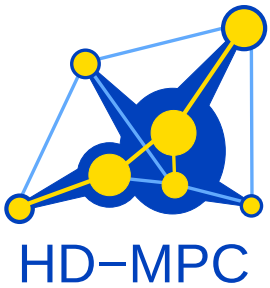
SEVENTH FRAMEWORK PROGRAMME
THEME – ICT
[Information and Communication Technologies]

Contract Number:	223854
Project Title:	Hierarchical and Distributed Model Predictive Control of Large-Scale Systems
Project Acronym:	HD-MPC



Deliverable Number:	D5.2
Deliverable Type:	Report
Contractual Date of Delivery:	March 1, 2011
Actual Date of Delivery:	**March 1, 2011**
Title of Deliverable:	**Intermediate report on new methods for distributed state and covariance estimation for large-scale interconnected systems**
Dissemination level:	Public
Workpackage contributing to the Deliverable:	WP5
WP Leader:	Riccardo Scattolini
Partners:	KUL, POLIMI, UNC, UWM
Author(s):	M. Farina, J. Espinosa, J. Garcia, R. Scattolini

Table of contents

Project co-ordinator

Name: Bart De Schutter
Address: Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 Delft, The Netherlands
Phone Number: +31-15-2785113
Fax Number: +31-15-2786679
E-mail: b.deschutter@tudelft.nl
Project web site: `http://www.ict-hd-mpc.eu`

Executive Summary

This report describes the research activity in the Seventh Framework Programme, Theme 3 "Information and Communication Technologies", STREP research project **Hierarchical and Distributed Model Predictive Control of Large Scale Systems- HD-MPC**, focusing on WP5 - "Distributed state estimation algorithms". Specifically, the report aims at presenting the main results achieved in Task 5.1 (State estimation) and Task 5.2 (Variance estimation).
The report is organized in four chapters:

- Chapter 1 presents a synopsis of the report, summarizes the content of the following chapters and, for each one of them, highlights the main results achieved.
- Chapter 2 first introduces the problem of distributed state estimation, i.e. the problem of estimating the state of the system by means of a network of sensors which can exchange information according to a given topology. Then, the problem is formally stated and a solution based on the use of Moving Horizon Estimators is proposed. The properties of the approach, in terms of convergence of the state estimates, are presented and a simulation example is shown to illustrate the potentials of the method.
- Chapter 3 describes some extensions of the distributed MHE algorithm presented in the previous chapter. Specifically, it is highlighted how the performance of the state estimation scheme depends upon various observability properties of the system, and the main convergence results are extended to consider different communication protocols. A discussion on how these protocols impact on the quality of the estimates is finally reported.
- Chapter 4 extends some results already sketched in Deliverable D5.1 concerning the analysis of the methods reported in the literature for the estimation of the noise variances affecting the system. The methods are critically compared and tested in a number of significant simulation examples. Some ideas for further improvements are also reported and discussed.

Chapter 1

Synopsis of the report

1.1 Synopsis of Chapter 2

In Chapter 2, the problem of distributed state estimation is formulated and solved by resorting to the Moving Horizon Estimation (MHE) technique. In order to properly define the problem, consider sensor networks composed by a set of electronic devices, with sensing and computational capabilities, which coordinate their activity through a communication network. Sensor networks can be employed in a wide range of applications, such as monitoring, exploration, surveillance or to track targets over specific regions; their diffusion is partly due to the recent developments in wireless communications and to the availability of low cost devices. Despite the recent developments in this field, many challenging problems have still to be tackled in order to fully exploit the potentialities of sensor networks. Among the open problems, their use for distributed state estimation is of paramount importance.
The problem can be described as follows. Assume that each sensor of the network measures some variables, computes a local estimate of the overall state of the system under monitoring, and transmits to the sensors connected to it both the measured values and the computed state estimate. Then, the main challenge is to provide a methodology which guarantees that all the sensors asymptotically reach a common reliable estimate of the state variables, i.e. the local estimates reach a *consensus*. This goal must be achieved even if the measurements performed by any sensor are not sufficient to guarantee observability of the process state (i.e., *local observability*), provided that all the sensors, if put together, guarantee such property (i.e., *collective observability*). The transmission of measurements and of estimates among the sensors must lead to the twofold advantage of enhancing the property of observability of the sensors and of reducing the uncertainty of state estimates computed by each node, respectively. In the literature, many consensus algorithms for distributed state estimation based on Kalman filters have recently been proposed; however, stability has not been proved for the discrete-time versions of these algorithms and optimality of the estimates has not been addressed. In general, the issue of distributed sensor fusion has been widely studied in the past years, while other studies focused on the design of decentralized Kalman filters based on system decomposition. Different solutions can be classified according to the model used by each subsystem for state-estimation purposes and the topology of the communication network among subsystems. However, none of the proposed solutions can handle constraints either on the state variables or on the disturbances affecting the state dynamics.
In Chapter 2 a novel MHE distributed algorithm is proposed. This approach has many advantages; first of all, the observer displays optimality properties, since a suitable minimization problem must be solved on-line at each time instant. Furthermore, it is proven that, under weak observability con-

ditions, convergence of the state estimates is guaranteed in a deterministic framework. Finally, constraints on the noise and on the state are taken into account, as it is common in receding horizon approaches in control and estimation.

The main assumptions of the developed method are the following:

- the system under investigation is described by a discrete time linear model affected by disturbances acting on the state and on the output measurements;
- the available M sensing nodes are connected by a directed graph with known structure and can exchange information (measurements and local state estimates) according to this graph structure.

Accordingly, it is possible to distinguish between *local*, *regional* and *collective* quantities. Specifically, for any node, a quantity (for example a set of measurements) is:

- local, if related to the node solely;
- regional, if referred to the node and its neighborhoods;
- collective, if referred to the whole network.

This leads to different definitions of observability, namely the system is:

- locally observable by any node if it is observable with respect to its local measurements;
- regionally observable by any node if it is observable with respect to its regional measurements;
- collectively observable if it is observable with respect to all its measurements.

The DMHE (Distributed Moving Horizon Estimation) algorithm, presented in details in Chapter 2, can be sketched at follows. At each time instant, every node:

- receives from its neighbors their measurements, state estimates and the corresponding covariances;
- computes a weighted average estimate of the state and of the corresponding covariance according to the graph topology;
- determines the new estimate of the state and of the state disturbance over a sliding window according to the Moving Horizon Estimation approach.

The proposed approach guarantees convergence of the DMHE observer under weak assumptions, in particular a given matrix (function of the system and of the algorithm parameters) is required to be Schur. This matrix collectively identifies the dynamics of the regionally unobservable modes of the sensors' estimation errors and it is easily computable. A necessary condition on the transmission graph guaranteeing that the eigenvalues of this matrix can be assigned at will is also reported. More specifically, each isolated strongly connected subgraph of the network (where the term isolated denotes the fact that no node of such a subnetwork has neighbors belonging to other subnetworks) must be collectively observable. Finally, conditions guaranteeing that the covariances of the computed estimates remain bounded are given.

The DMHE algorithm can be coupled with the state feedback distributed Model Predictive Control algorithms developed in the other workpackages, and in particular in WP3, to produce output feedback methods to be used in most practical cases, where the knowledge of the state is usually unavailable.

The chapter is structured as follows. After an initial review of the relevant literature, in Section 2.2 the observed dynamical system is introduced together with the structure of the sensor network, and the observability properties are defined. In Section 2.3 the distributed state estimation algorithm is described in detail. In Section 2.4 the convergence properties of the algorithm are investigated, and in Section 2.5 it is discussed how to select the design parameters in order to guarantee the applicability of the main results. Finally, in Section 2.6 a simulation example is presented. For the sake of clarity, the proofs are reported in Section 2.7.

1.2 Synopsis of Chapter 3

In Chapter 3 the results presented in Chapter 2 are extended to incorporate a number of significant practical considerations. Specifically, the main results of Chapter 2 are derived by considering different communication protocols, and it is discussed how these protocols impact on the quality of the estimates.
Assuming that the measurements taken by a sensor at time t are instantaneously transmitted to its adjacent (with reference to the network topology) neighboring agents, and letting $N_T \geq 1$ be the number of transmissions between two sensors within a sampling interval, two types of data communication protocols are considered:

P$_1$) For $N_T \geq 1$, at time t each sensor collects the sets of measurements taken from all other sensors that are connected to it by a path of length less than or equal to N_T.

P$_2$) For $N_T = 1$ and a sliding window $N \geq 1$ used in the DMHE algorithm, at time t each sensor i collects the sets of measurement from the other sensors which can transmit data to it within a frame of N steps, i.e. sensors that are connected by a path of length less than or equal to N.

Protocol P$_2$ is schematically illustrated in Figure 1.1. Note also that the protocols can be combined to obtain a more complex information transmission scheme. However, for simplicity, in Chapter 3 the two cases P$_1$ and P$_2$ have been addressed independently. The main results shown in this chapter can be summarized as follows.

A) In the case of communication protocol P$_1$ the fundamental property of regional observability can be enhanced by increasing the number N_T of data transmissions between agents within a sampling interval. This produces two accompanying effects.

   1) If all the isolated strongly connected subgraphs are collectively observable, there exists a threshold value for N_T (say $\bar{N}_t$) such that regional observability is satisfied by all the sensors for $N_T \geq \bar{N}_T$.

   2) It enhances the stability properties of the estimation algorithm (in terms of eigenvalues location of a properly defined dynamic matrix).

B) As for protocol P$_2$, it is shown how to enhance regional observability and the stability property of the estimation algorithm by increasing the estimation horizon N.

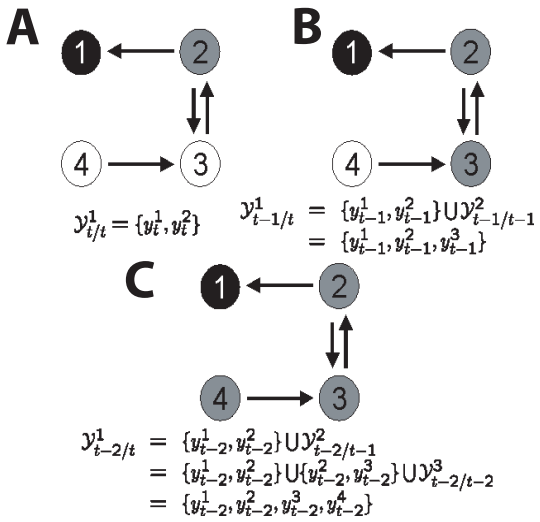


Figure 1.1: Illustration of the communication protocol P_2 for $N = 2$.

The chapter is structured as follows. In Section 3.1 the definition of different communication protocols is introduced and an analysis on how they affect the definition of regional observability is carried out. In Section 3.2 we highlight the main implications on the distributed state estimation algorithm when different transmission protocols are introduced. Finally, in Section 3.3, it is shown how the parameters of the communication protocols can be properly tuned, so as to enhance the performance of the estimation scheme.

1.3 Synopsis of Chapter 4

It is well known that, in Kalman filtering, the knowledge of the covariances Q_w and R_v of the noises affecting the state equation and the output measurements, respectively, is strictly required. In fact, the performance of the estimator can significantly deteriorate when the algorithm is fed with wrong values of these matrices. However, in many practical cases, Q_w and R_v are not a-priori known and quite often they are used as tuning parameters. The same knowledge is required in the MHE approach, i.e. the one used in the developments described in the previous chapters, where Q_w and R_v are weighting terms to be used in the performance index to be minimized, together with the (time-varying) matrix Π weighting the state error at the beginning of the considered time window.
For these reasons, there is practical interest in the estimation of the noise variances, and for the solution of this problem many methods have been proposed in the literature (see also the preliminary results reported in Deliverable D5.1). Among them, two appear to be the most interesting and promising ones. The first, which can be viewed as a seminal contribution in the field, is due to Mehra and traces back to the early '70s (see [32, 33]), while the second one has been recently developed (see [35], [1]). In order to compare their performance, these methods have been implemented and tested in a number of significant cases.

We denote by N_t the number of data sets used to assess the mean and variance of each covariance estimation method, and by $\hat{Q}_i$, and $\hat{R}_i$ the estimates of Q_w, and R_v respectively, obtained with the data generated in test i. In order to test the quality of the covariance matrix estimation performances, the following indices are used:

- The Root Mean Square (RMS) error,

$$RMS_q = \sqrt[2]{\frac{1}{N_t}\sum_{i=1}^{N_t}\|Q_w - \hat{Q}_i\|_2^2}, \qquad RMS_r = \sqrt[2]{\frac{1}{N_t}\sum_{i=1}^{N_t}\|R_v - \hat{R}_i\|_2^2}$$

- The mean of the ∞-norm of errors (MIE),

$$MIE_q = \frac{1}{N_t}\sum_{i=1}^{N_t}\|Q_w - \hat{Q}_i\|_\infty, \qquad MIE_r = \frac{1}{N_t}\sum_{i=1}^{N_t}\|R_v - \hat{R}_i\|_\infty$$

Three different test cases are studied in Chapter 4, two of which have been taken from the literature, while the third corresponds to the linearized model of a single reach of the hydro-power valley, which is one of the main benchmarks of the HD-MPC project. In all the cases, the results achieved have shown that the more recent methods proposed in [35] and [1] outperform the basic Mehra's algorithm ([32, 33]). However, all these algorithms require a very large data set to achieve reasonable performance, so that the problem of properly tuning the MHE (and Kalman filters) remains a critical practical problem.

In order to have a quick overview of the main results, one of the considered test cases refers to a linear discrete time model with state $x \in R^5$, disturbance $w \in R^3$, and output $y \in R^2$. Denoting by N_d the number of samples in the data set used for the estimation of the covariance matrices, the results obtained are summarized in the following tables:

Table 1.1: Mehra's algorithm

N_d	$\|RMS_q\|$	$\|RMS_r\|$	$\|MIE_q\|$	$\|MIE_r\|$
10^3	3.8078	4.2695	3.0682	3.4597
10^4	1.0694	1.2339	0.8821	1.0353
10^5	0.3597	0.4002	0.2981	0.3297

Table 1.2: ALS algorithm ([1])

N_d	$\|RMS_q\|$	$\|RMS_r\|$	$\|MIE_q\|$	$\|MIE_r\|$
10^3	1.6399	1.0288	1.3323	0.8369
10^4	0.7965	0.4961	0.6546	0.4092
10^5	0.3118	0.1942	0.2464	0.1533

It is apparent that acceptable performances can be obtained only with a very large data set, which turns out to be a very restrictive condition for a practical implementation of the methods. For this reason, it is believed that different tuning knobs that are easy to use would be required to make the MHE algorithms practically useful in industrial applications.

Chapter 2

Distributed Moving Horizon Estimation for linear systems

In this chapter, the problem of distributed state estimation is formally posed and a solution is proposed with the MHE approach. The content of this chapter is based on the paper [20].

2.1 Literature review

The problem of distributed state estimation with sensor networks has been widely discussed in the literature. Consensus algorithms for distributed state estimation based on Kalman filters have recently been proposed in [16, 6, 39, 37, 50, 38, 29]. In particular, in [39, 37, 50], *consensus on measurements* is used to reduce their uncertainty and Kalman filters are applied by each agent. In [38], three algorithms for distributed filtering are proposed. The first algorithm is similar to the one described in [37], save for the fact that sensors exploit only partial measurements of the state vector. The second approach relies on communicating the state estimates among neighboring agents (*consensus on estimates*). The third algorithm, named *iterative Kalman consensus filter*, is based on the discrete-time version of a continuous-time Kalman filter plus a *consensus step* on the state estimates, which is proved to be stable. However, stability has not been proved for the discrete-time version of the algorithm and optimality of the estimates has not been addressed. Recently, convergence in mean of the local state estimates obtained with the algorithm presented in [37] has been proved in [29], provided that the observed process is stable.
In [6] consensus on the estimates is used together with Kalman filters. The weights of the sensors' estimates in the consensus step and the Kalman gain are optimized in order to minimize the estimation error covariance. A two-step procedure is also used in [16], where the considered observed signal is a *random walk*. A two-step algorithm is proposed, where filtering and consensus are performed subsequently, and the estimation error is minimized with respect to both the observer gain and the consensus weights. This guarantees optimality of the solution.

More in general, the issue of distributed sensor fusion has been widely studied in the past years e.g., [14, 51]. The paper [14] provides an algorithm accounting for dynamically changing interconnections among sensors, unreliable communication links, and faults, where convergence of the estimates to the true values is proved, under suitable hypothesis of "dynamical" graph connectivity, while in [51] the authors propose a minimum variance estimator for distributed tracking of a noisy time-varying signal.

Other studies focused on the design of decentralized Kalman filters based on system decomposi-

tion. Different solutions can be classified according to the model used by each subsystem for state-estimation purposes and the topology of the communication network among subsystems. Early works, e.g. [23], [41] require all-to-all communication and assume each subsystem has full knowledge of the whole dynamics. Subsystems with overlapping states are also studied, *e.g.* in [31], where a fully decentralized scheme is presented.

Notation. I_n and $\mathbf{0}_{\nu\times\mu}$ denote the $n\times n$ identity matrix and the $\nu\times\mu$ matrix of zero elements, respectively. Given a set $\mathscr{S}$, $|\mathscr{S}|$ denotes its cardinality. The notation $\|z\|_S^2$ stands for z^TSz, where S is a symmetric positive-semidefinite matrix. The symbol $\otimes$ denotes the Kronecker product, and $\mathbb{1}_M$ is the M-dimensional column vector whose entries are all equal to 1. The matrix $\mathrm{diag}(\eta_1,\ldots,\eta_s)$ is block-diagonal with blocks η_i. Finally, we use the short-hand $\mathbf{v}=(v_1,\ldots,v_s)$ to denote a column vector with s (not necessarily scalar) components.

2.2 System and sensor network

We assume that the observed process obeys to the linear dynamics

$$x_{t+1}=Ax_t+w_t, \tag{2.1}$$

where $x_t\in\mathbb{X}\subseteq\mathbb{R}^n$ is the state vector and the term $w_t\in\mathbb{W}\subseteq\mathbb{R}^n$ represents a white noise with covariance equal to Q. We assume that the sets $\mathbb{X}$ and $\mathbb{W}$ are convex and contain the origin. The initial condition $x_0\in\mathbb{X}$ is a random variable with mean μ and covariance Π_0. The pair $(A,\sqrt{Q})$ is stabilizable. Measurements on the state vector are performed by M sensors, according to the sensing model (in general different from sensor to sensor)

$$y_t^i=C^ix_t+v_t^i\ ,\ i=1,...,M \tag{2.2}$$

where the term $v_t^i\in\mathbb{R}^{p_i}$ represents white noise with covariance equal to R_i.

The communication network among sensors is described by the directed graph $\mathscr{G}=(\mathscr{V},\mathscr{E})$, where the nodes in $\mathscr{V}=\{1,2,\ldots,M\}$ represent the sensors and the edge (j,i) in the set $\mathscr{E}\subseteq\mathscr{V}\times\mathscr{V}$ models that sensor j can transmit information to sensor i. We assume $(i,i)\in\mathscr{E}$, $\forall i\in\mathscr{V}$. We denote with $\mathscr{V}_i^k$ the set of k-th order neighbors to node i, *i.e.*, $\mathscr{V}_i^k=\{j\in\mathscr{V}:$ there is a path of length at most k from j to i in $\mathscr{G}\}$. We will also use the shorthand $\mathscr{V}_i=\mathscr{V}_i^1$ and we denote as M_i the number of nodes j satisfying $(i,j)\in\mathscr{E}$.

We introduce now the definition of *isolated* subgraph. If the graph $\mathscr{G}$ is not strongly connected (*i.e.*, it is *reducible*), one can partition $\mathscr{G}$ into l nonempty irreducible subgraphs $\mathscr{G}_i=(\mathscr{N}_i,\mathscr{A}_i)$, $i=1,\ldots,l$ (see e.g. [18]). If, for all $p\in\mathscr{N}_i$, $q\in\mathscr{V}_p$ implies that $q\in\mathscr{N}_i$ we say that $\mathscr{G}_i$ is *isolated*. Remark that if $\mathscr{G}$ is strongly connected, it is also isolated.

We associate to the graph $\mathscr{G}$ the stochastic matrix $K\in\mathbb{R}^{M\times M}$, with entries

$$k_{ij}\geq 0 \text{ if } (j,i)\in\mathscr{E} \tag{2.3a}$$

$$k_{ij}=0 \text{ otherwise} \tag{2.3b}$$

$$\sum_{j=1}^{M}k_{ij}=1,\ \forall i=1,...,M \tag{2.3c}$$

Any matrix K with entries satisfying (2.3) is said to be compatible with $\mathscr{G}$. Given a graph $\mathscr{G}$, there are many degrees of freedom for the choice of K, which will be exploited to guarantee the convergence of the state estimator described in the following and/or to reduce the uncertainty of the estimates.

It is assumed that, at a generic time instant t, sensor i collects the measurements produced by itself and its neighboring sensors. Moreover, each sensor transmits and receives information once within a sampling interval. This means that measurements available to node i are y_t^j, with $j \in \mathscr{V}_i$.
Three types of quantities can be distinguished: *local*, *regional*, and *collective*. Specifically, a quantity is *local* (with respect to sensor i) when it is related to the node i solely. A quantity is *regional* (with respect to sensor i) if it is related to the nodes in $\mathscr{V}_i$. Finally, a quantity is *collective*, if it is related to the whole network. For the sake of clarity, we use different notations for local, regional and collective variables. Namely, given a variable z, z^i represents its local version, $\vec{z}^i$ is its regional counterpart, and $\mathbf{z}$ the collective one. For instance, we refer to y_t^i in (2.2) as local measurement. On the other hand, if $\mathscr{V}_i = \{j_1^i, ..., j_{v_i}^i\}$, the regional measurement of node i is given by

$$\vec{y}_t^i = \bar{C}^i x_t + \vec{v}_t^i \tag{2.4}$$

where $\vec{y}_t^i = (y_t^{j_1^i}, \ldots, y_t^{j_{v_i}^i})$, $\bar{C}^i = [(C^{j_1^i})^T \ \ldots \ (C^{j_{v_i}^i})^T]^T$, and $\vec{v}_t^i = (v_t^{j_1^i}, \ldots, v_t^{j_{v_i}^i})$. The dimension of vectors $\vec{y}_t^i$ and $\vec{v}_t^i$, and the number of rows of matrix $\bar{C}^i$ is $\bar{p}_i = \sum_{k=1}^{v_i} p_{j_k^i}$. Furthermore, we denote by $\bar{R}_i$, the covariance matrix related to the regional noise $\vec{v}_t^i$ on sensor i, *i.e.*, $\bar{R}_i = \text{diag}(R_{j_1^i}, \ldots, R_{j_{v_i}^i})$.

According to the adopted terminology, three different observability notions can be introduced.

Definition 1 *The system is* locally observable *by sensor i (sensor i is locally observable) if the pair* (A, C^i) *is observable. The system is* regionally observable *by sensor i (sensor i is regionally observable) if the pair* $(A, \bar{C}^i)$ *is observable. The system is* collectively observable *if the pair* $(A, \mathbf{C}^*)$ *is observable, where* $\mathbf{C}^* = [(C^1)^T \ \ldots \ (C^M)^T]^T$. □

Notice that, for a given sensor i, local observability implies regional observability, and regional observability of any sensor implies collective observability, while all opposite implications are false. We partition the set $\mathscr{V}$ into the subsets $\mathscr{V}_O = \{j \in \mathscr{V} : (A, \bar{C}^j)$ is an observable pair$\}$, $\mathscr{V}_{NO} = \{j \in \mathscr{V} : (A, \bar{C}^j)$ is an unobservable pair$\}$.

Given a single sensor model (2.1)-(2.2), the i-th sensor regional observability matrix $\bar{\mathscr{O}}_n^i$ is

$$\bar{\mathscr{O}}_n^i = \begin{bmatrix} (\bar{C}^i)^T & (\bar{C}^i A)^T & \ldots & (\bar{C}^i A^{n-1})^T \end{bmatrix}^T \tag{2.5}$$

Let $\bar{P}_{NO}^i$ be the orthogonal projection matrix on $\ker(\bar{\mathscr{O}}_n^i)$, that is the regionally unobservable subspace. Similarly, let $\bar{P}_O^i$ be the orthogonal projection on the regional observability subspace $\ker(\bar{\mathscr{O}}_n^i)^\perp$. Next, we recall how $\bar{P}_O^i$ and $\bar{P}_{NO}^i$ can be computed. Let $r_i = \text{rank}(\bar{\mathscr{O}}_n^i)$ and denote with $\xi_{r_i+1}, \ldots, \xi_n$ an orthonormal basis of $\ker(\bar{\mathscr{O}}_n^i)$. Let also $\xi_1, \ldots, \xi_{r_i}$ be an orthonormal basis of $\ker(\bar{\mathscr{O}}_n^i)^\perp$ and define the orthonormal and non-singular matrix $\bar{T}^i = [\xi_1^i, ..., \xi_n^i]$. Defining the matrices $\bar{S}_O^i$ and $\bar{S}_{NO}^i$ as

$$\bar{S}_O^i = \begin{bmatrix} I_{r_i} \\ 0_{(n-r_i)\times r_i} \end{bmatrix}, \ \bar{S}_{NO}^i = \begin{bmatrix} 0_{r_i\times(n-r_i)} \\ I_{n-r_i} \end{bmatrix},$$

we have $\bar{P}_O^i = \bar{T}^i \bar{S}_O^i (\bar{S}_O^i)^T (\bar{T}^i)^{-1}$ and $\bar{P}_{NO}^i = \bar{T}^i \bar{S}_{NO}^i (\bar{S}_{NO}^i)^T (\bar{T}^i)^{-1}$. Furthermore, defining $\mathbf{T} = \text{diag}(\bar{T}_1, \ldots, \bar{T}_M)$, $\mathbf{S}_O = \text{diag}(\bar{S}_O^1, \ldots, \bar{S}_O^M)$, and $\mathbf{S}_{NO} = \text{diag}(\bar{S}_{NO}^1, \ldots, \bar{S}_{NO}^M)$, the collective projection matrices are $\mathbf{P}_O = \mathbf{T}\mathbf{S}_O\mathbf{S}_O^T\mathbf{T}^{-1}$ and $\mathbf{P}_{NO} = \mathbf{T}\mathbf{S}_{NO}\mathbf{S}_{NO}^T\mathbf{T}^{-1}$. Note that $\bar{S}_{NO}^i$ is empty when the system is regionally observable by sensor i. In this case we assume that $\bar{P}_{NO}^i = \mathbf{0}_{n\times n}$.

