

Università degli Studi di Firenze

Dipartimento di Ingegneria dell'Informazione

Dottorato di Ricerca in Informatica, Sistemi e
Telecomunicazioni

XXVI^o ciclo - anni 2011-2013

Indirizzo Ingegneria Informatica e dell'Automazione

Settore Scientifico Disciplinare: ING-INF/04



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DINFO
DIPARTIMENTO DI
INGEGNERIA
DELL'INFORMAZIONE

Stima distribuita dello stato
in reti di sensori

Giovanni Mugnai

Coordinatore: Prof. Luigi Chisci

Tutor: Proff. Luigi Chisci e Giorgio Battistelli

Firenze - Dicembre 2013

Vi sono più cose in cielo e in terra, Orazio,
di quante non sogni la tua filosofia.

— W. Shakespeare, Amleto, I. 5

Indice

Indice	v
Elenco delle figure	vii
Elenco delle tabelle	viii
Elenco degli algoritmi	ix
Introduzione	xvii
I Stima Bayesiana distribuita	1
1 Stima Bayesiana centralizzata	3
1.1 Predizione e correzione nella stima Bayesiana	4
1.1.1 Correzione	5
1.1.2 Predizione	6
1.1.3 Algoritmi per la stima Bayesiana	7
1.2 PDF di transizione e verosimiglianza	7
1.3 Caso lineare Gaussiano: Filtro di Kalman	8
1.4 Stima dello stato di un sistema non lineare	10
1.4.1 Extended Kalman Filter	10
1.4.2 Unscented Kalman Filter	12
1.4.3 Filtro a particelle	16
2 Stima Bayesiana distribuita	21
2.1 Modello della rete	21
2.2 Metriche per i grafi	24
2.3 Topologia di rete	25
2.4 Consenso sulla media di Kullback-Leibler	27
2.5 Filtraggio Bayesiano distribuito	33
2.5.1 Consenso a posteriori	34
2.5.2 Consenso sulla verosimiglianza	35
2.5.3 Consenso parallelo su verosimiglianza e informazione a priori	35
3 Algoritmi per la stima distribuita	37
3.1 Caso lineare Gaussiano	38
3.2 Caso non lineare	43

3.2.1	Consenso a posteriori	44
3.2.2	Consenso sulla verosimiglianza	45
3.2.3	Consenso parallelo su verosimiglianza e informazione a priori	47
3.3	Analisi di stabilità	50
3.3.1	Analisi di stabilità dell'algoritmo CP	51
3.3.2	Analisi di stabilità dell'algoritmo CLCP	53
3.4	Risultati simulativi	54
II	Stima distribuita in presenza di virus	61
4	I virus nelle reti di sensori	63
4.1	Modelli di diffusione dei virus nelle reti	63
4.1.1	Modelli macroscopici	65
4.1.2	Modello a catena di Markov singola esatta	66
4.1.3	Modello a catene di Markov intrecciate	69
4.2	Comportamento dei virus nel consenso	71
4.3	Virus e stima distribuita: analisi simulativa	72
III	Stima distribuita tra agenti a stato parzialmente sovrapposto	83
5	Consenso tra agenti a stati sovrapposti	85
5.1	Fusione di ordine ridotto	85
5.2	Risultati delle simulazioni	88
5.2.1	Sistemi lineari	90
5.2.2	Sistema non lineare	94
	Conclusioni	97
	Appendici	
	A	101
	B	113
	C	117
	D	121
	Bibliografia	123

Elenco delle figure

1.1	Effetto Verosimiglianza.	6
2.1	Modello della rete.	22
2.2	Rete geometrica casuale con $R_{max} = 780 (m)$	26
2.3	Rete geometrica casuale con $R_{max} = 1200 (m)$	27
3.1	Rete di sensori lineare.	56
3.2	Rete di sensori non-lineare.	57
3.3	Caso lineare: PRMSE per diversi valori di L	58
3.4	Caso non-lineare: PRMSE per diversi valori di L	59
4.1	Modello a catena di Markov per la diffusione dei virus.	64
4.2	Probabilità d'infezione per alcuni valori di β	67
4.3	Probabilità del tempo di recupero per alcuni valori di α	68
4.4	Reti di sorveglianza con 20 nodi.	74
4.5	Reti di sorveglianza con 50 nodi.	75
4.6	Reti di sorveglianza con 100 nodi.	76
4.7	Dinamiche temporali delle epidemie virali.	78
4.8	PRMSE nel caso della rete di 20 nodi.	80
4.9	PRMSE nel caso della rete di 50 nodi.	80
4.10	PRMSE nel caso della rete di 100 nodi.	81
5.1	Schema di FoU	87
5.2	Schema di $FoU - NI$	87
5.3	Schema di $FoLSS$	88
5.4	Schema di $FoLSS - NI$	89
5.5	Schema di FoI	89
5.6	Esempi di traiettorie del sistema lineare instabile.	91
5.7	Esempi di traiettorie del sistema lineare stabile.	92
5.8	Sistema lineare: RMSE medi dgl'agente 1.	93
5.9	Sistema lineare: RMSE medi dgl'agente 2.	93
5.10	Attrattore di Lorentz.	94
5.11	Posizioni dei sensori TOA.	96
5.12	Esempio di una traiettoria disturbata del sistema di Lorentz.	97
5.13	Sistema non-lineare: RMSE medi.	98
B.1	Rete lineare: PRMSE per $L = 1$ e $L = 2$	113
B.2	Rete lineare: PRMSE per $L = 3$ e $L = 4$	113

B.3	Rete lineare: PRMSE per $L = 5$ e $L = 6$	114
B.4	Rete lineare: PRMSE per valori di $L = 7$ e $L = 8$	114
B.5	Rete lineare: PRMSE per $L = 9$	114
B.6	Rete non-lineare: PRMSE per $L = 1$ e $L = 2$	115
B.7	Rete non-lineare: PRMSE per $L = 3$ e $L = 4$	115
B.8	Rete non-lineare: PRMSE per $L = 5$ e $L = 6$	115
B.9	Rete non-lineare: PRMSE per $L = 7$ e $L = 8$	116
B.10	Rete non-lineare: PRMSE per $L = 9$	116
C.1	Parametri di rete.	117
C.2	Soglie epidemiche.	118
C.3	Massima prevalenza epidemica per la rete di 20 nodi.	118
C.4	Massima prevalenza epidemica per la rete di 50 nodi.	119
C.5	Massima prevalenza epidemica per la rete di 100 nodi.	119
D.1	Sistema lineare, modello 1: RMSE.	121
D.2	Sistema lineare, modello 2: RMSE.	121
D.3	Sistema lineare, modello 3: RMSE.	122
D.4	Sistema lineare, modello 4: RMSE.	122
D.5	Sistema non-lineare: RMSE.	122

Elenco delle tabelle

3.1	Coppie d'informazione nella fase di correzione.	48
4.1	Valori di β corrispondenti a un picco epidemico.	79
5.1	Sistemi lineari: RMSE medi.	92
5.2	Sistema non-lineare: RMSE medi.	97

Elenco degli algoritmi

1	Algoritmo di stima Bayesiana.	7
2	Algoritmo del Filtro di Kalman.	9
3	Algoritmo EKF nella forma di covarianza.	12
4	Algoritmo UT.	15
5	Algoritmo UKF.	15
6	Algoritmo SIR-PF.	19
7	Filtro CP (Consensus on Posteriors) Bayesiano.	34
8	Filtro CL (Consensus on Likelihoods) Bayesiano.	35
9	Filtro CLCP (Consensus on Likelihoods & Posteriors) Bayesiano.	36
10	Algoritmo CP - Caso lineare Gaussiano.	40
11	Algoritmo CL - Caso lineare Gaussiano.	41
12	Algoritmo CLCP - Caso lineare Gaussiano.	42
13	Algoritmo CP - Caso non lineare.	44
14	Algoritmo CL - Caso non lineare.	47
15	Algoritmo CLCP - Caso non lineare.	49
16	Algoritmo CEKF/CUKF nella forma di informazione.	55

Sommario

Attualmente le reti di sensori rappresentano una tecnologia efficace per la realizzazione di sistemi di monitoraggio scalabili, flessibili, a costi contenuti e tolleranti ai guasti. La sorveglianza di un processo, di un oggetto o di un fenomeno richiede la stima dello stato del processo d'interesse mediante l'utilizzo di modelli matematici e l'effettuazione di misure di grandezze direttamente o indirettamente legate al suo stato. I sistemi di monitoraggio tradizionali sono costituiti da un insieme di sensori distribuiti nella regione di sorveglianza connessi con un unico centro di elaborazione che provvede a elaborare i dati raccolti. Negli ultimi tempi si stanno sempre più diffondendo reti in cui ogni nodo (o agente) è dotato di capacità di elaborazione, comunicazione e, in alcuni casi, rilevamento. Una rete di questo tipo è detta distribuita o anche *net-centrica*, in quanto ogni agente è in grado di effettuare una stima dello stato del processo monitorato, o di una sottoparte di esso.

Due diversi approcci per la stima distribuita di sistemi lineari, detti consenso sull'informazione a posteriori (CP: *Consensus on Posteriors*) e consenso sulla verosimiglianza (CL: *Consensus on Likelihoods*), vengono combinati per dar luogo ad una nuova classe di filtri di consenso su verosimiglianza e informazione a priori (CLCP: *Consensus on Likelihoods and Priors*), che beneficiano dei vantaggi complementari degli approcci CL e CP. Vengono presentati nuovi risultati teorici, limitatamente ai sistemi lineari, sulla stabilità dei filtri CLCP sotto condizioni di osservabilità collettiva e connettività di rete. Infine viene fatta una valutazione comparativa delle prestazioni di alcuni filtri basati sul consenso in un caso di studio relativo al tracciamento di un oggetto in movimento, sia con sensori lineari che non lineari.

Una delle principali vulnerabilità dei sistemi di sorveglianza net-centrici è la loro debolezza rispetto agli effetti distruttivi di virus informatici. Si sono studiati i principali modelli per la diffusione dei virus e l'insorgenza di un focolaio epidemico in relazione ai principali parametri di rete. Un ampio insieme di simulazioni su scenari realistici di sorveglianza ha permesso di analizzare il modo in cui i parametri di rete influenzano la diffusione del virus e la fusione distribuita dell'informazione, fornendo indicazioni utili per preservare la funzionalità di sorveglianza della rete in presenza di guasti e/o attacchi informatici.

Nelle reti di monitoraggio e sorveglianza non è detto che ciascun agente abbia una visione complessiva dello stato del sistema. Si è studiato il problema della stima distribuita nel caso in cui gli agenti abbiano una visione parziale del sistema complessivo facendo uno studio comparativo tra varie tecniche di fusione dei dati su scenari di studio sia lineari che non lineari.

Ringraziamenti

Prima di tutto voglio esprimere la mia sincera gratitudine al Prof. Luigi Chisci per l'opportunità di lavorare sotto la sua guida, il suo sostegno ed i suoi innumerevoli consigli durante questi tre anni. Sono profondamente grato a Giorgio Battistelli per la sua amicizia, i suggerimenti e le conversazioni che mi hanno aiutato a comprendere l'argomento. Voglio esprimere la mia gratitudine ai Dr. Alfonso Farina, Antonio Graziano e Alessio Liburdi di Selex ES per le stimolanti discussioni e gli spunti di riflessione su alcuni argomenti di questa tesi. Un sentito ringraziamento va anche ai colleghi del laboratorio Claudio Fantacci e Daniela Selvi, per la loro amicizia e disponibilità. Infine, sono grato ai miei genitori Marcella e Fernando, ai miei fratelli Marco e Paolo e gli amici, in particolare Sonia ed Enrico. Un ringraziamento speciale va a mia nipote Emma che con la sua solarità mi ha portato gioia ogni giorno. A lei dedico questa tesi.

Firenze, Dicembre 2013

G. M.

Acronimi

CE	Consensus on Estimates
CEKF	Centralized Extended Kalman Filter
CKF	Centralized Kalman Filter
CL	Consensus on Likelihoods
CLCP	Consensus on Likelihoods and Priors
COM	COMmunication node
CP	Consensus on Posteriors
CUKF	Centralized Unscented Kalman Filter
DCLCP	Distributed CLCP
DOA	Direction of Arrival
DSE	Distributed State Estimation
EKF	Extended Kalman Filter
EMD	Exponential Mixture Density
FoI	Fusion over Intersection
FoLSS	Fusion over the Local State Space
FoLSS-NI	Fusion over the Local State Space - Non Informative
FoU	Fusion over the Union
FoU-NI	Fusion over the Union - Non Informative
GCI	Generalized Covariance Intersection
IWC	Information Weighted Consensus
KF	Kalman Filter
KLA	Kullback-Leibler Average
KLD	Kullback-Leibler Divergence
MAP	Maximum A Posteriori
MC	Monte Carlo
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
PDF	Probability Density Function
PF	Particle Filter
PMDI	Principle of Minimum Discrimination Information
PRMSE	Position Root Mean Square Error
RMSE	Root Mean Square Error
SEN	SENsor node
SIR	Susceptible Infected Recovered
SIR-PF	Sequential Importance Resampling - Particle Filter
SIS	Susceptible Infected Susceptible
SMCF	Sequential Monte Carlo Filter
TOA	Time Of Arrival
UKF	Unscented Kalman Filter
UT	Unscented Transformation

Introduzione

In seguito ai recenti rapidi progressi delle tecnologie dell'informazione e delle comunicazioni, i sistemi di monitoraggio e sorveglianza basati su reti distribuite stanno sempre più affermandosi negli ambiti più diversificati, da quello ambientale a quelli tecnologico-industriali (reti di comunicazione, reti di distribuzione, sistemi di produzione) a quelli strettamente dedicati alla sicurezza sia in ambito civile che militare.

La recente evoluzione delle architetture dei sistemi di monitoraggio e sorveglianza ha portato alla definizione di un nuovo modello di rete in cui ogni elemento della rete, detto nodo o agente, è dotato di capacità di elaborazione, comunicazione e, in un certo numero di nodi della rete, di rilevamento e di misura. Si parla, in questo caso, di rete distribuita o decentralizzata o anche *net-centrica* perché ogni agente è in grado di effettuare una stima dello stato del sistema monitorato. L'architettura di rete distribuita è stata sviluppata in modo da conferire ai sistemi di sorveglianza le seguenti caratteristiche:

- **scalabilità**, cioè il carico elaborativo di ciascun nodo deve essere indipendente dalle dimensioni della rete;
- **flessibilità**, cioè la possibilità di introdurre nuovi nodi per estendere spazialmente o funzionalmente la rete senza dover riprogettare e ricostruire la rete;
- **stabilità**, cioè l'errore di stima dello stato deve essere limitato ed il più piccolo possibile in ogni nodo della rete;
- **robustezza**, cioè in presenza di nodi guasti e/o virus informatici le funzionalità di sorveglianza e/o monitoraggio della rete devono essere preservate.

Nelle architetture tradizionali la presenza di un centro di raccolta, fusione ed elaborazione dei dati rappresenta, per varie ragioni, un elemento di debolezza della rete che richiede particolari sforzi per la tutela e la manutenzione della rete stessa. Prima di tutto, un guasto nel nodo di elaborazione comporta la totale inutilizzabilità della rete. Secondo, l'approccio centralizzato non è scalabile dato che la complessità computazionale del nodo di fusione cresce con il numero di nodi della rete. Infine, in reti di grandi dimensioni i nodi sono organizzati, da un punto di vista topologico, secondo una struttura complessa che è incompatibile con la presenza di un unico centro di raccolta, fusione ed elaborazione. La realizzazione di reti distribuite è stata resa possibile soprattutto grazie al forte e rapido sviluppo delle tecnologie informatiche e di telecomunicazioni, in particolare nel settore delle comunicazioni wireless.

Il nuovo paradigma architetturale dei sistemi di sorveglianza e monitoraggio multiagente offre vantaggi in termini di scalabilità, flessibilità e tolleranza ai guasti (in quanto non ha bisogno di alcuna infrastruttura di coordinamento), ma pone anche alcune nuove questioni che sono attualmente oggetto di ricerca. In primo luogo è necessaria una sostanziale revisione dei metodi di stima, in modo che la fusione distribuita dei dati contrasti gli effetti deleteri del *data incest*, cioè il riuso inconsapevole delle informazioni che circolano nella rete. Il fenomeno del data incest si verifica quando più agenti comunicanti, senza un centro di coordinazione, cercano di stimare lo stato di un sistema dinamico. Infatti, in tal caso, è facile che, a causa della correlazione tra le stime dei diversi agenti, si abbia un degrado delle prestazioni, rispetto ad uno stimatore centralizzato, o se il processo oggetto di monitoraggio è instabile, si abbia addirittura la divergenza dell'errore di stima.

Tra gli approcci studiati per contrastare il fenomeno del *data incest*, meritano particolare attenzione i filtri basati sul consenso [1, 2] che utilizzano il metodo di fusione della *covariance intersection* (intersezione di covarianze) [3, 4]. I filtri basati sul consenso forniscono uno strumento generale per la media distribuita; nel contesto della stima distribuita dello stato, tali filtri sono stati sfruttati per fare la media delle innovazioni [5, 6] o delle stime dello stato [7, 8, 9, 10].

La fusione mediante intersezione di covarianze adotta un approccio a filtro d'informazione, propagando la matrice d'informazione (inversa della covarianza) ed il vettore d'informazione definito premoltiplicando la stima dello stato per la matrice d'informazione, e calcola le coppie d'informazione fuse come combinazione convessa delle coppie d'informazione locali (dell'agente in oggetto e dei suoi agenti vicini). Altri approcci di rilievo per la stima distribuita dello stato sono quelli basati sulla stima a orizzonte mobile [11] e sui processi di diffusione [12].

Per la stima distribuita su reti di sensori esistono vari tipi di algoritmi proposti in letteratura sia per il caso lineare che per quello non lineare. Una prima famiglia di tecniche, a cui si fa riferimento come consenso sulle stime (CE: *Consensus on Estimates*) si basa sull'idea di diffondere l'informazione sulla rete effettuando, ad ogni istante di tempo, un consenso che media le stime o le predizioni dello stato locale [8, 9, 10, 11, 12, 13, 14, 15]. La scelta di fare la media solo tra le stime dello stato ha lo scopo di mantenere lo scambio di informazioni tra i nodi vicini più limitato possibile. Tuttavia, poichè anche le matrici di covarianza contengono informazioni preziose che possono essere utilizzate per migliorare le prestazioni, sono stati proposti altri approcci più elaborati. Ad esempio, in [5, 6, 16], si propone di effettuare un consenso tra le misure locali e la covarianza delle innovazioni così da approssimare, in modo distribuito, la fase di correzione del filtro di Kalman centralizzato. Si farà riferimento a questo approccio come consenso sulla verosimiglianza (CL: *Consensus on Likelihoods*). In [17] è stato adottato un punto di vista diverso, proponendo il cosiddetto consenso sull'informazione a posteriori (CP: *Consensus on Posteriors*). Da un punto di vista algoritmico, nel caso lineare CP non è altro che il consenso applicato all'inversa della matrice di covarianza (matrice d'informazione) e al vettore d'informazione. Dal punto di vista della teoria dell'informazione, l'approccio CP può essere interpretato come consenso sulle PDF nel senso della media di Kullback-Leibler.

Un primo contributo di questo lavoro, capitoli 1-3, riguarda proprio la stima distribuita dello stato di un sistema dinamico su reti di sensori, sia per il caso lineare

che per quello non lineare [18] e [19]. Il nuovo approccio parte dall'osservazione che CL e CP hanno caratteristiche complementari. Infatti, CP garantisce la stabilità per qualunque numero di passi di consenso (anche per un solo passo), ma le sue prestazioni in termini di errore quadratico medio della stima vengono deteriorate per il fatto che la regola di fusione adotta un punto di vista conservativo assumendo che la correlazione tra le stime provenienti da nodi diversi sia completamente sconosciuta. CL, viceversa, fondendo solo l'informazione nuova, non fa alcuna ipotesi conservativa sulla correlazione, ma non garantisce stabilità se il numero di passi di consenso non è sufficientemente elevato. È stata quindi sviluppata una nuova famiglia di algoritmi per la stima distribuita dello stato che effettuano il consenso sia sulla verosimiglianza che sull'informazione a priori, chiamati consenso su verosimiglianza e informazione a priori (CLCP: *Consensus on Likelihoods and Priors*), in cui i due tipi di consenso CL e CP vengono combinati in modo da mantenere le caratteristiche positive di entrambi. Inoltre, viene analizzata la stabilità della famiglia di filtri CLCP nel caso di sistemi lineari, insieme con una discussione su possibili estensioni al caso non lineare.

Facendo seguito a [18], dove vengono presentati risultati preliminari sull'argomento, è stata effettuata una valutazione comparativa delle prestazioni di diversi filtri basati sul consenso su un caso di studio simulativo di *target tracking*, sia con sensori lineari che non lineari [19]. Infine, si evidenzia che, per una particolare scelta dei pesi, il CLCP è equivalente all'algoritmo *Information Weighted Consensus* (IWC), presentato in [20].

Il decentramento, a fronte degli indubbi vantaggi, presenta purtroppo anche debolezze. In particolare, i sistemi net-centrici sono soggetti a minacce intenzionali, come ad esempio l'attacco di un virus informatico. Un secondo contributo presentato in questa tesi riguarda appunto lo studio della vulnerabilità dei sistemi distribuiti a questo tipo di attacchi [21].

La seconda parte di questo lavoro, capitolo 4, si occupa della modellazione di una epidemia di virus su una rete di sensori, concentrandosi sia sui meccanismi di diffusione che sui possibili effetti sulla stima distribuita. Per quanto riguarda la diffusione dei virus, sono stati esaminati vari modelli caratterizzati da diversi livelli di complessità, inclusi *modelli macroscopici* [22]-[23] che descrivono il comportamento complessivo dell'epidemia virale mediante un insieme di equazioni differenziali ordinarie, e *modelli microscopici* [24]-[25] in cui tutti i nodi ed archi della rete sono presi in considerazione separatamente. In ogni caso, tutti i modelli considerati si basano su varianti del modello classico *suscettibile-infetto-rimosso* per la diffusione di un'epidemia in una popolazione. Questa tipologia di modelli di diffusione virale si basa su un'opportuna catena di Markov omogenea. In particolare, si suppone che ogni nodo della rete possa trovarsi in uno dei seguenti stati: *suscettibile* (S), cioè il nodo non è né infetto né immune; *infetto* (I), cioè il nodo è infetto e può trasmettere il virus; *recuperato* (R), cioè il nodo non è infetto ed è immune; *morto* (D), cioè il nodo ha smesso di funzionare a causa dell'infezione e non può infettare altri nodi. Inoltre, si assume che le transizioni tra gli stati possano verificarsi casualmente secondo certe probabilità.

Anche per modellare il comportamento dei virus nei nodi della rete si possono considerare varie alternative: il *modello di virus non-cooperativo*, in cui i nodi infetti smettono di prendere parte all'algoritmo di stima distribuita; il *modello di virus*

a *comportamento deviante*, in cui i nodi infetti prendono parte all'algoritmo di consenso con dati generati a caso, o il *modello di virus malizioso*, in cui i nodi prendono parte al consenso con dati generati appositamente per compromettere del tutto il funzionamento dell'algoritmo. Questo lavoro di tesi prende in considerazione solo il caso di virus non-cooperativo.

Sono state trovate relazioni tra la topologia della rete, la possibilità di un'esplosione epidemica e le prestazioni di algoritmi di fusione dell'informazione distribuiti. In particolare, si vedrà che una caratteristica importante di tutti i modelli di diffusione dei virus considerati è l'esistenza di una *soglia epidemica* che dipende dalla topologia della rete in funzione del *grado medio* o del *raggio spettrale*.

Questo consentirà di fornire misure significative della *robustezza di una rete* ad un'epidemia virale in termini di alcuni parametri topologici della rete (*connettività algebrica, connettività di nodo, grado medio, raggio spettrale, ecc.*) che, a sua volta, sarà la base delle linee guida proposte per la progettazione di una rete di sensori. Il capitolo si conclude con un'estensiva serie di simulazioni su un modello di rete realistico allo scopo di validare l'analisi ed i modelli proposti.

Il terzo contributo di questo lavoro, capitolo 5, riguarda la stima distribuita nel caso in cui gli agenti non abbiano la visione complessiva dello stato del sistema oggetto di monitoraggio ma solo di una sua parte. In generale, in questo caso, gli agenti avranno stati parzialmente sovrapposti e ciascun agente ha il compito di stimare solamente un sottoinsieme dello stato del sistema. Questa situazione si può presentare nel caso di sistemi di sorveglianza se i nodi della rete hanno una visione parziale della regione sorvegliata e/o degli oggetti da sorvegliare (target), ma anche nel caso di sistemi spazialmente distribuiti per i quali, a causa della grandissima estensione del dominio spaziale di interesse, ogni nodo della rete si può occupare esclusivamente di una sottoregione di tale dominio collocata nell'intorno del nodo stesso. Per fare fronte a queste situazioni di notevole rilevanza pratica, vengono proposti tre diversi metodi di fusione che operano rispettivamente sullo *spazio di stato unione* degli agenti (FoU = Fusion over the Union), sullo *spazio di stato locale* (FoLSS = Fusion over the Local State Space) o sullo *spazio di stato intersezione* degli agenti (FoI = Fusion over the Intersection). In attesa di studiare le proprietà di stabilità e consistenza di questi metodi, è stato effettuato uno studio comparativo in due diversi casi di studio, uno lineare ed uno non lineare. Questo studio preliminare, seppur ancora non supportato da risultati di tipo analitico, sembra indicare un buon comportamento dei metodi proposti ed auspicarne un efficace utilizzo nella stima dello stato di sistemi spazialmente distribuiti.

Parte I

Stima Bayesiana distribuita

Capitolo 1

Stima Bayesiana centralizzata

Quando si parla di stima centralizzata si fa riferimento ad un sistema costituito da uno o più sensori, distribuiti nella regione sotto osservazione, connessi con un unico centro di elaborazione che provvede a effettuare le elaborazioni per ottenere la stima. L'approccio Bayesiano trae il suo nome dall'utilizzo del Teorema di Bayes per aggiornare le stime condizionatamente alle osservazioni disponibili.

Si consideri un generico sistema dinamico a tempo-discreto:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \\ \mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{v}_t) \end{cases} \quad (1.1)$$

con

\mathbf{x}_t : stato del sistema;

\mathbf{y}_t : uscita osservata, che costituisce una misura (generalmente indiretta) dello stato del sistema;

\mathbf{u}_t : eventuale ingresso deterministico quale, ad esempio, un segnale di comando impresso al sistema da un operatore oppure da un sistema di controllo automatico;

$\mathbf{w}_t \sim wn(\mathbf{0}, \mathbf{Q}_t)$: disturbo di processo (o di sistema), che solitamente modella i fattori ambientali che interferiscono con il sistema;

$\mathbf{v}_t \sim wn(\mathbf{0}, \mathbf{R}_t)$: disturbo (o rumore) di misura, che tiene conto delle imprecisioni del sensore.

$wn(\mathbf{m}, \mathbf{\Sigma})$ denota un *rumore bianco* con media \mathbf{m} e covarianza $\mathbf{\Sigma}$.

L'obiettivo è quello di determinare, ad ogni istante, una stima $\hat{\mathbf{x}}_t$ dello stato \mathbf{x}_t del sistema.

Si noti che è sufficiente che una tra le funzioni $\mathbf{f}_t(\cdot)$ ed $\mathbf{h}_t(\cdot)$ sia non lineare perché il sistema sia non lineare.

Sia \mathbf{z} una generica variabile, di seguito si indicherà con \mathbf{z}_k il valore assunto da \mathbf{z} al tempo k , e con $\mathbf{z}^k = \{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_k\}$ la sequenza dei valori assunti da \mathbf{z} dall'istante iniziale all'istante k . Nel seguito si indicherà con $\hat{\mathbf{x}}_{t|k}$ la stima dello stato \mathbf{x}_t basata sulle osservazioni \mathbf{y}^k

Il procedimento di stima consta di due fasi:

- 1) **Correzione** (detta anche *filtraggio* o *aggiornamento alla misura*): partendo dalla stima $\hat{\mathbf{x}}_{t|t-1}$ e data la nuova misura, $\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{v}_t)$, si determina la stima aggiornata $\hat{\mathbf{x}}_{t|t}$ detta anche *stima corretta* o *filtrata*.
- 2) **Predizione** (o *aggiornamento temporale*): partendo da $\hat{\mathbf{x}}_{t|t}$ ed utilizzando il modello di stato $\mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)$ si determina la *stima predetta* o *predizione* $\hat{\mathbf{x}}_{t+1|t}$.

In ambito meteorologico ed oceanografico la correzione è chiamata *analisi* mentre la predizione è chiamata *forecast* o *simulazione*. In tali ambiti, l'intero processo di stima viene chiamato *assimilazione dati*.

1.1 Predizione e correzione nella stima Bayesiana

Secondo l'approccio Bayesiano [26], il problema di stima viene interpretato come determinazione della densità di probabilità (*PDF: Probability Density Function*) della quantità da stimare condizionatamente alle osservazioni. Con riferimento alla stima dello stato del sistema (1.1) si definiscono le seguenti PDF:

- $p_{t|t}(\mathbf{x}) = p(\mathbf{x}_t | \mathbf{y}^t)$ *PDF filtrata o corretta*
- $p_{t+1|t}(\mathbf{x}) = p(\mathbf{x}_{t+1} | \mathbf{y}^t)$ *PDF predetta.*

Nota la PDF, per determinare la stima si possono poi utilizzare diversi criteri, quali ad esempio il criterio del minimo errore quadratico medio (*MMSE: Minimum Mean Square Error*) oppure il criterio della massima probabilità a posteriori (*MAP: Max A Posteriori Probability*), così definiti:

- criterio MMSE: $\hat{\mathbf{x}}_{t|t} = E[\mathbf{x}_t | \mathbf{y}^t] = \int \mathbf{x} p_{t|t}(\mathbf{x}) d\mathbf{x}$
- criterio MAP: $\hat{\mathbf{x}}_{t|t} = \arg \max_{\mathbf{x}} p_{t|t}(\mathbf{x})$

e la corrispondente varianza sarà:

$$\mathbf{P}_{t|t} = \int (\mathbf{x} - \hat{\mathbf{x}}_{t|t})(\mathbf{x} - \hat{\mathbf{x}}_{t|t})^\top p_{t|t}(\mathbf{x}) d\mathbf{x}.$$

Quindi, partendo dalla PDF iniziale $p_{1|0}(\mathbf{x})$, la stima Bayesiana consta delle seguenti due fasi:

- **Correzione:** $p_{t|t-1}(\mathbf{x}) \xrightarrow[\text{eq. di misura}]{\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{v}_t)} p_{t|t}(\mathbf{x})$
- **Predizione:** $p_{t|t}(\mathbf{x}) \xrightarrow[\text{modello di stato}]{\mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)} p_{t+1|t}(\mathbf{x})$.

Nel prossimo paragrafo si determinano le equazioni ricorsive per le suddette PDF.

1.1.1 Correzione

Data $p_{t|t-1}(\mathbf{x}) = p(\mathbf{x}_t | \mathbf{y}^{t-1})$ si vuole determinare $p_{t|t}(\mathbf{x}) = p(\mathbf{x}_t | \mathbf{y}^t)$ utilizzando la misura $\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{v}_t)$. A tale proposito si utilizza la proprietà:

$$p(a | b, c) = \frac{p(b | a, c) p(a | c)}{p(b | c)}$$

che deriva da $p(a | b, c) p(b | c) = p(a, b | c) = p(b | a, c) p(a | c)$. Ponendo $a = \mathbf{x}_t$, $b = \mathbf{y}_t$ e $c = \mathbf{y}^{t-1}$, si ha

$$p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{y}^{t-1}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t, \mathbf{y}^{t-1}) p(\mathbf{x}_t | \mathbf{y}^{t-1})}{p(\mathbf{y}_t | \mathbf{y}^{t-1})}.$$

Si noti che:

- $p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{y}^{t-1}) = p_{t|t}(\mathbf{x})$
- $p(\mathbf{x}_t | \mathbf{y}^{t-1}) = p_{t|t-1}(\mathbf{x})$
- $p(\mathbf{y}_t | \mathbf{x}_t, \mathbf{y}^{t-1}) = p(\mathbf{y}_t | \mathbf{x}_t)$, detta *verosimiglianza* (*likelihood*),

infatti, il condizionamento in $p(\mathbf{y}_t | \mathbf{x}_t, \mathbf{y}^{t-1})$ rispetto a \mathbf{y}^{t-1} è irrilevante dato che la stessa è condizionata anche rispetto a \mathbf{x}_t . Quindi la verosimiglianza è una PDF indicata con

$$l_t(\mathbf{x}, \mathbf{y}_t) = p(\mathbf{y}_t = \mathbf{y} | \mathbf{x}_t = \mathbf{x}).$$

Si ha

$$p_{t|t}(\mathbf{x}) \propto l_t(\mathbf{x}, \mathbf{y}_t) \cdot p_{t|t-1}(\mathbf{x}).$$

Quanto più \mathbf{x} è verosimile con l'osservazione \mathbf{y}_t , tanto maggiore sarà $p_{t|t}(\mathbf{x})$. Il prodotto $l_t(\cdot) \cdot p_{t|t-1}(\cdot)$ sposta la PDF risultante, $p_{t|t}(\cdot)$, verso gli stati più verosimili, vedi figura 1.1. Risulta

$$p_{t|t}(\mathbf{x}) = \frac{l_t(\mathbf{x}, \mathbf{y}_t) \cdot p_{t|t-1}(\mathbf{x})}{c}$$

dove c è una costante tale che

$$\int \frac{l_t(\boldsymbol{\xi}, \mathbf{y}_t) \cdot p_{t|t-1}(\boldsymbol{\xi})}{c} d\boldsymbol{\xi} = 1 \quad \implies \quad c = \int l_t(\boldsymbol{\xi}, \mathbf{y}_t) \cdot p_{t|t-1}(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

Più $l_t(\cdot, \mathbf{y}_t)$ è stretta, cioè migliore è la qualità del sensore (\mathbf{R} piccola), più $l_t(\cdot, \mathbf{y}_t)$ è informativa.

L'equazione ricorsiva risultante per la correzione è quindi:

$$p_{t|t}(\mathbf{x}) = \frac{l_t(\mathbf{x}, \mathbf{y}_t) \cdot p_{t|t-1}(\mathbf{x})}{\int l_t(\boldsymbol{\xi}, \mathbf{y}_t) \cdot p_{t|t-1}(\boldsymbol{\xi}) d\boldsymbol{\xi}}. \quad (1.2)$$

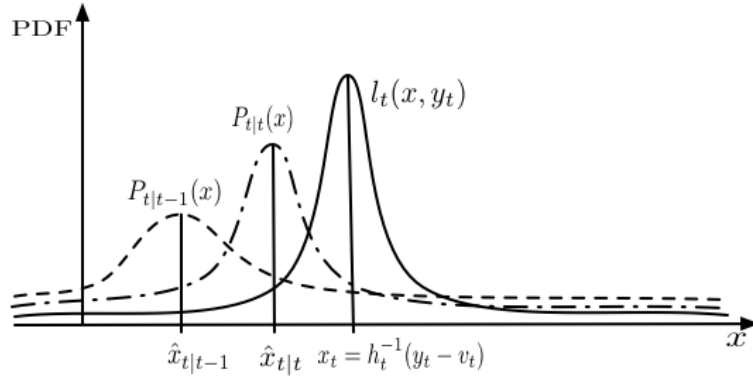


Figura 1.1: Effetto della verosimiglianza nella fase di correzione.

1.1.2 Predizione

Data $p_{t|t}(\mathbf{x}) = p(\mathbf{x}_t | \mathbf{y}^t)$, utilizzando $\mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{w}_t)$, si vuole determinare $p_{t+1|t}(\mathbf{x}) = p(\mathbf{x}_{t+1} | \mathbf{y}^t)$.

A tale proposito, si determina la PDF congiunta di \mathbf{x}_t ed \mathbf{x}_{t+1} e quindi la si marginalizza rispetto a \mathbf{x}_t :

$$p_{t+1|t}(\mathbf{x}) = p(\mathbf{x}_{t+1} | \mathbf{y}^t) = \int p(\mathbf{x}_{t+1}, \mathbf{x}_t | \mathbf{y}^t) d\mathbf{x}_t.$$

Sfruttando la proprietà: $p(a | c) = \int p(a, b | c) db = \int p(a | b, c) p(b | c) db$, con $a = \mathbf{x}_{t+1}$, $b = \mathbf{x}_t$ e $c = \mathbf{y}^t$, si ha

$$p_{t+1|t}(\mathbf{x}) = p(\mathbf{x}_{t+1} | \mathbf{y}^t) = \int p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{y}^t) p(\mathbf{x}_t | \mathbf{y}^t) d\mathbf{x}_t$$

Considerando $p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{y}^t)$, il condizionamento rispetto a \mathbf{x}_t rende irrilevante il condizionamento rispetto a \mathbf{y}^t , cioè

$$p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{y}^t) = p(\mathbf{x}_{t+1} | \mathbf{x}_t)$$

e, posto $\xi = \mathbf{x}_t$, si ha

$$p_{t+1|t}(\mathbf{x}) = \int \underbrace{p(\mathbf{x} | \xi)}_{\text{PDF di transizione}} \underbrace{p(\xi | \mathbf{y}^t)}_{p_{t|t}(\xi)} d\xi.$$

La quantità $p(\mathbf{x}_{t+1} | \mathbf{x}_t)$ descrive la transizione della PDF da t a $t + 1$ e viene per questo detta *PDF di transizione* ed indicata con la notazione $\varphi_{t+1|t}(\mathbf{x} | \xi)$,

$$\varphi_{t+1|t}(\mathbf{x} | \xi) \triangleq p(\mathbf{x}_{t+1} = \mathbf{x} | \mathbf{x}_t = \xi) \quad (1.3)$$

La PDF di transizione, $\varphi(\cdot)$, dipende da:

- il modello di evoluzione dello stato $\mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)$;
- la PDF del rumore di processo \mathbf{w}_t : $p_{\mathbf{w}_t}(\cdot)$.

L'equazione ricorsiva risultante per la predizione è:

$$p_{t+1|t}(\mathbf{x}) = \int \varphi_{t+1|t}(\mathbf{x} | \xi) p_{t|t}(\xi) d\xi. \quad (1.4)$$

1.1.3 Algoritmi per la stima Bayesiana

Nell'effettuare la stima i passi di predizione e correzione possono essere scambiati a seconda delle esigenze. L'Algoritmo 1, di seguito riportato, riassume un ciclo temporale del filtro Bayesiano nel caso in cui la stima ricorsiva viene effettuata prima con il passo di correzione e poi con quello di predizione.

Algoritmo 1: Algoritmo di stima Bayesiana.

Dati: PDF a priori $p_{1|0}(\cdot)$

for $t = 1, 2, \dots$ **do**

Correzione

si acquisisce la misura \mathbf{y}_t

$$p_{t|t}(\mathbf{x}) = \frac{l_t(\mathbf{x}, \mathbf{y}_t) \cdot p_{t|t-1}(\mathbf{x})}{\int l_t(\boldsymbol{\xi}, \mathbf{y}_t) \cdot p_{t|t-1}(\boldsymbol{\xi}) d\boldsymbol{\xi}}$$

Predizione

$$p_{t+1|t}(\mathbf{x}) = \int \varphi_{t+1|t}(\mathbf{x} | \boldsymbol{\xi}) p_{t|t}(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

end for

1.2 PDF di transizione e verosimiglianza

La determinazione della PDF di transizione $\varphi_{t|t-1}(\mathbf{x}|\boldsymbol{\xi})$ e della verosimiglianza $l_t(\mathbf{x}, \mathbf{y}_t)$ è generalmente complessa. Si assume, come è realistico supporre in molti contesti pratici, che i disturbi siano *additivi*, cioè che il sistema dinamico a tempo-discreto sia della forma:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t \\ \mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t) + \mathbf{v}_t \end{cases} \quad (1.5)$$

In tal caso si possono ricavare facilmente i disturbi conoscendo le altre variabili, cioè:

$$\begin{aligned} \mathbf{w}_t &= \mathbf{x}_{t+1} - \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t) \\ \mathbf{v}_t &= \mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t) \end{aligned} \quad (1.6)$$

Si possono così ricavare la PDF di transizione dello stato

$$\varphi_{t+1|t}(\mathbf{x} | \boldsymbol{\xi}) = p_{\mathbf{w}_t}(\mathbf{x} - \mathbf{f}_t(\boldsymbol{\xi}, \mathbf{u}_t)) \quad (1.7)$$

e la verosimiglianza

$$l_t(\mathbf{x}, \mathbf{y}_t) = p_{\mathbf{v}_t}(\mathbf{y}_t - \mathbf{h}_t(\mathbf{x})). \quad (1.8)$$

In generale, la soluzione della ricorsione Bayesiana è *infinito-dimensionale*, cioè non è esprimibile con un numero finito di parametri, per cui è impossibile da determinare analiticamente. Vi sono però alcune eccezioni, di seguito elencate.

1. Caso lineare Gaussiano, caratterizzato da funzioni $\mathbf{f}_t(\mathbf{x}, \mathbf{u}, \mathbf{w})$ e $\mathbf{h}_t(\mathbf{x}, \mathbf{v})$ lineari nonché da PDF $p_{1|0}(\cdot)$, $p_{\mathbf{w}_t}(\cdot)$ e $p_{\mathbf{v}_t}(\cdot)$ Gaussiane.

2. Caso lineare uniforme in assenza di rumore di processo [27].
3. Caso in cui lo stato \mathbf{x}_t è discreto e finito, i.e. $\mathbf{x}_t \in \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$ con $k < \infty$.
4. Alcune non linearità speciali, cioè funzioni $\mathbf{f}_t(\cdot)$ e $\mathbf{h}_t(\cdot)$ particolari, e.g. filtro di Benes [28] e filtro di Daum [29] che lo generalizza.

Caso di disturbi Gaussiani

Se si fa l'ulteriore ipotesi che i disturbi \mathbf{w}_t e \mathbf{v}_t siano Gaussiani, cioè

$$\begin{aligned} p_{\mathbf{w}_t}(\mathbf{w}) &= \mathcal{N}(\cdot, \mathbf{0}, \mathbf{Q}_t) = \frac{1}{(\det(2\pi\mathbf{Q}_t))^{1/2}} \exp\left(-\frac{1}{2}\mathbf{w}^\top \mathbf{Q}_t^{-1} \mathbf{w}\right) \\ p_{\mathbf{v}_t}(\mathbf{v}) &= \mathcal{N}(\cdot, \mathbf{0}, \mathbf{R}_t) = \frac{1}{(\det(2\pi\mathbf{R}_t))^{1/2}} \exp\left(-\frac{1}{2}\mathbf{v}^\top \mathbf{R}_t^{-1} \mathbf{v}\right) \end{aligned} \quad (1.9)$$

si possono ricavare le seguenti espressioni analitiche per la PDF di transizione dello stato e per la verosimiglianza

$$\begin{aligned} \varphi_{t+1|t}(\mathbf{x} | \boldsymbol{\xi}) &= \frac{1}{(\det(2\pi\mathbf{Q}_t))^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{f}_t(\boldsymbol{\xi}, \mathbf{u}_t))^\top \mathbf{Q}_t^{-1} (\mathbf{x} - \mathbf{f}_t(\boldsymbol{\xi}, \mathbf{u}_t))\right) \end{aligned} \quad (1.10)$$

$$l_t(\mathbf{x}, \mathbf{y}_t) = \frac{1}{(\det(2\pi\mathbf{R}_t))^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}))^\top \mathbf{R}_t^{-1} (\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}))\right) \quad (1.11)$$

1.3 Caso lineare Gaussiano: Filtro di Kalman

Si consideri un sistema dinamico a tempo-discreto *lineare*, eventualmente tempo-variante

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{D}_t \mathbf{w}_t \\ \mathbf{y}_t = \mathbf{C}_t \mathbf{x}_t + \mathbf{v}_t \end{cases} \quad (1.12)$$

che deriva dal sistema (1.1), nel caso in cui sia $\mathbf{f}_t(\cdot)$ che $\mathbf{h}_t(\cdot)$ sono lineari. Se

- $p_{1|0}(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{1|0}, \mathbf{P}_{1|0})$
- $\mathbf{w}_t, \mathbf{v}_t$ sono congiuntamente Gaussiani ed incorrelati, i.e. $\begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix} \sim \mathcal{N}\left(\cdot; \mathbf{0}, \begin{bmatrix} \mathbf{Q}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_t \end{bmatrix}\right)$

allora, per ogni $t \geq 1$, $p_{t|t}(\mathbf{x})$ e $p_{t|t-1}(\mathbf{x})$ risultano Gaussiane, vale a dire

- $p_{t|t-1}(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$
- $p_{t|t}(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$

con medie e covarianze $\hat{\mathbf{x}}_{t|t-1}, \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t-1}, \mathbf{P}_{t|t}$, che caratterizzano completamente tali PDF fornite dal Filtro di Kalman (KF, *Kalman Filter*). KF è soluzione ottima nel senso MMSE della stima Bayesiana nel caso di un sistema dinamico *lineare*

Gaussiano. Le equazioni ricorsive del Filtro di Kalman sono riportate in dettaglio nell'Algoritmo 2.

Algoritmo 2: Algoritmo del Filtro di Kalman.

Dati: $\hat{\mathbf{x}}_{1|0}$ e $\mathbf{P}_{1|0}$, se non si hanno informazioni si può porre $\hat{\mathbf{x}}_{1|0} = \mathbf{0}$ e $\mathbf{P}_{1|0} = \alpha \mathbf{I}_n$ con $\alpha \gg 1$ e \mathbf{I}_n matrice identità di ordine n

for $t = 1, 2, \dots$ **do**

Correzione

si acquisisce la misura \mathbf{y}_t

$\mathbf{e}_t = \mathbf{y}_t - \mathbf{C}_t \hat{\mathbf{x}}_{t|t-1}$; %errore di stima dell'uscita o *innovazione*

$\mathbf{S}_t = \mathbf{R}_t + \mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}_t^\top$; %covarianza del processo delle innovazioni

$\mathbf{L}_t = \mathbf{P}_{t|t-1} \mathbf{C}_t^\top \mathbf{S}_t^{-1}$; %guadagno di Kalman

$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{L}_t \mathbf{e}_t$

$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{L}_t \mathbf{S}_t \mathbf{L}_t^\top$

$\mathbf{u}_t = g(\hat{\mathbf{x}}_{t|t}, \dots)$; %in un sistema di controllo qui viene calcolato u_t

Predizione

$\hat{\mathbf{x}}_{t+1|t} = \mathbf{A}_t \hat{\mathbf{x}}_{t|t} + \mathbf{B}_t \mathbf{u}_t$

$\mathbf{P}_{t+1|t} = \mathbf{A}_t \mathbf{P}_{t|t} \mathbf{A}_t^\top + \mathbf{D}_t \mathbf{Q}_t \mathbf{D}_t^\top$

end for

Di seguito si elencano alcune delle proprietà significative del Filtro di Kalman.

- Il filtro di Kalman è non-polarizzato, cioè $E[\hat{\mathbf{x}}_{t|t}] = E[\mathbf{x}_t]$.
- **Stabilità:** sotto opportune ipotesi di osservabilità, la covarianza dell'errore di stima resta limitata anche se il sistema è instabile, cioè le successioni di matrici $\mathbf{P}_{t|t}$ e $\mathbf{P}_{t+1|t}$ sono limitate.
- **Ottimalità:** nel caso lineare Gaussiano, il filtro di Kalman è ottimo sia nel senso MMSE che MAP. Nel caso semplicemente lineare (non Gaussiano) il filtro di Kalman è, comunque, lo stimatore ottimo MMSE tra tutti i filtri lineari, anche se possono esistere filtri non lineari migliori.
- Il filtro di Kalman è ricorsivo nel senso che, all'istante t , è sufficiente conoscere $\hat{\mathbf{x}}_{t-1|t-1}$ e $\mathbf{P}_{t-1|t-1}$, e non i valori precedentemente assunti da queste variabili, per poter proseguire nell'elaborazione.
- Se $\dim(\mathbf{x}) = n$, il filtro di Kalman richiede un impiego di memoria $O(n^2)$ ed una complessità computazionale $O(n^3)$ per ogni ciclo temporale.

Il vero parametro di sintonizzazione del filtro di Kalman è la covarianza del rumore di processo \mathbf{Q}_t . Anche se talvolta non si hanno informazioni su \mathbf{Q}_t , assegnandogli un valore opportuno è possibile quantificare l'incertezza sul modello.

La matrice \mathbf{Q} ha anche un'interpretazione frequenziale. Anzitutto si osserva che le prestazioni del filtro non dipendono singolarmente dal valore assoluto delle componenti di \mathbf{Q} bensì dal loro valore relativo rispetto alle componenti della matrice di covarianza del rumore di misura \mathbf{R} . In altri termini, riscaldando della stessa

quantità positiva le matrici \mathbf{Q}_t ed \mathbf{R}_t non produce alcun effetto sulla successione delle stime $\hat{\mathbf{x}}_{t|t}$ e $\hat{\mathbf{x}}_{t+1|t}$ ma solo una analoga riscalatura delle matrici di covarianza $\mathbf{P}_{t|t}$ e $\mathbf{P}_{t+1|t}$. Tenendo conto che \mathbf{Q}_t ed \mathbf{R}_t hanno, in generale, dimensioni diverse si può comunque affermare che:

- se $\mathbf{Q} \gg \mathbf{R}$ (nel senso che le componenti diagonali di \mathbf{Q} sono molto maggiori delle componenti diagonali di \mathbf{R}) si ha un *filtro a banda elevata*, ovvero che risponde rapidamente alle misure
- se $\mathbf{Q} \ll \mathbf{R}$ (nel senso che le componenti diagonali di \mathbf{Q} sono molto minori delle componenti diagonali di \mathbf{R}), si ha un *filtro a banda stretta*, ovvero che risponde lentamente alle misure.

1.4 Stima dello stato di un sistema non lineare

Si consideri un sistema dinamico non lineare a tempo-discreto, eventualmente tempo-variante, con disturbi *additivi*

$$\begin{cases} \mathbf{x}_{t+1} &= \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t \\ \mathbf{y}_t &= \mathbf{h}_t(\mathbf{x}_t) + \mathbf{v}_t. \end{cases} \quad (1.13)$$

Nel caso di sistemi non lineari, una soluzione analitica del problema di filtraggio Bayesiano non è disponibile ma esistono diversi metodi approssimati (subottimi) di stima, tra i quali i *metodi di linearizzazione*, i *metodi di campionamento della PDF* ed i *metodi a modelli multipli*. Storicamente la prima estensione del Filtro di Kalman al caso non lineare è l'Extended Kalman Filter (EKF), ideato e sviluppato presso la NASA per le missioni Apollo [30]. EKF linearizza, ad ogni istante temporale, il sistema in un intorno della stima disponibile a quell'istante ed applica il filtro di Kalman standard al sistema linearizzato. Si tratta di un metodo semplice che, tuttavia, non sempre fornisce prestazioni soddisfacenti in quanto si basa sulla linearizzazione nell'intorno di una stima che potrebbe essere lontana dallo stato effettivo. I metodi di campionamento della PDF approssimano la PDF continua con una discreta. L'approssimazione può essere effettuata in modo deterministico come nel caso dell'*Unscented Kalman Filter* (UKF) oppure in modo stocastico come nel *filtro a particelle* (PF= Particle Filter), detto anche *Filtro Sequenziale Monte Carlo* (SMCF=Sequential Monte Carlo Filter). Infine i metodi a *modelli multipli* o a *mistura di Gaussiane* approssimano un sistema non lineare con più modelli lineari, a ciascuno dei quali corrisponde una componente Gaussiana. Ne risulta una PDF che è una combinazione lineare convessa (mistura) di Gaussiane. Di seguito, si esamineranno in maggior dettaglio i metodi di linearizzazione e quelli di campionamento della PDF.

1.4.1 Extended Kalman Filter

In particolare, il metodo EKF prevede lo sviluppo in serie di Taylor al primo ordine delle funzioni $\mathbf{f}_t(\cdot)$ e $\mathbf{h}_t(\cdot)$. La funzione di transizione ad un passo dello

stato $\mathbf{f}_t(\cdot)$ viene usata per fare predizione quando è nota la stima $\hat{\mathbf{x}}_{t|t}$, per cui viene sviluppata in un intorno di $\hat{\mathbf{x}}_{t|t}$ come segue

$$\mathbf{x}_{t+1} = \mathbf{f}_t(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t) + \frac{\partial \mathbf{f}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t) \cdot (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}) + \cdots + \mathbf{w}_t$$

Viceversa, la funzione di misura $\mathbf{h}_t(\cdot)$ viene usata per fare correzione quando è nota la stima $\hat{\mathbf{x}}_{t|t-1}$; pertanto viene sviluppata in un intorno di $\hat{\mathbf{x}}_{t|t-1}$, cioè

$$\mathbf{y}_t = \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1}) + \frac{\partial \mathbf{h}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t-1}) \cdot (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \cdots + \mathbf{v}_t.$$

Si devono quindi calcolare i Jacobiani:

$$\mathbf{A}_t = \frac{\partial \mathbf{f}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t)$$

prima della predizione e

$$\mathbf{C}_t = \frac{\partial \mathbf{h}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t-1})$$

prima della correzione. Si noti che le matrici \mathbf{A}_t e \mathbf{C}_t risultano tempo-varianti anche se il sistema non lineare originario è tempo-invariante poiché gli Jacobiani devono essere ricalcolati ad ogni istante t . Di seguito si esaminano in dettaglio le fasi di correzione e predizione di EKF.

