned} \tag{1.A.154}$$

e quindi riportata alla forma canonica (1.33) con le sostituzioni:

$$\begin{aligned} m_{LP} &= 2m; \quad n_{LP} = n + m \\ \mathbf{c} &= \begin{bmatrix} \mathbf{0}_n \\ \mathbf{1}_m \end{bmatrix} \in \mathbb{R}^{n_{LP}}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix} \in \mathbb{R}^{m_{LP}} \\ \mathbf{H} &= \begin{bmatrix} \mathbf{A} & -\mathbf{I}_m \\ -\mathbf{A} & -\mathbf{I}_m \end{bmatrix} \in \mathbb{R}^{m_{LP} \times n_{LP}}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{w} \end{bmatrix} \in \mathbb{R}^{n_{LP}} \\ \hat{\mathbf{x}}_1 &= [\hat{z}_1, \dots, \hat{z}_n]^\top \end{aligned} \tag{1.A.155}$$

Dove i vettori $\mathbf{0}_n, \mathbf{1}_n$ sono composti da n elementi nulli il primo ed uguali a 1 il secondo, e dove \mathbf{I}_m è la matrice identità di ordine m . $\hat{\mathbf{x}}_1$ è la soluzione di (1.29)

1.B $\mathcal{P}_{e_\infty/r_u}$, $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_\infty$ riscritto come LP

La minimizzazione in norma infinito (1.28), corrispondente al problema $\mathcal{P}_{e_\infty/r_u}$, può essere riscritta in termini di problema lineare per mezzo di uno scalare

decisionale ausiliare $w \in \mathbb{R}^+$

$$\begin{aligned} \min \quad & w \\ \text{s.t.} \quad & -w \leq \mathbf{Ax} - \mathbf{b} \leq w \\ & w \geq 0 \end{aligned} \tag{1.B.156}$$

e quindi riportata alla forma canonica (1.33) con le sostituzioni:

$$\begin{aligned} m_{LP} &= 2m; \quad n_{LP} = n + 1 \\ \mathbf{c} &= \begin{bmatrix} \mathbf{0}_n \\ 1 \end{bmatrix} \in \mathbb{R}^{n_{LP}}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix} \in \mathbb{R}^{m_{LP}} \\ \mathbf{H} &= \begin{bmatrix} \mathbf{A} & -\mathbf{1}_m \\ -\mathbf{A} & -\mathbf{1}_m \end{bmatrix} \in \mathbb{R}^{m_{LP} \times n_{LP}}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ w \end{bmatrix} \in \mathbb{R}^{n_{LP}} \\ \hat{\mathbf{x}}_\infty &= [\hat{z}_1, \dots, \hat{z}_n]^\top \end{aligned} \tag{1.B.157}$$

$\hat{\mathbf{x}}_\infty$ è la soluzione di (1.28)

1.C Esempi di matrici \mathbf{L}

Le due matrici \mathbf{L}_1 e \mathbf{L}_2 sono discretizzazioni di operatori differenziali:

$$\mathbf{L}_1 = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1) \times n}; \tag{1.C.158}$$

$$\mathbf{L}_2 = \begin{bmatrix} -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-2) \times n} \tag{1.C.159}$$

\mathbf{L}_1 è una discretizzazione dell'operatore derivata prima, il suo "Nucleo" $\mathcal{N}(\mathbf{L}_1)$ contiene i vettori "costanti" che di conseguenza non vengono abbattuti dalla scelta $\mathbf{L} = \mathbf{L}_1$.

$$\mathbf{L}_1 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{1.C.160}$$

Mentre \mathbf{L}_2 corrisponde alla discretizzazione, tramite il metodo alle differenze finite a 3 punti, dell'operatore Laplaciano mono-dimensionale. Al suo nucleo $\mathcal{N}(\mathbf{L}_2)$ appartengono sia i vettori "costanti" sia i vettori "lineari", ed è quindi

utilizzato per selezionare soluzioni “lisce” (*smoothed*).

$$\mathbf{L}_2 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \quad \mathbf{L}_2 \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (1.C.161)$$

1.D $\mathcal{P}_{r_0/e=0}$, (P0) $\min_{\mathbf{x}} \|\mathbf{x}\|_0$ s.t. $\mathbf{Ax} = \mathbf{b}$ riscritto come MILP

Per trasformare il problema (P0), (1.65) in un problema lineare a variabili decisionali miste, reali e binarie, oltre alle variabili decisionali di interesse $\mathbf{x} \in \mathbb{R}^n$, si sfruttano variabili decisionali binarie di appoggio, $\mathbf{w} \in \{0, 1\}^n$ e si aggiungono i vincoli:

$$\begin{cases} x_j & \leq \mathit{bigM} \cdot w_j, \quad j = 1, \dots, n \\ -x_j & \leq \mathit{bigM} \cdot w_j, \quad j = 1, \dots, n \end{cases} \quad (1.D.162)$$

dove bigM è legato alla norma infinito del vettore da ricostruire $\hat{\mathbf{x}}$ dalla relazione $\|\hat{\mathbf{x}}\|_\infty < \mathit{bigM}$, [33].

$$\mathbf{f.o.} \min \sum_{j=1}^n w_j \quad (1.D.163)$$

Grazie a questi due nuovi vincoli, e alla funzione obiettivo, si hanno le seguenti implicazioni logiche:

$$\begin{cases} (x_j = 0) \wedge \mathbf{f.o.} & \implies w_j = 0 \\ (x_j \neq 0) & \implies w_j = 1 \end{cases} \quad (1.D.164)$$

Si noti come la prima implicazione sia ottenuta grazie alla $\mathbf{f.o.}$ che impone la scelta di $w_j = 0$ quando $x_j = 0$. Il Problema quindi diventa:

$$\begin{aligned} & \min_{\mathbf{w}} \mathbf{c}^\top \mathbf{w} \\ & \text{s.t. } \mathbf{Hx} = \mathbf{b} \\ & \quad -\mathit{bigM} \cdot \mathbf{w} \leq \mathbf{x} \leq \mathit{bigM} \cdot \mathbf{w} \\ & \quad \mathbf{x} \in \mathbb{R}^n \\ & \quad \mathbf{w} \in \{0, 1\}^n \end{aligned} \quad (1.D.165)$$

con $\mathbf{c} = \mathbf{1}_n \in \mathbb{R}^n$

1.E $\mathcal{P}_{r_1/e=0}$, (P1) $\min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\mathbf{Ax} = \mathbf{b}$, Basis Pursuit, riscritto come LP

Il problema (P1), (1.67), può essere ricondotto ad un problema LP nella **Forma Standard** (equivalente alla **Forma Canonica** (1.33)):

$$\begin{aligned}
 & \min_{\mathbf{z}} \mathbf{c}^\top \mathbf{z} \\
 & \text{s.t. } \mathbf{Hz} \leq \mathbf{y} \\
 & \quad \mathbf{z} = \mathbf{0}_{n_{LP}} \\
 & \text{con:} \\
 & \quad \mathbf{H} \in \mathbb{R}^{m_{LP} \times n_{LP}} \\
 & \quad \mathbf{z}, \mathbf{c}, \mathbf{0}_{n_{LP}} \in \mathbb{R}^{n_{LP}} \\
 & \quad \mathbf{y} \in \mathbb{R}^{m_{LP}}
 \end{aligned} \tag{1.E.166}$$

dove con $\mathbf{0}_{n_{LP}}$ si intende il vettore con n_{LP} componenti nulle.

Prima utilizzando due vettori di variabili decisionali $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ Ponendo:

$$\begin{aligned}
 & m_{LP} = m; \quad n_{LP} = 2n \\
 & \mathbf{c} = \begin{bmatrix} \mathbb{1}_n \\ \mathbb{1}_n \end{bmatrix} \in \mathbb{R}^{n_{LP}}, \quad \mathbf{y} = \mathbf{b} \in \mathbb{R}_{LP}^m \\
 & \mathbf{H} = \begin{bmatrix} \mathbf{A} & -\mathbf{A} \end{bmatrix} \in \mathbb{R}^{m_{LP} \times n_{LP}}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \in \mathbb{R}^{n_{LP}} \\
 & \hat{\mathbf{x}}_{r_1} = [\hat{z}_1 - \hat{z}_{n+1}, \dots, \hat{z}_n - \hat{z}_{2n}]^\top
 \end{aligned} \tag{1.E.167}$$

Dove il vettore $\mathbb{1}_n$ è composto da n elementi uguali a 1. $\hat{\mathbf{x}}_{r_1}$ è la soluzione di (1.67)

1.F *Discrepancy Principle* per Tikhonov generalizzato

Si vuole determinare μ_g in (1.106) in modo che

$$\mu_g : \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{TGen}}\|_2 \leq \delta \tag{1.F.168}$$

Utilizzando la fattorizzazione GSVD (1.83), e sviluppando i calcoli si ha:

$$\begin{aligned}
 \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{TGen}} &= \mathbf{b} - \mathbf{U}\Sigma\mathbf{X}^\top\mathbf{X}^{-\top}(\Sigma^\top\Sigma + \mu_g\mathbf{\Lambda}^\top\mathbf{\Lambda})^{-1}\Sigma^\top\mathbf{U}^\top\mathbf{b} \\
 &= \mathbf{b} - \mathbf{U}\Sigma(\Sigma^\top\Sigma + \mu_g\mathbf{\Lambda}^\top\mathbf{\Lambda})^{-1}\Sigma^\top\mathbf{U}^\top\mathbf{b} \\
 &= \mathbf{U} [\mathbf{U}^\top\mathbf{b} - \Sigma(\Sigma^\top\Sigma + \mu_g\mathbf{\Lambda}^\top\mathbf{\Lambda})^{-1}\Sigma^\top\mathbf{U}^\top\mathbf{b}] = \\
 &= \mathbf{U} [\mathbf{I}_m - \Sigma(\Sigma^\top\Sigma + \mu_g\mathbf{\Lambda}^\top\mathbf{\Lambda})^{-1}\Sigma^\top] \mathbf{U}^\top\mathbf{b}
 \end{aligned} \tag{1.F.169}$$

quindi

$$\begin{aligned}
 \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{TGen}}\|_2^2 &= \|\mathbf{U} [\mathbf{I}_n - \boldsymbol{\Sigma}(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} + \mu_g \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})^{-1} \boldsymbol{\Sigma}^\top] \mathbf{U}^\top \mathbf{b}\|_2^2 \\
 &= \|\mathbf{U} [\mathbf{I}_n - \boldsymbol{\Sigma}(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} + \mu_g \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})^{-1} \boldsymbol{\Sigma}^\top] \mathbf{U}^\top \mathbf{b}\|_2^2 \\
 &= \sum_{j=1}^n \left[\left(1 - \frac{\sigma_j^2}{\sigma_j^2 + \mu_g \lambda_j^2} \right) \mathbf{u}_j^\top \mathbf{b} \right]^2 \\
 &= \sum_{j=1}^n \left[\left(\frac{\lambda_j^2}{\beta \sigma_j^2 + \lambda_j^2} \right) \mathbf{u}_j^\top \mathbf{b} \right]^2 \\
 &\text{dove } \beta = \frac{1}{\mu_g}
 \end{aligned} \tag{1.F.170}$$

Definiamo la funzione $\Psi_{\text{TGen}}(\beta)$:

$$\Psi_{\text{TGen}}(\beta) := \sum_{j=1}^n \left(\frac{\lambda_j^2 \mathbf{u}_j^\top \mathbf{b}}{\beta \sigma_j^2 + \lambda_j^2} \right)^2 - \delta^2 \tag{1.F.171}$$

in [30] si mostra come Ψ_{TGen} sia convessa e monotona decrescente, inoltre, per ragionevoli valori di δ (norma dell'errore di misura), si ha¹⁶:

$$\begin{aligned}
 \Psi_{\text{TGen}}(0) &= \sum_{j=1}^n (\mathbf{u}_j^\top \mathbf{b})^2 - \delta^2 > 0 \\
 \lim_{\beta \rightarrow \infty} \Psi_{\text{TGen}}(\beta) &= -\delta^2 < 0 \\
 \implies \exists \beta^* : \Psi_{\text{TGen}}(\beta^*) &= 0
 \end{aligned} \tag{1.F.172}$$

che possiamo trovare col metodo di Newton usando la derivata prima data da:

$$\Psi'_{\text{TGen}}(\beta) = -2 \sum_{j=1}^n \frac{\sigma_j^2 \lambda_j^2 (\mathbf{u}_j^\top \mathbf{b})^2}{(\beta \sigma_j^2 + \lambda_j^2)^3} \tag{1.F.173}$$

e poi scegliere $\mu_g^* = \frac{1}{\beta^*}$

¹⁶Se \mathbf{A} è quadrata e a rango pieno si ha $\Psi_{\text{TGen}}(0) = \|\mathbf{b}\|_2^2 - \delta^2$

Capitolo 2

Identificazione delle principali sorgenti armoniche nelle reti di distribuzione

2.1 Introduzione

Negli ultimi decenni si sta assistendo all'evoluzione delle reti di distribuzione elettrica verso reti di nuova generazione, conosciute come *Smart Grid* (SG). Tale evoluzione sta introducendo numerosi cambiamenti nei sistemi elettrici. In particolare, la crescente presenza di carichi non lineari e generatori distribuiti, con la relativa elettronica di potenza, comporta il costante incremento dell'inquinamento armonico nelle reti. La conseguenza è l'insorgenza di diversi problemi, come la riduzione della vita utile di dispositivi sensibili, l'aumento delle perdite e dei costi di gestione e così via [34].

In letteratura sono state proposte diverse soluzioni per monitorare lo stato delle reti di potenza in termini di componenti armoniche mediante tecniche di *Harmonic State Estimation*, si vedano ad esempio [35–38], ma la loro implementazione nelle reti di distribuzione non è sempre possibile. Tali tecniche si basano sul monitoraggio diretto del sistema ma, date le grandi dimensioni e l'elevato numero di nodi, questo non è ipotizzabile nei sistemi di distribuzione anche a causa degli alti costi di installazione e gestione dei dispositivi di misura della qualità dell'energia, i cosiddetti *Power Quality* (PQ) *meter*. Le potenziali sorgenti di inquinamento armonico dovrebbero quindi essere valutate a partire dall'utilizzo di un numero molto elevato di pseudo-misure altamente incerte ad integrazione delle poche misure disponibili. Nel rilevamento delle sorgenti di inquinamento, così come in qualsiasi altra applicazione in ambito PQ, gioca infatti

un ruolo chiave l'accuratezza delle misure. In questo contesto è importante evidenziare come il comportamento dei trasformatori di misura, di corrente (TA) e di tensione (TV), è tipicamente definito solo alla frequenza nominale, mentre le caratteristiche metrologiche di questi trasformatori tendono a degradare in presenza di segnali non sinusoidali, si vedano ad esempio, per quanto riguarda i TV, [39] e [40]. Recenti lavori in letteratura [41, 42] hanno infatti proposto metodologie per compensare l'impatto dei trasformatori di tensione di misura nelle misure armoniche.

Per la stima delle armoniche nei sistemi di potenza sono state proposte soluzioni basate sull'apprendimento automatico, comunemente chiamato *Machine Learning* (ML) [43, 44], sulla trasformata wavelet [45] o su altre tecniche [46].

L'approccio descritto in questo capitolo, che presenta i risultati dell'attività di ricerca che ha portato all'articolo [32], è del tutto differente. L'obiettivo principale non è quello di ricostruire lo stato armonico della rete di distribuzione, bensì quello di identificare le principali sorgenti di inquinamento armonico. Tale identificazione può consentire un'azione mirata sulle cause del problema da parte degli operatori del sistema. Questo approccio, chiamato stima delle sorgenti armoniche, *Harmonic Source Estimation* (HSoE), è stato presentato inizialmente in [47, 48], in modo specifico per le reti di distribuzione. Il problema viene modellato come la ricerca di generatori armonici di corrente in corrispondenza dei principali carichi inquinanti. Per i motivi prima elencati, da un lato il problema si presenta come sotto-determinato, a causa delle dimensioni delle reti di distribuzione e degli alti costi di un sistema capillare di PQ meter, dall'altro è però possibile assumere come sparso il vettore di stato da ricostruire. Infatti, è ragionevole pensare che, in un dato momento, il contributo armonico nella rete sia, in modo prevalente, dovuto solo a pochi nodi. Partendo da queste considerazioni, in [49] e [50] è stato proposto un approccio al HSoE basato sul *Compressive Sensing* (CS). Questa recente tecnica matematica, introdotta nel paragrafo 1.4, è stata applicata con risultati molto promettenti in diversi ambiti scientifici tra i quali, la ricostruzione di immagini per applicazioni mediche [51, 52], il *pattern recognition* [53], i sistemi di comunicazione [54], sino al posizionamento ottimale di sensori per localizzare le perdite nelle reti di distribuzione dell'acqua [55].

Nel contesto della PQ, gli algoritmi in [50] possono essere applicati sia ad analisi compressive (considerando più ordini armonici contemporaneamente) sia ad analisi specifiche (incentrate su un singolo ordine armonico). Queste soluzioni sono pensate per essere implementate in uno scenario SG, e quindi si basano sull'uso di moderni dispositivi di misura, in grado di fornire misure di fasori armonici sincronizzati, come quelli proposti in [56] e [57].

Poiché nell'ambito del CS un problema può essere affrontato con diverse tecniche, in [58] sono state valutate le prestazioni di due algoritmi basati sul CS

e in particolare son stati utilizzati l'algoritmo di tipo *greedy*, chiamato *Block Orthogonal Matching Pursuit* (BOMP) e la minimizzazione in ℓ_1 con vincolo di uguaglianza (noto anche come *Basis Pursuit* (BP) o algoritmo (P1)), applicati all'analisi HSoE, in diverse condizioni di incertezza di misura. Il BOMP è l'estensione a blocchi del *Orthogonal Matching Pursuit* (OMP) [59] e viene spesso scelto per la facilità di implementazione e il basso carico computazionale, mentre gli algoritmi di minimizzazione in ℓ_1 , a cui il BP o (P1) appartiene, benché più esigenti dal punto di vista computazionale, forniscono risultati più accurati. In [58] è stato evidenziato come l'algoritmo BOMP sia particolarmente sensibile agli errori di misura, soprattutto con riferimento agli errori nelle misure di fase, mentre l'algoritmo (P1), si è dimostrato più stabile al variare della incertezza di misura, suggerendo la minimizzazione in ℓ_1 come la più promettente per le applicazioni HSoE.

Per aumentare la robustezza della minimizzazione in ℓ_1 rispetto all'errore di misura, in [60] è stato rilassando il vincolo di uguaglianza del (P1), che corrisponde ad una formulazione $\mathcal{P}_{r_1/e=0}$, introducendo un vincolo quadratico sull'errore di approssimazione legato alla valutazione dell'energia del rumore. In [32] si è proposta una nuova formulazione del problema HSoE mediante un nuovo algoritmo di minimizzazione in ℓ_1 chiamato (P2), che, come visto nei Paragrafi 1.3 e 1.4, è la formulazione a norma minima con vincolo sull'errore di approssimazione (\mathcal{P}_{r_1/e_2}) della metodologia LASSO.

Il metodo proposto si fonda su appropriate valutazioni sui limiti degli errori di misura delle armoniche sincronizzate, integra nell'algoritmo il modello di incertezza e propone una trasformazione di sbiancamento, *Whitening Transformation*, in grado di recuperare l'informazione sulle distribuzioni degli errori di misura che viene alterata a seguito delle trasformazioni delle misure operate nelle equazioni di misura. La *Whitening Transformation* permette una valutazione accurata del limite dell'energia dell'errore di misura che è indispensabile per la determinazione del vincolo quadratico dell'algoritmo (P2) e, conseguentemente, per le sue prestazioni.

L'algoritmo (P2) supera le prestazioni, in termini di qualità dell'identificazione delle sorgenti armoniche, dell'algoritmo per HSoE presentato in [58] e basato sul (P1). Le sue prestazioni sono state verificate su reti di test di distribuzione simulate, ipotizzando la presenza di adeguati sistemi di misura e considerando diversi scenari di accuratezza.

Il capitolo è organizzato come segue:

- Il Paragrafo 2.2.1 descrive la "piattaforma" o "framework" di stima, descrive l'approccio HSoE e presenta gli algoritmi del CS presi in considerazione.

- Il Paragrafo 2.3 analizza nel dettaglio la valutazione dell'energia dell'errore di misura, affrontando gli aspetti teorici per poi validarli con opportune sperimentazioni.
- Il Paragrafo 2.4 è dedicato alla presentazione dei test condotti su due reti di distribuzione e alla discussione dei risultati ottenuti.
- Infine le Conclusioni si trovano nel Paragrafo 2.5.

2.2 Formulazione dell'algoritmo di stima

2.2.1 Modello di stima delle sorgenti armoniche

L'identificazione delle sorgenti armoniche in una SG, nell'ipotesi di disporre delle misure fasoriali armoniche, può essere effettuata stimando il contributo armonico iniettato in rete da ciascuna sorgente, carico o generatore, per tutti gli ordini armonici di interesse. Infatti, dato il generico ordine armonico h , è possibile formulare il problema con il seguente modello lineare:

$$\bar{\mathbf{y}}_h = \bar{\mathbf{A}}_h \bar{\mathbf{u}}_h + \bar{\mathbf{e}}_h \quad (2.1)$$

nel quale il pedice h indica l'ordine armonico, $\bar{\mathbf{y}}_h \in \mathbb{C}^M$ rappresenta il vettore delle M misure dei fasori armonici, di tensione e/o corrente, con riferimento all'ordine armonico h , i cui corrispondenti errori di misura sono rappresentati con il vettore $\bar{\mathbf{e}}_h \in \mathbb{C}^M$. La matrice di misura $\bar{\mathbf{A}}_h \in \mathbb{C}^{M \times N}$ mette in relazione le M misure con il vettore $\bar{\mathbf{u}}_h \in \mathbb{C}^N$ delle N correnti armoniche incognite che potrebbero essere iniettate da ciascuna potenziale sorgente. La matrice di misura è ottenuta considerando sia le linee sia i carichi, quindi i suoi elementi sono differenti per i differenti ordini armonici, perchè il loro comportamento è dipendente dalla frequenza. Il vettore $\bar{\mathbf{u}}_h$, detto vettore delle forzanti, viene popolato con i contributi armonici di ciascuna sorgente, i cui valori possono variare notevolmente a causa dei diversi livelli di inquinamento forniti da ciascuna sorgente. Di conseguenza, in base al livello di inquinamento iniettato, è possibile classificare le fonti di inquinamento in fonti scarsamente/non inquinanti e fonti inquinanti in modo significativo/prevalente. Il primo insieme è tipicamente quello più popolato, quindi, trascurando i contributi delle numerose sorgenti scarsamente inquinanti, è possibile trattare il vettore incognito come vettore sparso.

Il modello di misura in (2.1) è a valori complessi, sia per quanto riguarda la matrice di misura, sia per quanto riguarda i vettori: incognito, di misura e d'errore. Al fine della corretta valutazione dell'impatto dell'incertezza di misura, sviluppata in dettaglio nel Paragrafo 2.3, è opportuno trasformarlo in un problema nel campo reale, separando le componenti reali da quelle immaginarie.

In particolare, indicando con gli apici r e x la voce reale e quella immaginaria rispettivamente, il sistema in (2.1) diventa:

$$\mathbf{y}_h = \begin{bmatrix} \mathbf{y}_h^r \\ \mathbf{y}_h^x \end{bmatrix} = \begin{bmatrix} \mathbf{A}_h^r & -\mathbf{A}_h^x \\ \mathbf{A}_h^x & \mathbf{A}_h^r \end{bmatrix} \begin{bmatrix} \mathbf{u}_h^r \\ \mathbf{u}_h^x \end{bmatrix} + \begin{bmatrix} \mathbf{e}_h^r \\ \mathbf{e}_h^x \end{bmatrix} = \mathbf{A}_h \mathbf{u}_h + \mathbf{e}_h \quad (2.2)$$

Indicando con $\Re[\cdot]$ e $\Im[\cdot]$ rispettivamente gli operatori vettoriali di parte reale e parte immaginaria, il vettore $\mathbf{y}_h^r \triangleq \Re[\bar{\mathbf{y}}_h] = [y_{h,1}^r, \dots, y_{h,M}^r]^\top$ indica la parte reale del vettore di misura e $\mathbf{y}_h^x \triangleq \Im[\bar{\mathbf{y}}_h] = [y_{h,1}^x, \dots, y_{h,M}^x]^\top$ indica la corrispondente parte immaginaria. Le stesse considerazioni sono valide per tutti i termini, matrici e vettori, dell'equazione (2.2): $\mathbf{A}_h^r = \Re[\bar{\mathbf{A}}_h]$, $\mathbf{A}_h^x = \Im[\bar{\mathbf{A}}_h]$, $\mathbf{e}_h^r = \Re[\bar{\mathbf{e}}_h]$, $\mathbf{e}_h^x = \Im[\bar{\mathbf{e}}_h]$.

Nel seguito tutte le considerazioni sono riferite ad un singolo ordine armonico e il pedice h è omissso per semplicità di trattazione.

2.2.2 Algoritmi per il *Compressive Sensing*

Come discusso nel Capitolo 1, il *Compressive Sensing*, anche detto *Compressed Sensing*, è una tecnica matematica che consente di ricostruire segnali sparsi data una limitata quantità di informazioni in ingresso. Un generico vettore \mathbf{u} è sparso se la sua pseudo norma $\|\mathbf{u}\|_0$ è significativamente inferiore alla sua cardinalità¹, dove $\|\cdot\|_0$ è definita in (1.16) come la cardinalità del supporto di un vettore, quindi $\|\mathbf{u}\|_0 = \text{Card}\{i : u_i \neq 0\}$. In particolare, il numero S di componenti diversi da zero in \mathbf{u} indica il suo livello di sparsità e il vettore è definito *S-sparso*, se $\|\mathbf{u}\|_0 \leq S$ come da definizione (1.63). Inoltre, un vettore non sparso potrebbe ammettere una rappresentazione sparsa, quando rappresentato in una base diversa o quando è ben approssimato da un vettore *S-sparso* (si veda la definizione 1.4.2). Nel caso dell'identificazione delle sorgenti armoniche prevalenti, gli elementi del vettore incognito sono caratterizzati da ordini di grandezza molto diversi. Si ha quindi un vettore detto comprimibile o relativamente sparso e ben approssimabile da un vettore *S-sparso*, nel quale le componenti più piccole sono state trascurate. È quindi possibile sfruttare le tecniche del CS per ricostruire le componenti del vettore relative ai nodi della rete inquinanti in modo prevalente.