2.3 The distributed estimation algorithm

Our aim is to design, for a generic sensor $i \in \mathscr{V}$, an algorithm for computing an estimate of the system state based on regional measurements $\vec{y}_t^i$ and further pieces of information provided by sensors $j \in \mathscr{V}_i$.

The proposed solution relies on MHE, in view of its capability to handle state and noise constraints. More specifically, we propose a Distributed MHE (DMHE) scheme where each sensor solves a MHE problem.

2.3.1 The local minimization problem

For a given estimation horizon $N \geq 1$, each node $i \in \mathscr{V}$ at time t determines the estimates $\hat{x}^i$ and $\hat{w}^i$ of x and w, respectively, by solving the constrained minimization problem (*MHE-i*)

$$\Theta_t^{*i} = \min_{\hat{x}^i_{t-N}, \{\hat{w}^i_k\}_{k=t-N}^{t-1}} J^i(t-N, t, \hat{x}^i_{t-N}, \hat{w}^i, \hat{\tilde{v}}^i, \Gamma^i_{t-N}) \tag{2.6}$$

under the constraints

$$\hat{x}^i_{k+1} = A\hat{x}^i_k + \hat{w}^i_k, \ k = t-N, \ldots, t \tag{2.7a}$$
$$\bar{y}^i_k = \bar{C}^i \hat{x}^i_k + \hat{\tilde{v}}^i_k \tag{2.7b}$$
$$\hat{w}^i_k \in \mathbb{W} \tag{2.7c}$$
$$\hat{x}^i_k \in \mathbb{X} \tag{2.7d}$$

The local cost function J^i is given by

$$\begin{aligned} J^i(t-N, t, \hat{x}^i_{t-N}, \hat{w}^i, \hat{\tilde{v}}^i, \Gamma^i_{t-N}) = \tfrac{1}{2}\textstyle\sum_{k=t-N}^{t} \|\hat{\tilde{v}}^i_k\|^2_{R_i^{-1}} + \\ + \tfrac{1}{2}\textstyle\sum_{k=t-N}^{t-1} \|\hat{w}^i_k\|^2_{Q^{-1}} + \Gamma^i_{t-N}(\hat{x}^i_{t-N}; \hat{\tilde{x}}^i_{t-N/t-1}) \end{aligned} \tag{2.8}$$

We denote with $\hat{x}^i_{t-N/t}$ and with $\{\hat{w}^i_{k/t}\}_{k=t-N}^{t-1}$ the optimizers to (2.6) and with $\hat{x}^i_{k/t}$, $k = t-N, \ldots, t$ the local state sequence stemming from $\hat{x}^i_{t-N/t}$ and $\{\hat{w}^i_{k/t}\}_{k=t-N}^{t-1}$. Furthermore

$$\hat{\tilde{x}}^i_{t-N/t-1} = \sum_{j=1}^{M} k_{ij} \hat{x}^j_{t-N/t-1} \tag{2.9}$$

denotes the weighted average state estimates produced by sensors $j \in \mathscr{V}^i$. In (2.8), the function $\Gamma^i_{t-N}(\hat{x}^i_{t-N}; \hat{\tilde{x}}^i_{t-N/t-1})$ is the so called *initial penalty*, defined as follows

$$\Gamma^i_{t-N}(\hat{x}^i_{t-N}; \hat{\tilde{x}}^i_{t-N/t-1}) = \frac{1}{2}\|\hat{x}^i_{t-N} - \hat{\tilde{x}}^i_{t-N/t-1}\|^2_{(\Pi^i_{t-N/t-1})^{-1}} + \Theta^{*i}_{t-1} \tag{2.10}$$

where Θ^{*i}_{t-1} is the optimal cost defined in (2.6) and the positive-definite symmetric weighting matrix $\Pi^i_{t-N/t-1}$ appearing in (2.10) plays the role of a covariance matrix whose choice will be discussed in details in the next paragraphs. The term Θ^{*i}_{t-1} is a constant in (2.10) and could be neglected when solving (2.6). However, since it plays a major role in establishing the main convergence properties of DMHE, it is here maintained for clarity of presentation.

Note that, in view of the definition of k_{ij} in (2.3), $\Gamma^i(\cdot)$ depends only upon regional quantities and, since also the cost (2.8) and the constraints (2.7) depend only upon regional variables, the overall estimation scheme is decentralized. Finally, notice that $\Gamma^i(\cdot)$ embodies a *consensus-on-estimates* term, in the sense that it penalizes deviations of $\hat{x}^i_{t-N/t-1}$ from $\hat{\tilde{x}}^i_{t-N/t-1}$. Consensus, besides increasing accuracy of the local estimates, is fundamental to guarantee convergence of the state estimates to the state of the observed system even if regional observability does not hold. In other words, it allows sensor i to reconstruct components of the state that cannot be estimated by the i-th regional model.

2.3.2 The collective minimization problem

The local estimation problems (2.6)-(2.10) can be given a collective form more suitable for the following developments. To this end, let $\mathbf{J}$ be the collective cost function given by

$$\mathbf{J}(\cdot) = \sum_{i=1}^{M} J^i(t-N, t, \hat{x}^i_{t-N}, \hat{w}^i, \hat{\tilde{v}}^i, \Gamma^i_{t-N}) \tag{2.11}$$

Define the collective vectors $\hat{\mathbf{x}}_t = \left(\hat{x}^1_t, \ldots, \hat{x}^M_t\right)$, $\hat{\tilde{\mathbf{v}}}_t = \left(\hat{\tilde{v}}^1_t, \ldots, \hat{\tilde{v}}^M_t\right)$, $\hat{\mathbf{w}}_t = \left(\hat{w}^1_t, \ldots, \hat{w}^M_t\right)$, the quantities $\mathbf{\Theta}^*_{t-1} = \sum_{i=1}^{M} \Theta^{*i}_{t-1}$, $\mathbf{K} = K \otimes I_n$,

$$\mathbf{\Pi}_{t_1/t_2} = \operatorname{diag}\left(\Pi^1_{t_1/t_2}, \ldots, \Pi^M_{t_1/t_2}\right) \tag{2.12}$$

and the collective initial penalty

$$\mathbf{\Gamma}_{t-N}(\hat{\mathbf{x}}_{t-N}; \hat{\mathbf{x}}_{t-N/t-1}) = \mathbf{\Gamma}^o_{t-N}(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) + \mathbf{\Theta}^*_{t-1} \tag{2.13}$$

where $\mathbf{\Gamma}^o_{t-N}(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) = \frac{1}{2}\|\hat{\mathbf{x}}_{t-N} - \mathbf{K}\hat{\mathbf{x}}_{t-N/t-1}\|^2_{\mathbf{\Pi}^{-1}_{t-N/t-1}}$. Then, using the matrices $\bar{\mathbf{R}} = \operatorname{diag}\left(\bar{R}^1, \ldots, \bar{R}^M\right)$, $\mathbf{Q} = \operatorname{diag}(Q, \ldots, Q) \in \mathbb{R}^{nM \times nM}$, the collective cost function $\mathbf{J}(\cdot)$ can be rewritten as

$$\begin{array}{c} \mathbf{J}(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}, \hat{\tilde{\mathbf{v}}}, \mathbf{\Gamma}_{t-N}) = \frac{1}{2}\sum_{k=t-N}^{t} \|\hat{\tilde{\mathbf{v}}}_k\|^2_{\bar{\mathbf{R}}^{-1}} + \\ + \frac{1}{2}\sum_{k=t-N}^{t-1} \|\hat{\mathbf{w}}_k\|^2_{\mathbf{Q}^{-1}} + \mathbf{\Gamma}_{t-N}(\hat{\mathbf{x}}_{t-N}; \hat{\mathbf{x}}_{t-N/t-1}) \end{array} \tag{2.14}$$

Defining $\mathbf{A}$ =diag$(A, ..., A) \in \mathbb{R}^{nM \times nM}$ and $\bar{\mathbf{C}}$ =diag$(\bar{C}^1, ..., \bar{C}^M)$, also the constraints (2.7) can be written in the following collective form

$$\hat{\mathbf{x}}_{k+1} = \mathbf{A}\hat{\mathbf{x}}_k + \hat{\mathbf{w}}_k, \ k = t-N, \ldots, t \tag{2.15a}$$

$$\bar{\mathbf{y}}_k = \bar{\mathbf{C}}\hat{\mathbf{x}}_k + \hat{\tilde{\mathbf{v}}}_k \tag{2.15b}$$

$$\hat{\mathbf{w}}_k \in \mathbb{W}^M \tag{2.15c}$$

$$\hat{\mathbf{x}}_k \in \mathbb{X}^M \tag{2.15d}$$

It is important to note that solving the problem

$$\Theta^*_t = \min_{\hat{\mathbf{x}}_{t-N}, \{\hat{\mathbf{w}}_k\}_{k=t-N}^{t-1}} \left\{ \mathbf{J}(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}, \hat{\tilde{\mathbf{v}}}, \mathbf{\Gamma}_{t-N}) \text{ subject to (2.15)} \right\} \tag{2.16}$$

is equivalent to solve the *MHE-i* problems (2.6), in the sense that $\hat{x}^i_{t-N/t}, \{\hat{w}^i_{k/t}\}_{k=t-N}^{t-1}$ is a solution to (2.6) if and only if $\hat{\mathbf{x}}_{t-N/t}, \{\hat{\mathbf{w}}_{k/t}\}_{k=t-N}^{t-1}$ is a solution to (2.16), where $\hat{\mathbf{w}}_{k/t} = (\hat{w}^1_{k/t}, \ldots, \hat{w}^M_{k/t})$.

Let t_1 verify $t-N \le t_1 \le t$. We define the *transit cost* of a generic state $\mathbf{z} \in \mathbb{R}^{nM}$, computed at instant t as

$$\begin{array}{c} \mathbf{\Xi}_{t_1/t}(\mathbf{z}) = \min_{\hat{\mathbf{x}}_{t-N}, \{\hat{\mathbf{w}}_k\}_{k=t-N}^{t-1}} \left\{ \mathbf{J}(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}, \hat{\tilde{\mathbf{v}}}, \mathbf{\Gamma}_{t-N}) \right. \\ \left. \text{subject to (2.15) and } \hat{\mathbf{x}}_{t_1} = \mathbf{z} \right\} \end{array} \tag{2.17}$$

As discussed in [22], $\mathbf{\Xi}_{t_1/t}(\mathbf{z})$ provides a measure of the likelihood that $\hat{\mathbf{x}}_{t_1}$ is equal to $\mathbf{z}$ given the data $\bar{\mathbf{y}}_k$, $k = t-N, \ldots, t$ and the prior likelihood $\mathbf{\Gamma}_{t-N}(\cdot)$ on $\mathbf{x}_{t-N}$. Specifically, the lower $\mathbf{\Xi}_{t_1/t}(\mathbf{z})$, the more

likely the equality $\hat{\mathbf{x}}_{t_1} = \mathbf{z}$. The prior $\mathbf{\Gamma}_{t-N}(\cdot)$ can be interpreted as an approximation of $\mathbf{\Xi}_{t-N/t-1}(\cdot)$. The key condition involving these two terms, that will be fundamental for proving convergence of DMHE (see Section 2.7.2), is that, for all $z \in \mathbb{X}$

$$\mathbf{\Gamma}_{t-N}(\mathbb{1}_M \otimes z; \hat{\mathbf{x}}_{t-N/t-1}) \leq \mathbf{\Xi}_{t-N/t-1}(\mathbb{1}_M \otimes z) \tag{2.18}$$

Equation (2.18) is similar to the assumption (C2) in [45] for centralized MHE. However, in (2.18), the transit cost instead of the arrival cost appears. In fact $\mathbf{\Xi}_{t-N/t-1}$ is a smoothing term, since it takes into account data up to time t, in order to enforce consensus (in [44] this approach is called *smoothing update*).

An explicit formula for a lower bound to $\mathbf{\Xi}_{t-N/t-1}(\mathbf{z})$ (which coincides with $\mathbf{\Xi}_{t-N/t-1}(\mathbf{z})$ for unconstrained estimation problems, see the proof of Lemma 1 in Section 2.7.2) is given by a quadratic cost function, i.e.

$$\mathbf{\Xi}_{t-N/t-1}(\mathbf{z}) \geq \frac{1}{2}\|\mathbf{z} - \hat{\mathbf{x}}_{t-N/t-1}\|^2_{(\tilde{\mathbf{\Pi}}_{t-N/t-1})^{-1}} + \mathbf{\Theta}^*_{t-1} \tag{2.19}$$

for a suitable choice of $\tilde{\mathbf{\Pi}}_{t-N/t-1}$. The computation of $\tilde{\mathbf{\Pi}}_{t-N/t-1}$, and a procedure for updating the matrix $\mathbf{\Pi}_{t-N/t-1}$ in (2.13) satisfying (2.18) are given in the next section.

2.3.3 Update of the weighting matrices

As remarked in the previous section, the first step for updating matrices ${\Pi^i}_{t-N/t-1}$, is to compute $\tilde{\mathbf{\Pi}}_{t-N/t-1}$ in (2.19), with the following diagonal structure

$$\tilde{\mathbf{\Pi}}_{t-N/t-1} = \operatorname{diag}\left(\tilde{\Pi}^1_{t-N/t-1}, \ldots, \tilde{\Pi}^M_{t-N/t-1}\right) \tag{2.20}$$

where the update of $\tilde{\Pi}^i_{t-N/t-1}$ is carried out by the sensor i, based on regional pieces of information. For this reason, this step is denoted *regional weights update*. Specifically, the matrix $\tilde{\Pi}^i_{t-N/t-1}$, $i \in \mathscr{V}$, is given by one iteration of the difference Riccati equation associated to a Kalman filter for the system

$$\begin{cases} x_{t-N} &= Ax_{t-N-1} + w_{t-N-1} \\ \bar{z}^i_{t-N} &= \bar{\mathscr{O}}^i_N x_{t-N} + \bar{V}^i_{t-N} \end{cases}$$

where matrix $\bar{\mathscr{O}}^i_N$ is defined in (2.5). If we define

$$\mathscr{C}^i_N = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ \bar{C}^i & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{C}^i A^{N-2} & \bar{C}^i A^{N-3} & \ldots & \bar{C}^i \end{bmatrix} \in \mathbb{R}^{\bar{p}_i N \times n(N-1)} \tag{2.21}$$

$$\bar{R}^i_N = \operatorname{diag}\left(\bar{R}^i, \ldots, \bar{R}^i\right) \in \mathbb{R}^{\bar{p}_i N \times \bar{p}_i N} \tag{2.22}$$

$$Q_{N-1} = \operatorname{diag}(Q, \ldots, Q) \in \mathbb{R}^{n(N-1) \times n(N-1)} \tag{2.23}$$

$$\operatorname{Cov}[w_t] = Q \tag{2.24}$$

$$\operatorname{Cov}[\bar{V}^i_t] = \bar{R}^{*i}_N = \bar{R}^i_N + \mathscr{C}^i_N Q_{N-1} (\mathscr{C}^i_N)^T \tag{2.25}$$

and set the covariance of the estimate $\hat{x}^i_{t-N-1}$ as

$$\Pi^{*i}_{t-N-1/t-2} = ((\Pi^i_{t-N-1/t-2})^{-1} + (\bar{C}^i)^T (\bar{R}^i)^{-1} \bar{C}^i)^{-1} \tag{2.26}$$

the resulting Riccati recursion is given by

$$\begin{aligned}\tilde{\Pi}^i_{t-N/t-1} =& \mathscr{R}^i\left(\Pi^{*i}_{t-N-1/t-2}; Q, \bar{R}^{*i}_N\right) \\ =& A\Pi^{*i}_{t-N-1/t-2}A^T + Q - A\Pi^{*i}_{t-N-1/t-2}(\bar{\mathscr{O}}^i_N)^T \times \\ & \times \left(\bar{\mathscr{O}}^i_N \Pi^{*i}_{t-N-1/t-2}(\bar{\mathscr{O}}^i_N)^T + \bar{R}^{*i}_N\right)^{-1} \bar{\mathscr{O}}^i_N \Pi^{*i}_{t-N-1/t-2}A^T\end{aligned} \tag{2.27}$$

Once the matrices $\tilde{\Pi}^i_{t-N/t-1}$ have been computed, we perform a *consensus weights update*, in order to compute the matrices $\Pi^i_{t-N/t-1}$ appearing in (2.10), which must satisfy the fundamental inequality (2.18). As stated in Lemma 1 in Section 2.7.2, (2.18) is verified if $\mathbf{\Pi}_{t-N/t-1}$ fulfills the Linear Matrix Inequality (LMI)

$$\mathbf{\Pi}_{t-N/t-1} \geq \mathbf{K}\tilde{\mathbf{\Pi}}_{t-N/t-1}\mathbf{K}^T \tag{2.28}$$

The LMI (2.28) deserves a few comments. In order to make the initial penalty $\mathbf{\Gamma}_{t-N}(\cdot)$ a good approximation of the transit cost $\mathbf{\Xi}_{t-N/t-1}(\cdot)$, one would require the matrix $\mathbf{\Pi}_{t-N/t-1}$ to be "as close as possible" to $\mathbf{K}\tilde{\mathbf{\Pi}}_{t-N/t-1}\mathbf{K}^T$. Therefore, in our case, one would make the matrix $\mathbf{\Pi}_{t-N/t-1}$ "as small as possible", subject to the constraint (2.28). A way for achieving this is to solve the LMI problem

$$\min\left(\text{trace}(\mathbf{\Pi}_{t-N/t-1})\right), \text{ subject to (2.28)}, \tag{2.29}$$

where $\mathbf{\Pi}_{t-N/t-1}$ has the block-diagonal structure (2.12). Notice that (2.29) could be solved by each sensor since, similarly to the formula for updating covariances in Kalman filtering, the computation of $\Pi^i_{t-N/t-1}$ does not depend upon the collected measurements. However, problem (2.29) has a centralized flavor. This limitation is severe since, for instance, the LMI (2.28) has size $n \times M$ which implies that the computational burden for solving (2.29) scales with the number of sensors, hence hampering the application of DMHE to large networks. The next proposition provides a way to circumvent this problem.

Proposition 1 *The matrices* $\Pi^i_{t-N/t-1}$ *which satisfy,* $\forall i \in \mathscr{V}$

$$\Pi^i_{t-N/t-1} \geq \sum_{j=1}^{M} M_j k_{ij}^2 \tilde{\Pi}^j_{t-N/t-1} \tag{2.30}$$

also satisfy the LMI (2.28).

Proof 1 *See Section 2.7.1.*

Notice that, in the solution provided by Proposition 1, each node computes $\Pi^i_{t-N/t-1}$ solely on the basis of $\tilde{\Pi}^j_{t-N/t-1}$, provided by its neighbors, $j \in \mathscr{V}_i$. In view of this, the LMI (2.28) can be solved in a decentralized fashion by setting

$$\Pi^i_{t-N/t-1} = \sum_{j=1}^{M} M_j k_{ij}^2 \tilde{\Pi}^j_{t-N/t-1} \tag{2.31}$$

2.3.4 DMHE algorithm

In the following we sketch the steps that have to be carried out, in practice, in order to apply the proposed DMHE algorithm.

- Initialization: at $t = 0$ all nodes store the matrix Π_0 and the estimate $\hat{x}_{0/0} = \mu$ of x_0, where μ is given. Recall that Π_0 is the covariance matrix related to the initial condition x_0
- if $1 \leq t \leq N$, the estimation horizon N is reduced to $\tilde{N} = t$ and node $i \in \mathscr{V}$ performs the following steps
  - compute $\Pi^i_{t-\tilde{N}/t-1} = \Pi^i_{0/t-1}$ from Π_0 according to (2.31), for all $i \in \mathscr{V}$,
  - solve the problem *MHE-i*, with initial penalty

$$\Gamma^i_{t-\tilde{N}} = \frac{1}{2}\|\hat{x}^i_0 - \tilde{x}^i_{0/t-1}\|^2_{(\Pi^i_{0/t-1})^{-1}}$$

- if $t > N$, at each time instant, every node $i \in \mathscr{V}$,
  - computes $\Pi^i_{t-N/t-1}$ from $\mathbf{\Pi}_{t-N-1/t-2}$ according to (2.26), (2.27) and (2.31),
  - solves the problem *MHE-i*, with initial penalty

$$\Gamma^i_{t-N} = \frac{1}{2}\|\hat{x}^i_{t-N} - \tilde{x}^i_{t-N/t-1}\|^2_{(\Pi^i_{t-N/t-1})^{-1}}$$

2.4 Convergence properties of DMHE

The main purpose of this section is to extend the convergence results of [44] for centralized MHE to the proposed DMHE scheme.

Definition 2 *Let Σ be system* (2.1) *with $w = 0$ and denote by $x_\Sigma(t, x_0)$ the state reached by Σ at time t starting from initial condition x_0. Assume that the trajectory $x_\Sigma(t, x_0)$ is feasible,* i.e., *$x_\Sigma(t, x_0) \in \mathbb{X}$ for all t. DMHE is* convergent *if $\|\hat{x}^i_{t/t} - x_\Sigma(t, x_0)\| \xrightarrow{t\to\infty} 0$ for all $i \in \mathscr{V}$.* □

Note that, as in [44], convergence is defined assuming that the model generating the data is noiseless, but the possible presence of noise is taken into account in the state estimation algorithm. Now, defining the collective vector $\mathbf{x}_\Sigma(t, x_0) = \mathbb{1}_M \otimes x_\Sigma(t, x_0)$ and $\boldsymbol{\varepsilon}_{k/t} = \hat{\mathbf{x}}_{k/t} - \mathbf{x}_\Sigma(k, x_0)$, the following result can be stated.

Theorem 1 *If: (i) matrices $\Pi^i_{t-N/t-1}$ are computed according to* (2.26), (2.27) *and* (2.28), *(ii) $\Pi^i_{t-N/t-1}$ are bounded for all t, and for all $i \in \mathscr{V}$, (iii) $N \geq n-1$ and $N \geq 1$, then*

a) there exists an asymptotically vanishing sequence α_t (i.e., $\|\alpha_t\| \xrightarrow{t\to\infty} 0$) such that the dynamics of the state estimation error provided by the DMHE scheme is given by

$$\boldsymbol{\varepsilon}_{t-N/t} = \Phi\boldsymbol{\varepsilon}_{t-N-1/t-1} + \alpha_t \tag{2.32}$$

where $\Phi = \mathbf{P}_{NO}\mathbf{KAP}_{NO}$;

b) if (iv) Φ is Schur, then DMHE is convergent.

Proof 2 *See Section 2.7.2.*

In Section 2.5 we will provide a method to choose a matrix K compatible with $\mathscr{G}$ such that conditions (ii) and (iv) of Theorem 1 are satisfied.

We highlight that Condition (iv) does not require the asymptotic stability of system (2.1). Moreover, Theorem 1 does not hinge on observability properties. In fact, convergence of the estimation error can be achieved even if a weaker detectability property is satisfied, i.e. if matrix Φ inherits only stable eigenvalues of A. However, it is of interest to determine conditions guaranteeing that the matrix Φ does not inherit any (non-zero) eigenvalue of A. The reason is twofold. First, this is tantamount to requiring that the unobservable dynamics of all regional systems are affected by the communication network. Second, the study is a preliminary step towards the goal of choosing K and, when possible, the network topology, in order to assign the eigenvalues of Φ at will. Let λ_A^i and v_A^i be the eigenvalues and the eigenvectors of A, respectively, with $i = 1,...,n$. Then, the eigenvalues of $\mathbf{A}$ are λ_A^i $(i = 1,\dots,n)$, each one with multiplicity M. Moreover, denoting by e_j, $j = 1,...,M$ the canonical basis vectors of $\mathbb{R}^M$, the eigenspace related to λ_A^i is span$(e_1 \otimes v_A^i, ..., e_M \otimes v_A^i)$. In view of the previous discussion, we want to investigate the following property.

Property 1 *If λ_i is a non-zero eigenvalue of A, for all $\mathbf{x} \in$span$(e_1 \otimes v_A^i, ..., e_M \otimes v_A^i)$, λ_A^i and $\mathbf{x}$ are not an eigenvalue/eigenvector pair for Φ.* □

Conditions guaranteeing that Property 1 holds are given in the following Theorem, which is illustrated in Figure 2.1.