Correzione: l'innovazione vale

$$\mathbf{e}_t = \mathbf{y}_t - \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1}) = \mathbf{C}_t (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \cdots + \mathbf{v}_t, \quad (\text{con } \mathbf{C}_t: \text{ Jacobiano})$$

mentre il guadagno di correzione \mathbf{L}_t viene calcolato come nel filtro di Kalman standard utilizzando la matrice \mathbf{C}_t calcolata come Jacobiano della funzione di misura, cioè

$$\begin{aligned} \mathbf{S}_t &= \mathbf{R}_t + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^\top \\ \mathbf{L}_t &= \mathbf{P}_t \mathbf{C}_t^\top \mathbf{S}_t^{-1} \\ \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{L}_t \mathbf{S}_t \mathbf{L}_t^\top \end{aligned}$$

Nella correzione della stima

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{L}_t (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})$$

si approssima $\hat{\mathbf{y}}_{t|t-1}$ ponendo $\hat{\mathbf{y}}_{t|t-1} \cong \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1})$, da cui

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{L}_t (\mathbf{y}_t - \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1}))$$

Predizione: Nella predizione della stima, si approssima $\hat{\mathbf{x}}_{t+1|t}$ ponendo

$$\hat{\mathbf{x}}_{t+1|t} \cong \mathbf{f}_t(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t).$$

Per l'aggiornamento temporale della matrice di covarianza $\mathbf{P}_{t+1|t} = E[\tilde{\mathbf{x}}_{t+1|t} \tilde{\mathbf{x}}_{t+1|t}^\top]$, si fa uso della linearizzazione

$$\tilde{\mathbf{x}}_{t+1|t} = \mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t} \cong \mathbf{A}_t \tilde{\mathbf{x}}_{t|t} + \mathbf{w}_t$$

da cui si deduce la stessa equazione di aggiornamento del filtro di Kalman standard

$$\mathbf{P}_{t+1|t} = \mathbf{A}_t \mathbf{P}_{t|t} \mathbf{A}_t^\top + \mathbf{Q}_t.$$

in cui la matrice \mathbf{A}_t è calcolata come Jacobiano della funzione di transizione dello stato.

EKF è riassunto nell'Algoritmo 3.

Algoritmo 3: Algoritmo EKF nella forma di covarianza.

Dati: $\hat{\mathbf{x}}_{1|0}$ e $\mathbf{P}_{1|0}$, se non si hanno informazioni si può porre $\hat{\mathbf{x}}_{1|0} = \mathbf{0}$ e $\mathbf{P}_{1|0} = \alpha \mathbf{I}_n$ con $\alpha \gg 1$ e \mathbf{I}_n matrice identità di ordine n

Calcolo preliminare dei Jacobiani

$$\mathbf{A}_t = \frac{\partial \mathbf{f}_t}{\partial \mathbf{x}}(\cdot)$$

$$\mathbf{C}_t = \frac{\partial \mathbf{h}_t}{\partial \mathbf{x}}(\cdot)$$

for $t = 1, 2, \dots$ **do**

Correzione

si acquisisce la misura \mathbf{y}_t

$$\mathbf{C}_t = \frac{\partial \mathbf{h}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t-1})$$

$\mathbf{e}_t = \mathbf{y}_t - \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1});$ %innovazione

$\mathbf{S}_t = \mathbf{R}_t + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^\top;$ %covarianza del processo delle innovazioni

$\mathbf{L}_t = \mathbf{P}_t \mathbf{C}_t^\top \mathbf{S}_t^{-1};$ %guadagno di Kalman

$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{L}_t \mathbf{e}_t$

$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{L}_t \mathbf{S}_t \mathbf{L}_t^\top$

$\mathbf{u}_t = \mathbf{g}(\hat{\mathbf{x}}_{t|t}, \dots);$ %in un sistema di controllo qui viene calcolato \mathbf{u}_t

Predizione

$$\mathbf{A}_t = \frac{\partial \mathbf{f}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t)$$

$\hat{\mathbf{x}}_{t+1|t} = \mathbf{f}_t(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t)$

$\mathbf{P}_{t+1|t} = \mathbf{A}_t \mathbf{P}_{t|t} \mathbf{A}_t^\top + \mathbf{Q}_t$

end for

Esistono risultati di stabilità locale relativi a EKF [31], anche se questo algoritmo fornisce solitamente prestazioni inferiori ai metodi di campionamento della PDF e ai filtri a modelli multipli.

1.4.2 Unscented Kalman Filter

L'Unscented Kalman Filter è un metodo di campionamento deterministico che utilizza la cosiddetta *trasformazione unscented* (UT= Unscented Transform), una tecnica efficace per propagare i momenti di una variabile aleatoria (v.a.) attraverso una trasformazione non lineare. Data una v.a. $\mathbf{X} \sim (\bar{\mathbf{x}}, \mathbf{P}_\mathbf{x})$, di cui sono note media e varianza, se ne vuole approssimare la distribuzione con un numero finito di campioni, i *punti- σ* \mathbf{x}_i , la cui media e varianza campionaria coincidano esattamente con $\bar{\mathbf{x}}$ e, rispettivamente, $\mathbf{P}_\mathbf{x}$. Si dimostra che a tale proposito sono sufficienti $n + 1$ campioni, dove n è la dimensione di \mathbf{X} , anche se, per motivi di simmetria, se ne utilizzano solitamente $2n + 1$ indicati con \mathbf{x}_i per $i = -n, \dots, 0, \dots, n$, con relativi

pesi w_i . L'uguaglianza di media e covarianza campionarie dei punti- σ con media e covarianza della v.a. \mathbf{X} impone le relazioni:

$$\bar{\mathbf{x}} = \sum_{i=-n}^n w_i \mathbf{x}_i \quad (1.14)$$

$$\mathbf{P}_{\mathbf{x}} = \sum_{i=-n}^n w_i (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^\top \quad (1.15)$$

Considerando una generica trasformazione non lineare $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ della v.a. \mathbf{X} , ovvero

$$\mathbf{X} \sim (\bar{\mathbf{x}}, \mathbf{P}_{\mathbf{x}}) \xrightarrow{\mathbf{g}(\cdot)} \mathbf{Y} = \mathbf{g}(\mathbf{X}) \sim (\bar{\mathbf{y}}, \mathbf{P}_{\mathbf{y}})$$

si possono determinare i campioni trasformati

$$\mathbf{y}_i = \mathbf{g}(\mathbf{x}_i) \quad i = -n, \dots, 0, \dots, n$$

da cui si possono ricavare, in modo approssimato, i primi due momenti, media e varianza, della distribuzione di \mathbf{Y} tramite:

$$\bar{\mathbf{y}} \cong \sum_{i=-n}^n w_i \mathbf{y}_i \quad (1.16)$$

$$\mathbf{P}_{\mathbf{y}} \cong \sum_{i=-n}^n w_i (\mathbf{y}_i - \bar{\mathbf{y}}) (\mathbf{y}_i - \bar{\mathbf{y}})^\top \quad (1.17)$$

Si può dimostrare che se la distribuzione discreta $\{\mathbf{x}_i\}_{i=0}^{2n}$ fornisce esattamente i primi due momenti, media e varianza, di \mathbf{X} , allora la corrispondente distribuzione discreta $\{\mathbf{y}_i\}_{i=0}^{2n}$ approssima correttamente almeno al secondo ordine media e varianza di \mathbf{Y} . Si noti che, viceversa, EKF approssima tali momenti solo al primo ordine.

Generazione dei campioni

Data la v.a. $\mathbf{X} \sim (\bar{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$, in primo luogo si fattorizza la sua matrice di covarianza come

$$\mathbf{P}_{\mathbf{x}} = \mathbf{\Gamma} \mathbf{\Gamma}^\top = \sum_{i=1}^n \gamma_i \gamma_i^\top \quad (1.18)$$

dove $\mathbf{\Gamma} = [\gamma_1, \gamma_2, \dots, \gamma_n]$. Si può ottenere questa fattorizzazione in svariati modi, quali ad esempio:

- la fattorizzazione di Cholesky in cui $\mathbf{\Gamma}$ coincide con una matrice triangolare, inferiore o superiore;
- la decomposizione a valori singolari (autovalori-autovettori) $\mathbf{P}_{\mathbf{x}} = \mathbf{T} \mathbf{\Lambda} \mathbf{T}^\top$, con \mathbf{T} matrice ortogonale degli autovettori e $\mathbf{\Lambda}$ matrice diagonale degli autovalori, che fornisce $\mathbf{\Gamma} = \mathbf{T} \mathbf{\Lambda}^{1/2}$;
- la radice quadrata simmetrica $\mathbf{\Gamma} = \sqrt{\mathbf{P}_{\mathbf{x}}}$.

Confrontando (1.17) con (1.18) si ha:

$$\sum_{i=-n}^n w_i (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^\top = \sum_{i=1}^n \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^\top \quad (1.19)$$

Una possibile scelta dei punti- σ per imporre (1.19) è quindi

$$\mathbf{x}_i - \bar{\mathbf{x}} = \begin{cases} \alpha_i \boldsymbol{\gamma}_i, & i > 0 \\ \mathbf{0}, & i = 0 \\ -\alpha_i \boldsymbol{\gamma}_i, & i < 0 \end{cases} \implies \mathbf{x}_i = \begin{cases} \bar{\mathbf{x}} + \alpha_i \boldsymbol{\gamma}_i, & i > 0 \\ \bar{\mathbf{x}}, & i = 0 \\ \bar{\mathbf{x}} - \alpha_i \boldsymbol{\gamma}_i, & i < 0 \end{cases} \quad (1.20)$$

Sostituendo (1.20) in (1.19) e ponendo $w_i = w_{-i}$, si ha

$$\sum_{i=1}^n 2w_i \alpha_i^2 \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^\top = \sum_{i=1}^n \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^\top$$

da cui si deduce la relazione $2w_i \alpha_i^2 = 1$ e quindi

$$\alpha_i = \frac{1}{\sqrt{2w_i}} \quad i = 1, 2, \dots, n \quad (1.21)$$

Sostituendo (1.20) in (1.16), l'uguaglianza della media comporta la relazione

$$\sum_{i=-n}^n w_i = 1 \quad (1.22)$$

È prassi comune scegliere tutti i pesi (eccetto eventualmente w_0) uguali, cioè

$$w_{\pm i} = w \quad i = 1, 2, \dots, n$$

per cui la condizione (1.22) diventa $w_0 + 2nw = 1$. Pertanto si può scegliere il peso centrale w_0 e, conseguentemente, ricavare i pesi non centrali mediante $w = (1 - w_0)/2n$. Alternativamente si possono scegliere i pesi non centrali w e determinare conseguentemente $w_0 = 1 - 2nw$. In letteratura [32], sono proposte scelte del peso centrale w_0 che permettono di influenzare i momenti di ordine superiore della distribuzione dei punti- σ senza alterarne la media e la covarianza. Una buona scelta, in situazioni in cui non si ha nessuna conoscenza a-priori sul tipo di distribuzione della variabile da stimare, è:

$$w_0 = 0, \quad w = \frac{1}{2n} \implies \alpha_i = \alpha = \sqrt{n} \quad i = 1, \dots, n$$

È anche buona norma, per garantire consistenza cioè che la covarianza calcolata non sottostimi quella effettiva, utilizzare pesi w_i^c diversi per il calcolo della covarianza; una scelta consigliata in tal senso è

$$w_0^c = 2, \quad w_i^c = \frac{1}{2n} \quad i = \pm 1, \dots, \pm n$$

Riassumendo i precedenti sviluppi, la trasformazione unscented (UT) viene riportata nell'Algoritmo 4. Si noti che per il suo utilizzo nella stima è spesso richiesto anche il calcolo della cross-covarianza $\mathbf{P}_{\mathbf{xy}}$ fra le v.a. \mathbf{X} e $\mathbf{Y} = \mathbf{g}(\mathbf{X})$.

Algoritmo 4: Algoritmo UT.

Dati: $\bar{\mathbf{x}}$, \mathbf{P}_x e funzione $\mathbf{g}(\cdot)$ 1) Fattorizzazione $\mathbf{P}_x = \mathbf{\Gamma} \mathbf{\Gamma}^\top$ con $\mathbf{\Gamma} = [\gamma_1, \dots, \gamma_n]$

2) Calcolo dei pesi

$$w_0 = 0$$

$$w_0^c = 2$$

$$w_i = w_i^c = \frac{1}{2n} \quad i = \pm 1, \dots, \pm n$$

3) Calcolo dei σ -punti

$$\mathbf{x}_0 = \bar{\mathbf{x}}, \alpha = \sqrt{n}$$

for $i = 1, \dots, n$ **do**

$$\quad \mathbf{x}_i = \bar{\mathbf{x}} + \alpha \gamma_i$$

$$\quad \mathbf{x}_{-i} = \bar{\mathbf{x}} - \alpha \gamma_i$$

end for4) Calcolo di $\bar{\mathbf{y}}$, \mathbf{P}_y , \mathbf{P}_{xy} **for** $i = -n, \dots, 0, \dots, n$ **do**

$$\quad \mathbf{y}_i = \mathbf{g}(\mathbf{x}_i)$$

end for

$$\bar{\mathbf{y}} = \sum_{i=-n}^n w_i \mathbf{y}_i$$

$$\mathbf{P}_y = \sum_{i=-n}^n w_i^c (\bar{\mathbf{y}} - \mathbf{y}_i)(\bar{\mathbf{y}} - \mathbf{y}_i)^\top$$

$$\mathbf{P}_{xy} = \sum_{i=-n}^n w_i^c (\bar{\mathbf{x}} - \mathbf{x}_i)(\bar{\mathbf{y}} - \mathbf{y}_i)^\top.$$

Algoritmo 5: Algoritmo UKF.

Dati: $\hat{\mathbf{x}}_{1|0}$, $\mathbf{P}_{1|0}$ **for** $t = 1, 2, \dots$ **do**

Correzione

$$\left[\hat{\mathbf{y}}_{t|t-1}, \mathbf{P}_{t|t-1}^y, \mathbf{P}_{t|t-1}^{xy} \right] = \text{UT} \left(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}, \mathbf{h}_t(\cdot) \right)$$

$$\mathbf{S}_t = \mathbf{P}_{t|t-1}^y + \mathbf{R}_t$$

$$\mathbf{L}_t = \mathbf{P}_{t|t-1}^{xy} \mathbf{S}_t^{-1}$$

Si acquisisce la misura \mathbf{y}_t

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{L}_t (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})$$

Predizione

$$\left[\hat{\mathbf{x}}_{t+1|t}, \bar{\mathbf{P}}_{t+1|t}, \cdot \right] = \text{UT} \left(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}, \mathbf{f}_t(\cdot, \mathbf{u}_t) \right)$$

$$\mathbf{P}_{t+1|t} = \bar{\mathbf{P}}_{t+1|t} + \mathbf{Q}_t$$

end for

Nell'UKF (vedi algoritmo 5), invece di approssimare al 1° ordine le funzioni $\mathbf{f}_t(\cdot, \mathbf{u}_t)$ e $\mathbf{h}_t(\cdot)$ come nell'EKF (mediante linearizzazione in un intorno della stima),

si approssimano i momenti della PDF $p_{t|t}(\cdot)$, la stima $\hat{\mathbf{x}}_{t|t}$ e la covarianza $\mathbf{P}_{t|t}$, tramite i punti- σ .

1.4.3 Filtro a particelle

L'aumento della potenza di calcolo degli elaboratori ha reso sempre più praticabili i metodi di campionamento che approssimano la distribuzione continua con una distribuzione discreta costituita da un numero molto elevato, M , di campioni, i.e.:

$$p(\mathbf{z}) \cong \sum_{i=1}^M w_i \delta(\mathbf{z} - \mathbf{z}_i) \quad (1.23)$$

dove: $\delta(\mathbf{z} - \mathbf{z}_i)$ è la delta di Dirac centrata sul campione \mathbf{z}_i ; $w_i \geq 0$ è il *peso di importanza* relativo al campione \mathbf{z}_i con $\sum_{i=1}^M w_i = 1$; la coppia (w_i, \mathbf{z}_i) prende il nome di *particella*. Da (1.23) si approssimano i momenti della distribuzione, cioè la media

$$\bar{\mathbf{z}} \cong \int \mathbf{z} \sum_{i=0}^M w_i \delta(\mathbf{z} - \mathbf{z}_i) d\mathbf{z} = \sum_{i=0}^M w_i \underbrace{\int \mathbf{z} \delta(\mathbf{z} - \mathbf{z}_i) d\mathbf{z}}_{\mathbf{z}_i} = \sum_{i=0}^M w_i \mathbf{z}_i,$$

e la covarianza

$$\begin{aligned} \mathbf{P}_{\mathbf{z}} &\cong \int \sum_{i=0}^M w_i \delta(\mathbf{z} - \mathbf{z}_i) (\mathbf{z} - \bar{\mathbf{z}}) (\mathbf{z} - \bar{\mathbf{z}})^\top d\mathbf{z} \\ &= \sum_{i=0}^M w_i \underbrace{\int (\mathbf{z} - \bar{\mathbf{z}}) (\mathbf{z} - \bar{\mathbf{z}})^\top \delta(\mathbf{z} - \mathbf{z}_i) d\mathbf{z}}_{(\mathbf{z}_i - \bar{\mathbf{z}}) (\mathbf{z}_i - \bar{\mathbf{z}})^\top} \\ &= \sum_{i=0}^M w_i (\mathbf{z}_i - \bar{\mathbf{z}}) (\mathbf{z}_i - \bar{\mathbf{z}})^\top. \end{aligned}$$

Con riferimento al problema di filtraggio, si assume che, all'istante t , la PDF $p_{t|t-1}(\cdot)$ è parametrizzata dall'insieme delle particelle $\left\{ \left(w_{t|t-1}^i, \mathbf{x}_{t|t-1}^i \right) \right\}_{i=1}^M$ nel seguente modo

$$p_{t|t-1}(\mathbf{x}) = \sum_{i=1}^M w_{t|t-1}^i \delta(\mathbf{x} - \mathbf{x}_{t|t-1}^i)$$

Applicando la correzione Bayesiana, si ha

$$\begin{aligned}
p_{t|t}(\mathbf{x}) &= \frac{l_t(\mathbf{x}, \mathbf{y}_t) p_{t|t-1}(\mathbf{x})}{\int l_t(\boldsymbol{\xi}, \mathbf{y}_t) p_{t|t-1}(\boldsymbol{\xi}) d\boldsymbol{\xi}} \\
&= \frac{l_t(\mathbf{x}, \mathbf{y}_t) \sum_{i=1}^M w_{t|t-1}^i \delta(\mathbf{x} - \mathbf{x}_{t|t-1}^i)}{\int l_t(\boldsymbol{\xi}, \mathbf{y}_t) \sum_{i=1}^M w_{t|t-1}^i \delta(\boldsymbol{\xi} - \mathbf{x}_{t|t-1}^i) d\boldsymbol{\xi}} \\
&= \frac{\sum_{i=1}^M w_{t|t-1}^i l_t(\mathbf{x}_{t|t-1}^i, \mathbf{y}_t) \delta(\mathbf{x} - \mathbf{x}_{t|t-1}^i)}{\sum_{j=1}^M w_{t|t-1}^j l_t(\mathbf{x}_{t|t-1}^j, \mathbf{y}_t)} \\
&= \sum_{i=1}^M w_{t|t}^i \delta(\mathbf{x} - \mathbf{x}_{t|t}^i)
\end{aligned}$$

con

$$\begin{aligned}
\mathbf{x}_{t|t}^i &= \mathbf{x}_{t|t-1}^i \\
w_{t|t}^i &= \frac{l_t(\mathbf{x}_{t|t-1}^i, \mathbf{y}_t) w_{t|t-1}^i}{\sum_{j=1}^M l_t(\mathbf{x}_{t|t-1}^j, \mathbf{y}_t) w_{t|t-1}^j}
\end{aligned} \tag{1.24}$$

per $i = 1, 2, \dots, M$.

In altri termini, la correzione preserva la natura discreta della PDF originaria $p_{t|t-1}(\cdot)$ con gli stessi campioni e pesi di importanza aggiornati con la nuova misura \mathbf{y}_t in accordo alla seconda relazione in (1.24). Si considera adesso la predizione Bayesiana:

$$\begin{aligned}
p_{t+1|t}(\mathbf{x}) &= \int \varphi_{t+1|t}(\mathbf{x}|\boldsymbol{\xi}) p_{t|t}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&= \int \varphi_{t+1|t}(\mathbf{x}|\boldsymbol{\xi}) \sum_{i=1}^M w_{t|t}^i \delta(\boldsymbol{\xi} - \mathbf{x}_{t|t}^i) d\boldsymbol{\xi} \\
&= \sum_{i=1}^M w_{t|t}^i \int \varphi_{t+1|t}(\mathbf{x}|\boldsymbol{\xi}) \delta(\boldsymbol{\xi} - \mathbf{x}_{t|t}^i) d\boldsymbol{\xi} \\
&= \sum_{i=1}^M w_{t|t}^i \varphi_{t+1|t}(\mathbf{x}|\mathbf{x}_{t|t}^i)
\end{aligned} \tag{1.25}$$

Si noti che la PDF $p_{t+1|t}(\cdot)$ in (1.25) non è altro che una mistura delle PDF di transizione continue $\varphi_{t+1|t}(\cdot|\mathbf{x}_{t|t}^i)$ condizionate ai vari campioni $\mathbf{x}_{t|t}^i$. In particolare, nel caso di disturbo di processo additivo Gaussiano, $p_{t+1|t}(\cdot)$ si riduce ad una mistura

di Gaussiane. In ogni caso, la PDF $p_{t+1|t}(\cdot)$ risultante dalla predizione non è più discreta (mistura di PDF impulsive) come $p_{t|t}(\cdot)$. Un algoritmo di filtraggio a particelle (PF= Particle Filter) o Sequenziale Monte Carlo (SMCF= Sequential Monte Carlo Filter) deve, dunque, provvedere a campionare la PDF continua (1.25). Il campionamento di (1.25) può essere effettuato in svariati modi, i.e. utilizzando diverse densità di importanza, ciascuno dei quali caratterizza una diversa variante di filtro a particelle. Il metodo più comunemente usato è quello di generare $\mathbf{x}_{t+1|t}^i$ campionando direttamente la mistura di PDF, ovvero nel seguente modo

per $i = 1, \dots, M$:

$$\begin{aligned} \mathbf{w}_t^i &\sim p_{\mathbf{w}_t}(\cdot) \\ \mathbf{x}_{t+1|t}^i &= \mathbf{f}_t(\mathbf{x}_{t|t}^j, \mathbf{u}_t, \mathbf{w}_t^i) \quad \text{con probabilità } w_{t|t}^j \end{aligned}$$

Questo porta all'approssimazione

$$p_{t+1|t}(\mathbf{x}) \cong \sum_{i=1}^M w_{t+1|t} \delta(\mathbf{x} - \mathbf{x}_{t+1|t}^i)$$

con $w_{t+1|t} = 1/M$ per $i = 1, 2, \dots, M$. L'algoritmo risultante da questa scelta è denominato *filtro a particelle con ricampionamento sequenziale* (SIR-PF = Sequential Importance Resampling - Particel Filter) ed è riportato nell'algoritmo 6.

Un metodo generico di approssimazione consiste nell'utilizzo di una PDF di campionamento d'importanza generica $\mathbf{x}_{t+1|t}^i \sim q(\mathbf{x}, \mathbf{x}_{t|t}^i, \mathbf{y}_t)$, eventualmente tenendo conto delle misure.

Algoritmo 6: Algoritmo SIR-PF.

Dati: Per $i = 1, 2, \dots, M$ si generano i campioni $\mathbf{x}_{1|0}^i$ con la PDF $p_{1|0}(\cdot)$ e si pone $w_{1|0}^i = 1/M$

for $t = 1, 2, \dots$ **do**

Correzione:

si acquisisce \mathbf{y}_t

for $i = 1, 2, \dots, M$ **do**

$$w_{t|t}^i = \frac{l_t(\mathbf{x}_{t|t-1}^i, \mathbf{y}_t) w_{t|t-1}^i}{\sum_{j=1}^M l_t(\mathbf{x}_{t|t-1}^j, \mathbf{y}_t) w_{t|t-1}^j}$$

$$\mathbf{x}_{t|t}^i = \mathbf{x}_{t|t-1}^i;$$

end for

$$\hat{\mathbf{x}}_{t|t} = \sum_{i=1}^M w_{t|t}^i \mathbf{x}_{t|t}^i$$

$$\mathbf{P}_{t|t} = \sum_{i=1}^M w_{t|t}^i (\mathbf{x}_{t|t}^i - \hat{\mathbf{x}}_{t|t}) (\mathbf{x}_{t|t}^i - \hat{\mathbf{x}}_{t|t})^\top$$

Predizione:

for $i = 1, 2, \dots, M$ **do**

si estrae $j \in \{1, 2, \dots, M\}$ con probabilità $w_{t|t}^j$

si pone $\mathbf{x}_{t+1|t}^i \sim \varphi_{t+1|t}(\cdot | \mathbf{x}_{t|t}^j)$, che equivale a generare $\mathbf{x}_{t+1|t}^i$ con un

passo di simulazione a partire da $\mathbf{x}_{t|t}^j$, cioè $\mathbf{x}_{t+1|t}^i = \mathbf{f}_t(\mathbf{x}_{t|t}^j, \mathbf{u}_t, \mathbf{w}_t^i)$ con

$$\mathbf{w}_t^i \sim p_{\mathbf{w}_t}(\cdot)$$

$$w_{t+1|t}^i = 1/M$$

end for

$$\hat{\mathbf{x}}_{t+1|t} = \sum_{i=1}^M w_{t+1|t}^i \mathbf{x}_{t+1|t}^i$$

$$\mathbf{P}_{t+1|t} = \sum_{i=1}^M w_{t+1|t}^i (\mathbf{x}_{t+1|t}^i - \hat{\mathbf{x}}_{t+1|t}) (\mathbf{x}_{t+1|t}^i - \hat{\mathbf{x}}_{t+1|t})^\top$$

end for

Capitolo 2

Stima Bayesiana distribuita

In questo capitolo viene presentato un approccio Bayesiano, del tutto generale, al problema della stima distribuita dello stato di un sistema dinamico mediante una rete di sensori. Tale approccio utilizza il *consenso in media* [2, 33] sulla PDF del vettore di stato da stimare ed una definizione di media fra PDF derivante dalla teoria dell'informazione [17, 34]. Grazie a questa formulazione Bayesiana sarà possibile, nel capitolo successivo, derivare algoritmi distribuiti applicabili anche a sistemi dinamici non lineari e per tali algoritmi, restringendosi al caso lineare tempo-invariante, dimostrare proprietà di stabilità valide sotto i requisiti minimali di connettività della rete e di osservabilità collettiva del sistema dinamico da parte della rete.

Il capitolo è organizzato come segue. Nel paragrafo 2.1 viene presentato il modello della rete di sensori considerata e nel paragrafo 2.2 sono introdotte alcune metriche di interesse per valutare le proprietà di connettività della rete. Nel paragrafo 2.3 viene illustrata la topologia di rete wireless considerata per la generazione delle reti di prova utilizzate in questo lavoro di tesi allo scopo di valutare le prestazioni degli stimatori distribuiti proposti. Nel paragrafo 2.4 viene introdotta la *media di Kullback-Leibler* fra PDF ed il relativo consenso in media fra PDF. Sulla base degli strumenti del paragrafo 2.4, il paragrafo 2.5 presenta tre diverse metodologie di consenso per distribuire, in modo scalabile, la stima dello stato fra i vari nodi della rete. Un primo approccio opera sulle PDF a posteriori dei vari nodi. Un secondo approccio opera sulle funzioni di verosimiglianza (cioè sulla nuova informazione) dei nodi, mentre un terzo approccio effettua in parallelo un consenso sulle verosimiglianze ed un consenso sulle PDF a priori per combinare gli effetti benefici dei primi due approcci.

2.1 Modello della rete

In questo e nel successivo paragrafo verranno illustrate le reti distribuite e le loro caratteristiche. Il tipo di rete considerato in questo lavoro è schematizzato in fig. 2.1. Come si vede dalla figura, la rete consiste di due tipi di nodi: i nodi *sensore* (**SEN**) e i nodi di *comunicazione* (**COM**). Entrambi hanno capacità di *elaborazione* e *comunicazione*, cioè possono elaborare localmente i dati e scambiare informazioni con i nodi vicini. L'unica differenza sta nel fatto che soltanto i nodi sensore

Node Types



Node functions

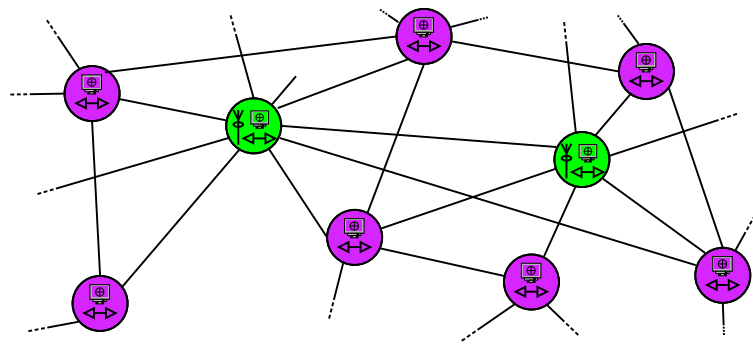
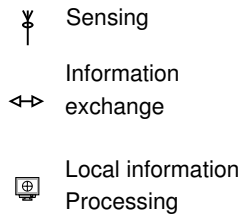


Figura 2.1: Modello della rete.

hanno capacità di rilevamento, cioè possono raccogliere informazioni dall'ambiente circostante. Si noti che i nodi **COM** non forniscono alcun tipo di informazione sull'ambiente, la loro presenza è necessaria unicamente per garantire e migliorare la connettività di rete.

Le reti d'interesse sono caratterizzate, oltre che dalla presenza di nodi **SEN** e **COM**, dalle seguenti ipotesi.

N1. La rete non presenta un *nodo centrale di fusione*, cioè ha una configurazione *decentralizzata* o *distribuita*. Ne segue che tutti i nodi della rete operano alla stessa maniera, cioè

- rilevano, eventualmente, informazioni dall'ambiente (solo i nodi **SEN**);
- ricevono dati dai nodi vicini;
- fondono opportunamente i dati rilevati, quelli ricevuti e locali;
- trasmettono i dati fusi ai nodi vicini.

N2. Nessun nodo conosce la topologia della rete, cioè il numero di nodi che la compongono e le loro connessioni.

N3. La topologia di rete può essere tempo variante a causa, ad esempio, dell'aggiunta o dell'eliminazione di nodi o di collegamenti.

Da un punto di vista matematico, la rete può essere descritta come un grafo diretto eventualmente tempo-variante $\mathcal{G}_t = (\mathcal{N}_t, \mathcal{A}_t)$ dove \mathcal{N}_t è l'insieme dei nodi al tempo t e $\mathcal{A}_t \subset \mathcal{N}_t \times \mathcal{N}_t$ l'insieme degli archi, che rappresentano i collegamenti (connessioni), al tempo t . In particolare, (i, j) appartiene a \mathcal{A}_t se il nodo j può ricevere dati dal nodo i al tempo t . L'insieme dei nodi al tempo t è partizionato in $\mathcal{N}_t = \mathcal{S}_t \cup \mathcal{C}_t$ dove \mathcal{S}_t e \mathcal{C}_t denotano i sottoinsiemi disgiunti di nodi sensori, e rispettivamente, nodi di comunicazione al tempo t . $N_{s,t} \triangleq |\mathcal{S}_t|$, $N_{c,t} \triangleq |\mathcal{C}_t|$ e $N_t \triangleq |\mathcal{N}_t| = N_{c,t} + N_{s,t}$

indicano il numero di nodi sensori, il numero di nodi di comunicazione ed il numero totale di nodi al tempo t . Per ogni nodo $i \in \mathcal{N}_t$, $\mathcal{N}_t^i \triangleq \{j \in \mathcal{N}_t : (j, i) \in \mathcal{A}_t\}$ indica l'insieme dei nodi vicini ad i al tempo t . Per definizione, $(i, i) \in \mathcal{A}_t$ per ogni istante t e ogni nodo $i \in \mathcal{N}_t$ e, quindi, $i \in \mathcal{N}_t^i$ per ogni i e t . Analogamente, $\mathcal{S}_t^i = \mathcal{N}_t^i \cap \mathcal{S}_t$ e $\mathcal{C}_t^i = \mathcal{N}_t^i \cap \mathcal{C}_t$ denotano gli insiemi di nodi sensori e, rispettivamente, di comunicazione, vicini del nodo i . Infine, $N_{s,t}^i$, $N_{c,t}^i$ e N_t^i denotano le cardinalità di \mathcal{S}_t^i , \mathcal{C}_t^i e, rispettivamente, \mathcal{N}_t^i .

In molti casi verranno considerate reti tempo-invarianti ma, anche se tempo-varianti, per semplicità si ometterà di indicare la dipendenza dal tempo t . Quindi la notazione per i grafi verrà semplificata come segue

$$\mathcal{G} = (\mathcal{N}, \mathcal{A}), \quad \mathcal{N} = \mathcal{S} \cup \mathcal{C}, \quad N \triangleq |\mathcal{N}| = N_s + N_c = |\mathcal{S}| + |\mathcal{C}|,$$

$$\mathcal{N}^i = \mathcal{S}^i \cup \mathcal{C}^i, \quad N^i \triangleq |\mathcal{N}^i| = N_s^i + N_c^i = |\mathcal{S}^i| + |\mathcal{C}^i|$$

dove:

\mathcal{N}	insieme dei nodi;
$\mathcal{A} \subset \mathcal{N} \times \mathcal{N}$	insieme degli archi;
$\mathcal{S} \subset \mathcal{N}$	insieme dei nodi SEN ;
$\mathcal{C} = \mathcal{N} \setminus \mathcal{S}$	insieme dei nodi COM ;
N	numero dei nodi;
N_s	numero di nodi SEN ;
N_c	numero di nodi COM ;
\mathcal{N}^i	insieme dei nodi connessi al nodo i (incluso i);
\mathcal{S}^i	insieme dei nodi SEN connessi al nodo i (incluso eventualmente i);
\mathcal{C}^i	insieme dei nodi COM connessi al nodo i (incluso eventualmente i);
N^i	numero di nodi connesso al nodo i (incluso i);
N_s^i	numero di nodi SEN connesso al nodo i (incluso eventualmente i);
N_c^i	numero di nodi COM connessi al nodo i (incluso eventualmente i).

Nel seguito la rete sarà indicata anche con la tripletta $(\mathcal{S}, \mathcal{C}, \mathcal{A})$.

Ogni nodo della rete è caratterizzato da alcuni parametri, tra cui:

- coordinate geografiche (nell'area di sorveglianza 1D, 2D o 3D);
- tipo di nodo, **SEN** o **COM**, e in caso di nodo **SEN** anche il tipo di sensore, dato che la rete può accogliere sensori eterogenei;
- lista dei nodi vicini.

Dato che l'obiettivo di questo lavoro è di considerare l'impiego delle reti per scopi di monitoraggio distribuito così come di studiare, in questo contesto, la robustezza delle reti a vari tipi di attacchi da parte di virus informatici o a guasti, è importante caratterizzare, possibilmente in modi diversi, la connettività della rete (grafo), che è legata alla capacità della rete di diffondere sia le informazioni che le epidemie virali. Nel prossimo paragrafo, verranno esaminate varie metriche per i grafi ed alcune relazioni matematiche tra di esse.

2.2 Metriche per i grafi

Questo paragrafo introduce alcune metriche per i grafi (reti) che ci serviranno per caratterizzare le reti in termini di connettività. In particolare, il nostro obiettivo sarà di analizzare il comportamento di uno stimatore dello stato net-centrico (che sarà introdotto successivamente) sotto varie condizioni di connettività e così mettere in relazione prestazioni e robustezza della stima con le metriche di rete. Un altro obiettivo sarà quello di analizzare gli effetti dei virus informatici sulle reti di sorveglianza e mettere in relazione la diffusione epidemica con adeguate metriche di rete.

Autovalori del Laplaciano e connettività algebrica

Una metrica significativa per le reti è la cosiddetta matrice Laplaciano del grafo associato alla rete [35]. Dato un grafo semplice $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ con $|\mathcal{N}| = N$ nodi, la sua *matrice di adiacenza* \mathbf{A} è definita come la matrice $N \times N$ di elementi a_{ij} tali che $a_{ij} = 1$ se $(i, j) \in \mathcal{A}$ altrimenti $a_{ij} = 0$. In particolare il raggio spettrale di \mathbf{A} , $\rho(\mathbf{A}) \triangleq \max \{|\lambda|, \lambda \in sp(\mathbf{A})\}$, rappresenta un'importante caratterizzazione della connettività della rete.

Inoltre sia \mathbf{D} una matrice diagonale $N \times N$ i cui elementi $d_{ii} = k_i$ sono pari al grado del nodo i , cioè il numero di nodi connessi ad i . Allora, la matrice Laplaciano, o semplicemente il Laplaciano, del grafo \mathcal{G} è definito come $\mathbf{L} \triangleq \mathbf{D} - \mathbf{A}$. Tale matrice \mathbf{L} risulta singolare e, nel caso di grafi semplici, simmetrica e semi-definita positiva. Più specificamente, gli autovalori del Laplaciano $\lambda_1, \lambda_2, \dots, \lambda_N$ soddisfano le disuguaglianze:

$$0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$$

In particolare, il secondo più piccolo autovalore $\lambda_2 \geq 0$ di L è chiamato *connettività algebrica* del grafo \mathcal{G} . Infatti, questo autovalore rappresenta un indicatore della connettività del grafo dato che il grafo risulta essere connesso se e solo se $\lambda_2 > 0$ [36]. In realtà, ci sono anche altre relazioni tra λ_2 e la connettività. A questo scopo, si introduce la *connettività di nodo* ν definita come il più piccolo numero di nodi che se rimossi da \mathcal{G} , lasciano un sottografo disconnesso o banale. Inoltre, sia ϵ la *connettività di arco* definita come il più piccolo numero di archi che, quando rimossi da \mathcal{G} lasciano un sottografo disconnesso o banale. Allora, come mostrato in [36], ne segue che $\lambda_2 \leq \nu \leq \epsilon$, cioè la connettività algebrica è un limite inferiore per la connettività di nodo che, a sua volta, è un limite inferiore per la connettività di arco. Siano k_{min} , k_{max} e \bar{k} , rispettivamente, il grado minimo, massimo e medio della rete. Si può mostrare [36] che, per una rete generica, le quantità sopra definite soddisfano le seguenti disuguaglianze:

$$\lambda_2 \leq \nu \leq \epsilon \leq k_{min} \leq \bar{k} \leq \rho(\mathbf{A}) \leq k_{max}. \quad (2.1)$$

Un grafo si dice k -regolare se $k_{min} = k_{max} = \bar{k} = k$, cioè se tutti i nodi hanno lo stesso grado k .

Altre relazioni interessanti sono state trovate tra la connettività algebrica e il diametro o distanza media di un grafo. A tale riguardo, si definisce la distanza di hop $\rho(i, j)$ tra due nodi $i, j \in \mathcal{N}$ del grafo come il numero di archi che costituiscono

il cammino più breve che unisce tali nodi. Allora, il diametro di un grafo è la massima distanza di hop tra due qualunque nodi del grafo, cioè

$$diam(\mathcal{G}) \triangleq \max_{i,j \in \mathcal{N}} \varrho(i, j).$$

Invece, la *distanza media* di un grafo è pari alla media di tutte le distanze di hop tra i nodi, cioè

$$\bar{\varrho}(\mathcal{G}) \triangleq \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N}, i \neq j} \varrho(i, j)$$

Mohar [35] ha dimostrato le seguenti limitazioni, inferiore e superiore, λ_2 -dipendenti sul diametro e sulla distanza media di un grafo:

$$\begin{aligned} \frac{4}{N\lambda_2} &\leq diam(\mathcal{G}) \leq 2 \left\lceil \frac{\Delta + \lambda_2}{4\lambda_2} \ln(N-1) \right\rceil \\ \frac{1}{N-1} \left\lceil \frac{2}{\lambda_2} + \frac{N-2}{2} \right\rceil &\leq \bar{\varrho}(\mathcal{G}) \leq \frac{N}{N-1} \left(\left\lceil \frac{\Delta + \lambda_2}{4\lambda_2} \ln(N-1) \right\rceil + \frac{1}{2} \right) \end{aligned} \quad (2.2)$$

dove $\Delta = \Delta(\mathcal{G})$ indica il massimo grado tra tutti i nodi del grafo \mathcal{G} , cioè

$$\Delta = \max_{i \in \mathcal{N}} \{ |\mathcal{N}^i| - 1 \}.$$

Si noti che i limiti sul diametro e la distanza media (2.2) si avvicinano al crescere della connettività algebrica λ_2 .

In [2, 33] è stato mostrato come la velocità di convergenza di un algoritmo di consenso sia legata alla connettività algebrica della rete λ_2 . Pertanto, la connettività algebrica è una metrica fondamentale in questo lavoro dato che il metodo del consenso viene ampiamente utilizzato negli algoritmi di stima distribuita considerati.

2.3 Topologia di rete

In questo lavoro di tesi si considerano, come reti di prova per la valutazione delle prestazioni degli stimatori distribuiti, reti *geografiche*, in particolare reti *geometriche casuali*. Si tratta di reti distribuite con nodi (sensori e di comunicazione) disposti casualmente nell'area d'interesse. Per quanto riguarda le interconnessioni tra i nodi, in un contesto di comunicazione senza fili (radio) la connettività dipende essenzialmente dalla posizione relativa dei nodi, dalla potenza di trasmissione e sensibilità di ricezione dei nodi.

Il più semplice modello di connettività per reti senza fili multihop prevede che un nodo i ed un nodo j siano connessi se la loro distanza non supera un valore di soglia assegnato. Sia $d(i, j)$ la distanza tra i nodi i e j . Allora, tipicamente, l'attenuazione del segnale dal nodo i al nodo j è proporzionale ad una certa potenza di quella distanza, cioè

$$\beta(i, j) \triangleq \frac{p_T(i)}{p_R(j)} = C d(i, j)^\alpha \quad (2.3)$$

dove: $p_T(i)$ è la potenza trasmessa dal nodo i ; $p_R(j)$ la potenza ricevuta dal nodo j ; C è una costante che rappresenta l'attenuazione del segnale a distanza unitaria; α è chiamato *pathloss exponent*, *esponente di attenuazione sul percorso*. Il nodo j riceve adeguatamente i dati trasmessi dal nodo i se la potenza ricevuta $p_R(j)$ è maggiore o uguale ad un valore di soglia assegnato $p_{R,min}(j)$ detto *sensibilità del ricevitore* del nodo j . In base a questo semplice modello, puramente geometrico, si può concludere che

$$(i, j) \in \mathcal{A} \iff d(i, j) \leq d_{max}(i, j) \triangleq \left[\frac{p_T(i)}{C p_{R,min}(j)} \right]^{1/\alpha} \quad (2.4)$$

La condizione di connettività (2.4) stabilisce che esiste un collegamento tra i e j se la distanza $d(i, j)$ non supera la soglia $d_{max}(i, j)$, che in generale dipende sia dalla potenza di trasmissione del nodo i che dalla sensibilità del ricevitore del nodo j . Se in particolare tutti i nodi hanno le stesse caratteristiche di comunicazione (cioè, potenza di trasmissione e sensibilità del ricevitore), allora $d_{max}(i, j) = R_{max}$ è la stessa per tutte le coppie (i, j) e la condizione di connettività (2.4) diventa

$$(i, j) \in \mathcal{A} \iff (j, i) \in \mathcal{A} \iff d(i, j) \leq R_{max}. \quad (2.5)$$

Una rete caratterizzata dalla condizione di connettività (2.5) da qui in avanti sarà chiamata *rete geometrica* con distanza di comunicazione (o raggio di connettività) R_{max} . Esempi di reti di questo tipo sono riportati nelle figg. 2.2 e 2.3. Si noti che,

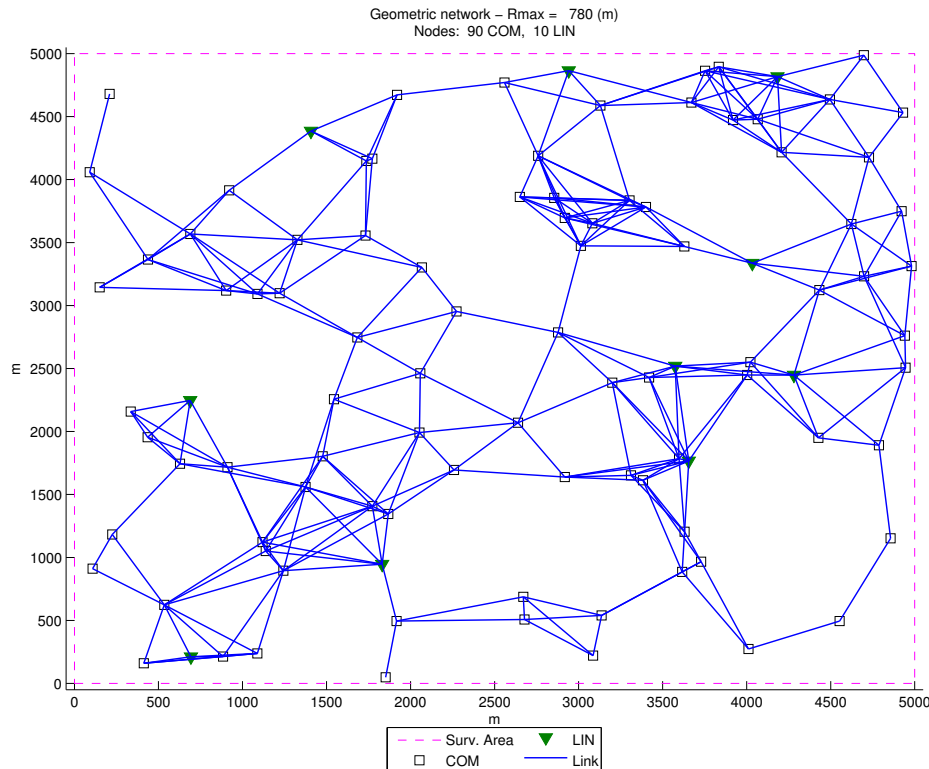


Figura 2.2: Rete geometrica casuale con $R_{max} = 780$ (m).

in virtù della simmetria di (2.5), R_{max} rappresenta sia la distanza di trasmissione che

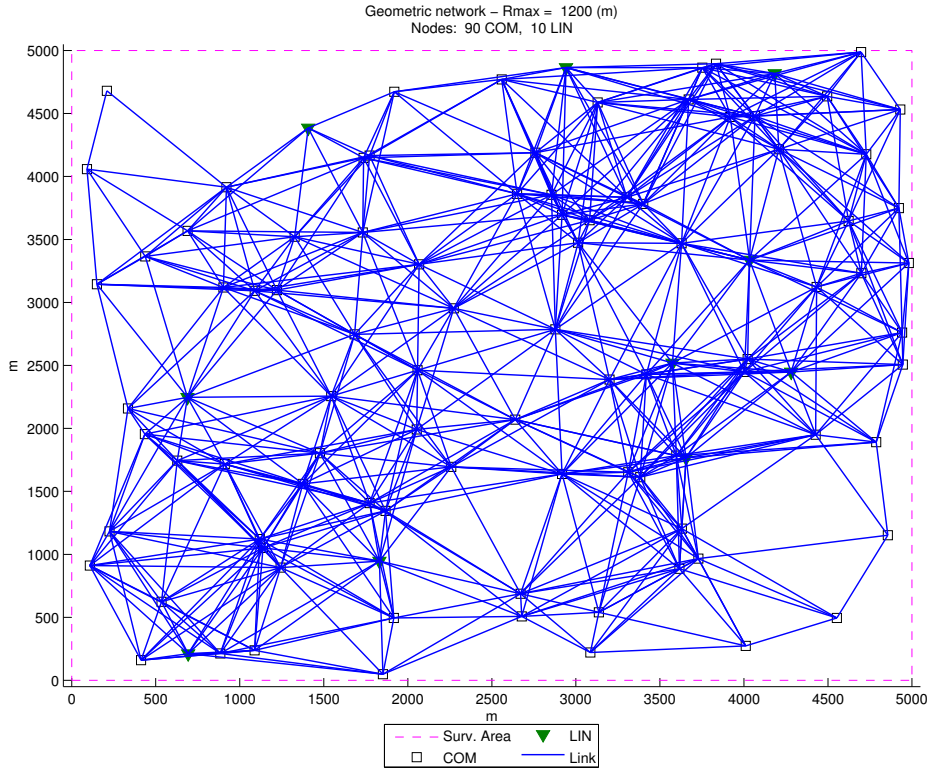


Figura 2.3: Rete geometrica casuale con $R_{max} = 1200 (m)$.

quella di ricezione di un nodo. Una rete geometrica può essere descritta mediante un grafo semplice.

Per costruire una rete geometrica si possono posizionare casualmente i nodi con una distribuzione uniforme nell'area di sorveglianza, e poi i collegamenti tra i nodi vengono definiti secondo (2.5) avendo fissato la distanza massima, come ad esempio le reti delle figg. 2.2 e 2.3 che presentano la stessa disposizione dei nodi ma raggi di connettività massima diversi. Entrambe le reti sono costituite da 90 nodi **COM** e 10 nodi **SEN**. Per quanto riguarda le connessioni, i due esempi mostrano due casi diversi, uno con $R_{max} = 780 (m)$ e l'altro con $R_{max} = 1200 (m)$. È evidente come aumentando R_{max} la connettività della rete aumenta sensibilmente.

2.4 Consenso sulla media di Kullback-Leibler

Si consideri una rete di N nodi, rappresentata mediante un grafo semplice $\mathcal{G} = (\mathcal{N}, \mathcal{A})$. Sia

$$\mathcal{P} \triangleq \left\{ p(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R} \mid \int_{\mathbb{R}^n} p(\mathbf{x}) d\mathbf{x} = 1 \text{ e } p(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^n \right\}$$

l'insieme delle PDF su \mathbb{R}^n . Si suppone che, ad ogni nodo della rete i , sia associata una PDF $p^i(\cdot) \in \mathcal{P}$ che rappresenta l'informazione locale su un vettore casuale $\mathbf{x} \in \mathbb{R}^n$. Ad esempio, tali PDF possono essere ottenute per mezzo di inferenza statistica o possono essere il risultato di qualche algoritmo di stima Bayesiano ricorsivo (come si vedrà in seguito).

Si vuole trovare un algoritmo di consenso che garantisca che tutti i nodi della rete raggiungano un accordo sulla PDF di \mathbf{x} . Per algoritmo di consenso s'intende un modo d'interazione (tra gli agenti) distribuito che specifichi:

1. lo scambio d'informazione tra un nodo e i suoi vicini;
2. il meccanismo per aggiornare la PDF locale in base alle informazioni ricevute.

Per questo, sia $p_\ell^i(\cdot) \in \mathcal{P}$ la PDF al nodo i all' ℓ -esima iterazione dell'algoritmo di consenso, inizializzata con $p_0^i(\cdot) = p^i(\cdot)$. Allora, si ha il seguente problema di consenso.

Problema 1. *Trovare un algoritmo tale che*

$$\lim_{\ell \rightarrow \infty} p_\ell^i(\mathbf{x}) = \bar{p}(\mathbf{x}), \quad \forall i \in \mathcal{N},$$

dove la PDF asintotica $\bar{p}(\cdot)$ rappresenta la media (in qualche senso) delle PDF iniziali $p^i(\cdot)$, $i \in \mathcal{N}$.