La formulazione base del CS² è data dalla formulazione (1.65) che è qui riportata con riferimento ai simboli del paragrafo precedente:

$$(P0) \quad \mathbf{u} : \min_{\mathbf{u}} \|\mathbf{u}\|_0 \text{ s.t. } \mathbf{A}\mathbf{u} = \mathbf{y} \quad (2.3)$$

Il problema è NP-hard con complessità di tipo combinatorio³ [61], quindi non

¹Il numero di elementi non nulli di un insieme, indicata con $\text{Card}\{\cdot\}$

²Maggiori dettagli nel paragrafo 1.4

³Per matrici e vettori reali è possibile riscriverlo come problema di programmazione lineare a variabili decisionali miste (MILP), come discusso nel paragrafo 1.4.2 e mostrato in dettaglio nell'Appendice 1.D.

trattabile per problemi di medie e grandi dimensioni. Usualmente vengono quindi preferite tecniche alternative.

Uno degli algoritmi più comunemente implementati per applicazioni del CS è l'*Orthogonal Matching Pursuit* (OMP) [62]. Questo algoritmo è di tipo *greedy*, ovvero ad ogni passo cerca la migliore soluzione locale e viene spesso scelto per la sua facilità di implementazione e il basso carico computazionale. Si tratta infatti di un algoritmo iterativo che esegue tante iterazioni quanti sono gli elementi del vettore \mathbf{u} diversi da zero che si vogliono recuperare. Durante ogni iterazione, l'OMP seleziona la colonna della matrice di misura più correlata ai residui, la aggiunge ad una matrice temporanea inizialmente vuota, che utilizza sia per stimare un nuovo elemento diverso da zero sia per aggiornare il valore degli elementi stimati nelle iterazioni precedenti.

Un approccio differente e più robusto è quello che consiste nell' approssimare la formulazione (2.3) con un problema di minimizzazione in ℓ_1 . In letteratura (si vedano [63] e [17]), è stato mostrato come, in assenza di errori di misura, qualora la matrice di misura soddisfi la proprietà di isometria ristretta, ovvero la *Restricted Isometry Property* (RIP) della definizione 1.4.4, sia possibile ricostruire un vettore sparso incognito⁴ risolvendo il problema (P1) (1.67), detto anche *Basis Pursuit* e qui di seguito riportato utilizzando la simbologia specifica del problema in esame e introdotta nel paragrafo precedente:

$$(P1) \quad \mathbf{u} : \min_{\mathbf{u}} \|\mathbf{u}\|_1 \text{ s.t. } \mathbf{A}\mathbf{u} = \mathbf{y} \quad (2.4)$$

Quando riferito a matrici e vettori in campo reale, è possibile riscrivere il problema (P1) come problema LP in **Forma Standard**, come mostrato in Appendice 1.E. Nella pratica, qualunque misura è affetta da errori e questi non possono essere trascurati. Quindi, per affrontare il problema, è necessario rilassare il vincolo di uguaglianza della formulazione (P1), che è di tipo $\mathcal{P}_{r_1/\epsilon=0}$, e sostituirlo con un vincolo sull'errore di approssimazione valutato in ℓ_2 (formulazione $\mathcal{P}_{r_1/\epsilon_2}$) pervenendo al problema chiamato (P2) in [11] e [60]⁵. La sua formulazione (1.68) viene qui riportata con la nuova simbologia:

$$(P2) \quad \min_{\mathbf{u}} \|\mathbf{u}\|_1 \text{ s.t. } \|\mathbf{A}\mathbf{u} - \mathbf{y}\|_2 \leq \epsilon \quad (2.5)$$

dove ϵ rappresenta il limite superiore della norma del vettore di errore. In questa formulazione la corretta definizione del parametro di ottimizzazione ϵ gioca un ruolo fondamentale ed è la chiave delle prestazioni dell'algoritmo per superare i limiti del (P1) in presenza di rumore, come mostrato in [60].

La definizione di ϵ non è banale e di seguito viene presentato un approccio teorico per la sua valutazione nello specifico ambito del HSoE, approccio poi

⁴Si veda il Teorema 1.4.

⁵Noto anche come *Basis Pursuit De-Noising* (BPDN) in [16], o LASSO in [9], si veda il Paragrafo 1.3.4.

validato tramite adeguate simulazioni.

2.3 Valutazione dell'energia dell'errore di misura

Nel paragrafo precedente sono state presentate due diverse formulazioni di minimizzazione in ℓ_1 , rispettivamente etichettate come (P1) e (P2). Queste formulazioni sono qui applicate al problema HSoE. In particolare, mentre (P1) affronta la minimizzazione in ℓ_1 di un sistema lineare sotto-determinato esatto ($\mathcal{P}_{r_1/\epsilon=0}$), la formulazione (P2) prende in considerazione un più realistico sistema lineare approssimato:

$$\mathbf{A}\mathbf{u} \simeq \mathbf{y} \quad (2.6)$$

nel quale l'errore di approssimazione, o *misfit*, tra il vettore delle misure armoniche \mathbf{y} ed il modello $\mathbf{A}\mathbf{u}$, è limitato in ℓ_2 ($\mathcal{P}_{r_1/\epsilon_2}$). Nello specifico, il vettore dell'errore di misura in (2.2) e il parametro che definisce il vincolo quadratico in (2.5) sono strettamente legati. Il "Principio di Discrepanza", *Discrepancy Principle* (DP), di Morozov [20] suggerisce di scegliere il parametro in modo che sia circa uguale alla norma dell'errore di misura, come da (1.69), che nella simbologia di questo capitolo diventa:

$$\xi_{\mathbf{e}} = \|\mathbf{e}\|_2^2 \leq \epsilon^2 \quad (2.7)$$

dove $\xi_{\mathbf{e}} = \|\mathbf{e}\|_2^2$ è l'energia del vettore dell'errore di misura del fasore armonico in componenti rettangolari, di seguito denominata «energia dell'errore». Inoltre, come suggerito in [60] e come discusso nella Sezione 1.6.1: sfruttando la *concentration inequality*, se si conoscono la media $\mu_{\xi_{\mathbf{e}}}$ e la deviazione standard $\sigma_{\xi_{\mathbf{e}}}$ della distribuzione dell'energia dell'errore, il limite corrispondente può essere scelto come $\epsilon^2 = \mu_{\xi_{\mathbf{e}}} + \lambda\sigma_{\xi_{\mathbf{e}}}$ dove λ permette di variare il livello di confidenza associato alla disuguaglianza $\|\mathbf{e}\|_2^2 \leq \mu_{\xi_{\mathbf{e}}} + \lambda\sigma_{\xi_{\mathbf{e}}}$ (in genere è necessario $\lambda \geq 2$ per avere un'alta probabilità di validità della disuguaglianza).

È necessario sottolineare che in [60] il calcolo della media e della deviazione standard del rumore è semplice, poiché si assume che tutti gli elementi nel vettore di errore seguano la stessa (e nota) distribuzione di probabilità, Gaussiana o uniforme, con lo stesso valore medio e deviazione standard. La stessa ipotesi non vale quando si valutano le armoniche nei sistemi elettrici di potenza, per i seguenti motivi:

1. Le misure armoniche sono caratterizzate da incertezze differenti a causa delle differenze nelle grandezze misurate (tensioni e correnti) e nei parametri misurati (ampiezza e angolo di fase);
2. A causa delle trasformazioni eseguite per ottenere (2.2), le distribuzioni delle componenti del vettore di misura differiscono dalle distribuzioni delle

componenti d'errore direttamente attribuibili agli strumenti di misura, e sono, in generale, sconosciute. Questo vale anche per le componenti del corrispondente vettore di errore di misura.

Il primo problema è ben noto ed è affrontato mediante la trasformazione di sbiancamento, *Whitening Transformation*, discussa qui di seguito.

2.3.1 *Whitening Transformation*

Per tenere adeguatamente conto dell'accuratezza di ogni misurazione, è necessario eseguire la cosiddetta *Whitening Transformation*. Questa procedura permette di pesare ciascuna misura secondo la propria incertezza, moltiplicando a sinistra l'equazione del modello di misura (2.2) con una matrice quadrata di sbiancamento \mathbf{W} (anche in questo caso, per semplicità, si omette il pedice h). Infatti gli elementi di \mathbf{W} sono definiti in modo tale che la trasformazione permetta di assegnare pesi minori alle misure con incertezze maggiori, e viceversa.

Eseguire una trasformazione di sbiancamento significa applicare una trasformazione lineare al vettore casuale \mathbf{y} in modo da ottenere un nuovo vettore casuale \mathbf{y}_w avente come matrice di covarianza la matrice identità, ovvero $\Sigma_{\mathbf{y}_w} = \mathbf{I}$. La matrice di covarianza di un vettore colonna casuale, a valore atteso nullo, è definita come:

$$\Sigma_{\mathbf{y}} = E[\mathbf{y}\mathbf{y}^T] \quad (2.8)$$

Quindi è possibile osservare che⁶:

$$\begin{aligned} \forall \mathbf{W} : \mathbf{W}^T \mathbf{W} &= \Sigma_{\mathbf{y}}^{-1} \\ \mathbf{y}_w \triangleq \mathbf{W}\mathbf{y} &\implies \Sigma_{\mathbf{y}_w} = \mathbf{I} \end{aligned} \quad (2.9)$$

Data la generica i -esima misura fasoriale, di ampiezza ρ_i e angolo di fase ϕ_i , le sue corrispondenti componenti reale e immaginaria, rispettivamente y_i^r e y_i^x , possono essere ottenute con la trasformazione da coordinate polari a coordinate rettangolari:

$$\begin{aligned} y_i^r &= \rho_i \cos(\phi_i) \\ y_i^x &= \rho_i \sin(\phi_i) \end{aligned} \quad (2.10)$$

la cui matrice Jacobiana, nell'intorno del punto di misura è \mathbf{J}_i :

$$\mathbf{J}_i = \begin{bmatrix} \cos \phi_i & -\rho_i \sin \phi_i \\ \sin \phi_i & \rho_i \cos \phi_i \end{bmatrix} \quad (2.11)$$

Considerando le deviazioni standard, sia di ampiezza che di fase, rispettivamente indicate con σ_{ρ_i} e σ_{ϕ_i} , è possibile ottenere la corrispondente matrice di covarianza 2×2 :

⁶Per maggiori dettagli e una dimostrazione, si rimanda al Teorema 1.6.

$$\Sigma_{\rho_i, \phi_i} = \begin{bmatrix} \sigma_{\rho_i}^2 & 0 \\ 0 & \sigma_{\phi_i}^2 \end{bmatrix} \quad (2.12)$$

e, attraverso l'applicazione della legge di propagazione dell'incertezza [31], la matrice di covarianza riferita alle corrispondenti componenti reale e immaginaria:

$$\begin{aligned} \Sigma_{y_i^r, y_i^x} &= \mathbf{J}_i \begin{bmatrix} \sigma_{\rho_i}^2 & 0 \\ 0 & \sigma_{\phi_i}^2 \end{bmatrix} \mathbf{J}_i^\top = \\ &= \begin{bmatrix} \rho_i^2 \sigma_{\phi_i}^2 \sin^2 \phi_i + \sigma_{\rho_i}^2 \cos^2 \phi_i & (\sigma_{\rho_i}^2 - \rho_i^2 \sigma_{\phi_i}^2) \sin \phi_i \cos \phi_i \\ (\sigma_{\rho_i}^2 - \rho_i^2 \sigma_{\phi_i}^2) \sin \phi_i \cos \phi_i & \rho_i^2 \sigma_{\phi_i}^2 \cos^2 \phi_i + \sigma_{\rho_i}^2 \sin^2 \phi_i \end{bmatrix} \end{aligned} \quad (2.13)$$

Per estensione, assumendo che gli errori di misura di ampiezza e angolo di fase, per tutti i fasori misurati, siano non correlati, è possibile ottenere la matrice di covarianza del vettore di misura complessivo $\Sigma_{\mathbf{y}}$ come:

$$\Sigma_{\mathbf{y}} = \mathbf{J} \Sigma_{\rho, \phi} \mathbf{J}^\top \quad (2.14)$$

dove $\Sigma_{\mathbf{y}}$ è la matrice di covarianza di tutte le parti reali e immaginarie delle misure fasoriali prese in considerazione, e \mathbf{J} è la matrice Jacobiana della trasformazione complessiva. Vale la pena sottolineare che, se tutti gli M fasori misurati sono considerati non correlati, la matrice di covarianza complessiva in ampiezza e fase, $\Sigma_{\rho, \phi}$, è diagonale. Pertanto, definendo l'inversa della sua radice quadrata:

$$\Sigma_{\rho, \phi}^{-\frac{1}{2}} = \begin{bmatrix} \sigma_{\rho_1}^{-1} & & & & & 0 \\ & \ddots & & & & \\ & & \sigma_{\rho_M}^{-1} & & & \\ & & & \sigma_{\phi_1}^{-1} & & \\ & & & & \ddots & \\ 0 & & & & & \sigma_{\phi_M}^{-1} \end{bmatrix} \quad (2.15)$$

è possibile mostrare come le matrici di sbiancamento che soddisfano la (2.9) sono infinite. Infatti, data una qualsiasi matrice ortogonale $\mathbf{Q} \in \mathbb{R}^{2M \times 2M}$, è possibile osservare che:

$$\begin{aligned} \Sigma_{\mathbf{y}}^{-1} &= \mathbf{J}^{-\top} \Sigma_{\rho, \phi}^{-\frac{1}{2}} \Sigma_{\rho, \phi}^{-\frac{1}{2}} \mathbf{J}^{-1} = \\ &= (\mathbf{J}^{-\top} \Sigma_{\rho, \phi}^{-\frac{1}{2}} \mathbf{Q}^\top) (\mathbf{Q} \Sigma_{\rho, \phi}^{-\frac{1}{2}} \mathbf{J}^{-1}) \\ &\quad \forall \mathbf{Q} : \mathbf{Q} \mathbf{Q}^\top = \mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_{2M} \end{aligned} \quad (2.16)$$

Le infinite matrici di sbiancamento possono quindi essere definite come:

$$\mathbf{W}_{\mathbf{Q}} \triangleq \mathbf{Q} \Sigma_{\rho, \phi}^{-\frac{1}{2}} \mathbf{J}^{-1} \quad (2.17)$$

In [50] e [58], la matrice che consente lo sbiancamento delle misure è stata ottenuta mediante la fattorizzazione di Cholesky della matrice di covarianza complessiva (2.14), $\mathbf{\Sigma}_y = \mathbf{U}_C^T \mathbf{U}_C$ e assegnando $\mathbf{W}_C = \mathbf{U}_C^{-T}$, che soddisfa la (2.9). Quindi, indicando con il pedice w_C il risultato della trasformazione di sbiancamento attraverso la fattorizzazione di Cholesky, il sistema di misura riferito alle misure sbiancate diventa:

$$\begin{aligned} \mathbf{y}_{w_C} &\triangleq \mathbf{W}_C \mathbf{y} = \mathbf{W}_C \mathbf{A} \mathbf{u} + \mathbf{W}_C \mathbf{e} = \\ &= \mathbf{A}_{w_C} \mathbf{u}_{w_C} + \mathbf{e}_{w_C} \end{aligned} \quad (2.18)$$

dove $\mathbf{e}_{w_C} = \mathbf{W}_C \mathbf{e}$ è il vettore dell'errore di misura sbiancato.

Tuttavia, questa particolare matrice di sbiancamento non consente di risolvere anche il problema della mancanza di conoscenza sulle distribuzioni degli elementi del vettore \mathbf{y}_{w_C} . Infatti, anche assumendo distribuzioni uniformi per gli errori di misura, sia per le ampiezze sia per gli angoli di fase di $\bar{\mathbf{y}}$, gli elementi del vettore di errore sbiancato \mathbf{e}_{w_C} non sono più uniformi. La causa principale di questa alterazione è la trasformazione da coordinate polari (2.1) a coordinate rettangolari (2.2) che si ripercuote sul vettore d'errore:

$$\mathbf{e} \simeq \mathbf{J} \mathbf{e}_{\rho, \phi} \quad (2.19)$$

dove $\mathbf{e}_{\rho, \phi} = [e_{\rho_1}, \dots, e_{\rho_M}, e_{\phi_1}, \dots, e_{\phi_M}]^T$. Infatti, è possibile osservare che l'informazione sulle distribuzioni delle componenti di errore viene alterata applicando la trasformazione Jacobiana. Inoltre, la successiva applicazione della matrice di sbiancamento \mathbf{W}_C non è in grado di recuperare le distribuzioni originarie.

Per dimostrarlo, si considerino due misure di fasori armonici V_1 e $I_{9,10}$, con riferimento allo scenario di test che verrà presentato nel Paragrafo 2.4.1, i cui errori sia per l'ampiezza sia per l'angolo di fase in $\mathbf{e}_{\rho, \phi} = [e_{\rho_{V_1}}, e_{\rho_{I_{9,10}}}, e_{\phi_{V_1}}, e_{\phi_{I_{9,10}}}]^T$ sono distribuiti uniformemente. I loro possibili valori sono stati ottenuti con 10^6 prove Monte Carlo (MC). Le componenti di $\mathbf{e}_{\rho, \phi}$ sono state estratte casualmente secondo distribuzioni uniformi, i cui limiti sono stati definiti utilizzando il livello di incertezza che verrà discusso in Sezione 2.4.1. Durante ogni prova, ai valori misurati è stata applicata la trasformazione polare-rettangolare (2.10), in modo da ottenere $\mathbf{e} = [e_{V_1}^r, e_{I_{9,10}}^r, e_{V_1}^x, e_{I_{9,10}}^x]^T$. Osservando gli istogrammi delle sue componenti, mostrati in Figura 2.1, si può notare come la trasformazione da coordinate polari a rettangolari alteri le distribuzioni degli errori di misura che non sono più né uniformi, né identiche. Infine, il vettore di errore è stato sbiancato mediante \mathbf{W}_C . In Figura 2.2 sono riportati gli istogrammi degli ingressi del vettore di errore sbiancato $\mathbf{e}_{w_C} = [e_1^{w_C}, e_2^{w_C}, e_3^{w_C}, e_4^{w_C}]^T$. È evidente che queste distribuzioni non sono uniformi, come affermato in precedenza.

Per ovviare a questo problema, e per poter determinare un limite ammissibile della norma del vettore dell'errore di misura, si introduce di seguito un'apposita matrice di sbiancamento [32].

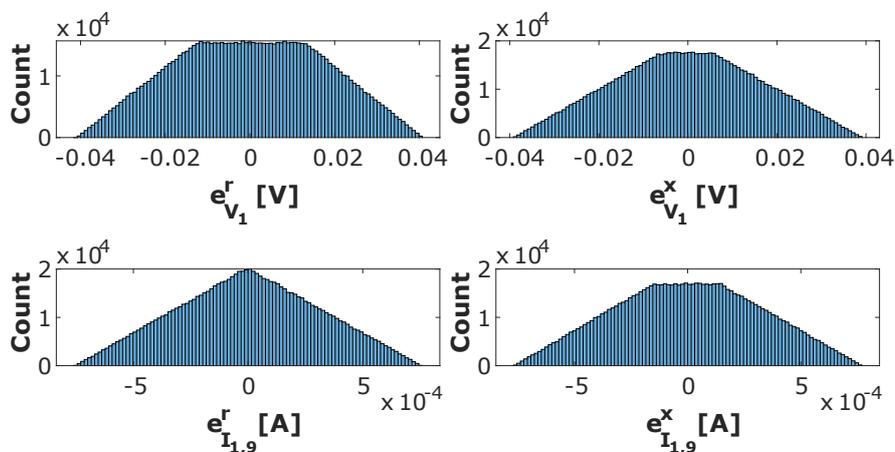


Figura 2.1: Istogrammi della componente reale e immaginaria delle misure di tensione e corrente NON sbiancate.

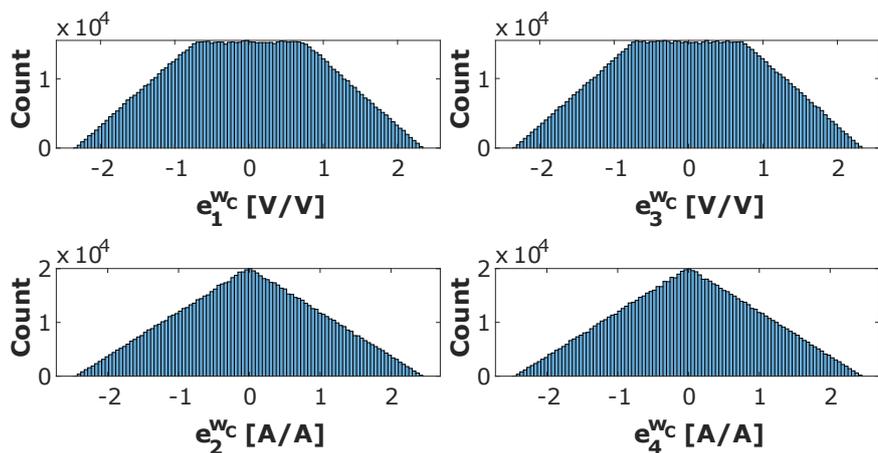


Figura 2.2: Istogrammi delle componenti delle misure rettangolari di tensione e corrente sbiancate tramite \mathbf{W}_C .

2.3.2 Matrice di sbiancamento proposta

Un opportuno sbiancamento può recuperare le informazioni sulle distribuzioni iniziali dell'errore di misura, che vengono modificate a causa della trasformazione da coordinate polari a coordinate rettangolari. La matrice di sbiancamento proposta sarà etichettata come \mathbf{W}_R , dove il pedice R indica che questa trasformazione ha la caratteristica di recuperare l'informazione.

Partendo dall'equazione (2.17), che rappresenta le infinite matrici che potrebbero soddisfare i requisiti, si propone l'utilizzo della seguente matrice di

sbiancamento⁷:

$$\mathbf{W}_R = \Sigma_{\rho,\phi}^{-\frac{1}{2}} \mathbf{J}^{-1} \quad (2.20)$$

Il nuovo vettore di errore sbiancato, indicato con \mathbf{e}_{w_R} , diventa:

$$\mathbf{e}_{w_R} = \mathbf{W}_R \mathbf{e} = \Sigma_{\rho,\phi}^{-\frac{1}{2}} \mathbf{J}^{-1} \mathbf{e} \quad (2.21)$$

e può essere messo in relazione diretta con il vettore di errore originale in coordinate polari:

$$\mathbf{e}_{w_R} \simeq \Sigma_{\rho,\phi}^{-\frac{1}{2}} \mathbf{J}^{-1} \mathbf{J} \mathbf{e}_{\rho,\phi} = \Sigma_{\rho,\phi}^{-\frac{1}{2}} \mathbf{e}_{\rho,\phi} \quad (2.22)$$

Osservando l'espressione (2.15), è immediato notare che in (2.22) la matrice $\Sigma_{\rho,\phi}^{-\frac{1}{2}}$ ha l'unico effetto di riscaldare ogni componente del vettore di errore $\mathbf{e}_{\rho,\phi}$, riferito sia all'errore di ampiezza sia all'errore di fase, in base alla propria deviazione standard (incertezza standard di misura). È quindi possibile affermare che attraverso la matrice di sbiancamento \mathbf{W}_R , a meno dell'approssimazione del primo ordine dovuta alla linearizzazione della propagazione dell'errore, è possibile ottenere un errore sbiancato che recupera le utili informazioni sulle distribuzioni dell'errore di misura originali, che sono state alterate nella trasformazione da coordinate polari a coordinate rettangolari (2.19).

Di conseguenza, partendo dalle stesse misure fasoriali armoniche precedentemente considerate nell'esempio, gli istogrammi dell'errore sbiancato $\mathbf{e}_{w_R} = [e_1^{w_R}, e_2^{w_R}, e_3^{w_R}, e_4^{w_R}]^T$, ottenuti mediante la matrice di sbiancamento proposta \mathbf{W}_R , sono mostrate in Figura 2.3 e evidenziano come gli errori sbiancati siano uniformi ed equidistribuiti.

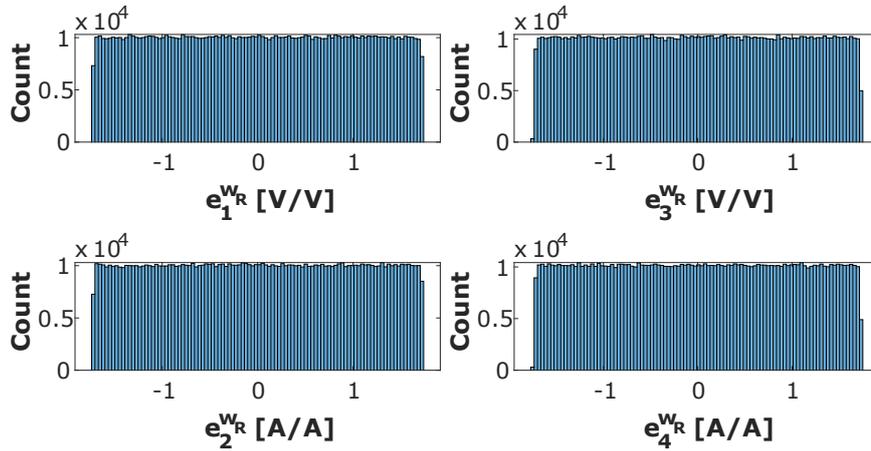


Figura 2.3: Istogrammi delle componenti delle misure rettangolari di tensione e corrente sbiancate tramite \mathbf{W}_R .

⁷Nella Sezione 1.6.3 è stato mostrato come sia possibile generalizzare la matrice proposta al generico caso di matrice Jacobiana quadrata non singolare.

Inoltre, osservando la formulazione (2.17) è possibile notare come tutti i vettori sbiancati, ottenuti applicando una qualsiasi trasformazione di sbiancamento, hanno lo stessa energia. Infatti, ricordando la proprietà delle matrici ortogonali di non alterare la norma 2 di un vettore (1.22), si ha⁸:

$$\begin{aligned} \forall \mathbf{Q} : \quad \mathbf{Q}\mathbf{Q}^\top &= \mathbf{Q}^\top\mathbf{Q} = \mathbf{I} \\ \|\mathbf{W}\mathbf{e}\|_2^2 &= \left\| \mathbf{Q}\boldsymbol{\Sigma}_{\rho,\Phi}^{-\frac{1}{2}}\mathbf{J}^{-1}\mathbf{e} \right\|_2^2 = \|\mathbf{Q}\|_2^2 \left\| \boldsymbol{\Sigma}_{\rho,\Phi}^{-\frac{1}{2}}\mathbf{J}^{-1}\mathbf{e} \right\|_2^2 = \\ &= \left\| \boldsymbol{\Sigma}_{\rho,\Phi}^{-\frac{1}{2}}\mathbf{J}^{-1}\mathbf{e} \right\|_2^2 = \|\mathbf{W}_R\mathbf{e}\|_2^2 \\ \implies \|\mathbf{e}_w\|_2^2 &= \|\mathbf{e}_{w_R}\|_2^2 \end{aligned} \quad (2.23)$$

dove $\mathbf{e}_w \triangleq \mathbf{W}\mathbf{e}$ è il generico vettore di errore sbiancato. Pertanto, se grazie ad una specifica trasformazione di sbiancamento è possibile stimare l'energia del vettore di errore sbiancato con la specifica matrice, tale stima resta valida per tutti i vettori di errore sbiancati che possono essere ottenuti dallo stesso vettore di errore.