Theorem 2 *Consider a partition of $\mathscr{G}$ into the irreducible subgraphs $\mathscr{G}_i = (\mathscr{N}_i, \mathscr{A}_i)$, $i = 1,\dots,l$. If for all the isolated strongly connected subgraphs $\mathscr{G}_i$ it holds*

$$\bigcap_{j \in \mathscr{N}_i} \ker(\bar{\mathscr{O}}_n^j) = 0 \tag{2.33}$$

then Property 1 is verified.

Proof 3 *See Section 2.7.3.*

In the case of strongly connected graphs we have the following result.

Corollary 1 *If $\mathscr{G}$ is strongly connected and the system is collectively observable, then Property 1 is verified.*

Proof 4 *See Section 2.7.3.*

As a trivial case, assume that all sensors are regionally observable and arranged in a strongly connected graph $\mathscr{G}$. This yields $\mathbf{P}_{NO} = \Phi = \mathbf{0}_{nM \times nM}$ and convergence of DMHE follows from Theorem 1. Moreover, Property 1 trivially holds.

2.5 Selection of the design parameter K

The key assumption of Theorem 1 that the sequence $\{\Pi_{t-N/t-1}\}_{t=0}^{\infty}$ is bounded is not a-priori guaranteed by formula (2.31). However, under weak assumptions, boundedness of $\{\Pi_{t-N/t-1}\}_{t=0}^{\infty}$ can be enforced by properly choosing the entries k_{ij}, $\forall (i,j) \in \mathscr{E}$ of K. Interestingly, we will also prove that the proposed choice of K results in assigning all the eigenvalues of Φ equal to zero, that guarantees convergence of DMHE and Property 1.

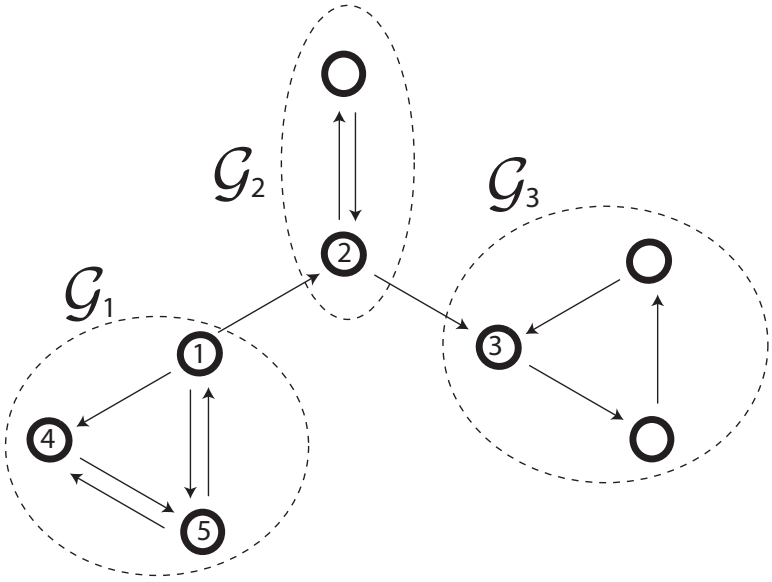


Figure 2.1: The graph is decomposed into three connected subgraphs $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$. Notice that the node 2 of $\mathcal{G}_2$ is a neighbor to node 3 of $\mathcal{G}_3$. Therefore, graph $\mathcal{G}_3$ is not isolated. Analogously, the subgraph $\mathcal{G}_2$ is not isolated, while the subgraph $\mathcal{G}_1$ is isolated. Condition (2.33) states that collective observability is required for the subgraph $\mathcal{G}_1$, *i.e.*, the pair $(A, \mathbf{C}^*_{\mathcal{G}_1})$ is observable, where $\mathbf{C}^*_{\mathcal{G}_1} = [(C^1)^T\ (C^4)^T\ (C^5)^T]^T$.

Theorem 3 *If $\mathcal{V}_O$ is non-empty and, for all $i \in \mathcal{V}_{NO}$, there exists $k > 0$ such that $\mathcal{V}_i^k \bigcap \mathcal{V}_O \neq \emptyset$, then there exists K, compatible with $\mathcal{G}$, such that matrices $\Pi^i_{t-N/t-1}$ ($i = 1,\ldots,M$), resulting from* (2.27) *and* (2.31) *are bounded for all $i \in \mathcal{V}$.*

Proof 5 *See Section 2.7.4.*

The assumption of Theorem 3 that, for each node in $\mathcal{V}_{NO}$, there exists an incoming directed path stemming from a node in $\mathcal{V}_O$ requires that at least one sensor is regionally observable. This condition, although not necessary to guarantee the existence of a suitable K, allows one to identify at least a "reference" node, which provides reliable estimates even without communication, see the proof of Theorem 3 in Section 2.7.4. The proof of Theorem 3 is constructive and leads to the following algorithm for computing the matrix K. Algorithm 1 is illustrated in Figure 2.2. Given the availability of methods for computing paths with a computational complexity that scales polynomially with $|\mathcal{V}|$ [10], the overall algorithm is polynomial. Moreover, if $\mathcal{G}$ is complete graph, Algorithm 1 provides a method for designing a not a-priori fixed communication network. Furthermore, Algorithm 1 implicitly provides a rule for connecting a new regionally observable/unobservable sensor to the network without spoiling the boundedness of the sequence $\{\Pi_{t-N/t-1}\}_{t=0}^{\infty}$.

Finally, by selecting K according to Algorithm 1, the following result holds.

Corollary 2 *Under the assumption of Theorem 3, if K is selected according to Algorithm 1, then Φ has all the eigenvalues equal to zero.*

Proof 6 *See Section 2.7.5.*

Algorithm 1 Selection of K

1) for each $i \in \mathscr{V}_O$, set $k_{ii} = 1$;
2) for each $i \in \mathscr{V}_{NO}$, select $k_{ii} < \frac{1}{\sqrt{M_i}\sigma^{(i)}(A)}$, where $\sigma^{(i)}(A)$=max$\{|\lambda_j^{(i)}(A)| : \lambda_j^{(i)}(A)$ is an unobservable eigenvalue for the pair $(A, \overline{C}^i)\}$;
3) for each $i \in \mathscr{V}_{NO}$ select a node $j \in \mathscr{V}_O$ and a path from j to i, in such a way that each node in the path has at most one neighbor. We denote with $\mathscr{E}^*$ the set of edges selected in this way;
4) for all edges $(i, j) \in \mathscr{E}^*$, choose $k_{ij} = 1 - k_{ii}$, while for all edges $(i, j) \in \mathscr{E} \setminus \mathscr{E}^*$, set $k_{ij} = 0$. □

A final remark is due. Under the assumption of Theorem 3, the choice of a matrix K is not unique and details on the available degrees of freedom in the definition of a suitable K (see Remark 1 after the proof of Theorem 3) can be used to reduce the conservativeness imposed by Algorithm 1. In fact the generated matrix K is lower triangular, up to a permutation of the node indexes. However, the same arguments of the proof of Theorem 3 can be used to show that boundedness of $\Pi^i_{t-N|t-1}$ is guaranteed also by any stochastic matrix $\overline{K}$ compatible with $\mathscr{G}$ with: (i) the same diagonal elements of the matrix K obtained with Algorithm 1; (ii) arbitrary (non a-priori zero) elements in the lower triangular part; (iii) sufficiently small (non a-priori zero) elements in the upper triangular part. Details on point (iii) are given in Remark 1 in Section 2.7.4. This choice allows for a full exploitation of the communication links. In view of this, and the fact that connected components of the graph produced by Algorithm 1 can be linked through arcs, one expects to increase convergence rates of the estimates to a common value. Moreover, the presence of more links results in an increased reliability against communication faults.

2.6 Example

We consider the fourth-order system

$$x_{t+1} = \begin{bmatrix} 0.9962 & 0.1949 & 0 & 0 \\ -0.1949 & 0.3819 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1.21 & 1.98 \end{bmatrix} x_t + w_t \tag{2.34}$$

where $x_t = [x_{1,t}\, x_{2,t}\, x_{3,t}\, x_{4,t}]^T$. Notice that the eigenvalues of the matrix A are 0.9264, 0.4517, $0.99 \pm 0.4795\,i$ and, since $|0.99 \pm 0.4795\,i| > 1$, the system is unstable.

Let $e_t \in \mathbb{R}^4$, be white noise with covariance $Q_e = \text{diag}\,(0.0012, 0.038, 0.0012, 0.038)$. In the following we consider two cases

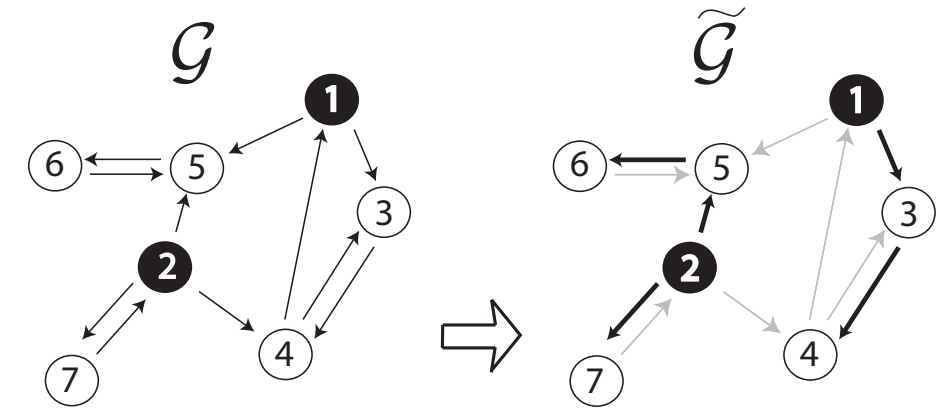
A. $w_t = e_t$, $Q = Q_e$ and $\mathbb{W} = \mathbb{R}^4$ (unconstrained input noise)

B. $w_t = |e_t|$, $Q = Q_e$ and $\mathbb{W} = \mathbb{R}^4_{\geq 0}$ (constrained input noise)

In both cases, we set $\mu = [0\ 0\ 0\ 0]^T$, $\Pi_0 = 100 I_n$ and $N = 2$ in the DMHE algorithm.

The state of (2.34) is measured by $M = 4$ sensors with sensing model

$$\begin{aligned} y^i_t &= [1\ 0\ 0\ 0]\, x_t + v^i_t \quad \text{if } i = 1,2 \\ y^i_t &= [0\ 0\ 1\ 0]\, x_t + v^i_t \quad \text{if } i = 3,4 \end{aligned}$$

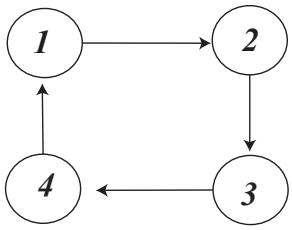


$$K = \begin{bmatrix} k_{11} & 0 & 0 & k_{14} & 0 & 0 & 0 \\ 0 & k_{22} & 0 & 0 & 0 & 0 & k_{27} \\ k_{31} & 0 & k_{33} & k_{34} & 0 & 0 & 0 \\ 0 & k_{42} & k_{43} & k_{44} & 0 & 0 & 0 \\ k_{51} & k_{52} & 0 & 0 & k_{55} & k_{56} & 0 \\ 0 & 0 & 0 & 0 & k_{65} & k_{66} & 0 \\ 0 & k_{72} & 0 & 0 & 0 & 0 & k_{77} \end{bmatrix} \qquad K = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \widetilde{k}_{31} & 0 & \widetilde{k}_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & \widetilde{k}_{43} & \widetilde{k}_{44} & 0 & 0 & 0 \\ 0 & \widetilde{k}_{52} & 0 & 0 & \widetilde{k}_{55} & 0 & 0 \\ 0 & 0 & 0 & 0 & \widetilde{k}_{65} & \widetilde{k}_{66} & 0 \\ 0 & \widetilde{k}_{72} & 0 & 0 & 0 & 0 & \widetilde{k}_{77} \end{bmatrix}$$

Figure 2.2: The original graph $\mathcal{G}$ (left panel) presents two types of sensors: nodes 1 and 2 are regionally observable (black circles), and nodes 3-7 are regionally unobservable (white circles). Therefore, $\mathcal{V}_O = \{1,2\}$ and $\mathcal{V}_{NO} = \{3,4,5,6,7\}$. The graph $\tilde{G}$ (right panel) is defined by selecting a subset of edges (black ones) of the original graph, according to step 2 of Algorithm 1. Below each graph we show the corresponding matrix K.

where $\mathrm{Var}(v_t^i) = R_i = 1$, $i = 1,\ldots,4$. Sensors are connected according to the graph in Figure 2.3, where the matrix K is also given. It is apparent that the information available, at each instant, to node 1 consists of the measurements of $x_{1,t}$ and $x_{3,t}$ (transmitted by sensor 4). Analogously, the information available to node 3 consists of $x_{1,t}$ (transmitted by sensor 2) and $x_{3,t}$. It is easy to check that the system is regionally observable by sensors 1 and 3. On the other hand, at each time instant sensor 2 can only use two different measurements of $x_{1,t}$ (produced by sensors 1 and 2). Similarly, sensor 4 can only use two different measures of $x_{3,t}$ (produced by sensors 3 and 4). Therefore, the system is not regionally observable by sensors 2 and 4. In fact, $\bar{P}^2_{NO} = \mathrm{diag}(0,0,1,1)$, $\bar{P}^4_{NO} = \mathrm{diag}(1,1,0,0)$. The eigenvalues of the matrix Φ are 0, 0.4632, 0.2258 and $0.4950 \pm 0.2397i$. Since Φ is Schur, convergence of DMHE is guaranteed by Theorem 1. Moreover, since the graph is strongly connected and collective observability holds, Corollary 1 guarantees that also Property 1 holds.

In Figure 2.4(a) the estimation errors produced by all sensors in the case A are shown. It is worth noticing that the estimates produced by sensors 2 [resp. 4], relative to states $x_{3,t}$, $x_{4,t}$ [resp. $x_{1,t}$, $x_{2,t}$] display big errors for $t < 6$. In fact, these states cannot be observed by these sensors using regional measurements. Nonetheless, the estimation errors of all sensors asymptotically tend to the same values, thanks to the consensus action embodied in the DMHE scheme. The estimation errors for case B are depicted in Figure 2.4(b). Analogously to case A, convergence of DMHE can be noticed. Figure 2.5 depicts the evolution of the eigenvalues of matrices $\Pi^i_{t/t+N-1}$ over time. Note that these



$$K = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Figure 2.3: Communication network and associated matrix K

matrices are the same in the cases A and B. Indeed, the update procedure described in Section 2.3.3 does not depend on the estimates and can be run off-line. Further simulation experiments have been performed (results not shown for space limitations), in order to assess the effect of the variation of the horizon length N on the estimation performances. As expected, the larger the horizon length, the more accurate the results. In fact, as N increases, a larger set of data is taken into account in the optimization problem. However, the need of increasing N for optimality reasons is conflicting with the need of reducing as much as possible the computational load.

2.7 Proofs

2.7.1 Proof of Proposition 1

Proof 7 *For all vectors* $\mathbf{x} = [x_1^T \ \dots \ x_M^T]^T \in \mathbb{R}^{nM}$, *from* (2.28) *it holds that*

$$\mathbf{x}^T \mathbf{\Pi}_{t-N/t-1} \mathbf{x} \geq \mathbf{x}^T \mathbf{K} \tilde{\mathbf{\Pi}}_{t-N/t-1} \mathbf{K}^T \mathbf{x} \tag{2.35}$$

Notice that the j*-th block of* $\mathbf{K}^T\mathbf{x}$ *corresponds to* $\sum_{i=1}^M k_{ij} x_i$ *so that, in view of* (2.20)*, the right-hand side of equation* (2.35) *can be written as*

$$\|\mathbf{K}^T \mathbf{x}\|^2_{\tilde{\mathbf{\Pi}}_{t-N/t-1}} = \sum_{j=1}^{M} \| \sum_{i=1}^{M} k_{ij} x_i \|^2_{\tilde{\Pi}^j_{t-N/t-1}}$$

Using the triangle inequality we obtain

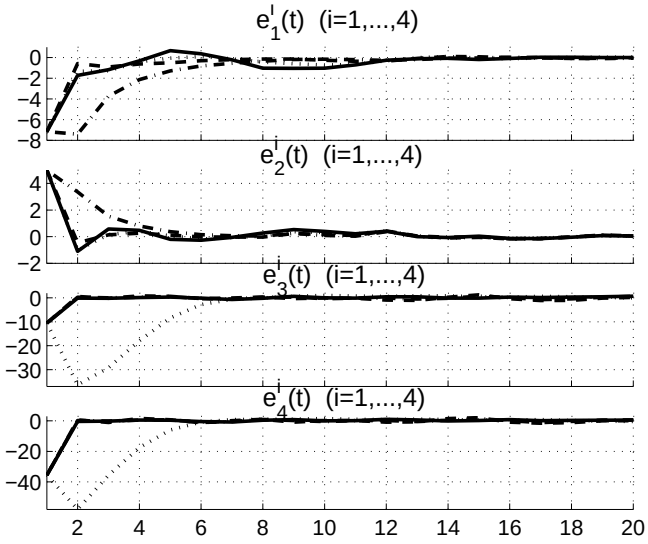
$$\begin{array}{rcl} \sum_{j=1}^M \| \sum_{i=1}^M k_{ij} x_i \|^2_{\tilde{\Pi}^j_{t-N/t-1}} & \leq & \sum_{j=1}^M M_j \sum_{i=1}^M k_{ij}^2 \|x_i\|^2_{\tilde{\Pi}^j_{t-N/t-1}} \\ & = & \sum_{i=1}^M \|x_i\|^2_{\sum_{j=1}^M M_j k_{ij}^2 \tilde{\Pi}^j_{t-N/t-1}} \end{array}$$

which proves that matrices $\Pi^i_{t-N/t-1}$ *verifying* (2.30) *also verify* (2.28)*.*

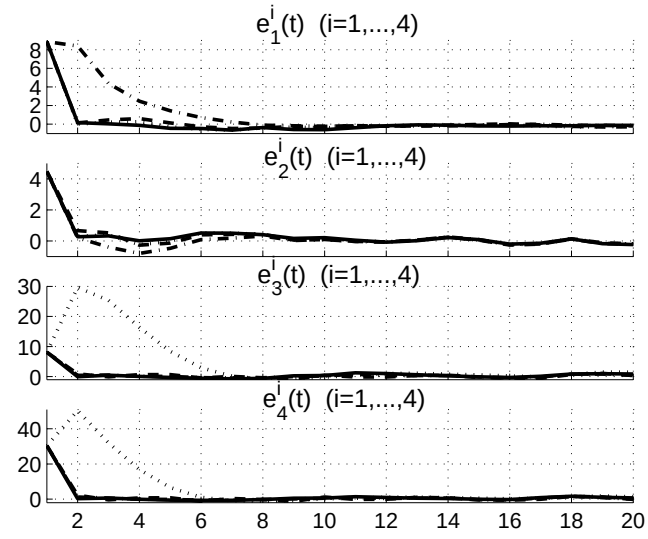
2.7.2 Proof of Theorem 1

The proof of Theorem 1 uses classical results for MHE, [43, 42, 44, 45] and additional results we provide next.

Lemma 1 *If* (2.28) *is satisfied then, for* $z \in \mathbb{X}$, (2.18) *is fulfilled.*



(a) Case A



(b) Case B

Figure 2.4: Components of the estimation error $[e^i_{1,t}, e^i_{2,t}, e^i_{3,t}, e^i_{4,t}]^T = x_t - \hat{x}^i_{t/t}$ of the different sensors. Solid line $i = 1$, dotted line $i = 2$, dashed line $i = 3$, dash-dotted line $i = 4$.

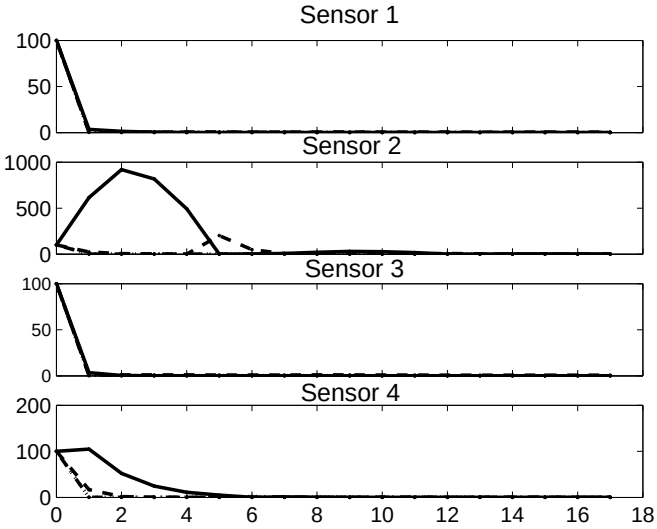


Figure 2.5: Evolution of the four eigenvalues (for each sensor) of the matrices $\Pi^i_{t/t+N-1}$, $i = 1,\dots,4$. Solid line $i = 1$, dotted line $i = 2$, dashed line $i = 3$, dash-dotted line $i = 4$.

Proof 8 (Proof of Lemma 1) *Let* $\mathbf{z} = \mathbb{1}_M \otimes z$. *We define the "unconstrained" transit cost as*

$$\begin{aligned}\Xi^u_{t-N+1/t}(\mathbf{z}) = \min_{\hat{\mathbf{x}}_{t-N},\{\hat{\mathbf{w}}_k\}_{k=t-N}^{t-1}} \{&\mathbf{J}(t-N,t,\hat{\mathbf{x}}_{t-N},\hat{\mathbf{w}},\hat{\tilde{\mathbf{v}}},\mathbf{\Gamma}_{t-N})\\ &\textit{subject to } (2.15a),\ (2.15b) \textit{ and } \hat{\mathbf{x}}_{t-N+1} = \mathbf{z}\}\end{aligned}$$

that, differently from $\Xi_{t-N+1/t}$ *in* (2.17)*, does not account for input and state constraints. Notice that*

$$\Xi^u_{t-N+1/t}(\mathbf{z}) = \sum_{i=1}^{M} \Xi^{i,u}_{t-N+1/t}(z) \tag{2.36}$$

where the unconstrained transit cost associated to sensor i is

$$\begin{aligned}\Xi^{i,u}_{t-N+1/t}(z) = \min_{\hat{x}^i_{t-N},\{\hat{w}^i_k\}_{k=t-N}^{t-1}} &\{J^i(t-N,t,\hat{x}^i_{t-N},\hat{w}^i,\hat{\tilde{v}}^i,\Gamma^i_{t-N})\\ &\textit{subject to } (2.7a),\ (2.7b) \textit{ and } \hat{x}^i_{t-N+1} = z\}\end{aligned} \tag{2.37}$$

We first compute explicitly $\Xi^{i,u}_{t-N+1/t}(z)$. *Recalling* (2.7) *we can write*

$$\bar{V}_i^{[t-N+1,t]} = \bar{Y}_i^{[t-N+1,t]} - \bar{\mathscr{O}}^i_N \hat{x}^i_{t-N+1} - \mathscr{C}^i_N W_i^{[t-N+1,t-1]}$$

where matrices $\mathscr{C}^i_N$ *and* $\bar{\mathscr{O}}^i_N$ *are defined in* (2.21) *and* (2.5)*, respectively,* $\bar{V}_i^{[t-N+1,t]} = [(\vec{\tilde{v}}^i_{t-N+1})^T,\dots,(\vec{\tilde{v}}^i_t)^T]^T$, $\bar{Y}_i^{[t-N+1,t]} = [(\vec{y}^i_{t-N+1})^T,\dots,(\vec{y}^i_t)^T]^T$, *and*

$W_i^{[t-N+1,t-1]} = [(\hat{w}_{t-N+1}^i)^T, \ldots, (\hat{w}_{t-1}^i)^T]^T$. *We can rewrite the i-th sensor's cost function as*

$$\begin{aligned} 2\,(J^i - \Theta_{t-1}^{*i}) &= \|\bar{Y}_i^{[t-N+1,t]} - \bar{\mathscr{O}}_N^i \hat{x}_{t-N+1}^i - \mathscr{C}_N^i W_i^{[t-N+1,t-1]}\|^2_{(\bar{R}_N^i)^{-1}} + \\ &+ \|W_i^{[t-N+1,t-1]}\|^2_{(Q_{N-1})^{-1}} + \|\hat{w}_{t-N}^i\|^2_{Q^{-1}} + \\ &+ \underbrace{\|\hat{x}_{t-N}^i - \sum_{j=1}^{M} k_{ij}\hat{x}_{t-N/t-1}^j\|^2_{(\Pi^i_{t-N/t-1})^{-1}} + \|\bar{y}_{t-N}^i - \bar{C}^i\hat{x}_{t-N}^i\|^2_{(\bar{R}^i)^{-1}}}_{(*)} \end{aligned} \tag{2.38}$$

where matrices $\bar{R}_N^i$ *and* Q_{N-1}^i *are defined in* (2.22) *and* (2.23)*, respectively. Minimizing the partial cost* $(*)$ *in* (2.38) *with respect to* $\hat{x}_{t-N}^i$ *gives*

$$\hat{x}_{t-N/t-1}^{*i} = \Pi_{t-N/t-1}^{*i}\Big[(\Pi^i_{t-N/t-1})^{-1}\sum_{j=1}^{M} k_{ij}\hat{x}^j_{t-N/t-1} + (\bar{C}^i)^T(\bar{R}^i)^{-1}\bar{y}^i_{t-N}\Big]$$

with $\Pi_{t-N/t-1}^{*i}$ *as in* (2.26)*. Therefore one has that the term* $(*)$ *in* (2.38) *is equal to*

$$\|\hat{x}_{t-N}^i - \hat{x}_{t-N/t-1}^{*i}\|^2_{(\Pi^{*i}_{t-N/t-1})^{-1}} \tag{2.39}$$

up to a constant term.