Prima di tutto occorre definire la PDF media $\bar{p}(\cdot)$ date le PDF degli agenti $p^i(\cdot)$, $i \in \mathcal{N}$. A questo scopo, è opportuno ricordare che, dato un insieme di punti $\{\mathbf{x}^1, \dots, \mathbf{x}^N\}$ appartenenti a \mathbb{R}^n ed i pesi loro associati $\{\pi^1, \dots, \pi^N\}$ che soddisfano

$$\pi^i \geq 0, \quad \sum_{i=1}^N \pi^i = 1, \quad (2.6)$$

la loro media pesata $\bar{\mathbf{x}} = \sum_{i=1}^N \pi^i \mathbf{x}^i$ soddisfa la proprietà variazionale di minimizzare la distanza quadratica media pesata da tali punti cioè,

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^N \pi^i \|\mathbf{x} - \mathbf{x}^i\|^2 \quad (2.7)$$

dove $\|\cdot\|$ è la norma Euclidea. Esistono molti modi di misurare la distanza tra due PDF $p(\cdot)$ e $q(\cdot)$ in \mathcal{P} . Dal punto di vista della teoria dell'informazione, la scelta più comune corrisponde alla *divergenza di Kullback-Leibler* (KLD: *Kullback-Leibler Divergence*), o *entropia relativa* definita come

$$D_{\text{KL}}(p||q) = \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}.$$

La KLD ha molte interpretazioni; ad esempio, nelle statistiche Bayesiane, può essere vista come il guadagno d'informazione che si ha passando da una PDF a priori $q(\cdot)$ ad una PDF a posteriori $p(\cdot)$. Vale anche la pena di ricordare che la KLD può essere vista come l'equivalente del quadrato della distanza Euclidea nello spazio delle PDF \mathcal{P} (infatti, sono entrambe esempi di divergenze di Bregman [37]). Con questa considerazione, e tenendo conto di (2.7), si introduce la seguente nozione.

Definizione 1. Date N PDF $p^i(\cdot) \in \mathcal{P}$, e i relativi pesi π^i che soddisfano (2.6), la loro *media di Kullback-Leibler* (KLA: *Kullback-Leibler Average*) è definita come segue

$$\bar{p} = \arg \inf_{p \in \mathcal{P}} \sum_{i=1}^N \pi^i D_{\text{KL}}(p||p^i). \quad (2.8)$$

In particolare, se $\pi^i = 1/N$ per $i = 1, \dots, N$, (2.8) fornisce la KLA (non ponderata) che media le PDF degli agenti dando a tutti lo stesso grado di fiducia (peso). In base alla definizione 1, la PDF media è quella che minimizza la somma dei guadagni d'informazione rispetto alle PDF iniziali. Così, questa scelta è coerente con il *principio di minima discriminazione dell'informazione* (PMDI: *Principle of Minimum Discrimination Information*), secondo il quale la PDF che meglio rappresenta lo stato attuale della conoscenza è quella che produce un guadagno d'informazione il più piccolo possibile (i lettori interessati possono consultare [38, 39] per una discussione su tale principio e le sue relazioni con il principio di Gauss e con la stima a massima verosimiglianza). In altre parole [40], “the probability assignment which most honestly describes what we know should be the most conservative assignment in the sense that it does not permit one to draw any conclusions not warranted by the data”, cioè la scelta della probabilità che meglio descrive ciò che sappiamo dovrebbe essere la scelta più conservativa nel senso che non permetta di trarre alcuna conclusione che non sia supportata dai dati. Come si vedrà nel prossimo paragrafo, tale principio porta ad un'interessante proprietà nel caso della stima distribuita dello stato di sistemi lineari Gaussiani.

Si può dimostrare il seguente risultato riguardo la KLA.

Teorema 1. *La KLA pesata definita in (2.8) è data da*

$$\bar{p}(\mathbf{x}) = \frac{\prod_{i=1}^N [p^i(\mathbf{x})]^{\pi^i}}{\int \prod_{i=1}^N [p^i(\mathbf{x})]^{\pi^i} d\mathbf{x}} \quad (2.9)$$

Dimostrazione: si veda l'Appendice A. \square

Il suddetto teorema afferma che la KLA (pesata) $\bar{p}(\cdot)$ non è altro che la media geometrica (pesata) normalizzata delle PDF degli agenti $p^i(\cdot)$, detta anche *mistura di densità esponenziale* (EMD: *Exponential Mixture Density*) in [41]. È importante evidenziare che la regola di fusione (2.9) coincide effettivamente con la cosiddetta *intersezione di covarianze generalizzata* (GCI: *Generalized Covariance Intersection*) che è stata introdotta [42, 43, 44, 45] come una generalizzazione della *Covariance Intersection*, ideata originariamente da [3] per le PDF Gaussiane. È ben noto che la GCI produce una fusione dell'informazione che è robusta rispetto alle correlazioni incognite tra le sorgenti d'informazione.

Per illustrare questo punto in modo chiaro, si considerano solo due sorgenti d'informazione Y^1 e Y^2 con associate PDF condizionate $p^i(\mathbf{x}) = p(\mathbf{x}|Y^i)$ per $i = 1, 2$. In generale, l'insieme complessivo delle informazioni $Y^1 \cup Y^2$ può essere decomposto nell'unione di tre insiemi d'informazione disgiunti (indipendenti) come segue $Y^1 \cup Y^2 = (Y^1 \setminus Y^2) \cup (Y^2 \setminus Y^1) \cup (Y^1 \cap Y^2)$. Da qui, la fusione ottima di $p^1(\cdot)$ e $p^2(\cdot)$ può essere ottenuta mediante:

$$p^{12}(\mathbf{x}) \propto p(\mathbf{x}|Y^1 \cup Y^2) \propto p(\mathbf{x}|Y^1 \setminus Y^2) p(\mathbf{x}|Y^2 \setminus Y^1) p(\mathbf{x}|Y^1 \cap Y^2) \propto \frac{p^1(\mathbf{x}) p^2(\mathbf{x})}{p(\mathbf{x}|Y^1 \cap Y^2)} \quad (2.10)$$

se la PDF $p(\mathbf{x}|Y^1 \cap Y^2)$ condizionata all'informazione comune $Y^1 \cap Y^2$ fosse nota (qui il simbolo \propto sta a significare "proporzionale a" e il fattore di proporzionalità è determinato imponendo che $p^{12}(\cdot)$ abbia integrale unitario su \mathbb{R}^n). Comunque, nell'ambito considerato, in cui i nodi fondono ripetutamente l'informazione che viene dai loro vicini senza nessuna conoscenza della topologia della rete, è impossibile individuare (isolare) l'informazione comune e quindi applicare (2.10). Per questo verranno adottate alcune strategie robuste di fusione subottima. Con questa considerazione, la fusione KLA (o GCI) fornisce

$$\bar{p}^{12}(\mathbf{x}) \triangleq \frac{[p^1(\mathbf{x})]^\pi [p^2(\mathbf{x})]^{1-\pi}}{\int [p^1(\xi)]^\pi [p^2(\xi)]^{1-\pi} d\xi} \propto \frac{[p(\mathbf{x}|Y^1 \setminus Y^2)]^\pi [p(\mathbf{x}|Y^2 \setminus Y^1)]^{1-\pi} p(\mathbf{x}|Y^1 \cap Y^2)}{[p(\mathbf{x}|Y^1 \setminus Y^2)]^\pi [p(\mathbf{x}|Y^2 \setminus Y^1)]^{1-\pi} p(\mathbf{x}|Y^1 \cap Y^2)}, \quad \pi \in [0, 1] \quad (2.11)$$

dove si può vedere che non si verifica alcun doppio conteggio dell'informazione comune. Il prezzo da pagare è un appiattimento conservativo dell'informazione indipendente $[p(Y^1 \setminus Y^2 | \mathbf{x})]^\pi [p(Y^2 \setminus Y^1 | \mathbf{x})]^{1-\pi}$. Perciò la fusione (2.9) con (2.6) risulta robusta rispetto al *data incest*.

Per ottenere la KLA (2.9) sull'intera rete in modo distribuito e scalabile, si può sfruttare l'approccio del consenso. A questo scopo, è opportuno introdurre due operatori che agiscono sulle PDF.

Più specificamente, date due PDF $p(\cdot)$ e $q(\cdot)$, viene definita la loro *somma* come

$$p(\mathbf{x}) \oplus q(\mathbf{x}) = \frac{p(\mathbf{x})q(\mathbf{x})}{\int p(\mathbf{x})q(\mathbf{x}) d\mathbf{x}}.$$

Inoltre, dato uno scalare positivo π e una PDF $p(\cdot)$, viene definito l'*operatore di moltiplicazione per uno scalare* come

$$\pi \odot p(\mathbf{x}) = \frac{[p(\mathbf{x})]^\pi}{\int [p(\mathbf{x})]^\pi d\mathbf{x}}.$$

In considerazione di tali definizioni, si può verificare che i due operatori \oplus e \odot soddisfano le seguenti proprietà:

- a) $(p \oplus q) \oplus r = p \oplus (q \oplus r) = p \oplus q \oplus r$
- b) $p \oplus q = q \oplus p$
- c) $(\alpha \beta) \odot p = \alpha \odot (\beta \odot p)$
- d) $1 \odot p = p$
- e) $\alpha \odot (p \oplus q) = (\alpha \odot p) \oplus (\alpha \odot q)$
- f) $(\alpha + \beta) \odot p = (\alpha \odot p) \oplus (\beta \odot p)$

dove p, q, r sono PDF e α, β scalari positivi. Inoltre, l'identità (2.9) con $\pi^i = \frac{1}{N}$ per ogni i può essere riscritta come

$$\bar{p}(\mathbf{x}) = \bigoplus_{i \in \mathcal{N}} \left(\frac{1}{N} \odot p^i(\mathbf{x}) \right). \quad (2.12)$$

Allora, la KLA collettiva (2.12), che richiederebbe la disponibilità delle PDF locali di tutti gli agenti, può essere calcolata in modo distribuito e scalabile applicando uno dei vari algoritmi di consenso esistenti per la media distribuita. Gli algoritmi più semplici consistono nell'aggiornare i dati locali per mezzo di una combinazione convessa con i dati ricevuti dagli agenti vicini [1, 2]. In questo caso, una tale scelta porta al seguente algoritmo di consenso:

$$p_{\ell+1}^i(\mathbf{x}) = \bigoplus_{j \in \mathcal{N}^i} (\pi^{i,j} \odot p_{\ell}^j(\mathbf{x})), \quad \forall i \in \mathcal{N} \quad (2.13)$$

dove $p^i(\mathbf{x})$ e $\pi^{i,j}$ sono opportuni pesi non-negativi, tali che

$$\sum_{j \in \mathcal{N}^i} \pi^{i,j} = 1, \quad \forall i \in \mathcal{N}.$$

Infatti, grazie alle proprietà a)-f), si può vedere che¹

$$p_{\ell}^i(\mathbf{x}) = \bigoplus_{j \in \mathcal{N}} (\pi_{\ell}^{i,j} \odot p^j(\mathbf{x})), \quad \forall i \in \mathcal{N} \quad (2.14)$$

dove $\pi_{\ell}^{i,j}$ è definito come l'elemento (i, j) della matrice $\mathbf{\Pi}^{\ell}$ e $\mathbf{\Pi}$ è la matrice di consenso il cui generico elemento (i, j) coincide con il peso del consenso $\pi^{i,j}$ (se $j \notin \mathcal{N}^i$ allora $\pi^{i,j}$ è posto uguale a 0). A tale riguardo, si ricorda che se i pesi di consenso $\pi^{i,j}$ sono scelti in modo da assicurare che la matrice $\mathbf{\Pi}$ sia primitiva e doppiamente stocastica², si ha

$$\lim_{\ell \rightarrow +\infty} \pi_{\ell}^{i,j} = \frac{1}{N}, \quad \forall i, j \in \mathcal{N}.$$

Per questo, quando il numero dei passi di consenso cresce, ogni PDF locale “tende” alla KLA collettiva (2.12).

Come è ben noto [46], condizione necessaria affinché la matrice $\mathbf{\Pi}$ sia primitiva è che il grafo \mathcal{G} associato con la rete dei sensori sia fortemente connesso, cioè per ogni coppia di nodi $i, j \in \mathcal{N}$ deve esserci un cammino da i a j e viceversa. Una possibile scelta, che garantisce, nel caso di grafi semplici, che la matrice dei pesi di consenso sia doppiamente stocastica, è quella dei cosiddetti *pesi di Metropolis* [1, 46]

$$\begin{aligned} \pi^{i,j} &= \frac{1}{\max\{|\mathcal{N}^i|, |\mathcal{N}^j|\}}, \quad i \in \mathcal{N}, j \in \mathcal{N}^i, i \neq j \\ \pi^{i,i} &= 1 - \sum_{j \in \mathcal{N}^i, j \neq i} \pi^{i,j}. \end{aligned}$$

¹Si noti che l'operatore di moltiplicazione per uno scalare è definito solo per scalari strettamente positivi. Comunque, nell'equazione (2.14) si ammette che qualcuno dei pesi scalari $\pi_{\ell}^{i,j}$ possa essere zero. S'intende che, qualora $\pi_{\ell}^{i,j}$ risulti nullo, la corrispondente PDF $p^j(\mathbf{x})$ sia omessa dalle somme. Questo può sempre essere fatto, dato che per ogni $i \in \mathcal{N}$ e ogni ℓ , almeno uno dei pesi $\pi_{\ell}^{i,j}$ è strettamente positivo.

²Si ricorda che una matrice quadrata non-negativa $\mathbf{\Pi}$ è doppiamente stocastica se tutte le sue righe e le sue colonne hanno somma pari a 1. Inoltre, è primitiva se esiste un intero m tale che tutti gli elementi di $\mathbf{\Pi}^m$ sono strettamente positivi.

Si noti che il consenso su PDF generiche introdotto in questo paragrafo costituisce un modo consistente, distribuito e scalabile per fondere l'informazione proveniente da più stimatori locali interconnessi; questo metodo ha applicabilità generale anche a problemi di stima distribuita non lineari e/o non Gaussiani. Inoltre, una caratteristica distintiva dell'approccio proposto, confrontata con le applicazioni usuali della GCI, è che qui le fusioni regionali sono effettuate con l'obiettivo di ottenere un risultato collettivo (cioè il calcolo della KLA collettiva).

Caso Gaussiano

Si focalizza adesso l'attenzione su una particolare famiglia parametrica di PDF, le distribuzioni normali. Comunque, considerazioni analoghe possono essere fatte anche per tutte quelle famiglie parametriche per le quali la KLD può essere calcolata in forma chiusa (tali famiglie includono quelle appartenenti alla classe esponenziale delle funzioni densità come la distribuzione esponenziale, la binomiale, ecc.).

Si suppone che tutte le PDF locali $p^i(\cdot)$ siano della forma

$$p^i(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i) \triangleq \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}^i)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}^i)^\top (\boldsymbol{\Sigma}^i)^{-1}(\mathbf{x}-\boldsymbol{\mu}^i)} \quad (2.15)$$

dove $\boldsymbol{\mu}^i \in \mathbb{R}^n$ è la media e $\boldsymbol{\Sigma}^i \in \mathbb{R}^{n \times n}$ è la matrice di covarianza (definita positiva). Allora, valgono i seguenti risultati.

Teorema 2. *Date PDF Gaussianhe come in (2.15) e i pesi π^i soddisfacenti (2.6), la loro KLA pesata in (2.9) ha la forma $\bar{p}(\cdot) = \mathcal{N}(\cdot; \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$ dove la media $\bar{\boldsymbol{\mu}}$ e la matrice di covarianza $\bar{\boldsymbol{\Sigma}}$ possono essere calcolate dalle seguenti relazioni algebriche*

$$\bar{\boldsymbol{\Sigma}}^{-1} = \sum_{i=1}^N \pi^i (\boldsymbol{\Sigma}^i)^{-1}, \quad (2.16)$$

$$\bar{\boldsymbol{\Sigma}}^{-1} \bar{\boldsymbol{\mu}} = \sum_{i=1}^N \pi^i (\boldsymbol{\Sigma}^i)^{-1} \boldsymbol{\mu}^i. \quad (2.17)$$

Dimostrazione: si veda l'Appendice A. \square

Ricordando che $\boldsymbol{\Omega}^i \triangleq (\boldsymbol{\Sigma}^i)^{-1}$ e $\mathbf{q}^i \triangleq (\boldsymbol{\Sigma}^i)^{-1} \boldsymbol{\mu}^i$ sono, rispettivamente, la *matrice d'informazione* ed il *vettore d'informazione*, associati con la PDF Gaussiana (2.15), il Teorema 2 afferma che la KLA di N PDF Gaussianhe può essere semplicemente ottenuto facendo la media delle matrici d'informazione e dei vettori d'informazione. Un'importante conseguenza di questo è che il Problema 1 può essere facilmente risolto iterando le medie pesate regionali attraverso l'algoritmo di consenso³

$$\boldsymbol{\Omega}_{\ell+1}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \boldsymbol{\Omega}_\ell^j, \quad i \in \mathcal{N}, \ell = 0, 1, \dots \quad (2.18)$$

$$\mathbf{q}_{\ell+1}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{q}_\ell^j, \quad i \in \mathcal{N}, \ell = 0, 1, \dots \quad (2.19)$$

³Seguendo il Teorema 2 si può vedere che le regole di aggiornamento (2.18) e (2.19) corrispondono al calcolo della KLA pesata tra la PDF locale e le PDF dei vicini.

Chiaramente, la ricorsione viene inizializzata ponendo

$$\mathbf{\Omega}_0^i = \mathbf{\Omega}^i, \quad i \in \mathcal{N} \quad (2.20)$$

$$\mathbf{q}_0^i = \mathbf{q}^i, \quad i \in \mathcal{N}. \quad (2.21)$$

Ricordando la precedente definizione di matrice di consenso $\mathbf{\Pi}$, la seguente proposizione deriva da ben noti risultati sul calcolo distribuito della media [1, 2].

Proposizione 1. *Sia la matrice di consenso $\mathbf{\Pi}$ primitiva e doppiamente stocastica. Allora, l'algoritmo di consenso (2.18)-(2.21) produce asintoticamente la KLA delle PDF locali iniziali*

$$\lim_{\ell \rightarrow \infty} \mathbf{\Omega}_\ell^i = \bar{\mathbf{\Omega}} \triangleq \frac{1}{N} \sum_{i=1}^N (\mathbf{\Sigma}^i)^{-1}$$

$$\lim_{\ell \rightarrow \infty} \mathbf{q}_\ell^i = \bar{\mathbf{q}} \triangleq \frac{1}{N} \sum_{i=1}^N (\mathbf{\Sigma}^i)^{-1} \boldsymbol{\mu}^i.$$

2.5 Filtraggio Bayesiano distribuito basato sul consenso

Il consenso sulle PDF può chiaramente essere sfruttato per effettuare, in modo distribuito e scalabile, il filtraggio Bayesiano su una rete di sensori. In questo paragrafo vengono presentati tre diversi approcci. Il primo, *consenso a posteriori* (CP: *Consensus on Posteriors*), viene ripreso da [17, 34]. Il secondo, *consenso sulla verosimiglianza* (CL: *Consensus on Likelihoods*), trae origine dal lavoro in [5, 6, 16]. Il terzo, *consenso parallelo su verosimiglianza e informazione a priori* (CLCP: *parallel Consensus on Likelihoods and Priors*) proposto in [47], evita qualunque ipotesi conservativa sulla correlazione fondendo separatamente l'informazione nuova e quella a priori.

Si considera il modello di rete introdotto nel paragrafo 2.1. Il problema della stima distribuita dello stato sulla rete di sensori $(\mathcal{S}, \mathcal{C}, \mathcal{A})$ può essere formulato nel modo seguente. Ogni nodo $i \in \mathcal{N}$ deve stimare, ad ogni istante di tempo $t \in \{1, 2, \dots\}$, lo stato \mathbf{x}_t del sistema dinamico

$$\mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t) + \mathbf{w}_t \quad (2.22)$$

date le misure locali

$$\mathbf{y}_k^i = \mathbf{h}_t^i(\mathbf{x}_k) + \mathbf{v}_k^i, \quad \text{se } i \in \mathcal{S} \quad (2.23)$$

per tutti i $k \geq 1$ fino al tempo t e i dati ricevuti da tutti i nodi adiacenti $j \in \mathcal{N}^i \setminus \{i\}$. Si suppone che $\mathbf{w}_t, \mathbf{v}_t^1, \mathbf{v}_t^2, \dots$ siano rumori bianchi a media nulla mutuamente indipendenti con PDF, eventualmente tempo-varianti, $p_{\mathbf{w}_t}(\cdot)$ e $\{p_{\mathbf{v}_t^i}(\cdot)\}_{i \in \mathcal{S}}$.

2.5.1 Consenso a posteriori

L'idea è di eseguire, ad ogni istante ed in ogni nodo della rete, un passo di correzione locale per determinare la PDF a posteriori locale seguito dal consenso su tali PDF a posteriori. Si supponga che, all'istante t , il nodo $i \in \mathcal{N}$ conosca la PDF a priori $p_{t|t-1}^i(\cdot)$. Allora, l'approccio del consenso a posteriori può essere riassunto nell'Algoritmo 13.

Come si vedrà nel capitolo successivo, nel caso lineare Gaussiano l'algoritmo CP implica un filtro di Kalman per i passi di correzione e predizione locali e l'intersezione delle covarianze per i passi di consenso. In questo caso, è stato dimostrato [17, 34] che, sotto l'ipotesi di connettività della rete e di osservabilità collettiva del sistema, l'algoritmo CP garantisce errore di stima limitato in media quadratica in tutti i nodi della rete e per un qualunque numero di passi di consenso. Nei casi non lineari e/o non Gaussiani, le PDF nei vari passi possono essere approssimate come Gaussiane e le corrispondenti medie e covarianze possono essere aggiornate per mezzo, ad esempio dell'EKF [30] o dell'UKF [32]. Vale la pena sottolineare che, comunque, ogni volta che il numero dei passi di consenso L tende all'infinito, il consenso a posteriori è incapace di trovare la soluzione del problema di filtraggio Bayesiano. Infatti, nell'approccio CP l'informazione nuova, venendo combinata nel consenso con l'informazione a priori, è inevitabilmente sottostimata.

Algoritmo 7: Filtro CP (Consensus on Posteriors) Bayesiano.

In ogni nodo $i \in \mathcal{N}$

Dati: $p_{1|0}^i(\mathbf{x})$

for $t = 1, 2, \dots$ **do**

Correzione locale

if $i \in \mathcal{S}$ **then**

 | si acquisisce la misura \mathbf{y}_t^i e si calcola

$$| p_{t|t}^i(\mathbf{x}) = p_{\mathbf{v}_t^i}(\mathbf{y}_t^i - \mathbf{h}_t^i(\mathbf{x})) \oplus p_{t|t-1}^i(\mathbf{x})$$

end if

if $i \in \mathcal{C}$ **then**

$$| p_{t|t}^i(\mathbf{x}) = p_{t|t-1}^i(\mathbf{x})$$

end if

Consenso

$$p_{t|t,0}^i(\mathbf{x}) = p_{t|t}^i(\mathbf{x})$$

for $\ell = 1, \dots, L$ **do**

$$| p_{t|t,\ell}^i(\mathbf{x}) = \bigoplus_{j \in \mathcal{N}^i} \left(\pi^{i,j} \odot p_{t|t,\ell-1}^j(\mathbf{x}) \right)$$

end for

Predizione

$$p_{t+1|t}^i(\mathbf{x}) = \int p_{\mathbf{w}_t}(\mathbf{x} - \mathbf{f}_t(\boldsymbol{\xi})) p_{t|t,L}^i(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

end for

2.5.2 Consenso sulla verosimiglianza

Un altro possibile approccio è quello di effettuare il consenso sull'informazione nuova (la verosimiglianza). I risultati sono poi combinati nel passo di correzione, segue quindi il passo di predizione. L'approccio proposto, di consenso sulla verosimiglianza, è riassunto nell'Algoritmo 14. Si noti che, nel passo di correzione, viene applicato al risultato del consenso sulla verosimiglianza un adeguato peso positivo $\rho_t^i > 0$, al fine di contrastare la sottoponderazione dell'informazione nuova.

Algoritmo 8: Filtro CL (Consensus on Likelihoods) Bayesiano.

In ogni nodo $i \in \mathcal{N}$

Dati: $p_{1|0}^i(\mathbf{x})$

for $t = 1, 2, \dots$ **do**

Consenso

if $i \in \mathcal{S}$ **then**

si acquisisce la misura \mathbf{y}_t^i e si calcola

$$l_{t,0}^i(\mathbf{x}) = p_{\mathbf{v}_t^i}(\mathbf{y}_t^i - \mathbf{h}_t^i(\mathbf{x}))$$

end if

if $i \in \mathcal{C}$ **then**

$$l_{t,0}^i(\mathbf{x}) = 1$$

end if

for $\ell = 1, \dots, L$ **do**

$$l_{t,\ell}^i(\mathbf{x}) = \bigoplus_{j \in \mathcal{N}^i} (\pi^{i,j} \odot l_{t,\ell-1}^j(\mathbf{x}))$$

end for

Correzione

$$p_{t|t}^i(\mathbf{x}) = p_{t|t-1}^i(\mathbf{x}) \oplus (\rho_t^i \odot l_{t,L}^i(\mathbf{x}))$$

Predizione

$$p_{t+1|t}^i(\mathbf{x}) = \int p_{\mathbf{w}_t}(\mathbf{x} - \mathbf{f}_t(\boldsymbol{\xi})) p_{t|t}^i(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

end for

2.5.3 Consenso parallelo su verosimiglianza e informazione a priori

Per rimediare agli inconvenienti degli approcci CP e CL, l'idea è di effettuare, in parallelo, il consenso, separatamente, sull'informazione nuova (verosimiglianza) e sull'informazione a priori. I risultati dei due consensi sono poi combinati nel passo di correzione, segue quindi il passo di predizione. L'approccio proposto, di consenso parallelo su verosimiglianza e informazione a priori è riassunto nell'Algoritmo 15. Si noti che, come nell'approccio CL, nel passo di correzione viene applicato al risultato del consenso sulla verosimiglianza un adeguato peso positivo $\rho_t^i > 0$ al fine di contrastare la sottoponderazione dell'informazione nuova.

Algoritmo 9: Filtro CLCP (Consensus on Likelihoods & Posteriors) Bayesiano.

In ogni nodo $i \in \mathcal{N}$

Dati: $p_{1|0}^i(\mathbf{x})$

for $t = 1, 2, \dots$ **do**

Consenso

if $i \in \mathcal{S}$ **then**

 | si acquisisce la misura \mathbf{y}_t^i e si calcola

 | $l_{t,0}^i(\mathbf{x}) = p_{\mathbf{v}_t^i}(\mathbf{y}_t^i - \mathbf{h}_t^i(\mathbf{x}))$

end if

if $i \in \mathcal{C}$ **then**

 | $l_{t,0}^i(\mathbf{x}) = 1$

end if

$p_{t|t,0}^i(\mathbf{x}) = p_{t|t}^i(\mathbf{x})$

for $\ell = 1, \dots, L$ **do**

 | $l_{t,\ell}^i(\mathbf{x}) = \bigoplus_{j \in \mathcal{N}^i} (\pi^{i,j} \odot l_{t,\ell-1}^j(\mathbf{x}))$

 | $p_{t|t-1,\ell}^i(\mathbf{x}) = \bigoplus_{j \in \mathcal{N}^i} (\pi^{i,j} \odot p_{t|t-1,\ell-1}^j(\mathbf{x}))$

end for

Correzione

$p_{t|t}^i(\mathbf{x}) = p_{t|t-1,L}^i(\mathbf{x}) \oplus (\rho_t^i \odot l_{t,L}^i(\mathbf{x}))$

Predizione

$p_{t+1|t}^i(\mathbf{x}) = \int p_{\mathbf{w}_t}(\mathbf{x} - \mathbf{f}_t(\boldsymbol{\xi})) p_{t|t}^i(\boldsymbol{\xi}) d\boldsymbol{\xi}$

end for

Si deve sottolineare che, sfortunatamente, i tre filtri Bayesiani precedentemente esposti (i.e. CP, CL, CLCP) non ammettono soluzione analitica esatta tranne che nel caso lineare Gaussiano. Algoritmi esatti (ottimi) per il caso lineare Gaussiano nonché algoritmi approssimati (subottimi) per il caso non lineare saranno esaminati nel capitolo successivo.

Capitolo 3

Algoritmi per la stima distribuita

Per la stima dello stato distribuita su reti di sensori esistono vari algoritmi proposti in letteratura, principalmente per il caso lineare ma anche per quello non lineare. Una prima famiglia di tecniche, a cui si fa riferimento come consenso sulle stime (CE: *Consensus on Estimates*), si basa sull'idea di diffondere l'informazione sulla rete effettuando, ad ogni istante di tempo, un consenso che media le stime/predizioni locali dello stato [8, 9, 10, 11, 12, 13, 14, 15]. La scelta di fare la media solo tra le stime dello stato ha lo scopo di mantenere lo scambio di informazioni tra i nodi vicini più limitato possibile. Tuttavia, poichè anche le matrici di covarianza contengono informazioni preziose che possono essere utilizzate per migliorare le prestazioni, sono stati proposti altri approcci più elaborati. Ad esempio, in [5, 6, 16], si propone di effettuare un consenso tra le misure locali e la covarianza delle innovazioni così da approssimare, in modo distribuito, la fase di correzione del filtro di Kalman centralizzato. Per coerenza con la terminologia introdotta precedentemente nel contesto Bayesiano, si farà riferimento a questo approccio come consenso sulla verosimiglianza (CL: *Consensus on Likelihoods*). In [17] è stato adottato un punto di vista diverso, proponendo il cosiddetto consenso sull'informazione a posteriori (CP: *Consensus on Posteriors*). Da un punto di vista algoritmico, nel caso lineare CP non è altro che il consenso applicato all'inversa della matrice di covarianza (matrice d'informazione) e al vettore d'informazione. Dal punto di vista della teoria dell'informazione, l'approccio CP può essere interpretato come consenso sulle PDF nel senso della media di Kullback-Leibler.

Si parte dall'osservazione che CL e CP mostrano possedere caratteristiche complementari. CP garantisce la stabilità per qualunque numero di passi di consenso (anche per un solo passo), ma le sue prestazioni in termini di errore quadratico medio della stima vengono limitate per il fatto che la regola di fusione adotta un punto di vista conservativo assumendo che la correlazione tra le stime provenienti da nodi diversi sia completamente sconosciuta. CL, viceversa, fondendo solo l'informazione nuova, non fa alcuna ipotesi conservativa sulla correlazione, ma non garantisce stabilità se il numero di passi di consenso non è sufficientemente elevato.

In questo capitolo vengono presentati alcuni algoritmi proposti in letteratura per la stima distribuita su reti di sensori sia per il caso lineare che per quello non lineare. In questo contesto, è stata sviluppata una nuova famiglia di algoritmi per la stima distribuita dello stato, chiamati algoritmi CLCP, in cui i due tipi di consenso CL e

CP vengono combinati in modo da mantenere le caratteristiche positive di entrambi. Si evidenzia che, per una particolare scelta dei pesi, l'approccio CLCP è equivalente all'algoritmo *Information Weighted Consensus* (IWC), presentato in [20]. Inoltre è stata effettuata un'analisi di stabilità, nel caso di sistemi lineari, sia del filtro CP che della famiglia di filtri CLCP insieme con una discussione su possibili estensioni di tale analisi al caso non lineare. Facendo seguito a [18], dove vengono presentati risultati preliminari sull'argomento, è stata effettuata una valutazione comparativa delle prestazioni di diversi filtri basati sul consenso relativamente ad un caso di studio simulativo di *target tracking*, con sensori sia lineari che non lineari.

Il capitolo è organizzato come segue. Nel paragrafo 3.1 vengono presentati algoritmi distribuiti basati sul consenso per il caso lineare Gaussiano. Nel successivo paragrafo 3.2, tali algoritmi vengono estesi al caso non lineare mediante gli approcci EKF ed UKF. Nel paragrafo 3.3 vengono studiate le proprietà di stabilità, limitatamente al caso lineare, degli algoritmi di consenso proposti. Infine, il paragrafo 3.4 presenta una valutazione comparativa dei vari algoritmi di consenso relativamente ad un caso di studio di tracking, con sensori lineari o non lineari.

3.1 Caso lineare Gaussiano

Si considera un sistema lineare tempo-discreto, in cui si suppone che la dinamica del sistema e l'equazione di misura del sistema siano lineari, ovvero

$$\mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{w}_t \quad (3.1)$$

il cui stato viene misurato da sensori lineari

$$\mathbf{y}_t^i = \mathbf{C}_t^i \mathbf{x}_t + \mathbf{v}_t^i, \quad i \in \mathcal{S}. \quad (3.2)$$

Inoltre, si suppone che lo stato iniziale, il disturbo di processo e tutti i rumori di misura siano distribuiti normalmente, cioè,

$$p_{1|0}(\cdot) = \mathcal{N}(\cdot; \hat{\mathbf{x}}_{1|0}, \mathbf{P}_{1|0}), \quad (3.3)$$

$$p(\cdot) = \mathcal{N}(\cdot; \mathbf{0}, \mathbf{Q}_t), \quad (3.4)$$

$$p_{\mathbf{v}_t^i}(\cdot) = \mathcal{N}(\cdot; \mathbf{0}, \mathbf{R}_t^i), \quad i \in \mathcal{S} \quad (3.5)$$

dove $\hat{\mathbf{x}}_{1|0}$ è un vettore noto e $\mathbf{P}_{1|0}$, \mathbf{Q}_t , \mathbf{R}_t^i sono matrici definite positive note. Come ben risaputo, in questo caso, la ricorsione di filtraggio Bayesiano ammette soluzione in forma chiusa dal fatto che, per ogni nodo i ed ogni istante $t \geq 1$, si ha

$$\begin{aligned} p_{t|t}^i(\mathbf{x}) &= \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{t|t}^i, \mathbf{P}_{t|t}^i), \\ p_{t+1|t}^i(\mathbf{x}) &= \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{t+1|t}^i, \mathbf{P}_{t+1|t}^i) \end{aligned}$$

dove i vettori $\hat{\mathbf{x}}_{t|t}^i$, $\hat{\mathbf{x}}_{t+1|t}^i$ e le matrici di covarianza $\mathbf{P}_{t|t}^i$, $\mathbf{P}_{t+1|t}^i$ possono essere calcolate ricorsivamente per mezzo del filtro di Kalman. Per semplificare la presentazione, è opportuno considerare il filtro di Kalman nella *forma di informazione* dove, invece delle stime e delle matrici di covarianza si propagano ricorsivamente le matrici di informazione

$$\mathbf{\Omega}_{t|t}^i = (\mathbf{P}_{t|t}^i)^{-1}, \quad \mathbf{\Omega}_{t+1|t}^i = (\mathbf{P}_{t+1|t}^i)^{-1}$$

e i vettori d'informazione

$$\mathbf{q}_{t|t}^i = \mathbf{\Omega}_{t|t}^i \hat{\mathbf{x}}_{t|t}^i, \quad \mathbf{q}_{t+1|t}^i = \mathbf{\Omega}_{t+1|t}^i \hat{\mathbf{x}}_{t+1|t}^i.$$

Con questa scelta, posto $\mathbf{V}_t^i \triangleq (\mathbf{R}_t^i)^{-1}$, il passo di correzione locale può essere scritto nel seguente modo

$$\mathbf{q}_{t|t}^i = \mathbf{q}_{t|t-1}^i + (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{y}_t^i, \quad (3.6)$$

$$\mathbf{\Omega}_{t|t}^i = \mathbf{\Omega}_{t|t-1}^i + (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i, \quad (3.7)$$

mentre, posto $\mathbf{W}_t \triangleq \mathbf{Q}_t^{-1}$, il passo di predizione prende la forma seguente

$$\mathbf{q}_{t+1|t}^i = \mathbf{A}_t^{-\top} \left[\mathbf{I} - \mathbf{\Omega}_{t|t}^i (\mathbf{\Omega}_{t|t}^i + \mathbf{A}_t^\top \mathbf{W}_t \mathbf{A}_t)^{-1} \right] \mathbf{q}_{t|t}^i, \quad (3.8)$$

$$\mathbf{\Omega}_{t+1|t}^i = \mathbf{A}_t^{-\top} \mathbf{\Omega}_{t|t}^i \mathbf{A}_t^{-1} - \mathbf{A}_t^{-\top} \mathbf{\Omega}_{t|t}^i (\mathbf{\Omega}_{t|t}^i + \mathbf{A}_t^\top \mathbf{W}_t \mathbf{A}_t)^{-1} \mathbf{\Omega}_{t|t}^i \mathbf{A}_t^{-1}. \quad (3.9)$$

Si noti che, perché le equazioni (3.8) e (3.9) abbiano senso, la matrice \mathbf{A}_t del sistema deve essere invertibile. Questa ipotesi è automaticamente soddisfatta nei sistemi a dati campionati in cui la matrice \mathbf{A}_t è ottenuta per discretizzazione della matrice di un sistema a tempo-continuo. Se la matrice \mathbf{A}_t non è invertibile ma è invertibile la matrice di covarianza del rumore di processo \mathbf{Q}_t , il passo di predizione (3.8)-(3.9) può essere effettuato mediante

$$\mathbf{q}_{t+1|t}^i = \mathbf{W}_t \mathbf{A}_t (\mathbf{\Omega}_{t|t}^i + \mathbf{A}_t^\top \mathbf{W}_t \mathbf{A}_t)^{-1} \mathbf{q}_{t|t}^i \quad (3.10)$$

$$\mathbf{\Omega}_{t+1|t}^i = \mathbf{W}_t - \mathbf{W}_t \mathbf{A}_t (\mathbf{\Omega}_{t|t}^i + \mathbf{A}_t^\top \mathbf{W}_t \mathbf{A}_t)^{-1} \mathbf{A}_t^\top \mathbf{W}_t \quad (3.11)$$

Se entrambe le matrici \mathbf{A}_t e \mathbf{Q}_t sono singolari, si può ricorrere ad algoritmi di filtraggio dell'informazione diversi, quali ad esempio il filtro d'informazione fattorizzato proposto in [48].

Sfruttando il suddetto filtro d'informazione e l'algoritmo di consenso (2.18)-(2.19), è possibile ottenere l'Algoritmo 10.

L'inizializzazione proposta nell'algoritmo presuppone che tutti i nodi della rete condividano la stessa informazione a priori sullo stato iniziale \mathbf{x}_1 . Si tratta di un'ipotesi piuttosto irrealistica; un'inizializzazione più realistica (e pratica) può essere fatta ponendo

$$\mathbf{\Omega}_{1|0}^i = 0, \quad i \in \mathcal{N} \quad (3.12)$$

$$\mathbf{q}_{1|0}^i = 0, \quad i \in \mathcal{N}, \quad (3.13)$$

che equivale a supporre che non sia disponibile alcuna informazione a priori.

Algoritmo 10: Algoritmo CP - Caso lineare Gaussiano.

Per ogni nodo $i \in \mathcal{N}$

Dati: $\mathbf{\Omega}_{1|0}^i = \mathbf{P}_{1|0}^{-1}$ e $\mathbf{q}_{1|0}^i = \mathbf{P}_{1|0}^{-1} \hat{\mathbf{x}}_{1|0}$

for $t = 1, 2, \dots$ **do**

Correzione:

if $i \in \mathcal{S}$ **then**

 si acquisisce la misura \mathbf{y}_t^i

$$\mathbf{\Omega}_{t|t,0}^i = \mathbf{\Omega}_{t|t-1}^i + (\mathbf{C}_t^i)^T \mathbf{V}_t^i \mathbf{C}_t^i$$

$$\mathbf{q}_{t|t,0}^i = \mathbf{q}_{t|t-1}^i + (\mathbf{C}_t^i)^T \mathbf{V}_t^i \mathbf{y}_t^i$$

end if

if $i \in \mathcal{C}$ **then**

$$\mathbf{\Omega}_{t|t,0}^i = \mathbf{\Omega}_{t|t-1}^i$$

$$\mathbf{q}_{t|t,0}^i = \mathbf{q}_{t|t-1}^i$$

end if

Consenso:

for $\ell = 1, \dots, L$ **do**

$$\mathbf{\Omega}_{t|t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{\Omega}_{t|t,\ell-1}^j$$

$$\mathbf{q}_{t|t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{q}_{t|t,\ell-1}^j$$

end for

Predizione:

 Si calcolano $\mathbf{q}_{t+1|t}^i$ e $\mathbf{\Omega}_{t+1|t}^i$ mediante (3.8)-(3.9) oppure mediante (3.10)-(3.11)

end for

È importante sottolineare che l'idea di fondere l'informazione ricevuta da più sensori facendo una combinazione convessa delle matrici e dei vettori d'informazione come nell'esecuzione dei passi di consenso ($\mathbf{\Omega}_{t|t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{\Omega}_{t|t,\ell-1}^j$ e $\mathbf{q}_{t|t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{q}_{t|t,\ell-1}^j$) non è nuova nella letteratura sulla fusione dell'informazione (sebbene la derivazione qui proposta sia molto diversa da quelle già esistenti). Infatti, questa regola di fusione è nota in letteratura come *Covariance Intersection*, [3, 4]. Ma la derivazione qui proposta mostra che una regola di fusione così diffusamente utilizzata discende in modo naturale come passo di un algoritmo di consenso per calcolare, in modo distribuito, la KLA delle PDF locali a posteriori.

Questo risultato potrebbe essere un punto di partenza per ricavare una generalizzazione di questa regola di fusione nei casi non lineare e/o non Gaussiano.

Si noti che l'algoritmo CP proposto per la stima distribuita dello stato presenta alcune differenze importanti se confrontato con altri approcci basati sul consenso presenti in letteratura. Infatti, le tecniche esistenti possono essere principalmente divisi in due famiglie: un primo gruppo [9, 10], in cui si ricerca il consenso tra i nodi della rete facendo la media solo sulle stime dello stato, trascurando così le matrici di covarianza così come la forma della PDF locale nei casi non lineare e/o non Gaussiano. Le tecniche del secondo gruppo [5, 6, 12] adottano un diverso punto di vista effettuando il consenso sulle coppie d'innovazione locale

$((\mathbf{C}_t^i)^\top (\mathbf{R}_t^i)^{-1} \mathbf{y}_t^i, (\mathbf{C}_t^i)^\top (\mathbf{R}_t^i)^{-1} \mathbf{C}_t^i)$ in modo da approssimare, in maniera distribuita, la fase di correzione del filtro di Kalman centralizzato. In questo ambito si colloca l'approccio CL presentato nel precedente capitolo che, nel caso lineare Gaussiano, dà luogo all'Algoritmo 11.

Algoritmo 11: Algoritmo CL - Caso lineare Gaussiano.

In ogni nodo $i \in \mathcal{N}$

Dati: $\mathbf{q}_{1|0}^i, \Omega_{1|0}^i$

for $t = 1, 2, \dots$ **do**

if $i \in \mathcal{S}$ **then**

 si acquisisce la misura \mathbf{y}_t^i

$$\delta \mathbf{q}_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{y}_t^i$$

$$\delta \Omega_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i$$

end if

if $i \in \mathcal{C}$ **then**

$$\delta \mathbf{q}_{t,0}^i = \mathbf{0}$$

$$\delta \Omega_{t,0}^i = \mathbf{0}$$

end if

Consenso

for $\ell = 1, 2, \dots, L$ **do**

$$\delta \mathbf{q}_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta \mathbf{q}_{t,\ell-1}^j$$

$$\delta \Omega_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta \Omega_{t,\ell-1}^j$$

end for

Correzione

$$\mathbf{q}_{t|t}^i = \mathbf{q}_{t|t-1}^i + \rho_t^i \delta \mathbf{q}_{t,L}^i$$

$$\Omega_{t|t}^i = \Omega_{t|t-1}^i + \rho_t^i \delta \Omega_{t,L}^i$$

Predizione:

 Si calcolano $\mathbf{q}_{t+1|t}^i$ e $\Omega_{t+1|t}^i$ mediante (3.8)-(3.9) oppure mediante

$$(3.10)-(3.11)$$

end for

La principale limitazione degli approcci CL è che la stabilità può essere garantita solo quando viene eseguito un numero sufficientemente elevato di passi di consenso ad ogni istante, così che l'informazione locale fornita dalle coppie d'innovazione abbia modo di diffondersi in tutta la rete. Inoltre, questo tipo di approccio si basa sull'assunzione che gli errori di misura relativi a sensori diversi siano mutuamente indipendenti e possono essere applicati a filtri tipo-Kalman. Come verrà anche chiarito nei prossimi paragrafi, l'approccio CP proposto non soffre di queste limitazioni e gode di forti proprietà di stabilità, grazie al fatto che le PDF a posteriori nel loro complesso sono combinate secondo il paradigma KLA (piuttosto che solo le stime dello stato o le coppie di innovazione). D'altra parte, non ci si

può aspettare che l'algoritmo proposto si avvicini al filtro di Kalman centralizzato in quanto la regola di fusione adottata segue una strategia cautelativa in modo da ottenere robustezza rispetto all'incesto dati. Quest'ultima considerazione suggerisce l'opportunità di combinare gli approcci CL e CP. A tale proposito l'approccio CLCP, introdotto nel precedente capitolo nel contesto generale del filtraggio Bayesiano, si riduce, nel caso lineare Gaussiano, all'Algoritmo 12.

Algoritmo 12: Algoritmo CLCP - Caso lineare Gaussiano.

In ogni nodo $i \in \mathcal{N}$

Dati: $\mathbf{q}_{1|0}^i, \mathbf{\Omega}_{1|0}^i$

for $t = 1, 2, \dots$ **do**

Consenso

if $i \in \mathcal{S}$ **then**

 si acquisisce la misura \mathbf{y}_t^i

$$\delta\mathbf{\Omega}_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i$$

$$\delta\mathbf{q}_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \bar{\mathbf{y}}_t^i$$

end if

if $i \in \mathcal{C}$ **then**

$$\delta\mathbf{\Omega}_{t,0}^i = \mathbf{0}$$

$$\delta\mathbf{q}_{t,0}^i = \mathbf{0}$$

end if

for $\ell = 1, \dots, L$ **do**

$$\delta\mathbf{\Omega}_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta\mathbf{\Omega}_{t,\ell-1}^j$$

$$\delta\mathbf{q}_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta\mathbf{q}_{t,\ell-1}^j$$

$$\mathbf{\Omega}_{t|t-1,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{\Omega}_{t|t-1,\ell-1}^j$$

$$\mathbf{q}_{t|t-1,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{q}_{t|t-1,\ell-1}^j$$

end for

Correzione

$$\mathbf{\Omega}_{t|t}^i = \mathbf{\Omega}_{t|t-1,L}^i + \rho_t^i \delta\mathbf{\Omega}_{t|t-1,L}^i$$

$$\mathbf{q}_{t|t}^i = \mathbf{q}_{t|t-1,L}^i + \rho_t^i \delta\mathbf{q}_{t|t-1,L}^i$$

Predizione

Si calcolano $\mathbf{q}_{t+1|t}^i$ e $\mathbf{\Omega}_{t+1|t}^i$ mediante (3.8)-(3.9) oppure mediante (3.10)-(3.11)

end for

Il paragrafo successivo è dedicato all'estensione degli algoritmi CP, CL e CLCP al caso di sistemi non lineari.

3.2 Caso non lineare

D'ora in poi si ipotizzerà che i rumori siano Gaussiani, cioè $p_{\mathbf{w}_t}(\cdot) = \mathcal{N}(\cdot; \mathbf{0}, \mathbf{Q}_t)$ e $p_{\mathbf{v}_t^i}(\cdot) = \mathcal{N}(\cdot; \mathbf{0}, \mathbf{R}_t^i)$, ma che il sistema, cioè $\mathbf{f}_t(\cdot)$ e/o $\mathbf{h}_t^i(\cdot)$, possa essere non lineare.

La questione principale è come operare con i sensori non lineari nel consenso sulla verosimiglianza dato che per le altre operazioni (cioè il consenso sull'informazione a priori, la correzione e la predizione) è ben noto come gestire le non linearità sfruttando, ad esempio, l'EKF o l'UKF o il filtro a particelle. Nell'ipotesi di Gaussianità, la verosimiglianza locale prende la forma

$$l_t^i(\mathbf{x}) = p_{\mathbf{v}_t^i}(\mathbf{y}_t^i - \mathbf{h}_t^i(\mathbf{x})) = \mathcal{N}(\mathbf{y}_t^i - \mathbf{h}_t^i(\mathbf{x}); \mathbf{0}, \mathbf{R}_t^i)$$

Ogni volta che il sensore i è lineare, cioè $\mathbf{h}_t^i(\mathbf{x}) = \mathbf{C}_t^i \mathbf{x}$,

$$l_t^i(\mathbf{x}) \propto \exp \left[-\frac{1}{2} (\mathbf{x}^\top \delta \boldsymbol{\Omega}_t^i - 2\mathbf{x}^\top \delta \mathbf{q}_t^i) \right]$$

dove

$$\delta \boldsymbol{\Omega}_t^i \triangleq (\mathbf{C}_t^i)^\top (\mathbf{R}_t^i)^{-1} \mathbf{C}_t^i, \quad \delta \mathbf{q}_t^i \triangleq (\mathbf{C}_t^i)^\top (\mathbf{R}_t^i)^{-1} \mathbf{y}_t^i \quad (3.14)$$

Allora è chiaro che moltiplicando, o elevando a potenza per un peso opportuno, le verosimiglianze $l_t^j(\cdot)$ è equivalente ad aggiungere, o moltiplicare per tali pesi, le corrispondenti $\delta \boldsymbol{\Omega}_t^i$ e $\delta \mathbf{q}_t^i$ definite in (3.14). Quindi, nel caso di sensori lineari Gaussiani, il consenso sulle verosimiglianze si riduce al consenso sulle coppie d'informazione definite in (3.14). Per sensori non lineari, un approccio ragionevole sembra pertanto di approssimare la funzione di misura non lineare $\mathbf{h}_t^i(\cdot)$ con una lineare affine, cioè

$$\mathbf{h}_t^i(\mathbf{x}) \cong \mathbf{C}_t^i (\mathbf{x} - \hat{\mathbf{x}}_{t|t-1}^i) + \hat{\mathbf{y}}_{t|t-1}^i \quad (3.15)$$

e poi sostituire l'equazione di misura $\mathbf{y}_t^i = \mathbf{h}_t^i(\mathbf{x}_t) + \mathbf{v}_t^i$ con $\bar{\mathbf{y}}_t^i = \mathbf{C}_t^i \mathbf{x}_t + \mathbf{v}_t^i$ per una pseudo-misura opportunamente definita

$$\bar{\mathbf{y}}_t^i = \mathbf{y}_t^i - \hat{\mathbf{y}}_{t|t-1}^i + \mathbf{C}_t^i \hat{\mathbf{x}}_{t|t-1}^i \quad (3.16)$$

Si vedrà adesso come effettuare la linearizzazione del sensore (3.15) in due diversi modi suggeriti, rispettivamente, dall'EKF e dall'UKF. Si supponga che all'istante t nel nodo i , sia presente una stima a priori $\hat{\mathbf{x}}_{t|t-1}^i$, allora la linearizzazione dell'EKF di $\mathbf{h}_t^i(\cdot)$ in un intorno di $\hat{\mathbf{x}}_{t|t-1}^i$ dà luogo a:

$$\mathbf{C}_t^i = \frac{\partial \mathbf{h}_t^i}{\partial \mathbf{x}} (\hat{\mathbf{x}}_{t|t-1}^i), \quad \hat{\mathbf{y}}_{t|t-1}^i = \mathbf{h}_t^i (\hat{\mathbf{x}}_{t|t-1}^i). \quad (3.17)$$

Sfruttando la trasformazione unscented, da $\hat{\mathbf{x}}_{t|t-1}^i$ e $\mathbf{P}_{t|t-1}^i$, si possono ottenere i σ -points $\hat{\mathbf{x}}_{t|t-1}^{i,j}$ e i relativi pesi w_j per $j = \pm 1, \dots, \pm n$ ($n = \dim \mathbf{x}$), e quindi calcolare

$$\begin{aligned} \hat{\mathbf{y}}_{t|t-1}^i &= \sum_{j=-n}^n w_j \hat{\mathbf{y}}_{t|t-1}^{i,j}, \quad \hat{\mathbf{y}}_{t|t-1}^{i,j} = \mathbf{h}_t^i (\hat{\mathbf{x}}_{t|t-1}^{i,j}) \\ \mathbf{P}_{t|t-1}^{\mathbf{y}\mathbf{x},i} &= \sum_{j=-n}^n w_j \left(\hat{\mathbf{y}}_{t|t-1}^i - \hat{\mathbf{y}}_{t|t-1}^{i,j} \right) \left(\hat{\mathbf{x}}_{t|t-1}^i - \hat{\mathbf{x}}_{t|t-1}^{i,j} \right)^\top \\ \mathbf{C}_t^i &= \mathbf{P}_{t|t-1}^{\mathbf{y}\mathbf{x},i} \left(\mathbf{P}_{t|t-1}^i \right)^{-1} \end{aligned} \quad (3.18)$$

3.2.1 Consenso a posteriori

La regola di fusione dell'intersezione di covarianze [3] suggerisce un metodo di consenso per la stima distribuita dello stato: il consenso sulla coppia d'informazione. Se al tempo t , ogni nodo $i \in \mathcal{N}$ dispone di una coppia d'informazione locale $(\Omega_{t|t-1}^i, \mathbf{q}_{t|t-1}^i)$, allora l'approccio CP, può essere riassunto dall'Algoritmo 13 da eseguire in ogni nodo $i \in \mathcal{N}$.