Vale la pena evidenziare che calcolare \mathbf{W}_R è immediato perché la matrice è disponibile direttamente dalla costruzione di $\boldsymbol{\Sigma}_y$.

2.3.3 Valutazione dei limiti dell'energia dell'errore

Una volta note le distribuzioni degli ingressi del vettore dell'errore di misura, è possibile determinare il limite della sua energia. Come riportato nel Paragrafo 2.2.2, questo parametro gioca un ruolo chiave nell'applicazione del problema (P2), ed una sua errata valutazione potrebbe portare a una diminuzione delle prestazioni dell'algoritmo.

Si prendano in considerazione le M misure dei fasori armonici i cui errori, in ampiezza e angolo di fase, siano uniformemente distribuite. La i -esima componente di \mathbf{e}_{w_R} è una variabile casuale uniforme con deviazione standard unitaria e, indicandola con $e_{w_R,i}$, è possibile scrivere:

$$e_{w_R,i} \sim \mathcal{U}(-\sqrt{3}, \sqrt{3}) \quad \forall i = 1, \dots, 2M \quad (2.24)$$

per cui, l'energia del vettore di errore sbiancato è la somma di variabili aleatorie uniformi al quadrato:

$$\xi_{\mathbf{e}_w} \triangleq \|\mathbf{e}_w\|_2^2 = \|\mathbf{e}_{w_R}\|_2^2 = \sum_{i=1}^{2M} e_{w_R,i}^2 \quad (2.25)$$

La somma di n variabile aleatorie uniformi al quadrato $\sim \mathcal{U}(-\Delta, \Delta)$ ha media $\frac{n\Delta^2}{3}$ e deviazione standard $\frac{\Delta^2 2\sqrt{n}}{3\sqrt{5}}$ [60]. Nel caso in esame, dalla (2.24) segue:

⁸Il risultato è stato generalizzato con il Teorema 1.8.

$\Delta = \sqrt{3}$ e $n = 2M$, che conduce a:

$$\begin{aligned}\mu_{\xi_{\mathbf{e}_w}} &= E[\xi_{\mathbf{e}_w}] \simeq 2M \\ \sigma_{\xi_{\mathbf{e}_w}} &= \text{std}(\xi_{\mathbf{e}_w}) = E[(\xi_{\mathbf{e}_w} - \mu_{\xi_{\mathbf{e}_w}})^2] \simeq 2\sqrt{\frac{2}{5}}M\end{aligned}\quad (2.26)$$

Considerando l'equazione (2.7) e la menzionata *concentration inequality* è possibile stimare il limite dell'energia dell'errore come:

$$\epsilon_w^2 = \mu_{\xi_{\mathbf{e}_w}} + \lambda \sigma_{\xi_{\mathbf{e}_w}} \quad (2.27)$$

scegliendo λ in funzione del richiesto livello di confidenza [60].

Al fine di verificare la validità della relazione (2.27), partendo da un insieme di 11 misure di fasori armonici, aventi le stesse caratteristiche delle misure considerate nel Paragrafo 2.4.1, è stata valutata la distribuzione dell'energia dell'errore sbiancato. Inoltre, per mostrare graficamente che tutti i vettori di errore sbiancato sono caratterizzati dalla stessa energia, gli istogrammi dell'energia degli errori sbiancati \mathbf{e}_{w_C} e \mathbf{e}_{w_R} , ottenuti rispettivamente con le due trasformazioni presentate in precedenza, ovvero (\mathbf{W}_C e \mathbf{W}_R), e valutati tramite 10^6 prove MC, sono riportati in Figura 2.4. Il 95-esimo e il 99-esimo percentile

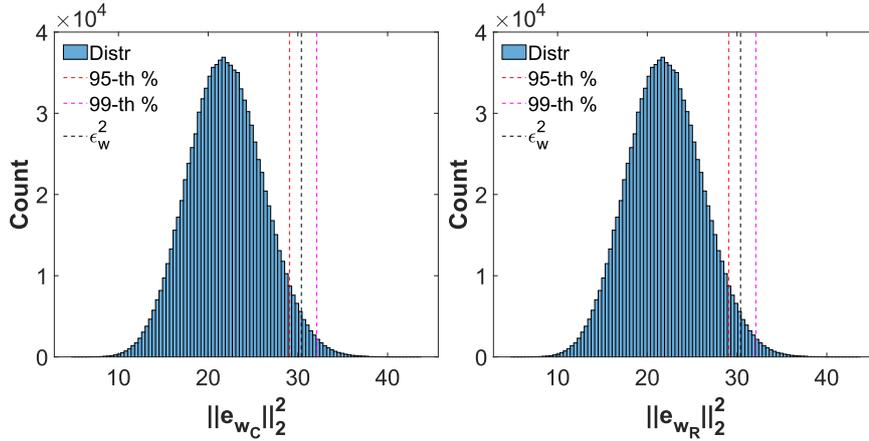


Figura 2.4: Istogrammi di $\xi_{\mathbf{e}_w} = \|\mathbf{e}_{w_C}\|_2^2$ e $\xi_{\mathbf{e}_w} = \|\mathbf{e}_{w_R}\|_2^2$, nei casi di misure sbiancate con le matrici di sbiancamento \mathbf{W}_C (a sinistra) e \mathbf{W}_R (a destra).

di ciascuna distribuzione dell'energia dell'errore sbiancato sono rappresentati, rispettivamente, con linee tratteggiate rosse e magenta, mentre il valore di ϵ_w^2 (stimato scegliendo $\lambda = 2$) è indicato con la linea tratteggiata nera. Come previsto, è possibile osservare che le due distribuzioni sono sostanzialmente identiche, con il 95-esimo percentile pari a 29.0630 e il 99-esimo percentile pari a 32.0938. Inoltre, il valore ϵ_w^2 calcolato mediante la stima (2.27) è compreso tra il 95-esimo e 99-esimo percentile della distribuzione dell'energia dell'errore sbiancato (ed è

pari a 30.3905). Quindi $\lambda = 2$ appare un buon compromesso tra il rischio di sottovalutazione di ξ_{e_w} e il rischio di un eccessivo rilassamento del vincolo sull'errore di approssimazione. Conseguentemente, nel seguito viene scelto $\lambda = 2$ per la valutazione del limite dell'energia dell'errore ϵ utilizzato nella formulazione (P2).

2.4 HSoE: prove e risultati

Al fine di confrontare le diverse prestazioni delle due formulazioni del problema di minimizzazione in ℓ_1 , denominate (P1) e (P2), entrambi gli algoritmi sono stati implementati all'interno della piattaforma HSoE ed applicati alle reti di distribuzione di test simulate.

2.4.1 Configurazione di test

La prima rete di prova utilizzata è una piccola rete di distribuzione, mostrata in Figura 2.5, a frequenza nominale di 60 Hz e tensione nominale $V_n = 4.16$ kV. Il sistema è composto da 13 nodi con un carico (*load*) potenzialmente inquinante connesso a ciascuno di essi, con la sola eccezione del nodo 1, dove è collegata la sottostazione, *Substation*.

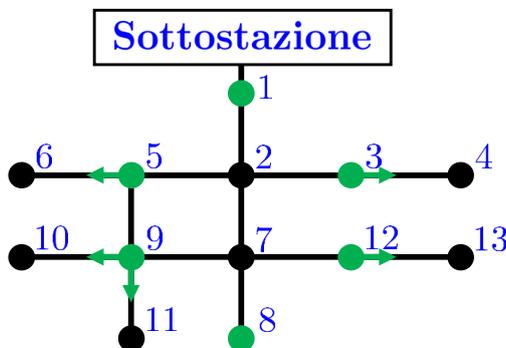


Figura 2.5: Schema della rete di test a 13 nodi.

I carichi sono modellati tramite un'impedenza ohmico-induttiva, i cui parametri sono definiti secondo le potenze nominali riportate nella Tabella 2.1, collegata in parallelo con una sorgente di corrente ideale che rappresenta la non linearità del carico e quindi l'inquinamento armonico iniettato nel sistema, come mostrato in Figura 2.6.

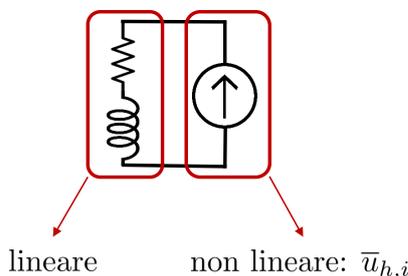


Figura 2.6: Modello di carico ohmico-induttivo con forzante armonica.

Tabella 2.1: Configurazione dei carichi nominali

Carichi	P_n [kW]	Q_n [kvar]
$L2$	160	110
$L3$	120	90
$L4$	170	125
$L5$	380	220
$L6$	120	90
$L7$	70	60
$L8$	290	210
$L9$	230	130
$L10$	490	190
$L11$	130	90
$L12$	380	220
$L13$	170	150

La corrente iniettata dal carico l -esimo si ottiene sommando i contributi armonici di tutti gli ordini armonici presi in considerazione. In questo studio sono stati modellati gli ordini armonici dispari dal terzo al 15-esimo ed è stata definita l'entità del livello di inquinamento in percentuale rispetto alle correnti nominali di ciascun carico. I valori di default per ogni ordine armonico sono riportati nella Tabella 2.2. Le prove sono state effettuate assumendo due sorgenti

Tabella 2.2: Configurazione delle Correnti Armoniche Forzanti

Ordine Armonico h	3	5	7	9	11	13	15
Ampiezza [%]	5	5	3	3	3	1	1

contemporanee di inquinamento armonico, fissando quindi il livello di sparsità $S = 2$, e simulando tutte le possibili $N_c = \binom{12}{2} = 66$ coppie di carichi inquinanti, per tutte le armoniche riportate in Tabella 2.2. Per concentrarsi sulle prestazioni delle soluzioni proposte e sull'impatto delle sorgenti di incertezza sulle prestazioni stesse, in tutte le prove sono stati utilizzati i carichi nominali.

CAPITOLO 2. IDENTIFICAZIONE DELLE PRINCIPALI SORGENTI
ARMONICHE NELLE RETI DI DISTRIBUZIONE

Le analisi sono state condotte ipotizzando un plausibile scenario in ambiente SG ed è stato quindi considerato che le misure sincronizzate dei fasori armonici, tensioni di nodo (V_i , con riferimento al nodo i -esimo) e correnti di ramo ($I_{i,j}$, con riferimento al ramo delimitato dai nodi i e j) derivati da dispositivi specifici per la PQ simili alle PMU, siano disponibili in 6 nodi della rete: nodo 1 (V_1), nodo 3 (V_3 e $I_{3,4}$), nodo 5 (V_5 e $I_{5,6}$), nodo 8 (V_8), nodo 9 (V_9 , $I_{9,10}$ e $I_{9,11}$) e nodo 12 (V_{12} e $I_{12,13}$). Ulteriori analisi sul posizionamento (*placement*) ottimale dei dispositivi di misurazione non rientrano nell'ambito di questa analisi e pertanto non sono nel seguito prese in considerazione.

Per certificare la validità della metodologia proposta, come prova iniziale entrambi gli algoritmi ($P1$) e ($P2$) sono stati utilizzati in uno scenario ideale, trascurando quindi gli errori di misura. Questi test iniziali, ovviamente, non corrispondono ad una realistica condizione operativa del sistema, ma sono essenziali per stabilire le migliori prestazioni possibili degli algoritmi in un contesto ideale. In presenza delle condizioni ideali per la configurazione di misura considerata, entrambi gli algoritmi sono stati in grado di identificare correttamente entrambi i carichi inquinanti in ciascuna delle N_c combinazioni possibili.

Tuttavia, come detto in precedenza, e discusso ad esempio in [50], le incertezze di misura assumono un ruolo chiave nel HSoE, e il loro impatto sulle prestazioni degli algoritmi di stima deve essere adeguatamente investigato. Di seguito viene valutato l'impatto delle incertezze di misura sulle prestazioni delle due diverse formulazioni ($P1$) e ($P2$) mediante prove MC. Per ogni fasore armonico misurato, sia per l'ampiezza sia per l'angolo di fase, è stato preso in considerazione un errore di misura additivo estratto da distribuzioni di tipo uniforme, i cui limiti sono stati definiti secondo tre diversi livelli di incertezza chiamati semplicemente basso, medio e alto, *Low*, *Medium*, *High*, sintetizzati in Tabella 2.3, nella quale l'errore di ampiezza è riportato in percentuale rispetto all'ampiezza del valore da misurare.

Tabella 2.3: Errori di Misura Massimi

	<i>Max Errore di Ampiezza</i>	<i>Max Errore di Fase</i>
	[%]	[crad]
Low	0.5	0.6
Medium	2.5	3
High	5	6

Tutte le simulazioni sono state eseguite in ambiente MATLAB, utilizzando le routine `l1eq_pd` e `l1qc_logbarrier`, per risolvere rispettivamente i problemi ($P1$) e ($P2$), che fanno parte della libreria ℓ_1 -MAGIC presentata in [11]. I vettori recuperati con gli algoritmi di minimizzazione in ℓ_1 sono nel campo reale, perchè riferiti al sistema (2.2), e non sono esattamente sparsi, poiché ogni

ingresso del vettore ricostruito è in generale diverso da zero. La soluzione dei problemi (P1) e (P1) viene prima ricomposta come vettore complesso, quindi le S sorgenti corrispondenti alle sue componenti con modulo massimo vengono classificate come inquinanti. Il valore ϵ in (2.5) è stato stimato mediante l'approccio di valutazione dell'energia del vettore di errore presentato nel Paragrafo 2.3, utilizzando $\lambda = 2$.

2.4.2 Risultati delle prove

In questa sezione vengono presentate e confrontate le prestazioni dei due algoritmi. I risultati dell'identificazione saranno espressi mediante la metrica statistica chiamata *Recall*, nota anche come sensibilità, comunemente usata nel *Pattern Recognition*, *Data Mining*, ML, ecc., e che nel nostro caso corrisponde al corretto tasso di identificazione delle sorgenti di inquinamento armonico. Dato il numero delle sorgenti inquinanti NS e denotando, rispettivamente, con NT e NF il numero di identificazioni corrette ed errate delle sorgenti inquinanti fornite dall'algoritmo, la *Recall* può essere definita come:

$$Recall = \frac{NT}{NT + NF} = \frac{NT}{NS} \quad (2.28)$$

È immediato verificare che $Recall \in [0, 1]$, dove al valore 1 corrisponde la perfetta identificazione.⁹ In questo studio si ipotizza il caso di due fonti contemporaneamente inquinanti su 12 sorgenti potenziali, sono quindi possibili 66 diverse combinazioni non ripetitive. Per ognuna delle possibili combinazioni di coppie di sorgenti inquinanti, indicata con il pedice c , $c = \{1, \dots, N_c\}$, e date $N_{MC} = 1000$ prove MC, è possibile definire la corrispondente *Recall* media al variare delle prove MC nel seguente modo:

$$Recall_c \triangleq \frac{1}{N_{MC}} \sum_{t=1}^{N_{MC}} \frac{NT_{c,t}}{NS_{c,t}} = \frac{\sum_{t=1}^{N_{MC}} NT_{c,t}}{N_{MC} NS} \quad (2.29)$$

tenuto conto del fatto che NS è costante per tutte le prove.

Una volta definita $Recall_c$ come *Recall* media per singolo caso (2.29), è possibile definire due parametri come la media e la deviazione standard di $Recall_c$ al variare delle varie N_c combinazioni, da utilizzare come indicatori di prestazione,

⁹La metrica *Recall* è solitamente utilizzata in associazione alla metrica *Precision* che nel caso in esame è stata omessa perché, data la configurazione di test presa in considerazione, è ridondante. Infatti le due metriche *Recall* e *Precision* differiscono per il denominatore. In particolare esprimono la percentuale la identificazioni corrette, nel primo caso rispetto a quelle rilevanti mentre nel secondo caso rispetto alla totalità delle rilevazioni effettuate. In questo lavoro, per ogni prova, il numero di sorgenti inquinanti e il numero di identificazioni sono costanti e pari in entrambi i casi a $NS = 2$, quindi $Recall = Precision$.

sintetici e complessivi, degli algoritmi:

$$\mu_{R_c} = \frac{\sum_{c=1}^{N_c} Recall_c}{N_c}$$

$$\sigma_{R_c} = \sqrt{\frac{\sum_{c=1}^{N_c} (Recall_c - \mu_{R_c})^2}{N_c - 1}}$$
(2.30)

La Figura 2.7 riporta gli intervalli assunti dai valori dell'indice $Recall_c$ insieme ai valori medi per entrambi gli algoritmi e per ogni livello di incertezza, con riferimento al 3°, 5° e 7° ordine armonico. È possibile osservare che, a causa del fatto che $Recall_c$ ha un supporto compatto $[0, 1]$, le sue medie sono prossime ai valori massimi quando l'incertezza è bassa, e gli intervalli sono fortemente non simmetrici.

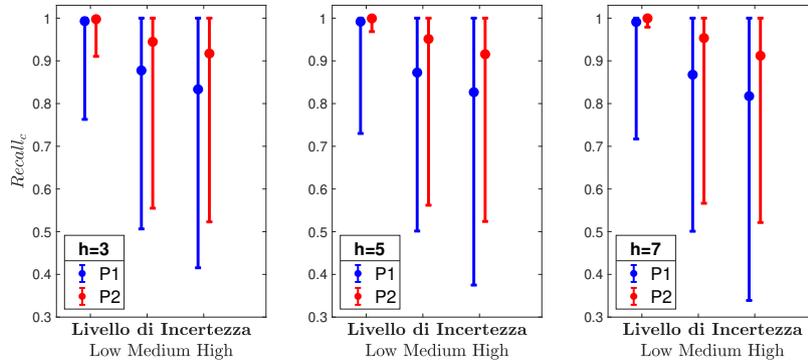


Figura 2.7: Intervalli e medie per $Recall_c$, per gli ordini armonici $h \in \{3, 5, 7\}$.

Nella Tabella 2.4, sono riportati i valori di μ_{R_c} e σ_{R_c} ottenuti per l'analisi degli ordini armonici 3°, 5° e 7° per entrambi gli algoritmi di minimizzazione in ℓ_1 , con i tre livelli di incertezza presi in considerazione nelle tre coppie di colonne. Gli indicatori del (P2) sono migliori rispetto a quelli del (P1) in tutte le configurazioni. Per entrambi gli algoritmi, come prevedibile, gli indicatori di prestazione peggiorano all'aumentare del livello di incertezza (*Low*, *Medium*, *High*) quindi i valori di μ_{R_c} diminuiscono mentre aumenta la deviazione standard σ_{R_c} . Inoltre, vale la pena sottolineare come anche le differenze tra le prestazioni dei due algoritmi aumentino all'aumentare dell'incertezza di misura: questo è dovuto alla differente formulazione dei due algoritmi di minimizzazione. Infatti, come discusso in precedenza, la possibilità di includere informazioni sull'errore di misura nella formulazione dell'algoritmo (P2), attraverso la definizione del limite dell'energia dell'errore ϵ , è la chiave per ottenere prestazioni migliori.

Va evidenziato che, mentre la valutazione del μ_{R_c} fornisce informazioni sulle prestazioni complessive dei due algoritmi, il valore σ_{R_c} rivela informazioni sulla stabilità delle prestazioni nelle diverse condizioni. In particolare, i valori σ_{R_c}

CAPITOLO 2. IDENTIFICAZIONE DELLE PRINCIPALI SORGENTI
ARMONICHE NELLE RETI DI DISTRIBUZIONE

Tabella 2.4: Media e deviazione standard di $Recall_c$ per i due algoritmi (rete di test a 13 nodi)

Ordine armonico	Livello di incertezza						
	Low		Medium		High		
	μ_{R_c}	σ_{R_c}	μ_{R_c}	σ_{R_c}	μ_{R_c}	σ_{R_c}	
$h = 3$	(P1)	0.993	0.032	0.878	0.187	0.833	0.219
	(P2)	0.998	0.013	0.945	0.128	0.917	0.151
$h = 5$	(P1)	0.992	0.037	0.873	0.194	0.827	0.222
	(P2)	0.999	0.004	0.951	0.112	0.916	0.148
$h = 7$	(P1)	0.991	0.039	0.868	0.200	0.817	0.223
	(P2)	0.999	0.003	0.953	0.106	0.912	0.148

più elevati per (P1) suggeriscono che l'identificazione fornita dall'algoritmo è meno stabile rispetto variare delle sorgenti inquinanti.

Focalizzando l'attenzione sui risultati di identificazione di ciascuna sorgente di inquinamento armonico, è possibile evidenziare ulteriormente la diversità in termini di prestazioni dei due algoritmi (P1) e (P2). A tal proposito, in Figura 2.8 è riportata la rilevazione del carico $L2$, con riferimento al 5° ordine armonico, in percentuale rispetto alle 1000 prove MC utilizzate per ogni possibile combinazione di carichi inquinanti. In ascissa è riportato il secondo carico (oltre al carico $L2$) inquinante per ogni scenario (livello di incertezza), mentre le linee blu e rossa indicano il tasso medio di rilevamento del carico $L2$ ottenuto rispettivamente con i metodi (P1) e (P2). Osservando il grafico in alto della figura, riferito al livello di incertezza più basso simulato nei test, è possibile vedere come entrambi gli algoritmi identificano correttamente il carico $L2$ quando è combinato con qualsiasi altro carico della rete. Al contrario, osservando gli altri due grafici (rispettivamente al centro e in basso nella figura), che si riferiscono rispettivamente al livello di incertezza medio e alto, è evidente che le prestazioni di entrambi gli algoritmi diminuiscono, in modo più marcato per dall'algoritmo (P1). Quando si assume il livello di incertezza massimo tra quelli simulati, l'algoritmo (P2) rileva correttamente il carico inquinante $L2$ in oltre il 55% dei casi, per ogni possibile combinazione di sorgenti inquinanti a cui è associato, mentre la rilevazione del (P1) raggiunge questo risultato solamente per una delle 11 combinazioni prese in considerazione.

Nelle Figure 2.9 e 2.10 è riportata la percentuale di identificazione rispettivamente dei carichi $L7$ e $L11$, quando iniettano inquinamento armonico assieme ad un altro carico, considerando tutti i carichi possibili in accoppiamento. Questi dati sottolineano come le prestazioni di entrambi gli algoritmi possano variare in modo significativo al variare dei carichi inquinanti. Mentre il carico $L11$ viene

CAPITOLO 2. IDENTIFICAZIONE DELLE PRINCIPALI SORGENTI ARMONICHE NELLE RETI DI DISTRIBUZIONE

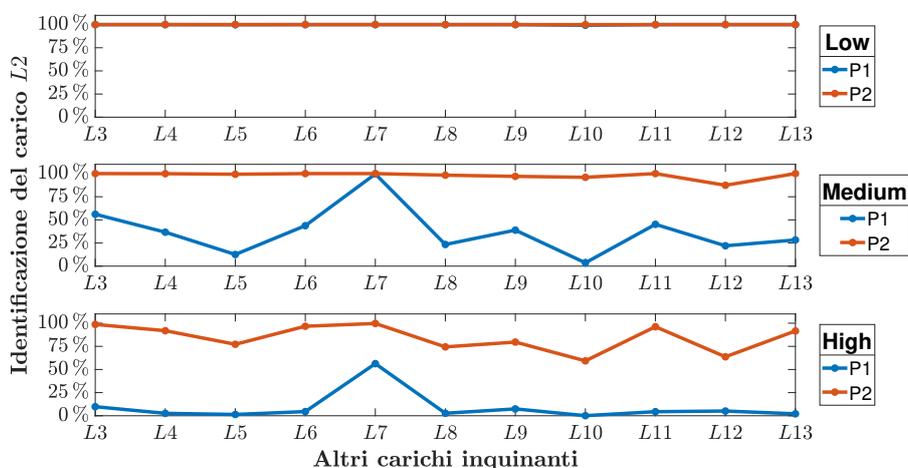


Figura 2.8: Identificazione del carico inquinante $L2$ in tutte le configurazioni, per l'ordine armonico $h = 5$.

identificato quasi perfettamente, in modo scarsamente dipendente dal livello di incertezza considerato, l'identificazione del carico $L7$ è più sensibile all'aumento dell'incertezza di misura. In particolare, è evidente come, anche quando si tiene correttamente conto degli effetti delle incertezze di misura, le caratteristiche metrologiche degli strumenti di misura giocano un ruolo fondamentale nella rilevazione delle sorgenti inquinanti, e strumenti ad alta accuratezza potrebbero portare a tassi di corretta identificazione significativamente migliori.

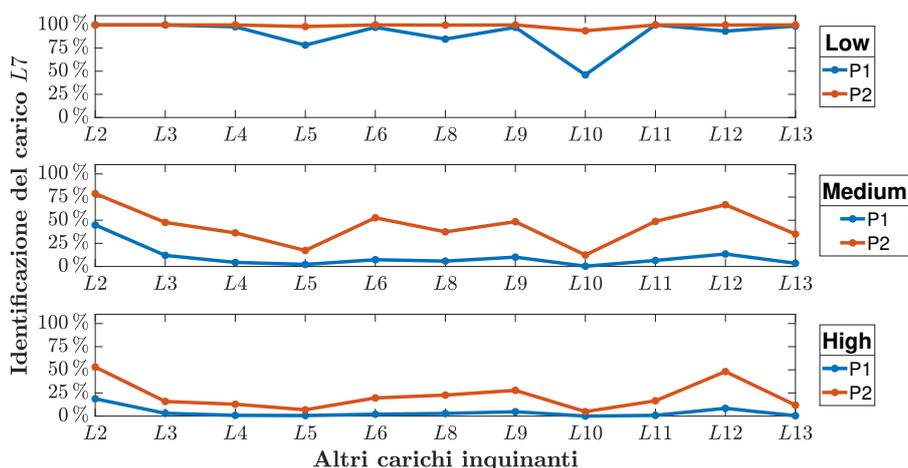


Figura 2.9: Identificazione del carico inquinante $L7$ in tutte le configurazioni, per l'ordine armonico $h = 5$.