We denote with $L^i(\cdot)$ *the minimum of* $J^i(\cdot)$ *with respect to vector* $W_i^{[t-N+1,t-1]}$, i.e.,

$$L^i = \min_{\{\hat{w}_k\}_{k=t-N+1}^{t-1}} J^i(\cdot) \tag{2.40}$$

We compute $\frac{\partial J^i(\cdot)}{\partial W_i^{[t-N+1,t-1]}} = 0$*. The vector* $W_i^{[t-N+1,t-1],opt}$ *which solves the minimization problem* (2.40) *is*

$$\begin{aligned} W_i^{[t-N+1,t-1],opt} &= \left((\mathscr{C}_N^i)^T(\bar{R}_N^i)^{-1}\mathscr{C}_N^i + (Q_{N-1}^i)^{-1}\right)^{-1} \times \\ &\times (\mathscr{C}_N^i)^T(\bar{R}_N^i)^{-1}\left(\bar{Y}_i^{[t-N+1,t]} - \bar{\mathscr{O}}_N^i\hat{x}_{t-N+1}^i\right) \end{aligned} \tag{2.41}$$

Replacing (2.41) *into* (2.38) *and using* (2.39) *one obtains*

$$\begin{aligned} L^i \;=\; & \|\bar{Y}_i^{[t-N+1,t]} - \bar{\mathscr{O}}_N^i\hat{x}_{t-N+1}^i\|^2_{(\bar{R}_N^i + \mathscr{C}_N^i Q_{N-1}(\mathscr{C}_N^i)^T)^{-1}} + \\ & + \|\hat{w}_{t-N}^i\|^2_{Q^{-1}} + \|\hat{x}_{t-N}^i - \hat{x}_{t-N/t-1}^{*i}\|^2_{(\Pi^{*i}_{t-N/t-1})^{-1}} \end{aligned}$$

up to an additive constant term. The solution of the optimization problem (2.37) *can be computed through a Kalman filter recursion with respect to the modified dynamical system*

$$\left\{\begin{array}{lcl} \hat{x}_{t-N+1}^i & = & A\hat{x}_{t-N}^i + w_{t-N} \\ \bar{Y}_i^{[t-N+1,t]} & = & \bar{\mathscr{O}}_N^i\hat{x}_{t-N+1}^i + \bar{V}_i^{[t-N+1,t]} \end{array}\right. \tag{2.42}$$

where w_t *has covariance equal to* Q*, the covariance of* $\bar{V}_i^{[t-N+1,t]}$ *is* $\bar{R}_N^i + \mathscr{C}_N^i Q_{N-1}(\mathscr{C}_N^i)^T$*, and* $\Pi_{t-N/t-1}^{*i}$ *in* (2.26) *is the uncertainty of the initial condition guess. In this way we can write the unconstrained transit cost as follows (see [22])*

$$\Xi_{t-N+1/t}^{i,u}(z) = \frac{1}{2}\|z - \hat{x}_{t-N+1/t}^{i,u}\|^2_{(\bar{\Pi}^i_{t-N+1/t})^{-1}} + \Theta_t^{*,i,u} \tag{2.43}$$

where $\hat{x}^{i,u}_{t-N+1/t}$ *minimizes the unconstrained problem, and* $\Theta^{*,i,u}_t$ *is the optimal solution of the unconstrained minimization problem, and* $\tilde{\Pi}^i_{t-N+1/t}$ *is computed as in* (2.27)*. Remark that the regionally unobservable subspaces of system* (2.42) *and system* (2.1)-(2.2) *coincide.*

From (2.43) *and* (2.36) *one has that*

$$\Xi^u_{t-N+1/t}(\mathbf{z}) = \frac{1}{2}\|\mathbf{z}-\hat{\mathbf{x}}^u_{t-N+1/t}\|^2_{(\tilde{\mathbf{\Pi}}_{t-N+1/t})^{-1}} + \mathbf{\Theta}^{*,u}_t \tag{2.44}$$

where $\mathbf{\Theta}^{*,u}_t = \sum_{i=1}^M \Theta^{*,i,u}_t$ *and* $\tilde{\mathbf{\Pi}}_{t-N+1/t} = diag(\tilde{\Pi}^1_{t-N+1/t},...,\tilde{\Pi}^M_{t-N+1/t})$*. We also define* $\mathbf{\Theta}_{t-N+1/t}(\mathbf{x};\hat{\mathbf{x}}^u_{t-N+1/t}) = \frac{1}{2}\|\mathbf{x}-\hat{\mathbf{x}}^u_{t-N+1/t}\|^2_{(\tilde{\mathbf{\Pi}}_{t-N+1/t})^{-1}}$ *in such a way that* $\Xi^u_{t-N+1/t}(\mathbf{z}) = \mathbf{\Theta}_{t-N+1/t}(\mathbf{z};\hat{\mathbf{x}}^u_{t-N+1/t}) + \mathbf{\Theta}^{*,u}_t$*. Let us finally consider the case of constrained estimation. Following the rationale of the proof of Lemma 4 in [44] one has that, since* $\mathbf{z}$ *lies in the feasibility region by assumption, one obtains* (2.19)*. Notice that the initial penalty term* $\mathbf{\Gamma}_{t-N+1}(\cdot)$*, computed as in* (2.13) *in* $\mathbf{z}$*, is*

$$\begin{aligned}\mathbf{\Gamma}_{t-N+1}(\mathbf{z};\hat{\mathbf{x}}_{t-N+1/t}) &= \frac{1}{2}\|\mathbf{z}-\mathbf{K}\hat{\mathbf{x}}_{t-N+1/t}\|^2_{\mathbf{\Pi}^{-1}_{t-N+1/t}} + \mathbf{\Theta}^*_t = \\ &= \frac{1}{2}\|\mathbf{z}-\hat{\mathbf{x}}_{t-N+1/t}\|^2_{\mathbf{K}^T\mathbf{\Pi}^{-1}_{t-N+1/t}\mathbf{K}} + \mathbf{\Theta}^*_t\end{aligned} \tag{2.45}$$

where the second equality holds because $\mathbf{Kz} = \mathbf{z}$*.*

Using Schur complement, the LMI (2.28) *is equivalent to*

$$\begin{bmatrix}\mathbf{\Pi}_{t-N/t-1} & \mathbf{K} \\ \mathbf{K}^T & \tilde{\mathbf{\Pi}}^{-1}_{t-N/t-1}\end{bmatrix} \geq 0 \tag{2.46}$$

and, being matrices $\mathbf{\Pi}_{t-N/t-1}$ *and* $\tilde{\mathbf{\Pi}}_{t-N/t-1}$ *positive definite,* (2.46) *is equivalent to*

$$\mathbf{K}^T\mathbf{\Pi}^{-1}_{t-N/t-1}\mathbf{K} \leq \tilde{\mathbf{\Pi}}^{-1}_{t-N/t-1} \tag{2.47}$$

From (2.19) *and* (2.45), (2.47) *implies* (2.18)*.*

Lemma 2 *If* (2.28) *is satisfied, then*

$$\Theta^*_t \leq \mathbf{\Gamma}_0(\mathbf{x}_0;\mathbf{x}_{0/0}) \text{ for all } t \geq 0 \tag{2.48}$$

where $\mathbf{x}_0 = [x_0^T \ \dots \ x_0^T]^T \in \mathbb{R}^{nM}$ *and* $\mathbf{x}_{0/0} = \mathbb{1}_M \otimes \mu$*.*

Proof 9 (Proof of Lemma 2) *First notice that, in view of Definition 2, the sequence* $\mathbf{x}_\Sigma(t,x_0)$ *verifies the constraints* (2.15d)*. In view of Lemma 1, equation* (2.18) *holds for* $\mathbf{z} = \mathbf{x}_\Sigma(t,x_0)$*, for all* t*. By optimality, we have*

$$\Theta^*_t \leq \Xi_{t-N+1/t}(\mathbf{x}_\Sigma(t-N+1,x_0)),\ \forall t \geq 0$$

Furthermore

$$\Xi_{t-N+1/t}(\mathbf{x}_\Sigma(t-N+1,x_0)) \leq \mathbf{J}(t-N,t,\mathbf{x}_\Sigma(t-N,x_0),0,0,\mathbf{\Gamma}_{t-N})$$

Note that, from (2.14)*, one has* $\mathbf{J}(t-N,t,\mathbf{x}_\Sigma(t-N,x_0),0,0,\mathbf{\Gamma}_{t-N}) = \mathbf{\Gamma}_{t-N}(\mathbf{x}_\Sigma(t-N,x_0);\hat{\mathbf{x}}_{t-N/t-1})$ *and in view of (2.18),* $\Theta^*_t \leq \Xi_{t-N/t-1}(\mathbf{x}_\Sigma(t-N,x_0)) \leq \mathbf{\Gamma}_{t-N-1}(\mathbf{x}_\Sigma(t-N-1,x_0);\hat{\mathbf{x}}_{t-N-1/t-2})$*. We can further iterate this procedure in order to prove* (2.48)*.*

Lemma 3 *Assume that* (a) $N \geq n-1$, *with* $N \geq 1$, (b) $\exists \bar{\Pi}$ *such that* $\Pi^i_{t-N/t-1} < \bar{\Pi}$, *for all* t, *for all* $i \in \mathscr{V}$, *and* (c)

$$\max_{k=t-N,\ldots,t} (\|\tilde{\hat{\mathbf{v}}}_{k/t}\|, \|\hat{\mathbf{w}}_{k/t}\|, \mathbf{\Gamma}^o_{t-N}(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1})) \xrightarrow{t\to\infty} 0 \tag{2.49}$$

Then the dynamics of the state estimation error provided by the DMHE scheme is given by (2.32).

Proof 10 (Proof of Lemma 3) *In the noiseless case, for any sensor* $i \in \mathscr{V}$ *at any* t, *the output signal is* $\vec{y}^i_k = \bar{C}^i x_\Sigma(t, x_0)$. *Similarly to Lemma 4.3 in [43],*

$$\begin{aligned} &\sum_{k=t-N}^{t} \|\hat{\vec{v}}^i_{k/t}\| = \sum_{k=t-N}^{t} \|\vec{y}^i_k - \bar{C}^i \hat{x}^i_{k/t}\| \geq \\ &\geq \sum_{k=t-N}^{t} (\|\vec{y}^i_k - \bar{C}^i A^{k-(t-N)} \hat{x}^i_{t-N/t}\| - \|\bar{C}^i A^{k-(t-N)} \hat{x}^i_{t-N/t} - \bar{C}^i \hat{x}^i_{k/t}\|) \end{aligned} \tag{2.50}$$

The first term at the right-hand side of (2.50) *is*

$$\sum_{k=t-N}^{t} \|\vec{y}^i_k - \bar{C}^i A^{k-(t-N)} \hat{x}^i_{t-N/t}\| = \|\bar{\mathscr{O}}^i_{N+1}(x_\Sigma(t-N, x_0) - \hat{x}^i_{t-N/t})\| \tag{2.51}$$

where $\bar{\mathscr{O}}^i_{N+1}$ *is the "extended" regional observability matrix of* $N+1$ *rows defined by replacing* n *with* $N+1$ *in* (2.5). *From* (2.7), *one has*

$$\hat{x}^i_{k/t} = A^{k-(t-N)} \hat{x}^i_{t-N/t} + \sum_{j=1}^{k-(t-N)} A^j \hat{w}^i_{k-j/t}$$

The second term at the right-hand side of (2.50) *can be bounded as*

$$\begin{aligned} &\Sigma_{k=t-N}^{t} \|\bar{C}^i A^{k-(t-N)} \hat{x}^i_{t-N/t} - \bar{C}^i \hat{x}^i_{k/t}\| \leq \\ &\leq \Sigma_{k=t-N}^{t} \|\bar{C}^i \Sigma_{j=1}^{k-(t-N)} A^j \hat{w}^i_{k-j/t}\| \\ &\leq \|\bar{C}^i\| \Sigma_{k=t-N}^{t} \Sigma_{j=1}^{k-(t-N)} \|A\|^j \|\hat{w}^i_{k-j/t}\| \end{aligned} \tag{2.52}$$

By replacing equations (2.51) *and* (2.52) *into* (2.50), *one obtains*

$$\begin{aligned} &\|\bar{\mathscr{O}}^i_{N+1}(\hat{x}^i_{t-N/t} - x_\Sigma(t-N, x_0))\| \leq \\ &\leq \sum_{k=t-N}^{t} \|\hat{\vec{v}}^i_{k/t}\| + \|\bar{C}^i\| \sum_{k=t-N}^{t} \sum_{j=1}^{k-(t-N)} \|A\|^j \|\hat{w}^i_{k-j/t}\| \end{aligned} \tag{2.53}$$

Note that the matrix $\bar{\mathscr{O}}^i_{N+1}$ *at the left-hand side of* (2.53) *selects the observable part of* $(\hat{x}^i_{t-N/t} - x_\Sigma(t-N, x_0))$. *Therefore, from* (2.49), *equation* (2.53) *leads to*

$$\|\mathbf{P}_0(\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N, x_0))\| \xrightarrow{t\to\infty} 0, \tag{2.54}$$

$$\mathbf{\Gamma}^o_{t-N}(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) \xrightarrow{t\to\infty} 0 \tag{2.55}$$

In view of assumption (b), *it follows that*

$$\mathbf{\Gamma}^o_{t-N}(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) \geq \|\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N, x_0)\|_{diag(\bar{\Pi}^{-1},\ldots,\bar{\Pi}^{-1})}$$

Hence, from (2.55)

$$\|\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N,x_0)\| \xrightarrow{t\to\infty} 0 \tag{2.56}$$

Note that, for $k = t-N,\ldots,t-1$

$$\hat{\mathbf{x}}_{k+1/t} = \mathbf{A}\hat{\mathbf{x}}_{k/t} + \hat{\mathbf{w}}_{k/t} \tag{2.57}$$

and that, in view of (2.49)*, one has* $\hat{\mathbf{w}}_{k/t} \to 0$ *as* $t \to \infty$*. Therefore one also has*

$$\hat{\mathbf{x}}_{k+1/t} - \mathbf{A}\hat{\mathbf{x}}_{k/t} \to 0 \text{ as } t \to \infty \tag{2.58}$$

From now on, we introduce, for simplicity of notation, terms α_t^j *to indicate asymptotically vanishing variables,* i.e., $\|\alpha_t^j\| \xrightarrow{t\to\infty} 0$*, for all* $j \in \mathscr{V}$*. Formulae* (2.54)*,* (2.56) *and* (2.57) *are equivalent to*

$$\mathbf{P}_O\hat{\mathbf{x}}_{t-N/t} = \mathbf{P}_O\mathbf{x}_\Sigma(t-N,x_0) + \alpha_t^1 \tag{2.59a}$$
$$\hat{\mathbf{x}}_{t-N/t} = \mathbf{K}\hat{\mathbf{x}}_{t-N/t-1} + \alpha_t^2 \tag{2.59b}$$
$$\hat{\mathbf{x}}_{t-N+1/t} = \mathbf{A}\hat{\mathbf{x}}_{t-N/t} + \alpha_t^3 \tag{2.59c}$$

Recall that, by definition, $\mathbf{P}_O + \mathbf{P}_{NO} = I$*. Therefore,*

$$\hat{\mathbf{x}}_{t-N/t} = \mathbf{P}_O\hat{\mathbf{x}}_{t-N/t} + \mathbf{P}_{NO}\hat{\mathbf{x}}_{t-N/t} \tag{2.60}$$

In (2.60)*, we replace terms* $\mathbf{P}_O\hat{\mathbf{x}}_{t-N/t}$ *and* $\mathbf{P}_{NO}\hat{\mathbf{x}}_{t-N/t}$ *according to* (2.59a) *and* (2.59b)*, premultiplied by* $\mathbf{P}_{NO}$*, respectively, we get*

$$\hat{\mathbf{x}}_{t-N/t} = \mathbf{P}_O\mathbf{x}_\Sigma(t-N,x_0) + \mathbf{P}_{NO}\mathbf{K}\hat{\mathbf{x}}_{t-N/t-1} + \alpha_t^4 \tag{2.61}$$

Since $\mathbf{P}_O + \mathbf{P}_{NO} = I$*, we write* $\mathbf{P}_O\mathbf{x}_\Sigma(t-N,x_0) = \mathbf{x}_\Sigma(t-N,x_0) - \mathbf{P}_{NO}\mathbf{x}_\Sigma(t-N,x_0)$*, and obtain*

$$\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N,x_0) = \mathbf{P}_{NO}(\mathbf{K}\hat{\mathbf{x}}_{t-N/t-1} - \mathbf{x}_\Sigma(t-N,x_0)) + \alpha_t^4 \tag{2.62}$$

First recall that, since $\mathbf{K}$ *is stochastic and* $\mathbf{x}_\Sigma(t-N,x_0) = \mathbb{1}_M \otimes x_\Sigma(t-N,x_0)$*,* $\mathbf{K}\mathbf{x}_\Sigma(t-N,x_0) = \mathbf{x}_\Sigma(t-N,x_0)$*. Then notice that* $\mathbf{x}_\Sigma(t-N,x_0) = \mathbf{A}\mathbf{x}_\Sigma(t-N-1,x_0)$*. From* (2.59c) *one obtains*

$$\begin{aligned}\boldsymbol{\varepsilon}_{t-N/t} &= \mathbf{P}_{NO}\mathbf{K}\mathbf{A}\boldsymbol{\varepsilon}_{t-N-1/t-1} + \alpha_t^5 \\ &= \mathbf{P}_{NO}\mathbf{K}\mathbf{A}(\mathbf{P}_{NO} + \mathbf{P}_O)\boldsymbol{\varepsilon}_{t-N/t} + \alpha_t^5\end{aligned} \tag{2.63}$$

Equation (2.59a) *implies that the term* $\mathbf{P}_O\boldsymbol{\varepsilon}_{t-N/t}$ *is asymptotically vanishing and equation* (2.32) *follows from* (2.63)*.*

Proof 11 (Proof of Theorem 1) *By direct calculation, for all* $t \geq 0$ *one has*

$$\Theta_t^* - \Theta_{t-1}^* = \tfrac{1}{2}\textstyle\sum_{k=t-N}^{t} \|\hat{\tilde{\mathbf{v}}}_{k/t}\|^2_{\mathbf{R}^{-1}} + \tfrac{1}{2}\sum_{k=t-N}^{t-1} \|\hat{\mathbf{w}}_{k/t}\|^2_{\mathbf{Q}^{-1}} + \boldsymbol{\Gamma}^o_{t-N}(\hat{\mathbf{x}}_{t-N/t};\hat{\mathbf{x}}_{t-N/t-1})$$

Furthermore, (2.48) *follows from Lemma 2 and* (2.28)*. Therefore it follows that* $\frac{1}{2}\sum_{k=t-N}^{t} \|\hat{\tilde{\mathbf{v}}}_{k/t}\|^2_{\mathbf{R}^{-1}} + \frac{1}{2}\sum_{k=t-N}^{t-1} \|\hat{\mathbf{w}}_{k/t}\|^2_{\mathbf{Q}^{-1}} + \boldsymbol{\Gamma}^o_{t-N}(\hat{\mathbf{x}}_{t-N/t};\hat{\mathbf{x}}_{t-N/t-1}) \xrightarrow{t\to\infty} 0$ *and hence* (2.49) *holds. This, in turn, implies (using Lemma 3) that the dynamics of state estimation error provided by the DMHE scheme is given by* (2.32)*.*

Furthermore, from (2.32)*, convergence of the error to zero is guaranteed if* Φ *is Schur.*

2.7.3 Proof of Theorem 2 and Corollary 1

Proof 12 (Proof of Theorem 2) *If the graph $\mathcal{G}$ is not strongly connected it can be partitioned into k irreducible subgraphs $\mathcal{G}_1^*$, $\mathcal{G}_2^*$, ..., $\mathcal{G}_k^*$ of cardinality m_1, ..., m_k, and $\sum_{i=1}^k m_i = M$. Without loss of generality, (i.e. by permuting sensor indexes) the matrix K can be brought in a block lower triangular form (with k square diagonal blocks K_{11},...K_{kk}, of dimensions m_1, ..., m_k, respectively).*

Notice that the block K_{ii} is stochastic if and only if $K_{ij} = 0$ for $j < i$. In this case, the nodes of the subgraph $\mathcal{G}_i^*$ have no neighbors belonging to other subgraphs and $\mathcal{G}_i^*$ is isolated*. Moreover, if a subgraph $\mathcal{G}_i^*$ is isolated, the block K_{ii} is stochastic and it has a single Frobenius eigenvalue equal to 1. On the other hand, if a graph $\mathcal{G}_i^*$ is not isolated, K_{ii} is irreducible but not stochastic (specifically, the sum of the entries of at least a row is smaller than* 1*) and its Frobenius eigenvalue has absolute value smaller than* 1[1]*. Notice that the eigenvalues of K are the eigenvalues of K_{11},...K_{kk}. So, the number of eigenvalues of K equal to* 1 *equals the number of isolated graphs in the network.*

Note that $\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{A}_{KO} = diag(A_{KO}^1, \ldots, A_{KO}^M)$ where A_{KO}^i is the "regional" observability Kalman decomposition of A associated to sensor i, that is

$$A_{KO}^i = \begin{bmatrix} A_O^i & 0 \\ A_{21}^i & A_{NO}^i \end{bmatrix} \tag{2.64}$$

Since $\mathbf{P}_{NO} = \mathbf{T}\mathbf{S}_{NO}\mathbf{S}_{NO}^T\mathbf{T}^{-1}$ one has

$$\Phi = \mathbf{P}_{NO}\mathbf{K}\tilde{\mathbf{A}}$$

where $\tilde{\mathbf{A}} = \mathbf{T}\tilde{\mathbf{A}}_{KO}\mathbf{T}^{-1}$, $\tilde{\mathbf{A}}_{KO} = diag(\tilde{A}_{KO}^1, \ldots, \tilde{A}_{KO}^M)$, and

$$\tilde{A}_{KO}^i = \begin{bmatrix} 0 & 0 \\ 0 & A_{NO}^i \end{bmatrix}$$

Now we prove that $\mathbf{x} \in span(e_1 \otimes v_A^i, \ldots, e_M \otimes v_A^i)$ is not an eigenvector of Φ associated to a non-zero eigenvalue λ_A^i. In general, given a vector $\alpha \in \mathbb{R}^M$, with $\alpha \neq 0$, one has that the eigenvector $\mathbf{x}$ of $\mathbf{A}$ can be written as $\mathbf{x} = \alpha \otimes v_A^i$. We obtain

$$\tilde{\mathbf{A}}\mathbf{x} = diag(\tilde{A}^1, \ldots, \tilde{A}^M) \begin{bmatrix} \alpha_1 v_A^i \\ \vdots \\ \alpha_M v_A^i \end{bmatrix} = \begin{bmatrix} \alpha_1 \tilde{A}^1 v_A^i \\ \vdots \\ \alpha_M \tilde{A}^M v_A^i \end{bmatrix}$$

By construction, $\tilde{A}^j v_A^i = \lambda_A^i v_A^i$ if v_A^i belongs to the regionally unobservable subspace of sensor j. Otherwise $\tilde{A}^j v_A^i = 0$. We write, in general $\tilde{A}^j v_A^i = f_{ij}\lambda_A^i v_A^i$, where

$$f_{ij} = \begin{cases} 1 & \text{if } v_A^i \text{ belongs to the regionally unobservable} \\ & \text{subspace of sensor } j \\ 0 & \text{otherwise.} \end{cases}$$

Defining $f_i = [f_{i1}, \ldots, f_{iM}]^T$, we can write

$$\tilde{\mathbf{A}}\mathbf{x} = \lambda_A^i (diag(f_i)\alpha) \otimes v_A^i$$

[1] This follows from the third Gershgorin theorem [53], dealing with irreducible matrices. Specifically, an eigenvalue of an irreducible matrix (in our case K_{ii}), which is on the boundary of a Gershgorin circle, is located on the boundary of all the Gershgorin circles. Since there is at least a row of K_{ii} such that the sum of its entries is smaller than 1, 1 cannot be an eigenvalue of K_{ii}, and hence all the eigenvalues of K_{ii} are strictly inside the unit circle (from the first Gershgorin theorem).