Si noti che l'Algoritmo 13 costituisce una generalizzazione al caso non lineare dell'Algoritmo 10.

Algoritmo 13: Algoritmo CP - Caso non lineare.

In ogni nodo $i \in \mathcal{N}$

Dati: $\mathbf{q}_{1|0}^i, \Omega_{1|0}^i$

for $t = 1, 2, \dots$ **do**

Correzione

if $i \in \mathcal{S}$ **then**

 si acquisisce la misura \mathbf{y}_t^i e si calcolano \mathbf{C}_t^i e $\hat{\mathbf{y}}_{t|t-1}^i$ per mezzo di (3.17)
 o di (3.18), e $\bar{\mathbf{y}}_t^i$ per mezzo di (3.16);

$$\Omega_{t|t,0}^i = \Omega_{t|t-1}^i + (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i$$

$$\mathbf{q}_{t|t,0}^i = \mathbf{q}_{t|t-1}^i + (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \bar{\mathbf{y}}_t^i$$

end if

if $i \in \mathcal{C}$ **then**

$$\Omega_{t|t,0}^i = \Omega_{t|t-1}^i$$

$$\mathbf{q}_{t|t,0}^i = \mathbf{q}_{t|t-1}^i$$

end if

Consenso

for $\ell = 1, 2, \dots, L$ **do**

$$\mathbf{q}_{t|t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{q}_{t|t,\ell-1}^j$$

$$\Omega_{t|t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \Omega_{t|t,\ell-1}^j$$

end for

Predizione

$$\mathbf{P}_{t|t}^i = \left(\Omega_{t|t,L}^i \right)^{-1}$$

$$\hat{\mathbf{x}}_{t|t}^i = \left(\Omega_{t|t,L}^i \right)^{-1} \mathbf{q}_{t|t,L}^i$$

da $\hat{\mathbf{x}}_{t|t}^i$ e $\mathbf{P}_{t|t}^i$ mediante EKF o UKF, si calcolano $\hat{\mathbf{x}}_{t+1|t}^i$ e $\mathbf{P}_{t+1|t}^i$

$$\Omega_{t+1|t}^i = \left(\mathbf{P}_{t+1|t}^i \right)^{-1} ; \quad \mathbf{q}_{t+1|t}^i = \left(\mathbf{P}_{t+1|t}^i \right)^{-1} \hat{\mathbf{x}}_{t+1|t}^i$$

end for

3.2.2 Consenso sulla verosimiglianza

In [5, 6] è stato proposto di sfruttare il consenso per calcolare in modo distribuito le quantità $\partial\boldsymbol{\Omega}_t \triangleq \sum_{i \in \mathcal{S}} (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i$ e $\partial\mathbf{q}_t \triangleq \sum_{i \in \mathcal{S}} (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{y}_t^i$ per il filtraggio alla Kalman lineare e distribuito. Come mostrato in [18], questo approccio, a cui si farà riferimento come *consenso sulla verosimiglianza* (CL: *Consensus on Likelihoods*), può essere esteso a sistemi non lineari. A questo scopo, può essere sfruttato il filtro d'informazione dell'Algoritmo 16, con la sola differenza che le misure virtuali $\bar{\mathbf{y}}_t^j$ devono essere ridefinite in termini di predizioni dello stato locale $\hat{\mathbf{x}}_{t|t-1}^j$ invece che della stima centralizzata $\hat{\mathbf{x}}_{t|t-1}$ (che non è disponibile in un contesto distribuito), per una spiegazione dettagliata si veda [18]. L'approccio DSE basato sul consenso sulla verosimiglianza è implementato nell'Algoritmo 14.

Si fa per il momento l'ipotesi irrealistica che un generico nodo i possa accedere alle misure di tutti i nodi $j \neq i$. In questo caso, posto $\mathbf{V}_t^j = (\mathbf{R}_t^j)^{-1}$, la correzione della stima dello stato può essere effettuata come segue

$$\hat{\mathbf{x}}_{t|t}^i = \hat{\mathbf{x}}_{t|t-1}^i + \left[(\mathbf{P}_{t|t-1}^i)^{-1} + \sum_{j \in \mathcal{S}} (\mathbf{C}_t^{j,i})^\top \mathbf{V}_t^j \mathbf{C}_t^{j,i} \right]^{-1} \times \sum_{j \in \mathcal{S}} (\mathbf{C}_t^{j,i})^\top \mathbf{V}_t^j [\mathbf{y}_t^j - \mathbf{h}_t^j(\hat{\mathbf{x}}_{t|t-1}^i)] \quad (3.19)$$

dove per ogni coppia di nodi sensori i e j , $\mathbf{C}_t^{j,i}$ vale

$$\mathbf{C}_t^{j,i} = \frac{\partial \mathbf{h}_t^j}{\partial \mathbf{x}} (\hat{\mathbf{x}}_{t|t-1}^i).$$

Inoltre, si considera l'approssimazione di Taylor al primo ordine:

$$\mathbf{h}_t^j(\hat{\mathbf{x}}_{t|t-1}^i) \cong \mathbf{h}_t^j(\hat{\mathbf{x}}_{t|t-1}^j) + \mathbf{C}_t^j (\hat{\mathbf{x}}_{t|t-1}^i - \hat{\mathbf{x}}_{t|t-1}^j) \quad (3.20)$$

$$\mathbf{C}_t^{j,i} \cong \mathbf{C}_t^j \triangleq \mathbf{C}_t^{j,j} = \frac{\partial \mathbf{h}_t^j}{\partial \mathbf{x}} (\hat{\mathbf{x}}_{t|t-1}^j). \quad (3.21)$$

Sostituendo le quantità approssimate (3.20) e (3.21) nell'equazione (3.19), si ottiene la seguente relazione:

$$\hat{\mathbf{x}}_{t|t}^i = \hat{\mathbf{x}}_{t|t-1}^i + \left[(\mathbf{P}_{t|t-1}^i)^{-1} + \sum_{j \in \mathcal{S}} (\mathbf{C}_t^j)^\top \mathbf{V}_t^j \mathbf{C}_t^j \right]^{-1} \times \sum_{j \in \mathcal{S}} (\mathbf{C}_t^j)^\top \mathbf{V}_t^j [\bar{\mathbf{y}}_t^j - \mathbf{C}_t^j \hat{\mathbf{x}}_{t|t-1}^i] \quad (3.22)$$

dove

$$\bar{\mathbf{y}}_t^j = \mathbf{y}_t^j - \mathbf{h}_t^j(\hat{\mathbf{x}}_{t|t-1}^j) + \mathbf{C}_t^j \hat{\mathbf{x}}_{t|t-1}^j. \quad (3.23)$$

Per avere la correzione nella forma del filtro d'informazione, è sufficiente moltiplicare a sinistra entrambi i membri di (3.22) per la matrice d'informazione $\boldsymbol{\Omega}_{t|t}^i$,

ottenendo così le seguenti equazioni di aggiornamento per il vettore e per la matrice d'informazione

$$\mathbf{q}_{t|t}^i = \mathbf{q}_{t|t-1}^i + \sum_{j \in \mathcal{S}} (\mathbf{C}_t^j)^\top \mathbf{V}_t^j \bar{\mathbf{y}}_t^j \quad (3.24)$$

$$\mathbf{\Omega}_{t|t}^i = \mathbf{\Omega}_{t|t-1}^i + \sum_{j \in \mathcal{S}} (\mathbf{C}_t^j)^\top \mathbf{V}_t^j \mathbf{C}_t^j. \quad (3.25)$$

Si noti che il consenso, quando converge, produce i valori medi $\partial \mathbf{\Omega}_t / |\mathcal{N}|$ e $\partial \mathbf{q}_t / |\mathcal{N}|$, dove $|\mathcal{N}|$ indica la cardinalità di \mathcal{N} (numero di nodi della rete), mentre l'aggiornamento del filtro d'informazione di fatto richiede $\partial \mathbf{\Omega}_t$ e $\partial \mathbf{q}_t$. Questo svantaggio può essere in parte compensato moltiplicando il risultato del consenso per adeguati pesi scalari ρ_t^i . Una scelta ragionevole, suggerita in [5, 6], consiste nel porre $\rho_t^i = |\mathcal{N}|$, così che al tendere all'infinito del numero dei passi di consenso, si può ritrovare il filtro EKF centralizzato. In pratica questa soluzione presenta alcuni svantaggi: (i) il numero dei nodi della rete può essere sconosciuto o difficile da determinare in modo distribuito; (ii) quando si effettua un numero finito di passi di consenso, la moltiplicazione per $|\mathcal{N}|$ può condurre in certi nodi a sovrastimare $\partial \mathbf{\Omega}_t$, una situazione da evitare se si vuol preservare la consistenza di ogni filtro locale¹. Una soluzione alternativa è quella di sfruttare il consenso in modo da calcolare, in modo distribuito, un fattore di normalizzazione capace di migliorare le prestazioni del filtro preservando al contempo la consistenza di ogni filtro locale. Ad esempio, una stima della frazione $|\mathcal{S}|/|\mathcal{N}|$ dei nodi sensori nella rete può essere calcolata per mezzo dell'algoritmo di consenso

$$b_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} b_{t,\ell-1}^j, \text{ for } \ell = 1, \dots, L \quad (3.26)$$

con l'inizializzazione

$$b_{t,0}^i = \begin{cases} 1, & \text{se } i \in \mathcal{S}, \\ 0, & \text{altrimenti.} \end{cases}$$

In tal caso, la scelta

$$\rho_t^i = \begin{cases} 1/b_{t,L}^i, & \text{se } b_{t,L}^i \neq 0 \\ 1 & \text{altrimenti} \end{cases} \quad (3.27)$$

ha la proprietà di preservare la consistenza di ogni filtro locale (si veda l'analisi riportata nel Paragrafo 3.3).

¹Si ricorda che un filtro si dice *consistente* quando la sua stima della covarianza dell'errore di stima è un limite superiore (nel senso della positiva definitezza) della covarianza dell'errore vero [3].

Algoritmo 14: Algoritmo CL - Caso non lineare.

In ogni nodo $i \in \mathcal{N}$

Dati: $\mathbf{q}_{1|0}^i, \mathbf{\Omega}_{1|0}^i$

for $t = 1, 2, \dots$ **do**

if $i \in \mathcal{S}$ **then**

 si acquisisce la misura \mathbf{y}_t^i e si calcolano \mathbf{C}_t^i e $\hat{\mathbf{y}}_{t|t-1}^i$ per mezzo di (3.17)
 o di (3.18), e $\bar{\mathbf{y}}_t^i$ per mezzo di (3.16);

$$\delta\mathbf{\Omega}_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i$$

$$\delta\mathbf{q}_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \bar{\mathbf{y}}_t^i$$

end if

if $i \in \mathcal{C}$ **then**

$$\delta\mathbf{q}_{t,0}^i = \mathbf{0}$$

$$\delta\mathbf{\Omega}_{t,0}^i = \mathbf{0}$$

end if

Consenso

for $\ell = 1, 2, \dots, L$ **do**

$$\delta\mathbf{q}_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta\mathbf{q}_{t,\ell-1}^j$$

$$\delta\mathbf{\Omega}_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta\mathbf{\Omega}_{t,\ell-1}^j$$

end for

Correzione

$$\mathbf{q}_{t|t}^i = \mathbf{q}_{t|t-1}^i + \rho_t^i \delta\mathbf{q}_{t,L}^i$$

$$\mathbf{\Omega}_{t|t}^i = \mathbf{\Omega}_{t|t-1}^i + \rho_t^i \delta\mathbf{\Omega}_{t,L}^i$$

Predizione:

$$\mathbf{P}_{t|t}^i = \left(\mathbf{\Omega}_{t|t}^i \right)^{-1}$$

$$\hat{\mathbf{x}}_{t|t}^i = \left(\mathbf{\Omega}_{t|t}^i \right)^{-1} \mathbf{q}_{t|t}^i$$

da $\hat{\mathbf{x}}_{t|t}^i$ e $\mathbf{P}_{t|t}^i$ mediante EKF o UKF, si calcolano $\hat{\mathbf{x}}_{t+1|t}^i$ e $\mathbf{P}_{t+1|t}^i$

$$\mathbf{\Omega}_{t+1|t}^i = \left(\mathbf{P}_{t+1|t}^i \right)^{-1} ; \quad \mathbf{q}_{t+1|t}^i = \left(\mathbf{P}_{t+1|t}^i \right)^{-1} \hat{\mathbf{x}}_{t+1|t}^i$$

end for

3.2.3 Consenso parallelo su verosimiglianza e informazione a priori

Il consenso sulla verosimiglianza (CL) ed il consenso a posteriori (CP) hanno caratteristiche complementari positive e negative. Per vedere questo, si indica con $\mathbf{\Pi}$ la matrice di consenso, i cui elementi sono i pesi di consenso $\pi^{i,j}$, $i, j \in \mathcal{N}$. Inoltre, sia $\pi_\ell^{i,j}$ l' (i, j) -esimo elemento di $\mathbf{\Pi}^\ell$, cioè l' ℓ -esima potenza della matrice di consenso $\mathbf{\Pi}$. Allora, il passo di correzione per gli algoritmi considerati può essere riscritto come in tab. 3.1. Si noti che, per i termini di innovazione $\delta\mathbf{q}_t^i$ e $\delta\mathbf{\Omega}_t^i$, le

Tabella 3.1: Coppie d'informazione nella fase di correzione dopo il consenso per gli algoritmi considerati.

	$\mathbf{q}_{t t}^i$	$\Omega_{t t}^i$
CP	$\sum_{j \in \mathcal{N}} \pi_L^{i,j} \mathbf{q}_{t t-1}^j + \sum_{j \in \mathcal{S}} \pi_L^{i,j} \delta \mathbf{q}_t^j$	$\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t t-1}^j + \sum_{j \in \mathcal{S}} \pi_L^{i,j} \delta \Omega_t^j$
CL	$\mathbf{q}_{t t-1}^i + \rho_t^i \sum_{j \in \mathcal{S}} \pi_L^{i,j} \delta \mathbf{q}_t^j$	$\Omega_{t t-1}^i + \rho_t^i \sum_{j \in \mathcal{S}} \pi_L^{i,j} \delta \Omega_t^j$
CLCP	$\sum_{j \in \mathcal{N}} \pi_L^{i,j} \mathbf{q}_{t t-1}^j + \rho_t^i \sum_{j \in \mathcal{S}} \pi_L^{i,j} \delta \mathbf{q}_t^j$	$\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t t-1}^j + \rho_t^i \sum_{j \in \mathcal{S}} \pi_L^{i,j} \delta \Omega_t^j$

sommatorie si estendono esclusivamente ai nodi sensori dato che, per tutti i nodi di comunicazione, si ha $\delta \mathbf{q}_t^j = 0$ e $\delta \Omega_t^j = 0$.

Come si può vedere dalla tab. 3.1, nel CP si fa una combinazione convessa dell'informazione locale a priori. Come mostrato in [17], questo è importante per garantire la limitatezza dell'errore di stima per un qualsiasi numero di passi di consenso (anche per uno solo). Al contrario, nell'algoritmo CL non si effettua alcun consenso sull'informazione locale a priori dato che viene combinata soltanto l'innovazione, $\delta \mathbf{q}_t^j$ and $\delta \Omega_t^j$. Per questo, la stabilità di questa famiglia di filtri distribuiti dipende in modo cruciale dal numero di passi di consenso L in ogni intervallo di campionamento. Infatti, se si indica con $\mathcal{S}^i(L)$ l'insieme di nodi sensori per i quali $\pi_L^{i,j}$ è diverso da 0, allora è immediato vedere che CL può dare un errore di stima limitato su un dato nodo i solo se si ha osservabilità (o almeno rilevabilità) da parte dell'insieme dei sensori $\mathcal{S}^i(L)$. In generale, questo è vero per un numero sufficientemente elevato di passi di consenso (si veda il paragrafo 3.3).

D'altra parte, con CP l'innovazione è inevitabilmente sottopesata, mentre con CL la moltiplicazione per i pesi scalari ρ_t^i può, almeno parzialmente, compensare questo svantaggio, migliorando così le prestazioni via via che cresce il numero di passi di consenso. Per avere i benefici e neutralizzare gli svantaggi degli approcci CP e CL, si possono combinare entrambi in modo da dar luogo ad un approccio a consenso ibrido, a cui si farà riferimento con l'acronimo CLCP (*Hybrid Consensus on Measurement + Consensus on Information*). CLCP effettua in parallelo i due tipi di consenso CP e CL già visti, ad ogni passo di consenso. L'implementazione di CLCP è riportata nell'Algoritmo 15.

Si noti che l'Algoritmo 15 rappresenta una famiglia di filtri distribuiti ciascuno corrispondente a una diversa scelta dei pesi scalari ρ_t^i . Come già detto, possibili scelte per i pesi scalari sono ad esempio $\rho_t^i = |\mathcal{N}|$ o quella dell'equazione (3.27). Quando si fa la prima scelta, l'algoritmo che ne risulta è equivalente a quello dell'*information weighted consensus* proposto in [49] nel caso lineare. Tale scelta ha l'interessante caratteristica di dar luogo ad un algoritmo distribuito che converge a quello centralizzato quando il numero dei passi di consenso L tende all'infinito. Inoltre, in questo caso, è possibile eseguire congiuntamente i due metodi di consenso dell'Algoritmo 15 in modo da risparmiare banda [49] (questo è vero ogni volta che i pesi ρ_t^i sono indipendenti dai nodi). Quando, invece, i pesi ρ_t^i sono calcolati come in (3.27), viene preservata la consistenza del filtro. Questo può essere visto guardando

l'ultima riga della tab. 3.1 e notando che l'innovazione non è mai sovrappesata.

Algoritmo 15: Algoritmo CLCP - Caso non lineare.

In ogni nodo $i \in \mathcal{N}$

Dati: $\mathbf{q}_{1|0}^i, \Omega_{1|0}^i$

for $t = 1, 2, \dots$ **do**

Consenso

if $i \in \mathcal{S}$ **then**

 si acquisisce la misura \mathbf{y}_t^i e si calcolano \mathbf{C}_t^i e $\hat{\mathbf{y}}_{t|t-1}^i$ per mezzo di (3.17)

 o di (3.18), e $\bar{\mathbf{y}}_t^i$ per mezzo di (3.16);

$$\delta\Omega_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i$$

$$\delta\mathbf{q}_{t,0}^i = (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \bar{\mathbf{y}}_t^i$$

end if

if $i \in \mathcal{C}$ **then**

$$\delta\Omega_{t,0}^i = \mathbf{0}$$

$$\delta\mathbf{q}_{t,0}^i = \mathbf{0}$$

end if

$$\Omega_{t|t-1,0}^i = \Omega_{t|t-1}^i$$

$$\mathbf{q}_{t|t-1,0}^i = \mathbf{q}_{t|t-1}^i$$

for $\ell = 1, \dots, L$ **do**

$$\delta\Omega_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta\Omega_{t,\ell-1}^j$$

$$\delta\mathbf{q}_{t,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \delta\mathbf{q}_{t,\ell-1}^j$$

$$\Omega_{t|t-1,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \Omega_{t|t-1,\ell-1}^j$$

$$\mathbf{q}_{t|t-1,\ell}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{q}_{t|t-1,\ell-1}^j$$

end for

Correzione

$$\Omega_{t|t}^i = \Omega_{t|t-1,L}^i + \rho_t^i \delta\Omega_{t,L}^i$$

$$\mathbf{q}_{t|t}^i = \mathbf{q}_{t|t-1,L}^i + \rho_t^i \delta\mathbf{q}_{t,L}^i$$

Predizione

$$\mathbf{P}_{t|t}^i = \left(\Omega_{t|t}^i \right)^{-1}$$

$$\hat{\mathbf{x}}_{t|t}^i = \left(\Omega_{t|t}^i \right)^{-1} \mathbf{q}_{t|t}^i$$

da $\hat{\mathbf{x}}_{t|t}^i$ e $\mathbf{P}_{t|t}^i$ mediante EKF o UKF, si calcolano $\hat{\mathbf{x}}_{t+1|t}^i$ e $\mathbf{P}_{t+1|t}^i$

$$\Omega_{t+1|t}^i = \left(\mathbf{P}_{t+1|t}^i \right)^{-1}; \quad \mathbf{q}_{t+1|t}^i = \left(\mathbf{P}_{t+1|t}^i \right)^{-1} \hat{\mathbf{x}}_{t+1|t}^i$$

end for

Infatti, i termini $\delta\mathbf{q}_t^j$ e $\delta\Omega_t^j$, per $i \in \mathcal{S}$, vengono moltiplicati per un peso $\rho_t^i \pi_L^{i,j} =$

$\pi_L^{i,j}/b_t^i(L)$ che, per costruzione, è garantito non superiori 1, dato che $b_t^i(L) = \sum_{i \in \mathcal{S}} \pi_L^{i,j}$.

3.3 Analisi di stabilità

Si esamineranno adesso le proprietà di stabilità, nel caso lineare tempo-invariante, degli algoritmi di stima distribuita dello stato CP e CLCP proposti. Pertanto si considera il sistema (3.1)-(3.5) con matrici $\mathbf{A}_t \equiv \mathbf{A}$, $\mathbf{C}_t^i \equiv \mathbf{C}^i$, $\mathbf{Q}_t \equiv \mathbf{Q}$, $\mathbf{R}_t^i = \mathbf{R}^i$ costanti.

Per l'analisi di stabilità si fanno le seguenti ipotesi preliminari.

A1. La matrice \mathbf{A} del sistema è invertibile.

A2. Il sistema è collettivamente osservabile dalla rete dei sensori, cioè è osservabile la coppia (\mathbf{A}, \mathbf{C}) con $\mathbf{C} \triangleq \text{col}(\mathbf{C}^1, \dots, \mathbf{C}^N)$.

Con $\text{col}(\mathbf{Z}^1, \dots, \mathbf{Z}^N)$ viene indicata la matrice ottenuta impilando una sopra l'altra le matrici $\mathbf{Z}^1, \dots, \mathbf{Z}^N$.

Occorre fare alcuni commenti sull'ipotesi A2. La ragione di richiedere l'osservabilità anziché la rilevabilità collettiva serve a semplificare l'analisi di stabilità. Tuttavia, questa può essere ripetuta senza perdita di generalità anche nel caso di rilevabilità collettiva, dato che per mezzo di un'opportuna trasformazione lineare di variabili

$$\mathbf{x}_t = \mathbf{T}\mathbf{z}_t$$

si può sempre considerare una decomposizione di osservabilità dello stato della forma

$$\begin{aligned} \mathbf{z}_{t+1} &= \tilde{\mathbf{A}}\mathbf{z}_t + \tilde{\mathbf{w}}_t, \\ \mathbf{y}_t^i &= \tilde{\mathbf{C}}^i\mathbf{z}_t + \mathbf{v}_t^i, \quad i \in \mathcal{N} \end{aligned}$$

con

$$\mathbf{z}_t = \begin{bmatrix} \mathbf{z}_t^o \\ \mathbf{z}_t^{\text{uo}} \end{bmatrix}, \quad \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}^o & 0 \\ \mathbf{A}^m & \mathbf{A}^{\text{uo}} \end{bmatrix}, \quad \tilde{\mathbf{C}}^i = [\mathbf{C}^{i,o} \mathbf{0}] ,$$

e la coppia $(\mathbf{A}^o, \mathbf{C}^o)$ con $\mathbf{C}^o \triangleq \text{col}(\mathbf{C}^{1,o}, \dots, \mathbf{C}^{N,o})$, osservabile. Successivamente, si può applicare l'algoritmo di stima distribuita proposto al solo sottosistema collettivamente osservabile per stimare \mathbf{z}_t^o . Quindi, si può calcolare ricorsivamente una stima dello stato inosservabile \mathbf{z}_t^{uo} mediante

$$\hat{\mathbf{z}}_{k|k}^{\text{uo}} = \mathbf{A}^{\text{uo}}\hat{\mathbf{z}}_{k-1|k-1}^{\text{uo}} + \mathbf{A}^m\hat{\mathbf{z}}_{k-1|k-1}^o .$$

Dato che, sotto rilevabilità globale, la matrice \mathbf{A}^{uo} è Schur-stabile, cioè i suoi autovalori sono nel cerchio unitario, la stabilità in errore quadratico medio di $\hat{\mathbf{z}}_{t|t}^{\text{uo}} - \mathbf{z}_t^{\text{uo}}$ implica automaticamente la medesima proprietà anche per $\hat{\mathbf{z}}_{t|t}^o - \mathbf{z}_t^o$.

3.3.1 Analisi di stabilità dell'algoritmo CP

Una delle caratteristiche fondamentali di cui deve godere un algoritmo di stima dello stato ricorsivo è che gli errori di stima veri siano consistenti con le loro statistiche predette. Infatti, come ben noto [50], la divergenza dell'errore di stima può avvenire quando la matrice di covarianza diventa troppo piccola o ottimistica (perché in questo caso le osservazioni successive tendono ad essere ignorate). A questo proposito, si può introdurre la seguente definizione [3, 50].

Definizione 2. Si consideri un vettore casuale \mathbf{x} . Inoltre, sia $\hat{\mathbf{x}}$ una stima non polarizzata di \mathbf{x} e \mathbf{P} una stima della corrispondente matrice di covarianza. Allora, la coppia $(\hat{\mathbf{x}}, \mathbf{P})$ si dice *consistente* se vale quanto segue

$$\mathbb{E} \{ (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\top \} \leq \mathbf{P}. \quad (3.28)$$

In altre parole, secondo la diseuguaglianza (3.28), consistenza significa richiedere che la covarianza dell'errore stimata \mathbf{P} sia un limite superiore (nel senso di positiva definitezza della matrice) della covarianza dell'errore vera. Questa proprietà diventa ancor più importante nella stima distribuita dello stato dal momento che l'inconsapevolezza del riuso degli stessi dati per la presenza di cicli nella rete così come la possibile correlazione tra misure di sensori diversi possono portare a inconsistenza o divergenza. Infatti, questa è stata la motivazione principale che ha portato allo sviluppo della regola di fusione *covariance intersection*.

Si noti che, se si considera la coppia d'informazione $(\mathbf{q}, \mathbf{\Omega}) = (\mathbf{P}^{-1}\hat{\mathbf{x}}, \mathbf{P}^{-1})$, la diseuguaglianza (3.28) può essere riscritta come

$$\mathbf{\Omega} \leq (\mathbb{E} \{ (\mathbf{x} - \mathbf{\Omega}^{-1}\mathbf{q})(\mathbf{x} - \mathbf{\Omega}^{-1}\mathbf{q})^\top \})^{-1}.$$

Allora, tenendo conto della relazione tra l'Algoritmo 10 e la *covariance intersection*, si può dimostrare il seguente risultato.

Teorema 3. Si suppone che valga l'ipotesi A1 e che venga usato l'Algoritmo 10 di stima distribuita dello stato con l'inizializzazione (3.12)-(3.13). Allora, per ogni istante $t = 0, 1, \dots$ e ogni nodo $i \in \mathcal{N}$, la coppia d'informazione $(\mathbf{q}_{t|t}^i, \mathbf{\Omega}_{t|t}^i)$ è consistente in quanto

$$\mathbf{\Omega}_{t|t}^i \leq (\mathbb{E} \{ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i)(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i)^\top \})^{-1}. \quad (3.29)$$

con $\hat{\mathbf{x}}_{t|t}^i \triangleq (\mathbf{\Omega}_{t|t}^i)^+ \mathbf{q}_{t|t}^i$.²

Dimostrazione: si veda l'Appendice A. \square

Il Teorema 3 evidenzia che, nel caso lineare, grazie all'adesione al PMDI, l'algoritmo di stima distribuita dello stato proposto evita il doppio conteggio delle informazioni e preserva la consistenza di tutte le coppie d'informazione. Inoltre,

²Qui viene considerata la pseudoinversa di Moore-Penrose della matrice d'informazione $\mathbf{\Omega}_{t|t}^i$, indicata come $(\mathbf{\Omega}_{t|t}^i)^+$, per tener conto del fatto che, in caso di PDF a priori non informative (3.12)-(3.13), tale matrice è singolare nei primi istanti di tempo.

in considerazione del Teorema 3, per dimostrare la limitatezza della covarianza dell'errore $\mathbb{E} \left\{ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i)(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i)^\top \right\}$ è sufficiente mostrare che asintoticamente la matrice d'informazione $\mathbf{\Omega}_{t|t}^i$ è limitata inferiormente da qualche matrice definita positiva (o, equivalentemente, mostrare che $\mathbf{P}_{t|t}^i = (\mathbf{\Omega}_{t|t}^i)^{-1}$ è asintoticamente limitata superiormente da qualche matrice costante). A questo proposito si può dimostrare il seguente teorema.

Teorema 4. *Si suppone che valgano le ipotesi A1-A2 e che venga usato l'Algoritmo 10 di stima distribuita dello stato con l'inizializzazione (3.12)-(3.13). Allora, se la matrice di consenso $\mathbf{\Pi}$ è primitiva e il numero dei passi di consenso L è maggiore di 0, esiste un istante di tempo \bar{t} e una matrice definita positiva $\tilde{\mathbf{\Omega}}$ tali che*

$$\mathbf{0} < \tilde{\mathbf{\Omega}} \leq \mathbf{\Omega}_{t|t}^i, \quad \forall i \in \mathcal{N} \quad \text{e} \quad \forall t \geq \bar{t}.$$

Come conseguenza, l'errore di stima è asintoticamente limitato in media quadratica dato che

$$\limsup_{t \rightarrow \infty} \mathbb{E} \left\{ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i)^\top (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i) \right\} \leq \text{tr}\{\tilde{\mathbf{\Omega}}^{-1}\}.$$

Dimostrazione: si veda l'Appendice A. \square

È opportuno fare alcune osservazioni sul teorema 4. Prima di tutto, è importante sottolineare che, per quanto a nostra conoscenza, questo è il primo risultato sulla stabilità per la stima distribuita dello stato basata sul consenso che si basa solo sull'ipotesi di osservabilità *collettiva*, cioè osservabilità dalla rete nel suo complesso. In realtà, anche le più recenti dimostrazioni di stabilità (riguardanti altri algoritmi di stima) richiedono qualche sorta di condizione di osservabilità (o di rilevabilità) *locale*, imponendo così forti ipotesi sulla qualità dei sensori e limitando l'applicabilità pratica (si veda, ad esempio, [9, 10, 12]). Tale avanzamento è dovuto principalmente al fatto che, nell'algoritmo proposto, le PDF a posteriori locali vengono combinate così che anche le matrici di covarianza sono aggiornate e mantenute consistenti con le stime combinate.

Vale anche la pena di notare che un tale risultato vale indipendentemente dalla possibile correlazione tra le misure provenienti da diversi sensori (che si suppone essere sconosciuta). Un'altra caratteristica di particolare rilievo è data dal fatto che è sufficiente un solo passo di consenso ad ogni iterazione per la stabilità.

Un'altra importante caratteristica dell'algoritmo di stima dello stato distribuita è che, quando le dinamiche del sistema così come le equazioni di misura sono senza rumore, è possibile dimostrare che tutti gli errori di stima locali $\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i$ vanno asintoticamente a zero. Per vedere questo, è utile considerare le funzioni di Lyapunov quadratiche tempo varianti

$$\mathcal{L}_t^i(\mathbf{x}) \triangleq \mathbf{x}^\top \mathbf{\Omega}_{t|t-1}^i \mathbf{x}, \quad i \in \mathcal{N}$$

e definire il vettore

$$\mathcal{L}_t = \text{col} \left(\mathcal{L}_t^1(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^1), \dots, \mathcal{L}_t^N(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^N) \right).$$

Allora, si può dimostrare il seguente teorema.

Teorema 5. *Siano le dinamiche del sistema (3.1) e le equazioni di misura (3.2) prive di rumori, cioè,*

$$\mathbf{w}_t = \mathbf{0}, \quad \mathbf{v}_t^i = \mathbf{0}, \quad i \in \mathcal{N}$$

per $t = 0, 1, \dots$. Allora, sotto le stesse ipotesi del Teorema 4, esiste un istante di tempo \bar{t} e un numero reale positivo $\beta < 1$ tali che

$$\mathcal{L}_{t+1} \leq \beta \mathbf{\Pi}^L \mathcal{L}_t,$$

per $t = \bar{t}, \bar{t} + 1, \dots$

Dimostrazione: si veda l'Appendice A. \square

Dato che, per costruzione, la matrice di consenso $\mathbf{\Pi}$ è stocastica per righe, il suo raggio spettrale è sempre 1. Come conseguenza, per $\beta < 1$, la matrice $\beta \mathbf{\Pi}^L$ risulta essere Schur-stabile. Questo, a sua volta, implica che tutte le componenti del vettore \mathcal{L}_t svaniscono quando t va all'infinito. Quindi, dato che tutte le matrici $\mathbf{\Omega}_{t|t-1}^i$ sono asintoticamente limitate dal basso da qualche matrice definita positiva (questo è diretta conseguenza del Teorema 4), si può dimostrare il seguente corollario al Teorema 5.

Corollario. *Sotto le stesse ipotesi del Teorema 5, l'algoritmo di stima distribuita dello stato proposto fornisce un osservatore asintotico in ogni nodo della rete, in quanto*

$$\lim_{t \rightarrow \infty} \mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i = 0, \quad \forall i \in \mathcal{N}.$$

3.3.2 Analisi di stabilità dell'algoritmo CLCP

In aggiunta alle precedenti assunzioni A1-A2, si considerano le seguenti ipotesi.

A3. La matrice di consenso $\mathbf{\Pi}$ è stocastica per righe e primitiva³.

A4. Esistono due scalari positivi $\underline{\rho}$ e $\bar{\rho}$ tali che $0 < \underline{\rho} \leq \rho_t^i \leq \bar{\rho}$, per ogni $i \in \mathcal{N}$ e $t \geq 0$.

Si noti che l'ipotesi A3 può sempre essere soddisfatta a condizione che la rete sia connessa. Ad esempio, in questo caso, i pesi di Metropolis [1, 46] soddisfano A3. Infine, si noti che entrambe le scelte proposte in precedenza per i pesi ρ_t^i soddisfano l'ipotesi A4.

Si introduce adesso il principale risultato sulla stabilità dell'algoritmo CLCP.

Teorema 6. *Si suppone che valgano le ipotesi A1-A4. Inoltre, si consideri l'algoritmo CLCP e sia $L \geq 1$. Infine, si ipotizza che, per ogni nodo $i \in \mathcal{N}$, la matrice d'informazione a priori $\mathbf{\Omega}_{1|0}^i$ sia definita positiva. Allora, l'errore di stima $\mathbf{e}_t^i = \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^i$ è asintoticamente limitato in media quadratica, cioè,*

$$\limsup_{t \rightarrow \infty} \mathbb{E} \{ \|\mathbf{e}_t^i\|^2 \} < +\infty,$$

per ogni $i \in \mathcal{N}$.

³Si ricorda che una matrice quadrata non negativa $\mathbf{\Pi}$ è stocastica per righe se la somma delle componenti di ogni sua riga è pari a 1. Inoltre, è primitiva se esiste un intero m tale che tutti gli elementi di $\mathbf{\Pi}^m$ sono strettamente positivi.

Dimostrazione: si veda l'Appendice A. \square

3.4 Risultati simulativi

Un modo per valutare le prestazioni di un qualunque algoritmo di stima distribuita è quello di confrontarlo con il corrispondente filtro centralizzato. Nel nostro caso, dato che si considerano sistemi non lineari, verrà utilizzato EKF o, in alternativa, UKF centralizzato (CEKF/CUKF), che si suppone possa elaborare contemporaneamente tutte le misure fatte dai vari sensori $\{\mathbf{y}_t^i, i \in \mathcal{S}\}$. Per ragioni di convenienza nelle situazioni multisensore, verrà adottato il filtro nella forma di informazione che propaga invece della stima $\hat{\mathbf{x}}_{t|t-1}$ e della covarianza $\mathbf{P}_{t|t-1}$, la coppia d'informazione, ovvero le *matrici d'informazione* (inverse delle matrici di covarianza)

$$\boldsymbol{\Omega}_{t|t-1} \triangleq \mathbf{P}_{t|t-1}^{-1}, \quad \boldsymbol{\Omega}_{t|t} \triangleq \mathbf{P}_{t|t}^{-1}$$

ed i *vettori d'informazione*

$$\mathbf{q}_{t|t-1} \triangleq \mathbf{P}_{t|t-1}^{-1} \hat{\mathbf{x}}_{t|t-1}, \quad \mathbf{q}_{t|t} \triangleq \mathbf{P}_{t|t}^{-1} \hat{\mathbf{x}}_{t|t}.$$

Introducendo le matrici d'informazione dei rumori $\mathbf{W}_t \triangleq \mathbf{Q}_t^{-1}$ e $\mathbf{V}_t^i \triangleq (\mathbf{R}_t^i)^{-1}$, si può ricavare CEKF/CUKF nella forma di informazione, cioè l'Algoritmo 16 che generalizza il filtro di Kalman, nella forma di informazione, a sistemi (2.22) e/o sensori (2.23) non lineari per mezzo del paradigma di linearizzazione (CEKF) oppure della trasformata unscented (CUKF).

Algoritmo 16: Algoritmo CEKF/CUKF nella forma di informazione.

In ogni nodo $i \in \mathcal{N}$;

Dati: Se si dispone di informazioni a priori le si utilizzano per inizializzare $\hat{\mathbf{x}}_{1|0}$, $\mathbf{P}_{1|0}$, $\boldsymbol{\Omega}_{1|0} = \mathbf{P}_{1|0}^{-1}$ e $\mathbf{q}_{1|0} = \boldsymbol{\Omega}_{1|0} \hat{\mathbf{x}}_{1|0}$, altrimenti si pone $\boldsymbol{\Omega}_{1|0} = \mathbf{0}$ e $\mathbf{q}_{1|0} = \mathbf{0}$

for $t = 1, 2, \dots$ **do**

Correzione;

In ogni nodo $i \in \mathcal{S}$ si acquisiscono le misure \mathbf{y}_t^i e le si raccolgono nel centro di fusione;

si calcolano \mathbf{C}_t^i e $\hat{\mathbf{y}}_{t|t-1}^i$ mediante (3.17) o (3.18), e $\bar{\mathbf{y}}_t^i$ per mezzo di (3.16);

$$\boldsymbol{\Omega}_{t|t} = \boldsymbol{\Omega}_{t|t-1} + \sum_{i \in \mathcal{S}} (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \mathbf{C}_t^i;$$

$$\mathbf{q}_{t|t} = \mathbf{q}_{t|t-1} + \sum_{i \in \mathcal{S}} (\mathbf{C}_t^i)^\top \mathbf{V}_t^i \bar{\mathbf{y}}_t^i;$$

Predizione;

$$\mathbf{P}_{t|t} = (\boldsymbol{\Omega}_{t|t})^{-1};$$

$$\hat{\mathbf{x}}_{t|t} = (\boldsymbol{\Omega}_{t|t})^{-1} \mathbf{q}_{t|t};$$

da $\hat{\mathbf{x}}_{t|t}$ e $\mathbf{P}_{t|t}$ per mezzo dell'EKF o dell'UKF, si calcolano $\hat{\mathbf{x}}_{t+1|t}$ e $\mathbf{P}_{t+1|t}$;

$$\boldsymbol{\Omega}_{t+1|t} = (\mathbf{P}_{t+1|t})^{-1}; \quad \mathbf{q}_{t+1|t} = (\mathbf{P}_{t+1|t})^{-1} \hat{\mathbf{x}}_{t+1|t};$$

end for

Uno degli obiettivi di questo lavoro è di indagare come tipi diversi di stimatori distribuiti dello stato basati sul consenso si confrontano tra loro e con lo stimatore centralizzato. A questo scopo, si considera un caso di studio relativo al tracciamento (tracking) di un singolo oggetto in movimento (target), in cui il moto di tale oggetto è descritto da un modello lineare (a velocità quasi costante) (vedi [17]) con matrici

$$\mathbf{A} = \begin{bmatrix} 1 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T_s \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \frac{T_s^3}{3} & \frac{T_s^2}{2} & 0 & 0 \\ \frac{T_s^2}{2} & T_s & 0 & 0 \\ 0 & 0 & \frac{T_s^3}{3} & \frac{T_s^2}{2} \\ 0 & 0 & \frac{T_s^2}{2} & T_s \end{bmatrix} q$$

dove $\mathbf{x}_t = [x_t, \dot{x}_t, y_t, \dot{y}_t]^\top$ è lo stato cinematico del target all'istante di campionamento t costituito dalle coordinate Cartesiane della posizione (x_t, y_t) e della velocità (\dot{x}_t, \dot{y}_t) ; l'intervallo di campionamento è stato fissato a $T_s = 1$ s e la varianza delle fluttuazioni casuali della velocità del target, q , a $0.5 \text{ m}^2/\text{s}^3$. Sono stati considerati due diversi scenari di simulazione corrispondenti a due diverse reti di sensori.

Nel primo scenario, si considera un'area di sorveglianza quadrata di 5000 m di lato; su tale area è distribuita la rete di monitoraggio, come illustrato in fig. 3.1. Si tratta di una rete geometrica con raggio di connettività massima $R_{max} = 799 \text{ m}$, e composta da 100 nodi di comunicazione e 5 nodi sensori di posizione lineari, caratterizzati dalla funzione di misura

$$\mathbf{h}^i(\mathbf{x}_t) = \mathbf{C}^i \mathbf{x}_t \quad \text{se } i \in \mathcal{S}$$

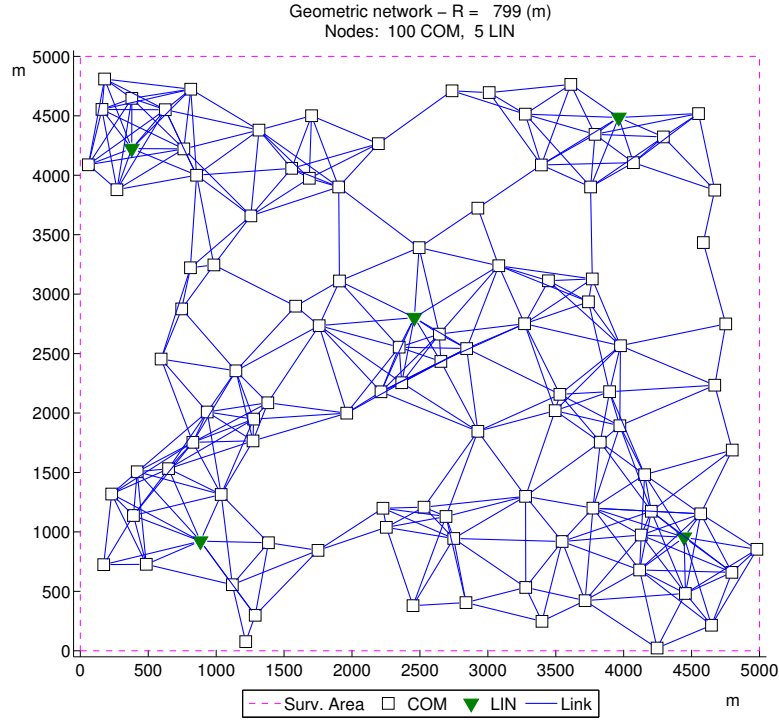


Figura 3.1: Rete di sensori lineare usata nelle simulazioni.

con

$$\mathbf{C}^i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Nel secondo scenario si ha la stessa regione di sorveglianza e una rete di monitoraggio geometrica con raggio di connettività massima $R_{max} = 810 m$ con nodi sensori di due tipi diversi, si tratta di sensori di posizione non lineari che misurano angoli o distanze. Questi due tipi di sensori, a cui si farà in seguito riferimento con gli acronimi DOA (Direction Of Arrival) e TOA (Time Of Arrival), sono caratterizzati dalle seguenti funzioni di misura:

$$\mathbf{h}^i(\mathbf{x}) = \begin{cases} \text{atan2}(x_1 - x_1^i, x_2 - x_2^i), & \text{se } i \text{ è un sensore DOA} \\ \sqrt{(x_1 - x_1^i)^2 + (x_2 - x_2^i)^2}, & \text{se } i \text{ è un sensore TOA} \end{cases}$$

dove atan2 è la funzione arcotangente a 4-quadranti e (x_1^i, x_2^i) indica la posizione, supposta nota, dell' i -esimo sensore. La rete consiste, complessivamente di 100 nodi di comunicazione, 5 nodi sensori TOA, e 5 nodi sensori DOA, come mostrato in fig. 3.2.

Per confrontare gli stimatori distribuiti tra loro e con quello centralizzato (utilizzato come termine di confronto) mediante esperimenti simulativi sono stati implementati sia i tre algoritmi proposti (i.e. CP, CL, CLCP) che lo stimatore centralizzato, CKF (Centralized Kalman Filter) nel caso lineare e CEKF (*Centralized EKF*) nel caso di sensori non lineari.

Si assume che il rumore di misura abbia deviazione standard $\sigma_{x_1} = \sigma_{x_2} = 20 m$ per i sensori lineari, $\sigma_\theta = 2^\circ$ per i sensori DOA, e $\sigma_r = 10 m$ for i sensori TOA.

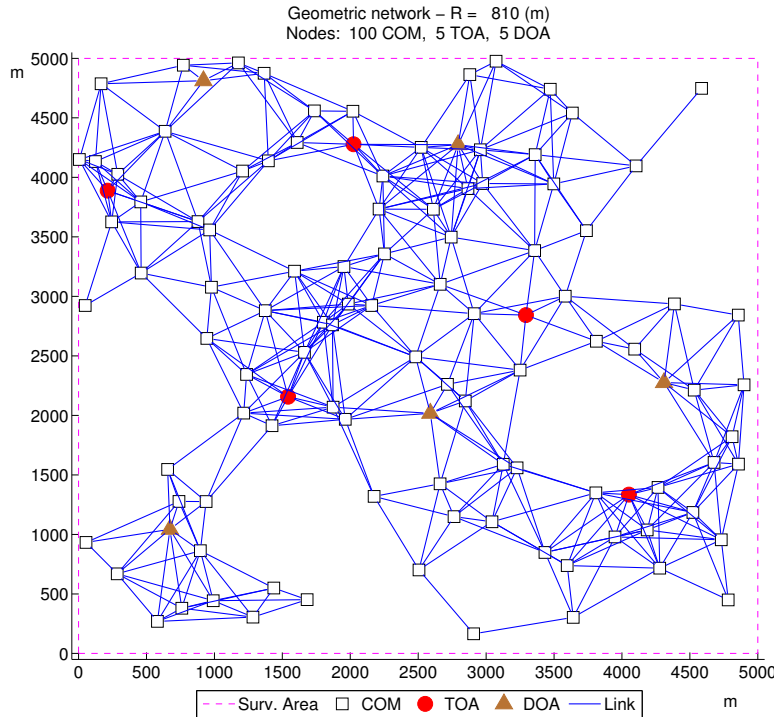


Figura 3.2: Rete di sensori non-lineare usata nelle simulazioni.

Infine, nelle simulazioni, come pesi di consenso sono stati scelti i pesi di Metropolis e come valori del numero di passi di consenso sono stati considerati valori di L da 1 a 9 in entrambi i casi di rete lineare e non lineare.

Per ciascuno degli scenari di simulazione considerati sono state eseguite 200 prove Monte Carlo indipendenti e si è calcolato come indice di prestazione la radice dell'errore quadratico medio di posizione (PRMSE: *Position Root Mean Square Error*) mediato su tutti i nodi della rete. Nelle figg. B.1-B.5 in Appendice B sono stati riportati i PRMSE nel caso di rete lineare per i valori di L considerati; la scala verticale è stata scelta in modo da evidenziare le differenze tra i vari filtri a regime.

Nelle figg. B.6-B.10 in Appendice B sono riportati i PRMSE nel caso di rete non lineare per i valori di L considerati. La scala verticale è stata scelta opportunamente in modo da evidenziare le differenze tra i vari filtri a regime; quando il PRMSE del CL non è riportato significa che il suo valore è fuori scala.

Infine, come indice di prestazione sintetico si è calcolato il PRMSE mediato sul tempo e su tutti i nodi della rete. Le figg. 3.3 e 3.4 mostrano il confronto tra i seguenti stimatori: CEKF, CL con pesi ρ_t^i come in (3.27), CP, CLCP con pesi ρ_t^i come in (3.27), IWC (corrispondente a CLCP con $\rho_t^i = |\mathcal{N}|$).

Si noti che il minimo numero di passi di consenso per i quali l'algoritmo CL mostra un comportamento stabile nelle due reti, quella con sensori lineari e quella con sensori non lineari, è $L = 4$ e, rispettivamente, $L = 5$. Conseguentemente i valori del PRMSE di CL per valori di L inferiori vengono omessi nei grafici in quanto tali valori risultano fuori scala. Si noti anche che un altro algoritmo di consenso CE di [12, 13], è stato considerato nelle simulazioni, ma per tale algoritmo i

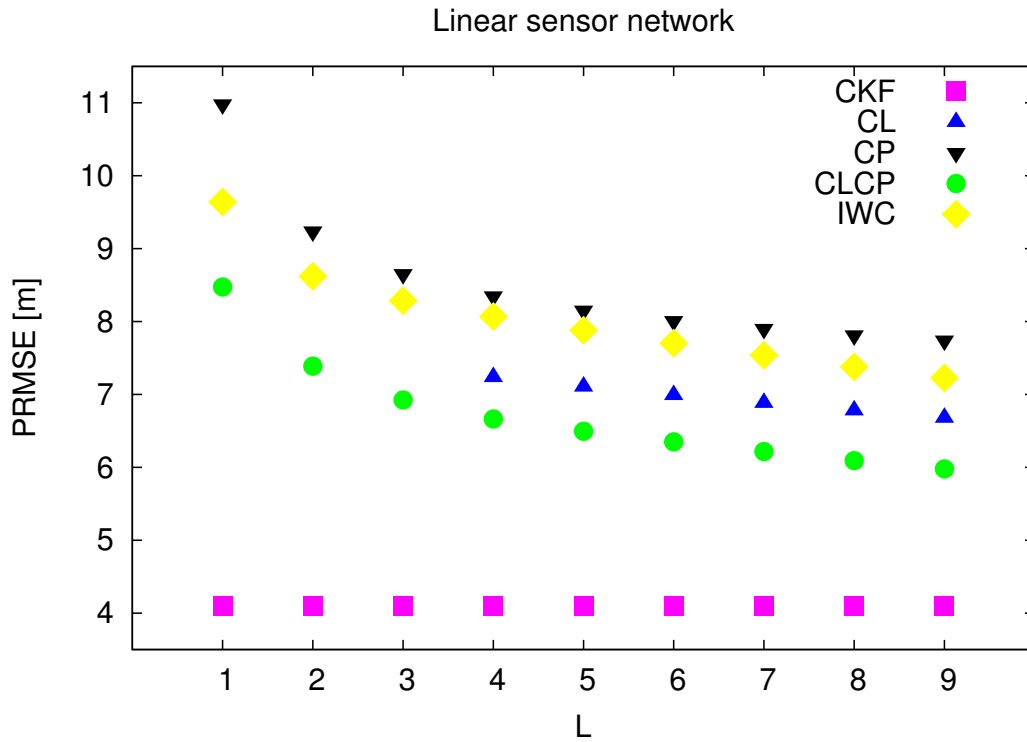


Figura 3.3: Caso lineare: PRMSE dei filtri considerati per diversi valori di L .

risultanti valori di PRMSE non sono stati riportati nei grafici dato che erano sempre maggiori di $10^2 m$ in tutte le configurazioni considerate. Come ci si aspettava, sia CLCP che IWC forniscono prestazioni soddisfacenti già per valori bassi di L con miglioramento nell'accuratezza della stima via via che cresce il numero dei passi di consenso. Inoltre confrontando CLCP e IWC, si può vedere che la scelta $\rho_t^i = |\mathcal{N}|$, che è asintoticamente ottimale (cioè quando $L \rightarrow +\infty$), non è necessariamente la scelta migliore quando si possono eseguire solo un numero limitato di passi di consenso ad ogni ciclo temporale.