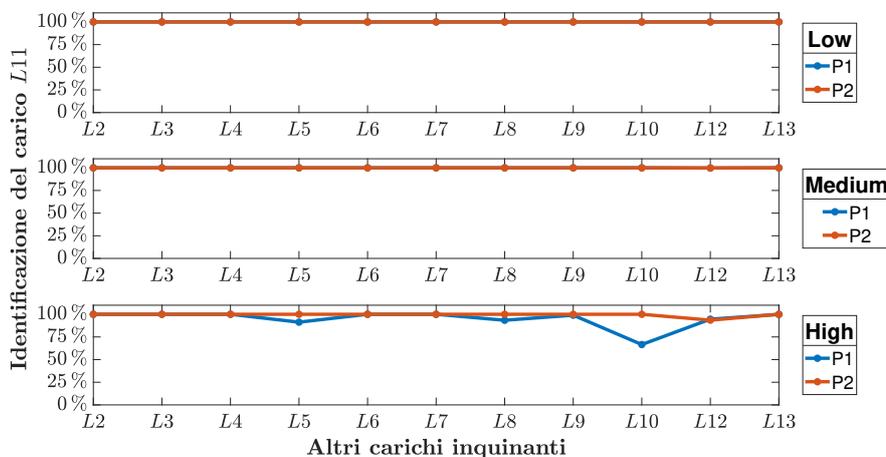


Figura 2.10: Identificazione del carico inquinante $L11$ in tutte le configurazioni, per l'ordine armonico $h = 5$.

Prove ulteriori su una rete piú grande

Per verificare la validità della tecnica descritta su uno scenario differente, sono state condotte nuove analisi su una rete di distribuzione piú ampia, caratterizzata da 49 nodi e 20 posizioni di misura, le cui caratteristiche sono riportate in Appendice 2.A. Inoltre, per dare una prospettiva piú larga, le prestazioni di entrambi gli algoritmi di minimizzazione in ℓ_1 , ovvero ($P1$) e ($P2$), sono state messe a confronto anche con le prestazioni di una tecnica basata sui minimi quadrati pesati (WLS per brevità nel seguito) a norma minima¹⁰, già proposta in [50] come termine di confronto.

Nella Tabella 2.5, sono riportati i risultati del confronto eseguito seguendo lo stesso approccio utilizzato per le prove precedentemente discusse (ovvero eseguendo prove in modo esaustivo su tutte le $N_c = \binom{48}{2} = 1128$ possibili coppie inquinanti per tutti e tre livelli di incertezza e 1000 prove MC), con riferimento agli ordini armonici $h \in \{3, 5, 7\}$, mediante la media e la deviazione standard dell'indicatore $Recall_c$, come definito in (2.30). I risultati mostrano in modo inequivocabile come l'algoritmo ($P2$) proposto, grazie ad una corretta modellazione dell'incertezza, supera nettamente gli altri due metodi in tutte le condizioni analizzate.

2.5 Conclusioni

Questo capitolo ha presentato un metodo per l'identificazione delle sorgenti armoniche nelle reti di distribuzione. Tale identificazione rappresenta uno stru-

¹⁰Corrispondente ad una formulazione pesata $\mathcal{P}_{r_2/e=0}$.

CAPITOLO 2. IDENTIFICAZIONE DELLE PRINCIPALI SORGENTI
ARMONICHE NELLE RETI DI DISTRIBUZIONE

Tabella 2.5: Media e deviazione standard di $Recall_c$ per i tre algoritmi (rete a 49 nodi)

Ordine armonico	Livello di incertezza						
		Low		Medium		High	
		μ_{R_c}	σ_{R_c}	μ_{R_c}	σ_{R_c}	μ_{R_c}	σ_{R_c}
$h = 3$	WLS	0.823	0.259	0.815	0.259	0.743	0.243
	(P1)	0.945	0.114	0.829	0.240	0.743	0.242
	(P2)	0.989	0.058	0.941	0.135	0.893	0.169
$h = 5$	WLS	0.823	0.258	0.813	0.259	0.733	0.241
	(P1)	0.945	0.119	0.826	0.242	0.731	0.239
	(P2)	0.989	0.055	0.946	0.125	0.894	0.165

mento importante per il gestore del sistema, che può consentire di agire direttamente sulla causa del problema, per ridurne gli effetti negativi. L'utilizzo di algoritmi progettati ad hoc per questo tipo di reti, come le metodologie basate sul CS, permette di superare i problemi legati alla limitata disponibilità di strumenti di misura per la PQ. Tuttavia, le prestazioni degli algoritmi di identificazione dipendono dalle incertezze di misura e il loro impatto non può essere trascurato.

Nel lavoro presentato in [32], che è la base della ricerca presentata in questo capitolo, partendo dagli aspetti teorici coinvolti nella valutazione delle incertezze di misura e nella riduzione del loro impatto sugli algoritmi HSoE, è stata proposta una nuova formulazione del problema di stima, basata sulla minimizzazione in ℓ_1 con vincolo sull'errore di approssimazione in ℓ_2 , denominata (P2), corrispondente ad una formulazione \mathcal{P}_{r_1/e_2} . Al fine di massimizzare le prestazioni della soluzione proposta, è stata presentata una matrice di sbiancamento che consente il recupero di informazioni sulle distribuzioni degli errori di misura, e quindi la stima dei relativi limiti, cardine per le prestazioni dell'algoritmo. L'efficacia della soluzione presentata è stata testata confrontando le sue prestazioni in termini di identificazione delle sorgenti di inquinamento, sia rispetto alla formulazione della minimizzazione in ℓ_1 con vincolo di uguaglianza, denominata (P1), sia con un approccio basato sui minimi quadrati pesati a norma (quadratica) minima, corrispondenti a formulazioni $\mathcal{P}_{r_1/e=0}$ e $\mathcal{P}_{r_2/e=0}$ rispettivamente. Le prove sono state eseguite in ambiente controllato attraverso simulazioni eseguite su due reti di distribuzione, tenendo conto di differenti scenari per i possibili sistemi di misura disponibili sulle SG. I risultati sottolineano come la formulazione proposta, integrando nel modello le opportune incertezze di misura, raggiunga le migliori prestazioni in termini di identificazione delle sorgenti di inquinamento e di stabilità al variare sia dell'incertezza sia delle sor-

CAPITOLO 2. IDENTIFICAZIONE DELLE PRINCIPALI SORGENTI ARMONICHE NELLE RETI DI DISTRIBUZIONE

genti di inquinamento da rilevare. Questa modellazione appare quindi come il passaggio necessario per rendere gli algoritmi più robusti in vista della futura implementazione sul campo da parte degli operatori di sistema.

Appendice

2.A Modello della rete di test a 49 nodi

Le principali caratteristiche della rete di prova a 49 nodi sono riportate in Figura 2.11, Tabella 2.6 e Tabella 2.7.

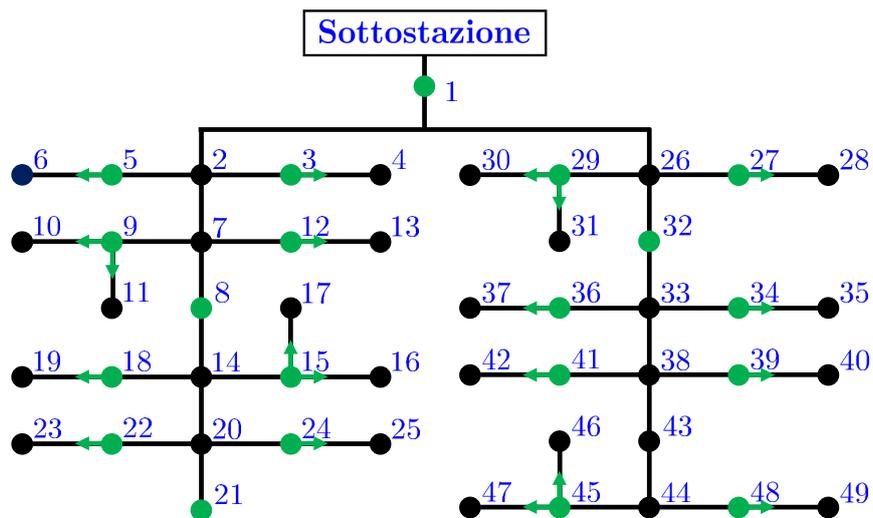


Figura 2.11: Topologia della rete di prova a 49 nodi

CAPITOLO 2. IDENTIFICAZIONE DELLE PRINCIPALI SORGENTI
ARMONICHE NELLE RETI DI DISTRIBUZIONE

Tabella 2.6: Carichi nominali della rete di prova a 49 nodi

Carichi	P_n [kW]	Q_n [kvar]
<i>L3, L21, L34, L39</i>	36	27
<i>L4, L22, L35, L40</i>	51	38
<i>L5, L23, L36, L41</i>	81	48
<i>L6, L24, L37, L42</i>	36	27
<i>L7, L14, L26, L44</i>	27	18
<i>L8, L25, L32, L43</i>	42	24
<i>L9, L15, L29, L45</i>	69	39
<i>L10, L16, L30, L46</i>	78	39
<i>L11, L17, L31, L47</i>	39	27
<i>L12, L18, L27, L48</i>	90	45
<i>L13, L19, L28, L49</i>	51	45

Tabella 2.7: Configurazione di misura della rete ti prova a 49 nodi

Tensioni di Nodo	$V_1, V_3, V_5, V_8, V_9, V_{12}, V_{15}, V_{18}, V_{21}, V_{22}$ $V_{24}, V_{27}, V_{29}, V_{32}, V_{34}, V_{36}, V_{39}, V_{41}, V_{45}, V_{48}$
Correnti di Ramo	$I_{3,4}, I_{5,6}, I_{9,10}, I_{9,11}, I_{12,13}, I_{15,16}$ $I_{15,17}, I_{18,19}, I_{22,23}, I_{24,25}, I_{27,28}, I_{29,30}$ $I_{29,31}, I_{34,35}, I_{36,37}, I_{39,40}, I_{41,42}, I_{45,46}, I_{45,47}, I_{48,49}$

Capitolo 3

Stima dei parametri di linea e compensazione degli errori sistematici in reti di trasmissione

3.1 Introduzione

Nel contesto dei sistemi di gestione e controllo delle moderne reti elettriche di trasmissione è essenziale conoscere accuratamente i parametri di linea delle reti. I valori reali dei parametri possono infatti differire in modo significativo, anche solo a causa dell'invecchiamento, da quelli nominali salvati nei database dei sistemi di gestione dell'energia, o *Energy Management System*, (EMS). Questo impatta negativamente non solo sulle funzionalità direttamente collegate agli EMS ma anche sull'efficienza ed efficacia del sistema di dispacciamento, con conseguenti ricadute tecnico-economiche negative [64].

Le unità di misura fasoriali, *Phasor Measurement Unit* (PMU), forniscono misure accurate di tensione e corrente sincronizzate rispetto ad un comune riferimento temporale, tipicamente il *Coordinated Universal Time* (UTC), e grazie all'elevata frequenza di misura costituiscono gli strumenti più avanzati per superare le difficoltà legate ai tradizionali sistemi di monitoraggio. Negli ultimi anni, infatti, le PMU sono state largamente adottate da parte di molti operatori di trasmissione dell'energia, *Transmission System Operator* (TSO), in tutto il mondo.

È comunque importante sottolineare il fatto che, nonostante l'alta accura-

tezza delle PMU, le misure fasoriali subiscono le sorgenti di incertezza di tutta la catena di misura e, in particolare, dei trasformatori di misura, come i trasformatori di corrente (TA) e i trasformatori di tensione (TV), che possono impattare in modo significativo sull'accuratezza delle misure fasoriali.

Per una determinazione efficace dei valori dei parametri di rete, è necessario quindi applicare un processo di valutazione che tenga conto opportunamente delle incertezze di misura. Inoltre, i metodi di stima devono essere basati su appropriati modelli delle sorgenti di incertezza, in modo tale che i loro effetti possano essere adeguatamente considerati e mitigati. Nel contesto della sistemi di protezione della rete elettrica, dove gli algoritmi di localizzazione dei guasti comunemente utilizzano i parametri di trasmissione disponibili nei database, è possibile citare [65], dove viene utilizzata la legge di propagazione dell'incertezza [31] per evidenziare come l'efficacia della localizzazione sia influenzata dall'incertezza di misura. Il problema dell'incertezza dei parametri della linea di trasmissione è affrontato anche in [66], dove viene derivata un'espressione analitica per definire i limiti dell'errore nella valutazione dei parametri della linea di trasmissione calcolati in modo diretto dalle misure delle PMU. Il medesimo algoritmo di calcolo diretto viene utilizzato in [67] sulla base di prove sul campo.

Nel contesto della stima dei parametri di linea, sono stati presentati numerosi studi indirizzati a diversi aspetti del problema. La maggior parte dei lavori prendono in considerazione la sola incertezza delle PMU, ad esempio [68]. In [69] viene proposta la stima dell'impedenza di una linea di trasmissione trifase, stima basata sulle PMU e fatta prendendo in considerazione la loro incertezza, benché si riconosca che l'accuratezza dei trasformatori di misura dovrebbe essere considerata durante la valutazione per una efficace applicazione della proposta in un contesto di rete reale. Altri articoli considerano la presenza dei trasformatori di corrente e tensione, attribuendo ad essi un dato contributo casuale come in [70] e [71], mentre in altri casi, ad esempio in [72], i trasformatori di misura vengono considerati pre-calibrati.

In [73] viene affrontato il problema della stima simultanea dei parametri della rete e della compensazione degli errori sistematici di misura trascurando però sia gli errori sistematici di fase delle tensioni, sia gli errori di ampiezza delle correnti. In [74] vengono discussi gli inconvenienti che insorgono quando si trascurano gli errori sistematici, quindi viene proposto un nuovo metodo che sfrutta l'aggiunta al modello di regressione di parametri extra che tengano conto delle polarizzazioni dovute ai trasformatori di misura non calibrati. L'articolo [75] presenta un metodo per la calibrazione in esercizio dei trasformatori di tensione mediante misurazioni sincrofasoriali, con il posizionamento ottimale di nuove misure di accuratezza elevata al fine di portare l'errore di calibrazione al di sotto di una soglia assegnata. In [76] viene sviluppato un metodo che

sfrutta la presenza di trasformatori di misura calibrati in un singolo nodo della rete, per propagare la calibrazione ai trasformatori installati negli altri nodi sfruttando la conoscenza della topologia della rete. In [77] e [78] viene affrontato il problema della stima simultanea dei parametri di linea e della correzione degli errori sistematici per un singolo ramo, utilizzando misure fasoriali compensate, a partire dai test di rilevamento degli errori sistematici nei trasformatori di misura.

In questo contesto, il capitolo presenta un lavoro di ricerca che ha portato anche alla stesura di un articolo sottoposto alla rivista IEEE Transactions on Instrumentation and Measurement e recentemente accettato [79]. La ricerca affronta il problema di un'efficace stima simultanea dei parametri di linea e degli errori sistematici di ampiezza e fase dovuti ai TA e TV nelle misure sincrofascoriali di corrente e tensione, senza alcuna assunzione sulla presenza di dispositivi precalibrati e prendendo in considerazione più condizioni operative. In particolare, si è sviluppata e migliorata la procedura proposta in [80], dove il problema è stato affrontato nel senso dei minimi quadrati pesati WLS, inizialmente in un ramo singolo e poi prendendo in considerazione più rami simultaneamente, ma sfruttando le misure acquisite in poche differenti condizioni operative.

Come sottolineato ad esempio in [81], il problema della stima dei parametri può essere sotto-determinato, è quindi utile sfruttare misurazioni relative a diverse condizioni operative, abbastanza differenti da stabilizzare il sistema, ma sufficientemente ravvicinate da poter considerare invariati i parametri di linea e la risposta degli strumenti di misura, rendendo il sistema sovra-determinato. Inoltre, in [82] viene evidenziato l'effetto positivo sulla stima ottenuto con l'aggiunta di misure acquisite in istanti temporali diversi e corrispondenti a diverse condizioni operative, sottolineando così l'importanza di includere più insiemi di misure nel metodo di stima.

Il lavoro di ricerca presentato in questo capitolo generalizza il metodo basato sul WLS, presentato in [80], inquadrandolo come caso particolare all'interno della regolarizzazione di Tikhonov, e da qui presenta un metodo per migliorare l'accuratezza della stima affrontando la gestione di molteplici condizioni operative nel processo di stima. I principali contributi presentati sono: una piattaforma per l'ottimizzazione della stima simultanea dei parametri di linea e degli errori sistematici in presenza di più condizioni operative, un criterio di configurazione per il problema di stima e l'introduzione di un algoritmo avanzato multi-ramo per sfruttare le capacità delle PMU in questo contesto. Il lavoro si basa su un'analisi dettagliata dell'incertezza di stima sfruttando gli strumenti di analisi ed ottimizzazione disponibili all'interno della regolarizzazione in ℓ_2 presentata nel Paragrafo 1.5.

Il Capitolo è organizzato come segue:

- Il Paragrafo 3.2 formula il modello di stima dei parametri di linea e degli errori sistematici derivanti dalla catena di misura. Il problema viene inquadrato all'interno della regolarizzazione di Tikhonov.
- Il Paragrafo 3.3 sfrutta gli strumenti matematici della regolarizzazione di Tikhonov per effettuare una accurata valutazione dell'incertezza della stima ed aumentarne le prestazioni gestendo in modo adeguato le misure derivanti da molteplici casi operativi.
- Il Paragrafo 3.4 è dedicato alla presentazione dei test condotti sulla rete di trasmissione IEEE a 14 nodi e alla discussione dei risultati ottenuti.
- Infine le conclusioni si trovano nel Paragrafo 3.5.

3.2 Formulazione del modello di stima

3.2.1 Modello di misura e vincoli di rete

Il generico ramo (i, j) di una rete di trasmissione è stato considerato con un modello a π come in Figura 3.1. In questo studio, per una più semplice e chiara introduzione all'approccio proposto, come ad esempio in [76], viene utilizzato un modello monofase equivalente del sistema trifase. Si assume che a ciascun capo della linea sia disponibile un'unità di misura sincrofasoriale che consente di misurare due sincrofasori di tensione (v_i e v_j rispettivamente per il nodo iniziale e finale) e due sincrofasori di corrente di ramo (i_{ij} e i_{ji}). Con $z_{ij} = R_{ij} + jX_{ij}$ si indica l'impedenza di linea nel modello π , mentre $B_{sh,ij}$ rappresenta la suscettanza posta in derivazione (*shunt*), che si assume divisa equamente ai due lati del ramo. Le misure sincronizzate possono essere allineate nel tempo e rappresentare quindi un insieme coordinato di misure riferite allo stesso istante temporale t . Il modello di linea definisce un modello di misura che collega l'insieme delle misure ai parametri di linea e agli errori che interessano ogni valore misurato.

Ciascun sincrofasore misurato può essere espresso in funzione dei valori di riferimento (indicati nelle seguenti equazioni con l'apice R ad indicare *Reference value*) e degli errori di misura, come segue:

$$\begin{aligned}
 v_h &= V_h e^{j\varphi_h} = V_h^r + jV_h^x \\
 &= (1 + \xi_h^{sys} + \xi_h^{rnd}) V_h^R e^{j(\varphi_h^R + \alpha_h^{sys} + \alpha_h^{rnd})} \\
 i_{ij} &= I_{ij} e^{j\theta_{ij}} = I_{ij}^r + jI_{ij}^x \\
 &= (1 + \eta_{ij}^{sys} + \eta_{ij}^{rnd}) I_{ij}^R e^{j(\theta_{ij}^R + \psi_{ij}^{sys} + \psi_{ij}^{rnd})}
 \end{aligned} \tag{3.1}$$

dove V_h e φ_h (con $h \in \{i, j\}$) sono, rispettivamente, le misure di ampiezza e di fase della tensione al nodo h , mentre I_{ij} e θ_{ij} sono le misure di ampiezza e di

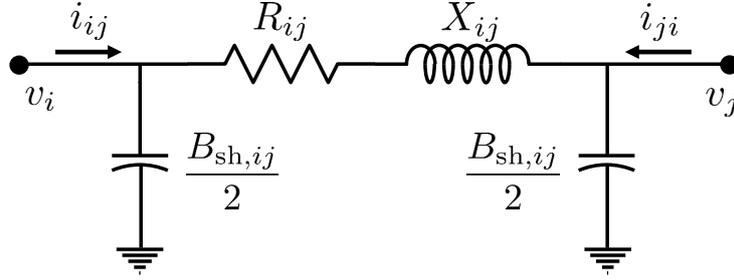


Figura 3.1: Modello a π per un generico ramo di rete di trasmissione e relativi parametri.

fase della corrente che fluisce dal nodo i verso il nodo j . Il fasore di corrente misurato i_{ji} (vedi Figura 3.1) può essere espresso in modo simile. Gli apici r e x sono utilizzati per indicare le parti reale e immaginaria dei corrispondenti fasori. Gli errori di misura possono essere sistematici o casuali (rispettivamente indicati con i suffissi sys e rnd) e interessano sia le ampiezze sia gli angoli di fase dei fasori. Infatti, i parametri ξ_h e η_{ij} si riferiscono agli errori di rapporto, mentre α_h e ψ_{ij} sono gli errori di fase per le quantità in (3.1). La principale differenza tra le due tipologie di errore è che, mentre gli errori sistematici non variano tra misurazioni ripetute¹, gli errori casuali invece danno un contributo differente per ogni osservazione. In questo lavoro, come in [80], l'idea alla base dell'approccio di stima è che gli errori sistematici nella catena di misura non possono essere trascurati nel modello e sono quindi trattati come parametri incogniti aggiuntivi da stimare insieme ai parametri di linea.

L'algoritmo di stima si basa sulla definizione dei vincoli dati dalle leggi di Kirchhoff, che consentono di scrivere le relazioni tra parametri di linea, valori misurati ed errori di misura. Poiché le leggi di Kirchhoff sono valide tra i valori effettivi, le equazioni in (3.1) devono essere riscritte per esplicitare i valori di riferimento in funzione delle misurazioni e degli errori (si veda (3.A.27) nell'Appendice 3.A per i dettagli).

L'applicazione delle due leggi di Kirchhoff, sulla caduta di tensione al ramo e sull'equilibrio delle correnti al nodo, danno luogo alle due equazioni complesse:

$$(v_i^R - v_j^R) = z_{ij} \left(i_{ij}^R - j \frac{B_{sh,ij}}{2} v_i^R \right) \quad (3.2)$$

$$(i_{ij}^R + i_{ji}^R) = \frac{B_{sh,ij}}{2} (v_i^R + v_j^R) \quad (3.3)$$

nelle quali i parametri di linea possono essere espressi in funzione dei loro valori

¹Non possono quindi essere mitigati tramite da misurazione ripetute.

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

nominali, assunti noti e disponibili nei database dei TSO, qui indicati con l'apice ⁰, e delle rispettive deviazioni rispetto ad essi:

$$\begin{aligned} z_{ij} &= R_{ij}^0 (1 + \gamma_{ij}) + jX_{ij}^0 (1 + \beta_{ij}) \\ B_{sh,ij} &= B_{sh,ij}^0 (1 + \rho_{ij}) \end{aligned} \quad (3.4)$$

dove γ_{ij} , β_{ij} e ρ_{ij} sono le deviazioni relative dei valori effettivi (veri) rispetto ai valori nominali. Queste deviazioni sono sconosciute e devono essere stimate insieme alle altre incognite. Tale rappresentazione permette, sotto ipotesi realistiche, la linearizzazione delle relazioni di misura e consente di introdurre in modo diretto ed omogeneo le informazioni a priori come di seguito discusso.

Sostituendo (3.4) e (3.A.27) in (3.2) e (3.3) e trascurando i termini del secondo ordine (vedi anche [80]), per ogni ramo si ottengono quattro equazioni a valori reali, la cui derivazione e le cui espressioni dettagliate sono riportate nell'Appendice 3.A. Da (3.A.28)-(3.A.31) si evince che ogni vincolo può essere visto come una misura equivalente che dipende: dagli errori sistematici incogniti (indicati col suffisso ^{sys}), dagli scostamenti dei parametri di linea dai corrispondenti valori nominali per il ramo considerato e dagli errori di misura casuali (indicati coll'apice ^{rnd}). Considerando un singolo ramo (i, j) e un singolo istante di misura t (associato alla marca temporale, *timetag*, delle misure della PMU), si può definire il seguente sistema di equazioni:

$$\begin{aligned} \mathbf{b}_{ij,t} &= \mathbf{H}_{ij,t} \begin{bmatrix} \xi_i^{sys} \\ \alpha_i^{sys} \\ \xi_j^{sys} \\ \alpha_j^{sys} \\ \eta_{ij}^{sys} \\ \psi_{ij}^{sys} \\ \eta_{ji}^{sys} \\ \psi_{ji}^{sys} \\ \gamma_{ij} \\ \beta_{ij} \\ \rho_{ij} \end{bmatrix} + \mathbf{E}_{ij,t} \begin{bmatrix} \xi_i^{rnd} \\ \alpha_i^{rnd} \\ \xi_j^{rnd} \\ \alpha_j^{rnd} \\ \eta_{ij}^{rnd} \\ \psi_{ij}^{rnd} \\ \eta_{ji}^{rnd} \\ \psi_{ji}^{rnd} \end{bmatrix} \\ &= \mathbf{H}_{ij,t} \mathbf{x}_{ij} + \mathbf{E}_{ij,t} \mathbf{e}_{ij,t} = \mathbf{H}_{ij,t} \mathbf{x}_{ij} + \mathbf{e}_{ij,t} \end{aligned} \quad (3.5)$$

dove $\mathbf{b}_{ij,t}$ è il vettore 4×1 di misure equivalenti al tempo t , $\mathbf{H}_{ij,t}$ è la matrice di misura che esprime le funzioni lineari di misura, mentre $\mathbf{E}_{ij,t}$ è la matrice di trasformazione (matrice Jacobiana) che collega il vettore degli errori casuali delle misure equivalenti $\mathbf{e}_{ij,t}$ al vettore di errori casuali di misura in $\mathbf{e}_{ij,t}$. \mathbf{x}_{ij} è il vettore delle incognite che devono essere stimate.