From $\mathbf{K} = K \otimes I_n$ *we obtain*

$$\Phi\mathbf{x} = \lambda_A^i \mathbf{P}_{NO}(K \otimes I_n)\left[(diag(f_i)\alpha) \otimes v_A^i\right]$$

Recall that $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ *and hence*

$$\Phi\mathbf{x} = \lambda_A^i \mathbf{P}_{NO}(Kdiag(f_i)\alpha) \otimes v_A^i = \lambda_A^i (diag(f_i)Kdiag(f_i)\alpha) \otimes v_A^i$$

Finally, we obtain

$$\Phi(\alpha \otimes v_A^i) = \lambda_A^i (diag(f_i)Kdiag(f_i)\alpha) \otimes v_A^i$$

from which it is apparent that $\alpha \otimes v_A^i$ *is an eigenvector of* Φ*, with eigenvalue* $\lambda_A^i \neq 0$ *if and only if* $diag(f_i)Kdiag(f_i)\alpha = \alpha$*. Moreover, there exists* α *satisfying the previous equation if and only if* $diag(f_i)Kdiag(f_i)$ *has at least one eigenvector equal to* 1*. This occurs if and only if* $f_{ij} = 1$ *for all* j *belonging to an isolated subgraph. This means that all the sensors of an isolated subgraph have at least a common regionally unobservable eigenvector. Hence,* $\mathbf{x} \in span(e_1 \otimes v_A^i, ..., e_M \otimes v_A^i)$ *can not be an eigenvector of* Φ *if* (2.33) *holds. This completes the proof.*

Proof 13 (Proof of Corollary 1) *Recalling Definition 1, collective observability holds if and only if the observability matrix* $\mathbf{O}^*$ *of the pair* $(A, \mathbf{C}^*)$ *is such that*

$$\ker(\mathbf{O}^*) = 0 \tag{2.65}$$

Notice that, up to a permutation of the rows of $\mathbf{O}^*$*, we have* $\left[(\bar{\mathscr{O}}_n^1)^T \quad \dots \quad (\bar{\mathscr{O}}_n^M)^T\right]^T$*. Therefore* (2.65) *is equivalent to*

$$\bigcap_{i \in \mathscr{V}} \ker(\bar{\mathscr{O}}_n^i) = 0$$

which is equivalent to (2.33) *when the graph is strongly connected. This concludes the proof.*

2.7.4 Proof of Theorem 3

To prove Theorem 3, a number of intermediate results are needed. First, we address the problem of the stability of Riccati equations with respect to perturbations. This problem has been scarcely explored in the literature, with the exception of [52] where stability is proved with respect to small perturbations. In the following, we explore the issue under the lead of Theorem 4.1 in [13], and provide global stability results.

Given a pair (A,C), and matrices $Q \geq 0$, $R > 0$ of appropriate size, consider the following Riccati equation, affected by an exogenous perturbation term Δ_k

$$\Pi_{k+1}^o = (A - G_k^o C)(\Pi_k^o + \Delta_k)(A - G_k^o C)^T + Q + G_k^o R (G_k^o)^T \tag{2.66}$$

where Π_0^o is the initial condition and matrix G_k^o is the Kalman gain

$$G_k^o = A(\Pi_k^o + \Delta_k)C^T \left(C(\Pi_k^o + \Delta_k)C^T + R\right)^{-1} \tag{2.67}$$

Assuming that the pair (A,C) is detectable and that the pair $(A, \sqrt{Q})$ is stabilizable, there exists a unique solution $\bar{\Pi} \geq 0$ of the algebraic Riccati equation associated to (2.66) with $\Delta_k = 0$. In the sequel, we will denote with $\overrightarrow{X}_\tau$ the sequence of matrices X_k, with $k = 0, \dots, \tau$. In [30] the following definition of $\mathscr{L}$-stability of system (2.66) is given.

Definition 3 *System* (2.66) *is $\mathscr{L}$-stable from input Δ_k if, for a given norm $\mathscr{L}$, there exist $\gamma > 0$ and $\beta > 0$ such that*

$$\|\overrightarrow{\Pi}_\tau - \overline{\Pi}\|_{\mathscr{L}} \leq \gamma \|\overrightarrow{\Delta}_\tau\|_{\mathscr{L}} + \beta \,,\ \forall \Delta \in \mathscr{L}, \forall \tau \in [0,\infty).$$

From now on, we denote with $\|\overrightarrow{X}_\tau\|_{\mathscr{L}_\infty}$ the ∞-norm of the sequence $\|X_t\|_2$, with $t = 0,\ldots,\tau$.

Lemma 4 *Given a detectable pair (A,C), system* (2.66) *is $\mathscr{L}_\infty$-stable from a positive semi-definite input $\Delta_k \geq 0$.*

Proof 14 *We define a sequence Π_k (with $\Pi_0 = \Pi_0^o$) as follows*

$$\Pi_{k+1} = (A - GC)(\Pi_k + \Delta_k)(A - GC)^T + Q + GRG^T, \tag{2.68}$$

where G is an arbitrary gain such that $F = A - GC$ is Hurwitz. Notice that G always exists, since (A,C) is detectable. From (2.68)*, we obtain, for $k \geq 1$,*

$$\Pi_{k+1} - \Pi_k = F(\Pi_k - \Pi_{k-1})F^T + F(\Delta_k - \Delta_{k-1})F^T$$

and hence, for $i \geq 1$,

$$\Pi_{i+1} - \Pi_i = F^i(\Pi_1 - \Pi_0)(F^T)^i + \sum_{j=1}^{i} F^j(\Delta_{i+1-j} - \Delta_{i-j})(F^T)^j$$

Then, for $k > 1$,

$$\begin{array}{rcl} \Pi_k & = & \Pi_0 + \sum_{i=0}^{k-1}(\Pi_{i+1} - \Pi_i) \\ & = & \sum_{i=0}^{k-1} F^i(\Pi_1 - \Pi_0)(F^T)^i + \Pi_0 + \\ & & + \sum_{i=1}^{k-1}\sum_{j=1}^{i} F^j(\Delta_{i+1-j} - \Delta_{i-j})(F^T)^j \end{array} \tag{2.69}$$

Notice that, assuming $\Delta_0 = 0$ in (2.69) *one has*

$$\sum_{i=1}^{k-1}\sum_{j=1}^{i} F^j(\Delta_{i+1-j} - \Delta_{i-j})(F^T)^j = \sum_{i=1}^{k-1} F^i \Delta_{k-i}(F^T)^i$$

and (2.69) *gives*

$$\Pi_k = \sum_{i=0}^{k-1} F^i(\Pi_1 - \Pi_0)(F^T)^i + \Pi_0 + \sum_{i=1}^{k-1} F^i \Delta_{k-i}(F^T)^i \tag{2.70}$$

Let us set $\|\Pi_1 - \Pi_0\|_2 = \alpha$ and $\|\overrightarrow{\Delta}_\infty\|_{\mathscr{L}_\infty} = \bar{\delta}$. Since F is Hurwitz, there exists $\mu > 0$ and $0 < \nu < 1$ such that $\|F^i\|_2 \leq \mu\nu^i$. Remark that, since $\Delta_k \geq 0$, from optimality of Π_k^o [13] one has $0 \leq \Pi_k^o \leq \Pi_k$, $\forall k \geq 0$, and hence $\|\Pi_k^o\|_2 \leq \|\Pi_k\|_2$. Furthermore, from (2.70)

$$\begin{array}{rcl} \|\Pi_k^o\|_2 \leq \|\Pi_k\|_2 & \leq & \sum_{i=0}^{k-1} \|F^i\|_2^2 \|\Pi_1 - \Pi_0\|_2 + \|\Pi_0\|_2 + \\ & & + \sum_{i=1}^{k-1} \|F^i\|_2^2 \|\Delta_{k-i}\|_2 \\ & \leq & \alpha\mu^2 \sum_{i=0}^{k-1} \nu^{2i} + \|\Pi_0\|_2 + \bar{\delta}\mu^2 \sum_{i=1}^{k-1} \nu^{2i} \\ & \leq & \alpha\mu^2 \frac{1-\nu^{2k}}{1-\nu^2} + \|\Pi_0\|_2 + \bar{\delta}\mu^2\nu^2 \frac{1-\nu^{2(k-1)}}{1-\nu^2} \\ & \leq & \alpha\mu^2 \frac{1}{1-\nu^2} + \|\Pi_0\|_2 + \mu^2\nu^2 \frac{1}{1-\nu^2}\bar{\delta} \end{array} \tag{2.71}$$

The proof is concluded by applying Definition 3 with $\beta = \alpha\mu^2 \frac{1}{1-\nu^2} + \|\Pi_0\|_2 + \|\overline{\Pi}\|_2$ and $\gamma = \mu^2\nu^2 \frac{1}{1-\nu^2}$.

The proof of Theorem 3 can now be completed by applying Lemma 4 and the small gain result for interconnected systems reported in [17].

Proof 15 (Proof of Theorem 3) *First we show, by applying* (2.27) *and* (2.31)*, that it holds that:*

$$\tilde{\Pi}^i_{t-N/t-1} \leq \mathscr{R}^i(\sum_{j=1}^{M} M_j k_{ij}^2 \tilde{\Pi}^j_{t-N-1/t-2}; Q, \bar{R}^{*i}_N) \tag{2.72}$$

From (2.26)*, one has* $\Pi^{*i}_{t-N-1/t-2} \leq \Pi^i_{t-N-1/t-2}$*. Then, by applying* (2.27) *and by optimality [13],*

$$\tilde{\Pi}^i_{t-N/t-1} = \mathscr{R}^i\left(\Pi^{*i}_{t-N-1/t-2}; Q, \bar{R}^{*i}_N\right) \leq \mathscr{R}^i\left(\Pi^i_{t-N-1/t-2}; Q, \bar{R}^{*i}_N\right)$$

and from (2.31) *we obtain* (2.72)*. Now, with reference to the i-th sensor characterized by the pair* $(A, \bar{\mathscr{O}}^i_N)$*, we define the following sequence of matrices* Π^i_k

$$\Pi^i_{k+1} = \mathscr{R}^i\left(\textstyle\sum_{j=1}^{M} M_j k_{ij}^2 \Pi^j_k; Q, \bar{R}^{*i}_N\right) \tag{2.73}$$

with initial condition $\Pi^i_0 = \tilde{\Pi}^i_{0/N-1}$*. From optimality we obtain that* $\tilde{\Pi}^i_{k-N/k-1} \leq \Pi^i_{k-N}$ *for all* $k \geq N$*. Therefore, in order to prove boundedness of* $\Pi^i_{k-N/k-1}$*, it is sufficient to show that the sequence* Π^i_k *is bounded. This is the aim of the remainder of the proof.*

If we define $\Delta^i_k = \frac{1}{M_i k_{ii}^2} \sum_{j=1_{j\neq i}}^{M} M_j k_{ij}^2 \Pi^j_k$*,* (2.73) *can be written as*

$$\begin{aligned} \Pi^i_{k+1} &= (A - G^i_k \bar{\mathscr{O}}^i_N) M_i k_{ii}^2 \left(\Pi^i_k + \Delta^i_k\right) (A - G^i_k \bar{\mathscr{O}}^i_N)^T + \\ &\quad + Q + G^i_k \bar{R}^{*i}_N (G^i_k)^T \\ &= (\sqrt{M_i} k_{ii} A - G^i_k \sqrt{M_i} k_{ii} \bar{\mathscr{O}}^i_N) \left(\Pi^i_k + \Delta^i_k\right) \times \\ &\quad \times (\sqrt{M_i} k_{ii} A - G^i_k \sqrt{M_i} k_{ii} \bar{\mathscr{O}}^i_N)^T + Q + G^i_k \bar{R}^{*i}_N (G^i_k)^T \end{aligned} \tag{2.74}$$

where G^i_k *is the optimal Kalman gain computed as*

$$G^i_k = A(\Pi^i_k + \Delta^i_k)(\bar{\mathscr{O}}^i_N)^T \left(\bar{\mathscr{O}}^i_N (\Pi^i_k + \Delta^i_k)(\bar{\mathscr{O}}^i_N)^T + \bar{R}^{*i}_N\right)^{-1}$$

First we show that system (2.74) *is* $\mathscr{L}_\infty$*-stable. To this aim, we use Lemma 4. In order to satisfy the assumption of Lemma 4, one must guarantee that the pairs* $(\sqrt{M_i} k_{ii} A, \sqrt{M_i} k_{ii} \bar{\mathscr{O}}^i_N)$ *are detectable, for all* $i \in \mathscr{V}$*, which turns out to be a condition on the pairs* $(A, \bar{C}^i)$*, and on the weights* k_{ii}*. First notice that, by definition of* $\bar{\mathscr{O}}^i_N$*, the pair* $(\sqrt{M_i} k_{ii} A, \sqrt{M_i} k_{ii} \bar{\mathscr{O}}^i_N)$ *is detectable if and only if the pair* $(\sqrt{M_i} k_{ii} A, \sqrt{M_i} k_{ii} \bar{C}^i)$ *is detectable. The assumption of Theorem 3 is sufficient to guarantee that, for any regionally unobservable nodes, there exists a path stemming from a regionally observable node i.e., for which the assumption of Lemma 4 is satisfied for any arbitrary value of* k_{ii}*. In particular, in step 1 of Algorithm 1* $k_{ii} = 1$ *is chosen, for all* $i \in \mathscr{V}_O$*.*
On the other hand, if $(A, \bar{C}^i)$ *is not observable, the assumption of Lemma 4 can be verified if the pair* $(\sqrt{M_i} k_{ii} A, \bar{C}^i)$ *is detectable. This leads to the choice of* k_{ii} *in step 2 of Algorithm 1.*

Then, by Lemma 4, (2.74) *is a finite gain* $\mathscr{L}_\infty$*-stable system from input* $\Delta^i_k \geq 0$*, and there exist* $\gamma_i > 0, \beta_i > 0$ *such that*

$$\|\Pi^i_k - \bar{\Pi}^i\|_2 \leq \gamma_i \|\overrightarrow{\Delta}^i_k\|_{\mathscr{L}_\infty} + \beta_i\,,\ \forall \Delta^i_k \in \mathscr{L}_\infty, \forall k \in [0, \infty) \tag{2.75}$$

From the definition of Δ_k^i we get

$$\|\Pi_k^i - \overline{\Pi}^i\|_2 \leq \frac{\gamma_i}{M_i k_{ii}^2} \sum_{j=1_{j\neq i}}^{M} M_j k_{ij}^2 \|\overrightarrow{\Pi}_k^j\|_{\mathscr{L}_\infty} + \beta_i \tag{2.76}$$

$\forall \Pi_k^j \in \mathscr{L}_\infty$, $\forall k \in [0,\infty)$, $\forall j \in \mathscr{V}^i$. *Given* (2.76), *we resort to Theorem 8* [2] *in [17] for guaranteeing that* Π_k^i *are bounded if the matrix*

$$\Psi = diag(\frac{\gamma_1}{M_1 k_{11}^2}, ..., \frac{\gamma_M}{M_M k_{MM}^2}) \left(K \odot K diag(M_1, \ldots, M_M) - diag(M_1 k_{11}^2, ..., M_M k_{MM}^2)\right) \tag{2.77}$$

is Schur. In (2.77) *the symbol* $\odot$ *represents the element-wise matrix product.*

To conclude the proof, we show that, under the assumptions of Theorem 3, it is possible to find a matrix K, compatible with the graph topology, such that Ψ *is Schur.*

First, from the graph $(\mathscr{V},\mathscr{E})$, *we derive a subgraph* $\mathscr{G}^* = (\mathscr{V},\mathscr{E}^*)$, *by selecting edges* $(i,j) \in \mathscr{E}^* \subseteq \mathscr{E}$ *according to Algorithm 1.*

By construction, the graph $\mathscr{G}^*$ *is a forest [12], i.e. a graph composed by a number of mutually disjoint trees. Moreover, the root of each tree is a regionally observable node while all other nodes are regionally unobservable. It follows that each row of the matrix K produced by Algorithm 1 has only one off-diagonal element that is different from zero*[3].

Up to a permutation of the node indexes, K is lower triangular (see, e.g. Figure 2.2). It follows that the matrix Ψ *defined in* (2.77) *is triangular, with zero diagonal entries and hence, for any choice of* γ_i, $I = 1,\ldots,M$, Ψ *is Schur. This concludes the proof.*

Remark 1 *The matrix K generated by Algorithm 1 is lower triangular, up to a permutation of the node indexes. The same arguments of the above proof can be used to show, by continuity, that boundedness of* $\Pi_{t-N|t-1}^i$ *is guaranteed by any stochastic matrix* $\overline{K}$ *compatible with* $(\mathscr{V},\mathscr{E})$ *with: (i) the same diagonal elements of the matrix K; (ii) arbitrary elements in the lower triangular part; (iii) sufficiently small elements in the upper triangular part so as to guarantee the matrix* Ψ *defined in* (2.77) *is Schur.*

2.7.5 Proof of Corollary 2

Proof 16 *Recall that, from Algorithm 1, K is lower triangular up to a permutation of the sensor indexes. Hence,* $\mathbf{K} = K \otimes I_n$ *is a block lower triangular matrix. Recalling that* $\mathbf{P}_{NO}$ *and* $\mathbf{A}$ *are block diagonal matrices,* $\Phi = \mathbf{P}_{NO}\mathbf{K}\mathbf{A}\mathbf{P}_{NO}$ *is a block lower triangular matrix as well. Accordingly, the eigenvalues of* Φ *correspond to the eigenvalues of the M diagonal blocks of* Φ, *denoted as* Φ_i, $i \in \mathscr{V}$, *and defined as*

$$\Phi_i = k_{ii}\overline{T}^i(\overline{S}_{NO}^i)^T \overline{S}_{NO}^i (\overline{T}^i)^{-1} A \overline{T}^i (\overline{S}_{NO}^i)^T \overline{S}_{NO}^i (\overline{T}^i)^{-1}$$

Let A_{KO}^i *be defined as in* (2.64). *One has*

$$\Phi_i = k_{ii}\overline{T}^i(\overline{S}_{NO}^i)^T \overline{S}_{NO}^i A_{KO}^i (\overline{S}_{NO}^i)^T \overline{S}_{NO}^i (\overline{T}^i)^{-1} \tag{2.78}$$

[2] Note that, Theorem 8 in [17] can be directly applied when the constant β_i in (2.75) is replaced by a $\mathscr{K}_\infty$ function of the initial conditions. Furthermore, Theorem 8 deals with global stability rather than $\mathscr{L}_\infty$-stability. However, a careful examination of the proof reveals that Theorem 8 holds with reference to $\mathscr{L}_\infty$- stability when β is constant, provided that the gain map (represented by matrix Ψ in our context) is linear. We also highlight that, although Theorem 8 is for continuous-time systems, Proposition 15 in [17] guarantees that it holds for discrete-time systems as well.

[3] We also highlight that the matrix K produced by Algorithm 1 is compatible with the graph $(\mathscr{V},\mathscr{E})$.

and according to the definition of $\bar{S}^i_{NO}$*, one also obtains*

$$\tilde{A}^i_{KO} = (\bar{S}^i_{NO})^T \bar{S}^i_{NO} A^i_{KO} (\bar{S}^i_{NO})^T \bar{S}^i_{NO} = \begin{bmatrix} 0 & 0 \\ 0 & A^i_{NO} \end{bmatrix}$$

Therefore, from (2.78)*,* $\Phi_i = \bar{T}^i \frac{1}{k_{ii}} \tilde{A}^i_{KO} (\bar{T}^i)^{-1}$*. It is thus clear that the non-zero eigenvalues of* Φ *are also eigenvalues of* A^i_{NO}*, for some* $i \in \mathscr{V}_{NO}$*. If* $i \in \mathscr{V}_O$*,* $\Phi_i = \tilde{A}^i_{KO} = \mathbf{0}_{n\times n}$*. On the other hand, if* $i \in \mathscr{V}_{NO}$*, recall that, from step 2 of Algorithm 1, we have* $k_{ii} < \frac{1}{\sqrt{M_i}\sigma^{(i)}(A)}$*. Therefore,* $|\lambda_j(\Phi_i)| \leq k_{ii}\sigma^{(i)}(A) < \frac{1}{\sqrt{M_i}} < 1$*, for all* $j = 1,\ldots,n$*, for all* $i \in \mathscr{V}_{NO}$*. The Schureness of* Φ *then follows from the Schureness of* Φ_i*.*

Finally, notice that the assumptions of Theorem 3 imply that, in all the isolated strongly connected subgraphs of $\mathscr{G}$*, there is at least one observable node, and hence* (2.33) *holds. Therefore, by Theorem 2, Property 1 is verified. This concludes the proof.*

Chapter 3

DMHE extensions and analysis of the impact of different communication protocols

This chapter extends the results presented in Chapter 2 by considering different communication protocols among the agents of the sensor network. The effect of these protocols on the quality of the estimates is also studied. The content of this chapter is based on the paper [21].

3.1 Communication protocols and models

The first assumption on the communication network is that measurements taken by a sensor at time t are instantaneously transmitted to its first-order neighboring agents, i.e. to the agents directly connected to it. Secondly, we let $N_T \geq 1$ be the number of transmissions between two sensors within a sampling interval. Two types of data communication protocols can be assumed:

P$_1$) For $N_T \geq 1$, at time t sensor i collects the sets of measurement $\mathscr{Y}^i_{k/t} = \left\{ y^j_k,\, j \in \mathscr{V}_i^{N_T} \right\}$, for all $k \in [t-N, t]$.

P$_2$) For $N_T = 1$ and given $N \geq 1$, at time t sensor i collects the sets of measurement $\mathscr{Y}^i_{k/t} = \left\{ y^j_k,\, j \in \mathscr{V}_i^{t-k+1} \right\}$, for all $k \in [t-N, t]$.

Note that, even if $\mathscr{G}$ contains loops, measurements in the sets $\mathscr{Y}^i_{k/t}$ are considered just once. In the case of protocol P$_2$, the elements of the sets $\mathscr{Y}^i_{k/t}$ are illustrated in Figure 3.1. Note also that the protocols can be combined to obtain a more complex information transmission scheme. However, for simplicity, in the following the two cases P$_1$ and P$_2$ will be addressed independently.

We introduce now suitable notations for describing measurements available at node i at time t with both protocols. Let $\vec{y}^i_{k/t}$ be the vector of measurements in $\mathscr{Y}^i_{k/t}$ [1]. We denote with $\vec{p}^i_{t-k}$ the dimension of $\vec{y}^i_{k/t}$. Apparently, from matrices C^i one can build matrices $\vec{C}^i_{t-k} \in \mathbb{R}^{p^i_{t-k} \times n}$ such that

$$\vec{y}^i_{k/t} = \vec{C}^i_{t-k} x_k + \vec{v}^i_{k/t}, \quad t-N \leq k \leq t \tag{3.1}$$

[1] Note that the order in which elements of $\mathscr{Y}^i_{k/t}$ are listed in $\vec{y}^i_{k/t}$ does not play any particular role.

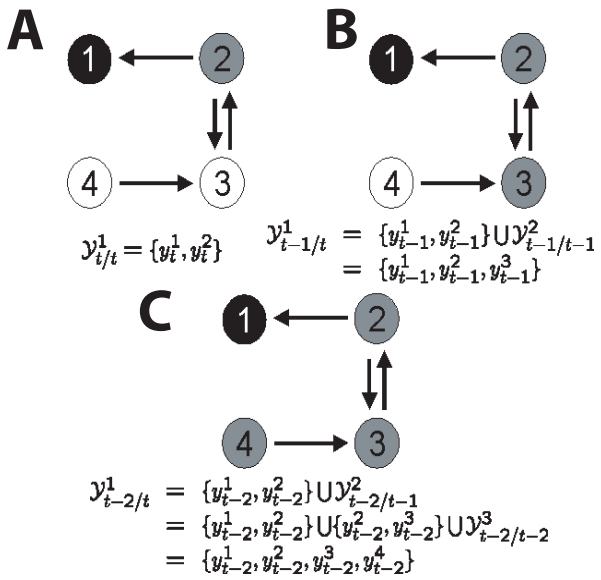


Figure 3.1: Illustration of the communication protocol P_2 for $N = 2$. The information available to node 1 at time t consists of $\mathscr{Y}^1_{k/t}$, $k = t-2, t-1, t$. The set $\mathscr{Y}^1_{t/t}$ (panel A) collects the data measured by nodes 1 and 2 at time t. The set $\mathscr{Y}^1_{t-1/t}$ (panel B) contains data measured by nodes 1 and 2 at time $t-1$ and data collected by node 2 at time $t-1$, that is $\mathscr{Y}^2_{t-1/t-1} = \{y^2_{t-1}, y^3_{t-1}\}$. Analogously, $\mathscr{Y}^1_{t-2/t}$ contains (panel C) the measurements y^j_{t-2}, $j = 1, \ldots, 4$.

where $\vec{v}^i_{k/t}$ collects noise samples affecting the measurements $\vec{y}^i_{k/t}$. Note also that in case of protocol P_1, matrices $\bar{C}^i_{t-k}$ are all identical.

We can now redefine a *regional* quantity (with respect to sensor i) as a quantity which is related to the sensor i and the nodes in $\mathscr{V}_i^{N_T}$ and $\mathscr{V}_i^{t-k+1}$ for protocols P_1 and P_2, respectively. Similarly to Section 2.2, $\vec{y}^i_{k/t}$ in (3.1) will be referred to as a regional measurement. Furthermore, we denote by $\bar{R}^i_{t-k} = \text{cov}[\vec{v}^i_{k/t}] \in \mathbb{R}^{\vec{p}^i_{t-k} \times \vec{p}^i_{t-k}}$, the covariance matrix of the regional noise $\vec{v}^i_{k/t}$, *i.e.*, $\bar{R}^i_{t-k} = \text{diag}\{R_j, j \in \mathscr{V}_i^{N_T}\}$ and $\bar{R}^i_{t-k} = \text{diag}\{R_j, j \in \mathscr{V}_i^{t-k+1}\}$ for protocols P_1 and P_2, respectively. The concept of regional observability can also be re-defined, similarly to the previous chapter. However, besides depending on the local observability properties of the single sensors, it is intimately linked to the estimation methodology adopted and to the communication graph properties. In fact, the *regional models* (2.1) and (3.1) are time-varying models, since the output equation (3.1) depends upon k, and the definition of regional observability will refer to these kinds of models. In this context, the most suitable observability definition is that of uniform observability [45], which easily applies to time-varying systems, as well as to nonlinear systems. For this reason, the following definitions depend upon the size of the estimation horizon and, in particular, upon the number of output-dependent terms $\|\vec{v}^i_k\|^2_{(\bar{R}^i_{t-k})^{-1}}$, see (2.8) (*i.e.*, $N+1$).