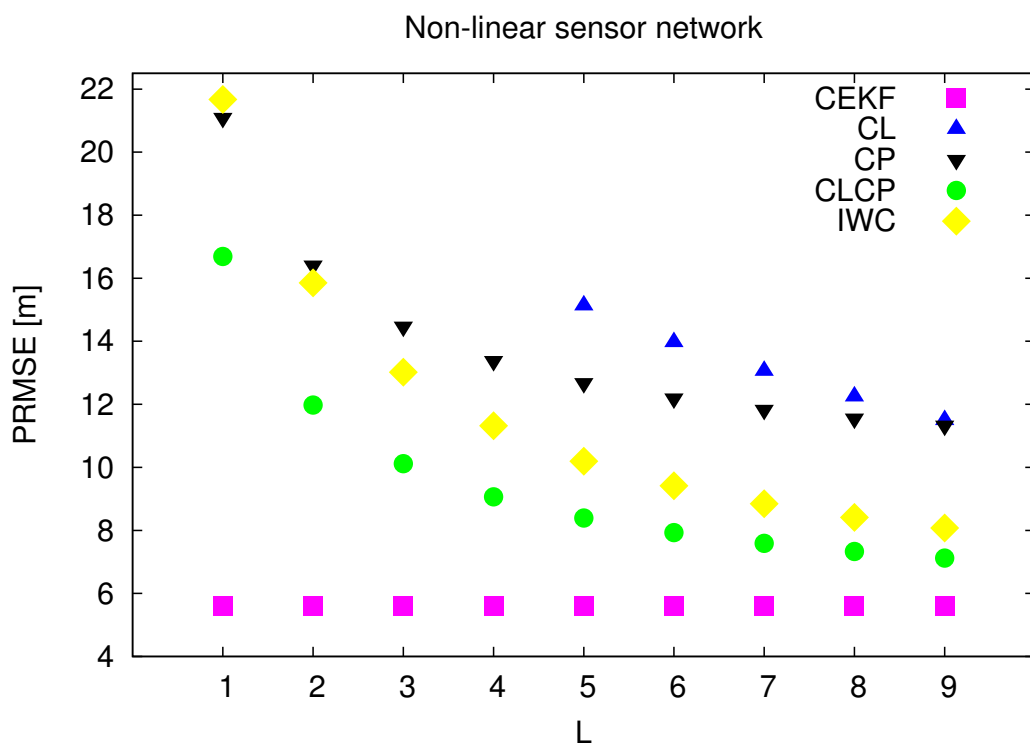


Figura 3.4: Caso non-lineare: PRMSE dei filtri considerati per diversi valori di L .

Parte II

Stima distribuita in presenza di virus

Capitolo 4

I virus nelle reti di sensori

Le reti di sorveglianza wireless rappresentano una tecnologia efficace per la realizzazione di sistemi di monitoraggio scalabili, flessibili, a basso costo e tolleranti ai guasti. Purtroppo, però, i sistemi di sorveglianza distribuiti sono suscettibili di attacchi distruttivi da parte di virus informatici. In questo capitolo viene affrontato lo studio di questa problematica da più punti di vista. In primo luogo viene trattata la modellazione di un'epidemia virale in una rete di monitoraggio e si studiano i suoi effetti su un algoritmo di stima distribuita. In secondo luogo, saranno analizzate le correlazioni tra la connettività di rete, le prestazioni di un algoritmo di stima distribuita, e la diffusione del virus. Infine, è stata effettuata un'estesa campagna di simulazioni su uno scenario di sorveglianza realistico allo scopo di investigare il modo in cui i parametri della rete influenzano la diffusione dei virus e la fusione dell'informazione distribuita. Questo ha fornito utili linee guida per preservare le funzionalità di una rete di monitoraggio soggetta a guasti e/o attacchi virali.

4.1 Modelli di diffusione dei virus nelle reti

La prima questione da affrontare è quella di modellare la diffusione virale nelle reti. A questo proposito, in letteratura sono stati proposti diverse classi di modelli di differente complessità. Ad esempio, si possono considerare *modelli macroscopici* [22, 23] che descrivono il comportamento globale di un'epidemia virale per mezzo di equazioni differenziali ordinarie o, invece, *modelli microscopici* [24, 25] dove tutti i nodi e gli archi della rete vengono considerati separatamente. A sua volta, quest'ultimo può essere suddiviso in *modelli a catena di Markov singola* esatti, che comportano un numero di stati esponenziale, e *modelli a catena di Markov incrociata* semplificati.

Tuttavia, quasi tutti i modelli esistenti sono basati su varianti del modello classico suscettibile-infetto-recuperato per la diffusione delle epidemie nelle popolazioni. Questo tipo di modelli di diffusione dei virus [51] si basa sulla catena di Markov riportata in fig. 4.1. In particolare, si suppone che ciascun nodo della rete possa essere in uno dei seguenti stati:

- *suscettibile* (S), se il nodo non è né infetto né immune;
- *infetto* (I), se il nodo è infetto e può trasmettere il virus;

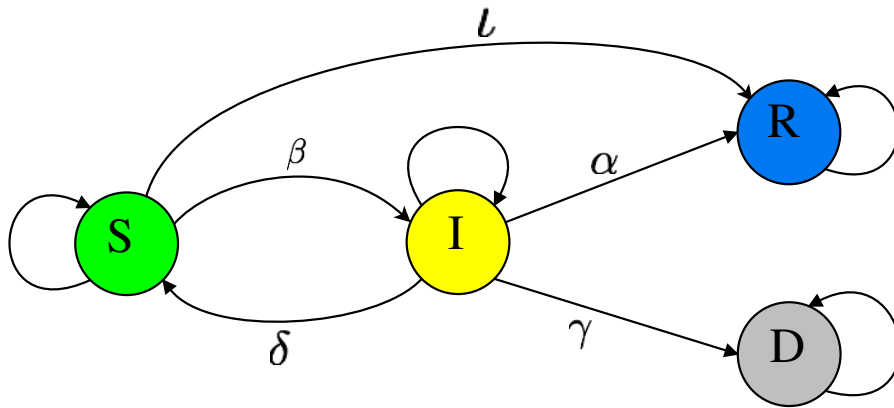


Figura 4.1: Modello a catena di Markov per la diffusione dei virus.

- *recuperato* (R), se il nodo non è infetto ed è immune;
- *morto* (D), se il nodo ha smesso di funzionare a causa dell'infezione virale e non può più trasmettere l'epidemia.

Inoltre, le transizioni tra gli stati sono governate dalle seguenti probabilità

- *probabilità d'infezione* β , associata alla transizione dallo stato S ad I ;
- *probabilità d'immunizzazione* ι , associata alla transizione dallo stato S ad R ;
- *probabilità di guarigione* δ , associata alla transizione dallo stato I ad S ;
- *probabilità di recupero* α , associata alla transizione dallo stato I ad R ;
- *probabilità di morte* γ , associata alla transizione dallo stato I a D .

Tipicamente, vengono considerati modelli semplificati in cui alcuni dei fenomeni suddetti vengono ignorati. Ad esempio, il *modello SIS* considera che i nodi possano essere negli stati S o I : un nodo suscettibile può divenire infetto con probabilità β e poi guarire con probabilità δ . Naturalmente, questo coincide col porre $\alpha = 0$, $\iota = 0$ e $\gamma = 0$ nella catena di Markov di fig. 4.1. Un'altro modello che ha ricevuto attenzione speciale è il *modello SIR*, ottenuto da quello generale di fig. 4.1 ponendo $\delta = 0$, $\iota = 0$ e $\gamma = 0$. Questo significa che un nodo infetto, una volta guarito, risulta immunizzato al virus entrando nello stato terminale R .

Si possono considerare varie alternative per modellare come si manifestino i virus nei nodi della rete: il *modello di virus non cooperativo*, in cui i nodi infetti smettono di prendere parte alla stima distribuita; il *modello di virus a comportamento deviante*, in cui i nodi infetti prendono parte a all'algoritmo di consenso con dati generati casualmente; e il *modello di virus malizioso*, in cui i nodi infetti prendono parte al consenso con dati generati opportunamente per compromettere il funzionamento dell'algoritmo di consenso. In questo lavoro saranno presi in considerazione solo virus non cooperativi.

4.1.1 Modelli macroscopici

Sia $i(t)$ la frazione di nodi infetti al tempo t . L'idea su cui si basano i modelli macroscopici è che, quando la dimensione della rete N è sufficientemente grande, $i(t)$ può essere vista come una quantità continua la cui evoluzione temporale può essere descritta da equazioni differenziali ordinarie. Ad esempio, per una rete con grado medio \bar{k} e con modello epidemico SIS, Kephart e White [22] hanno proposto di rappresentare l'evoluzione temporale di $i(t)$ per mezzo dell'equazione differenziale

$$\frac{di(t)}{dt} = \beta \bar{k} i(t) [1 - i(t)] - \delta i(t). \quad (4.1)$$

Questo modello può essere ricavato ipotizzando che per ogni nodo suscettibile la probabilità d'infezione sia ottenibile moltiplicando la probabilità d'infezione per collegamento β con il numero di nodi vicini infetti che, a sua volta, può essere approssimato con $\bar{k} i(t)$.

Nonostante la sua semplicità, questo modello si è rivelato sorprendentemente utile nel predire l'esistenza di una *soglia epidemica* (cioè, una condizione in cui un virus può prendere il sopravvento e creare un'epidemia invece che estinguersi). Infatti, si può vedere che l'equazione (4.1) dà luogo a due diversi comportamenti a seconda del valore del rapporto β/δ . Più specificamente, quando $\beta/\delta < 1/\bar{k}$ l'epidemia svanisce asintoticamente dato che $i = 0$ è un punto di equilibrio asintoticamente stabile con dominio di attrazione che include $[0, 1]$; invece, quando $\beta/\delta > 1/\bar{k}$, la frazione di nodi infetti converge al valore di regime $1 - \delta/(\beta\bar{k})$. Quindi, si può vedere che per il modello di Kephart-White la soglia epidemica risulta essere

$$\frac{\beta}{\delta} \geq \frac{1}{\bar{k}}.$$

Il modello di Kephart-White (4.1) può portare a predizioni ragionevoli per reti omogenee in cui tutti i nodi hanno approssimativamente lo stesso grado, ma in generale è inadeguato a modellare la diffusione epidemica in reti non omogenee. Così, in letteratura sono state proposte varie modifiche per tener conto della topologia della rete. Ad esempio, in [52, 53] è stato proposto di usare modelli diversi per le dinamiche di nodi aventi grado diverso. Più specificamente, sia $P(k)$ la frazione di nodi con grado k e $i_k(t)$ la frazione di tali nodi che è infetta al tempo t . Allora, si suppone che l'evoluzione temporale di $i_k(t)$ possa essere descritta per mezzo dell'equazione differenziale

$$\frac{di_k(t)}{dt} = \beta k \Theta(t) [1 - i_k(t)] - \delta i_k(t) \quad (4.2)$$

dove $\Theta(t)$ rappresenta la probabilità che un arco colleghi ad un nodo infetto ed è data da

$$\Theta(t) = \frac{\sum_k (k-1) P(k) i_k(t)}{\sum_k k P(k)}.$$

Sfortunatamente, come evidenziato in [23], anche questo modello macroscopico topologico è intrinsecamente inaccurato a causa di un'implicita assunzione di miscelazione omogenea. Infatti la variabile $\Theta(t)$, supponendo implicitamente che i nodi infetti siano uniformemente distribuiti nella rete, non tiene conto di come

tali nodi sono connessi tra loro: se sono raggruppati insieme o sparpagliati nella rete. Per questo motivo, come mostrato in [23] per mezzo di un'estesa analisi di simulazioni, questo tipo di modelli tende a sovrastimare la velocità di propagazione dell'epidemia.

Quanto al modello di diffusione virale SIR, al posto di (4.1) si possono considerare le due equazioni differenziali accoppiate

$$\frac{di(t)}{dt} = \beta \bar{k} i(t) s(t) - \delta i(t), \quad (4.3)$$

$$\frac{dr(t)}{dt} = \delta i(t) \quad (4.4)$$

dove $r(t)$ e $s(t)$ indicano, rispettivamente, la frazione dei nodi recuperati e quella dei nodi suscettibili al tempo t con $s(t) + i(t) + r(t) = 1$, per ogni t . Una prima differenza importante tra il modello SIR descritto dalle equazioni (4.3)-(4.4) e quello SIS è che nel precedente l'epidemia si estingue sempre via via che il tempo passa (va all'infinito), a prescindere dalle condizioni iniziali e dalle probabilità d'infezione e di recupero. Questo significa che si deve introdurre una diversa nozione di soglia epidemica. A tal riguardo, una scelta ragionevole è di definire la soglia epidemica come la condizione in cui la frazione di nodi infetti $i(t)$ cresce (o, meglio, ci si aspetta che cresca) rispetto al valore iniziale $i(0)$. Adottando questa definizione, risulta che la soglia epidemica dipende dalle condizioni iniziali in quanto il *picco epidemico* $\max_t i(t)$ supera il valore iniziale $i(0)$ se e solo se $\beta/\alpha > 1/[\bar{k}s(0)]$.

Supponendo, tuttavia, che la dimensione della rete N sia sufficientemente grande e che, al tempo 0, solo pochi nodi siano infetti, si può approssimare $s(0)$ con 1 e concludere che la condizione di esplosione epidemica sia

$$\frac{\beta}{\alpha} \geq \frac{1}{\bar{k}}$$

che è della stessa forma di quella del modello SIS con la probabilità di guarigione δ sostituita da quella di recupero α .

4.1.2 Modello a catena di Markov singola esatta

In questo modello, la diffusione dell'epidemia virale nella rete è descritto mediante una sola catena di Markov il cui spazio degli stati è rappresentato da tutte le possibili configurazioni in cui i nodi della rete possono trovarsi. Come dovrebbe essere evidente, questo modello implica un numero esponenziale di stati possibili: 4^N stati per il modello generale di fig. 4.1, 3^N per il modello SIR, 2^N stati per quello SIS. Si sottolinea il fatto che questa complessità esponenziale è inevitabile per fornire una rappresentazione esatta delle dinamiche epidemiche virali. Infatti non è possibile considerare i nodi della rete come entità separate indipendenti dato che, in generale, la probabilità β_i che un nodo suscettibile i venga infettato dovrebbe dipendere dal numero di nodi vicini infetti N_i^I . Ad esempio, supponendo l'indipendenza tra gli eventi, tale probabilità può essere espressa come

$$\beta_i = 1 - (1 - \beta)^{N_i^I} \quad (4.5)$$

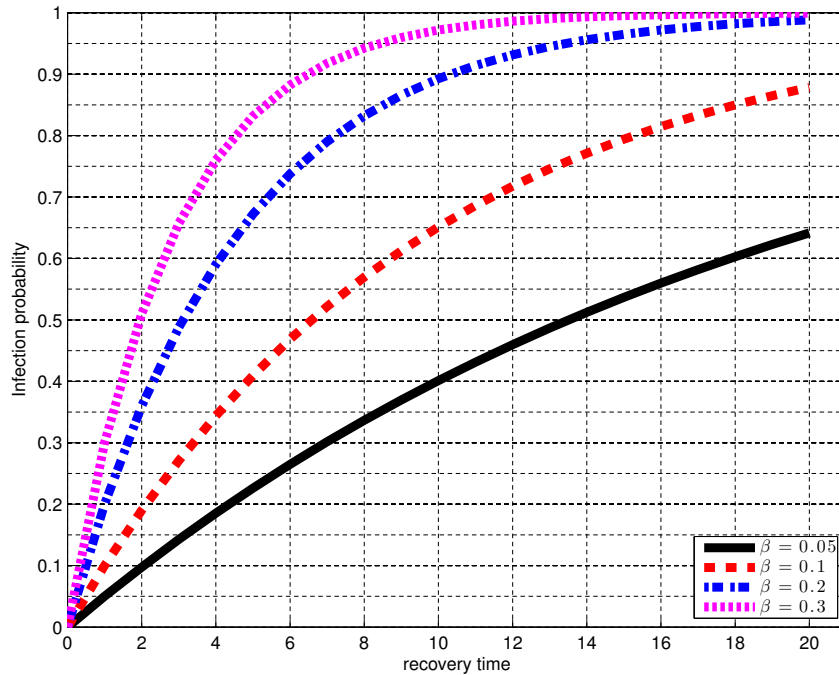


Figura 4.2: Probabilità d'infezione in funzione di N_i^I per alcuni valori di β .

dove β è la probabilità d'infezione da ciascun vicino (infetto). La dipendenza delle probabilità d'infezione in funzione di N_i^I per alcuni valori di β è raffigurata in fig. 4.2.

Quanto alla probabilità di recupero α_i di un nodo infetto i , in generale essa dipende dal tempo trascorso da quando il nodo è stato infettato $t - T_i^I$ (dove T_i^I è l'istante d'infezione del nodo i). Tipici modelli semplificati sono il modello a *tempo di recupero costante*, dove $\alpha_i = 1$ se $t - T_i^I$ supera una certa soglia e $\alpha_i = 0$ altrimenti, e il modello a *tempo di recupero distribuito geometricamente*, dove ad ogni istante di tempo discreto la probabilità di recupero istantanea è costante ed uguale a α . D'ora in poi, si farà sempre riferimento a quest'ultimo modello, per il quale il tempo di recupero ha un valore atteso pari a $1/\alpha$ e varianza $(1 - \alpha)/\alpha^2$. La fig. 4.3 mostra la probabilità del tempo di recupero per alcuni valori di α . Analoghe considerazioni possono essere fatte anche per la probabilità di guarigione δ_i .

In linea di principio, il modello a catena di Markov esatta contiene tutte le informazioni necessarie per predire il comportameneto epidemico. Sfortunatamente, il numero degli stati della catena di Markov è esponenziale e così l'applicazione dell'analisi della catena di Markov standard diventa computazionalmente intrattabile per reti non banali. Si possono fare alcune considerazioni.

Si considera anzitutto un modello virale SIS. In questo caso, è facile verificare che la catena di Markov ha un unico stato assorbente corrispondente alla configurazione in cui tutti i nodi sono suscettibili. Infatti, una volta che l'epidemia virale è svanita e la rete è entrata in quello stato, non si può verificare alcun cambiamento di configurazione. Dato che tutti gli altri stati della catena di Markov sono transitori e la catena di Markov stessa è irriducibile (cioè, tutti gli stati sono raggiungibili da

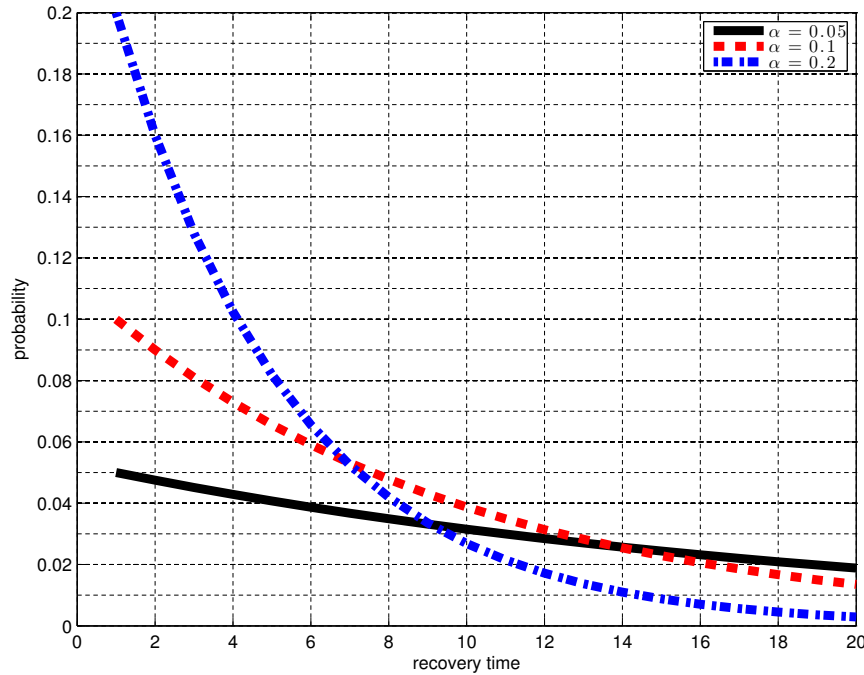


Figura 4.3: Probabilità del tempo di recupero per alcuni valori di α nel caso di tempo di recupero distribuito geometricamente.

ciascun altro), allora si può concludere che prima o poi la catena di Markov entrerà nell'unico stato assorbente con probabilità 1 (cioè, lo stato assorbente è lo stato stazionario). Quando detto può essere riassunto come segue.

Fatto: Per il modello SIS, c'è un solo stato assorbente (corrispondente a quando tutti i nodi sono nello stato S). Come conseguenza, l'epidemia si estingue asintoticamente con un tasso esponenziale a prescindere dalle condizioni iniziali e dalla topologia della rete.

Osservazione (Il paradosso dello stato metastabile). Chiaramente, un tale comportamento asintotico rappresenta uno scostamento sostanziale da quello del modello di Kephart-White (4.1) secondo il quale l'epidemia può essere persistente in certe condizioni. È comunque importante sottolineare che, in pratica, la convergenza della catena di Markov al suo stato assorbente può essere estremamente lenta al punto che può essere impossibile osservarla per mezzo di simulazioni. Infatti, come sarà anche confermato dai risultati delle simulazioni nel prossimo capitolo, quando le condizioni per la soglia epidemica sono soddisfatte, il modello SIS mostra uno stato apparentemente di regime caratterizzato dalla persistenza del virus nella rete, che è coerente con le predizioni del modello di Kephart-White. Uno stato di questo genere, apparentemente di regime, è spesso chiamato “stato metastabile” dato che, malgrado le apparenze, in un intervallo di tempo sufficientemente lungo è sicuro che scomparirà con probabilità 1 [54].

La situazione è piuttosto diversa quando si considera un modello virale SIR. Infatti, in questo caso tutti gli stati in cui nessuno dei nodi è nello stato I sono stati assorbenti. Quindi esistono complessivamente 2^N stati assorbenti, corrispondenti a tutte le possibili configurazioni di nodi suscettibili e recuperati. Dato che tutti gli altri stati sono transitori e che la catena di Markov è irriducibile, si ha la seguente proprietà.

Fatto: Per il modello SIR, ci sono 2^N stati assorbenti (corrispondenti a quando nessun nodo è nello stato I). Come conseguenza, l'epidemia si estingue asintoticamente a tasso esponenziale a prescindere dalle condizioni iniziali e dalla topologia della rete.

Si noti che, in questo caso, il comportamento asintotico del modello a catena di Markov esatta è in accordo con le previsioni del modello macroscopico (4.3) e con i risultati simulativi.

Si riportano di seguito alcune considerazioni sulla soglia epidemica per il modello SIR. A questo scopo, si suppone che al tempo $t = 0$ un solo nodo della rete sia infetto mentre tutti gli altri sono suscettibili. Allora, è facile calcolare il numero atteso, \bar{I} , di nodi infetti agli istanti di tempo successivi. Infatti, osservando che $P(k)$ rappresenta la probabilità che il nodo infetto abbia grado k , si può scrivere

$$\bar{I} = \sum_k P(k) \bar{I}_k + 1 - \delta$$

dove \bar{I}_k è il numero atteso di nodi vicini infetti di grado k che, a sua volta, può essere scritto come

$$\bar{I}_k = \sum_{j=1}^k j \binom{k}{j} \beta^j (1 - \beta)^{k-j} = k \beta.$$

Quindi, combinando le ultime due equazioni, si ha

$$\bar{I} = \sum_k P(k) k \beta + 1 - \delta = \bar{k} \beta + 1 - \delta.$$

Questo significa che, se ci si attiene alla definizione introdotta per il modello macroscopico (4.3) imponendo $\bar{I} \geq 1$, la soglia epidemica risulta essere la stessa prevista dal modello macroscopico.

4.1.3 Modello a catene di Markov intrecciate

Sfortunatamente, è difficile trarre ulteriori conclusioni analizzando i modelli a catena di Markov esatta a causa della loro complessità esponenziale. Per questo motivo sono stati proposti modelli semplificati che, pur mantenendo una prospettiva microscopica, sono caratterizzati da una complessità moderata. In particolare ci si concentrerà sul modello a N catene di Markov intrecciate ottenuto facendo un'ipotesi d'indipendenza [24]. Per brevità, il modello sarà presentato solo per il caso SIS anche se è possibile una sua generalizzazione a diversi modelli epidemici

[55]. Sia $\pi_i^S(t)$ la probabilità che il nodo i sia nello stato S al tempo t . Allora, si ha chiaramente

$$\pi_i^S(t+1) = [1 - \beta_i(t)] \pi_i^S(t) + \delta [1 - \pi_i^S(t)] \quad (4.6)$$

dove $\beta_i(t)$ è la probabilità che il nodo i venga infettato dai suoi vicini al tempo t . Dato che il calcolo esatto di $\beta_i(t)$ comporterebbe la soluzione del modello a catena di Markov singola esatta, l'idea di [24] è di approssimarla ponendo

$$\beta_i(t) = 1 - \prod_{j \in \mathcal{N}_i} \{1 - \beta [1 - \pi_j^S(t)]\} \quad (4.7)$$

dove, in accordo con l'ipotesi di indipendenza, si ipotizza che le probabilità $\pi_j^S(t)$ siano indipendenti tra loro.

Dato che le equazioni (4.6)-(4.7) devono essere ripetute per ogni nodo della rete, l'approssimazione considerata dà luogo ad un sistema a tempo-discreto non lineare N -dimensionale la cui dinamica può essere analizzata per mezzo degli strumenti classici della teoria dei sistemi. Ad esempio, si possono facilmente verificare i seguenti risultati.

Fatto: Lo stato $\pi_1^S = \dots = \pi_N^S = 1$, corrispondente all'estinzione dell'epidemia, è un punto di equilibrio per il sistema non lineare (4.6)-(4.7), per $i = 1, \dots, N$.

Fatto: [24] Sia $\rho(\mathbf{A})$ il **raggio spettrale** (il più grande autovalore) della matrice di adiacenza \mathbf{A} della rete. Allora, se $\beta/\delta < 1/\rho(\mathbf{A})$, l'epidemia si estingue asintoticamente a prescindere dalle condizioni iniziali e dalla topologia della rete. Altrimenti, l'epidemia non scompare mai dato che il punto di equilibrio $\pi_1^S = \dots = \pi_N^S = 1$ non è asintoticamente stabile.

Tali risultati permettono di concludere che la condizione di soglia epidemica per il modello approssimato (4.6)-(4.7) è

$$\frac{\beta}{\delta} \geq \frac{1}{\rho(\mathbf{A})}. \quad (4.8)$$

È opportuno fare alcune osservazioni. La prima è che la soglia che può anche essere ottenuta con altre approssimazioni (ad esempio, la teoria del campo medio e i modelli linearizzati [56]), è generalmente diversa da quella ricavata mediante il modello macroscopico di Kephart-White. Infatti, per una rete generica, si ha

$$\bar{k} \leq \rho(\mathbf{A}) \leq k_{\max}$$

dove k_{\max} è il massimo grado di un nodo presente nella rete. Così, le due soglie coincidono soltanto quando $\rho(\mathbf{A}) = \bar{k}$ (questo si verifica ad esempio quando la rete è regolare, cioè tutti i nodi hanno lo stesso grado), mentre negli altri casi la soglia (4.8) tende ad essere meno stringente, il che significa che prevederà la diffusione di epidemie virali più spesso.

Ad esempio, per una rete power-law (scale-free), con distribuzione del grado $P(k) \propto k^{-\eta}$, il raggio spettrale, per una rete di dimensione sufficientemente grande, è

$$\rho(\mathbf{A}) \cong \begin{cases} \sqrt{k_{\max}} & \text{if } \eta > 5/2 \\ \sigma_k^2/\bar{k} & \text{if } 2 < \eta < 5/2 \end{cases}$$

dove σ_k^2 è la varianza del grado della rete.

Si noti che l'analisi precedente può essere estesa, in linea di principio, anche a modelli diversi. Ad esempio, in riferimento al modello SIR, la condizione di soglia epidemica per il modello a N catene di Markov intrecciate risulta essere [55]

$$\frac{\beta}{\alpha} > \frac{1}{\rho(\mathbf{A})}.$$

Questo risultato tuttavia è meno supportato da prove ed è stato recentemente criticato [25]. Come ultima osservazione, si precisa che, per quanto riguarda la soglia epidemica, l'introduzione nel modello SIR dell'ulteriore stato assorbente D come in fig. 4.1 non porterebbe alcuna sostanziale modifica al comportamento descritto. Infatti, dato che i due stati D e R giocano lo stesso ruolo, è sufficiente considerare $\alpha + \gamma$ al posto di α .

4.2 Comportamento dei virus nel consenso

Come già discusso, esistono vari modelli per descrivere la diffusione di un'epidemia virale in una rete di sensori. Una caratteristica importante di tutti questi modelli è l'esistenza di una soglia epidemica che dipende dalla topologia della rete sia in funzione del grado medio \bar{k} che del raggio spettrale $\rho(\mathbf{A})$. Quindi sembra naturale considerare queste due quantità come possibili misure della *suscettibilità di una rete* ad un'epidemia virale. Tuttavia, per i nostri scopi, quest'analisi non è sufficiente. Infatti, un'altra questione di primaria importanza che deve essere risolta è quanto sia distruttiva un'epidemia virale per gli algoritmi di stima distribuita. Naturalmente, la risposta a questa domanda non è univoca e dipende in modo cruciale dal come si comportano i nodi infetti.

Per completare il modello delle epidemie virali nelle reti di sensori, è anche necessario specificare come si comporta il virus nei nodi della rete. A questo scopo, il modello più semplice, usato anche nelle analisi simulative, è quello *non cooperativo*. In questo modello si assume che, una volta infetto, un nodo perda tutti i suoi dati, smetta di raccogliere nuovi dati, e smetta di prendere parte all'algoritmo di stima distribuita. Poi, una volta risanato dall'infezione, il nodo riprende il suo comportamento normale con la stima locale inizializzata ad un valore arbitrario e la covarianza locale inizializzata ad un valore estremamente elevato.

Come è evidente, per un modello virale non cooperativo, l'effetto principale dell'epidemia è la riduzione della connettività di rete. Questa, a sua volta, implica che un'epidemia virale sia meno distruttiva quando la rete è più robusta riguardo la connettività. Quindi i parametri topologici che sembrano più adatti a riassumere la robustezza della rete rispetto agli effetti di un'epidemia virale sono la connettività algebrica λ_2 e la connettività di nodo ν . Infatti, più grandi sono λ_2 e ν , più difficile è

dividere un grafo in componenti disconnesse. Ricordando che per una rete generica si ha

$$\lambda_2 \leq \nu \leq k_{\min} \leq \bar{k} \leq \rho(\mathbf{A}) \leq k_{\max}$$

dove k_{\min} è il grado minimo di un qualunque nodo della rete, è immediato concludere che, almeno in prima approssimazione, le reti il cui funzionamento è meno soggetto a essere affetto da un'epidemia virale sono quelle caratterizzate da raggi spettrali $\rho(\mathbf{A})$ piccoli e da piccole distanze tra la connettività di nodo ν (o, alternativamente, il grado medio \bar{k}) ed il raggio spettrale $\rho(\mathbf{A})$. D'altra parte, poichè la convergenza degli algoritmi di consenso, su cui si basano tutti gli algoritmi di stima distribuita proposti, cresce al crescere della connettività algebrica λ_2 , possibili linee di progetto per una rete di sensori sono di rendere $\rho(\mathbf{A})$ il più piccolo possibile mantenendo al contempo valori sufficientemente grandi per λ_2 e ν . Ad esempio, le reti cosiddette *scale-free* non sono molto robuste alle epidemie virali essendo caratterizzate da un ampio intervallo $\rho(\mathbf{A}) - \nu$. D'altra parte, per reti regolari è teoricamente possibile rendere questa distanza arbitrariamente piccola dato che $k_{\min} = \bar{k} = \rho(\mathbf{A}) = k_{\max}$.

Riassumendo, un virus non cooperativo può essere tollerato se la rete ha un'elevata connettività algebrica o di nodo. La situazione è più problematica quando si considerano altri effetti dei virus. Per vedere questo, si considera ad esempio un *modello di virus a comportamento deviante*, in cui i nodi infetti prendono parte all'algoritmo di consenso con dati generati a caso, o un *modello di virus malizioso*, in cui i nodi prendono parte al consenso con dati generati appositamente per compromettere del tutto il funzionamento dell'algoritmo. Allora, in quei casi, le epidemie virali possono essere estremamente distruttive. Infatti, come è ben noto, se non vengono prese contromisure perfino un nodo a comportamento deviante può compromettere il funzionameneto dell'algoritmo di consenso [57]. Inoltre, perfino prendendo contromisure, pochi nodi devianti o maliziosi possono compromettere il funzionamento dell'algoritmo di consenso. Infatti, come mostrato in [58], una rete di consenso necessita di avere connettività di nodo almeno uguale a $2k + 1$ (e, rispettivamente $k + 1$) per k nodi maliziosi (e rispettivamente, devianti) per essere rilevabile ed identificabile da ogni agente ben funzionante. Infine, anche quando è possibile il rilevamento di nodi infetti, il progetto di algoritmi distribuiti computazionalmente efficienti per eseguire un tale compito è ancora un problema aperto. In conclusione, l'epidemia di un virus deviante o malizioso può essere molto distruttiva. Se possibile, la rete dovrebbe essere progettata in modo da non superare la soglia epidemica.

4.3 Virus e stima distribuita: analisi simulativa

Si vogliono analizzare adesso gli effetti dei paramentri di rete sulla diffusione dei virus e la fusione dell'informazione distribuita per mezzo di un'estesa campagna di simulazioni in uno scenario di sorveglianza realistico. A questo scopo, si farà riferimento al problema di tracciamento di un oggetto che si muove in una regione di sorveglianza di $5\text{ km} \times 5\text{ km}$. Si assume che il moto dell'oggetto possa essere

descritto dal modello cinematico a velocità quasi costante:

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_k + \mathbf{w}_k$$

dove $\Delta = 1 \text{ s}$ è l'intervallo di campionamento e il vettore di stato ignoto è dato dalle componenti di posizione e velocità lungo gli assi coordinati, cioè, $\mathbf{x} = [p_x, v_x, p_y, v_y]'$. La matrice di covarianza del rumore di processo si assume pari a $\mathbf{Q} = \mathbf{G} q^2$ con

$$\mathbf{G} = \begin{bmatrix} \Delta^3/3 & \Delta^2/2 & 0 & 0 \\ \Delta^2/2 & \Delta & 0 & 0 \\ 0 & 0 & \Delta^3/3 & \Delta^2/2 \\ 0 & 0 & \Delta^2/2 & \Delta \end{bmatrix}$$

e $q = 0.5 \text{ m/s}^2$.

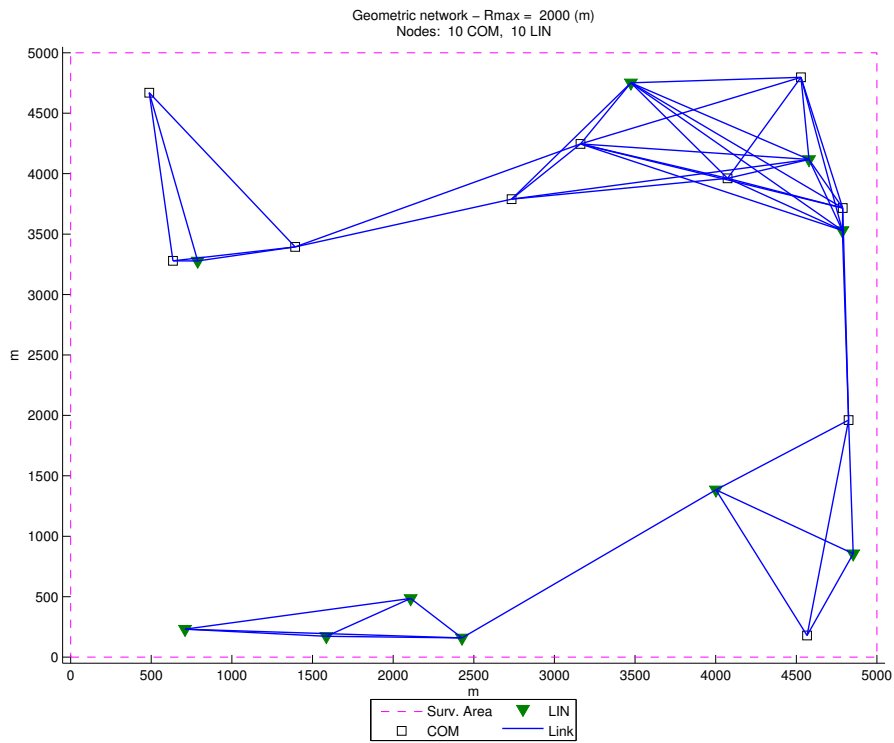
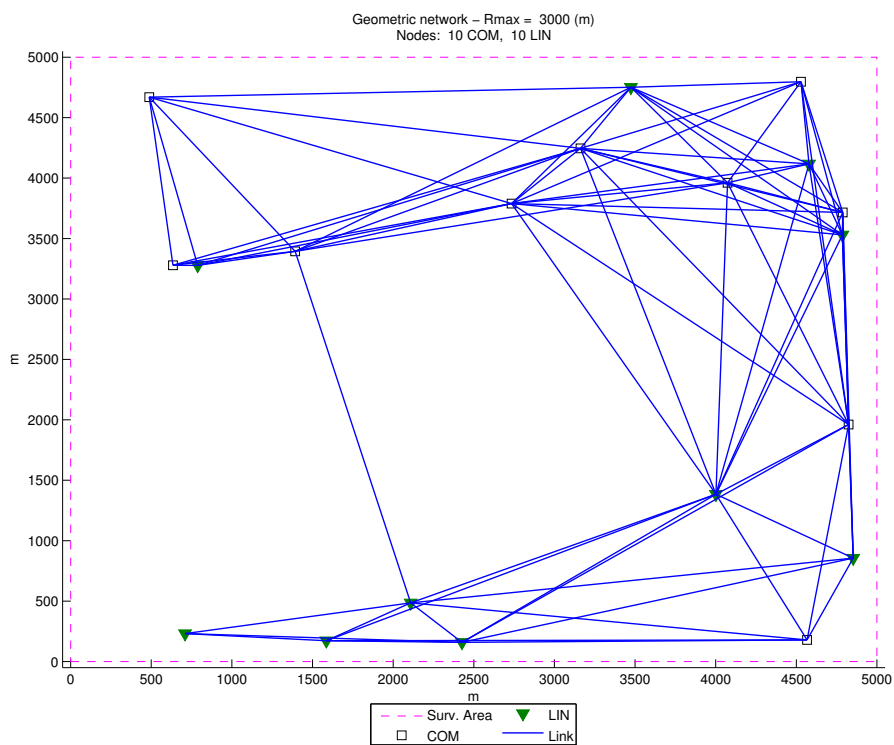
Si suppone siano disponibili un totale di 10 sensori, che forniscono misure della posizione dell'oggetto in coordinate cartesiane, cioè,

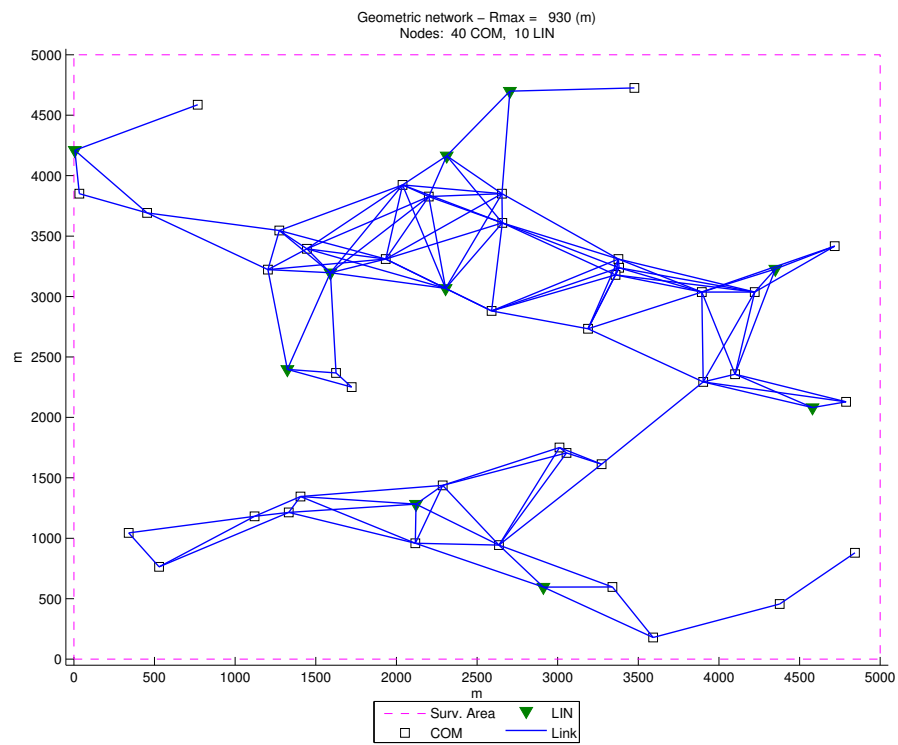
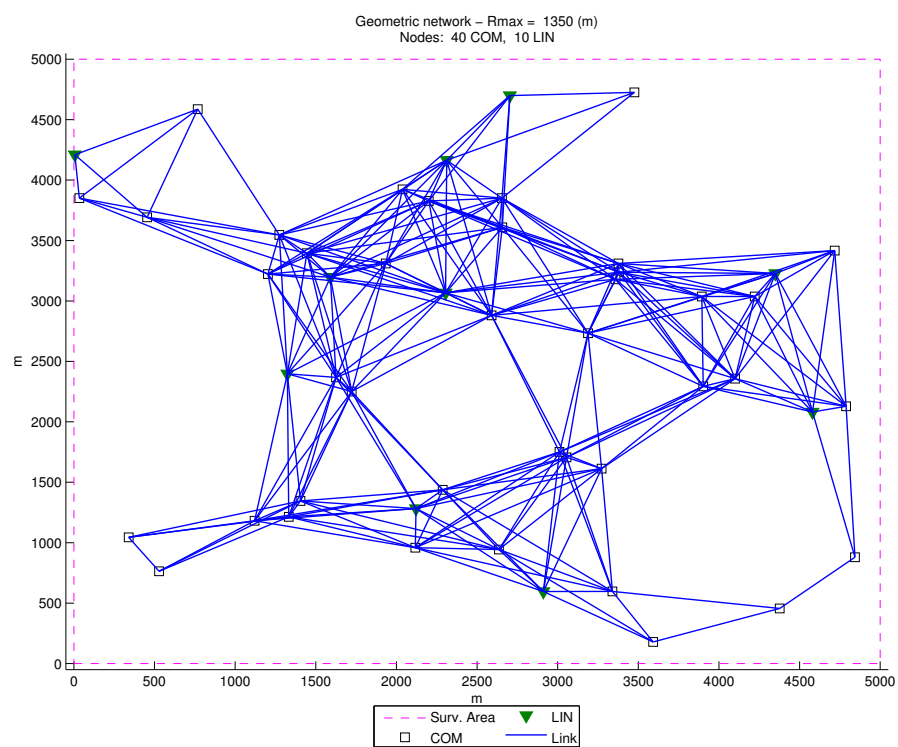
$$\mathbf{y}_k^i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k^i.$$

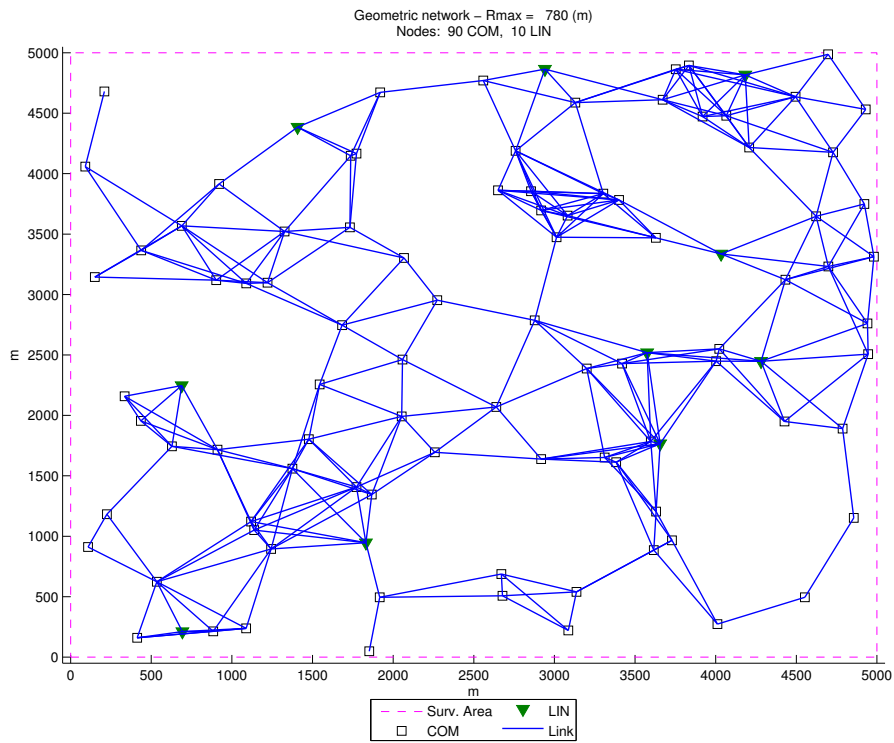
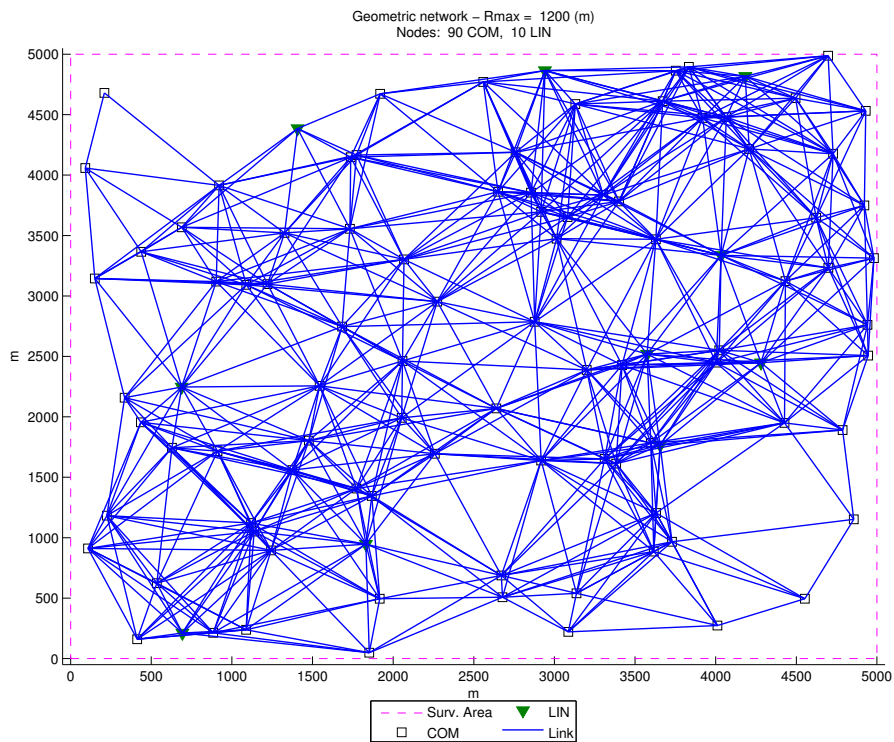
Le matrici di covarianza del rumore di misura sono poste uguali a $\mathbf{R}^i = \text{diag}(r^2, r^2)$ con $r = 20 \text{ m}$.

Oltre ai 10 sensori, la rete di sorveglianza è costituita da un certo numero di nodi di comunicazione. Specificamente, si considerano tre diverse configurazioni con rispettivamente 10, 40 e 90 nodi di comunicazione (le coordinate dei nodi vengono generate casualmente nell'area di sorveglianza con distribuzione uniforme). Per quanto riguarda il collegamenti tra i nodi della rete, si è considerato un modello di collegamento *geometrico*. Per ogni configurazione dei nodi, sono stati considerati valori diversi del raggio di connettività per analizzare l'effetto di questo parametro sulla diffusione dei virus. Nelle figg. 4.4-4.6, sono riportate le reti di sorveglianza risultanti.

Si noti che, per le tre configurazioni dei nodi considerate (20, 50 e 100 nodi), i raggi di connettività minimi che assicurano che la rete sia connessa sono, rispettivamente, 2000 m , 930 m , and 780 m . Dalle figg. 4.4-4.6, è evidente come l'incremento del 50% del raggio di connettività implica un numero notevolmente superiore di collegamenti che, da una parte, rendono la rete più robusta nei confronti della disconnessione dei nodi infetti ma, dall'altra, favoriscono anche la diffusione dei virus nella rete. Nella fig. C.1 in Appendice C, si vede come variano alcuni dei più importanti parametri di rete (il raggio spettrale della matrice di adiacenza $\rho(\mathbf{A})$, il grado medio \bar{k} , la connettività di nodo ν , e la connettività algebrica λ_2) al crescere del raggio di connettività R_{max} . Si noti che tutti i parametri, come ci si aspetterebbe, sono funzioni monotone non decrescenti del raggio di connettività. In particolare, le due curve del raggio spettrale della matrice di adiacenza e del grado medio sono quasi parallele, cioè questi due parametri vanno soggetti a variazioni simili. Inoltre, la connettività di nodo, che è un parametro intero, subisce incrementi graduali a partire dal valore minimo di 1 in corrispondenza del raggio minimo di

(a) $R_{max} = 2000$ m.(b) $R_{max} = 3000$ m.**Figura 4.4:** Reti di sorveglianza con 20 nodi e due diversi raggi di connettività, R_{max} .

(a) $R_{max} = 930$ m.(b) $R_{max} = 1350$ m.**Figura 4.5:** Reti di sorveglianza con 50 nodi e due diversi raggi di connettività, R_{max} .

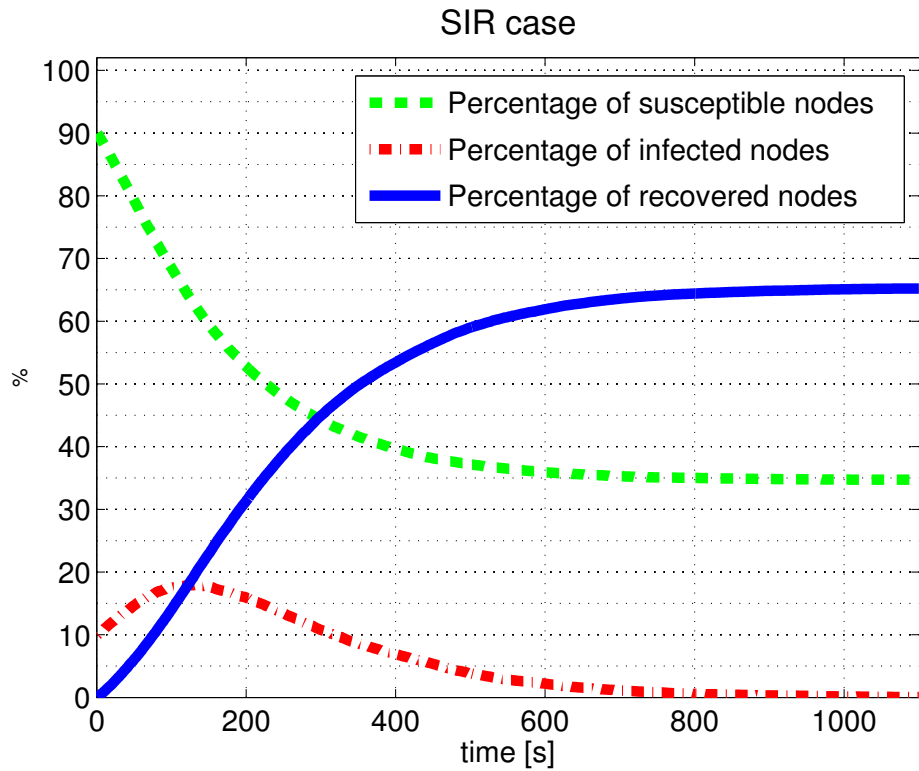
(a) $R_{max} = 780$ m.(b) $R_{max} = 1200$ m.**Figura 4.6:** Reti di sorveglianza con 50 nodi e due diversi raggi di connettività, R_{max} .

connettività della rete. Infine, la connettività algebrica, che caratterizza principalmente la velocità di convergenza degli algoritmi di consenso (cioè la diffusione dell'informazione nelle rete) aumenta solo leggermente con il raggio di connettività per questo tipo di reti casuali (irregolari).