Prendendo in considerazione più rami contemporaneamente (un insieme di rami Γ di cardinalità l), il problema può essere definito considerando tutti i

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

vincoli di tensione (3.2) e le equazioni di corrente (3.3) per tutti i rami $(i, j) \in \Gamma$. Le incognite sono quindi tutti gli errori sistematici delle misure dei sincrofasori di tensione e di corrente coinvolti e tutte le deviazioni dei parametri di linea dei rami. Il sistema in (3.5) cresce fino a diventare un insieme di $4l$ equazioni con n incognite (vettore \mathbf{x}_Γ):

$$\mathbf{b}_{\Gamma,t} = \begin{bmatrix} \mathbf{b}_{(i_1,j_1),t} \\ \vdots \\ \mathbf{b}_{(i_l,j_l),t} \end{bmatrix} = \mathbf{H}_{\Gamma,t}\mathbf{x}_\Gamma + \mathbf{E}_{\Gamma,t}\mathbf{e}_{\Gamma,t} = \mathbf{H}_{ij,t}\mathbf{x}_\Gamma + \mathbf{e}_{\Gamma,t} \quad (3.6)$$

dove il pedice Γ indica che incognite, equazioni e misure, sono relative allo specifico insieme di rami Γ . È importante notare che il numero di incognite non aumenta linearmente con l (numero di rami) poiché ci sono nodi, e quindi misure, condivisi tra più rami. Questo è infatti il motivo per cui la soluzione di (3.6) non è il risultato di una semplice giustapposizione di problemi di singoli rami come (3.5), ma consente di diminuire il rapporto incognite/vincoli con la conseguenza di migliorare la stima.

Il problema di stima definito da (3.6) è tipicamente sotto-determinato. Tuttavia, come menzionato nel Paragrafo 3.1, possono essere utilizzati più istanti temporali t_1, \dots, t_{N_t} , cioè più insiemi di misure sincronizzate, e può essere definito un nuovo sistema lineare sovra-determinato come segue:

$$\begin{aligned} \mathbf{b}_\Gamma &= \begin{bmatrix} \mathbf{b}_{\Gamma,t_1} \\ \vdots \\ \mathbf{b}_{\Gamma,t_{N_t}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{\Gamma,t_1} \\ \vdots \\ \mathbf{H}_{\Gamma,t_{N_t}} \end{bmatrix} \mathbf{x}_\Gamma + \begin{bmatrix} \mathbf{E}_{\Gamma,t_1} & & \\ & \ddots & \\ & & \mathbf{E}_{\Gamma,t_{N_t}} \end{bmatrix} \begin{bmatrix} \mathbf{e}_{\Gamma,t_1} \\ \vdots \\ \mathbf{e}_{\Gamma,t_{N_t}} \end{bmatrix} \\ &= \mathbf{H}_\Gamma \mathbf{x}_\Gamma + \mathbf{E}_\Gamma \mathbf{e}_\Gamma = \mathbf{H}_\Gamma \mathbf{x}_\Gamma + \mathbf{e}_\Gamma \end{aligned} \quad (3.7)$$

dove è importante evidenziare che il vettore incognito \mathbf{x}_Γ è comune a tutti gli istanti temporali, mentre la matrice di misura e gli errori casuali cambiano con il *timestamp*. A causa dell'elevato *reporting rate* (tasso di occorrenza delle misure) delle PMU, fino a centinaia di misurazioni al secondo, è possibile classificare i diversi insiemi di misurazioni in due categorie: possono essere considerati come il risultato di osservazioni ripetute della stessa condizione di carico di rete oppure come diverse istantanee relative a diverse condizioni operative della rete. Il primo scenario sarà indicato nel seguito come “misure ripetute”, mentre il secondo sarà indicato come “casi operativi” o semplicemente “casi”, poiché riflette variazioni dei carichi, delle potenze generate, etc. Misure ripetute acquisite in ristretto intervallo di tempo (es. corrispondente a 1s o anche meno), possono essere mediate e utilizzate per definire una versione media di (3.7), con matrici e vettori mediati. Questo approccio, che oltre a minimizzare gli effetti dei contributi dell'errore casuale, consente di ridurre la dimensione del sistema, si è dimostrato valido in [83] e verrà qui adottato. Pertanto, la definizione del

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

sistema da qui in avanti si basa su casi diversi e misurazioni ripetute medie per ciascun caso.

Oltre alle informazioni fornite da diversi casi, vengono utilizzate anche le informazioni preliminari sulle incognite (informazioni a priori o *prior*). Questo consente di includere le conoscenze pregresse disponibili, ad esempio, sulla classe di accuratezza dei trasformatori di misura e sulla massima variabilità dei parametri rispetto ai valori nominali. Considerando gli l rami e i $N_\Gamma \leq 2l$ nodi in Γ , combinando C diversi casi operativi (facendo la media delle misure ripetute) e aggiungendo le informazioni a priori sulle incognite, si ottiene un sistema lineare sovra-determinato e aumentato (con i vincoli dati dalle prior):

$$\begin{aligned} \mathbf{b}_{tot} &= \begin{bmatrix} \mathbf{b}_\Gamma \\ \mathbf{0}_{n \times 1} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_\Gamma \\ \mathbf{I}_n \end{bmatrix} \mathbf{x}_\Gamma + \begin{bmatrix} \boldsymbol{\epsilon}_\Gamma \\ \mathbf{e}_{prior} \end{bmatrix} \\ &= \mathbf{H}_{tot} \mathbf{x}_\Gamma + \boldsymbol{\epsilon}_{tot} \end{aligned} \quad (3.8)$$

con $\mathbf{b}_\Gamma \in \mathbb{R}^m$, $\mathbf{H}_\Gamma \in \mathbb{R}^{m \times n}$, $\mathbf{H}_{tot} \in \mathbb{R}^{(m+n) \times n}$, $\mathbf{b}_{tot} \in \mathbb{R}^{m+n}$, dove m è il numero vincoli di Kirchhoff (considerando parte reale e immaginaria separatamente) coinvolti. Quindi, $m = 4lC$ è il numero delle incognite che include entrambi gli insiemi, $2N_\Gamma + 4l$ errori sistematici della catena di misura e $3l$ deviazioni dei parametri di linea dai rispettivi valori nominali². Le n variabili incognite possono essere stimate tramite WLS, dove la matrice dei pesi viene scelta come l'inversa della matrice di covarianza delle misure:

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{tot}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_\Gamma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{e}_{prior}} \end{bmatrix} \quad (3.9)$$

che include:

- $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_\Gamma} \in \mathbb{R}^{m \times m}$, la matrice di covarianza di tutte le misure equivalenti, definita applicando la legge di propagazione dell'incertezza [31] alla matrice $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_\Gamma}$, che esprime le incertezze delle misure delle PMU, assunte non correlate tra i vari casi.

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_\Gamma} = \mathbf{E}_\Gamma \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_\Gamma} \mathbf{E}_\Gamma^T \quad (3.10)$$

È importante sottolineare che \mathbf{E}_Γ , la matrice Jacobiana, è una matrice rettangolare a rango pieno (pari al numero delle righe, inferiore a quello delle colonne), quindi $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_\Gamma}$ è invertibile.

- $\boldsymbol{\Sigma}_{\mathbf{e}_{prior}} \in \mathbb{R}^{n \times n}$ è la matrice diagonale composta da tutte le varianze a priori delle incognite.

²nell'ipotesi che tutti i parametri di linea siano disponibili per il modello del ramo, per cui sia possibile aggiungere al modello di stima le deviazioni γ , β e ρ .

Il vettore di stato stimato $\hat{\mathbf{x}}$ è quindi la soluzione del problema WLS (si usa un pedice per ricordarlo):

$$(\mathbf{H}_{tot}^T \mathbf{W}_{tot} \mathbf{H}_{tot}) \hat{\mathbf{x}}_{WLS} = (\mathbf{H}_{tot}^T \mathbf{W}_{tot}) \mathbf{b}_{tot} \quad (3.11)$$

dove

$$\mathbf{W}_{tot} = \Sigma_{\epsilon_{tot}}^{-1} \quad (3.12)$$

3.2.2 WLS aumentato inquadrate come problema di regolarizzazione di Tikhonov

Di seguito viene mostrato come il metodo di stima possa essere inquadrate come caso particolare della regolarizzazione di Tikhonov. Questo approccio fornisce un modo per migliorare l'accuratezza della stima gestendo correttamente molteplici condizioni di carico, che possono essere comunemente osservate nelle reti elettriche, come descritto nel Paragrafo 3.3.

Il problema WLS aumentato, dato dalla (3.11), può essere trasformato in un problema ai minimi quadrati LS per mezzo della trasformazione di sbiancamento, *Whitening Transformation*³:

$$\begin{aligned} \forall \mathbf{W}_C \in \mathbb{R}^{m \times m} : \mathbf{W}_C^T \mathbf{W}_C &= \Sigma_{\epsilon_r}^{-1} \\ \mathbf{b}_w \triangleq \mathbf{W}_C \mathbf{b}_r &\implies \Sigma_{\mathbf{b}_w} = \mathbf{I}_m \end{aligned} \quad (3.13)$$

Scegliendo come matrice di sbiancamento la matrice $\mathbf{W}_C = \mathbf{U}_C^{-T}$, dove \mathbf{U}_C è la matrice triangolare superiore ottenuta attraverso la decomposizione di Cholesky della matrice di covarianza Σ_{ϵ_r} , e definendo $\mathbf{L} = \Sigma_{\epsilon_{prior}}^{-\frac{1}{2}}$ (che è quindi una matrice diagonale) si ottiene l'equivalente problema ai minimi quadrati:

$$\hat{\mathbf{x}}_{WLS} = \arg \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{W}_C \mathbf{H}_\Gamma \\ \mathbf{L} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{W}_C \mathbf{b}_r \\ \mathbf{0}_{n \times 1} \end{bmatrix} \right\|_2^2 \quad (3.14)$$

Quindi, definendo

$$\begin{aligned} \mathbf{y} &= \mathbf{Lx} \\ \mathbf{A} &= \mathbf{W}_C \mathbf{H}_\Gamma \mathbf{L}^{-1} \end{aligned} \quad (3.15)$$

ed introducendo un parametro di regolarizzazione $\mu \in \mathbb{R}^+$, il problema LS diventa:

$$\hat{\mathbf{y}}_\mu = \arg \min_{\mathbf{y}} \left\| \begin{bmatrix} \mathbf{A} \\ \sqrt{\mu} \mathbf{I}_n \end{bmatrix} \mathbf{y} - \begin{bmatrix} \mathbf{b}_w \\ \mathbf{0}_{n \times 1} \end{bmatrix} \right\|_2^2 \quad (3.16)$$

che è equivalente alla regolarizzazione di Tikhonov (vedi anche 1.5.4) in forma

³Per un approfondimento si rimanda alla Sezione 1.6.2.

standard:

$$\hat{\mathbf{y}}_\mu = \arg \min_{\mathbf{y}} \|\mathbf{A}\mathbf{y} - \mathbf{b}_w\|_2^2 + \mu \|\mathbf{y}\|_2^2 \quad (3.17)$$

come è immediato verificare applicando le equazioni normali alla (3.16). Una volta stimato $\hat{\mathbf{y}}_\mu$, lo stato $\hat{\mathbf{x}}$, che include le incognite di interesse, può essere ricostruito con un semplice ridimensionamento (invertendo la prima equazione in (3.15))⁴.

Infine, si sottolinea che è possibile inquadrare il problema WLS aumentato (3.11) come un caso particolare del problema di regolarizzazione di Tikhonov in cui il parametro di regolarizzazione μ è impostato a 1 e

$$\hat{\mathbf{x}}_{\text{WLS}} = \mathbf{L}^{-1} \hat{\mathbf{y}}_{\mu=1} \quad (3.18)$$

ritrovando quindi la soluzione WLS da $\hat{\mathbf{y}}_{\mu=1}$ come appena affermato.

3.3 Gestione di casi operativi multipli

Da un lato, [81] sottolinea la necessità di considerare più misurazioni in diverse condizioni operative (casi) e [82] evidenzia l'effetto positivo sui risultati della stima dei parametri di linea dell'utilizzo di un numero maggiore di vincoli. D'altro canto, le PMU consentono di acquisire diverse misurazioni al secondo. Lo scopo di questa sezione è mostrare come sia possibile ottimizzare la gestione di più casi all'interno del processo di stima, sfruttando così l'alto *reporting rate* fornito dalle PMU.

3.3.1 Analisi dell'incertezza della stima

Le equazioni (3.16) e (3.17) affrontano lo stesso problema. Mentre la prima è formulata come un LS aumentato, la seconda è una formulazione con penalizzazione a due termini⁵:

- $\|\mathbf{A}\mathbf{y} - \mathbf{b}_w\|_2^2$ è la norma al quadrato del vettore residuo ed è un indicatore generale dell'accordo tra il vettore soluzione \mathbf{y} e il vettore delle misure equivalenti \mathbf{b}_w ⁶.

⁴Introducendo il parametro di regolarizzazione μ nella (3.14) e definendo $\mathbf{A}_G = \mathbf{W}_C \mathbf{H}_G$ è possibile stimare direttamente il vettore incognito \mathbf{x} risolvendo il problema Tikhonov generalizzato $\hat{\mathbf{x}}_\mu = \arg \min_{\mathbf{x}} \|\mathbf{A}_G \mathbf{x} - \mathbf{b}_w\|_2^2 + \mu \|\mathbf{L}\mathbf{x}\|_2^2$ visto nella Sezione 1.5.5, affrontabile con la decomposizione GSVD, le cui equazioni si trovano in (1.5.1). Tuttavia, per le considerazioni inerenti la complessità computazionale, quando possibile, è preferibile riportare il problema generalizzato di Tikhonov alla forma standard, come discusso nella Sezione 1.5.5.

⁵Per un approfondimento sul significato dei due termini all'interno della formulazione, e per una interpretazione statistica, si rimanda al Paragrafo 1.3.

⁶Termine legato alla verosimiglianza, *likelihood* come visto nel Paragrafo 1.3.

- $\|\mathbf{y}\|_2^2$ è il quadrato della norma della soluzione, che valuta l'“energia” del vettore soluzione \mathbf{y} , cioè del vettore di stato \mathbf{x} pesato con le informazioni a priori.

In entrambe le formulazioni è possibile impostare il parametro di regolarizzazione μ in modo da sfruttare le informazioni a priori sul problema da risolvere. Applicando le equazioni normali all'equazione (3.16), segue:

$$(\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I}_n) \hat{\mathbf{y}}_\mu = \mathbf{A}^\top \mathbf{b}_w \quad (3.19)$$

e grazie alla decomposizione SVD⁷ della matrice $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, è possibile esprimere la soluzione analitica del problema di stima al variare del parametro di regolarizzazione μ :

$$\begin{aligned} \hat{\mathbf{y}}_\mu &= \mathbf{V} (\mathbf{\Sigma}^\top \mathbf{\Sigma} + \mu \mathbf{I}_n)^{-1} \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{b}_w \\ &= \sum_{\sigma_j > 0} \Phi_j \frac{\mathbf{u}_j^\top \mathbf{b}_w}{\sigma_j} \mathbf{v}_j = \sum_{\sigma_j > 0} g_j \mathbf{v}_j = \hat{\mathbf{y}}_{\text{Tikhonov}} \end{aligned} \quad (3.20)$$

dove

$$\Phi_j = \frac{\sigma_j^2}{\sigma_j^2 + \mu} \quad (3.21)$$

mentre \mathbf{u}_j e \mathbf{v}_j sono il j -esimo vettore colonna in \mathbf{U} e \mathbf{V} , rispettivamente. Il vettore stimato $\hat{\mathbf{y}}_\mu$ è ottenuto come combinazione lineare dei vettori \mathbf{v}_j , i termini $\mathbf{u}_j^\top \mathbf{b}_w$ sono anche chiamati coefficienti di Fourier e Φ_j è un filtro passa basso (“low-pass” filter) che abbatte le componenti corrispondenti con $\sigma_j^2 \ll \mu$. All'interno di questo quadro, è ora possibile studiare il comportamento dell'algoritmo di stima quando sono disponibili diversi casi e l'effetto dovuto alla scelta di diversi valori per il parametro μ .

Per eseguire tale analisi sono stati effettuati test su una parte della rete denominata IEEE 14 bus system [84] mostrata in Figura 3.2 e, in particolare, sulla porzione di rete presa in considerazione in [80], quindi sui rami da 1 a 6 e sui nodi corrispondenti, utilizzando l'approccio dei rami multipli e le stesse ipotesi che sono riportate in dettaglio nel seguente Paragrafo 3.4. I test sono stati eseguiti considerando diversi scenari con diverso numero di casi. Per valutare le prestazioni dell'algoritmo di stima, viene utilizzato l'indicatore *Relative Root Square Error* (RRSE):

$$\text{RRSE} = \frac{\|\hat{\mathbf{y}}_\mu - \mathbf{y}\|_2}{\|\mathbf{y}\|_2} \quad (3.22)$$

La Figura 3.3 mostra l'andamento in doppia scala logaritmica del RRSE in funzione di $\mu \in [0, \dots, 1000]$ in tre diversi scenari, utilizzando $C \in \{10, 50, 500\}$ diversi casi operativi rispettivamente. È importante sottolineare che in [80] è stato preso in considerazione solo lo scenario con $C = 10$. I valori RRSE sono

⁷ Si veda Sezione 1.5.1.

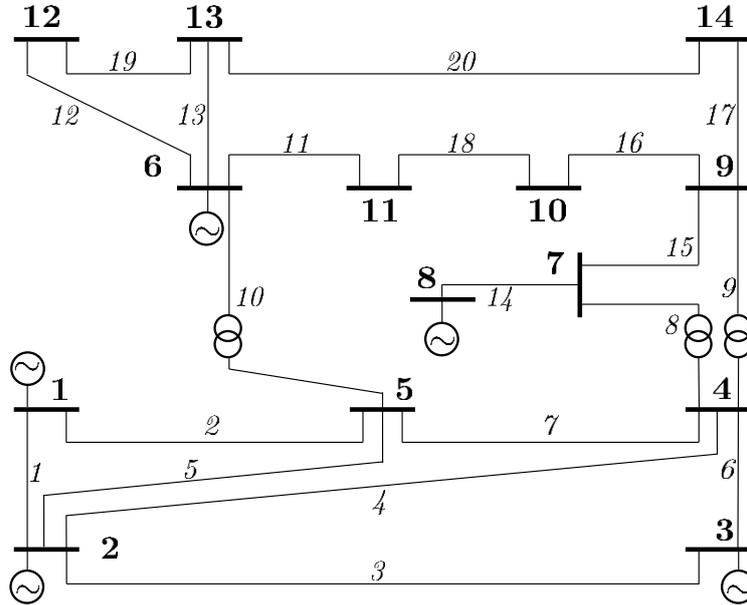


Figura 3.2: Rete di test IEEE 14 bus system

mediati su $N_{MC} = 5000$ prove Monte Carlo (MC) corrispondenti a diverse estrazioni sia dei parametri di linea sia degli errori sistematici nella catena di misura. Dalla figura appare chiaro come la stima sia del tutto inattendibile quando non vengono utilizzate informazioni a priori, ovvero quando $\mu = 0$. Per μ positivi i valori di RRSE scendono rapidamente fino a raggiungere una zona stabile, quindi un minimo, diverso nei tre scenari considerati; per un numero maggiore di casi il minimo è più basso e spostato verso destra, corrispondente quindi ad un valore μ più alto (μ_{min} nel seguito). La zona stabile (quasi piatta) della curva attorno al minimo indica che l'algoritmo è abbastanza robusto rispetto alle informazioni a priori sulle incognite, come mostrato in [85]. L'inaffidabilità dell'algoritmo di stima per μ molto bassi evidenzia come il problema di stima inizi come sotto-determinato, diventi sovra-determinato sfruttando diversi casi operativi (si veda, ad esempio, [81]) ma non sia ancora utilizzabile, anche per C maggiori, senza l'utilizzo del termine di regolarizzazione, ovvero le *prior*, adeguatamente pesate dal parametro μ . La linea tratteggiata verticale, etichettata con $\mu = 1$, indica il valore medio di RRSE ottenuto con l'approccio WLS. È possibile osservare che, quando si sfruttano 10 casi come in [80], $\mu = 1$ è molto vicino a μ_{min} ma, quando si sfruttano più casi è necessario utilizzare un μ maggiore. Infatti, se si considera uno scenario con 500 differenti casi operativi, il valore di RRSE ottenuto tramite WLS è superiore al 100%, quindi l'incertezza della stima è superiore rispetto a quella iniziale, nota a priori, indicata nella

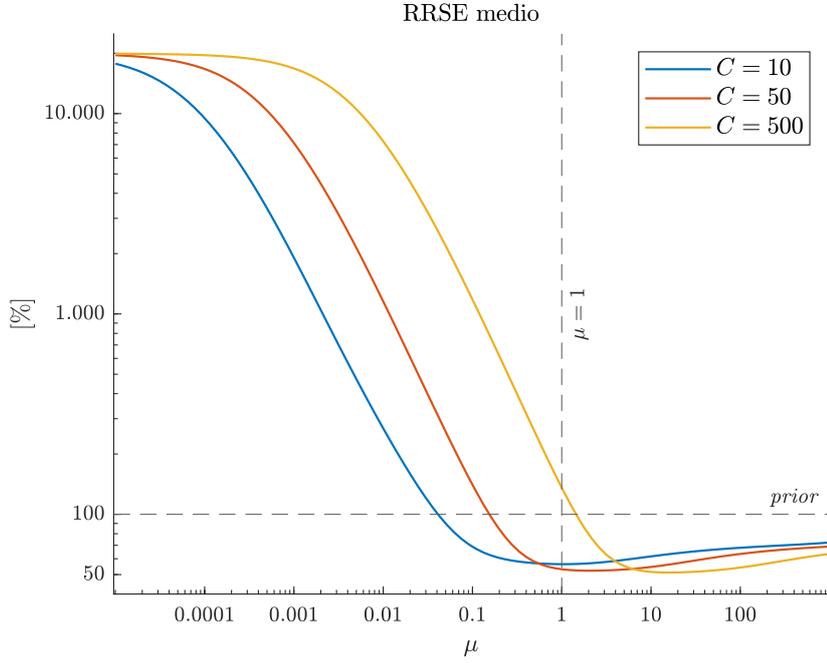


Figura 3.3: IEEE 14 - 6 rami, RRSE medio in funzione di μ in tre diversi scenari: $C \in \{10, 50, 500\}$

Figura 3.3 con la linea tratteggiata orizzontale e l’etichetta “*prior*”.

Focalizzando l’attenzione sulle cause di questo comportamento, l’ipotesi *naive* porterebbe a pensare al numero di condizionamento della matrice, tuttavia, come mostra la Figura 3.4 e come affermato in [81], un numero maggiore di casi porta a migliorare, seppur leggermente, il condizionamento della matrice, indicato con $\kappa(\mathbf{A})$ in figura. Tuttavia, il profilo dei valori singolari si sposta significativamente verso valori più alti, quindi è necessario un μ maggiore per consentire a (3.21) di agire come filtro passa basso, necessario per soddisfare la condizione discreta di Picard [28], [24]. La condizione di Picard afferma che i coefficienti di Fourier in (3.20) dovrebbero diminuire più velocemente dei corrispondenti valori singolari σ_j per avere una soluzione di (3.16) limitata. A tal proposito, è quindi possibile utilizzare questo paradigma per una stima più accurata, come spiegato nel seguito.

3.3.2 Metodo proposto e regolarizzazione

L’approccio proposto consiste nell’utilizzare un valore di μ che sia sintonizzato sulle esigenze del problema in esame e che dipenda dai suoi parametri. Test condotti con diverse configurazioni hanno mostrato che i valori di σ_j aumentano

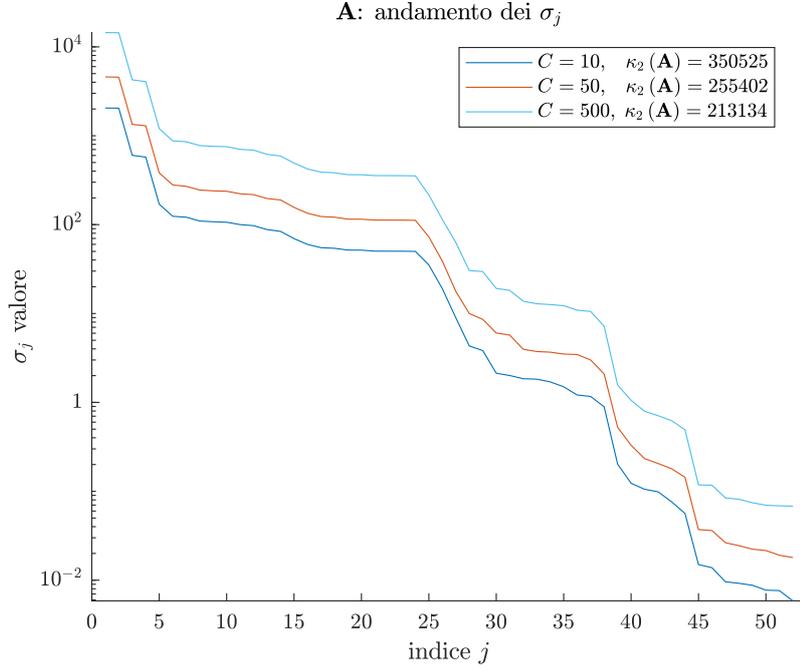


Figura 3.4: IEEE 14 - 6 rami, profilo dei valori singolari σ_j e del numero di condizionamento $\kappa_2(\mathbf{A})$ della matrice \mathbf{A} , in tre diversi scenari: $C \in \{10, 50, 500\}$

con la radice quadrata del numero di vincoli incorporati nella matrice di misura \mathbf{A} ⁸. Per questo motivo il parametro di regolarizzazione è stato scelto come:

$$\mu_{sqr} = \sqrt{\frac{m}{n}} \quad (3.23)$$

per adattarlo all'andamento dei valori singolari al variare del numero di vincoli, e quindi di righe, utilizzati nella matrice di misura. La Figura 3.5 mostra i grafici delle condizioni di Picard relativi alla soluzione in assenza di errore di misura (cerchi rossi), mentre gli asterischi viola, blu e verde rappresentano i coefficienti g_j ottenuti rispettivamente senza utilizzare le prior ($\mu = 0$), con WLS ($\mu = 1$) e con la soluzione proposta μ_{sqr} . È possibile osservare come il parametro di regolarizzazione proposto μ sia più efficace nel limitare i coefficienti.