Given a single sensor model (2.1) and (3.1) and the considered communication protocol, the s step

regional observability matrix $\bar{\mathscr{O}}_s^i$ for sensor i is

$$\bar{\mathscr{O}}_s^i = \begin{bmatrix} \bar{C}_{s-1}^i \\ \bar{C}_{s-2}^i A \\ \vdots \\ \bar{C}_0^i A^{s-1} \end{bmatrix} \tag{3.2}$$

Definition 4 *The system is* regionally observable *by sensor i (or, equivalently, the sensor i is* regionally observable*) on horizon N, if $\ker(\bar{\mathscr{O}}_{N+1}^i) = 0$.* □

Now that the matrix $\bar{\mathscr{O}}_{N+1}^i$ has been re-defined, the matrices $\bar{P}_O^i$, $\bar{P}_{NO}^i$, $\mathbf{P}_O$, and $\mathbf{P}_{NO}$ are defined equivalently as in Section 2.2.

3.2 The distributed estimation algorithm and main results

For a given estimation horizon $N \geq 1$, each node $i \in \mathscr{V}$ at time t solves the constrained minimization problem *MHE-i* (2.6), under the constraints (2.7) where (2.7b) is replaced by the following one

$$\bar{y}_{k/t}^i = \bar{C}_{t-k}^i \hat{x}_k^i + \tilde{v}_k^i \tag{3.3}$$

where $k = t-N, \ldots, t$ and the local cost function J^i is given by

$$\begin{aligned} J^i(t-N,t,\hat{x}_{t-N}^i,\hat{w}^i,\tilde{v}^i,\Gamma_{t-N}^i) &= \tfrac{1}{2}\textstyle\sum_{k=t-N}^{t} \|\tilde{v}_k^i\|^2_{(\bar{R}_{t-k}^i)^{-1}} + \\ &+ \tfrac{1}{2}\textstyle\sum_{k=t-N}^{t-1} \|\hat{w}_k^i\|^2_{Q^{-1}} + \Gamma_{t-N}^i(\hat{x}_{t-N}^i; \tilde{\hat{x}}^i(t-N/t-1)) \end{aligned} \tag{3.4}$$

We denote with $\hat{x}^i(t-N/t)$ and with $\left\{\hat{w}^i(k/t)\right\}_{k=t-N}^{t-1}$ the optimizers to (2.6) and with $\hat{x}^i(k/t)$, $k = t-N,\ldots,t$ the local state sequence stemming from $\hat{x}^i(t-N/t)$ and $\left\{\hat{w}^i(k/t)\right\}_{k=t-N}^{t-1}$. Furthermore, $\tilde{\hat{x}}^i(t-N/t-1)$ denotes the weighted average state estimate

$$\tilde{\hat{x}}^i(t-N/t-1) = \sum_{j=1}^{M} k_{ij}^* \hat{x}^j(t-N/t-1) \tag{3.5}$$

where k_{ij}^* are the entries of the stochastic matrix K^* compatible with the graph $\mathscr{G}^*$ induced by K^{N_T}. Of course, the choice $K^* = K^{N_T}$ is always possible. However, note that agent i can set the nonzero entries k_{ij}^* autonomously, as far as $\sum_{j=1}^{M} k_{ij}^* = 1$. For instance, a possible choice is $k_{ij}^* = 1/|\mathscr{V}_i^{N_T}|$, and this highlights that the choice of coefficients k_{ij}^* can be done in a distributed fashion. In (3.4), the function $\Gamma_{t-N}^i(\hat{x}_{t-N}^i; \tilde{\hat{x}}^i(t-N/t-1))$ is the *initial penalty* defined, analogously to (2.10), as

$$\Gamma_{t-N}^i(\hat{x}_{t-N}^i; \tilde{\hat{x}}^i(t-N/t-1)) = \frac{1}{2}\|\hat{x}_{t-N}^i - \tilde{\hat{x}}^i(t-N/t-1)\|^2_{(\Pi_{t-N/t-1}^i)^{-1}} + \Theta_{t-1}^{*i} \tag{3.6}$$

where Θ_{t-1}^{*i} is the optimal cost defined in (2.6).

From now on, the proposed algorithm is equivalent to the one proposed in Chapter 2. The only further difference lies in the fact that k_{ij}^* and K^* are used in (2.31) and in the definition of Φ instead of k_{ij} and K. For instance, about the latter

$$\Phi = \mathbf{P}_{NO}\mathbf{K}^*\mathbf{A}\mathbf{P}_{NO}$$

where $\mathbf{K}^* = K^* \otimes I_n$. The scalar M_i is now defined as the number of N_T-th order neighbors of sensor i. Having said this, the results presented in Section 2.4 still hold without any additional assumption or further definition.

3.3 Collective observability and convergence of DMHE

In this section we analyze some key implications of collective observability (see Definition 1). Specifically, we investigate how the assumptions of Theorems 3 and 1 can be fulfilled, by properly tuning N and N_T, and provided that the assumptions of Theorem 2 are satisfied. First we consider the assumptions of Theorem 3. Then, we prove that N and N_T can be chosen in such a way that Φ is Schur.

I) First of all, for sufficiently large values of N and N_T, the assumptions of Theorem 2 imply the assumptions of Theorem 3. In fact, there exists a threshold value $\bar{N}_T$ [resp. $\bar{N}$] such that, at least one node in each isolated irreducible subgraph $\mathscr{G}_j$ ($j = 1,\dots,l$) is observable for $N_T \geq \bar{N}_T$ [resp. $N \geq \bar{N}$].

II) We now study the matrix Φ. As a limit case, assume that all sensors enjoy regional observability. This yields $\mathbf{P}_{NO} = \Phi = \mathbf{0}_{nM\times nM}$ and convergence of DMHE follows from Theorem 1. Consider now the two data transmission protocols mentioned in Section 3.1.
In the case P_1, regional observability can be enhanced by increasing the number N_T of data transmissions between agents within a sampling interval. The increase of N_T has two accompanying effects.

1) If all the isolated strongly connected subgraphs are collectively observable, there exists a threshold value for N_T (say $\bar{N}_t$) such that regional observability is satisfied by all the sensors for $N_T \geq \bar{N}_T$. So, for $N_T \geq \bar{N}_T$ one has $\mathbf{P}_{NO} = \mathbf{0}_{nM\times nM}$.
2) If $K^* = K^{N_T}$, the modulus of the eigenvalues of matrix Φ decrease as N_T increases. In fact, the eigenvalues of $\mathbf{K}^*$ are equal to the eigenvalues of K^{N_T}. Since K is stochastic, it has l (being l the number of non empty irreducible subgraphs in which one can partition $\mathscr{G}$) eigenvalues equal to 1, and $M-l$ eigenvalues with modulus strictly less than 1. We denote with λ_j the j-th eigenvalues of K. The corresponding eigenvalues of K^* verify $|\lambda_j^{N_T}| \leq |\lambda_j|$, resulting in a decrease of the eigenvalues of Φ.

On the other hand, if the communication protocol P_2 is employed, we can enhance regional observability and the Schureness of Φ by increasing the estimation horizon N. As a limit case, if all the isolated strongly connected subgraphs are collectively observable, there exists a value $\bar{N}$ such that, $N \geq \bar{N}$ implies that $\mathbf{P}_{NO} = \mathbf{0}_{nM\times nM}$, which guarantees that Φ is Schur.

Chapter 4

Covariance estimation - a critical analysis of existing algorithms and some new ideas

Kalman filtering has been a widely used estimation technique since the '60s. It guarantees optimal estimation performances (in that it aims to minimize the estimation error variance) assuming that disturbances affect the system dynamics and the system measurements.

The covariances of the noisy disturbances entering in the model and the measurements are considered as the tuning parameters of the filter. However, in many applications the true values of such noise covariances are not known. In this scenario, the optimality of the Kalman filter can not be guaranteed.

Hence, in order to use the optimal filter, we need to know the covariances of the disturbances affecting the system, from which the optimal Kalman filter gain can be computed. In this report the problem of estimating the noise covariances for linear, time-invariant systems is addressed.

First a statement of the problem is given. Then the state of the art is presented. Two main algorithms are presented and tested on two well-known academic examples. Finally, the selected scheme is tested on a realistic problem.

4.1 Problem statement

Consider a linear, time-invariant, discrete-time model:

$$\begin{aligned} x_{k+1} &= Fx_k + Bu_k + Gw_k \\ z_k &= Hx_k + v_k \end{aligned} \tag{4.1}$$

where $x \in \mathbb{R}^n$ is the system state, $F \in \mathbb{R}^{n\times n}$ is the transition matrix, $B \in \mathbb{R}^{n\times m}$ is the control matrix, $G \in \mathbb{R}^{n\times g}$ is the disturbance matrix, $z \in \mathbb{R}^p$ is the observation vector, and $H \in \mathbb{R}^{p\times n}$ is the observation matrix. Note that, $\{u_k\}_{k=0}^{N_d}$, $\{w_k\}_{k=0}^{N_d}$, and $\{v_k\}_{k=0}^{N_d}$ are the control, the state uncertainty vector (or process-noise), and the measurement noise sequences respectively, with N_d the size of the sequences. The disturbances v and w are zero-mean Gaussian white noises with R_v and Q_w as covariance matrices, respectively.

State estimates of the system are computed using a linear, time-invariant state estimator:

$$\begin{aligned} \hat{x}_{k+1|k} &= F\hat{x}_{k|k} + Bu_k \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + L\left[z_k - H\hat{x}_{k|k-1}\right] \end{aligned} \tag{4.2}$$

where L is the estimator gain, which is not necessarily the optimal gain. The residuals of the output equations $z_k - H\hat{x}_{k|k-1}$ are the so-called L-innovations, since these are calculated using a state estimator with gain L. The variance estimation problem is reduced to find the true matrices Q_w and R_v using real data from the innovations, with the final goal of computing the optimal estimator's error covariance matrix and the optimal filter gain.

4.1.1 The effect of erroneous covariance matrices on the filter optimality

Let consider the effect of using a non-optimal gain on the performance of the filter due to the assumption of erroneous covariance matrices Q_w and R_v [24]. It will be shown that the use of the erroneous covariance matrices gives a suboptimal solution of the filtering problem.

First, let L_k be the optimal Kalman gain calculated with the real covariance matrices Q_w and R_v at instant k. Consider the use of this gain in a filter as in (4.2). The error in the estimation is defined as $\varepsilon_{k|k-1} = x_k - \hat{x}_{k|k-1}$. Then, an expression for the error dynamics is given by:

$$\begin{aligned}\varepsilon_{k+1|k} &= Fx_k + Bu_k + Gw_k - F\left[\hat{x}_{k|k-1} + L_k\left(z_k - H\hat{x}_{k|k-1}\right)\right] - Bu_k \\ &= Fx_k + Gw_k - F(I - L_kH)\hat{x}_{k|k-1} - FL_kz_k \\ &= Fx_k + Gw_k - F(I - L_kH)\hat{x}_{k|k-1} - FL_k(Hx_k + v_k)\end{aligned} \tag{4.3}$$

and after some calculations,

$$\varepsilon_{k+1|k} = F(I - L_kH)\varepsilon_{k|k-1} + Gw_k - FL_kv_k \tag{4.4}$$

The estimation error covariance is defined as:

$$M_k = E\left\{\varepsilon_{k|k-1}\varepsilon_{k|k-1}^T\right\} \tag{4.5}$$

where $E\{\cdot\}$ denotes the statistical expectation. Using (4.4) in the last equation we obtain

$$M_{k+1} = E\left\{\left[F(I - L_kH)\varepsilon_{k|k-1} + Gw_k - FL_kv_k\right]\left[F(I - L_kH)\varepsilon_{k|k-1} + Gw_k - FL_kv_k\right]^T\right\} \tag{4.6}$$

which after some computations allows to write the expression for the estimation error covariance matrix in a recursive way as:

$$M_{k+1} = FM_kF^T - FM_kH^TL_k^TF^T - FL_kHM_KF^T + FL_kHM_KH^TL_k^TF^T + GQ_wG^T + FL_kR_vL_k^TF^T \tag{4.7}$$

Equation (4.7) can be rearranged as:

$$M_{k+1} = F\left[M_k - M_kH^TL_k^T - L_kHM_k + L_k(HM_kH^T + R_v)L_k^T\right]F^T + GQ_wG^T \tag{4.8}$$

Minimizing (4.8) with respect to the gain L_k we obtain:

$$\min_{L_k} M_{k+1} = \min_{L_k}\left\{F\left[M_k - M_kH^TL_k^T - L_kHM_k + L_k(HM_kH^T + R_v)L_k^T\right]F^T + GQ_wG^T\right\} \tag{4.9}$$

which gives the classic result of the Kalman filter,

$$L_k = M_k H^T \left(HM_k H^T + R_v\right)^{-1} \tag{4.10}$$

On the other hand, a suboptimal gain L_k^e can be computed by using the erroneous covariance matrices Q_w^e and R_v^e, where the super index $(\cdot)^e$ denotes a erroneous quantity. Let us assume that the gain used in the filter deviates from the optimal gain by a quantity δL_k, *i.e*,

$$L_k^e = L_k + \delta L_k \tag{4.11}$$

Then the suboptimal estimation error covariance matrix can be written as:

$$M_{k+1}^e = F\left[M_k^e - M_k^e H^T \left(L_k^e\right)^T - L_k^e H M_k^e + L_k^e (HM_k^e H^T + R_v^e)\left(L_k^e\right)^T\right] F^T + GQ_w^e G^T \tag{4.12}$$

From (4.8) and (4.12) we can verify that, while L_k minimizes M_k in (4.8), the gain

$$L_k^e = M_k^e H^T \left(HM_k^e H^T + R_v^e\right)^{-1} \tag{4.13}$$

minimizes (4.12), but the desired (4.8) as claimed.

4.2 State of the art

The problem of estimating the covariances from open-loop data has long been subject of research in the field of adaptive filtering, and according to [35] the methods can be divided into four general categories: Bayesian [7, 25], maximum likelihood [11, 28], covariance matching [34], and correlation techniques. Bayesian and maximum likelihood methods have fallen out of favor because of their sometimes excessive computation times. They may be well suited to a multi-model approach as in [8]. Covariance matching is the computation of the covariances from the residuals of the state estimation problem. Covariance matching techniques have been shown to give biased estimates of the true covariances. The fourth category is correlation techniques, largely pioneered by Mehra [32, 33] and Carew and Bélanger [9, 15]. In [35] an alternative method to the one presented in [32, 33] is described, where necessary and sufficient conditions for uniqueness of the estimated covariances are also given. This method, called Autocovariance Least Squares (ALS) outperforms significantly the results obtained by the algorithm proposed in [32] as it will be shown later.

In the sequel the algorithm proposed in [32] and the ALS algorithm presented in [35] will be presented in detail. The first is presented in order to show the starting point in the literature about the correlation-based covariance estimation schemes. Then, a deep explanation of the ALS method is given.

4.2.1 The pioneering work of Mehra [32].

A covariance estimation scheme was proposed by Mehra at the beginning of the ’70s. In [32] a three step covariance estimation procedure using the L-innovations of the system from a sub-optimal tuning of the filter is shown. The preliminary step consists of an optimality test in order to verify when a innovation sequence originates from a suboptimal filter. The optimality test is described in detail in the following:

Optimality test. First a test of optimality is made to the L-innovation sequence. If the test gives positive results about the optimality of the innovation sequence, the next steps are discarded and the assumed covariances can be taken as the real ones. Otherwise the real covariances must be estimated. To show how the test of optimality works, consider the normalized autocorrelation coefficients of the L-innovation sequence,

$$[\hat{\rho}_k]_{ij} = \frac{\left[\hat{C}_k\right]_{ij}}{\left\{\left[\hat{C}_0\right]_{ii}\left[\hat{C}_0\right]_{jj}\right\}^{1/2}} \tag{4.14}$$

where $\left[\hat{C}_k\right]_{ij}$ denotes the element in the i-th row and the j-th column of the matrix $\hat{C}_k$, which is known as the estimated autocorrelation function of the L-innovation sequence and can be estimated as follows:

$$\hat{C}_k = \frac{1}{N_d}\sum_{i=k}^{N_d} \mathscr{Z}_i \mathscr{Z}_{i-k}^T \tag{4.15}$$

with $\mathscr{Z}_k = z_k - H\hat{x}_{k|k-1}$ as the L-innovation at the k-th time instant. Therefore, the test consist of analyzing the sequences $\{\hat{\rho}_k\}_{ii}$, $k > 0$ and checking the number of times they lie outside the band $\pm(1.96/N_d^{1/2})$, assuming a confidence band of 95%. If this number is less than 5% of the total, the sequence $\{\mathscr{Z}_1, ..., \mathscr{Z}_{N_d}\}$ is white [26].

The three steps to perform the covariance estimation algorithm are the following:

1. **Computation of $\hat{M}\hat{H}^T$**. In order to compute MH^T, consider the autocorrelation function, which is formally defined as

$$C_k = E\left\{\mathscr{Z}_i \mathscr{Z}_{i-k}^T\right\} \tag{4.16}$$

The L-innovation can be written as a function of the estimation error $\varepsilon_i = x_i - \hat{x}_{i|i-1}$ as:

$$\mathscr{Z}_i = H\varepsilon_i + v_i \tag{4.17}$$

Using (4.16) and (4.17), the autocorrelation function can be written as:

$$\begin{aligned} C_k &= E\left\{(H\varepsilon_i + v_i)(H\varepsilon_{i-k} + v_{i-k})^T\right\} \\ &= H\underbrace{E\left\{\varepsilon_i \varepsilon_{i-k}^T\right\}}_{(i)}H^T + H\underbrace{E\left\{\varepsilon_i v_{i-k}^T\right\}}_{(ii)} + \underbrace{E\left\{v_i \varepsilon_{i-k}^T\right\}H^T}_{=0} + \underbrace{E\left\{v_i v_{i-k}^T\right\}}_{=0} \end{aligned} \tag{4.18}$$

The computation of (i) can be performed using the recursive equation (4.4), where $L_k = L$ (the filter is considered in stationary conditions).

By iterating (4.4), we obtain that

$$\varepsilon_i = [F(I-LH)]^k \varepsilon_{i-k} - \sum_{j=1}^{k}[F(I-LH)]^{j-1} F L v_{i-j} + \sum_{j=1}^{k}[F(I-LH)]^{j-1} G w_{i-j} \tag{4.19}$$

Equation (4.19) is post multiplied by ε_{i-k}^T in order to get an expression for (i),

$$E\left\{\varepsilon_i \varepsilon_{i-k}^T\right\} = \left[F(I-LH)\right]^k M \tag{4.20}$$

where M is the steady state error covariance matrix for the suboptimal case. An expression for M can be obtained directly from (4.4) as:

$$\begin{aligned} M &= E\left\{\varepsilon_i \varepsilon_i^T\right\} \\ &= F(I-LH)M(I-LH)^T F^T + FLRL^T F^T + GQ_w G^T \end{aligned} \tag{4.21}$$

The computation of (ii) is based on (4.19) post multiplied by v_{i-k}^T,

$$E\left\{\varepsilon_i v_{i-k}^T\right\} = -\left[F(I-LH)\right]^{k-1} FLR_v \tag{4.22}$$

Hence, replacing (4.20) and (4.22) in (4.18) it follows that:

$$\begin{aligned} C_k &= H\left[F(I-LH)\right]^k MH - H\left[F(I-LH)\right]^{k-1} FLR_v \\ &= H\left[F(I-LH)\right]^{k-1} F\left[MH^T - L(HMH^T + R_v)\right] \end{aligned} \tag{4.23}$$

Hence

$$\begin{aligned} C_k &= HMH^T + R_v, & k = 0 \\ &= H\left[F(I-LH)\right]^{k-1} F\left[MH^T - LC_0\right], & k > 0 \end{aligned} \tag{4.24}$$

From (4.24) the term MH^T can be obtained in two alternative ways, i.e:

$$MH^T = B^\dagger \begin{bmatrix} C_1 + HFLC_0 \\ C_2 + HFLC_1 + HF^2LC_0 \\ \vdots \\ C_n + HFLC_{n-1} + \cdots + HF^nLC_0 \end{bmatrix} \tag{4.25}$$

or

$$MH^T = KC_0 + A^\dagger \begin{bmatrix} C_1 & \cdots & C_n \end{bmatrix}^T \tag{4.26}$$

where $(\cdot)^\dagger$ denotes the Moore Penrose pseudoinverse of $(\cdot)$, B is the product between the observability matrix[1] of the system and the transition matrix,

$$B \equiv \mathscr{O}(F,H) \cdot F \tag{4.27}$$

[1] the observability matrix is defined as $\mathscr{O}(F,H) = \begin{bmatrix} H^T & (HF)^T & \cdots & \left[HF^{n-1}\right]^T \end{bmatrix}^T$

and,

$$A = \begin{bmatrix} HF \\ HF(I-LH)F \\ \vdots \\ H\left[F(I-LH)^{n-1}\right]F \end{bmatrix} \tag{4.28}$$

Then an estimate of the product MH^T (denoted $\hat{M}\hat{H}^T$) can be calculated using the estimated values of the autocorrelation function using either (4.25), (4.26). The author suggests to use (4.26) since he claims that it is numerically better conditioned than (4.25).

2. **Estimation of** R_v. The covariance of the state noise can be computed directly using (4.24) for $k = 0$,

$$\hat{R}_v = \hat{C}_0 - H(\hat{M}\hat{H}^T) \tag{4.29}$$

3. **Estimation of** Q_w. This step becomes more complicated since matrix Q_w is only involved in the error covariance matrix equation (4.21). Then, only $n \times p$ linear relationships for the unknown parameters in Q_w are available from this equation. If the number of unknown parameters are greater than $n \times p$, an alternative procedure must be used. Moreover, as the estimates for MH^T and R_v are not helpful in the way the equation (4.21) is written, a derived equation from (4.21) must be found to get advantage of the estimations. In order to explain how it works, consider that (4.21) can be rewritten as follows:

$$\begin{aligned} M &= F(I-LH)M(I-LH)^T F^T + FLR_v L^T F^T + GQ_w G^T \\ &= FMF^T + \Omega + GQ_w G^T \end{aligned} \tag{4.30}$$

where $\Omega = F\left[-LHM - MH^T L^T + LC_0 L^T\right]F^T$. Substituting back for M in (4.30) n times and separating the terms involving Q_w, we have that

$$\sum_{j=0}^{k-1} F^j G Q_w G^T (F^j)^T = M - F^k M (F^k)^T - \sum_{j=0}^{k-1} F^j \Omega (F^j)^T, \quad k = 1,\ldots,n \tag{4.31}$$

Premultiplying both sides of last equation by H and post-multiplying by $(F^{-k})^T H^T$

$$\begin{aligned} \sum_{j=0}^{k-1} HF^j G Q_w G^T (F^{j-k})^T H^T =& HM(F^{-k})^T H^T - HF^k M H^T \\ & - \sum_{j=0}^{k-1} HF^j \Omega (F^{j-k})^T H^T, \quad k = 1,\ldots,n. \end{aligned} \tag{4.32}$$

Note that the right-hand side of (4.32) is completely determined by MH^T and C_0. Then the estimations of these terms allow to determine the components of the matrix Q_w. One of the drawbacks in the estimation of the components of matrix Q_w is that the equations described by means of (4.32) are not all linearly independent. Then, a linear independent subset of these equations must be chosen.

4.2.2 Autocovariance Least Squares -ALS- [35]

In [35] a method (constrained Autocovariance Least Squares -ALS-) to estimate the variances of the disturbances entering the process using routine operating data is presented. In this paper the results are compared with those presented in [32], which is considered as a seminal paper in this subject. As a result of this comparison, Odelson *et al.* stated the inability of the previous contribution to tackle the covariance estimation in certain cases. Indeed, it is shown that in specific examples, Mehra's scheme does not work properly. Moreover, one of the main criticisms on the Mehra's algorithm is the use of a three-step procedure to compute the covariances. The method presented in [35] uses a one-step procedure, which yields covariance estimates with lower uncertainty on all tested examples. The formulation used in this paper provides necessary and sufficient conditions for uniqueness of the estimated covariances, previously not available in the literature. Finally, the authors give a formulation to avoid negative definite estimates with a convex optimization problem using a barrier term.