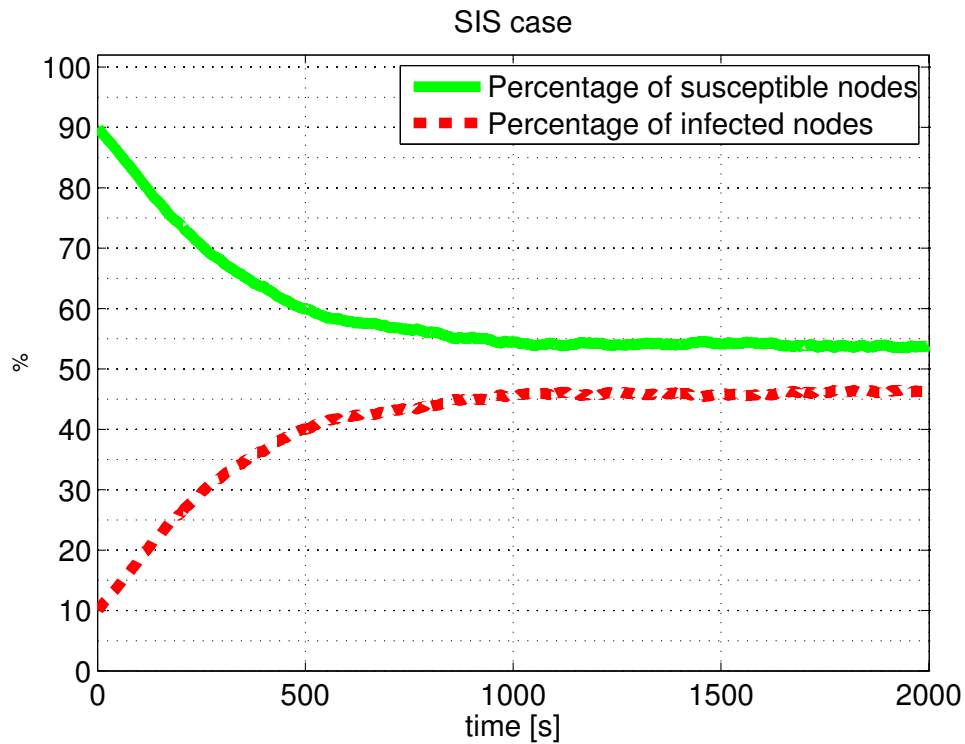
La diffusione di un virus non cooperativo nelle reti di monitoraggio considerate è stata simulata considerando entrambi i modelli SIR e SIS a catena di Markov singola esatta visti nel Paragrafo 4.1.2 con probabilità d'infezione di un nodo che dipende dal numero di nodi vicini infetti secondo (4.5). Quando non diversamente specificato, la probabilità di recupero α (caso SIR) e la probabilità di guarigione δ (caso SIS) sono state poste pari a 0.01. La prima parte dell'analisi simulativa è stata dedicata alla determinazione empirica delle soglie epidemiche. I risultati sono riportati nella fig. C.2 in Appendice C. Si possono trarre alcune conclusioni interessanti: per i due modelli virali le soglie epidemiche risultano piuttosto simili; in generale, la soglia epidemica teorica $1/\rho(\mathbf{A})$ è il limite inferiore di quella reale; come ci si aspetterebbe, la soglia epidemica tende a decrescere al crescere del raggio di connettività, indicando così che più una rete è connessa più è soggetta a epidemie virali. La prevalenza dell'epidemia virale nelle diverse configurazioni di rete può essere analizzata osservando la fig. C.3 in Appendice C dove, per ciascuno dei casi considerati, viene riportata la massima percentuale di nodi infetti in funzione del raggio di connettività (i grafici sono mediati su 500 prove Monte Carlo indipendenti). Si può vedere che, per tutte le reti, un incremento del raggio di connettività porta ad un incremento della prevalenza epidemica, specialmente per i valori più alti della probabilità d'infezione β . Tale incremento risulta, tuttavia, essere più moderato per la rete di 20 nodi, che di conseguenza risulta essere più robusta rispetto alla diffusione dei virus.

Per quanto riguarda le dinamiche epidemiche, come anticipato i due modelli di diffusione dei virus, SIR e SIS, mostrano comportamenti piuttosto diversi. Infatti, nel caso SIR l'epidemia svanisce sempre dopo un certo intervallo di tempo, mentre nel caso SIS l'epidemia è apparentemente persistente ogni qual volta viene superata la soglia epidemica. Gli andamenti tipici delle dinamiche epidemiche sono riportati in fig. 4.7 (i grafici sono mediati su 500 prove Monte Carlo indipendenti). Pertanto, si può concludere che assicurando alla rete di sensori capacità di recupero nei confronti del virus si può ridurre sostanzialmente l'effetto di un'epidemia virale. Sfortunatamente, in pratica questo non è sempre possibile dato che richiederebbe la presenza di meccanismi anti-virus in tutti i nodi della rete, una richiesta che potrebbe essere in contrasto con le limitate capacità dei dispositivi wireless disponibili.

Si rivolge adesso l'attenzione agli effetti di un'epidemia virale sull'algoritmo di stima distribuita considerato. A questo scopo si possono trarre alcune conclusioni sulle prestazioni dell'algoritmo di fusione distribuita proposto, dall'osservazione delle figg. 4.8-4.10 in cui viene riportato l'andamento temporale del Position Root Mean Square Error (PRMSE) per reti diverse in presenza di un'epidemia virale. Il modello virale è il SIS con probabilità di guarigione $\delta = 0.1$ e probabilità d'infezione scelta per ciascuna rete in modo da avere una prevalenza epidemica pari al 30%. I corrispondenti valori di β sono riportati in tab. 4.1. Il PRMSE è mediato su tutti i nodi non infetti della rete e su 200 prove Monte Carlo indipendenti. Chiaramente, la presenza di un'epidemia virale porta in generale ad un certo degrado delle prestazioni di stima. Tale degrado è, tuttavia, di gran lunga meno evidente per le



(a) Caso SIR.



(b) Caso SIS.

Figura 4.7: Dinamiche temporali delle epidemie virali per la rete con 100 nodi. Il raggio di connettività R_{max} e la probabilità d'infezione β sono poste, rispettivamente, pari a 1000 ed a 0.002.

N. di nodi	raggio di connettività R_{max}	probabilità d'infezione β
20	2000	0.052
20	3000	0.026
50	930	0.038
50	1350	0.016
100	780	0.031
100	1200	0.011

Tabella 4.1: Valori di β corrispondenti a un picco epidemico del 30% di nodi infetti nelle reti considerate (δ è fissato a 0.1).

reti caratterizzate da raggi di connettività più elevati, mentre il numero dei nodi della rete ha solo un effetto secondario. Questo stato di cose può essere compreso osservando che un incremento del raggio di connettività ha l'effetto di aumentare la connettività algebrica della rete, rendendo così la rete più robusta nei confronti del malfunzionamento di certi nodi e permettendo una maggiore velocità di convergenza del consenso che, a sua volta, porta a migliori prestazioni di stima. Naturalmente, si deve sempre tener presente il fatto che, come già discusso, una rete più connettiva è più vulnerabile alla diffusione dei virus, essendo caratterizzata da una soglia epidemica più bassa e da una prevalenza epidemica più elevata. In pratica, la scelta del livello di connettività di una rete di sensori deve essere fatta con attenzione tenendo conto di questi due effetti contrastanti, oltre ai principali aspetti tecnologici, in modo da ottenere il compromesso desiderato di caratteristiche prestazionali.

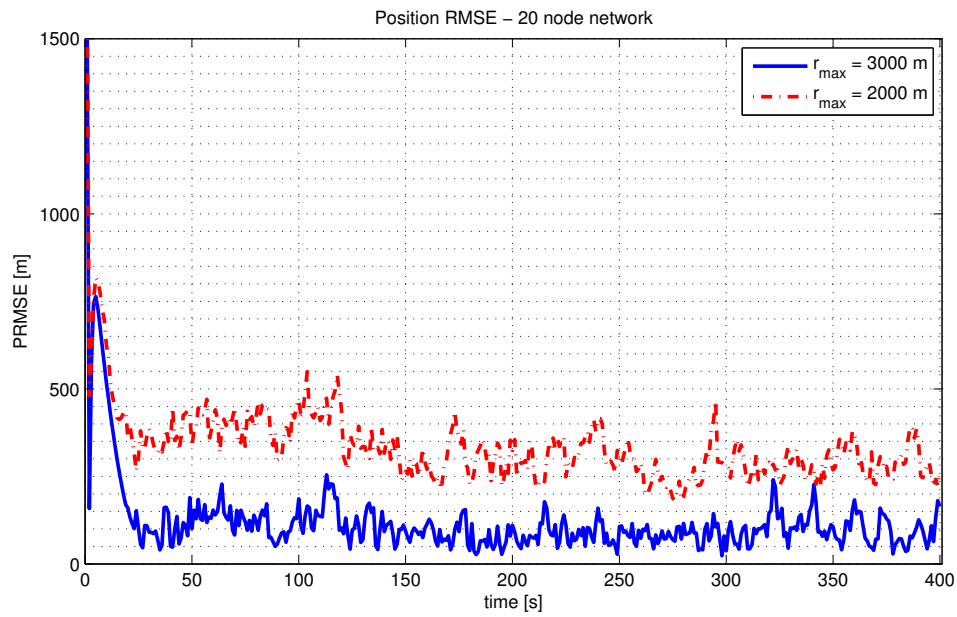


Figura 4.8: Prestazioni, in termini di Position Root Mean Square Error (PRMSE), dell’algoritmo di stima distribuita considerato per la rete di 20 nodi con raggi di connettività di 2000 m e di 3000 m.

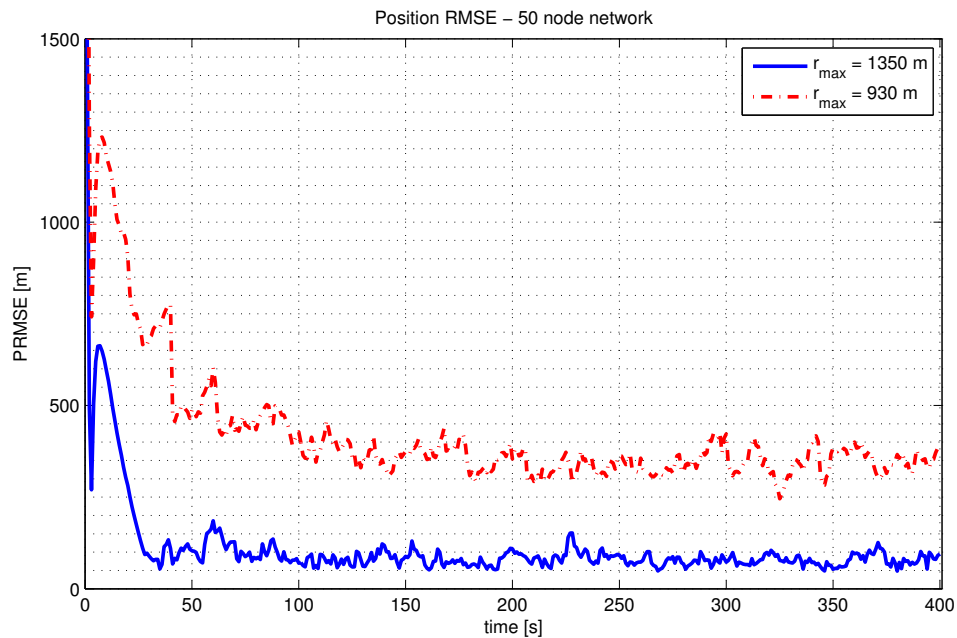


Figura 4.9: Prestazioni, in termini di Position Root Mean Square Error (PRMSE), dell’algoritmo di stima distribuita considerato per la rete di 50 nodi con raggi di connettività di 930 m e di 1350 m.

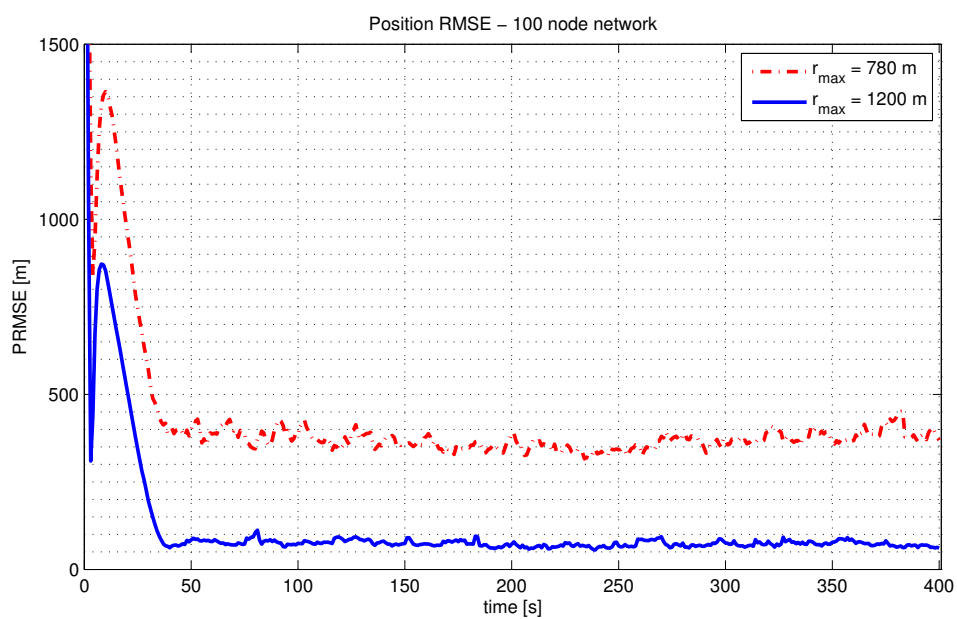


Figura 4.10: Prestazioni, in termini di Position Root Mean Square Error (PRMSE), dell'algoritmo di stima distribuita considerato per la rete di 100 nodi con raggi di connettività di 780 m e di 1200 m.

Parte III

Stima distribuita tra agenti a stato parzialmente sovrapposto

Capitolo 5

Consenso tra agenti a stati sovrapposti

Un problema di interesse in diversi ambiti applicativi è quello della stima dello stato di un sistema dinamico quando i nodi, o agenti, della rete non stimano tutto lo stato del sistema ma solo una sua parte. In questa situazione si avranno, in generale, agenti con stato parzialmente sovrapposto. Questo può accadere, ad esempio, nel caso di sistemi distribuiti di sorveglianza se i nodi della rete hanno una visione parziale della regione sorvegliata e/o degli oggetti di interesse in tale regione (target), oppure nel caso si stia monitorando un sistema spazialmente distribuito per il quale, a causa della grandissima estensione del dominio spaziale di interesse, ogni nodo della rete si può occupare esclusivamente di una sottoregione di tale dominio collocata nell'intorno del nodo stesso.

Il problema in oggetto è noto in letteratura come stima dello stato di ordine ridotto [59].

5.1 Fusione di ordine ridotto

Per semplicità di esposizione, ma anche senza perdita di generalità, si considerano solo due agenti, Ag^1 ed Ag^2 , ciascuno dei quali ha uno stato rappresentativo di una parte del sistema complessivo oggetto del monitoraggio. Si suppone anche che i due agenti abbiano un sottoinsieme proprio dello stato in comune. Si può quindi scomporre lo stato complessivo del sistema \mathbf{x} in tre parti:

\mathbf{x}_1 , parte dello stato aggiornata dall'agente 1 ma non dall'agente 2;

\mathbf{x}_2 , parte dello stato comune ai due agenti;

\mathbf{x}_3 , parte dello stato aggiornata dall'agente 2 ma non dall'agente 1.

ovvero

$$\mathbf{x} = [\mathbf{x}_1^\top, \mathbf{x}_2^\top, \mathbf{x}_3^\top]^\top \text{ stato del sistema}$$

$$\mathbf{x}^1 = [\mathbf{x}_1^\top, \mathbf{x}_2^\top]^\top \text{ stato di } Ag^1$$

$$\mathbf{x}^2 = [\mathbf{x}_2^\top, \mathbf{x}_3^\top]^\top \text{ stato di } Ag^2.$$

Quindi, ciascuno dei due agenti aggiorna solo una parte dello stato del sistema, ricevendo informazioni dai propri sensori e dall'altro agente. L'obiettivo è quello di effettuare la fusione tra le informazioni di cui dispongono i due agenti in modo che abbiano una visione condivisa della parte dello stato a comune.

Agenti che condividono un sottoinsieme di variabili di stato possono fondere le loro informazioni in svariati modi. Poiché l'informazione aggiornata da ciascun agente riguarda vettori di stato diversi, seppur parzialmente sovrapposti, per poter effettuare la fusione è necessario ricondursi, attraverso un'opportuna rielaborazione dell'informazione disponibile, ad uno stato comune. A tale proposito, nei paragrafi che seguono verranno presentati tre diversi metodi di fusione che operano rispettivamente sullo *spazio di stato unione* degli agenti (FoU = Fusion over the Union), sullo *spazio di stato locale* (FoLSS = Fusion over the Local State Space) o sullo *spazio di stato intersezione* degli agenti (FoI = Fusion over the Intersection). Nel seguito, si indicherà con $p^i(\cdot \cdot \cdot)$ una generica PDF relativa all'agente i . Siano $p^1(\mathbf{x}_1, \mathbf{x}_2)$ e $p^2(\mathbf{x}_2, \mathbf{x}_3)$ le PDF (a priori), $\bar{p}^1(\mathbf{x}_1, \mathbf{x}_2)$ e $\bar{p}^2(\mathbf{x}_2, \mathbf{x}_3)$ le PDF fuse (a posteriori) degli agenti 1 e 2. Infine, siano

$$p^1(\mathbf{x}_1) = \int p^1(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2$$

$$p^1(\mathbf{x}_2) = \int p^1(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1$$

$$p^2(\mathbf{x}_2) = \int p^2(\mathbf{x}_2, \mathbf{x}_3) d\mathbf{x}_3$$

$$p^2(\mathbf{x}_3) = \int p^2(\mathbf{x}_2, \mathbf{x}_3) d\mathbf{x}_2$$

le PDF marginali degli agenti 1 e 2. I metodi considerati operano sulle PDF ed utilizzano diverse operazioni, e.g., marginalizzazione, scambio di PDF fra agenti, fusione di PDF su uno stato comune, combinazione di PDF relative a stati disgiunti.

Fusione sull'unione

Il metodo FoU (fusione sull'unione) opera sullo spazio di stato unione degli agenti coinvolti nella fusione. Relativamente al caso di due agenti, tale fusione è schematizzata in due varianti diverse nelle figg. 5.1 e 5.2. Per comodità di esposizione, nel seguito si farà riferimento allo stato $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ come stato unione dei due agenti. Viceversa, $(\mathbf{x}_1, \mathbf{x}_2)$ e \mathbf{x}_3 rappresentano lo stato locale e lo stato esogeno per l'agente 1; analogamente $(\mathbf{x}_2, \mathbf{x}_3)$ e \mathbf{x}_1 sono lo stato locale e lo stato esogeno per l'agente 2. Dopo essersi scambiate le rispettive PDF $p^1(\mathbf{x}_1, \mathbf{x}_2)$ e $p^2(\mathbf{x}_2, \mathbf{x}_3)$, ciascun agente estende la propria PDF allo stato unione combinando quest'ultima con la PDF ottenuta dalla PDF dell'altro agente tramite marginalizzazione rispetto allo stato esogeno (schema FoU di fig. 5.1), o alternativamente con una PDF non informativa (e.g., Gaussiana con matrice di informazione nulla) come illustrato nella variante FoU-NI di fig. 5.2.

Successivamente, ciascun agente effettua la fusione delle due PDF ottenute e ne marginalizza il risultato al proprio spazio di stato locale.

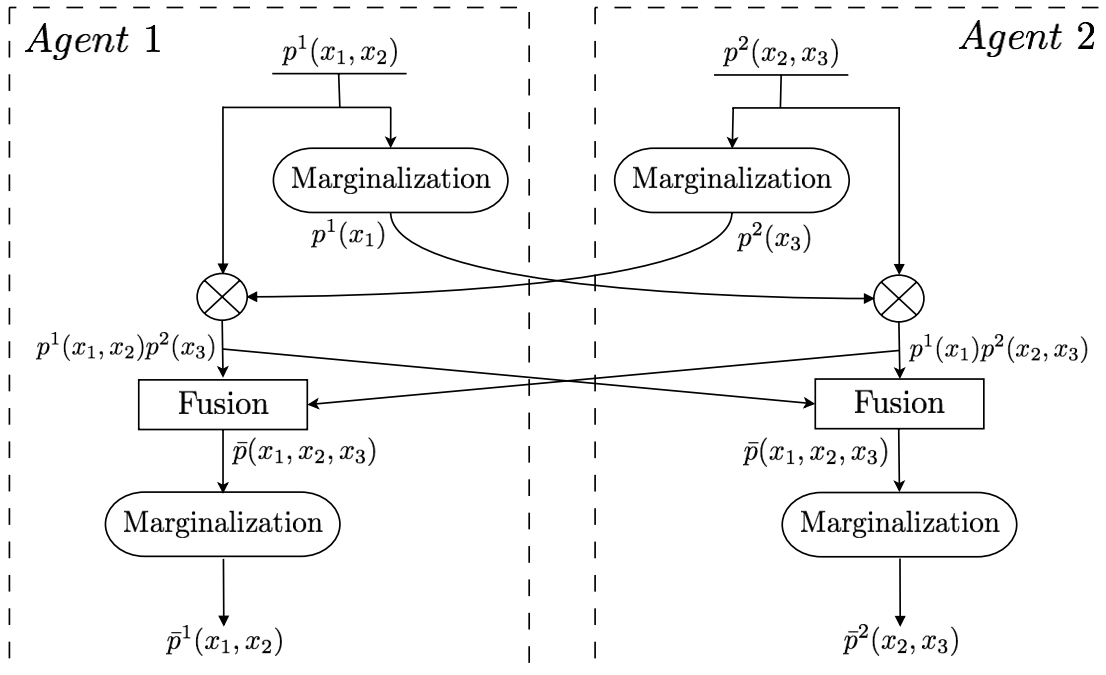


Figura 5.1: Fusione sull'unione.

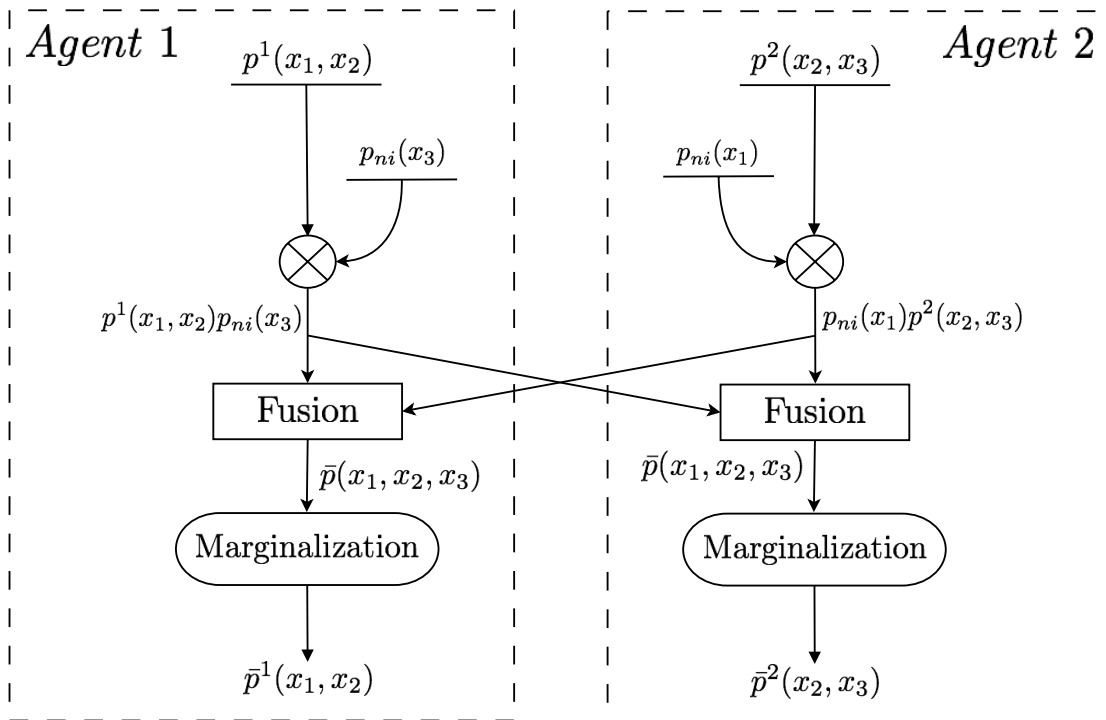


Figura 5.2: Fusione sull'unione con PDF non informativa.

Fusione sullo spazio di stato locale

Il metodo FoLSS opera la fusione sullo spazio di stato locale di ciascun agente. Due varianti di questo metodo sono schematizzate nelle figg. 5.3 e 5.4. Nello schema di fig. 5.3 l'agente 1 genera la PDF $\bar{p}(\mathbf{x}_1, \mathbf{x}_2)$ mediante fusione fra la propria PDF $p^1(\mathbf{x}_1, \mathbf{x}_2)$ e la PDF $p^1(\mathbf{x}_1)p^2(\mathbf{x}_2)$ ottenuta combinando le PDF marginali

$p^1(\mathbf{x}_1)$ dell'agente 1 e $p^2(\mathbf{x}_2)$ dell'agente 2. In modo analogo procede l'agente 2 per determinare $\bar{p}^2(\mathbf{x}_2, \mathbf{x}_3)$ fondendo $p^2(\mathbf{x}_2, \mathbf{x}_3)$ e $p^1(\mathbf{x}_2)p^2(\mathbf{x}_3)$. Nella variante FoLSS-NI di fig. 5.4, la PDF marginale $p^1(\mathbf{x}_1)$ dell'agente 1 viene sostituita da una PDF non informativa e analogamente $p^2(\mathbf{x}_3)$ per l'agente 2.

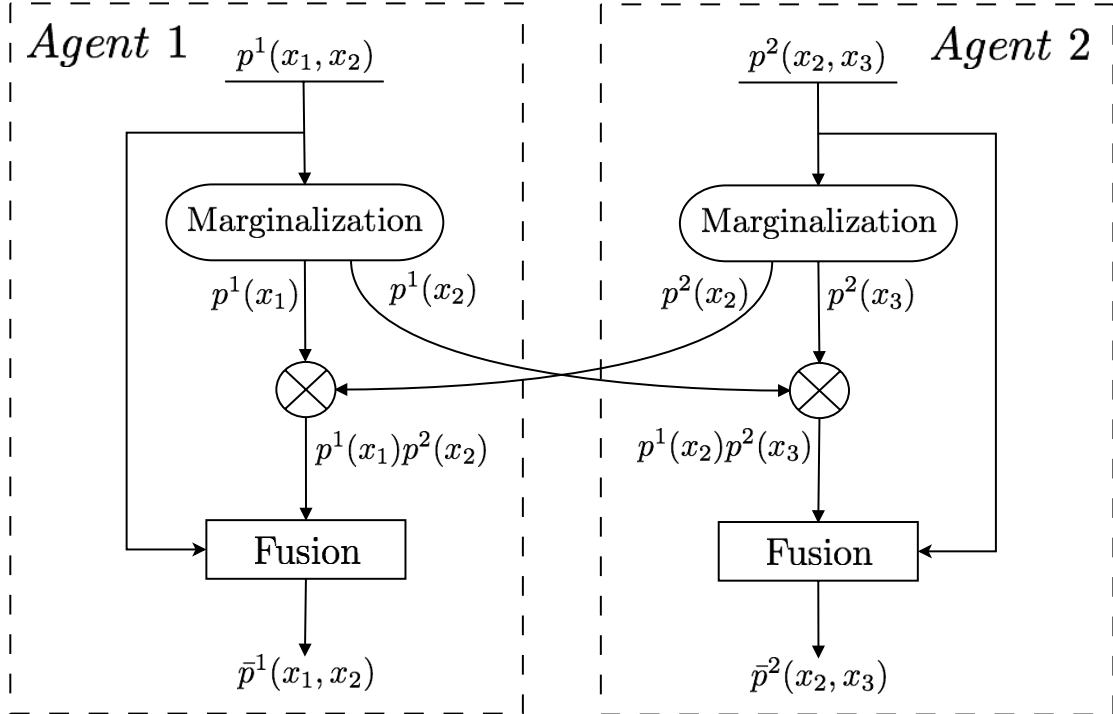


Figura 5.3: Fusione sullo spazio dello stato locale.

Fusione sull'intersezione

Il metodo FoI opera la fusione sullo spazio di stato intersezione degli agenti. Come schematizzato in fig. 5.5, ciascun agente marginalizza la propria PDF rispetto sia allo stato intersezione che al suo complementare rispetto allo stato locale. Viene poi effettuata la fusione sullo stato intersezione e successivamente una combinazione della PDF $\bar{p}(\mathbf{x}_2)$ risultante dalla fusione con la PDF marginale $p^1(\mathbf{x}_1)$ per l'agente 1 o $p^2(\mathbf{x}_3)$ per l'agente 2, in modo da ottenere una PDF definita sullo spazio di stato locale. Si noti come questo approccio, di complessità computazione inferiore rispetto ai precedenti, introduce tuttavia un disaccoppiamento forzato fra \mathbf{x}_1 e \mathbf{x}_2 per l'agente 1, fra \mathbf{x}_2 e \mathbf{x}_3 per l'agente 2.

5.2 Risultati delle simulazioni

Al fine di valutare i metodi di fusione proposti è stata effettuata una campagna di simulazioni considerando sistemi dinamici e sensori di tipo diverso. In particolare si sono considerati alcuni sistemi lineari a tempo-discreto, ed un sistema non-lineare basato sulla discretizzazione delle equazioni di Lorentz [60].

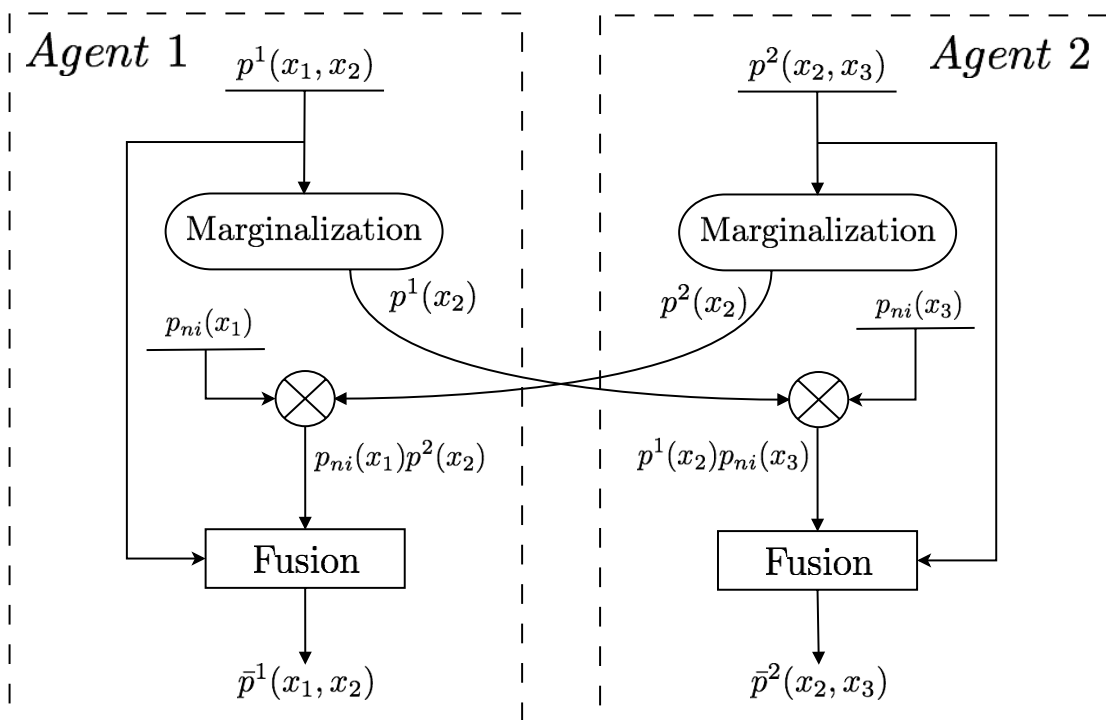


Figura 5.4: Fusione sullo spazio dello stato locale con PDF non informativa.

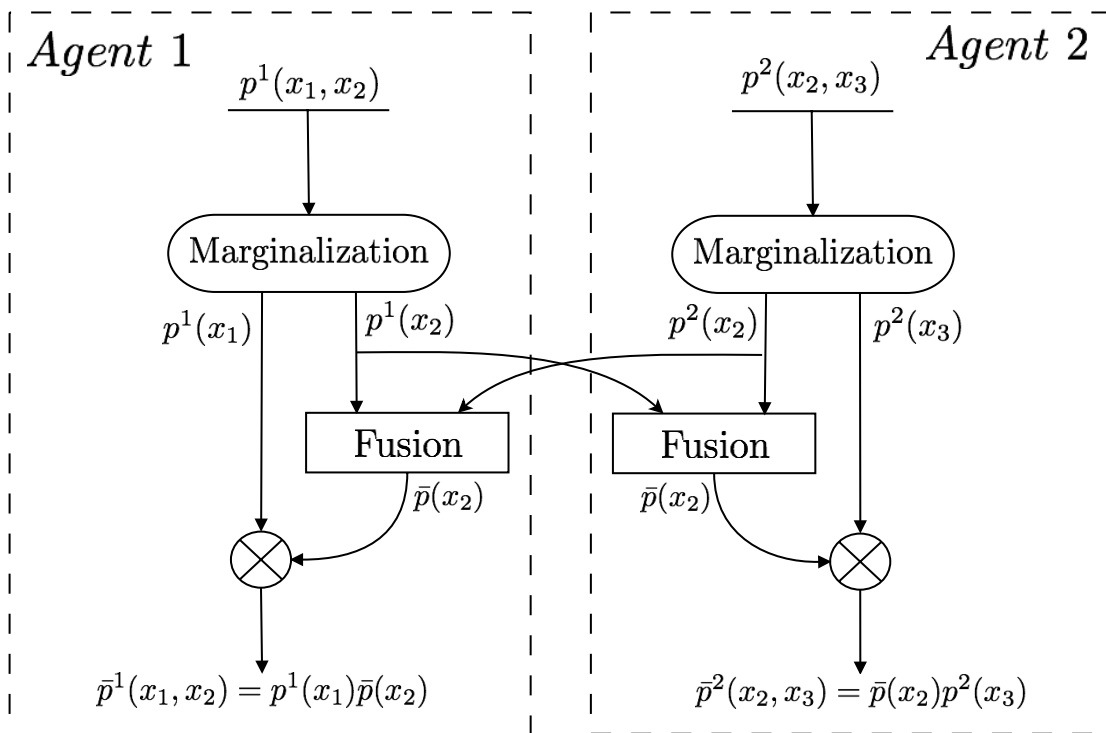


Figura 5.5: Fusione sull'intersezione.

5.2.1 Sistemi lineari

Si è considerato inizialmente il sistema dinamico lineare tempo-invariante a tempo-discreto

$$\begin{cases} \mathbf{x}_{k+1} &= \mathbf{A} \mathbf{x}_k + \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{C} \mathbf{x}_k + \mathbf{v}_k \end{cases} \quad (5.1)$$

con

$$\begin{aligned} \mathbf{x}_k &= [x_1, x_2, x_3]^\top \text{ stato del sistema al tempo } k \\ \mathbf{w}_k &\sim \text{wgn}(\mathbf{0}, \mathbf{Q}) \text{ disturbo di processo} \\ \mathbf{y}_k &\text{ misura al tempo } k \\ \mathbf{v}_k &\sim \text{wgn}(\mathbf{0}, \mathbf{R}) \text{ disturbo di misura} \\ \mathbf{x}_0 &\sim (\hat{\mathbf{x}}_0, \mathbf{P}_0) \text{ stato iniziale} \\ &\mathbf{x}_0, \mathbf{w}_k \text{ e } \mathbf{v}_k \text{ sono incorrelati.} \end{aligned}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & 1 & 0 \\ 0 & a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} q^2, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} r^2.$$

Per verificare le proprietà dei metodi di fusione proposti sono stati considerati diversi valori delle componenti della matrice di transizione \mathbf{A} (cioè a_{11} , a_{12} , a_{32} , a_{33}) corrispondenti a sistemi stabili o instabili, con diverso grado di accoppiamento. In particolare sono state considerate le seguenti scelte

- 1: sistema instabile (U: Unstable) con componenti debolmente accoppiate (WC: Weakly Coupled), U-WC:

$$a_{11} = 1.02, a_{12} = 0.5, a_{32} = 0.25, a_{33} = 1.05$$

- 2: sistema instabile (U: unstable) con componenti fortemente accoppiate (SC: Strongly Coupled), U-SC:

$$a_{11} = 1.02, a_{12} = 10, a_{32} = 10, a_{33} = 1.05$$

- 3: sistema stabile (S: stable) con componenti debolmente accoppiate (WC: Weakly Coupled), S-WC:

$$a_{11} = 0.92, a_{12} = 0.5, a_{32} = 0.25, a_{33} = 0.98$$

- 4: sistema stabile (S: stable) con componenti fortemente accoppiate (SC: Strongly Coupled), S-SC:

$$a_{11} = 0.92, a_{12} = 10, a_{32} = 10, a_{33} = 0.98.$$

I due agenti, $Ag^1, Ag^2 \in \mathcal{A}$, con \mathcal{A} insieme degli agenti, sono definiti in modo da avere una componente in comune, in particolare se \mathbf{x}^1 e \mathbf{x}^2 sono gli stati, rispettivamente, dell'agente 1 e dell'agente 2, si ha

$$Ag^1: \mathbf{x}^1 = [x_1, x_2]^\top,$$

$$Ag^2: \mathbf{x}^2 = [x_2, x_3]^\top.$$

Si noti che in questo esempio non c'è accoppiamento diretto per le componenti x_1 e x_3 . Le equazioni dei sottosistemi degli agenti sono

$$\begin{cases} \mathbf{x}_{k+1}^i &= \mathbf{A}^i \mathbf{x}_k^i + \mathbf{w}_k^i \\ \mathbf{y}_k^i &= \mathbf{C}^i \mathbf{x}_k^i + \mathbf{v}_k^i \end{cases} \quad (5.2)$$

dove, per ogni $i \in \mathcal{A}$ si ha

\mathbf{x}_k^i stato dell'agente i al tempo k
 $\mathbf{w}_k^i \sim wgn(\mathbf{0}, \mathbf{Q}^i)$ disturbo di processo per l'agente i
 \mathbf{y}_k^i misura del sensore per l'agente i al tempo k
 $\mathbf{v}_k^i \sim wgn(\mathbf{0}, \mathbf{R}^i)$ disturbo di misura per l'agente i
 $\mathbf{x}_0^i \sim (\hat{\mathbf{x}}_0^i, \mathbf{P}_0^i)$ condizioni iniziali, v. a. Gaussiana
 $\mathbf{x}_0^i, \mathbf{w}_k^i$ e \mathbf{v}_k^i sono incorrelati

Le matrici dei sottosistemi degli agenti sono:

$$Ag^1: \mathbf{A}^1 = \begin{bmatrix} a_{11} & a_{12} \\ 0 & 1 \end{bmatrix}, \quad \mathbf{Q}^1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} q^2, \quad \mathbf{C}^1 = [1 \quad 0], \quad \mathbf{R}^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} r^2.$$

$$Ag^2: \mathbf{A}^2 = \begin{bmatrix} 1 & 0 \\ a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{Q}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} q^2, \quad \mathbf{C}^2 = [0 \quad 1], \quad \mathbf{R}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} r^2.$$

Si è ipotizzata la presenza di un solo sensore lineare che comunica al filtro centralizzato le misure relative alla prima ed alla terza componente di stato, ed agli agenti Ag^1 ed Ag^2 , le misure della prima e, rispettivamente, della terza componente di stato.

Per ciascun sistema lineare considerato sono state eseguite 10000 prove Monte Carlo. In ciascuna prova viene generata una traiettoria della durata di 300 sec la cui origine viene scelta casualmente nel volume a forma di parallelepipedo centrato nell'origine degli assi coordinati e di lati $200 \times 200 \times 20$. In tutte le simulazioni effettuate le varianze dei rumori di processo e di misura sono state scelte come $q = 1$ ed $r = 10$. In tutti i metodi di fusione considerati, si è inoltre posto $\omega = 0.5$. Esempi di traiettorie dei modelli a tempo-discreto considerati sono riportati nelle figg. 5.6-5.7.

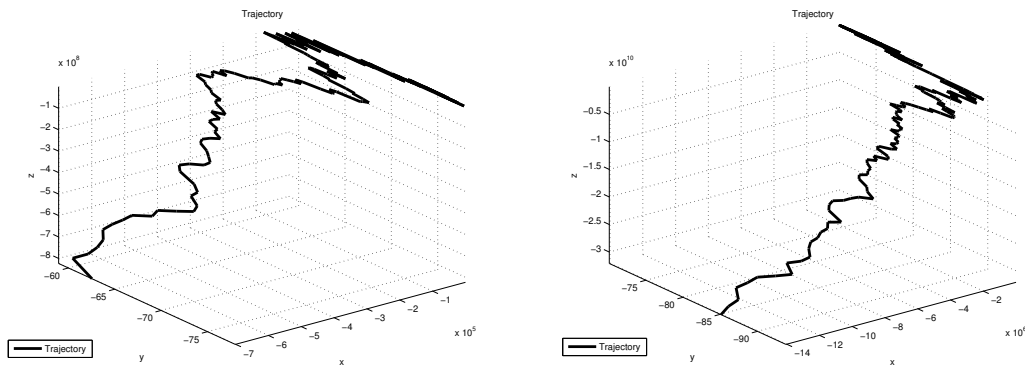


Figura 5.6: Esempio di traiettoria del sistema lineare instabile: debolmente accoppiato (modello 1, a sinistra) e ben accoppiato (modello 2 a destra).

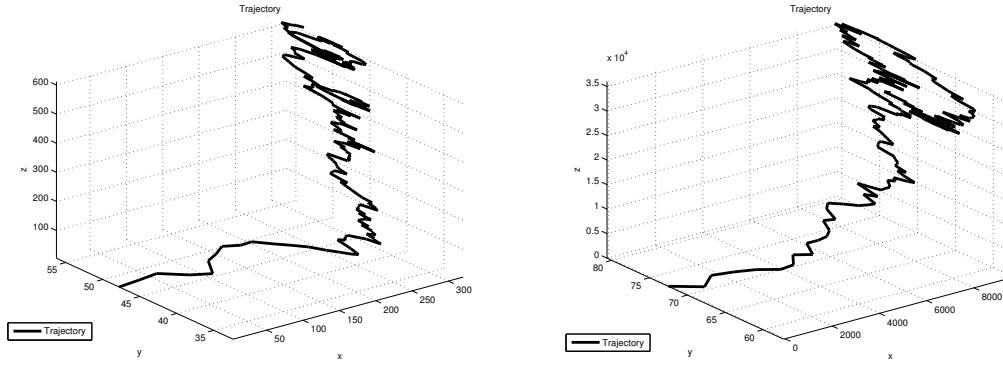


Figura 5.7: Esempio di traiettoria del sistema lineare stabile: debolmente accoppiato (modello 3, a sinistra) e ben accoppiato (modello 4 a destra).

(a) Agente Ag^1 .

Modello	CKF	FoU	FoU – NI	FoLSS	FoLSS – NI	FoI
U-WC	4.60	6.23	6.18	5.64	6.18	5.23
U-SC	5.54	16.98	8.39	15.26	8.39	8.09
S-WC	4.62	9.06	6.71	7.05	6.71	5.18
S-SC	5.50	18.30	9.00	16.31	9.00	8.49

(b) Agente Ag^2 .

Modello	CKF	FoU	FoU – NI	FoLSS	FoLSS – NI	FoI
U-WC	4.21	4.47	6.60	4.38	6.60	4.44
U-SC	5.82	11.72	7.86	11.58	7.86	8.17
S-WC	3.60	7.05	6.81	6.07	6.81	5.13
S-SC	5.72	11.45	7.32	11.32	7.32	7.87

Tabella 5.1: Sistemi lineari: RMSE medi degli agenti Ag^1 , Ag^2 per i metodi di fusione considerati e del filtro centralizzato.

I filtri Bayesiani, sia quello centralizzato che quelli degli agenti per i vari metodi di fusione considerati, vengono sempre inizializzati ponendo lo stato pari al vettore nullo, $\hat{\mathbf{x}}_0 = [0, 0, 0]^\top$ per il centralizzato e $\hat{\mathbf{x}}_0^i = [0, 0]^\top$ per gli agenti; le corrispondenti matrici di covarianza vengono poste a $\mathbf{P}_0 = \alpha \mathbf{I}_n$ con $\alpha = 10^6$ e \mathbf{I}_n opportuna matrice identità di ordine n , con $n = 3$ per il filtro centralizzato ed $n = 2$ per i filtri relativi agli agenti.

Nelle figg. D.1-D.4 in Appendice D sono riportati i valori di RMSE, mediati su tutte le prove Monte Carlo, degli agenti Ag^1 ed Ag^2 per i vari metodi di fusione presi in esame e del filtro centralizzato per i quattro modelli considerati. Infine, come indice di prestazione sintetico si sono considerati gli stessi RMSE mediati sul tempo, riportati nelle figg. 5.8 e 5.9 ed anche nella tab. 5.1 per meglio evidenziare le differenze.

Dall'esame dei risultati ottenuti si possono trarre alcune considerazioni prelimi-

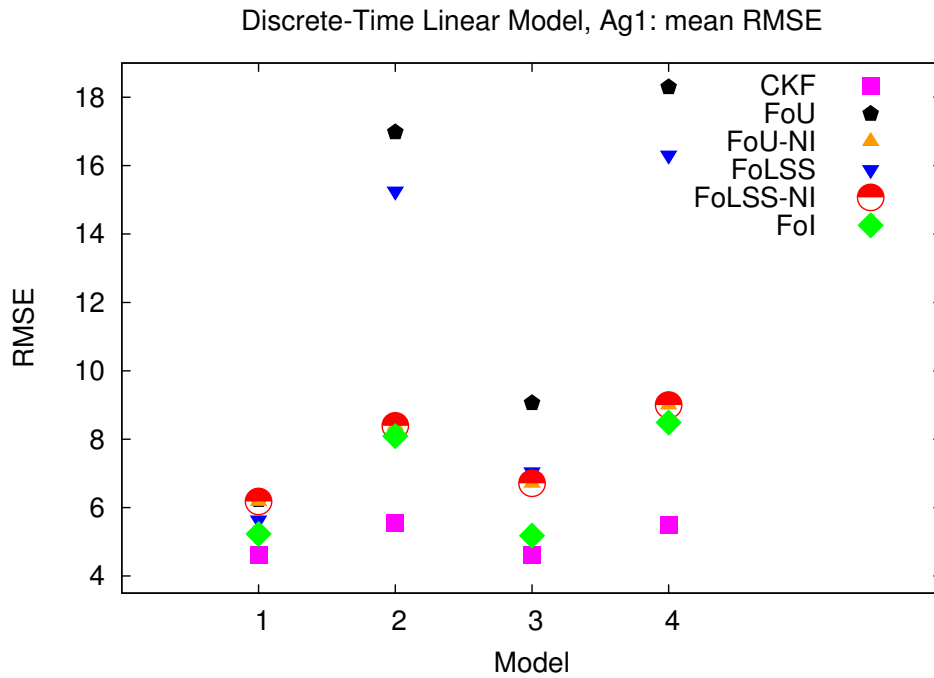


Figura 5.8: Sistema lineare: RMSE medi dell'agente 1 per i metodi di fusione considerati e del filtro centralizzato.

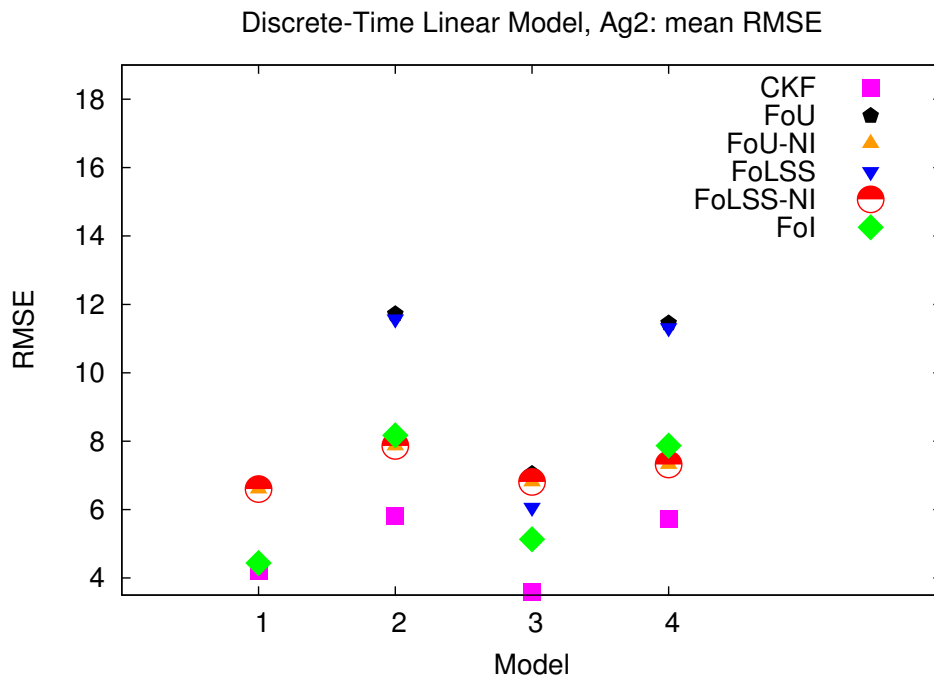


Figura 5.9: Sistema lineare: RMSE medi dell'agente 2 per i metodi di fusione considerati e del filtro centralizzato .

nari. Tutti i metodi di fusione considerati hanno maggiore difficoltà a stimare lo stato del sistema quando le sue componenti sono più fortemente accoppiate. Questo è particolarmente vero per i metodi che utilizzano la PDF diffusiva, $FoU - NI$ e

$FoLSS - NI$. Nel caso di componenti debolmente accoppiate, $FoLSS$ presenta generalmente errori di stima inferiori a quelli di $FoLSS - NI$; mentre nel caso di fusione sull'unione è $FoU - NI$ ad avere generalmente errori di stima inferiori a quelli di FoU . Infine si noti che, FoI presenta, nei sistemi considerati, errore di stima molto simile o inferiore a tutti gli altri metodi. La variabilità temporale dell'RMSE, generalmente, è più marcata in $FoU - NI$ e $FoLSS - NI$ nel caso di sistemi debolmente accoppiati, mentre è più marcata in FoU e $FoLSS$ nel caso di sistemi più fortemente accoppiati. La variabilità temporale dell'RMSE di FoI è simile o meno marcata di quella degli altri metodi di fusione.

5.2.2 Sistema non lineare

Per valutare i metodi di fusione proposti si è scelto un sistema fortemente non lineare tempo-continuo, il cosiddetto sistema di Lorenz [60], una cui traiettoria priva di disturbi è riportata in fig. 5.10, soggetto a un disturbo di processo additivo bianco e Gaussiano, e si è immaginato di poterne rilevare la dinamica nello spazio delle fasi con sensori non lineari. Si sono considerati cinque sensori che misurano la distanza di un punto della traiettoria dalla posizione del sensore (sensori di tipo TOA: *Time Of Arrival*); tali sensori sono soggetti a disturbo di misura additivo bianco e Gaussiano.

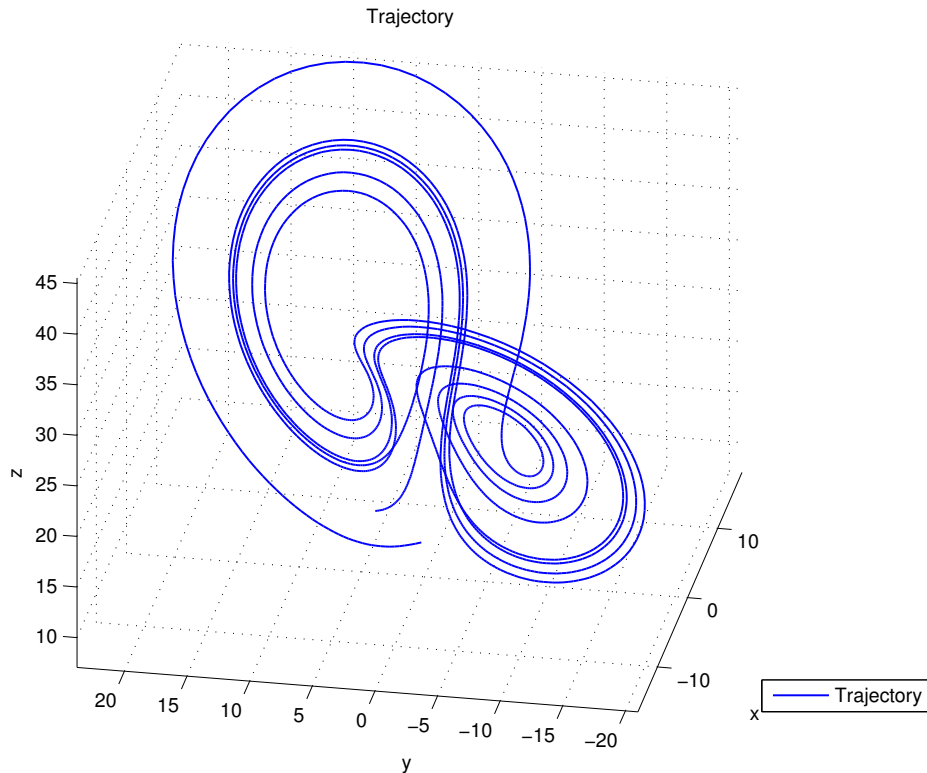


Figura 5.10: Esempio di traiettoria indisturbata del sistema non lineare.