Il metodo proposto è stato confrontato con i classici algoritmi utilizzati per la scelta del parametro di regolarizzazione nel problema di Tikhonov⁹. Il criterio basato sulla *L-curve* (LCC), la Cross Validation Generalizzata (GCV) e il criterio basato sul *Discrepancy Principle* (DP) [28], [29]. La *L-curve* confronta

⁸In statistica e nel ML è pratica comune normalizzare i valori singolari di una matrice dividendo la stessa per la radice quadrata del numero di campioni. In questo modo è possibile confrontare dataset di diverse dimensioni [86].

⁹Per maggiori dettagli si rimanda alla Sezione 1.5.6.

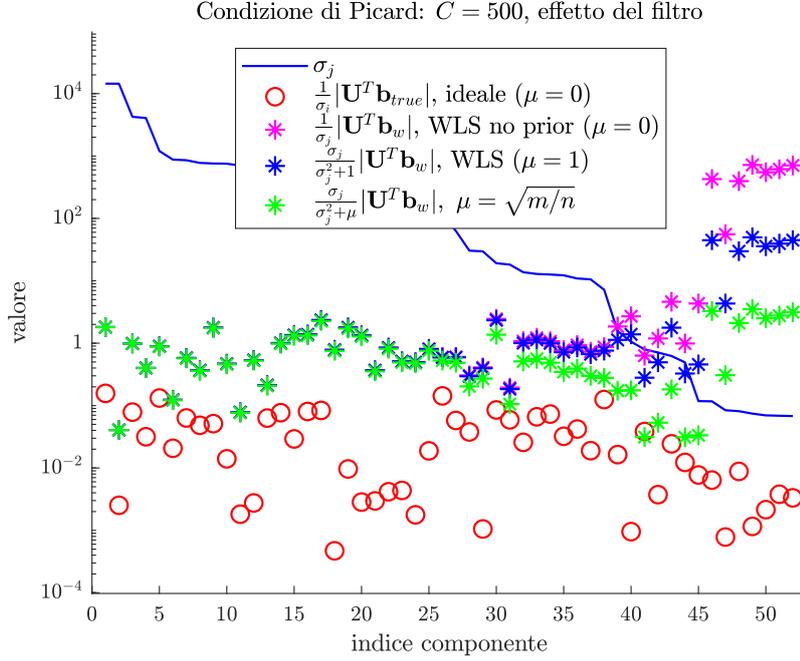


Figura 3.5: IEEE 14 - 6 rami, effetto del filtro.

in doppia scala logaritmica la norma del termine di regolarizzazione rispetto alla relativa norma del termine residuo, per tutti i μ considerati, ed assume l'aspetto di una L, da cui il nome, come da Figura 1.7 nel Paragrafo 1.5 sulla regolarizzazione in ℓ_2 . Il criterio LCC individua il miglior compromesso tra i due termini nell'angolo della curva, e seleziona il corrispondente μ come parametro di regolarizzazione. Sia per LCC che per GCV sono state utilizzate le implementazioni in [28].

Il criterio basato sul DP di Morozov [20] è l'unico metodo basato sulla conoscenza della norma dell'errore, o sulla disponibilità di una sua buona stima, e suggerisce di scegliere il valore μ in modo che la norma residua (errore di approssimazione) sia vicina alla norma dell'errore di misura. Quindi:

$$\mu_{\text{DP}} : \|\mathbf{b}_w - \mathbf{A}\hat{\mathbf{y}}_{\mu_{\text{DP}}}\|_2 \approx \tau \|\boldsymbol{\epsilon}_w\|_2, \tau \geq 1 \quad (3.24)$$

dove $\boldsymbol{\epsilon}_w = \mathbf{W}_C \boldsymbol{\epsilon}_\Gamma$ è il vettore di errore di misura sbiancato, avente la matrice identità come matrice di covarianza. Per ottenere una adeguata stima della norma dell'errore di misura, è stato utilizzato il metodo illustrato in [32]¹⁰ adattato al problema specifico. Per affrontare il problema (3.24) è stato seguito il metodo proposto in [30] e descritto nella Sezione 1.5.7. Inoltre, al fine di avere

¹⁰La procedura è stata presentata in modo generalizzato nel Paragrafo 1.6 e ripresa nel Paragrafo 2.3.

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

un punto di riferimento, *benchmark*, per le prestazioni dell'algoritmo di stima nell'ambito del problema di regolarizzazione di Tikhonov, è stato definito il μ ottimale per ogni prova MC:

$$\mu_{opt} = \arg \min_{\mu} \text{RRSE} \quad (3.25)$$

ed utilizzato nelle prove. Come si evince dalla definizione, μ_{opt} è utilizzato solo come riferimento perché non è disponibile nella pratica dato che il suo calcolo richiede una conoscenza preventiva di \mathbf{y} (valori veri).

La Figura 3.6 mostra gli andamenti medi dei RRSE quando C varia da 10 a 500. I valori RRSE medi del WLS, corrispondente a $\mu = 1$, all'inizio migliorano leggermente, raggiungendo un minimo con circa 50 casi, poi aumentano diventando presto inaffidabili per un numero elevato di casi (si veda anche Figura 3.3). I metodi di regolarizzazione classici (LCC, GCV, DP) non sono in grado di raggiungere buone prestazioni di stima in quanto il problema di fondo non è dovuto a problemi di condizionamento e, di conseguenza, i parametri di regolarizzazione selezionati mediante questi criteri introducono un effetto di filtro passa basso troppo forte. Infine, la figura mostra come il metodo di selezione del parametro proposto μ_{sqr} in (3.23) ottiene prestazioni prossime alle prestazioni del benchmark μ_{opt} . La Figura 3.7 mostra uno schema a blocchi del metodo

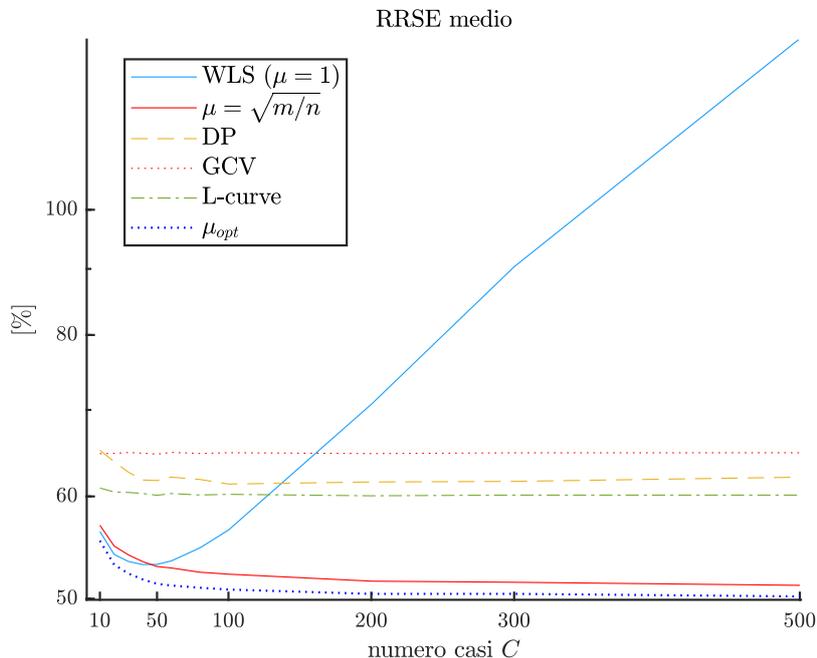


Figura 3.6: IEEE 14 - 6 rami: RRSE medio (5000 prove MC realizzate) con un numero di condizioni operative C variabile da 10 sino a 500.

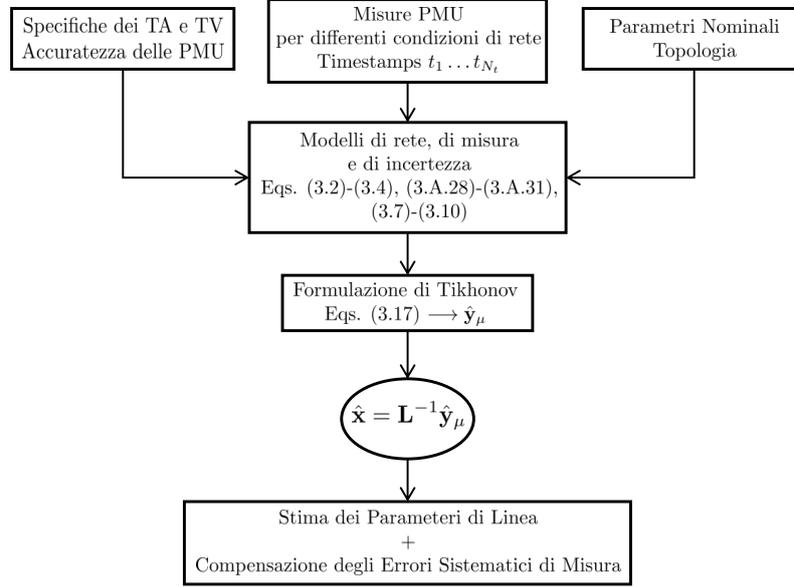


Figura 3.7: Flow chart della procedura proposta.

proposto.

Infine, vengono riportate alcune osservazioni sul peso computazionale del metodo. Ci sono due punti importanti da considerare, ovvero la costruzione della matrice di sbiancamento \mathbf{W}_{Ch} e la decomposizione SVD della matrice \mathbf{A} . Per ridurre il costo computazionale della costruzione della matrice di sbiancamento, sono state sfruttate le proprietà della decomposizione di Cholesky, che è implementabile a blocchi, e delle matrici triangolari che costituiscono un'algebra, in particolare, l'inversa di una matrice triangolare superiore (inferiore) è ancora una matrice triangolare superiore (inferiore). La matrice di sbiancamento quindi, è costruita a blocchi, cioè eseguendo la fattorizzazione di Cholesky delle piccole sotto-matrici di $\mathbf{\Sigma}_{\epsilon_r}$ corrispondenti a un caso singolo, invertendo le piccole matrici triangolari calcolate, e infine integrando i blocchi. Per quanto riguarda la SVD, anche se il numero delle righe della matrice \mathbf{A} aumentano con C , \mathbf{A} diventa una matrice molto sottile ($n \ll m$) e la *economy-size* SVD (si veda la Sezione 1.5.1) consente un calcolo veloce.

3.4 Prove e Risultati

In questa Sezione sono riportati i test di valutazione delle prestazioni per determinare la validità dell'approccio proposto. Tutti i test sono stati eseguiti

utilizzando la rete a 14 nodi IEEE 14 bus system [84], riportata in Figura 3.2. I metodi di regolarizzazione classici sono stati confrontati con quello proposto (come discusso nel Paragrafo 3.3.2) e viene utilizzato un diverso numero di rami coinvolti in diverse prove. Sono stati eseguiti test su diversi scenari, caratterizzati da $C \in \{10, 20, \dots, 1000\}$ casi (come accennato in precedenza, per tenere conto delle diverse condizioni operative di rete) e $M = 10$ misurazioni ripetute per ciascun caso. Le $N_{MC} = 5000$ prove MC eseguite sono state utilizzate per convalidare statisticamente i risultati. Per ogni prova MC sono stati simulati diversi errori sistematici. Gli errori sistematici sono attribuiti ai trasformatori di misura, quindi definiti in base alla loro classe di accuratezza, mentre gli errori casuali sono considerati associati alle PMU ed estratti di conseguenza. Per ogni istante di misura e per ogni caso, vengono estratti errori casuali e aggiunti, insieme agli errori sistematici della prova considerata, alle tensioni e alle correnti di riferimento derivanti da una procedura di *power flow*, per ottenere le misure di ampiezza e di fase del sincrofasore (di tensione e di corrente). Nelle prove vengono utilizzate le seguenti ipotesi, al fine di avere una configurazione realistica:

1. La massima deviazione dei parametri di linea R_{ij} , X_{ij} e $B_{sh,ij}$ è stata assunta pari $\pm 15\%$.
2. I trasformatori di misura, TA e TV, sono stati assunti di classe 0.5, per cui si è usato 0.5% per il massimo errore di rapporto, sia della tensione sia della corrente, 0.9 grad per il massimo errore di fase dei TA e 0.6 grad per il massimo errore di fase dei TV. Nel seguito, per ogni test, gli errori sono stati estratti da distribuzioni di tipo uniforme per definire gli errori sistematici effettivi per ogni prova MC.
3. Per gli errori associati alle PMU, sono stati utilizzati un errore massimo di ampiezza di 0.1% e un massimo errore di fase di 0.1 grad.
4. La variabilità dei carichi e dei generatori tra i diversi casi operativi è stata ipotizzata di $\pm 10\%$ rispetto ai valori nominali, sia per la potenza attiva che per la potenza reattiva.

Per valutare le prestazioni sia della stima dei parametri di linea sia della compensazione degli errori sistematici della catena di misura, viene utilizzato l'errore quadratico medio (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_{MC}} (\hat{\nu} - \nu)^2}{N_{MC}}} \quad (3.26)$$

dove $\hat{\nu}$ indica la quantità stimata. In (3.26) ν rappresenta una generica incognita in \mathbf{x} e può quindi essere (considerando un generico ramo (i, j)) uguale a ξ_i^{sys} ,

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

$\alpha_i^{sys}, \xi_j^{sys}, \alpha_j^{sys}, \eta_{ij}^{sys}, \psi_{ij}^{sys}, \eta_{ji}^{sys}, \psi_{ji}^{sys}, \gamma_{ij}, \beta_{ij}$ o ρ_{ij} . Nel seguito, per semplicità e quando non è necessario specificare altrimenti, il ramo sarà indicato con un unico indice k come in Figura 3.2 e quindi i parametri di linea associati saranno $\gamma_k, \beta_k, \rho_k$.

3.4.1 Stima su un insieme di rami selezionati

La prima serie di test è condotta su una parte della rete IEEE 14, cioè sui nodi di alta tensione da 1 a 5 in Figura 3.2, considerando l'approccio multi-ramo. I test sono stati eseguiti utilizzando tutti i metodi presentati nella Sezione 3.3 ma, secondo l'analisi riassunta nella Figura 3.6, da qui in avanti i risultati sono principalmente focalizzati sul confronto delle prestazioni del WLS aumentato ($\mu = 1$, [80]) con il metodo proposto (metodo μ_{sqr}). In ogni prova è stato utilizzato anche il risultato ottenuto con il valore ideale μ_{opt} in maniera da avere un limite superiore alle prestazioni globali (ovvero considerando l'intero vettore di stato, pesato con le prior, \mathbf{y}) ottenibili all'interno del paradigma della regolarizzazione di Tikhonov. Entrambi i risultati, WLS e metodo μ_{sqr} , vengono presentati scegliendo lo scenario migliore per ciascun metodo, ovvero il valore di C che consente le migliori prestazioni (C_{best})¹¹. Inoltre, viene presentato anche lo scenario base $C = 10$ per il metodo WLS, questo per avere un chiaro confronto con i risultati della stima multi-ramo presentati in [80] dove questo era l'unico scenario preso in considerazione. La Tabella 3.1 riporta gli scenari considerati e i corrispondenti risultati RRSE mediati su 5000 prove.

Nel seguito, tutti i risultati di stima presentati in termini di RMSE per una specifica incognita devono essere confrontati con la corrispondente deviazione standard a priori che ne costituisce l'iniziale incertezza, incertezza alla quale si dovrebbe fare riferimento in assenza di un processo di stima. Considerando gli errori sistematici di misura, la deviazione standard a priori è $\Delta_\xi/\sqrt{3} = \Delta_\eta/\sqrt{3} \simeq 0.29\%$ per l'errore di ampiezza delle misure di tensione e di corrente (Δ indica la massima deviazione), $\Delta_\alpha/\sqrt{3} \simeq 0.35$ crad e $\Delta_\psi/\sqrt{3} \simeq 0.52$ crad rispettivamente per l'errore di fase delle misure di tensione e corrente. Considerando i parametri di rete, la deviazione standard dei relativi errori è $\Delta_\gamma/\sqrt{3} = \Delta_\beta/\sqrt{3} = \Delta_\rho/\sqrt{3} \simeq 8.66\%$.

In Figura 3.8 sono riportati i risultati degli RMSE per la stima delle resistenze di ramo, considerando gli scenari della Tabella 3.1. Si può notare come i risultati ottenuti secondo [80] (rappresentati da simboli "più" verdi) possono essere migliorati mediamente di circa 32,2% quando si utilizza la stessa configurazione WLS ma sfruttando 40 casi (più blu). Tuttavia, la migliore configurazione WLS è superata dal metodo di regolarizzazione proposto (asterischi rossi),

¹¹I risultati ideali con μ_{opt} sono presentati usando lo stesso C di μ_{sqr} .

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

con un ulteriore miglioramento medio del 25.5%. Va anche detto che, in questo caso, il metodo μ_{sqr} è molto vicino ai risultati ottenuti con μ_{opt} . È importante evidenziare che l'indicatore RRSE, minimizzando il quale μ_{opt} è definito, è un indice riassuntivo che considera complessivamente tutti gli errori sui parametri incogniti (vedi (3.22) e (3.25)). Per questa ragione, gli RMSE ottenuti con μ_{opt} , solo per alcune quantità specifiche e/o alcuni rami, potrebbero non essere i più bassi. Questo è il caso della Figura 3.8, dove il metodo μ_{sqr} è anche leggermente migliore per γ , ma per altre incognite (vedi i risultati relativi a ξ nel seguito) μ_{opt} è più accurato.

Il quadro complessivo della stima dei parametri di linea è mostrato in Figura 3.9 dove vengono considerati tutti i tipi di parametri. È chiaro che il metodo proposto (barre in primo piano con bordo a linea continua) migliora significativamente i migliori risultati WLS (barre sullo sfondo con bordo tratteggiato) per tutti i parametri di linea e per tutti i rami. I miglioramenti medi (riduzione dell'errore percentuale) sono 34.9% e 57.2% rispettivamente per β_k e ρ_k .

Tabella 3.1: IEEE 14 - 6 rami: migliori RRSE medi (su 5000 prove MC) per metodi differenti

Metodo	RRSE medio [%]
WLS ($\mu = 1$), $C = 10$ [80]	56.35
WLS ($\mu = 1$), $C_{best} = 40$	53.14
$\mu = \sqrt{m/n}$, $C_{best} = 500$	51.21
μ_{opt} , $C = 500$	50.20

I risultati RMSE per la stima di ξ , che esprimono le prestazioni di compensazione dell'ampiezza della tensione, sono mostrati in Figura 3.10, in funzione del numero di casi C , per tutti i metodi di regolarizzazione presentati nel Paragrafo 3.3. È possibile evidenziare che i risultati di WLS (più verdi) e, per più di 50 casi, DP (triangoli blu) tendono a divergere in modo significativo, andando anche oltre le incertezze a priori (linea nera tratteggiata). In particolare, il metodo WLS può portare a risultati inaffidabili all'aumentare del numero di casi, confermando i risultati precedentemente mostrati nelle Figure 3.3 e 3.6. Le considerazioni fatte per l'approccio multi-ramo sono simili a quelle che si possono derivare per il caso mono-ramo (vedi [80]). La Tabella 3.2 riporta, ad esempio, un confronto dei parametri RMSE per ξ_1 quando vengono utilizzati sia $\mu = 1$ che $\mu = \sqrt{m/n}$ con $C \in \{10, 500\}$. Pur confermando il significativo miglioramento del metodo multi-ramo (vedi Figura 4 in [80]), la Tabella 3.2 mostra che la regolarizzazione introduce evidenti benefici quando è possibile sfruttare un numero elevato di casi operativi (C è alto) anche per la stima su singolo ramo e

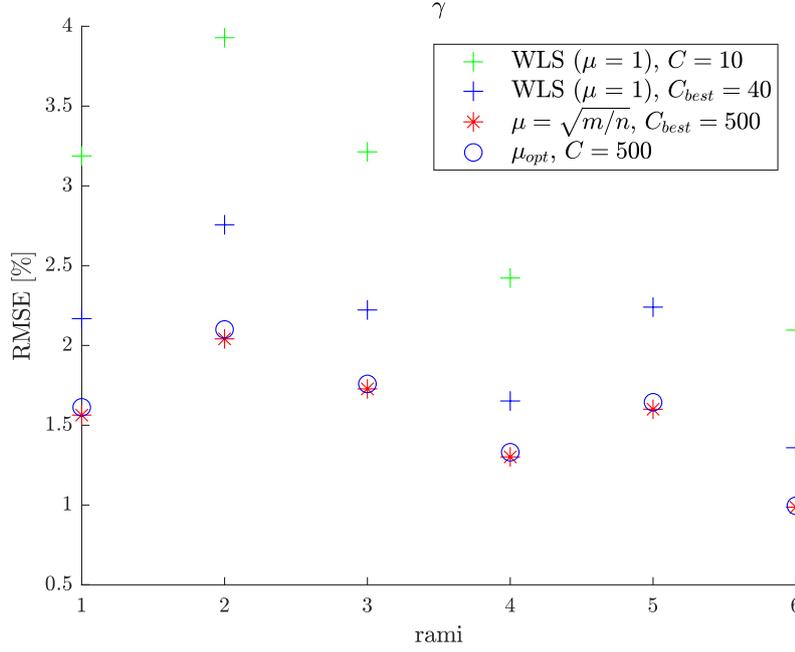


Figura 3.8: IEEE 14 - 6 rami: risultati di stima per il parametro γ di ognuno dei 6 rami.

previene condizioni critiche per la ξ come quelle già individuate in Figura 3.10.

Tabella 3.2: IEEE 14 - Nodo 1: RMSE (5000 prove MC) per ξ_1 - Confronto tra gli approcci Multi-Ramo e Ramo-Singolo

Metodo	C	RMSE [%]	
		Multi-Ramo	Ramo-Singolo
$\mu = \sqrt{m/n}$	10	0.16	0.22
$\mu = \sqrt{m/n}$	500	0.16	0.21
WLS ($\mu = 1$)	10	0.16	0.22
WLS ($\mu = 1$)	500	1.01	0.46

3.4.2 Stima simultanea sull'intera rete

Ulteriori analisi sono state effettuate sull'intera rete di test IEEE 14 nodi, utilizzando l'approccio multi-ramo. Sono stati simulati diversi scenari considerando un numero variabile di casi ($C \in \{10, 20, \dots, 1000\}$) e confrontando i metodi WLS e μ_{sqr} . Per entrambi i metodi è stato selezionato lo scenario migliore come fatto nei test precedenti. C_{best} dipende dalla configurazione e, in particolare,

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

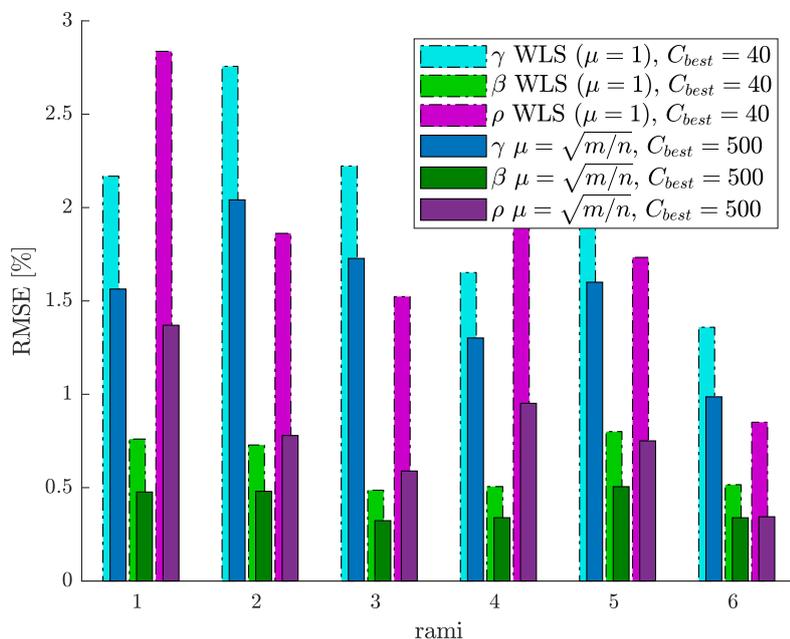


Figura 3.9: IEEE 14 - 6 rami: risultati di stima per tutti i parametri di linea.

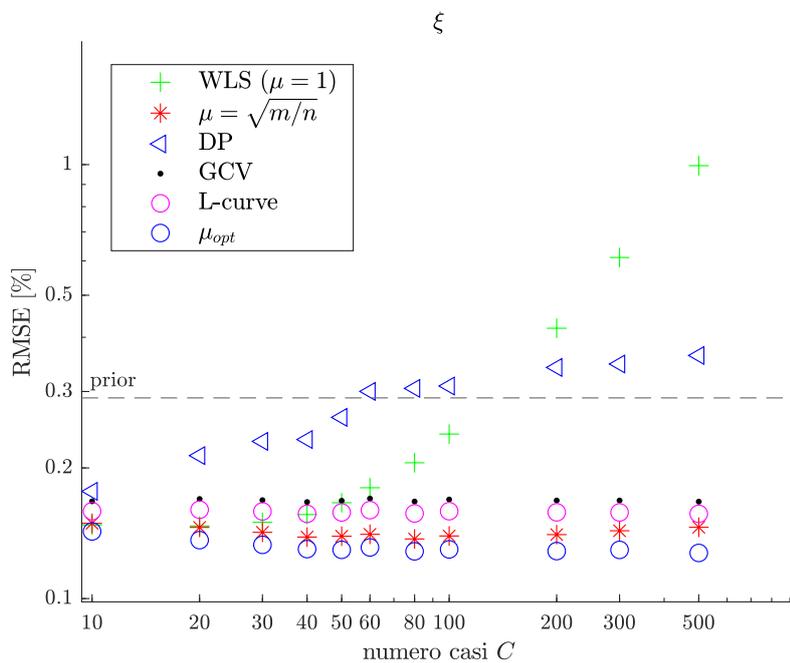


Figura 3.10: IEEE 14 - 6 rami: risultati di stima per ξ in funzione del numero di casi C per tutti i metodi presi in considerazione.

ripetendo sull'intera rete l'analisi e l'ottimizzazione descritta nel Paragrafo 3.3, si ottiene $C_{best} = 30$ per WLS e $C_{best} = 200$ per il metodo μ_{sqrt} . La Figura 3.11 mostra gli RMSE per le stime β_k (concentrandosi su $k = \{1, \dots, 6\}$, ovvero la reattanza dei rami da 1 a 6 in Figura 3.2) valutata con i due metodi e due diverse matrici di misura multi-ramo: la prima si ottiene utilizzando solo i vincoli dovuti ai primi 6 rami ($l = 6$), mentre la seconda integra i vincoli dell'intera rete ($l = 20$). I risultati ottenuti in [80] utilizzando il WLS con $C = 10$ casi sono riportati con i segni più verdi. La figura mostra i vantaggi di sfruttare in modo adeguato un maggior numero di vincoli, dovuti sia ad un maggior numero di casi sia ad un insieme più vasto di rami monitorati contemporaneamente. È inoltre possibile evidenziare che il miglioramento medio, che va dal WLS applicato in [80] ai migliori risultati ottenuti con $\mu = \sqrt{m/n}$, $C_{best} = 200$ e sfruttando i vincoli dell'intera rete (asterischi blu), è di circa 66.6%.