In order to show the algorithm, consider the dynamic evolution of the state estimation error, $\varepsilon_k = x_k - \hat{x}_{k|k-1}$, from (4.4):

$$\varepsilon_{k+1} = \underbrace{(F - FLH)}_{\bar{F}}\varepsilon_k + \underbrace{\begin{bmatrix} G & -FL \end{bmatrix}}_{\bar{G}} \underbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}_{\bar{w}(k)} \tag{4.33}$$

Then, the state-space model of the L-innovations is defined as:

$$\begin{aligned} \varepsilon_{k+1} &= \bar{F}\varepsilon_k + \bar{G}\bar{w}_k \\ \mathscr{Z}_k &= H\varepsilon_k + v_k \end{aligned} \tag{4.34}$$

In the sequel, the following conditions are assumed to hold:

- The pair (F,H) is detectable.
- The transition matrix of the estimation error dynamics is stable.
- $E(\varepsilon_0) = 0, \quad \mathrm{Cov}(\varepsilon_0) = M_0^-$

The last assumption gives the possibility of using the Lyapunov equation to guarantee a recursion for the covariance of the estimation error:

$$M_j^- = \bar{F}M_{j-1}^-\bar{F}^T + \bar{G}\bar{Q}_w\bar{G}, \quad j = 1,...,k. \tag{4.35}$$

Using (4.35) and the error dynamics, the expected values of the innovations can be written algebraically:

$$\begin{aligned} E\left\{\mathscr{Z}_k\mathscr{Z}_k^T\right\} &= HM^-H^T + R_v \\ E\left\{\mathscr{Z}_{k+j}\mathscr{Z}_k^T\right\} &= H\bar{F}^jM^-H^T - H\bar{F}^{j-1}FLR_v, \quad j \geq 1 \end{aligned} \tag{4.36}$$

Moreover, the autocovariance matrix (ACM) is defined as:

$$\mathscr{R}(N) = E\begin{bmatrix} C_0 & \cdots & C_{N-1} \\ \vdots & \ddots & \vdots \\ C_{N-1}^T & \cdots & C_0 \end{bmatrix} \tag{4.37}$$

where C_k is defined in (4.16) and N is a user-defined parameter. Using (4.36) and (4.37) the ACM of the L-innovations can be written as:

$$\mathscr{R}(N) = \mathscr{O}_{ALS} M^{-} \mathscr{O}_{ALS}^{T} + \Gamma \left[\bigoplus_{i=1}^{N} \bar{G} \bar{Q}_w \bar{G}^T \right] \Gamma^T + \Psi \left[\bigoplus_{j=1}^{N} R_v \right] + \left[\bigoplus_{j=1}^{N} R_v \right] \Psi^T + \left[\bigoplus_{j=1}^{N} R_v \right] \tag{4.38}$$

where,

$$\mathscr{O}_{ALS} = \begin{bmatrix} H \\ H\bar{F} \\ \vdots \\ H\bar{F}^{N-1} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 \\ H & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots \\ H\bar{F}^{N-2} & \cdots & H & 0 \end{bmatrix}, \quad \Psi = \Gamma \left[\bigoplus_{j=1}^{N} -FL \right]$$

Also, the covariance of the whole noise is,

$$E\left[\bar{w}(k)(\bar{w}(k))^T\right] = \bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}$$

In order to show the problem formulation as a Least-Squares problem, (4.38) is given in stacked form. Henceforth, $(\cdot)_s$ denotes the outcome to apply the *vec* operator to $(\cdot)$.(4.38) is written in a stacked way using the standard definitions [35] of the Kronecker sum $\oplus$, Kronecker product $\otimes$, and direct sum $\bigoplus$ as:

$$\begin{aligned} [\mathscr{R}(N)]_s = & \left[(\mathscr{O}_{ALS} \otimes \mathscr{O}_{ALS})(I_{n^2} - \bar{F} \otimes \bar{F})^{-1} + (\Gamma \otimes \Gamma)\mathscr{I}_{n,N} \right] (G \otimes G)(Q_w)_s \\ & + \left\{ \left[(\mathscr{O}_{ALS} \otimes \mathscr{O}_{ALS})(I_{n^2} - \bar{F} \otimes \bar{F})^{-1} + (\Gamma \otimes \Gamma)\mathscr{I}_{n,N} \right] (FL \otimes FL) + \left[\Psi \oplus \Psi + I_{p^2N^2} \right] \mathscr{I}_{p,N} \right\} (R_v)_s \end{aligned} \tag{4.39}$$

where also the Lyapunov equation is written as:

$$M_s^{-} = (\bar{F} \otimes \bar{F}) M_s^{-} + (\bar{G} \bar{Q}_w \bar{G}^T)_s \tag{4.40}$$

Equation (4.39) can be written as a LS problem, considering that $\mathscr{R}(N)_s$ can be estimated from (4.37) using the acquired data.

Given $\mathscr{A}x = b$, with

$$\begin{aligned} D &= \begin{bmatrix} (\mathscr{O}_{ALS} \otimes \mathscr{O}_{ALS}(I_{n^2} - \bar{F} \otimes \bar{F})^{-1} + (\Gamma \otimes \Gamma)\mathscr{I}_{n,N}) \end{bmatrix} \\ \mathscr{A} &= \begin{bmatrix} D(G \otimes G) & D(FL \otimes FL) + \left[\Psi \oplus \Psi + I_{p^2N^2} \right] \mathscr{I}_{p,N} \end{bmatrix} \\ x &= \begin{bmatrix} (Q_w)_s^T & (R_v)_s^T \end{bmatrix}^T \\ b &= \mathscr{R}(N)_s \end{aligned}$$

where $\mathscr{I}_{p,N}$ is a permutation matrix to convert the direct sum to a vector, i.e $\mathscr{I}_{p,N}$ is the $(pN)^2 \times p^2$ matrix of zeros and ones satisfying:

$$\left(\bigoplus_{j=1}^{N} R_v \right)_s = \mathscr{I}_{p,N} (R_v)_s$$

We define the ALS estimate as follows:

$$\begin{aligned} &\hat{x} = \arg\min_{x} \|\mathscr{A}x - b\|_2^2 \\ &s.t \;\; Q_w, R_v \geq 0 \end{aligned} \tag{4.41}$$

in which $\hat{x} = \left[\begin{array}{cc} (Q_w)_s^T & (R_v)_s^T \end{array} \right]^T$, and $\hat{b} = \hat{\mathscr{R}}(N)_s$. These steps are summarized in **Algorithm 2**.

Algorithm 2 ALS Algorithm

for $j = 1$ to $N-1$ **do**
$\hat{C}_j = \frac{1}{N_d - j} \sum_{i=1}^{N_d - j} \mathscr{Z}_i \mathscr{Z}_{i+j}^T$
end for
Compute $\hat{b} = \hat{\mathscr{R}}(N)_s$ from (4.37)
Solve $\left[\begin{array}{c} \hat{Q}_w \\ \hat{R}_v \end{array} \right] = \arg\min_{Q_w, R_v} \left\| \mathscr{A} \left[\begin{array}{c} (Q_w)_s \\ (R_v)_s \end{array} \right] - \hat{b} \right\|_2^2$ s.t. $Q_w \geqslant 0, \;\; R_v \geqslant 0$

The following Lemmas and Theorem guarantee the existence, uniqueness and unbiasedness nature of the estimates [35]:

Lemma 5 *The ALS estimate given for the Algorithm 2 exists and is unique if and only if $\mathscr{A}$ has full column rank.*

Lemma 6 *The expectation of the estimated autocovariance $\hat{C}_j$ is equal to the autocovariance C_j for all j, and the variance tends to zero as N_d tends to ∞.*

$$\begin{aligned} &E\left\{\hat{C}_j\right\} = C_j, \quad j = 0, ..., N \\ &Cov(\hat{C}_j) = O\left(\frac{1}{N_d - j}\right) \end{aligned}$$

Theorem 4 *Given $\mathscr{A}$ has full column rank, the ALS noise covariance estimates $(\hat{Q}_w, \hat{R}_v)$ are unbiased for all samples sizes and converge asymptotically to the true covariances (Q_w, R_v) as $N_d \to \infty$.*

4.2.3 Advanced schemes

In [1] a generalization of the ALS method for estimating the noise covariances even assuming cross covariances is presented. It is shown that equivalent results are obtained with both the predicting and the filtering form of the Kalman filter. The original algorithm is reformulated assuming that

$$\left[\begin{array}{c} w_k \\ v_k \end{array} \right] \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} Q_w & S_{wv} \\ S_{wv}^T & R_v \end{array} \right] \right) \tag{4.42}$$

In [2] an advanced ALSel method for estimating the noise covariances from real data is also discussed. The covariance estimation problem is stated as a least-squares problem, which is solved as a symmetric semidefinite least-squares problem. In this paper, Åkesson *et al.* address the problem following the analysis from [35] and generalized in [1] for systems with correlated process and measurement noises. Two contributions can be highlighted: the generalization of the autocovariance least-square method to systems with correlated noise, and the interior-point predictor-corrector algorithm for solving the symmetric semidefinite least-squares problem. The need of including an integrating disturbance model to

the linear state-space model in order to achieve offset-free control is also stated, especially in MPC applications. This model can be written as follows:

$$\begin{aligned} x_{k+1} &= Fx_k + Bu_k + G_1 d_k + G_2 w_k \\ d_{k+1} &= d_k + \xi_k \\ z_k &= H_1 x_k + H_2 d_k + v_k \end{aligned} \tag{4.43}$$

where $G_1 \in \mathbb{R}^{n_x \times n_d}$, G_2 is as G in (4.1), $H_2 \in \mathbb{R}^{n_z \times n_d}$. The white noise ξ_k has covariance Q_ξ which is uncorrelated with the process noise and the measurement noise. Then, an augmented model is formed from (4.43):

$$\begin{aligned} \bar{x}_{k+1} &= A\bar{x}_k + B_n u_k + G_n \omega_k \\ y_k &= C\bar{x}_k + v_k \end{aligned} \tag{4.44}$$

with $\bar{x}_k = \begin{bmatrix} x_k & d_k \end{bmatrix}^T$, $\omega_k = \begin{bmatrix} w_k & \xi_k \end{bmatrix}^T$, and

$$A = \begin{bmatrix} F & G_1 \\ 0 & I \end{bmatrix}, \quad B_n = \begin{bmatrix} B \\ 0 \end{bmatrix}, \quad G_n = \begin{bmatrix} G_2 & 0 \\ 0 & I \end{bmatrix}, \quad C = \begin{bmatrix} H_1 & H_2 \end{bmatrix} \tag{4.45}$$

Then, the covariances of the noises are defined as:

$$\bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & Q_\xi \end{bmatrix}, \quad \bar{S}_{wv} = \begin{bmatrix} S_{wv} & 0 \end{bmatrix} \tag{4.46}$$

The problem of optimal tuning of the Kalman filter is reduced to the estimation of the true covariance matrices in the augmented system (4.44). Once these matrices are estimated, the optimal filter or predictor can be implemented as follows:

Predictor:

$$\hat{\bar{x}}_{k+1|k} = A\hat{\bar{x}}_{k|k-1} + B_n u_k + K_p \left[y_k - C\hat{\bar{x}}_{k|k-1} \right] \tag{4.47}$$

Filter:

$$\begin{aligned} &\textit{Time update} \\ &\hat{\bar{x}}_{k+1|k} = A\hat{\bar{x}}_{k|k-1} + B_n u_k + G_n \hat{\omega}_{k|k} \\ &\\ &\textit{Measurement update} \\ &\hat{\bar{x}}_{k|k} = \hat{\bar{x}}_{k|k-1} + K_{fx}(y_k - C\hat{\bar{x}}_{k|k-1}) \\ &\hat{\omega}_{k|k} = K_{fw}(y_k - C\hat{\bar{x}}_{k|k-1}) \end{aligned} \tag{4.48}$$

with the Kalman gain for the prediction K_p and filtering K_{fx}, K_{fw} as:

$$\begin{aligned} K_p &= (AP_pC^T + G_n\bar{S}_{wv})(CP_pC^T + R_v)^{-1} \\ K_{fx} &= P_pC^T(CP_pC^T + R_v)^{-1} \\ K_{fw} &= \bar{S}_{wv}(CP_pC^T + R_v)^{-1} \end{aligned} \tag{4.49}$$

and P_p is the covariance of the state prediction error, what is obtained as the solution of the Riccati equation.

In [36] the ALS method proposed in [35] was tested on two chemical reactor control problems. This method uses closed-loop process data to recover the covariances of the disturbances entering the process. Moreover, in this contribution the ALS method is used with integrating white noise disturbances, which are required for offset-free control in most of the MPC applications. In fact, the authors highlight the considerably improved results obtained by using the optimal tuning of the Kalman filter.

Later, the ALS scheme has been modified in [48]. The authors claim that significant improvements to the original ALS method are presented, highlighting the following: (i) new and simpler necessary and sufficient conditions for the uniqueness of the covariance estimates, (ii) an optimal weighting applied to the least-squares formulation in order to minimize the variance of the estimates, and (iii) the estimation of the stochastic disturbance structure affecting the states.

Under the assumptions of observability and the use of the innovations data from the steady state response of the system, the ALS method is reformulated avoiding redundant definition of the lagged covariances. Hence, instead using (4.16) and (4.37) the ALS scheme is rewritten using only the first block column of the autocovariance matrix:

$$\mathscr{R}_1(N) = E\begin{bmatrix} \mathscr{Z}_k\mathscr{Z}_k^T \\ \vdots \\ \mathscr{Z}_{k+N-1}\mathscr{Z}_k^T \end{bmatrix} \tag{4.50}$$

As a result (by analogy with the original scheme) the following positive definite constrained least-squares problem in the symmetric elements of the covariances GQ_wG^T and R_v is defined:

$$\begin{aligned} \Phi = \min_{GQ_wG^T,R_v} & \left\| \mathscr{A}\begin{bmatrix} \mathscr{D}_n(GQ_wG^T)_{ss} \\ (R_v)_{ss} \end{bmatrix} - \hat{b} \right\|_W^2 \\ \text{subject to,} \quad & GQ_wG^T, R_v \geqslant 0, \quad R_v = R_v^T \end{aligned} \tag{4.51}$$

where $(R_v)_{ss}$ denote the column-wise stacking of only the symmetric $p(p+1)/2$ elements of the matrix R_v. In other words, there exists an unique matrix $\mathscr{D}_p \in \mathbb{R}^{p^2 \times \frac{p(p+1)}{2}}$ called the *duplication matrix* containing zeros and ones such that,

$$(R_v)_s = \mathscr{D}_p(R_v)_{ss}$$

and W is a weighting matrix in order to guarantee the minimum variance among all the unbiased estimators.

The first contribution is made assuming $W = I$, and the complete knowledge of matrix G. Using (4.36), and the Lyapunov equation (4.35), (4.50) can be written as follows:

$$\mathscr{R}_1(N) = \mathscr{O}PH^T + \Gamma R_v \tag{4.52}$$

in which

$$\mathscr{O} = \begin{bmatrix} H \\ H\bar{F} \\ \vdots \\ H\bar{F}^{N-1} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} I_p \\ -HFL \\ \vdots \\ -H\bar{F}^{N-2}FL \end{bmatrix} \tag{4.53}$$

Then, (4.52) is stacked as in the original ALS contribution:

$$\begin{aligned} [\mathscr{R}_s(N)]_s = &\left[(H\otimes\mathscr{O})(I_{n^2}-\bar{F}\otimes\bar{F})^{-1}\right](G\otimes G)(Q_w)_s \\ &+\left[(H\otimes\mathscr{O})(I_{n^2}-\bar{F}\otimes\bar{F})^{-1}\right](FL\otimes FL)+(I_p\otimes\Gamma)\big](R_v)_s \end{aligned} \tag{4.54}$$

Considering

$$\mathscr{R}_1(N) = E\left[\begin{array}{cccc} C_1 & \cdots & C_{N-1} \end{array}\right]^T \tag{4.55}$$

$\mathscr{R}_1(N)$ can be estimated using (4.18). Then the new LS problem is given as follows:

$$\begin{aligned} &\min_x \|\mathscr{A}x-\hat{b}\|_2^2 \\ &\text{s.t } Q_w,\ R_v \geq 0, \quad Q_w = Q_w^T, R_v = R_v^T \end{aligned} \tag{4.56}$$

$$\begin{aligned} \mathscr{A} &= \left[\begin{array}{cc} (H\otimes\mathscr{O})(I_{n^2}-\bar{F}\otimes\bar{F})^{-1}(G\otimes G) & (H\otimes\mathscr{O})(I_{n^2}-\bar{F}\otimes\bar{F})(FL\otimes FL)+(I_p\otimes\Gamma) \end{array}\right] \\ x &= \left[\begin{array}{cc} (Q_w)_s^T & (R_v)_s^T \end{array}\right]^T \\ \hat{b} &= \hat{\mathscr{R}}_1(N)_s \end{aligned} \tag{4.57}$$

New necessary and sufficient conditions for uniqueness of the estimates are derived taking into account the new formulation. These results can be summarized as follows:

Theorem 5 *If (F,H) is observable and F is non-singular, the optimization (4.51) with the above assumptions has a unique solution iff $dim[Null(M)] = 0$, where*

$$M = (C\otimes I_n)(I_{n^2}-\bar{F}\otimes\bar{F})^{-1}(G\otimes G)\mathscr{D}_g$$

The second contribution deals with a relationship to compute the weighting matrix W guaranteeing the minimum variance of the estimates among all constrained linear unbiased estimators. Namely, an iterative methodology to find the optimal weighting matrix is provided. However, the authors made the following claims: (i) the convergence of the iterative scheme could not be tested because of the numerical burden; (ii) large data sets are needed to guarantee a reliable weight estimation; (iii) the computation of such a weighting matrix could become prohibitively large even for a small dimensional problem with large data sets.

For these reasons, this algorithm will not be considered in the following experimental tests.

Finally, assuming $W = I$ in (4.51) the problem can be transformed to the problem of identifying the minimal disturbance structure affecting the system state, also with the unknown covariances. This problem is stated as follows

$$\begin{aligned} \Phi_* = \min_{Q,R_v} & \underbrace{\left\| \mathscr{A} \begin{bmatrix} (Q)_s \\ (R_v)_s \end{bmatrix} - \hat{b} \right\|_2^2}_{\Phi} + \rho \cdot \mathrm{rank}(Q) \\ \text{subject to,} \quad & Q, R_v \geqslant 0, \quad Q = Q^T, \quad R_v = R_v^T \end{aligned} \tag{4.58}$$

where $Q = GQ_wG^T$, and ρ is a tuning parameter. Since the constraints are in the form of convex linear matrix inequalities and the optimization problem is convex, the complexity of the problem is mainly due to the term $rank(Q)$. In fact, as the rank term only assumes integer values, this makes the problem computationally NP hard. This drawback can be tackled changing the functional as follows:

$$\begin{aligned} \Phi_* = \min_{Q,R_v} & \underbrace{\left\| \mathscr{A} \begin{bmatrix} (Q)_s \\ (R_v)_s \end{bmatrix} - \hat{b} \right\|_2^2}_{\Phi} + \rho \mathrm{Tr}(Q) \\ \text{subject to,} \quad & Q, R_v \geqslant 0, \quad Q = Q^T, \quad R_v = R_v^T \end{aligned} \tag{4.59}$$

Finally, as the above optimization problem can be transformed into an autocovariance least-squares with semidefinite programming problem (henceforth ALS-SDP) the following Theorem can be stated:

Theorem 6 *A solution $(\hat{Q}, \hat{R}_v)$ to the ALS-SDP in* (4.59) *is unique iff $dim[Null(M)] = 0$, where M is as in Theorem 5. Moreover G is any full column rank decomposition of $\hat{Q} = GG^T$*

Although a way to estimate the minimum structure of the model disturbance term is provided, the change from the original problem to the above one is only justified when black-box models are used, i.e, in systems where the disturbance structure is completely unknown.

4.3 Case studies

In this section the schemes presented in [32, 35] are tested using two known academic examples. To quantify the performance achieved in the application of the presented techniques, different tests are performed, each corresponding to a batch of data randomly generated accordingly to (4.1). Denote with Q_d, and R_d the vectors containing the true main diagonals of the covariances of the modeling and measurement noises respectively. Moreover, N_t is the number of data sets used to assess the mean and variance of each covariance estimation method, $\hat{Q}_i$, and $\hat{R}_i$ are the estimates of Q_d, and R_d respectively, obtained with the data generated in test i. To test the quality of the covariance matrix estimation performances, the following indexes are used:

- The Root Mean Square (RMS) error,

$$RMS_q = \sqrt[2]{\frac{1}{N_t}\sum_{i=1}^{N_t} \|Q_d - \hat{Q}_i\|_2^2}, \qquad RMS_r = \sqrt[2]{\frac{1}{N_t}\sum_{i=1}^{N_t} \|R_d - \hat{R}_i\|_2^2}$$

- The mean of the ∞-norm of errors (MIE),

$$MIE_q = \frac{1}{N_t}\sum_{i=1}^{N_t}\|Q_d - \hat{Q}_i\|_\infty, \qquad MIE_r = \frac{1}{N_t}\sum_{i=1}^{N_t}\|R_d - \hat{R}_i\|_\infty$$

4.3.1 Mehra's example

The example presented in [32] is tested with both Mehra's and the ALS algorithms. Consider the following linear time-invariant discrete-time system:

$$F = \begin{bmatrix} 0.75 & -1.74 & -0.3 & 0 & -0.15 \\ 0.09 & 0.91 & -0.0015 & 0 & -0.008 \\ 0 & 0 & 095 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0 & 0.905 \end{bmatrix}, \quad G = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 24.64 & 0 & 0 \\ 0 & 0.835 & 0 \\ 0 & 0 & 1.83 \end{bmatrix} \tag{4.60}$$
$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The data are generated according to the following distributions:

$$w(k) \sim N\left(0, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right), \qquad v(k) \sim N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \tag{4.61}$$

Note that

$$Q_d = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad R_d = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The initial values of Q_w and R_v, i.e, the covariances used to generate the innovations, are:

$$Q_0 = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.75 \end{bmatrix}, \qquad R_0 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix} \tag{4.62}$$

Mehra's algorithm

- **First Scenario**: the Mehra's algorithm is tested with $N_t = 100$ and $N_d = 1000$ samples in each data set. As mean estimates of the main diagonals we have:

$$Q_m = \begin{bmatrix} 1.0007 \\ 1.4340 \\ 1.0794 \end{bmatrix}, \quad R_m = \begin{bmatrix} 1.3054 \\ 0.9922 \end{bmatrix}$$

Table 4.1: Performance measurements using different amount of data. Mehra's algorithm

N_d	$\|RMS_q\|$	$\|RMS_r\|$	$\|MIE_q\|$	$\|MIE_r\|$
10^3	3.8078	4.2695	3.0682	3.4597
10^4	1.0694	1.2339	0.8821	1.0353
10^5	0.3597	0.4002	0.2981	0.3297

- **Second Scenario**: the Mehra's algorithm is tested with $N_t = 100$ and $N_d = 10^4$. As mean estimates of the main diagonals we have:

$$Q_m = \begin{bmatrix} 1.0010 \\ 1.0149 \\ 0.9537 \end{bmatrix}, \quad R_m = \begin{bmatrix} 1.0230 \\ 0.9962 \end{bmatrix}$$

- **Third Scenario**: the Mehra's algorithm is tested with $N_t = 100$ and $N_d = 10^5$. As mean estimates of the main diagonals we have:

$$Q_m = \begin{bmatrix} 0.9996 \\ 0.9887 \\ 0.9774 \end{bmatrix}, \quad R_m = \begin{bmatrix} 1.0331 \\ 1.0043 \end{bmatrix}$$

The performance indexes are presented in Table 4.1. Notice that errors less than 5% are achieved using a considerable amount of data ($N_d > 10^4$).

ALS algorithm

ALS algorithm is tested using the Algorithm 2. Thanks to the possibility to define constraints, only numerical values greater or equal to zero are expected on the main diagonals.

- **First Scenario**: the ALS algorithm is tested with $N_t = 100$ and $N_d = 10^3$. The tuning parameter is chosen as $N = 10$. As mean estimates of the main diagonals we have:

$$Q_m = \begin{bmatrix} 0.9901 \\ 1.4340 \\ 1.0032 \end{bmatrix}, \quad R_m = \begin{bmatrix} 1.1643 \\ 0.6790 \end{bmatrix}$$

- **Second Scenario**: the ALS algorithm is tested with $N_t = 100$ and $N_d = 10^4$. The tuning parameter is chosen as $N = 10$. As mean estimates of the main diagonals we have:

$$Q_m = \begin{bmatrix} 0.9973 \\ 1.0457 \\ 0.9847 \end{bmatrix}, \quad R_m = \begin{bmatrix} 1.0192 \\ 0.9667 \end{bmatrix}$$

- **Third Scenario**: the ALS algorithm is tested with $N_t = 100$ and $N_d = 10^5$. The tuning parameter is chosen as $N = 10$. As mean estimates of the main diagonals we have:

Table 4.2: Performance measurements using different amount of data. ALS algorithm

N_d	$\|RMS_q\|$	$\|RMS_r\|$	$\|MIE_q\|$	$\|MIE_r\|$
10^3	1.6399	1.0288	1.3323	0.8369
10^4	0.7965	0.4961	0.6546	0.4092
10^5	0.3118	0.1942	0.2464	0.1533

$$Q_m = \begin{bmatrix} 0.9994 \\ 1.0404 \\ 0.9962 \end{bmatrix}, \quad R_m = \begin{bmatrix} 1.0043 \\ 0.9773 \end{bmatrix}$$

The performance indexes for the three scenarios are presented in Table 4.2. It is apparent that errors less than 5% are reached with a considerable amount of data, but an order of magnitude less than with the Mehra's algorithm. In Figure 4.1 we show the variation of the root mean square norm as the number of measured data increases for both algorithms.