Il sistema considerato è, quindi, il seguente

$$\begin{cases} \dot{x}_1 = \sigma(x_2 - x_1) \\ \dot{x}_2 = \rho x_1 - x_1 x_3 - x_2 + w \\ \dot{x}_3 = x_1 x_2 - \beta x_3 \\ y = h(x_1, x_2, x_3) + v \end{cases} \quad (5.3)$$

con $w \sim wgn(0, q)$ disturbo di processo, y osservazione, $h(x_1, x_2, x_3)$ funzione di misura e $v \sim wgn(0, r)$ disturbo di misura. Si assume che w , v e \mathbf{x}_0 siano mutualmente incorrelati.

Da questo sistema si è ricavato, mediante la discretizzazione temporale di Eulero [61], il seguente sistema a tempo-discreto:

$$\begin{cases} x_{1,k+1} = \Delta t [\sigma(x_{2,k} - x_{1,k})] \\ x_{2,k+1} = \Delta t [\rho x_{1,k} - x_{1,k} x_{3,k} - x_{2,k} + w_k] \\ x_{3,k+1} = \Delta t [x_{1,k} x_{2,k} - \beta x_{3,k}]. \end{cases} \quad (5.4)$$

Si suppone di disporre di cinque sensori, indicati con i ($1 \leq i \leq 5$), la cui disposizione è visibile in fig. 5.11. Dato il j -esimo sensore TOA, sia $\mathbf{x}_k = [x_{1,k}, x_{2,k}, x_{3,k}]^\top$ il vettore di stato e $\mathbf{p}^j = [p_1^j, p_2^j, p_3^j]^\top$ la posizione di tale sensore, allora la sua equazione di misura è

$$y_k^j = h(\mathbf{x}_k, \mathbf{p}^j) + v_k = \sqrt{(x_{1,k} - p_1^j)^2 + (x_{2,k} - p_2^j)^2 + (x_{3,k} - p_3^j)^2} + v_k \quad (5.5)$$

I sensori TOA, sono posizionati in:

$$\mathbf{p}^1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{p}^2 = \begin{bmatrix} 16000 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{p}^3 = \begin{bmatrix} 8000 \\ 8000 \\ 8000 \end{bmatrix}, \quad \mathbf{p}^4 = \begin{bmatrix} 0 \\ 16000 \\ 16000 \end{bmatrix}, \quad \mathbf{p}^5 = \begin{bmatrix} 16000 \\ 16000 \\ 16000 \end{bmatrix}.$$

Lo stato degli agenti, Ag^1 e Ag^2 risulta:

$$Ag^1: \mathbf{x}^1 = [x_1, x_2]^\top$$

$$Ag^2: \mathbf{x}^2 = [x_2, x_3]^\top$$

I sottosistemi relativi agli agenti sono quelli sotto riportati per l'agente 1 e, rispettivamente, per l'agente 2.

$$Ag^1: \begin{cases} x_{1,k+1} = \Delta t [\sigma(x_{2,k} - x_{1,k})] \\ x_{2,k+1} = \Delta t [\rho x_{1,k} - x_{1,k} u_{1,k} - x_{2,k} + w] \\ y_k = h(x_{1,k}, x_{2,k}, u_{1,k}) + v_k \end{cases} \quad (5.6)$$

$$Ag^2: \begin{cases} x_{2,k+1} = \Delta t [\rho u_{2,k} - u_{2,k} x_{3,k} - x_{2,k} + w] \\ x_{3,k+1} = \Delta t [u_{2,k} x_{2,k} - \beta x_{3,k}] \\ y_k = h(u_{2,k}, x_{2,k}, x_{3,k}) + v_k \end{cases} \quad (5.7)$$

Si noti che l'agente 1 non stima x_3 e, quindi, utilizza la stima di x_3 proveniente dall'agente 2 come fosse un ingresso esterno u_1 . Analogamente, l'agente 2 non stima x_1 ma utilizza la stima di x_1 proveniente dall'agente 1 come fosse un ingresso esterno u_2 .

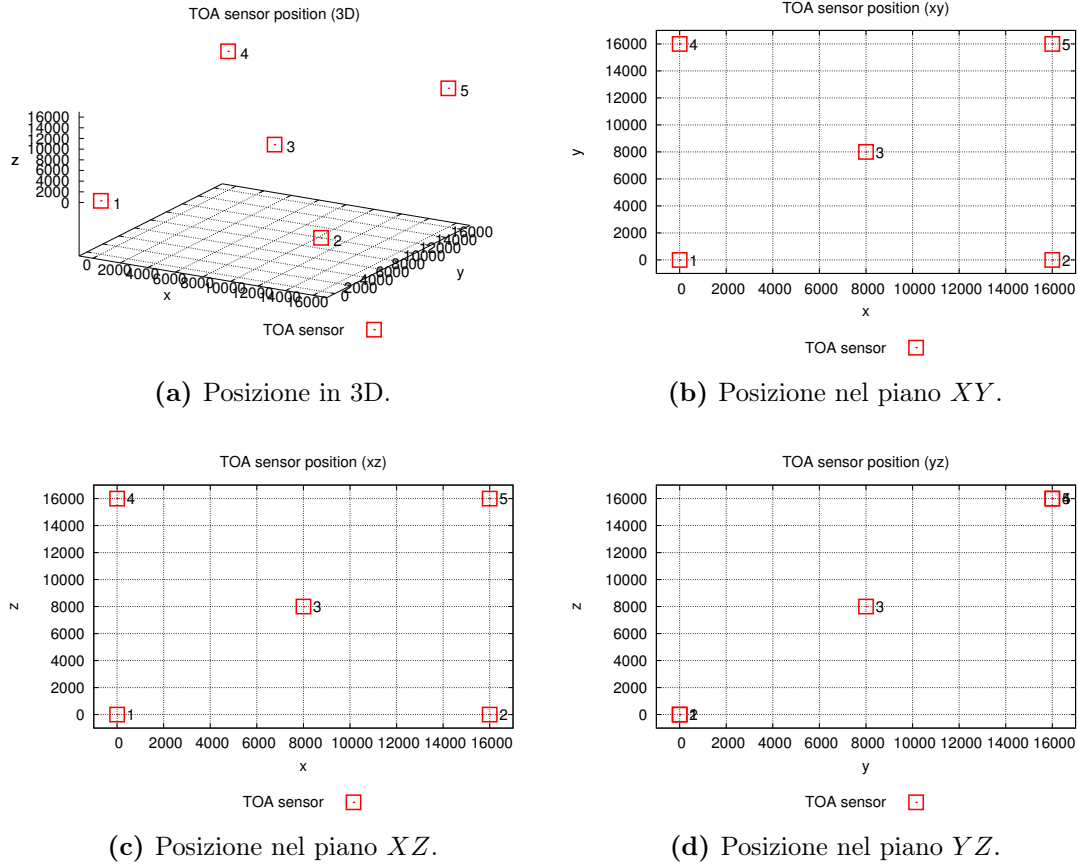


Figura 5.11: Posizioni dei cinque sensori TOA: vista in tre dimensioni e delle tre proiezioni sui piani degli assi coordinati.

Si suppone che il filtro centralizzato riceva ed elabori le misure dei cinque sensori, mentre l'agente 1 riceve le misure dai sensori 1, 2 e 3; e l'agente 2 dai sensori 3, 4 e 5. Inoltre i due agenti ricevono dai sensori le proiezioni delle misure sul sottospazio di stato di interesse, cioè (x_1, x_2) per l'agente 1 e (x_2, x_3) per l'agente 2.

Sono state effettuate 5000 prove Monte Carlo. In ciascuna prova viene generata una traiettoria della durata di 2000 passi la cui origine viene scelta casualmente nel volume a forma di parallelepipedo i cui punti hanno tutte le coordinate che possono variare nell'intervallo $[1, 11]$ in modo da evitare l'origine. In tutte le simulazioni sono stati scelti i valori $q = 1$, $r = 0.1$ e $\Delta t = 0.005$. In tutti i metodi di fusione considerati si è posto $\omega = 0.5$. Un esempio di traiettoria, sottoposta a disturbo, è riportato in fig. 5.12.

I filtri Bayesiani, sia quello centralizzato che quelli degli agenti, per i vari metodi di fusione considerati, vengono sempre inizializzati ponendo lo stato pari a un vettore nullo, cioè $\hat{\mathbf{x}}_0 = [0, 0, 0]^T$ per il centralizzato e $\hat{\mathbf{x}}_0^i = [0, 0]^T$ per gli agenti; le corrispondenti matrici di covarianza vengono poste a $\mathbf{P}_0 = \alpha \mathbf{I}_n$ con $\alpha = 10^6$ e \mathbf{I}_n opportuna matrice identità di ordine n , con $n = 3$ per il filtro centralizzato ed $n = 2$ per i filtri relativi agli agenti.

Nella fig. D.5 in Appendice D si sono riportati gli RMSE, mediati su tutte le prove Monte Carlo, degli agenti Ag^1 , Ag^2 per i vari metodi di fusione considerati

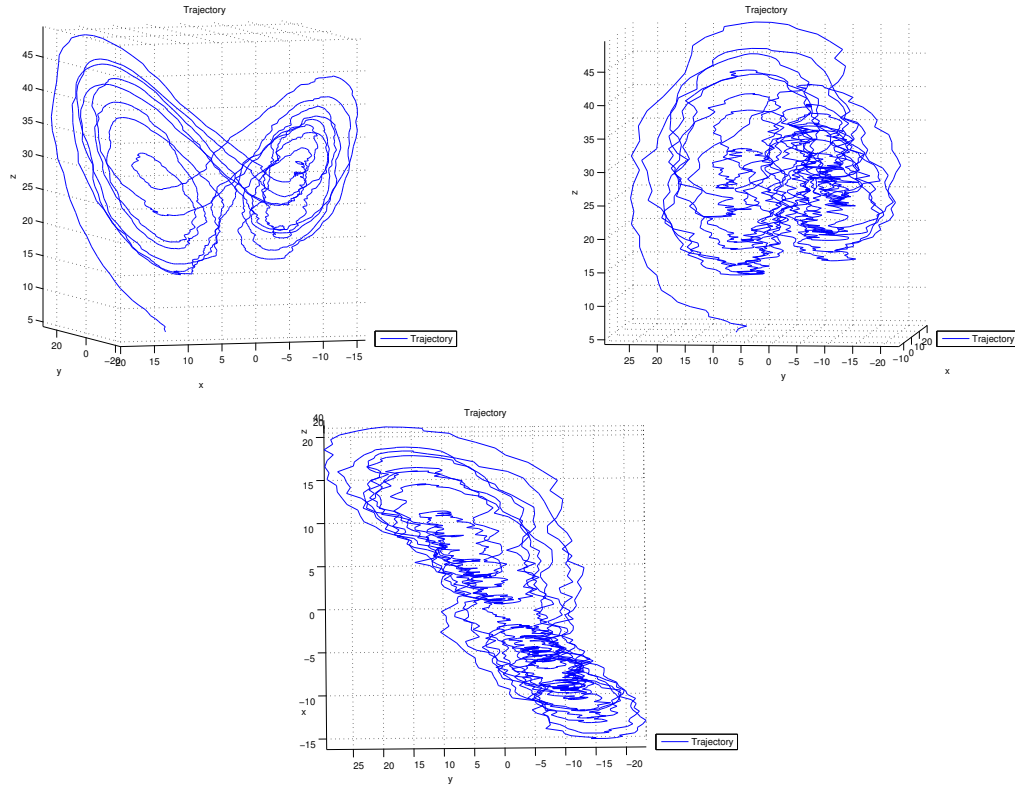


Figura 5.12: Esempio di una traiettoria disturbata del sistema non-lineare considerato da tre diversi punti di vista.

	<i>CUKF</i>	<i>FoU</i>	<i>FoU – NI</i>	<i>FoLSS</i>	<i>FoLSS – NI</i>	<i>FoI</i>
Ag^1	0.14	0.77	0.53	0.69	0.53	0.50
Ag^2	0.15	1.11	0.70	1.11	0.70	1.03

Tabella 5.2: Sistema non-lineare: RMSE medi degli agenti e del filtro centralizzato per i metodi di fusione considerati.

e del filtro centralizzato. Anche in questo caso si è considerato come indice di prestazione sintetico il valore dell’RMSE mediato sul tempo, riportato in fig. 5.13 ed in tab. 5.2 per meglio evidenziare le differenze.

Le prestazioni migliori, in termini di RMSE, sono quelle dei metodi *FoU – NI* e *FoLSS – NI*. Il metodi *FoI* ha prestazioni comparabili con i due precedenti nel caso dell’agente 1 ma peggiori nel caso dell’agente 2. Gli altri due metodi *FoU* e *FoLSS* hanno prestazioni nettamente peggiori degli altri.

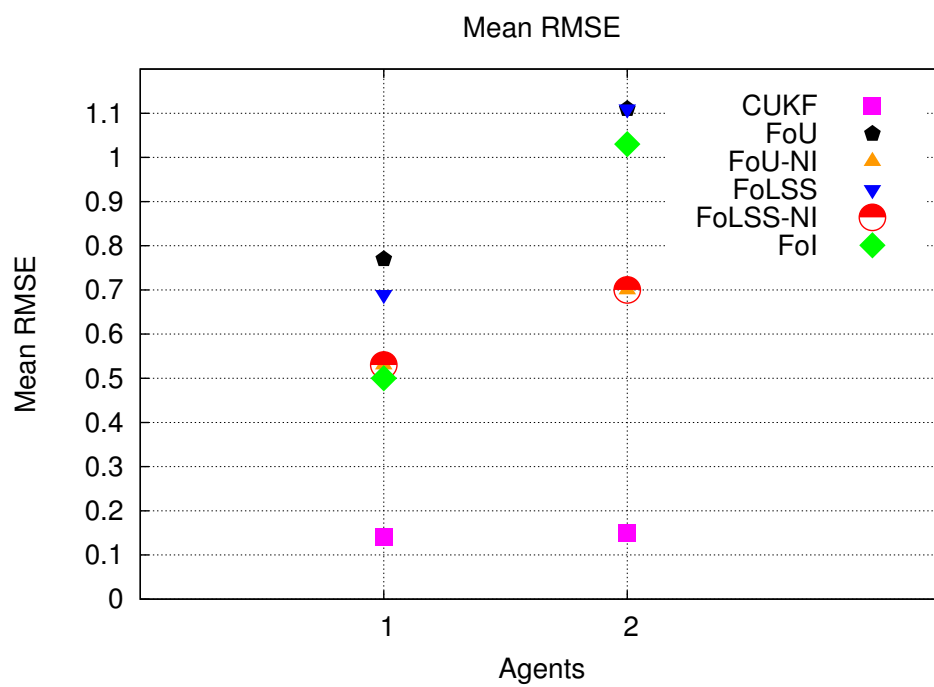


Figura 5.13: Sistema non-lineare: RMSE medi degli agenti e del filtro centralizzato per i metodi di fusione considerati.

Conclusioni

In questo lavoro di tesi sono stati presi in esame alcuni problemi aperti legati alla realizzazione di sistemi di monitoraggio e sorveglianza distribuiti, costituiti dall'interconnessione di nodi di elaborazione, comunicazione e misura (sensing). In particolare, è stata affrontata la stima distribuita dello stato di un sistema dinamico, sia lineare che non lineare, ricorrendo al ben noto approccio del consenso. In tale ambito, sono stati forniti alcuni nuovi contributi: l'estensione di tecniche di consenso esistenti per il filtraggio alla Kalman, cioè il consenso sulla verosimiglianza [5, 6] ed il consenso sull'informazione a posteriori [17], ai sistemi non lineari; è stato inoltre proposto un nuovo approccio al consenso, che combina i due tipi di consenso per fornire prestazioni migliori; è stata effettuata un'analisi di stabilità per il nuovo approccio di consenso proposto nel caso lineare. Infine, un insieme di esperimenti simulativi relativi ad un problema di inseguimento di un bersaglio, con sensori lineari e/o non lineari, ha dimostrato l'efficacia degli algoritmi proposti.

Come possibile sviluppo futuro si osserva che anche se l'analisi di stabilità effettuata si applica solo al caso lineare tempo-invariante, in realtà potrebbe essere utilizzata come punto di partenza per lo studio della stabilità locale della dinamica dell'errore di stima anche nel contesto dei sistemi non lineari.

Un altro contributo di questo lavoro ha riguardato lo studio della diffusione di un virus informatico su una rete di sensori ed i suoi effetti sulla stima distribuita. A tale proposito, sono state stabilite connessioni tra la topologia della rete, la possibilità di sviluppo di un focolaio epidemico, e le prestazioni di algoritmi di fusione dell'informazione distribuita. È stato mostrato, mediante simulazioni realistiche, che l'aumento della connettività di rete, ottenuta ad esempio mediante aumento della potenza di trasmissione del nodo, comporta due effetti contrastanti: i) si rende la rete più robusta rispetto al malfunzionamento di alcuni nodi; ii) si rende la rete più incline alla diffusione del virus (si abbassa la soglia epidemica e si ha una maggiore prevalenza epidemica). Inoltre, è stato mostrato che i nodi con capacità di recupero sono da preferirsi poiché un'epidemia SIR ha soltanto un effetto limitato nel tempo, mentre un'epidemia SIS è quasi persistente.

I risultati analizzati finora riguardano reti casuali e virus non-cooperativi. Possibili sviluppi futuri riguarderanno reti regolari e/o virus a comportamento deviante o malizioso. A tale proposito, ci si aspetta che le reti regolari possano essere meno soggette agli effetti del virus, in quanto per tali reti risulta possibile ridurre la distanza tra connettività algebrica e raggio spettrale della matrice di adiacenza. Per quanto riguarda, viceversa, virus a comportamento deviante e/o malizioso, l'aumento della connettività potrebbe risultare più problematico, dal momento che, per queste tipologie di virus, pochi nodi infetti potrebbero compromettere seriamente il

funzionamento di un algoritmo di fusione dell'informazione distribuito. In questo caso, un obiettivo stimolante è quello di sviluppare algoritmi per il rilevamento dei virus, il loro isolamento e la loro sistemazione nel contesto di una rete distribuita.

Un'ulteriore prospettiva di lavoro futuro è quella di validare i risultati discussi in questa tesi per mezzo di simulazioni ancora più realistiche ottenute con un simulatore di rete basato sulla comunicazione (ad esempio NS2 o OPNET) e/o con esperimenti reali su una rete di sensori di test.

Infine, un ulteriore contributo della tesi è stato quello di studiare il problema della fusione fra agenti a stato parzialmente sovrapposto, situazione che si presenta in molti casi di interesse pratico relativi a sistemi di sorveglianza e/o monitoraggio con agenti (nodi) distribuiti sul territorio. In tale ambito, cinque diversi metodi di fusione (FoU , $FoU - NI$, $FoLSS$, $FoLSS - NI$ e FoI) sono stati validati e confrontati in due scenari distinti relativi ad un sistema lineare e, rispettivamente, non lineare. I risultati simulativi ottenuti con filtri di consenso basati su tali metodi di fusione, seppur al momento non ancora supportati da risultati analitici di stabilità dei filtri, evidenziano buone prestazioni di stima e ne suggeriscono un potenziale efficace impiego per il monitoraggio di sistemi spazialmente distribuiti. A tale proposito, un immediato sviluppo di questo tema di ricerca sarà quello di applicare queste tecniche a sistemi a parametri distribuiti, definiti da equazioni differenziali alle derivate parziali utilizzando, ad esempio, il metodo degli elementi finiti.

Appendice A

Dimostrazione del Teorema 1:

Sia

$$\tilde{p}(\mathbf{x}) = \prod_i [p^i(\mathbf{x})]^{\pi^i}, \quad c = \int \tilde{p}(\mathbf{x}) d\mathbf{x}$$

così che la PDF in (2.9) possa essere scritta come $\bar{p}(\mathbf{x}) = \tilde{p}(\mathbf{x})/c$. Allora, il costo da minimizzare mediante la KLA è

$$\begin{aligned} J(p) &\triangleq \sum_{i=1}^N \pi^i D_{KL}(p \parallel p^i) = \sum_{i=1}^N \pi^i \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{p^i(\mathbf{x})} d\mathbf{x} \\ &= \int p(\mathbf{x}) \left[\sum_{i=1}^N \pi^i \log \frac{p(\mathbf{x})}{p^i(\mathbf{x})} \right] d\mathbf{x} \\ &= \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\prod_{i=1}^N p^i(\mathbf{x})^{\pi^i}} d\mathbf{x} \\ &= \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\tilde{p}(\mathbf{x})} d\mathbf{x} \\ &= \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{c \bar{p}(\mathbf{x})} d\mathbf{x} \\ &= \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\bar{p}(\mathbf{x})} d\mathbf{x} - \log(c) \int p(\mathbf{x}) d\mathbf{x} \\ &= D_{KL}(p \parallel \bar{p}) - \log(c) \end{aligned} \tag{A.1}$$

Dato che l'ultimo termine $-\log c$ non dipende da p e la KLD di due PDF è sempre non negativa e uguale a zero se e solo se le due PDF coincidono quasi ovunque, si può ottenere la minimizzazione di $J(p)$ rispetto a p azzerando il primo termine $D_{KL}(p \parallel \bar{p})$, cioè ponendo $p = \bar{p}$. Quindi, la KLA pesata $\bar{p}(\cdot)$ definita da (2.8) coincide con la media geometrica (pesata) normalizzata data da (2.9).

□

Dimostrazione del Teorema 2:

Per dimostrare il risultato è opportuno sfruttare le seguenti proprietà elementari

delle funzioni Gaussiane:

$$[\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})]^\pi \propto \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \pi^{-1}\boldsymbol{\Sigma}) \quad (\text{A.2})$$

$$\prod_{i=1}^N \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i) \propto \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{dove} \quad (\text{A.3})$$

$$\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^N (\boldsymbol{\Sigma}^i)^{-1} \quad \text{e} \quad \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \sum_{i=1}^N (\boldsymbol{\Sigma}^i)^{-1} \boldsymbol{\mu}^i \quad (\text{A.4})$$

La verifica della (A.2) discende direttamente dalla definizione di funzione Gaussiana, infatti

$$[\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})]^\pi \propto \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top [\pi^{-1}\boldsymbol{\Sigma}]^{-1} (\mathbf{x} - \boldsymbol{\mu})\right] \propto \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \pi^{-1}\boldsymbol{\Sigma})$$

Per verificare (A.4), si noti che

$$\prod_{i=1}^N \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i) \propto \exp\left\{-\frac{1}{2}\left[\mathbf{x}^\top \sum_{i=1}^N (\boldsymbol{\Sigma}^i)^{-1} \mathbf{x} - \left(\sum_{i=1}^N (\boldsymbol{\Sigma}^i)^{-1} \boldsymbol{\mu}^i\right)^\top \mathbf{x} - \mathbf{x}^\top \sum_{i=1}^N (\boldsymbol{\Sigma}^i)^{-1} \boldsymbol{\mu}^i + \sum_{i=1}^N (\boldsymbol{\mu}^i)^\top (\boldsymbol{\Sigma}^i)^{-1} \boldsymbol{\mu}^i\right]\right\}$$

Allora, sfruttando le relazioni di $\boldsymbol{\Sigma}^{-1}$ e di $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ in (A.4), è immediato ottenere

$$\prod_{i=1}^N \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i) \propto \exp\left\{-\frac{1}{2}\left[\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}\right]\right\} \propto \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right] \propto \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Ora, supponendo che (2.15) e sfruttando (A.2)-(A.4),

$$\bar{p}(\mathbf{x}) \propto \prod_{i=1}^N [\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i)]^{\pi^i} \propto \prod_{i=1}^N \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}^i, \frac{\boldsymbol{\Sigma}^i}{\pi^i}\right) \propto \mathcal{N}(\mathbf{x}; \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$$

dove $\bar{\boldsymbol{\mu}}$ e $\bar{\boldsymbol{\Sigma}}$ soddisfano banalmente (2.16) e (2.17). Quindi, $\bar{p}(\mathbf{x}) = c \mathcal{N}(\mathbf{x}; \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$ per qualche costante di normalizzazione $c \in \mathbb{R}$. Dato che

$$\int \bar{p}(\mathbf{x}) d\mathbf{x} = c \int \mathcal{N}(\mathbf{x}; \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}) d\mathbf{x} = c = 1$$

il risultato è dimostrato. □

Per brevità si riscrive il passo di predizione sulla matrice d'informazione (3.9) come

$$\Omega_{k+1|k}^i = \Psi(\Omega_{k|k}^i)$$

con

$$\Psi(\Omega) \triangleq \mathbf{A}^{-\top} \Omega \mathbf{A}^{-1} - \mathbf{A}^{-\top} \Omega (\Omega + \mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A})^{-1} \Omega \mathbf{A}^{-1}.$$

Dimostrazioni dei Teoremi 3 e 4:

Per dimostrare i Teoremi 3 e 4, conviene avvalersi del seguente Lemma.

Lemma 1. *Nell'Ipotesi A1 valgono i seguenti fatti.*

- i) *La funzione $\Psi(\cdot)$ è monotona non decrescente, cioè, date due matrici semidefinite positive Ω_1 e Ω_2 con $\Omega_1 \leq \Omega_2$ si ha $\mathbf{0} \leq \Psi(\Omega_1) \leq \Psi(\Omega_2)$.*
- ii) *Per ogni matrice semidefinita positiva $\check{\Omega}$, esiste un numero reale strettamente positivo $\check{\beta} \leq 1$ tale che $\Psi(\Omega) \geq \check{\beta} \mathbf{A}^{-\top} \Omega \mathbf{A}^{-1}$ per ogni $\Omega \leq \check{\Omega}$.*
- iii) *Per ogni matrice definita positiva $\tilde{\Omega}$, esiste un numero reale positivo $\tilde{\beta} < 1$ tale che $\Psi(\Omega) \leq \tilde{\beta} \mathbf{A}^{-\top} \Omega \mathbf{A}^{-1}$ per ogni $\Omega \geq \tilde{\Omega}$.*

Dimostrazione del Lemma 1:

i) Si ricorda che, per ogni matrice non singolare Ω , si ha

$$\Psi(\Omega) = (\mathbf{A} \Omega^{-1} \mathbf{A}^\top + \mathbf{Q})^{-1}. \quad (\text{A.5})$$

Si considerano ora due matrici Ω_1 e Ω_2 con $\Omega_1 \leq \Omega_2$ e si definisce $\Omega_1(\alpha) = \Omega_1 + \alpha \mathbf{I}$, $\Omega_2(\alpha) = \Omega_2 + \alpha \mathbf{I}$. Usando (A.5), è immediato vedere che, per ogni $\alpha > 0$,

$$\Psi(\Omega_1(\alpha))^{-1} - \Psi(\Omega_2(\alpha))^{-1} = \mathbf{A} (\Omega_1(\alpha)^{-1} - \Omega_2(\alpha)^{-1}) \mathbf{A}^\top \geq \mathbf{0}$$

e, quindi, $\Psi(\Omega_2(\alpha)) - \Psi(\Omega_1(\alpha)) \geq \mathbf{0}$. Allora la dimostrazione del fatto i) segue per questioni di continuità quando $\alpha \rightarrow 0^+$.

ii) Si considera una generica $\Omega \leq \check{\Omega}$ e sia $\Omega(\alpha) = \Omega + \alpha \mathbf{I}$ con $\alpha > 0$. Usando di nuovo (A.5), si può scrivere

$$\Psi(\Omega(\alpha)) = \mathbf{A}^{-\top} (\Omega(\alpha)^{-1} + \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-\top})^{-1} \mathbf{A}^{-1}. \quad (\text{A.6})$$

Si noti ora che, per ogni fissato $\check{\alpha}$, si ha $\Omega(\alpha)^{-1} \geq (\check{\Omega} + \check{\alpha} \mathbf{I})^{-1}$ per ogni $0 < \alpha \leq \check{\alpha}$ e $\mathbf{0} \leq \Omega \leq \check{\Omega}$. Questo, a sua volta, implica che esiste uno scalare positivo $\check{\gamma}$ tale che

$$\mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-\top} \leq \check{\gamma} \Omega(\alpha)^{-1} \quad (\text{A.7})$$

per ogni $0 < \alpha \leq \check{\alpha}$ e $\mathbf{0} \leq \Omega \leq \check{\Omega}$. Combinando (A.6) con (A.7), si ottiene

$$\Psi(\Omega(\alpha)) \geq \frac{1}{1 + \check{\gamma}} \mathbf{A}^{-\top} \Omega(\alpha) \mathbf{A}^{-1}.$$

Allora, la dimostrazione del fatto ii) può essere ottenuta ponendo $\check{\beta} = 1/(1 + \check{\gamma})$ e notando che l'ultima disuguaglianza deve valere anche al limite per $\alpha \rightarrow 0^+$.

iii) Data una generica $\mathbf{\Omega} \geq \tilde{\mathbf{\Omega}} > \mathbf{0}$, si può scrivere

$$\Psi(\mathbf{\Omega}) = \mathbf{A}^{-\top} (\mathbf{\Omega}^{-1} + \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-\top})^{-1} \mathbf{A}^{-1}.$$

Dato che $\mathbf{Q} > \mathbf{0}$, esiste sempre uno scalare positivo $\tilde{\gamma}$ tale che

$$\mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-\top} \geq \tilde{\gamma} \tilde{\mathbf{\Omega}}^{-1} \geq \tilde{\gamma} \mathbf{\Omega}^{-1}$$

e, da qui,

$$\Psi(\mathbf{\Omega}) \leq \frac{1}{1 + \tilde{\gamma}} \mathbf{A}^{-\top} \mathbf{\Omega} \mathbf{A}^{-1}.$$

Allora la dimostrazione del fatto iii) può essere ottenuta ponendo $\check{\beta} = 1/(1 + \check{\gamma})$. □

Dimostrazione del Teorema 3:

Può essere ottenuta per induzione. A questo scopo, si suppone che, ad un generico istante di tempo k ,

$$\mathbf{\Omega}_{k|k-1}^i \leq (\mathbb{E} \{ (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i)^\top \})^{-1} \quad (\text{A.8})$$

per ogni $i \in \mathcal{N}$. Chiaramente, definendo $\mathcal{I}^i \triangleq (\mathbf{C}^i)^\top (\mathbf{R}^i)^{-1} \mathbf{C}^i$, (A.8) implica anche

$$\begin{aligned} & (\mathbb{E} \{ (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k,0}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k,0}^i)^\top \})^{-1} = \\ & (\mathbb{E} \{ (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i)^\top \})^{-1} + \mathcal{I}^i \geq \mathbf{\Omega}_{k|k-1}^i + \mathcal{I}^i = \mathbf{\Omega}_{k|k,0}^i. \end{aligned}$$

Allora, dato che la regola di fusione della Covariance Intersection conserva le proprietà di consistenza [3, 4], cioè

$$\begin{aligned} & \left(E \left\{ \tilde{\mathbf{x}}_{k|k,\ell}^i (\tilde{\mathbf{x}}_{k|k,\ell}^i)^\top \right\} \right)^{-1} \geq \mathbf{\Omega}_{k|k,\ell} \implies \\ & \left(E \left\{ \tilde{\mathbf{x}}_{k|k,\ell+1}^i (\tilde{\mathbf{x}}_{k|k,\ell+1}^i)^\top \right\} \right)^{-1} \geq \mathbf{\Omega}_{k|k,\ell+1} \quad \text{per ogni } \ell = 0, \dots, L-1, \end{aligned}$$

si può concludere che (3.29) vale all'istante k con $\hat{\mathbf{x}}_{k|k}^i = \hat{\mathbf{x}}_{k|k,L}^i$ e $\mathbf{\Omega}_{k|k}^i = \mathbf{\Omega}_{k|k,L}^i$. Inoltre, in virtù del fatto i) del Lemma 1, è immediato vedere che

$$\begin{aligned} \mathbf{\Omega}_{k+1|k}^i &= \Psi(\mathbf{\Omega}_{k|k}^i) \leq \Psi \left((\mathbb{E} \{ (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)^\top \})^{-1} \right) = \\ & \left(E \left\{ (\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}^i)(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}^i)^\top \right\} \right)^{-1}. \end{aligned}$$

La dimostrazione può essere conclusa notando che, con l'inizializzazione (3.12)-(3.13), la disuguaglianza (A.8) vale automaticamente per $k = 0$. □

Dimostrazione del Teorema 4:

Per brevità la dimostrazione viene data solo per il caso $L = 1$. L'estensione ad un generico $L \geq 1$ è semplice.

Quando $L = 1$, ogni matrice d'informazione fusa può essere scritta come

$$\mathbf{\Omega}_{k|k}^i = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathbf{\Omega}_{k|k-1}^j + \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathcal{I}^j = \sum_{j \in \mathcal{N}^i} \pi^{i,j} \boldsymbol{\psi} \left(\mathbf{\Omega}_{k-1|k-1}^j \right) + \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathcal{I}^j. \quad (\text{A.9})$$

Inoltre, in considerazione della disuguaglianza (3.29), si ha che

$$\mathbf{\Omega}_{k|k}^i \leq \left(\mathbb{E} \left\{ (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)^\top \right\} \right)^{-1} \leq \mathbf{\Omega}_{k|k}$$

dove $\mathbf{\Omega}_{k|k}$ è la matrice d'informazione del filtro di Kalman centralizzato (che, in questo caso, rappresenta il limite superiore di Cramer-Rao per la matrice d'informazione di un filtro non polarizzato). Dato che \mathbf{Q} e \mathbf{R}^i , $i \in \mathcal{N}$, sono definite positive, la sequenza $\mathbf{\Omega}_{k|k}$ può essere limitata uniformemente da una matrice costante $\tilde{\mathbf{\Omega}}$. Così, si può utilizzare il fatto ii) del Lemma 1 ed ottenere il limite inferiore

$$\mathbf{\Omega}_{k|k}^i \geq \sum_{j \in \mathcal{N}^i} \pi^{i,j} \check{\beta} \mathbf{A}^{-\top} \mathbf{\Omega}_{k-1|k-1}^j \mathbf{A}^{-1} + \sum_{j \in \mathcal{N}^i} \pi^{i,j} \mathcal{I}^j \quad (\text{A.10})$$

per qualche numero reale positivo $\check{\beta}$. Applicando ricorsivamente le equazioni (A.9) e (A.10) un certo numero, \bar{k} , di volte, è possibile scrivere

$$\mathbf{\Omega}_{k|k}^i \geq \check{\beta}^{\bar{k}} \left(\mathbf{A}^{-\bar{k}} \right)^\top \left(\sum_{j \in \mathcal{N}} \pi_{\bar{k}}^{i,j} \mathbf{\Omega}_{k-\bar{k}|k-\bar{k}}^j \right) \mathbf{A}^{-\bar{k}} + \sum_{\tau=1}^{\bar{k}} \check{\beta}^{\tau-1} \left(\mathbf{A}^{-\tau+1} \right)^\top \left(\sum_{j \in \mathcal{N}} \pi_{\tau}^{i,j} \mathcal{I}^j \right) \mathbf{A}^{-\tau+1}.$$

dove $\pi_{\tau}^{i,j}$ indica l'(i, j)-esimo elemento della matrice $\mathbf{\Pi}^\tau$. Allora, la dimostrazione può essere terminata definendo

$$\tilde{\mathbf{\Omega}} \triangleq \sum_{\tau=1}^{\bar{k}} \check{\beta}^{\tau-1} \left(\mathbf{A}^{-\tau+1} \right)^\top \left(\sum_{j \in \mathcal{N}} \pi_{\tau}^{i,j} \mathcal{I}^j \right) \mathbf{A}^{-\tau+1}.$$

Infatti, quando la matrice di consenso $\mathbf{\Pi}$ è primitiva, gli elementi $\pi_{\tau}^{i,j}$ sono tutti positivi per τ maggiore di un certo m . Allora, è facile verificare che, in condizioni di osservabilità collettiva, la matrice $\tilde{\mathbf{\Omega}}$ è definita positiva per \bar{k} maggiore di $m + n$. \square

Dimostrazione del Teorema 5:

Per la dimostrazione del Teorema 5 è utile il seguente Lemma.

Lemma 2. *Dati un intero $N \geq 2$, N matrici definite positive $\mathbf{M}_1, \dots, \mathbf{M}_N$, e N vettori $\mathbf{v}_1, \dots, \mathbf{v}_N$, valgono le seguenti disuguaglianze*

$$\left(\sum_{i=1}^N \mathbf{M}_i \mathbf{v}_i \right)^\top \left(\sum_{i=1}^N \mathbf{M}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{M}_i \mathbf{v}_i \right) \leq \sum_{i=1}^N \mathbf{v}_i^\top \mathbf{M}_i \mathbf{v}_i. \quad (\text{A.11})$$

Dimostrazione del Lemmma 2: La dimostrazione è svolta per induzione. Si considera prima il caso $N = 2$. Allora, il membro sinistro di (A.11) diventa

$$\begin{aligned} (\mathbf{M}_1 \mathbf{v}_1 + \mathbf{M}_2 \mathbf{v}_2)^\top (\mathbf{M}_1 + \mathbf{M}_2)^{-1} (\mathbf{M}_1 \mathbf{v}_1 + \mathbf{M}_2 \mathbf{v}_2) = \\ \mathbf{v}_1^\top \mathbf{M}_1 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_1 \mathbf{v}_1 + \mathbf{v}_2^\top \mathbf{M}_2 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_2 \mathbf{v}_2 + \\ \mathbf{v}_1^\top \mathbf{M}_1 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_2 \mathbf{v}_2 + \mathbf{v}_2^\top \mathbf{M}_2 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_1 \mathbf{v}_1. \end{aligned}$$

Dato che la matrice

$$\mathbf{M}_1 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_2 = (\mathbf{M}_1^{-1} + \mathbf{M}_2^{-1})^{-1} = \mathbf{M}_2 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_1$$

è definita positiva, si ha

$$\begin{aligned} \mathbf{v}_1^\top \mathbf{M}_1 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_2 \mathbf{v}_2 + \mathbf{v}_2^\top \mathbf{M}_2 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_1 \mathbf{v}_1 \leq \\ \mathbf{v}_1^\top \mathbf{M}_1 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_2 \mathbf{v}_1 + \mathbf{v}_2^\top \mathbf{M}_2 (\mathbf{M}_1 + \mathbf{M}_2)^{-1} \mathbf{M}_1 \mathbf{v}_2 \end{aligned}$$

e da qui

$$(\mathbf{M}_1 \mathbf{v}_1 + \mathbf{M}_2 \mathbf{v}_2)^\top (\mathbf{M}_1 + \mathbf{M}_2)^{-1} (\mathbf{M}_1 \mathbf{v}_1 + \mathbf{M}_2 \mathbf{v}_2) \leq \mathbf{v}_1^\top \mathbf{M}_1 \mathbf{v}_1 + \mathbf{v}_2^\top \mathbf{M}_2 \mathbf{v}_2 \quad (\text{A.12})$$

che dimostra l'affermazione per $N = 2$. Si considera ora un generico N e si suppone che (A.11) valga per $N - 1$, cioè,

$$\left(\sum_{i=1}^{N-1} \mathbf{M}_i \mathbf{v}_i \right)^\top \left(\sum_{i=1}^{N-1} \mathbf{M}_i \right)^{-1} \left(\sum_{i=1}^{N-1} \mathbf{M}_i \mathbf{v}_i \right) \leq \sum_{i=1}^{N-1} \mathbf{v}_i^\top \mathbf{M}_i \mathbf{v}_i. \quad (\text{A.13})$$

Definendo

$$\mathbf{M} = \sum_{i=1}^{N-1} \mathbf{M}_i, \quad \mathbf{v} = \mathbf{M}^{-1} \left(\sum_{i=1}^{N-1} \mathbf{M}_i \mathbf{v}_i \right),$$

si può scrivere

$$\begin{aligned} \left(\sum_{i=1}^N \mathbf{M}_i \mathbf{v}_i \right)^\top \left(\sum_{i=1}^N \mathbf{M}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{M}_i \mathbf{v}_i \right) = \\ (\mathbf{M} \mathbf{v} + \mathbf{M}_N \mathbf{v}_N)^\top (\mathbf{M} + \mathbf{M}_N)^{-1} (\mathbf{M} \mathbf{v} + \mathbf{M}_N \mathbf{v}_N) \leq \\ \mathbf{v}^\top \mathbf{M} \mathbf{v} + \mathbf{v}_N^\top \mathbf{M}_N \mathbf{v}_N \end{aligned}$$

dove l'ultima disuguaglianza segue da (A.12). La dimostrazione si conclude affermando che

$$\mathbf{v}^\top \mathbf{M} \mathbf{v} = \left(\sum_{i=1}^{N-1} \mathbf{M}_i \mathbf{v}_i \right)^\top \left(\sum_{i=1}^{N-1} \mathbf{M}_i \right)^{-1} \left(\sum_{i=1}^{N-1} \mathbf{M}_i \mathbf{v}_i \right)$$

e, quindi, applicando (A.13).

□

Per dimostrare il Teorema 5 si indicano, in forma compatta, l'errore di predizione e di stima al nodo i con, rispettivamente,

$$\begin{aligned}\tilde{\mathbf{x}}_{k|k-1}^i &\triangleq \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i \quad \text{e} \\ \tilde{\mathbf{x}}_{k|k}^i &\triangleq \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i.\end{aligned}\tag{A.14}$$

In considerazione del Teorema 4 e del fatto iii) del Lemma 1, si ha

$$\begin{aligned}\mathcal{L}_{k+1}^i(\tilde{\mathbf{x}}_{k+1|k}^i) &= (\tilde{\mathbf{x}}_{k+1|k}^i)^\top \boldsymbol{\Omega}_{k+1|k}^i \tilde{\mathbf{x}}_{k+1|k}^i = \\ &= [(\tilde{\mathbf{x}}_{k+1|k}^i)^\top \boldsymbol{\Psi} (\boldsymbol{\Omega}_{k|k}^i \tilde{\mathbf{x}}_{k+1|k}^i)]^{1/2} \\ &\leq \tilde{\beta} [(\tilde{\mathbf{x}}_{k+1|k}^i)^\top \mathbf{A}^{-\top} \boldsymbol{\Omega}_{k|k}^i \mathbf{A}^{-1} \tilde{\mathbf{x}}_{k+1|k}^i]\end{aligned}$$

per qualche numero reale positivo $\tilde{\beta} < 1$ e per $k \geq \bar{k}$. Questo, a sua volta, implica

$$\mathcal{L}_{k+1}^i(\tilde{\mathbf{x}}_{k+1|k}^i) \leq \tilde{\beta} [(\tilde{\mathbf{x}}_{k|k}^i)^\top \boldsymbol{\Omega}_{k|k}^i \tilde{\mathbf{x}}_{k|k}^i]\tag{A.15}$$

dato che $\tilde{\mathbf{x}}_{k+1|k}^i = \mathbf{A} \tilde{\mathbf{x}}_{k|k}^i$.

Si noti ora che il passo locale di correzione dell'Algoritmo 10 senza rumore, può essere scritto come

$$\boldsymbol{\Omega}_{k|k,0}^j \hat{\mathbf{x}}_{k|k,0}^j = \boldsymbol{\Omega}_{k|k-1}^j \hat{\mathbf{x}}_{k|k-1}^j + \boldsymbol{\mathcal{I}}^j \mathbf{x}_k\tag{A.16}$$

dove $\boldsymbol{\mathcal{I}}^j \triangleq (\mathbf{C}^j)^\top (\mathbf{R}^j)^{-1} \mathbf{C}^j$. La precedente espressione è stata ottenuta da (3.6) con

$$\begin{aligned}\mathbf{y}_k^j &= \mathbf{C}^j \mathbf{x}_k \quad \text{al posto di } \mathbf{y}_k^i, \\ \mathbf{q}_{k|k,0}^j &= \boldsymbol{\Omega}_{k|k,0}^j \hat{\mathbf{x}}_{k|k,0}^j \quad \text{al posto di } \mathbf{q}_{k|k}^i, \\ \mathbf{q}_{k|k-1}^j &= \boldsymbol{\Omega}_{k|k-1}^j \hat{\mathbf{x}}_{k|k-1}^j \quad \text{al posto di } \mathbf{q}_{k|k-1}^i.\end{aligned}\tag{A.17}$$

Si ricorda anche che, per mezzo del consenso,

$$\begin{aligned}\boldsymbol{\Omega}_{k|k}^i &\triangleq \boldsymbol{\Omega}_{k|k,L} = \sum_{j \in \mathcal{N}} \pi_L^{i,j} \boldsymbol{\Omega}_{k|k,0}^j \\ \boldsymbol{\Omega}_{k|k}^i \tilde{\mathbf{x}}_{k|k}^i &\triangleq \boldsymbol{\Omega}_{k|k,L}^i \tilde{\mathbf{x}}_{k|k,L}^i = \sum_{j \in \mathcal{N}} \pi_L^{i,j} \boldsymbol{\Omega}_{k|k,0}^j \tilde{\mathbf{x}}_{k|k,0}^j.\end{aligned}\tag{A.18}$$

Quindi, pre-moltiplicando la prima equazione di (A.18) per $(\boldsymbol{\Omega}_{k|k}^i)^{-1}$ e post-moltiplicandola per \mathbf{x}_k , si ottiene

$$\mathbf{x}_k = (\boldsymbol{\Omega}_{k|k}^i)^{-1} \sum_{j \in \mathcal{N}} \pi_L^{i,j} \boldsymbol{\Omega}_{k|k,0}^j \mathbf{x}_k.\tag{A.19}$$

Analogamente, pre-moltiplicando la seconda equazione di (A.18) per $(\boldsymbol{\Omega}_{k|k}^i)^{-1}$, si ottiene

$$\hat{\mathbf{x}}_{k|k}^i = (\boldsymbol{\Omega}_{k|k}^i)^{-1} \sum_{j \in \mathcal{N}} \pi_L^{i,j} \boldsymbol{\Omega}_{k|k,0}^j \hat{\mathbf{x}}_{k|k,0}^j.\tag{A.20}$$

E ancora, sottraendo (A.20) da (A.19), si ottiene

$$\tilde{\mathbf{x}}_{k|k}^i \triangleq \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i = (\boldsymbol{\Omega}_{k|k}^i)^{-1} \sum_{j \in \mathcal{N}} \pi_L^{i,j} \boldsymbol{\Omega}_{k|k,0}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k,0}^j).\tag{A.21}$$

Dato che nel passo di correzione locale $\Omega_{k|k,0}^j = \Omega_{k|k-1}^j + \mathcal{I}^j$, si ha anche

$$\Omega_{k|k,0}^j \mathbf{x}_k = \Omega_{k|k-1}^j \mathbf{x}_k + \mathcal{I}^j \mathbf{x}_k. \quad (\text{A.22})$$

Allora, sfruttando (A.16) e (A.22) in (A.21), l'errore di stima può essere riscritto come

$$\tilde{\mathbf{x}}_{k|k}^i = (\Omega_{k|k}^i)^{-1} \sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{k|k-1}^j \tilde{\mathbf{x}}_{k|k-1}^j.$$

Sfruttando l'identità precedente, (A.15) può essere riscritto come

$$\begin{aligned} \mathcal{L}_{k+1}^i(\tilde{\mathbf{x}}_{k+1|k}^i) &\leq \\ &\tilde{\beta} \left[\left(\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{k|k-1}^j \tilde{\mathbf{x}}_{k|k-1}^j \right)^\top (\Omega_{k|k}^i)^{-1} \left(\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{k|k-1}^j \tilde{\mathbf{x}}_{k|k-1}^j \right) \right]. \end{aligned} \quad (\text{A.23})$$

Poiché $\Omega_{k|k}^i \geq \sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{k|k-1}^j$, (A.23) implica anche

$$\begin{aligned} \mathcal{L}_{k+1}^i(\tilde{\mathbf{x}}_{k+1|k}^i) &\leq \\ &\tilde{\beta} \left[\left(\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{k|k-1}^j \tilde{\mathbf{x}}_{k|k-1}^j \right)^\top \left(\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{k|k-1}^j \right)^{-1} \left(\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{k|k-1}^j \tilde{\mathbf{x}}_{k|k-1}^j \right) \right]. \end{aligned} \quad (\text{A.24})$$

Infine, applicando il Lemma 2 al membro destro dell'equazione (A.24), si può vedere che

$$\mathcal{L}_{k+1}^i(\tilde{\mathbf{x}}_{k+1|k}^i) \leq \tilde{\beta} \sum_{j \in \mathcal{N}} \pi_L^{i,j} (\tilde{\mathbf{x}}_{k|k-1}^j)^\top \Omega_{k|k-1}^j \tilde{\mathbf{x}}_{k|k-1}^j = \tilde{\beta} \sum_{j \in \mathcal{N}} \pi_L^{i,j} \mathcal{L}_k^j(\tilde{\mathbf{x}}_{k|k-1}^j).$$

□

Dimostrazione del Teorema 6:

Per ricavare il risultato di stabilità suddetto, un primo passo importante è lo studio delle proprietà delle matrici d'informazione $\Omega_{t|t}^i$.

Lemma 3. *Si assume la validità delle stesse ipotesi del Teorema 6. Allora, esistono matrici definite positive $\underline{\Omega}$, $\bar{\Omega}$, $\underline{\Omega}^+$, e $\bar{\Omega}^+$ tali che $0 < \underline{\Omega} \leq \Omega_{t|t}^i \leq \bar{\Omega}$ e $0 < \underline{\Omega}^+ \leq \Omega_{t+1|t}^i \leq \bar{\Omega}^+$ per ogni $i \in \mathcal{N}$ e $t \geq 1$.*

Dimostrazione: Concentriamoci anzitutto sulle matrici $\Omega_{t|t}^i$. Ricordando l'identità nell'ultima riga della tab. 3.1 e il fatto che $\delta \Omega_t^j = (\mathbf{C}^j)^T \mathbf{V}^j \mathbf{C}^j$, si può vedere che l'ipotesi A4 implica che

$$\Omega_{t|t}^i \leq \sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t|t-1}^i + \sum_{j \in \mathcal{S}} \pi_L^{i,j} \bar{\rho} (\mathbf{C}^j)^T \mathbf{V}^j \mathbf{C}^j, \quad (\text{A.25})$$

$$\Omega_{t|t}^i \geq \sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t|t-1}^i + \sum_{j \in \mathcal{S}} \pi_L^{i,j} \underline{\rho} (\mathbf{C}^j)^T \mathbf{V}^i \mathbf{C}^j. \quad (\text{A.26})$$

Si suppone ora che venga eseguito l'algoritmo CP con matrici di covarianza dei rumori \mathbf{R}^j rimpiazzata da $\mathbf{R}^j/\bar{\rho}$ e siano $\bar{\Omega}_{t|t}^i$ e $\bar{\Omega}_{t|t-1}^i$ le risultanti matrici d'informazioni. Analogamente, siano $\underline{\Omega}_{t|t}^i$ e $\underline{\Omega}_{t|t-1}^i$ le matrici d'informazione risultanti da un algoritmo CP con matrici \mathbf{R}^j rimpiazzate da $\mathbf{R}^j/\underline{\rho}$. Ricordando la prima riga della tab. 3.1, si può facilmente vedere che

$$\bar{\Omega}_{t|t}^i = \sum_{j \in \mathcal{N}} \pi_L^{i,j} \bar{\Omega}_{t|t-1}^i + \sum_{j \in \mathcal{S}} \pi_L^{i,j} \bar{\rho} (\mathbf{C}^j)^T \mathbf{V}^j \mathbf{C}^j, \quad (\text{A.27})$$

$$\underline{\Omega}_{t|t}^i = \sum_{j \in \mathcal{N}} \pi_L^{i,j} \underline{\Omega}_{t|t-1}^i + \sum_{j \in \mathcal{S}} \pi_L^{i,j} \underline{\rho} (\mathbf{C}^j)^T \mathbf{V}^j \mathbf{C}^j. \quad (\text{A.28})$$

Quindi, a condizione che si adottino le stesse inizializzazioni per tutti gli algoritmi, cioè $\bar{\Omega}_{1|0}^i = \underline{\Omega}_{1|0}^i = \Omega_{1|0}^i$, si può vedere con semplici considerazioni induttive che $\underline{\Omega}_{t|t}^i \leq \Omega_{t|t}^i \leq \bar{\Omega}_{t|t}^i$ per ogni $i \in \mathcal{N}$ e $t \geq 1$ (si ricordi anche che il passo di predizione del filtro di Kalman è monotono e non decrescente nel senso che $\underline{\Omega}_{t|t}^i \leq \Omega_{t|t}^i \leq \bar{\Omega}_{t|t}^i$ implica $\underline{\Omega}_{t+1|t}^i \leq \Omega_{t+1|t}^i \leq \bar{\Omega}_{t+1|t}^i$). Allora, la prima parte della dimostrazione si conclude osservando che, nelle ipotesi fatte, esistono due matrici definite positive $\underline{\Omega}$ e $\bar{\Omega}$ tali che $\underline{\Omega}_{t|t}^i \geq \underline{\Omega}$ e $\bar{\Omega}_{t|t}^i \leq \bar{\Omega}$ (in virtù del Teorema 2 di [17] e del fatto che $\Omega_{1|0}^i > 0$). Per quanto riguarda i limiti di $\Omega_{t+1|t}^i$, si può semplicemente prendere $\bar{\Omega}^+ = \mathbf{W}$ e $\underline{\Omega}^+ = (\mathbf{A}\bar{\Omega}^{-1}\mathbf{A}^T + \mathbf{Q})^{-1}$.