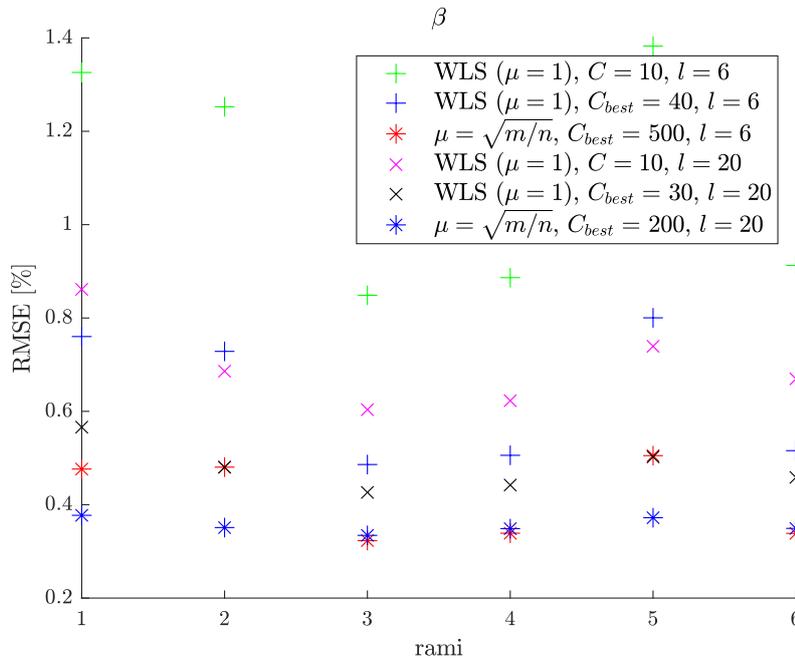


Figura 3.11: Risultati di stima di β utilizzando l'approccio multi-ramo, applicato ad una porzione di rete e all'intera rete, utilizzando diversi metodi e configurazioni.

La Figura 3.12 mostra i risultati RMSE per tutti i parametri di linea, ottenuti monitorando l'intera rete, utilizzando C_{best} per il WLS (figura in alto) e il metodo μ_{sqrt} (figura in basso). È importante sottolineare che nei test vengono analizzati tutti i rami e quindi vengono presi in considerazione anche i rami che possono essere modellati utilizzando la sola reattanza di linea o trascurando la

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

suscettanza in derivazione (alcuni parametri non sono quindi presenti in figura). La figura evidenzia come una stima significativamente migliore può essere ottenuta con il metodo di regolarizzazione proposto. Infatti le stime di γ_k , β_k and ρ_k hanno un miglioramento medio del 16.0 %, 23.9 % e 44.4 %, rispettivamente.

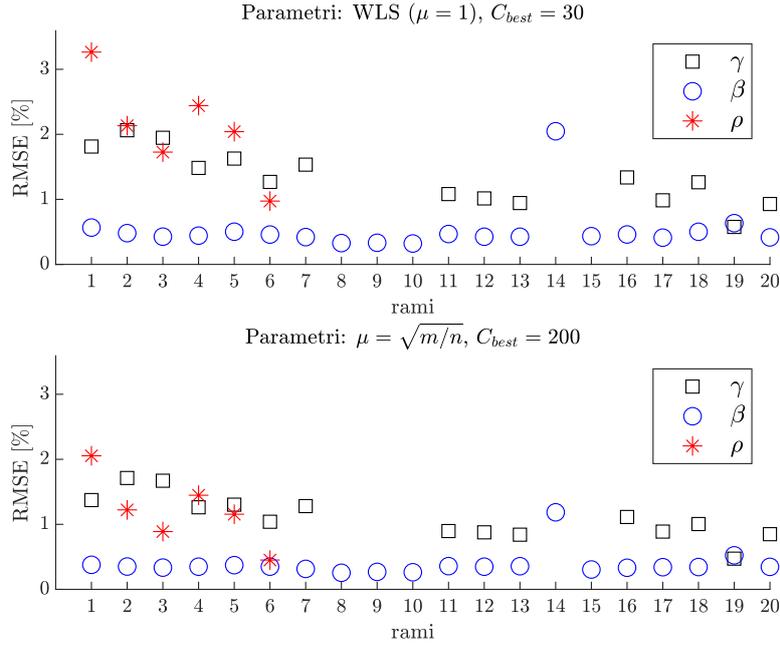


Figura 3.12: IEEE 14 - intera rete: risultati di stima per tutti i parametri di linea (deviazioni).

Infine, benché non sia possibile un confronto omnicomprensivo con altri algoritmi della letteratura, un esempio dei risultati RMSE ottenibili con metodi diversi utilizzando le stesse misurazioni è riportato nella Tabella 3.3. Il metodo proposto, riassunto in Figura 3.7, è stato confrontato con due metodi. In particolare, il primo è un metodo di stima diretta (Metodo A) per i parametri di linea basato sulle equazioni in [66] (le stesse equazioni si possono trovare in [67]). Per un confronto equo, le stime ottenute per un singolo caso sono state mediate e le misurazioni ripetute sono state quindi valutate come nel metodo proposto. Il secondo è il metodo presentato in [78, Sec. IV] (Metodo B) che si propone di stimare i parametri di una linea e i fattori di compensazione per i trasformatori di misura al nodo di arrivo. Sono stati utilizzati due diversi scenari operativi, che sfruttano quindi un numero diverso C di casi operativi: $C = 10$ è il valore base mentre $C = 200$ corrisponde ai test precedentemente illustrati in questa sezione. Per il confronto viene scelto il ramo 1 della rete, ma risultati simili si

possono trovare sugli altri rami. La tabella 3.3 riporta i risultati RMSE per la resistenza, reattanza e per gli errori sistematici delle misure del nodo di arrivo. Il Metodo A non consente di stimare gli errori sistematici e il Metodo B non consente il calcolo degli errori sistematici al nodo di partenza del ramo, quindi, anche per il metodo proposto non vengono riportati gli RMSE corrispondenti. Per valutare anche la robustezza del metodo, l'accuratezza della PMU è stata variata rispetto ai test precedenti ipotizzando anche un altro scenario di misura. In particolare, sono state effettuate le prove simulando anche un errore massimo di ampiezza e un errore massimo di fase di 0.2% e 0.2 crad, rispettivamente.

I risultati nella Tabella 3.3 mostrano come il metodo proposto sia significativamente più accurato degli altri, in ogni condizione e configurazione considerata. Il Metodo A ottiene per γ_1 una stima sempre peggiorativa dell'incertezza a priori, poiché gli errori sistematici influenzano fortemente le sue stime. È interessante notare che i suoi RMSE variano solo leggermente al variare dei C casi e dell'accuratezza delle PMU, perché gli errori sistematici sono la principale fonte di errore di stima per lo specifico algoritmo. Il Metodo B sfrutta il numero di casi utilizzati per ridurre gli errori dei parametri di linea, ma gli RMSE degli errori sistematici sulle misurazioni di tensione e corrente non migliorano, o sono anche peggiori dei valori ottenuti con $C = 10$. Inoltre, il metodo proposto è più robusto del Metodo B quando l'incertezza della PMU aumenta, come evidenziato dal caso con accuratezza della PMU (0.2%, 0.2 crad) e $C = 10$. Questi e altri risultati hanno dimostrato che considerare tutte le fonti di incertezza, utilizzare un approccio multi-ramo e un algoritmo di stima robusto, è la chiave per migliorare le prestazioni di stima.

3.5 Conclusioni

Questo capitolo ha presentato un metodo a rami multipli per stimare contemporaneamente i parametri di linea nelle linee di trasmissione e gli errori sistematici nelle misure fasoriali sincronizzate, che è ottimizzato in presenza di più insiemi di misure ottenute da un sistema di monitoraggio su area vasta (WAMS) e corrispondenti a più istanti temporali, i timestamp associati alle misure. Il metodo proposto, basato su una definizione completa del modello di incertezza, fornisce le informazioni necessarie per una gestione efficace della rete di trasmissione. Il modello adottato consente di tenere conto dei diversi contributi di errore in una catena di misurazione realistica e delle informazioni preliminari sull'accuratezza dei trasformatori di misura e delle PMU. Le informazioni possono essere ottenute dalle schede tecniche degli strumenti e dalle procedure di caratterizzazione. Il metodo è stato progettato nell'ambito della regolarizzazione di Tikhonov, che corrisponde ad una formulazione $\mathcal{P}_{e_2+r_2}$, consentendo così

CAPITOLO 3. STIMA DEI PARAMETRI DI LINEA E COMPENSAZIONE
DEGLI ERRORI SISTEMATICI IN RETI DI TRASMISSIONE

un'ottimizzazione dell'algoritmo e la definizione di un criterio di configurazione per il problema specifico. Questa generalizzazione consente di sfruttare le potenzialità delle PMU, ovvero l'elevata accuratezza e l'elevata frequenza di misura (*reporting rate*), per migliorare la stima dei parametri di linea e per affinare la compensazione degli errori sistematici di misura. I test di simulazione eseguiti utilizzando la rete di test IEEE 14 hanno confermato la validità dell'approccio proposto e hanno evidenziato i vantaggi della stima multi-ramo, nella quale tutte le incognite vengono stimate contemporaneamente.

Tabella 3.3: IEEE 14 - Ramo 1: Confronto tra diversi metodi

Metodo	C	PMU accuratezza · [%], ∠[crad]	RMSE					
			γ_1 [%]	β_1 [%]	ξ_2 [%]	α_2 [crad]	η_{21} [%]	ψ_{21} [crad]
Metodo Proposto	10	0.1, 0.1	2.53	0.88	0.10	0.10	0.20	0.37
		0.2, 0.2	3.73	1.53	0.11	0.14	0.21	0.38
	200	0.1, 0.1	1.38	0.38	0.10	0.09	0.20	0.36
		0.2, 0.2	2.08	0.76	0.11	0.11	0.20	0.37
Metodo A [66]	10	0.1, 0.1	13.15	5.16	-	-	-	-
		0.2, 0.2	13.16	5.16	-	-	-	-
	200	0.1, 0.1	13.08	5.10	-	-	-	-
		0.2, 0.2	13.08	5.10	-	-	-	-
Metodo B [78, Sec. IV]	10	0.1, 0.1	7.34	2.48	0.37	0.42	0.36	0.56
		0.2, 0.2	14.99	5.71	0.55	0.65	0.53	0.68
	200	0.1, 0.1	2.55	0.97	0.29	0.36	0.29	0.53
		0.2, 0.2	4.84	3.05	0.31	0.46	0.30	0.53

Appendice

3.A Derivazione della matrice di misura e della rappresentazione dell'incertezza

Supponendo che i valori assoluti di tutti gli errori siano piccoli, molto inferiori a uno (es. $|\xi|, |\alpha|, |\eta|, |\psi| \ll 1$), assunzione del tutto realistica per le accuratèzze tipiche dei trasduttori e delle PMU, le equazioni in (3.1) possono essere riscritte in maniera tale che v_h^R e i_{ij}^R siano espressi in funzione dei sincrofasori misurati v_h e i_{ij} come segue:

$$\begin{aligned}
 v_h^R &\simeq (1 - \xi_h^{sys} - \xi_h^{rnd}) (V_h^r + jV_h^x) \\
 &\quad \cdot (1 - j\alpha_h^{sys} - j\alpha_h^{rnd}) \\
 i_{ij}^R &\simeq (1 - \eta_{ij}^{sys} - \eta_{ij}^{rnd}) (I_{ij}^r + jI_{ij}^x) \\
 &\quad \cdot (1 - j\psi_{ij}^{sys} - j\psi_{ij}^{rnd})
 \end{aligned} \tag{3.A.27}$$

dove, usando l'approssimazione $(1 + x)^{-1} \simeq 1 - x$ per $x \simeq 0$, gli errori di ampiezza che sarebbero al denominatore nel termine a destra sono portati al numeratore. Usando l'approssimazione $e^{-jx} \simeq 1 - jx$, le funzioni esponenziali in (3.1) possono essere prima coniugate e poi linearizzate per trovare (3.A.27).

Utilizzando (3.1) e (3.4) in (3.2) e (3.3), trascurando i termini del secondo ordine e dividendo parti reali e immaginarie delle equazioni, si ottengono le equazioni (3.A.28)-(3.A.31). In ogni equazione, il primo termine dipende da grandezze note (valori misurati e nominali) e il secondo termine è una combinazione lineare di errori sistematici e casuali, i cui coefficienti sono noti, perché ancora espressi dai valori misurati e nominali.

$$\begin{aligned}
& V_i^r - V_j^r - I_{ij}^r R_{ij}^0 + I_{ij}^x X_{ij}^0 - \frac{B_{sh,ij}^0}{2} (V_i^r X_{ij}^0 + V_i^x R_{ij}^0) \\
& \simeq \left[V_i^r - \frac{B_{sh,ij}^0}{2} (V_i^r X_{ij}^0 + V_i^x R_{ij}^0) \right] (\xi_i^{sys} + \xi_i^{rnd}) \\
& \quad + \left[-V_i^x - \frac{B_{sh,ij}^0}{2} (V_i^r R_{ij}^0 - V_i^x X_{ij}^0) \right] (\alpha_i^{sys} + \alpha_i^{rnd}) \\
& \quad - V_j^r (\xi_j^{sys} + \xi_j^{rnd}) + V_j^x (\alpha_j^{sys} + \alpha_j^{rnd}) \\
& \quad + (-I_{ij}^r R_{ij}^0 + I_{ij}^x X_{ij}^0) (\eta_{ij}^{sys} + \eta_{ij}^{rnd}) \\
& \quad + (I_{ij}^r X_{ij}^0 + I_{ij}^x R_{ij}^0) (\psi_{ij}^{sys} + \psi_{ij}^{rnd}) \\
& \quad + R_{ij}^0 \left(I_{ij}^r + \frac{B_{sh,ij}^0}{2} V_i^x \right) \gamma_{ij} + X_{ij}^0 \left(-I_{ij}^x + \frac{B_{sh,ij}^0}{2} V_i^r \right) \beta_{ij} \\
& \quad + \frac{B_{sh,ij}^0}{2} (V_i^r X_{ij}^0 + V_i^x R_{ij}^0) \rho_{ij}
\end{aligned} \tag{3.A.28}$$

$$\begin{aligned}
& V_i^x - V_j^x - I_{ij}^r X_{ij}^0 - I_{ij}^x R_{ij}^0 + \frac{B_{sh,ij}^0}{2} (V_i^r R_{ij}^0 - V_i^x X_{ij}^0) \\
& \simeq \left[V_i^x + \frac{B_{sh,ij}^0}{2} (V_i^r R_{ij}^0 - V_i^x X_{ij}^0) \right] (\xi_i^{sys} + \xi_i^{rnd}) \\
& \quad + \left[V_i^r + \frac{B_{sh,ij}^0}{2} (-V_i^r X_{ij}^0 + V_i^x R_{ij}^0) \right] (\alpha_i^{sys} + \alpha_i^{rnd}) \\
& \quad - V_j^x (\xi_j^{sys} + \xi_j^{rnd}) - V_j^r (\alpha_j^{sys} + \alpha_j^{rnd}) \\
& \quad + (-I_{ij}^r X_{ij}^0 - I_{ij}^x R_{ij}^0) (\eta_{ij}^{sys} + \eta_{ij}^{rnd}) \\
& \quad + (-I_{ij}^r R_{ij}^0 + I_{ij}^x X_{ij}^0) (\psi_{ij}^{sys} + \psi_{ij}^{rnd}) \\
& \quad + R_{ij}^0 \left(I_{ij}^x - \frac{B_{sh,ij}^0}{2} V_i^r \right) \gamma_{ij} + X_{ij}^0 \left(I_{ij}^r + \frac{B_{sh,ij}^0}{2} V_i^x \right) \beta_{ij} \\
& \quad + \frac{B_{sh,ij}^0}{2} (V_i^x X_{ij}^0 - V_i^r R_{ij}^0) \rho_{ij}
\end{aligned} \tag{3.A.29}$$

$$\begin{aligned}
& I_{ij}^r + I_{ji}^r + \frac{B_{sh,ij}^0}{2} (V_i^x + V_j^x) \\
& \simeq I_{ij}^r (\eta_{ij}^{sys} + \eta_{ij}^{rnd}) - I_{ij}^x (\psi_{ij}^{sys} + \psi_{ij}^{rnd}) \\
& \quad + I_{ji}^r (\eta_{ji}^{sys} + \eta_{ji}^{rnd}) - I_{ji}^x (\psi_{ji}^{sys} + \psi_{ji}^{rnd}) \\
& \quad + \frac{B_{sh,ij}^0}{2} [V_i^x (\xi_i^{sys} + \xi_i^{rnd}) + V_j^x (\xi_j^{sys} + \xi_j^{rnd}) \\
& \quad + V_i^r (\alpha_i^{sys} + \alpha_i^{rnd}) + V_j^r (\alpha_j^{sys} + \alpha_j^{rnd}) - (V_i^x + V_j^x) \rho_{ij}]
\end{aligned} \tag{3.A.30}$$

$$\begin{aligned}
& I_{ij}^x + I_{ji}^x - \frac{B_{sh,ij}^0}{2} (V_i^r + V_j^r) \\
& \simeq I_{ij}^x (\eta_{ij}^{sys} + \eta_{ij}^{rnd}) + I_{ij}^r (\psi_{ij}^{sys} + \psi_{ij}^{rnd}) \\
& \quad + I_{ji}^x (\eta_{ji}^{sys} + \eta_{ji}^{rnd}) + I_{ji}^r (\psi_{ji}^{sys} + \psi_{ji}^{rnd}) \\
& \quad + \frac{B_{sh,ij}^0}{2} [-V_i^r (\xi_i^{sys} + \xi_i^{rnd}) - V_j^r (\xi_j^{sys} + \xi_j^{rnd}) \\
& \quad + V_i^x (\alpha_i^{sys} + \alpha_i^{rnd}) + V_j^x (\alpha_j^{sys} + \alpha_j^{rnd}) + (V_i^r + V_j^r) \rho_{ij}]
\end{aligned} \tag{3.A.31}$$

Conclusioni

Le moderne reti di trasmissione e distribuzione stanno vivendo una fase di evoluzione che richiede strumenti di gestione e controllo sempre più complessi, flessibili ed efficaci. Tali strumenti si basano su dati ottenuti da sistemi di monitoraggio, costituiti da strumenti di misura dalle caratteristiche sempre più avanzate ma comunque caratterizzati da incertezza, e mediante metodologie di stima influenzate da tutte le incertezze della catena di misura. È quindi sempre più importante disporre di strumenti matematici in grado di modellare efficacemente i problemi di stima, sfruttando appieno le caratteristiche degli strumenti di misura e le conoscenze a priori sul dominio.

In questa tesi è stato presentato un paradigma basato su tecniche di ottimizzazione per la stima da misure sincronizzate e declinato per due importanti applicazioni di monitoraggio di sistemi elettrici di potenza: la stima delle principali sorgenti di inquinamento armonico in una rete di distribuzione e la stima dei parametri di linea in una rete di trasmissione simultanea alla stima degli errori sistematici della catena di misura. Sono stati affrontati gli aspetti teorici e pratici relativi all'integrazione di possibili informazioni a priori nei problemi sotto indagine: informazioni sul dominio, sul sistema di misura e sulla sua incertezza e sulle caratteristiche note della soluzione incognita.

La ricerca delle principali sorgenti armoniche è stata modellata come un problema a soluzione sparsa ed affrontata con un nuovo algoritmo, formulato come problema a minima norma in ℓ_1 con vincolo sull'errore di approssimazione (formulazione \mathcal{P}_{r_1/e_2}). Si è mostrato come l'algoritmo proposto possa raggiungere ottime prestazioni in termini di identificazione delle sorgenti di inquinamento e di stabilità delle stesse al variare sia dell'incertezza sia delle sorgenti da rilevare, grazie all'integrazione nella formulazione di un'accurata modellazione dell'incertezza.

La stima simultanea dei parametri di linea e degli errori sistematici della catena di misura è stata affrontata mediante un nuovo metodo basato su misure fasoriali sincronizzate. Il metodo è stato concepito per sfruttare un approccio a rami multipli e un numero potenzialmente elevato di misure ottenute da un

sistema di tipo WAMS e corrispondenti a più condizioni operative. Il metodo è stato progettato nell'ambito della regolarizzazione di Tikhonov (formulazione $\mathcal{P}_{e_2+r_2}$) e consente di sfruttare più efficacemente le potenzialità delle PMU, ovvero l'elevata accuratezza e l'elevato *reporting rate*, per migliorare la stima dei parametri di linea e per affinare la compensazione degli errori sistematici di misura. I risultati ottenuti in simulazione hanno evidenziato i vantaggi dell'approccio proposto rispetto alla letteratura.

La validità del paradigma proposto è stata quindi confermata da tutte le applicazioni e sperimentazioni condotte. La flessibilità delle tecniche di ottimizzazione discusse in questa tesi ha infatti permesso di modellare le diverse informazioni a priori sul dominio in esame e sullo stato da ricostruire, sfruttando ogni tipo di conoscenza disponibile sull'errore di misura. Tali tecniche possono quindi essere considerate un valido strumento, flessibile ed efficace, per affrontare i nuovi problemi di misura posti dalle moderne reti elettriche di potenza.

Bibliografia

- [1] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. New Heaven, CT: Yale University Press, 1923.
- [2] G. Rodriguez, *Algoritmi numerici*. Bologna, Italy: Pitagora Editrice, 2008.
- [3] Å. Björck, *Numerical methods for least squares problems*. Philadelphia: SIAM, 1996.
- [4] C. Clason, "L[∞] fitting for inverse problems with uniform noise," *Inverse Problems*, vol. 28, no. 10, p. 104007, Oct. 2012. [Online]. Available: <http://stacks.iop.org/0266-5611/28/i=10/a=104007?key=crossref.322b065ee01ac62c3bd9659d8819ace3>
- [5] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004.
- [6] T. E. Dielman, "Least absolute value regression: recent contributions," *Journal of Statistical Computation and Simulation*, vol. 75, no. 4, pp. 263–286, 2005. [Online]. Available: <https://doi.org/10.1080/0094965042000223680>
- [7] A. Abur and A. Gómez Expósito, *Power system state estimation: theory and implementation*, ser. Power engineering. New York, NY: Marcel Dekker, 2004.
- [8] R. J. Marks, G. L. Wise, D. G. Haldeman, and J. L. Whited, "Detection in laplace noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-14, no. 6, pp. 866–872, 1978.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009. [Online]. Available: <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>
- [10] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*. New York, NY: John Wiley & Sons, 2013.

-
- [11] E. Candes and J. Romberg, “l1-magic : Recovery of sparse signals via convex programming,” 2005. [Online]. Available: <https://statweb.stanford.edu/~candes/software/l1magic/downloads/l1magic.pdf>
- [12] T. A., “Solution of incorrectly formulated problems and the regularization method,” *Soviet Math. Dokl.*, vol. 5, p. 1035/1038, 1963. [Online]. Available: <https://ci.nii.ac.jp/naid/10004315593/en/>
- [13] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, pp. 55–67, 1970.
- [14] A. E. Hoerl and R. W. Kennard, “Ridge regression: Applications to nonorthogonal problems,” *Technometrics*, vol. 12, pp. 69–82, 1970.
- [15] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: <http://www.jstor.org/stable/2346178>
- [16] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.
- [17] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [18] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, ser. Applied and numerical harmonic analysis. New York: Birkhäuser, 2013, oCLC: ocn860992971.
- [19] E. Candes and M. Wakin, “An Introduction To Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008. [Online]. Available: <http://ieeexplore.ieee.org/document/4472240/>
- [20] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*. New York, NY: Springer New York, 1984.
- [21] G. W. Stewart, *Matrix algorithms*. Philadelphia: Society for Industrial and Applied Mathematics, 1998.
- [22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: The John Hopkins University Press, 1996.
- [23] C. F. Van Loan, “Generalizing the singular value decomposition,” *SIAM Journal on Numerical Analysis*, vol. 13, no. 1, pp. 76–83, 1976. [Online]. Available: <http://www.jstor.org/stable/2156468>

- [24] P. C. Hansen, “The discrete picard condition for discrete ill-posed problems,” *BIT Numerical Mathematics*, vol. 30, pp. 658–672, 1990.
- [25] L. Reichel, F. Sgallari, and Q. Ye, “Tikhonov regularization based on generalized Krylov subspace methods,” *Applied Numerical Mathematics*, vol. 62, no. 9, pp. 1215–1228, Sep. 2012. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0168927410001765>
- [26] S. Gazzola, “Regularization techniques based on krylov methods for ill-posed linear systems,” Ph.D. dissertation, Dept. of Mathematics, University of Padua, Italy, Padua, Italy, 2014.
- [27] A. Buccini, “Tikhonov-type iterative regularization methods for ill-posed inverse problems: theoretical aspects and applications,” Ph.D. dissertation, Dept. of Science and High Technology, University of Insubria, Italy, 2016.
- [28] P. C. Hansen, “REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems,” *Numerical Algorithms*, vol. 6, no. 1, pp. 1–35, Mar. 1994.
- [29] P. C. Hansen, “The L-curve and its use in the numerical treatment of inverse problems,” *P. Johnston (Ed.), Computational Inverse Problems in Electrocar-diography*, WIT Press, Southampton, pp. 119–142, 2001.
- [30] L. Reichel and A. Shyshkov, “A new zero-finder for Tikhonov regularization,” *BIT Numerical Mathematics*, vol. 48, no. 3, pp. 627–643, Sep. 2008.
- [31] JCGM, “Evaluation of data - guide to the expression of uncertainty in measurement,” *JCGM 100:2008*, Sep. 2008.
- [32] D. Carta, C. Muscas, P. A. Pegoraro, A. V. Solinas, and S. Sulis, “Compressive Sensing-Based Harmonic Sources Identification in Smart Grids,” *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [33] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau, “Exact Sparse Approximation Problems via Mixed-Integer Programming: Formulations and Computational Performance,” *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1405–1419, Mar. 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7313004/>
- [34] M. H. J. Bollen, “What is power quality?” *Electric Power Systems Research*, vol. 66, no. 1, pp. 5–14, Jul. 2003.