□

Il Lemma 3 garantisce che le matrici $\Omega_{t|t}^i$ siano non singolari e, quindi, che gli errori di stima \mathbf{e}_t^i siano ben definiti. Si può adesso stabilire il seguente risultato.

Proposizione 2. *Si suppone la validità delle stesse ipotesi del Teorema 6. Allora, gli errori di stima \mathbf{e}_t^i obbediscono alla ricorsione*

$$\mathbf{e}_{t+1}^i = \sum_{j \in \mathcal{N}} \Phi_t^{i,j} \mathbf{e}_t^j + \sum_{j \in \mathcal{N}} \Gamma_t^{i,j} \mathbf{v}_t^j + \mathbf{w}_t \quad (\text{A.29})$$

per ogni $i \in \mathcal{N}$, dove $\Phi_t^{i,j} = \mathbf{A} \left(\Omega_{t|t}^i \right)^{-1} \pi_L^{i,j} \Omega_{t|t-1}^j$ e $\Gamma_t^{i,j} = \mathbf{A} \left(\Omega_{t|t}^i \right)^{-1} \pi_L^{i,j} \omega_t^i \mathbf{V}_t^j$.

Dimostrazione: Si nota anzitutto che $\mathbf{e}_{t+1}^i = \mathbf{A}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i) + \mathbf{w}_t$. Inoltre, la stima $\hat{\mathbf{x}}_{t|t}^i$ può essere espressa come

$$\hat{\mathbf{x}}_{t|t}^i = \left(\Omega_{t|t}^i \right)^{-1} \left[\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t|t-1}^j \hat{\mathbf{x}}_{t|t-1}^j + \omega_t^i \sum_{j \in \mathcal{S}} \pi_L^{i,j} (\mathbf{C}^j)^T \mathbf{V}^j \mathbf{y}_t^j \right] \quad \text{con} \quad \mathbf{y}_t^j = \mathbf{C}^j \mathbf{x}_t + \mathbf{v}_t^j$$

(si ricordi di nuovo l'ultima riga della tab. 3.1). Allora, l'equazione (A.29) può essere derivata direttamente osservando che vale l'identità

$$\mathbf{x}_t = \left(\Omega_{t|t}^i \right)^{-1} \left[\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t|t-1}^j \mathbf{x}_t + \omega_t^i \sum_{j \in \mathcal{S}} \pi_L^{i,j} (\mathbf{C}^j)^T \mathbf{V}^j \mathbf{C}^j \mathbf{x}_t \right].$$

□

Si consideri ora la dinamica collettiva degli errori di stima in assenza di rumore, cioè

$$\mathbf{e}_{t+1} = \Phi_t \mathbf{e}_t \quad (\text{A.30})$$

dove $\mathbf{e}_t = \text{col}(\mathbf{e}_t^i, i \in \mathcal{N})$ e Φ_t è una matrice a blocchi i cui blocchi sono dati dalle matrici $\Phi_t^{i,j}$ definite nella Proposizione 2. Allora, vale il seguente risultato.

Lemma 4. *Si supponga la validità delle stesse ipotesi del Teorema 6. Allora il sistema tempo variante (A.30) è uniformemente esponenzialmente stabile.*

Dimostrazione: Sia \mathbf{p} l'autovettore sinistro di Perron-Frobenius della matrice Π^L . Si noti che, in virtù dell'ipotesi A3, tale autovettore ha componenti strettamente positive $p^i, i \in \mathcal{N}$, e soddisfa l'equazione $\mathbf{p}^T \Pi^L = \mathbf{p}^T$, cioè $\sum_{j \in \mathcal{N}} p^j \pi_L^{j,i} = p^i$. Si consideri ora la funzione di Lyapunov

$$\mathcal{V}_t(\mathbf{e}_t) = \sum_{i \in \mathcal{N}} p^i (\mathbf{e}_t^i)^T \Omega_{t|t-1}^i \mathbf{e}_t^i$$

candidata per il sistema (A.30). In considerazione del Lemma 3, è immediato vedere che esistono costanti strettamente positive α_1 e α_2 tali che $\alpha_1 \|\mathbf{e}_t\|^2 \leq \mathcal{V}_t(\mathbf{e}_t) \leq \alpha_2 \|\mathbf{e}_t\|^2$. Inoltre, sfruttando il fatto che $\Omega_{t+1|t}^i \leq \tilde{\beta} \mathbf{A}^{-T} \Omega_{t|t}^i \mathbf{A}^{-1}$ per qualche reale positivo $\tilde{\beta} < 1$ (vedi il punto iii) nel Lemma 1 di [34]), si trova che

$$\begin{aligned} (\mathbf{e}_{t+1}^i)^T \Omega_{t+1|t}^i \mathbf{e}_{t+1}^i &= \left(\sum_{j \in \mathcal{N}} \Phi_t^{i,j} \mathbf{e}_t^j \right)^T \Omega_{t+1|t}^i \sum_{j \in \mathcal{N}} \Phi_t^{i,j} \mathbf{e}_t^j \\ &\leq \tilde{\beta} \left(\sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t|t-1}^j \mathbf{e}_t^j \right)^T (\Omega_{t|t}^i)^{-1} \sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t|t-1}^j \mathbf{e}_t^j \\ &\leq \tilde{\beta} \sum_{j \in \mathcal{N}} \pi_L^{i,j} (\mathbf{e}_t^j)^T \Omega_{t|t-1}^j \mathbf{e}_t^j \end{aligned}$$

dove l'ultima disuguaglianza segue dal fatto che $\Omega_{t|t}^i \geq \sum_{j \in \mathcal{N}} \pi_L^{i,j} \Omega_{t|t-1}^j$ e dal Lemma 2 di [34]. Ne segue che è possibile scrivere

$$\begin{aligned} \mathcal{V}_{t+1}(\mathbf{e}_{t+1}) &= \sum_{i \in \mathcal{N}} p^i (\mathbf{e}_{t+1}^i)^T \Omega_{t+1|t}^i \mathbf{e}_{t+1}^i \\ &\leq \tilde{\beta} \sum_{i,j \in \mathcal{N}} p^i \pi_L^{i,j} (\mathbf{e}_t^j)^T \Omega_{t|t-1}^j \mathbf{e}_t^j \\ &= \tilde{\beta} \sum_{j \in \mathcal{N}} p^j (\mathbf{e}_t^j)^T \Omega_{t|t-1}^j \mathbf{e}_t^j = \tilde{\beta} \mathcal{V}_t(\mathbf{e}_t) \end{aligned}$$

da cui segue immediatamente la stabilità esponenziale uniforme di (A.30).

□

Quindi, il Teorema 6 segue immediatamente dal Lemma 4 e dal fatto che i termini $\sum_{j \in \mathcal{N}} \mathbf{\Gamma}_t^{i,j} \mathbf{v}_t^j + \mathbf{w}_t$ in (A.29) sono limitati in media quadratica in virtù del Lemma 3.

□

Appendice B

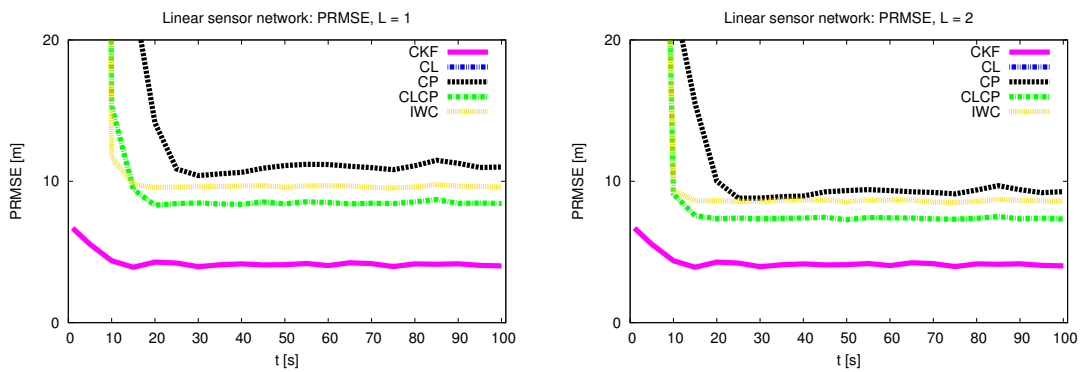


Figura B.1: Rete lineare: PRMSE dei filtri considerati per $L = 1$ (a sinistra) e $L = 2$ (a destra).

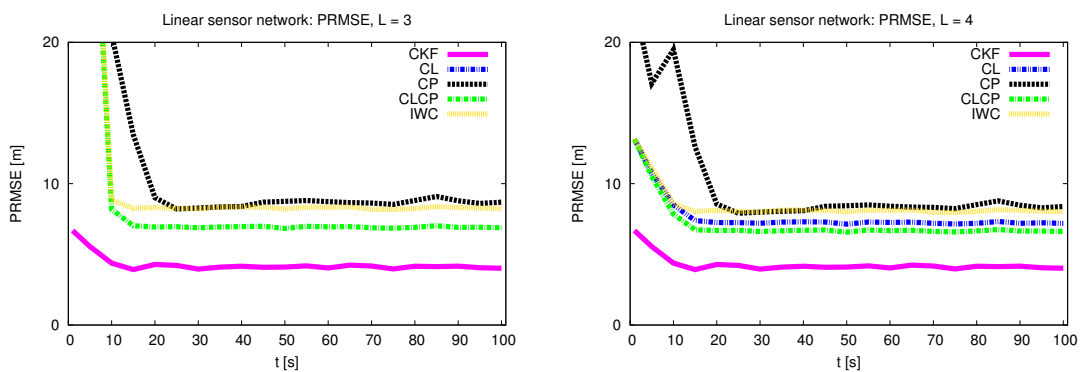


Figura B.2: Rete lineare: PRMSE dei filtri considerati per $L = 3$ (a sinistra) e $L = 4$ (a destra).

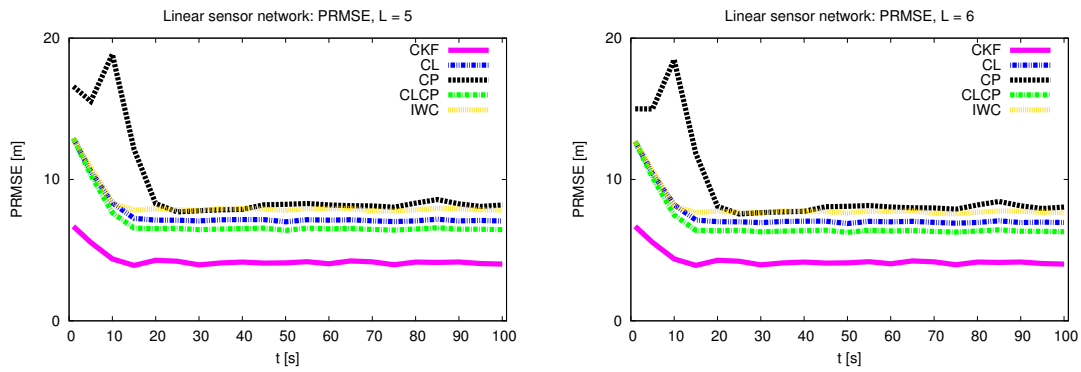


Figura B.3: Rete lineare: PRMSE dei filtri considerati per $L = 5$ (a sinistra) e $L = 6$ (a destra).

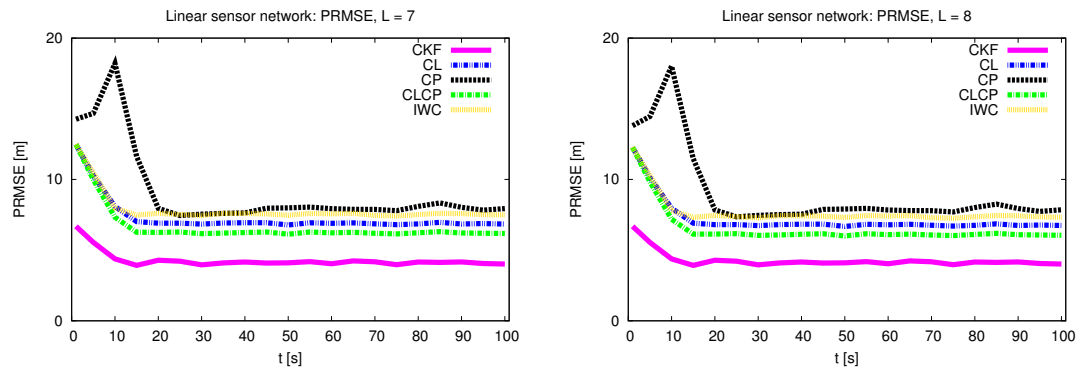


Figura B.4: Rete lineare: PRMSE dei filtri considerati per $L = 7$ (a sinistra) e $L = 8$ (a destra).

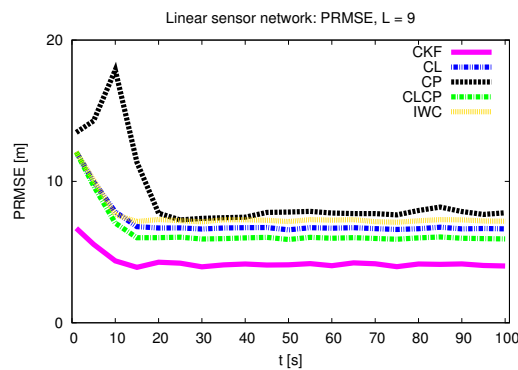


Figura B.5: Rete lineare: PRMSE dei filtri considerati per $L = 9$.

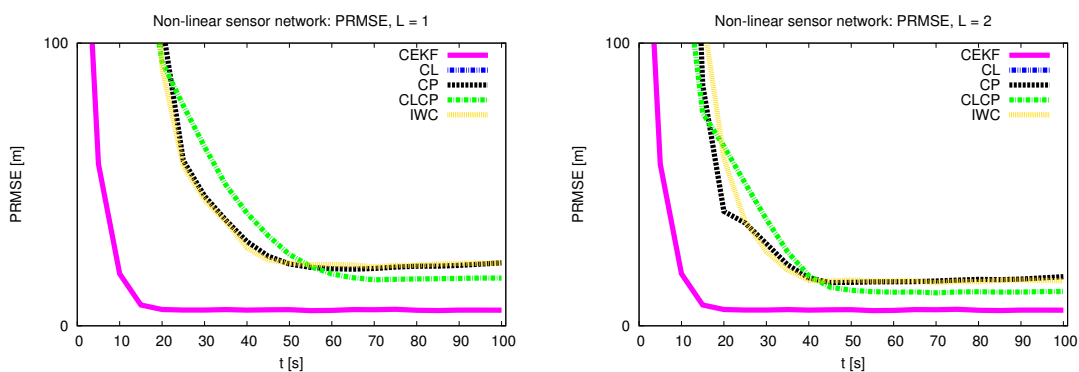


Figura B.6: Rete non-lineare: PRMSE dei filtri considerati per $L = 1$ (a sinistra) e $L = 2$ (a destra).

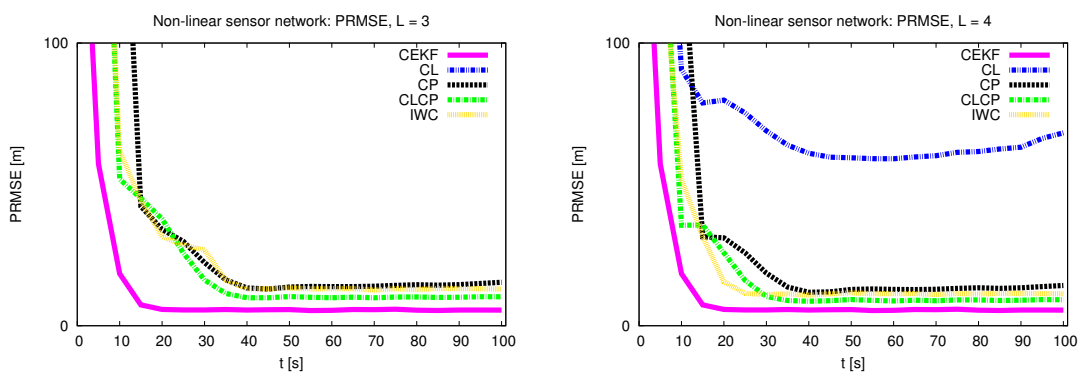


Figura B.7: Rete non-lineare: PRMSE dei filtri considerati per $L = 3$ (a sinistra) e $L = 4$ (a destra).

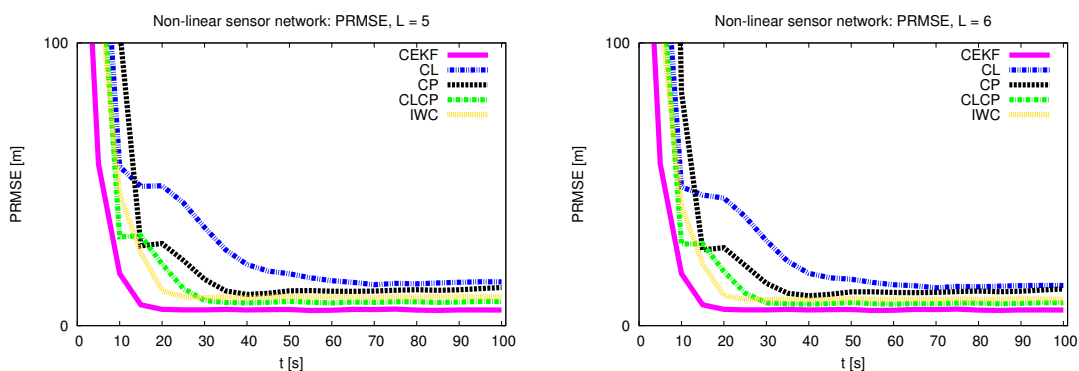


Figura B.8: Rete non-lineare: PRMSE dei filtri considerati per $L = 5$ (a sinistra) e $L = 6$ (a destra).

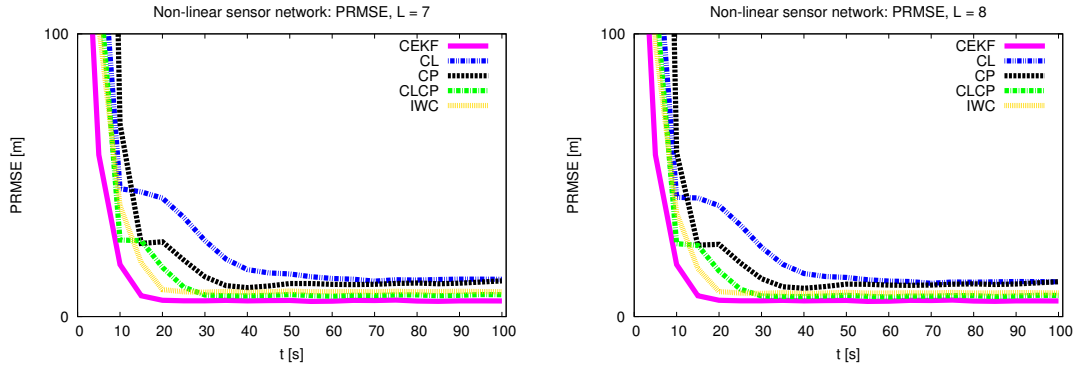


Figura B.9: Rete non-lineare: PRMSE dei filtri considerati per $L = 7$ (a sinistra) e $L = 8$ (a destra).

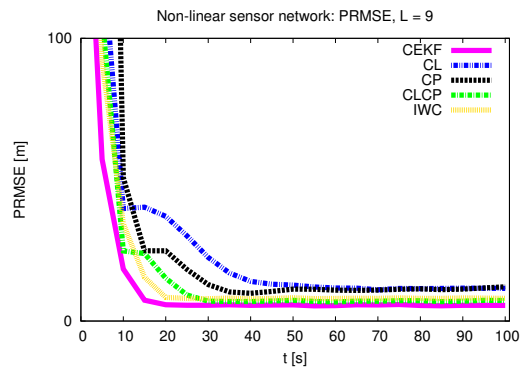


Figura B.10: Rete non-lineare: PRMSE dei filtri considerati per $L = 9$.

Appendice C

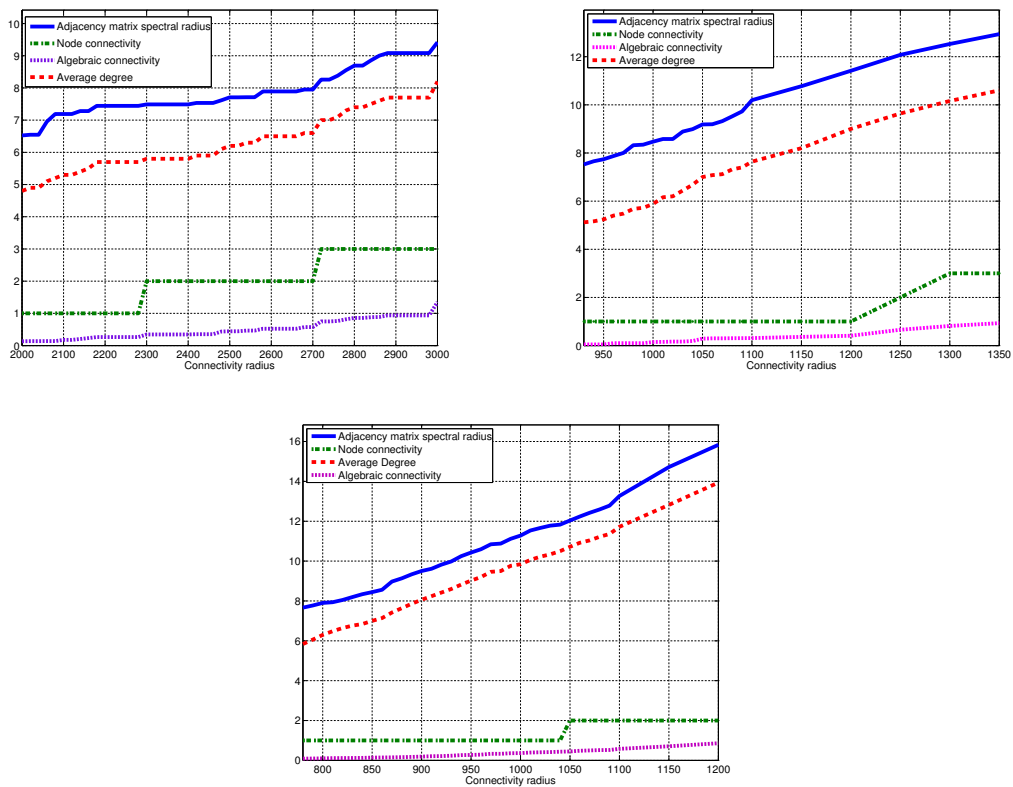


Figura C.1: Parametri di rete in funzione del raggio di connettività R_{max} per le reti di 20 nodi (in alto a sinistra), di 50 nodi (in alto a destra) e di 100 nodi (in basso).

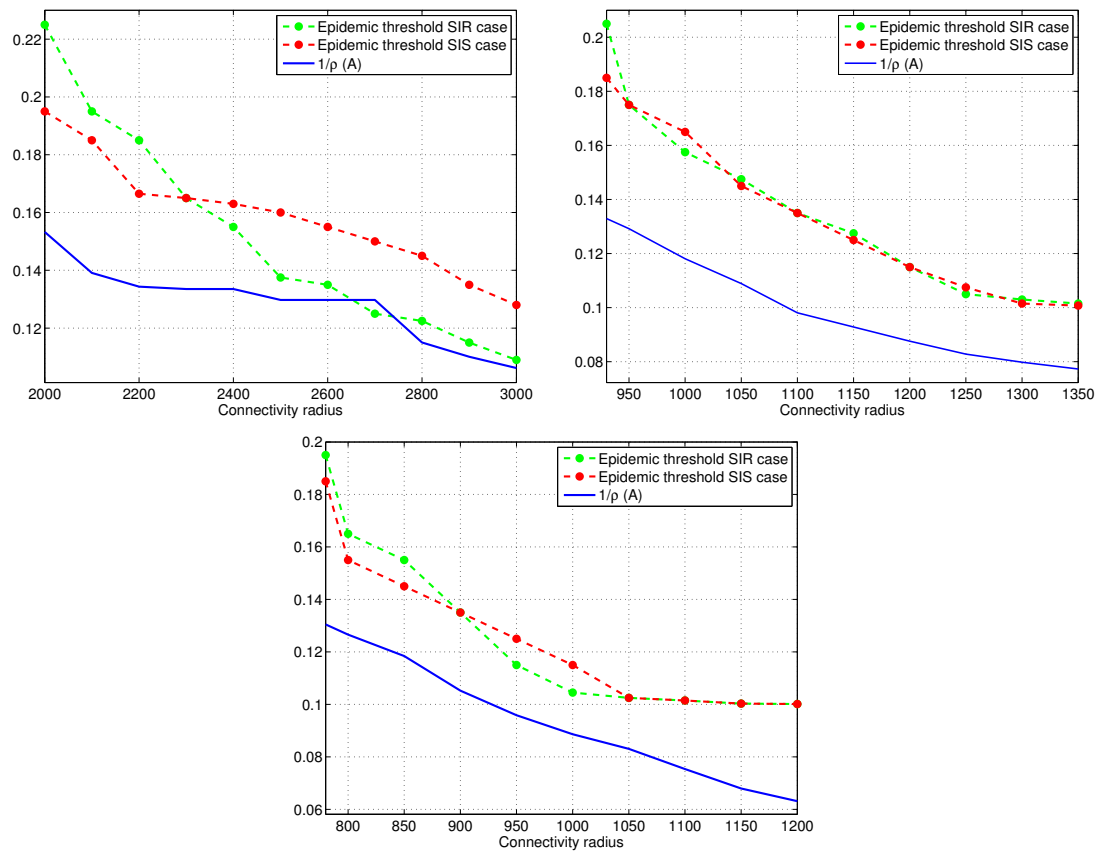


Figura C.2: Soglie epidemiche (teoriche ed empiriche) in funzione del raggio di connettività R_{max} per le reti di 20 (in alto a sinistra), di 50 nodi (in alto a destra) e di 100 nodi (in basso).

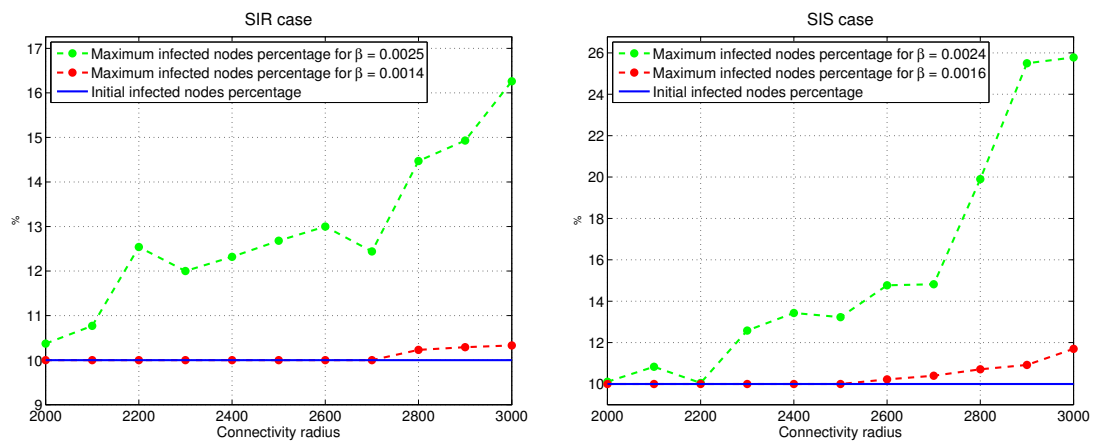


Figura C.3: Massima prevalenza epidemica in funzione del raggio di connettività R_{max} per la rete di 20 nodi, casi SIR (a sinistra) e SIS (a destra).

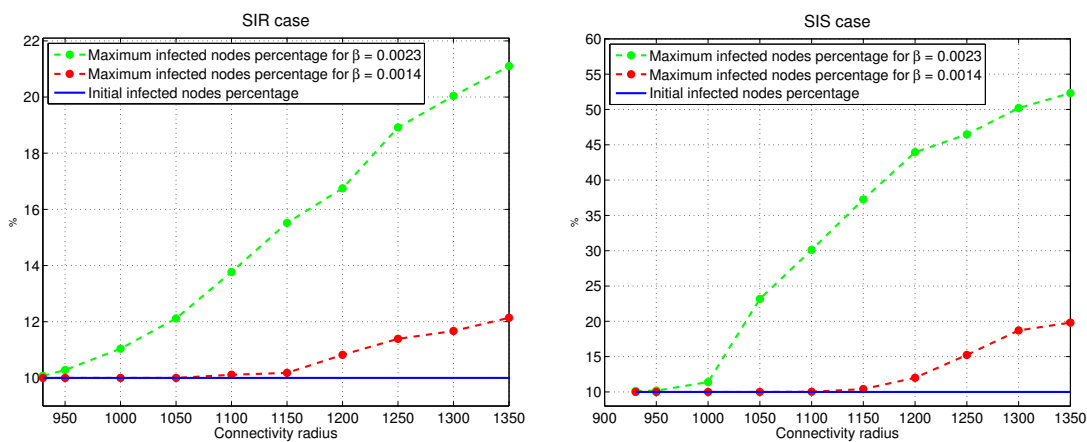


Figura C.4: Massima prevalenza epidemica in funzione del raggio di connettività R_{max} per la rete di 50 nodi, casi SIR (a sinistra) e SIS (a destra).

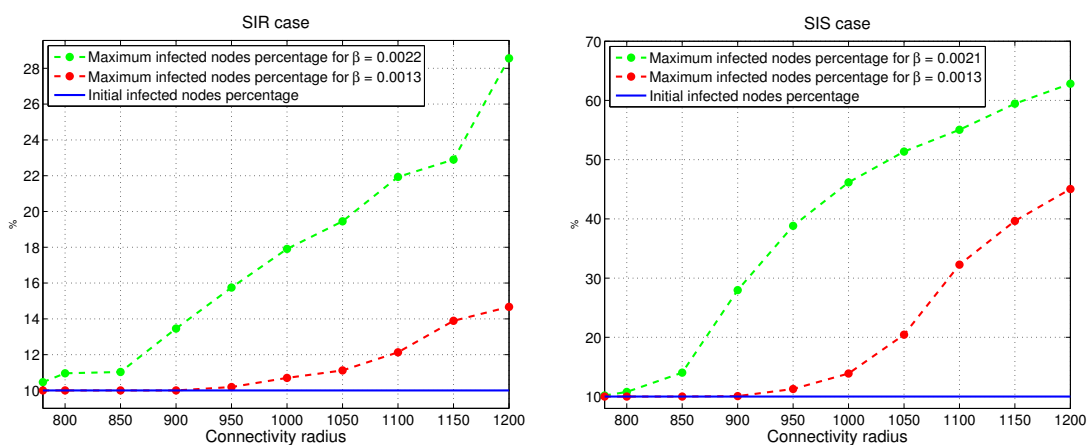


Figura C.5: Massima prevalenza epidemica in funzione del raggio di connettività R_{max} per la rete di 100 nodi, casi SIR (a sinistra) e SIS (a destra).

Appendice D

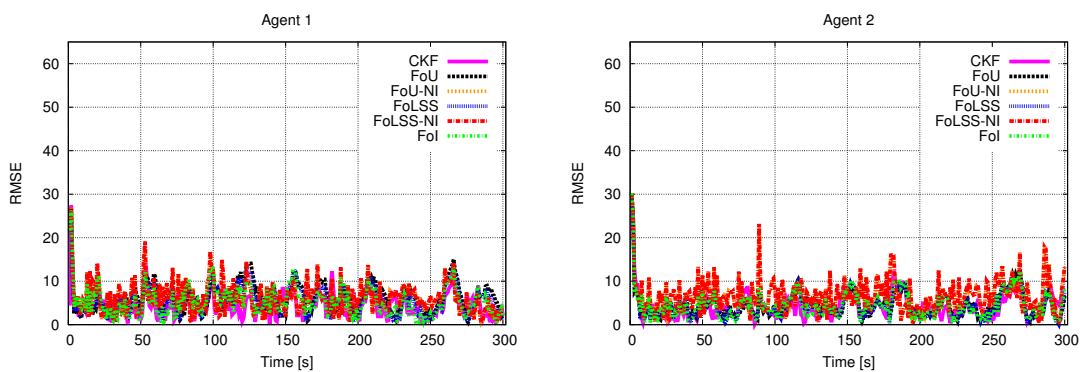


Figura D.1: Sistema lineare, modello 1: RMSE degli agenti 1 (a sinistra) e 2 (a destra) per il filtro centralizzato ed i metodi di fusione considerati.

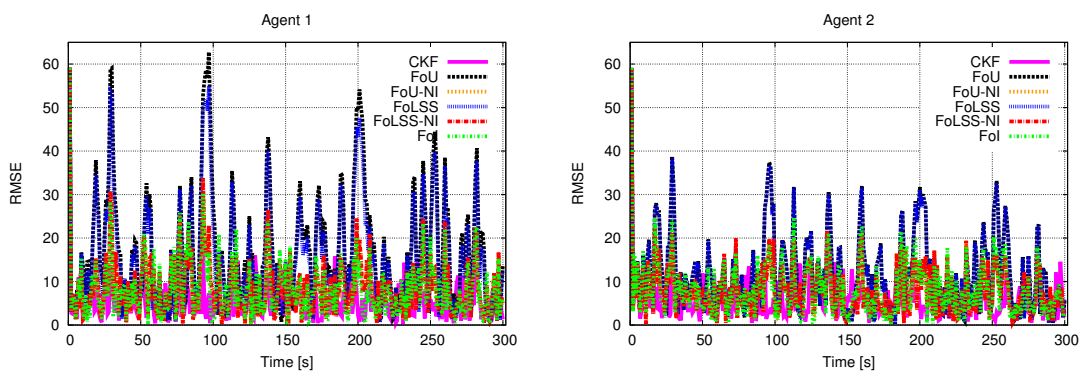


Figura D.2: Sistema lineare, modello 2: RMSE degli agenti 1 (a sinistra) e 2 (a destra) per il filtro centralizzato ed i metodi di fusione considerati.

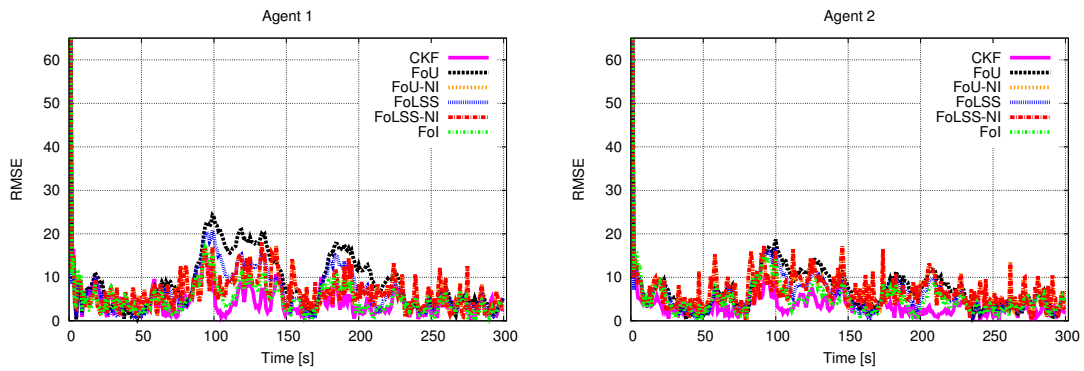


Figura D.3: Sistema lineare, modello 3: RMSE degli agenti 1 (a sinistra) e 2 (a destra) per il filtro centralizzato ed i metodi di fusione considerati.

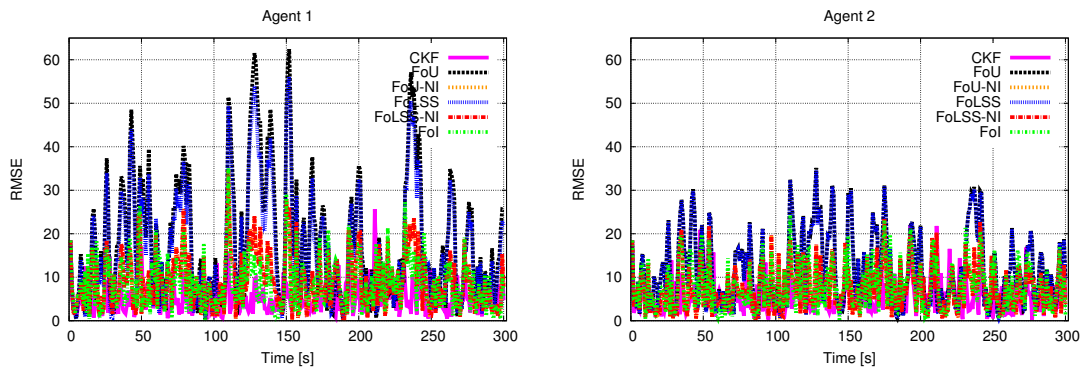


Figura D.4: Sistema lineare, modello 4: RMSE degli agenti 1 (a sinistra) e 2 (a destra) per il filtro centralizzato ed i metodi di fusione considerati.

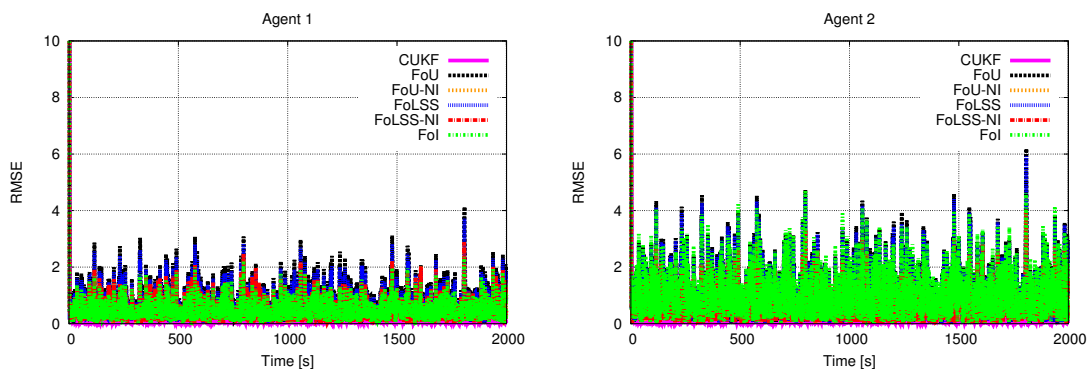


Figura D.5: Sistema non-lineare: RMSE degli agenti 1 (a sinistra) e 2 (a destra) per il filtro centralizzato ed i metodi di fusione considerati.

Bibliografia

- [1] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN)*, pp. 63–70, 2005.
- [2] R. Olfati-Saber, J. A. Fax, and R. M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 49–54, 2007.
- [3] S. Julier and J. Uhlmann, “A non-divergent estimation algorithm in the presence of unknown correlations,” *Proceedings of the American Control Conference*, vol. 4, pp. 2369–2373, 1997.
- [4] L. Chen, P. Arambel, and R. Mehra, “Estimation under unknown correlation: covariance intersection revisited,” *IEEE Transactions on Automatic Control*, vol. 47, no. 11, pp. 1879–1882, 2002.
- [5] R. Olfati-Saber, “Distributed Kalman filtering for sensor networks,” *Proceedings of the 46th IEEE Conference on Decision and Control*, pp. 5492–5498, 2007.
- [6] M. Kamgarpour and C. Tomlin, “Convergence properties of a decentralized Kalman filter,” *Proceedings of the 47th IEEE Conference on Decision and Control*, pp. 3205–3210, 2008.
- [7] P. Alriksson and A. Rantzer, “Distributed Kalman filtering using weighted averaging,” *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, 2006.
- [8] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, “Distributed Kalman filtering based on consensus strategies,” *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 622–633, 2008.
- [9] R. Olfati-Saber, “Kalman-consensus filter : Optimality, stability, and performance,” *Proceedings of the 48th IEEE Conference on Decision and Control, held jointly with the 28th Chinese Control Conference*, pp. 7036–7042, 2009.
- [10] S. S. Stankovic, M. S. Stankovic, and D. M. Stipanovic, “Consensus based overlapping decentralized estimation with missing observations and communication faults,” *Automatica*, vol. 45, no. 6, pp. 1397–1406, 2009.

- [11] M. Farina, G. Ferrari-Trecate, and R. Scattolini, "Distributed moving horizon estimation for linear constrained systems," *IEEE Transactions on Automatic Control*, vol. 55, pp. 2462–2475, 2010.
- [12] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069–2084, 2010.
- [13] F. Cattivelli and A. Sayed, "Distributed nonlinear Kalman filtering with applications to wireless localization," *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 3522–3525, 2010.
- [14] M. Farina, G. Ferrari-Trecate, and R. Scattolini, "Distributed moving horizon estimation for nonlinear constrained systems," *International Journal of Robust and Nonlinear Control*, vol. 22, no. 2, pp. 123–143, 2012.
- [15] U. Khan and A. Jadbabaie, "Networked estimation under information constraints," *submitted to IEEE Transactions on Automatic Control*, 2011. [Online]. Available: <http://arxiv.org/abs/1111.4580>
- [16] W. Li and Y. Jia, "Distributed consensus filtering for discrete-time nonlinear systems with non-Gaussian noise," *Signal Processing*, vol. 92, no. 10, pp. 2464–2470, 2012.
- [17] G. Battistelli, L. Chisci, S. Morrocchi, and F. Papi, "An information-theoretic approach to distributed state estimation," *Preprints of the 18th IFAC World Congress*, pp. 12 477–12 482, 2011.
- [18] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, and A. Graziano, "Consensus-based algorithms for distributed filtering," *Proc. 51st IEEE Conference on Decision and Control*, pp. 794–799, 2012.
- [19] G. Battistelli, L. Chisci, A. Farina, A. Graziano, and G. Mugnai, "Consensus-based linear and nonlinear filtering," *Submitted to IEEE Transactions on Automatic and Control*, 2013.
- [20] A. T. Kamal, J. A. Farrell, and A. K. Roy-Chowdhury, "Information weighted consensus," *Proceedings 51st IEEE Conference on Decision and Control*, pp. 2732–2737, 2012.
- [21] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, A. Graziano, and A. Liburdi, "Modeling and counteracting virus diffusion in sensor networks for net-centric surveillance systems," in *Effective Surveillance for Homeland Security*, F. Flammini, G. Franceschetti, and R. Setola, Eds. Chapman and Hall/CRC, 2013, ch. 14, pp. 355–386. [Online]. Available: <http://dx.doi.org/10.1201/b14839-17>
- [22] J. Kephart and S. White, "Directed-graph epidemiological models of computer viruses," *Proceedings IEEE Computer Society Symposium on Research in Security and Privacy*, vol. 3, pp. 343–359, 1991.

- [23] C. Zou, D. Towsley, and W. Gong, "Modeling and simulation study of the propagation and defense of internet e-mail worms," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, pp. 105–118, 2007.
- [24] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Transactions on Information and System Security*, vol. 10, pp. 1–26, 2008.
- [25] C. Castellano and R. Pastor-Satorras, "Thresholds for epidemic spreading in networks," *Physical Review Letters*, vol. 105, pp. 218 701–4, 2010.
- [26] A. Farina and F. Studer, *Radar data processing Vol. 1 - Introduction and Tracking*. Letchford: Research Studies Press LTD, 1985.
- [27] L. Servi and Y. Ho, "Recursive estimation in the presence of uniformly distributed measurement noise," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 563–565, 1981.
- [28] V. E. Benes, "Exact finite-dimensional filters for certain diffusions with non linear drift," *Stochastics*, vol. 5, no. 1, pp. 65–92, 1981.
- [29] F. E. Daum and J. Huang, "The curse of dimensionality for particle filters," *Proceedings of IEEE Conference on Aerospace*, vol. 4, pp. 1979–1993, 2003.
- [30] S. F. Schmidt, "Application of state-space methods to navigation problems," *Advanced Control Systems*, vol. 3, pp. 293–340, 1966.
- [31] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended kalman filter," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714–728, 1999.
- [32] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [33] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1520–1533, 2004.
- [34] G. Battistelli and L. Chisci, "Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, in print.
- [35] B. Mohar, "Eigenvalues, diameter and mean distance in graph," *Graph and Combinatorics*, vol. 7, pp. 53–64, 1991.
- [36] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, pp. 298–305, 1973.
- [37] A. Banerjee, X. Guo, and H. Wang, "On the optimality of conditional expectation as a Bregman predictor," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2664 –2669, 2005.

- [38] L. Campbell, “Equivalence of Gauss’s principle and minimum discrimination information estimation of probabilities,” *The Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 1011–1015, 1970.
- [39] H. Akaike, “Information theory and the extension of the maximum likelihood principle,” *Proceedings of the Second International Symposium on Information Theory*, pp. 267–281, 1973.
- [40] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.
- [41] M. Uney, D. E. Clark, and S. J. Julier, “Information measures in distributed multitarget tracking,” *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pp. 1–8, 2011.
- [42] R. P. S. Mahler, “Optimal/robust distributed data fusion: a unified approach,” *Proceedings of the SPIE Defense and Security Symposium*, 2000.
- [43] M. B. Hurley, “An information theoretic justification for covariance intersection and its generalization,” *Proceedings of the 5th International Conference on Information Fusion*, pp. 505–511, 2002.
- [44] S. J. Julier, T. Bailey, and J. K. Uhlmann, “Using exponential mixture models for suboptimal distributed data fusion,” *Proceedings of the IEEE Nonlinear Nonlinear Statistical Signal Processing Workshop (NSSPW’06)*, pp. 160–163, 2006.
- [45] S. J. Julier, “Fusion without independence,” *Proceedings of IET Seminar on Target Tracking and Data Fusion: Algorithms and Applications*, pp. 3–4, 2008.
- [46] G. C. Calafiore and F. Abrate, “Distributed linear estimation over sensor networks,” *International Journal of Control*, vol. 82, no. 5, pp. 868–882, 2009.
- [47] G. Battistelli, L. Chisci, and C. Fantacci, “Parallel consensus on likelihoods and priors for networked nonlinear filtering,” *submitted for publication*, 2013.
- [48] L. Chisci and G. Zappa, “Square-root Kalman filtering of descriptor systems,” *Systems & Control Letters*, vol. 19, no. 4, pp. 325–334, 1992.
- [49] A. T. Kamal, J. A. Farrell, and A. K. Roy-Chowdhury, “Information weighted consensus filters and their application in distributed camera networks,” *IEEE Transactions on Automatic Control*, print.
- [50] A. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [51] D. Pradip, L. Yonghe, and K. Sajal, “Modeling node compromise spread in wireless sensor networks using epidemic theory,” *IEEE Proceedings International Symposium on a World of Wireless, Mobile and Multimedia (WoWMoM’06)*, 2006.

- [52] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Physical Review Letters*, vol. 86, pp. 3200–3203, 2001.
- [53] M. Boguna, R. Pastor-Satorras, and A. Vespignani, “Epidemic spreading in complex networks with degree correlations,” *Statistical Mechanics of Complex Networks*, vol. 625, pp. 127–147, 2003.
- [54] P. Van Mieghem, J. Omic, and R. Kooij, “Virus spread in networks,” *IEEE/ACM Transactions on Networking*, vol. 17, pp. 1–14, 2009.
- [55] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and F. C., “Got the flu (or mumps)? check the eigenvalue!” *arXiv*, 1004.0060v1, 2010.
- [56] A. Ganesh, L. Massoulié, and D. Towsley, “The effect of network topology on the spread of epidemics,” *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pp. 1455–1466, 2005.
- [57] F. Pasqualetti, A. Bicchi, and F. Bullo, “Distributed intrusion detection for secure consensus computations,” *Proceedings of the IEEE Conference on Decision and Control*, pp. 5594–5599, 2007.
- [58] —, “Consensus computation in unreliable networks: a system theoretic approach,” *IEEE Transactions on Automatic Control*, vol. 57, pp. 90–104, 2012.
- [59] U. Khan and J. Moura, “Distributing the Kalman filter for large-scale systems,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4919–4935, 2008.
- [60] E. N. Lorentz, “Deterministic nonperiodic flow,” *Journal of the Atmospheric Sciences*, vol. 20, pp. 130–141, 1963.
- [61] W. Song and J. Liang, “Difference equation of Lorentz system,” *International Journal of Pure and Applied Mathematics*, vol. 83, pp. 101–110, 2013.
- [62] M. Uney, S. Julier, D. E. Clark, and B. Ristic, “Information measures in distributed multitarget tracking,” *Proceedings of Sensor Signal Processing for Defence (SSPD)*, 2010.
- [63] D. Smith and S. Singh, “Approaches to multisensor data fusion in target tracking: A survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 12, pp. 1696–1710, 2006.
- [64] K. C. Chang, C. Y. Chong, and S. Mori, “On scalable distributed sensor fusion,” *Proceedings of the 11th International Conference on Information Fusion*, pp. 1–8, 2008.
- [65] K. Åström and B. Bernhardsson, “Comparison of Riemann and Lebesgue sampling for first order stochastic systems,” *Proceedings of the 41th Conference on Decision and Control*, pp. 2011–2016, 2002.

- [66] G. Battistelli, A. Benavoli, and L. Chisci, "State estimation in a centralized sensor network under limited communication rate," *Proceedings of the European Control Conference*, 2009.
- [67] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 622–633, 2008.
- [68] L. M. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," *Proceedings IEEE Infocom*, pp. 1548–1557, 2001.
- [69] D. P. Spanos and R. M. Murray, "Distributed sensor fusion using dynamic consensus," *Proceedings of the 16th IFAC World Congress*, 2005.
- [70] J. K. Yook, D. M. Tilbury, and N. R. Soparkar, "Trading computation for bandwidth: Reducing communication in distributed control systems using state estimators," *IEEE Transactions on Control Systems Technology*, vol. 10, pp. 503–518, 2002.
- [71] X. Li and W. S. Wong, "State estimation with communication constraints," *Systems & Control Letters*, vol. 28, pp. 49–54, 1996.
- [72] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints - Part I: state estimation problem," *IEEE Transactions on Automatic Control*, vol. 42, pp. 1294–1299, 1997.
- [73] Y. Xu and J. P. Hespanha, "Estimation under uncontrolled and controlled communications in networked control systems," *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 842–847, 2005.
- [74] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of IEEE Special Issue on Technology of Networked Control Systems*, vol. 95, pp. 138–162, 2007.
- [75] V. Gupta, B. Hassibi, and M. Murray, "Optimal LQG control across packet-dropping links," *Systems & Control Letters*, vol. 56, pp. 439–446, 2007.
- [76] L. A. Montestruque and P. Antsaklis, "Stability of model-based networked control systems with time-varying transmission times," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1562–1572, 2004.
- [77] I. Bilinskis and A. Mikelsons, *Randomized Signal Processing*. Prentice Hall International, 1992.
- [78] V. Paxson, "End-to-end routing behavior in the internet," *ACM SIGCOMM Computer Communication Review*, vol. 26, pp. 25–38, 1996.
- [79] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1453–1464, 2004.

- [80] G. Battistelli, A. Benavoli, and L. Chisci, “State estimation with a remote sensor under limited communication rate,” *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing*, pp. 654–659, 2008.
- [81] X. R. Li, Y. M. Zhu, J. Wang, and C. Z. Han, “Optimal linear estimation fusion-part I: unified fusion rules,” *IEEE Transactions on Information Theory*, vol. 49, pp. 2192–2208, 2003.
- [82] Y. S. Suh, V. H. Nguyen, and Y. S. Ro, “Modified Kalman filter for networked monitoring systems employing a send-on-delta method,” *Automatica*, vol. 43, pp. 332–338, 2007.
- [83] P. A. Absil and R. Sepulchre, “Continuous dynamical systems that realize discrete optimization on the hypercube,” *Systems & Control Letters*, vol. 52, pp. 297–304, 2004.