- [35] K. K. C. Yu, N. R. Watson, and J. Arrillaga, "Error analysis in static harmonic State estimation: a statistical approach," *IEEE Trans. Power Del.*, vol. 20, no. 2, pp. 1045–1050, Apr. 2005.
- [36] C. F. M. Almeida and N. Kagan, "Harmonic state estimation through optimal monitoring systems," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 467–478, Mar. 2013.
- [37] C. Rakpenthai, S. Uatrongjit, N. R. Watson, and S. Premrudeepreechacharn, "On harmonic state estimation of power system with uncertain network parameters," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4829–4838, Nov. 2013.
- [38] Y. Xi *et al.*, "Harmonic estimation in power systems using an optimised adaptive kalman filter based on pso-ga," *IET Generation, Transmission Distribution*, vol. 13, no. 17, pp. 3968–3979, 2019.
- [39] N. Locci, C. Muscas, and S. Sulis, "Experimental comparison of mv voltage transducers for power quality applications," in *2009 IEEE Instrumentation and Measurement Technology Conference*, 2009, pp. 92–97.
- [40] G. Crotti, D. Gallo, D. Giordano, C. Landi, M. Luiso, and M. Modarres, "Frequency response of mv voltage transformer under actual waveforms," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 6, pp. 1146–1154, 2017.
- [41] M. Faifer, C. Laurano, R. Ottoboni, S. Toscani, and M. Zanoni, "Harmonic distortion compensation in voltage transformers for improved power quality measurements," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3823–3830, 2019.
- [42] A. Cataliotti *et al.*, "Compensation of nonlinearity of voltage and current instrument transformers," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 5, pp. 1322–1332, 2019.
- [43] J. Mazumdar, R. G. Harley, F. C. Lambert, G. K. Venayagamoorthy, and M. L. Page, "Intelligent tool for determining the true harmonic current contribution of a customer in a power distribution network," *IEEE Trans. Ind. Appl.*, vol. 44, no. 5, pp. 1477–1485, Sep. 2008.
- [44] X. Zhu, C. Shen, and X. Ren, "Harmonic detection method using APFFT and neural network," in *2013 5th Int. Conf. Intelligent Human-Machine Syst. and Cybernetics*, Hangzhou, China, Aug. 2013, pp. 356–359.

- [45] V. K. Tiwari and S. K. Jain, “Hardware implementation of polyphase-decomposition-based wavelet filters for power system harmonics estimation,” *IEEE Trans. Instrum. Meas.*, vol. 65, no. 7, pp. 1585–1595, 2016.
- [46] R. Sinvula, K. M. Abo-Al-Ez, and M. T. Kahn, “Harmonic source detection methods: A systematic literature review,” *IEEE Access*, vol. 7, pp. 74 283–74 299, 2019.
- [47] G. D’Antona, C. Muscas, P. A. Pegoraro, and S. Sulis, “Harmonic source estimation in distribution systems,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3351–3359, Oct. 2011.
- [48] G. D’Antona, C. Muscas, and S. Sulis, “Localization of nonlinear loads in electric systems through harmonic source estimation,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3423–3430, Oct. 2011.
- [49] D. Carta, C. Muscas, P. A. Pegoraro, and S. Sulis, “Harmonics detector in distribution systems based on compressive sensing,” in *IEEE Int. Workshop on Appl. Meas. for Power Syst. (AMPS)*, Liverpool, UK, Sep. 2017, pp. 1–5.
- [50] D. Carta, C. Muscas, P. A. Pegoraro, and S. Sulis, “Identification and estimation of harmonic sources based on compressive sensing,” *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 95–104, Jan. 2019.
- [51] A. Galli, G. Frigo, D. Chindamo, A. Depari, M. Gadola, and G. Giorgi, “Denosing ECG signal by CSTFM algorithm: Monitoring during motorbike and car races,” *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2433–2441, 2019.
- [52] J. Zhang, Z. L. Yu, Z. Gu, Y. Li, and Z. Lin, “Multichannel electrocardiogram reconstruction in wireless body sensor networks through weighted $\ell_{1,2}$ minimization,” *IEEE Trans. Instrum. Meas.*, vol. 67, no. 9, pp. 2024–2034, 2018.
- [53] S. Rousseau and D. Helbert, “Compressive color pattern detection using partial orthogonal circulant sensing matrix,” *IEEE Trans. Image Process.*, vol. 29, pp. 670–678, 2020.
- [54] E. Balestrieri, L. De Vito, F. Picariello, and I. Tudosa, “A method exploiting compressive sampling for localization of radio frequency emitters,” *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2325–2334, 2020.
- [55] X. Xie, Q. Zhou, D. Hou, and H. Zhang, “Compressed sensing based optimal sensor placement for leak localization in water distribution

-
- networks,” *Journal of Hydroinformatics*, vol. 20, no. 6, pp. 1286–1295, Nov. 2018. [Online]. Available: <https://iwaponline.com/jh/article/20/6/1286/38895/Compressed-sensing-based-optimal-sensor-placement>
- [56] K. Duda, T. P. Zieliński, A. Bień, and S. H. Barczentewicz, “Harmonic phasor estimation with flat-top fir filter,” *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2039–2047, 2020.
- [57] G. Frigo, A. Derviškadić, P. A. Pegoraro, C. Muscas, and M. Paolone, “Harmonic phasor measurements in real-world PMU-based acquisitions,” in *IEEE Int. Instr. and Meas. Technol. Conf. (I2MTC)*, Auckland, NZ, 2019, pp. 1–6.
- [58] D. Carta, C. Muscas, P. A. Pegoraro, A. V. Solinas, and S. Sulis, “Impact of measurement uncertainties on compressive sensing-based harmonic source estimation algorithms,” in *IEEE Int. Instr. and Meas. Technol. Conf. (I2MTC)*, Dubrovnik, Croatia, May 2020.
- [59] Y. Pati, R. Rezaifar, and P. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44 vol.1.
- [60] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Commun. on Pure and Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [61] D. L. Donoho, “For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution,” *Commun. on Pure and Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.
- [62] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [63] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [64] H. Zhang, Z. Diao, and Y. Cui, “Identification of Power Network Branch Parameters Based on State Space Transformation,” *IEEE Access*, vol. 7, pp. 91 720–91 730, Jul. 2019.
- [65] J. Fu, G. Song, and B. De Schutter, “Influence of Measurement Uncertainty on Parameter Estimation and Fault Location for Transmission Lines,” *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 1, pp. 337–345, Jan. 2021.

-
- [66] M. Asprou, E. Kyriakides, and M. M. Albu, "Uncertainty Bounds of Transmission Line Parameters Estimated From Synchronized Measurements," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2808–2818, Aug. 2019.
- [67] S. Vlahinic, D. Frankovic, M. Z. Durovic, and N. Stojkovic, "Measurement Uncertainty Evaluation of Transmission Line Parameters," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–7, Apr. 2021.
- [68] A. Xue *et al.*, "Robust Identification Method for Transmission Line Parameters That Considers PMU Phase Angle Error," *IEEE Access*, vol. 8, pp. 86 962–86 971, May 2020.
- [69] V. Milojević, S. Čalija, G. Rietveld, M. V. Ačanski, and D. Colangelo, "Utilization of PMU measurements for three-phase line parameter estimation in power systems," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 10, pp. 2453–2462, Oct. 2018.
- [70] M. Asprou and E. Kyriakides, "Identification and estimation of erroneous transmission line parameters using pmu measurements," *IEEE Trans. Power Del.*, vol. 32, no. 6, pp. 2510–2519, Dec. 2017.
- [71] A. Wehenkel, A. Mukhopadhyay, J.-Y. L. Boudec, and M. Paolone, "Parameter Estimation of Three-Phase Untransposed Short Transmission Lines From Synchrophasor Measurements," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6143–6154, Sep. 2020.
- [72] Z. Wu, L. T. Zora, and A. G. Phadke, "Simultaneous transmission line parameter and PMU measurement calibration," in *2015 IEEE Power & Energy Society General Meeting*, Denver, CO, USA, Jul. 2015, pp. 1–5.
- [73] Y. G. Kononov, O. S. Rybasova, and K. A. Sidirov, "Identification of overhead-line parameters from PMU data with compensation of systematic measurement errors," in *2018 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, Moscow, Russia, May 2018, pp. 1–5.
- [74] R. S. Singh, S. Babaev, V. Cuk, S. Cobben, and H. van den Brom, "Line Parameters Estimation in Presence of Uncalibrated Instrument Transformers," in *2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET)*, Split, Croatia, Apr. 2019, pp. 1–8.
- [75] A. Pal, P. Chatterjee, J. S. Thorp, and V. A. Centeno, "Online Calibration of Voltage Transformers Using Synchrophasor Measurements," *IEEE Trans. Power Del.*, vol. 31, no. 1, pp. 370–380, Feb. 2016.

- [76] C. Wang, V. A. Centeno, K. D. Jones, and D. Yang, "Transmission Lines Positive Sequence Parameters Estimation and Instrument Transformers Calibration Based on PMU Measurement Error Model," *IEEE Access*, vol. 7, pp. 145 104–145 117, Oct. 2019.
- [77] K. V. Khandeparkar, S. A. Soman, and G. Gajjar, "Detection and Correction of Systematic Errors in Instrument Transformers Along With Line Parameter Estimation Using PMU Data," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3089–3098, Jul. 2017.
- [78] H. Goklani, G. Gajjar, and S. A. Soman, "Instrument transformer calibration and robust estimation of transmission line parameters using PMU measurements," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 1761–1770, May 2021.
- [79] C. Sitzia, C. Muscas, P. A. Pegoraro, A. V. Solinas, and S. Sulis, "Enhanced PMU-based Line Parameters Estimation and Compensation of Systematic Measurement Errors in Power Grids Considering Multiple Operating Conditions," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
- [80] C. Muscas, P. A. Pegoraro, C. Sitzia, A. V. Solinas, and S. Sulis, "Compensation of systematic measurement errors in PMU-based Monitoring Systems for Transmission grids," in *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Glasgow, UK, May 2021, pp. 1–6.
- [81] P. Ren, H. Lev-Ari, and A. Abur, "Tracking Three-Phase Untransposed Transmission Line Parameters Using Synchronized Measurements," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4155–4163, Jul. 2018.
- [82] P. A. Pegoraro, K. Brady, P. Castello, C. Muscas, and A. von Meier, "Line impedance estimation based on synchrophasor measurements for power distribution systems," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 4, pp. 1002–1013, Apr. 2019.
- [83] P. A. Pegoraro, K. Brady, P. Castello, C. Muscas, and A. von Meier, "Compensation of systematic measurement errors in a PMU-based monitoring system for electric distribution grids," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3871–3882, Oct. 2019.
- [84] R. Abu-Hashim *et al.*, "Test systems for harmonics modeling and simulation," *IEEE Trans. Power Del.*, vol. 14, no. 2, pp. 579–587, Apr. 1999.

- [85] C. Muscas *et al.*, “Characterization of a PMU-based method for transmission line parameters estimation with systematic measurement error modeling,” in *2021 AEIT International Annual Conference (AEIT)*, 2021, pp. 1–6.
- [86] S. L. Brunton and J. N. Kutz, “Data Driven Science & Engineering,” p. 572, 2017.

Elenco delle figure

1.1	Sfere unitarie in \mathbb{R}^2 rispetto alle norme 1, 2 e ∞	6
1.2	Sfere unitarie in \mathbb{R}^2 rispetto alle norme $p = \{\frac{1}{2}, 1, 5\}$ con evidenza della convessità per norme 1 e 5 e della non convessità per la $\ \cdot\ _{\frac{1}{2}}$	8
1.3	Errori relativi della stima delle componenti del vettore di stato nelle configurazioni: errore Gaussiano con outliers a sinistra (stima con minimizzazioni in ℓ_2 e in ℓ_1), errore uniforme a destra (stima con minimizzazioni in ℓ_2 e in ℓ_∞). Visualizzata una sola prova MC.	16
1.4	Soluzione grafica del problema $\mathcal{P}_{r_q/e=0}$ a norma minima con vincolo di uguaglianza, $\hat{\mathbf{x}}_{r_q} = \mathbf{x} : \min_{\mathbf{x}} \ \mathbf{x}\ _q$ s.t. $\mathbf{A}\mathbf{x} = \mathbf{b}$, nel caso $\mathbf{A} \in \mathbb{R}^{1 \times 2}$ e $q \in \{0, 1, 2, \infty\}$	18
1.5	Soluzione grafica del problema a minimo errore quadratico $\min_{\mathbf{x}} \ \mathbf{A}\mathbf{x} - \mathbf{b}\ _2^2$ con vincolo di regolarizzazione $\ \mathbf{x}\ _1 \leq t$, e $\ \mathbf{x}\ _2^2 \leq t^2$ a destra sinistra.	21
1.6	Minima norma in ℓ_1 e in ℓ_2 con vincolo quadratico sul residuo, $\mathcal{P}_{r_q/e_2} : \min_{\mathbf{x}} \ \mathbf{x}\ _q$ s.t. $\ \mathbf{A}\mathbf{x} - \mathbf{b}\ _2 \leq \epsilon_2$ per $q \in \{1, 2\}$. $\hat{\mathbf{x}}_{r_1/e_2}$ e $\hat{\mathbf{x}}_{r_2/e_2}$ sono le soluzioni per $q = 1$ e $q = 2$ rispettivamente.	22
1.7	Forma generica della <i>L-curve</i>	38
2.1	Istogrammi della componente reale e immaginaria delle misure di tensione e corrente NON sbiancate.	62
2.2	Istogrammi delle componenti delle misure rettangolari di tensione e corrente sbiancate tramite \mathbf{W}_C	62
2.3	Istogrammi delle componenti delle misure rettangolari di tensione e corrente sbiancate tramite \mathbf{W}_R	63
2.4	Istogrammi di $\xi_{\mathbf{e}_w} = \ \mathbf{e}_{w_C}\ _2^2$ e $\xi_{\mathbf{e}_w} = \ \mathbf{e}_{w_R}\ _2^2$, nei casi di misure sbiancate con le matrici di sbiancamento \mathbf{W}_C (a sinistra) e \mathbf{W}_R (a destra).	65
2.5	Schema della rete di test a 13 nodi.	66
2.6	Modello di carico ohmico-induttivo con forzante armonica.	67

2.7	Intervalli e medie per $Recall_c$, per gli ordini armonici $h \in \{3, 5, 7\}$.	70
2.8	Identificazione del carico inquinante $L2$ in tutte le configurazioni, per l'ordine armonico $h = 5$.	72
2.9	Identificazione del carico inquinante $L7$ in tutte le configurazioni, per l'ordine armonico $h = 5$.	72
2.10	Identificazione del carico inquinante $L11$ in tutte le configurazioni, per l'ordine armonico $h = 5$.	73
2.11	Topologia della rete di prova a 49 nodi	76
3.1	Modello a π per un generico ramo di rete di trasmissione e relativi parametri.	82
3.2	Rete di test IEEE 14 bus system	89
3.3	IEEE 14 - 6 rami, RRSE medio in funzione di μ in tre diversi scenari: $C \in \{10, 50, 500\}$	90
3.4	IEEE 14 - 6 rami, profilo dei valori singolari σ_j e del numero di condizionamento $\kappa_2(\mathbf{A})$ della matrice \mathbf{A} , in tre diversi scenari: $C \in \{10, 50, 500\}$	91
3.5	IEEE 14 - 6 rami, effetto del filtro.	92
3.6	IEEE 14 - 6 rami: RRSE medio (5000 prove MC realizzate) con un numero di condizioni operative C variabile da 10 sino a 500.	93
3.7	Flow chart della procedura proposta.	94
3.8	IEEE 14 - 6 rami: risultati di stima per il parametro γ di ognuno dei 6 rami.	98
3.9	IEEE 14 - 6 rami: risultati di stima per tutti i parametri di linea.	99
3.10	IEEE 14 - 6 rami: risultati di stima per ξ in funzione del numero di casi C per tutti i metodi presi in considerazione.	99
3.11	Risultati di stima di β utilizzando l'approccio multi-ramo, applicato ad una porzione di rete e all'intera rete, utilizzando diversi metodi e configurazioni.	100
3.12	IEEE 14 - intera rete: risultati di stima per tutti i parametri di linea (deviazioni).	101

Elenco delle tabelle

1.1	RRSE medi, 10 casi e 1000 MC	15
2.1	Configurazione dei carichi nominali	67
2.2	Configurazione delle Correnti Armoniche Forzanti	67
2.3	Errori di Misura Massimi	68
2.4	Media e deviazione standard di $Recall_c$ per i due algoritmi (rete di test a 13 nodi)	71
2.5	Media e deviazione standard di $Recall_c$ per i tre algoritmi (rete a 49 nodi)	74
2.6	Carichi nominali della rete di prova a 49 nodi	77
2.7	Configurazione di misura della rete ti prova a 49 nodi	77
3.1	IEEE 14 - 6 rami: migliori RRSE medi (su 5000 prove MC) per metodi differenti	97
3.2	IEEE 14 - Nodo 1: RMSE (5000 prove MC) per ξ_1 - Confronto tra gli approcci Multi-Ramo e Ramo-Singolo	98
3.3	IEEE 14 - Ramo 1: Confronto tra diversi metodi	103

Glossario

BLUE Best Linear Unbiased Estimator

BOMP Block-Orthogonal Matching Pursuit

BP Basis Pursuit

BPDN Basis Pursuit De-Noising

CS Compressive Sensing

DP Discrepancy Principle

EMS Energy Management System

ETF Equiangular Tight Frame

GCV Generalized Cross-Validation

GSVD Generalized Singular Value Decomposition

HSoE Harmonic Source Estimation

IID Independent Identically Distributed

LAD Least Absolut Deviations

LASSO Least Absolute Shrinkage and Selection Operator

LAV Least Absolut Values

LCC L-curve criterion

LP Linear Programming

LS Least Squares

- MAP** Maximum a Posteriori Probability
- MC** Monte Carlo
- MILP** Mixed Integer Linear Programming
- MIP** Mixed Integer Programming
- ML** Machine Learning
- MLE** Maximum Likelihood Estimator
- NP** non-deterministic polynomial acceptable problems
- OMP** Orthogonal Matching Pursuit
- PCA** Principal Component Analysis
- PDF** Probability Density Function
- PMU** Phasor Measurement Unit
- PQ** Power Quality
- RIP** Restricted Isometry Property
- RMSE** Root Mean Square Error
- RRSE** Relative Root Square Error
- RSS** Residual Sum of Squares
- SG** Smart Grid
- SP** Signal Processing
- SVD** Singular Value Decomposition
- TA** Trasformatore di Corrente
- TSO** Transmission System Operator
- TSVD** Truncated Singular Value Decomposition
- TV** Trasformatore di Tensione
- UTC** Coordinated Universal Time
- WAMS** Wide Area Monitoring System
- WLS** Weighted Least Squares

Elenco dei simboli

($P0$) Minima norma in ℓ_0 , con vincolo di uguaglianza [16], equivalente al $\mathcal{P}_{r_0/e=0}$:
(1.65)

($P1$) Minima norma in ℓ_1 , con vincolo di uguaglianza [16]. Il problema è anche conosciuto come *Basis Pursuit* [16], equivalente al $\mathcal{P}_{r_1/e=0}$

($P2$) Minima norma in ℓ_1 , con vincolo quadratico [16], equivalente al \mathcal{P}_{r_1/e_2} . Il problema è riconducibile al *Basis Pursuit De-Nosing* (BPDN), [16], (1.68)

$P(\mathbf{x})$ Probabilità a priori, *Prior*. Nel caso Gaussiano (1.58). Nel caso Laplaciano (1.60).

$P(\cdot)$ Probabilità di (\cdot) .

$P(\mathbf{b}|\mathbf{x})$ Probabilità condizionata di \mathbf{b} dato \mathbf{x} .

$P(\mathbf{x}|\mathbf{b})$ Probabilità condizionata, di \mathbf{x} dato \mathbf{b} .

S-sparso Vettore con al massimo $S \in \mathbb{N}$ componenti diverse da zero. Definizione: 1.4.1

$\Im[\cdot]$ Operatore parte immaginaria del vettore complesso (\cdot)

$\Re[\cdot]$ Operatore parte reale del vettore complesso (\cdot)

ℓ_0 Equivalente a ℓ_0 -norm e $\|\cdot\|_0$. Formulazione: (1.16).

ℓ_1 Equivalente a ℓ_1 -norm e $\|\cdot\|_1$. Formulazione: (1.10).

ℓ_2 Equivalente a ℓ_2 -norm e $\|\cdot\|_2$. Formulazione: (1.11).

ℓ_p Norma vettoriale o matriciale, presentata in modo esteso in 1.2, equivalente a ℓ_p -norm e $\|\cdot\|_p$

ℓ_∞ Equivalente a ℓ_∞ -norm e $\|\cdot\|_\infty$. Formulazione: (1.12).

$\hat{\mathbf{x}}_1$ soluzione del problema \mathcal{P}_{e_1/r_u} a minimo errore di approssimazione in ℓ_1 con vincolo di regolarizzazione illimitato. Equivalente al metodo dei minimi valori assoluti LAV. Formulazione: (1.29).

$\hat{\mathbf{x}}_2$ soluzione del problema \mathcal{P}_{e_2/r_u} a minimo errore di approssimazione in ℓ_2 con vincolo di regolarizzazione illimitato. Equivalente al metodo dei minimi quadrati LS, (per matrici quadrate o sovra-determinate). Formulazione: (1.27).

$\hat{\mathbf{x}}_\infty$ soluzione del problema $\mathcal{P}_{e_\infty/r_u}$ a minimo errore di approssimazione in ℓ_∞ con vincolo di regolarizzazione illimitato. Formulazione: (1.28).

$\hat{\mathbf{x}}_p$ soluzione del problema \mathcal{P}_{e_p/r_u} a minimo errore di approssimazione in ℓ_p con vincolo di regolarizzazione illimitato. Formulazione: (1.26).

$\hat{}$ Indica la quantità stimata

$\kappa(\mathbf{A})$ Numero di condizionamento della matrice \mathbf{A}

$\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$, insieme dei reali positivi, escluso lo zero

$\mathcal{L}(\cdot)$ *Likelihood Function*, funzione di verosimiglianza, Formulazione: (1.51).

$\mathcal{N}(\mathbf{A})$ “Nucleo”, *Kernel* o **Null-Space** di una generica matrice $\mathbf{A} \in \mathbb{R}^{m \times n}$.
Definizione: 1.3.1

\mathcal{P}_{e_1/r_u} minimo errore di approssimazione in ℓ_1 con vincolo di regolarizzazione illimitato. Equivalente al metodo dei minimi valori assoluti LAV. Formulazione: (1.29).

$\mathcal{P}_{e_2+r_1}$ Minimi errore di regolarizzazione in ℓ_2 norma in ℓ_1 in formulazione Lagrangiana. Il problema è anche conosciuto come *Basis Pursuit De-Noising* (BPDN) [16] ed in statistica alla *LASSO Regression*, ed è equivalente a ed è equivalente a \mathcal{P}_{e_2/r_1} . Formulazione: 1.44.

$\mathcal{P}_{e_2+r_2}$ Minimi errore di regolarizzazione e norma in ℓ_2 in formulazione Lagrangiana. Il problema è anche conosciuto come regolarizzazione di Tikhonov ed in statistica alla *Ridge Regression*. Formulazione: 1.43.

\mathcal{P}_{e_2/r_1} Minimo errore di regolarizzazione in ℓ_2 con vincolo in ℓ_1 del termine di regolarizzazione. Il problema è anche conosciuto come *LASSO Regression* [15], ed è equivalente a $\mathcal{P}_{e_2+r_1}$. Formulazione: 1.45.

\mathcal{P}_{e_2/r_u} minimo errore di approssimazione in ℓ_2 con vincolo di regolarizzazione illimitato. Equivalente al metodo dei minimi quadrati LS, (per matrici quadrate o sovra-determinate). Formulazione: (1.27).

- $\mathcal{P}_{e_\infty/r_u}$ minimo errore di approssimazione in ℓ_∞ con vincolo di regolarizzazione illimitato. Formulazione: (1.28).
- \mathcal{P}_{e_p/r_u} minimo errore di approssimazione in norma- p con vincolo di regolarizzazione illimitato, (1.26)
- $\mathcal{P}_{r_0/e=0}$ Minima norma in ℓ_0 (del termine di regolarizzazione), con vincolo di uguaglianza, (1.65)
- \mathcal{P}_{r_1/e_2} Minima norma in ℓ_1 , con vincolo quadratico. Il problema è anche conosciuto come problema (P2) [11]. Il problema è riconducibile al *Basis Pursuit De-Nosing* (BPDN), [16], (1.68).
- $\mathcal{P}_{r_1/e=0}$ Minima norma in ℓ_1 , con vincolo di uguaglianza (per sistemi sotto-determinati). Il problema è anche conosciuto come *Basis Pursuit* [16] o problema (P1) [11], (1.67)
- $\mathcal{P}_{r_2/e=0}$ Minima norma in ℓ_2 , con vincolo di uguaglianza. Si tratta del metodo dei minimi quadrati a norma minima, ovvero il metodo dei minimi quadrati per sistemi sotto-determinati. Formulazione: (1.39).
- $\mathcal{P}_{r_q/e=0}$ Minima norma- q (del termine di regolarizzazione), con vincolo di uguaglianza, vedi Figura 1.4
- $\mathcal{R}(\mathbf{A})$ $\text{span}\{\cdot\}$ dei vettori colonna della matrice \mathbf{A} , ovvero il sottospazio vettoriale generato di vettori colonna di \mathbf{A} .
- \mathbf{A}^\dagger pseudo inversa di Monroe-Penrose, Formulazione (1.85)
- \mathbf{I}_n Matrice identità di ordine n
- $\|\cdot\|_p$ Norma vettoriale o matriciale, presentata in modo esteso in 1.2, equivalente a ℓ_p -norm, o più semplicemente ℓ_p
- \odot prodotto matriciale di Hadamard, ovvero prodotto eseguito elemento per elemento.
- $\text{Card}\{\cdot\}$ Cardinalità, indica il numero di componenti dell'insieme $\{\cdot\}$
- $E[\cdot]$ Aspettazione, *Expectation*, valore atteso di $[\cdot]$
- $\text{rank}(\mathbf{A})$ Rango della matrice \mathbf{A} , ovvero il numero massimo di righe o colonne linearmente indipendenti
- $\text{span}\{\cdot\}$ Spazio vettoriale generato dall'insieme di vettori $\{\cdot\}$