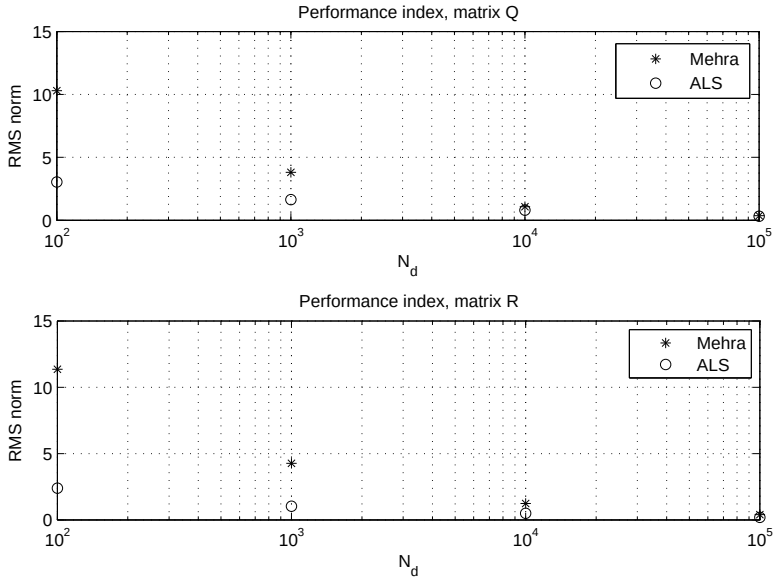


Figure 4.1: Root mean square norm (RMS) as the number of available data grows. Both methods applied to the Mehra's example.

4.3.2 Example from [35]

Consider the discrete-time system described for the following matrices:

$$F = \begin{bmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad H = \begin{bmatrix} 0.1 & 0.2 & 0 \end{bmatrix} \tag{4.63}$$

The data are generated using noise sequences with covariances $Q = 0.5$, $R = 0.1$. The L-innovations are calculated with a filter gain corresponding to incorrect noise variances $Q_0 = 0.2$, $R_0 = 0.4$.

This method is performed using $N_d = 1000$ data points. As in this example only one-element matrices must be estimated, the results can be presented graphically. In Figure 4.2 are shown the results provided by the ALS algorithm using a tuning parameter $N = 15$. The desired results are obtained after performing 200 simulations in order to illustrate the mean and variances of the estimator. The original contribution proposed $N = 15$, but similar results are obtained using other values of this parameter.

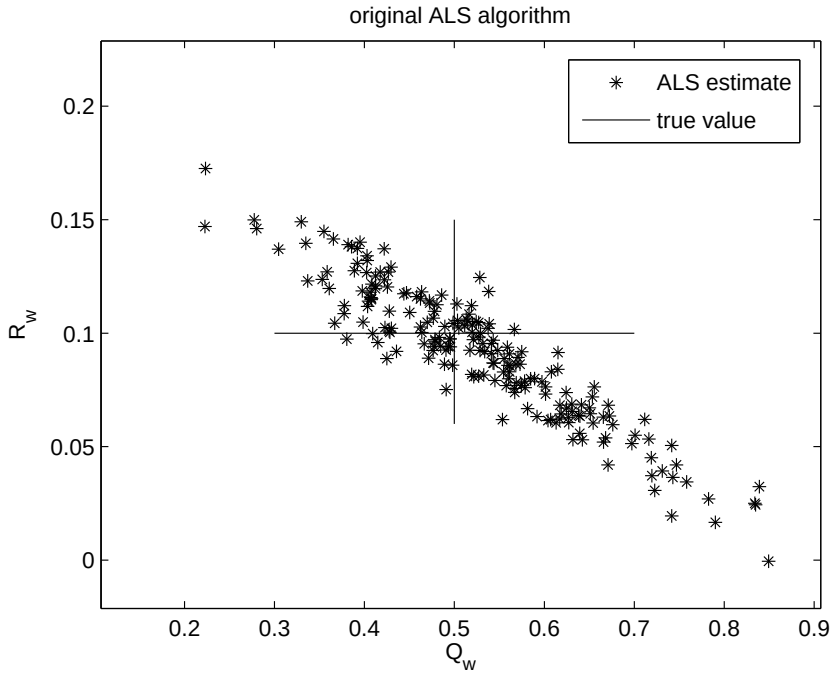


Figure 4.2: ALS method applied to a case study.

The algorithm has a good performance using different values of the parameter N, and even if the initial tuning is not close to the real values. The main drawback of this algorithm and those based on this is the need to perform a considerably amount of simulations in order to find the mean values of the entries of the covariance matrices. Moreover, although many values of the tuning parameter N make the algorithm work in a desired way, the choice of the best tuning parameter N is not a straightforward task. Also, there is not a criterion to know when the covariance matrices are well identified.

Mehra's algorithm was tested on the same example. Also in this simulation the method is performed using $N_d = 1000$ data points, and 200 data sets in order to compare the mean and variance values with the ALS results. The results of the estimates are shown in Figure 4.3. As it can be seen from this figure, the estimation of the Q_w parameter has a considerably bigger variance compared with the one obtained using the ALS scheme. In Figure 4.4 the evolution of the RMS measures with respect to the number of samples N_d is shown for both methods.

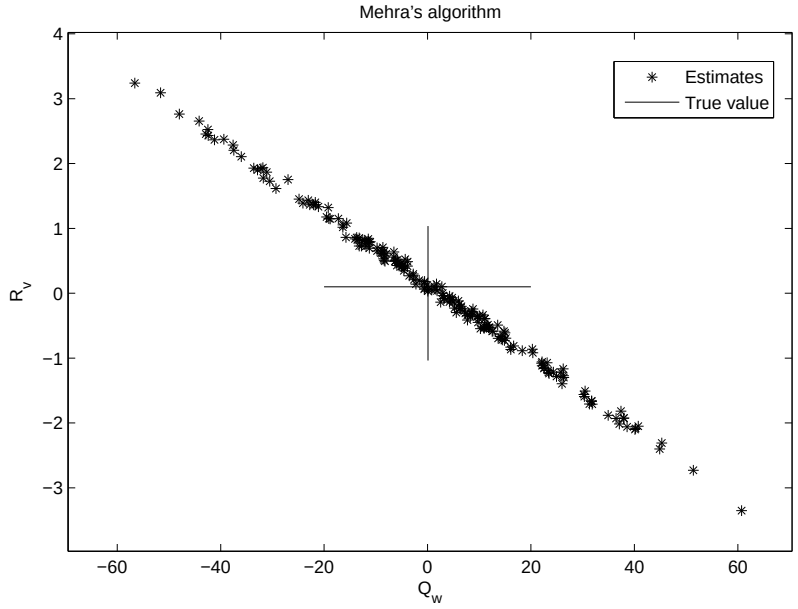


Figure 4.3: Mehra's algorithm applied to a case study.

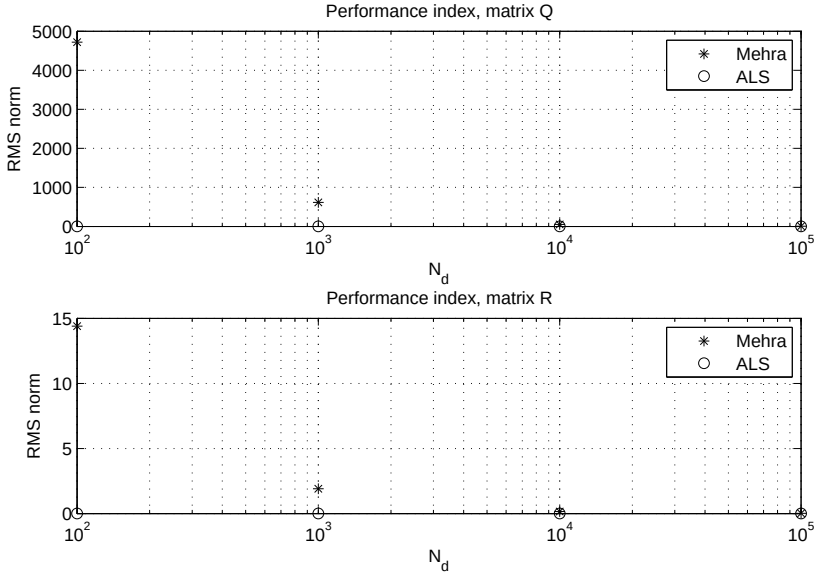


Figure 4.4: Performance index as the amount of data grows.

4.3.3 Application of the ALS Method to the model of a reach of a Hydro Power Valley Plant

The ALS-based methods have shown better results than previous covariance estimation schemes. Both the classic ALS scheme and the reduced ALS scheme described by (4.56) and (4.57) were tested on the model of one reach of a Hydro Power Valley [49]. Reduced ALS scheme is selected in view of its simplicity of implementation and reliability. First the reach model is presented, then results on covariance estimation using the reduced ALS scheme will be shown.

A hydro power valley is a system of lakes, reaches, ducts, penstocks, dams, pumps, valves and turbines which are interconnected together and controlled in order to generate electric power [40]. In this test, only a single reach model is considered.

The Saint Venant nonlinear, first-order system of partial differential equations (PDE) represents the state of the art for modeling one-dimensional river hydraulics with constant fluid density [49]. The hydraulic states of the river are described by two variables: the water depth $H(z,t)$ $[m]$, and the discharge across the section $Q(z,t)$ $\left[m^3/s\right]$, both varying as a function of space z and time t. The river dynamics are usually expressed as [49, 40]:

$$\begin{aligned} \frac{\partial Q}{\partial z} + \frac{\partial S}{\partial t} &= 0 \\ \frac{1}{g}\frac{\partial}{\partial t}\left(\frac{Q}{S}\right) + \frac{1}{2g}\frac{\partial}{\partial t}\left(\frac{Q^2}{S^2}\right) + \frac{\partial H}{\partial z} + I_f - I_0 &= 0 \end{aligned} \tag{4.64}$$

$$\begin{aligned} \frac{\partial Q}{\partial z} + B\frac{\partial H}{\partial t} &= 0 \\ \frac{1}{gB}\frac{\partial}{\partial t}\left(\frac{Q}{H}\right) + \frac{1}{2gB^2}\frac{\partial}{\partial t}\left(\frac{Q^2}{H^2}\right) + \frac{\partial H}{\partial z} + I_f - I_0 \end{aligned} \tag{4.65}$$

where $S(z,t)$ is the wetted area $\left[m^2\right]$, I_f is the friction slope, I_0 the bed slope and g the gravitational acceleration $\left[m/s^2\right]$. The friction slope I_f is defined by the Manning-Strickler formula:

$$I_f(z,t) = \frac{(Q/S)^2}{k_{str}^2 R^{4/3}}, \qquad R = \frac{S}{P} \tag{4.66}$$

where k_{str} is the Strickler coefficient $\left[m^{1/3}/s\right]$, $R(z,t)$ is the hydraulic radius, and $P(z,t)$ the wetted perimeter $[m]$.

The first Saint Venant equation of (4.65) originates from the conservation of mass principle while the second equation results from the conservation of momentum. All other parameters are derived from the river geometry as shown in Figure 4.5.
Assuming that the cross section of the river can be approximated as a rectangle and we consider the river width B constant along the river, then:

$$B(z,t) = B, \qquad P(z,t) = 2H(z,t) + B \tag{4.67}$$

A simple way to implement and simulate a PDE model is to discretize it into several ODE's (Ordinary Differential Equations), by substituting the space derivatives with their corresponding finite differences [40]. To obtain an ODE model we divide our river into $n+1$ small cross sections along the

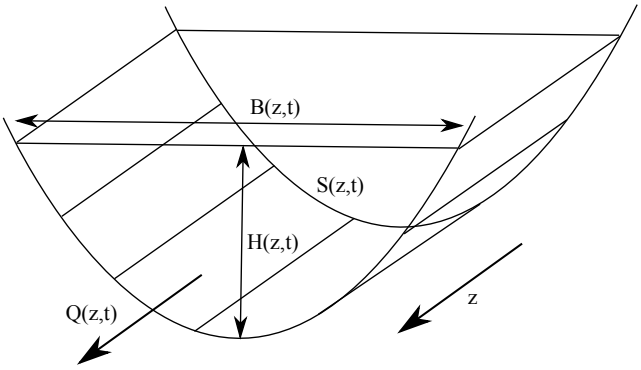


Figure 4.5: Variables definition in a river cross section [49, 40]

direction of the river, see Figure 4.6. To avoid unnecessary stiffness we overlap the crossings section of the different variables. The H variables are calculated at the crossing of each section and the variable Q is calculated in the middle of every section. The discretization is made by the finite difference method. To approximate out derivatives we use the first term in the Taylor series expansions of Q and H. Dividing the reach in $n+1$ cells we obtain the following ODE model where $\Delta z = L_c/n$, and L_c is the length of the reach:

$$\begin{aligned}
\frac{dH_1}{dt} &= -\frac{1}{B}\frac{Q_2 - Q_1}{\Delta z/2} \\
\frac{dQ_2}{dt} &= -\frac{2Q_2}{BH_2}\frac{Q_2 - Q_1}{\Delta z/2} + \left(\frac{Q_2^2}{BH_2^2} - gBH_2\right)\frac{H_3 - H1}{\Delta z} \\
&\quad - \frac{gBH_2}{k_{str}^2}\left(\frac{B+2H_2}{BH_2}\right)^{4/3}\left(\frac{Q_2}{BH_2}\right)^2 + gBI_oH_2 \\
&\vdots \\
\frac{dH_{2i-1}}{dt} &= -\frac{1}{B}\frac{Q_{2i} - Q_{2i-2}}{\Delta z} \\
\frac{dQ_{2i}}{dt} &= -\frac{2Q_{2i}}{BH_{2i}}\frac{Q_{2i} - Q_{2i-2}}{\Delta z} + \left(\frac{Q_{2i}^2}{BH_{2i}^2} - gBH_{2i}\right)\frac{H_{2i+1} - H2i - 1}{\Delta z} \\
&\quad - \frac{gBH_{2i}}{k_{str}^2}\left(\frac{B+2H_{2i}}{BH_{2i}}\right)^{4/3}\left(\frac{Q_{2i}}{BH_{2i}}\right)^2 + gBI_oH_{2i}, \qquad i = 2, \cdots, n \\
\frac{dH_{2n+1}}{dt} &= -\frac{1}{B}\frac{Q_{2n} - Q_{2n+1}}{\Delta z/2}
\end{aligned} \tag{4.68}$$

where H_i are referred as heights and Q_i are the flow rates. The even heights can be obtained by using a linear relationship:

$$H_{2i} = \frac{H_{2i+1} + H_{2i-1}}{2}, \qquad i = 1, ..., n$$

Assuming $n = 4$, previous model can be linearized around the following operating point [49]:

$$x_{op} = \begin{bmatrix} 3.8346 & 300 & 7.1073 & 300 & 10.4024 & 300 & 13.7008 & 300 & 17 \end{bmatrix}^T$$

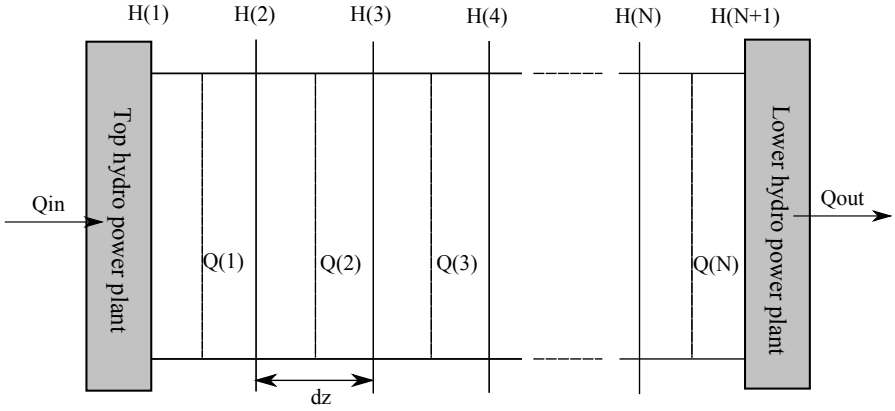


Figure 4.6: Spatial discretization [49, 40]

and then discretized using a sampling time of $1[s]$ to get the following linear time-invariant, discrete-time model:

$$F = \begin{bmatrix} 0.99 & 0 & 0.0001 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5.37 & 0.98 & -5.28 & 0 & -0.0001 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.99 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0019 & 0.0007 & 8.57 & 0.99 & -8.56 & 0 & -0.0002 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.99 & 0 & 0.0001 & 0 & 0 \\ 0 & 0 & 0.0023 & 0.0006 & 11.80 & 0.99 & -11.80 & 0.0001 & -0.0003 \\ 0 & 0 & 0 & 0 & 0.0001 & 0 & 0.99 & 0 & 0.0001 \\ 0 & 0 & 0 & 0 & 0.0026 & 0.0005 & 15.0335 & 0.9993 & -15.0329 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0002 & 0 & 0.99 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.0002 \\ 0 \end{bmatrix} \quad G = \begin{bmatrix} 0 \\ 0.0022 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

where the system state is defined as:

$$x = \begin{bmatrix} H1 & Q2 & H3 & Q4 & H5 & Q6 & H7 & Q8 & H9 \end{bmatrix}^T$$

As initial conditions:

$$Q_{2i}(0) = Q_0, \;\; i = 1, ..., n-1; \qquad H_{2i+1}(0) = H_N, \;\; i = 1, ..., n \tag{4.69}$$

where H_N is the normal height, and the boundary condition on the upstream and downstream of the reach:

$$Q_1(0) = Q_{in} \tag{4.70}$$

with Q_{in} considered as a system disturbance.
In this model the manipulated variable is the discharge through the turbine Q_t, which is related with the last flow rate as follows:

$$Q_{2n+1} = Q_t + Q_D(H_{2n+1})$$

where Q_D is a constant value discharge, known as the weir-discharge.

Noises (model and measurement) are generated accordingly to the following distributions:

$$w(k) \sim N(0,10), \qquad v(k) \sim N\left(0, \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right), \tag{4.71}$$

which correspond to the modeling uncertainty and the sensor variance, in this case two flow rate sensor and a level sensor.

As discussed, the estimation of the covariances using the constrained and reduced ALS scheme has the following degrees of freedom: the number of data in each data set N_d and the tuning parameter N. Therefore, the results are presented as follows:

- The amount of data is fixed as $N_d = 10^4$. Then, the results are evaluated as the tuning parameter is changed.
- The N parameter is chosen in such a way that any increase in it value does not generate a significant reduction on the variance of the estimates. Fixing this parameter, the simulations are performed using different amount of data in each data set.

In all the aforementioned cases we perform $N_t = 200$ different tests in order to obtain a reliable statistical characterization of the results.

In the first test, $N_t = 200$ sets of fixed data (each one with $N_d = 10^4$) are used to estimate the covariances as the N parameter is changed. In Table 4.3 the mean of the state noise covariance Q_w is shown together also with the corresponding performance indexes as N grows.
It is apparent that the accuracy of the estimation increases as N increases. However there is a point at which a considerable increase of N does not imply a considerable reduction of the estimation variance.

In the second test, the tuning parameter is taken as $N = 47$. The starting values of Q_w and R_v, i.e, the covariances used to generate the innovations are:

$$Q_0 = 9, \qquad R_0 = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} \tag{4.72}$$

Then, the following scenarios are considered:

Table 4.3: Performance indexes as N is varied

N	Q_m	$\|RMS_q\|$	$\|RMS_r\|$	$\|MAE_q\|$	$\|MAE_r\|$
2	21.6647	354.8952	2.8091×10^{-4}	14.1385	0.0161
10	11.0444	46.3145	2.3268×10^{-4}	5.3920	0.0146
20	10.0479	24.9942	2.3103×10^{-4}	4.0660	0.0145
30	9.9039	18.0471	2.2916×10^{-4}	3.5151	0.0145
40	9.6083	15.8991	2.2963×10^{-4}	3.2015	0.0145
50	9.4683	14.0892	2.2986×10^{-4}	3.0749	0.0145

Table 4.4: Performance measurements using different amount of data. Reduced ALS algorithm

N_d	$\|RMS_q\|$	$\|RMS_r\|$	$\|MIE_q\|$	$\|MIE_r\|$
10^3	55.6970	0.0020	6.1658	0.0428
10^4	12.5323	$2.1163 \times 10-4$	2.8934	0.0134
5×10^4	2.9040	4.6010×10^{-5}	1.3733	0.0064

- **First Scenario**: the covariances are estimated with $N_d = 10^3$. A graphical sketch of the estimation is presented in Figure 4.7 where each component of the main diagonal of R_v is plotted against the one-element matrix Q_w. As mean values of the main diagonals we have:

$$Q_m = 10.0725, \quad R_m = \begin{bmatrix} 0.1002 \\ 0.1002 \\ 0.9991 \end{bmatrix}$$

- **Second Scenario**: the covariances are estimated with $N_d = 10^4$. A graphical sketch of the estimated covariances is presented in Figure 4.8 where each component of the main diagonal of R_v is plotted against the one-element matrix Q_w. As mean values of the main diagonals we have:

$$Q_m = 9.7339, \quad R_m = \begin{bmatrix} 0.1001 \\ 0.0999 \\ 0.9992 \end{bmatrix}$$

- **Third Scenario**: the covariances are estimated with $N_d = 5 \times 10^4$. A graphical sketch of the estimated covariances is presented in Figure 4.9 where each component of the main diagonal of R_v is plotted against the one-element matrix Q_w. As mean values of the main diagonals we have:

$$Q_m = 10.0968, \quad R_m = \begin{bmatrix} 0.1000 \\ 0.1000 \\ 0.9996 \end{bmatrix}$$

The performance indexes for the aforementioned scenarios are presented in Table 4.4.
In this test the covariances have been estimated using a fixed number of data sets $N_t = 200$, the N parameter, the initial guesses, and varying the amount of data in each one of these sets. Figure 4.10 shows how the estimates are considerably improved as the amount of data is increased. However, it

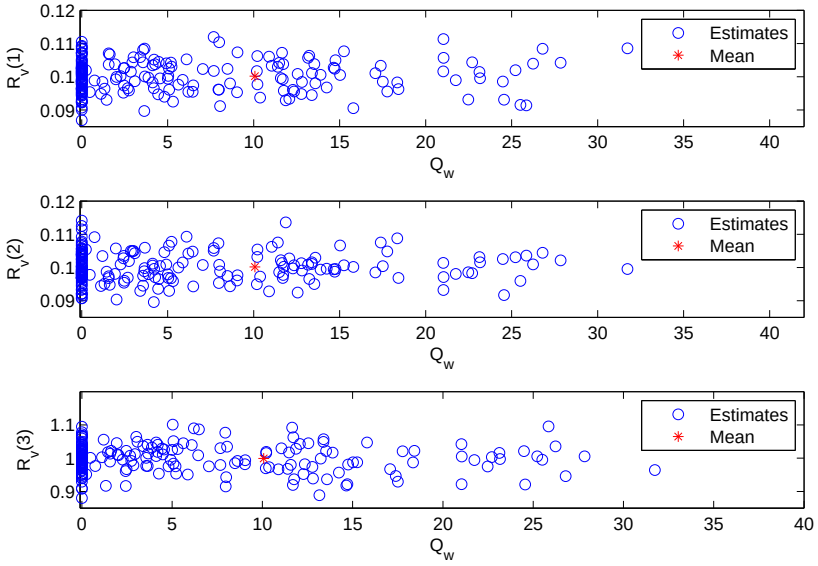


Figure 4.7: Covariance estimation in the single reach model. First scenario. Each element of R_v is plotted against the Q_w

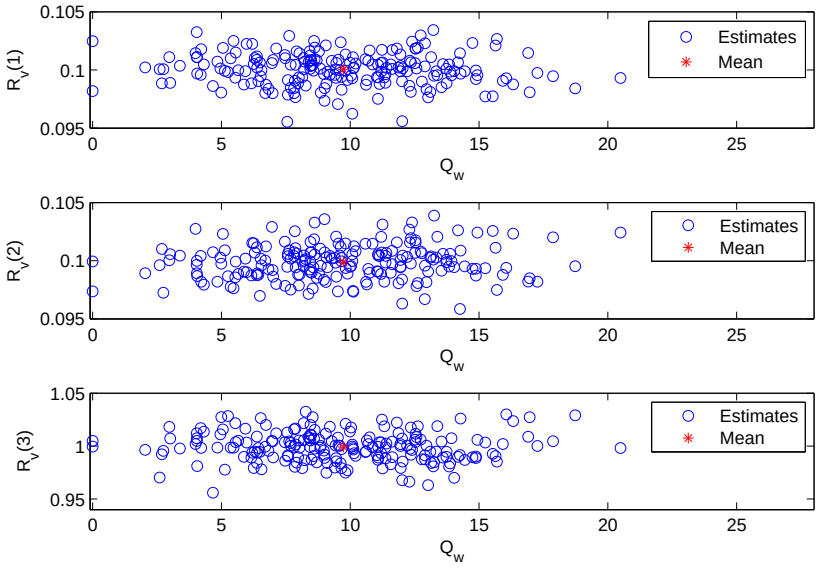


Figure 4.8: Covariance estimation in the single reach model. Second scenario. Each element of R_v is plotted against the Q_w

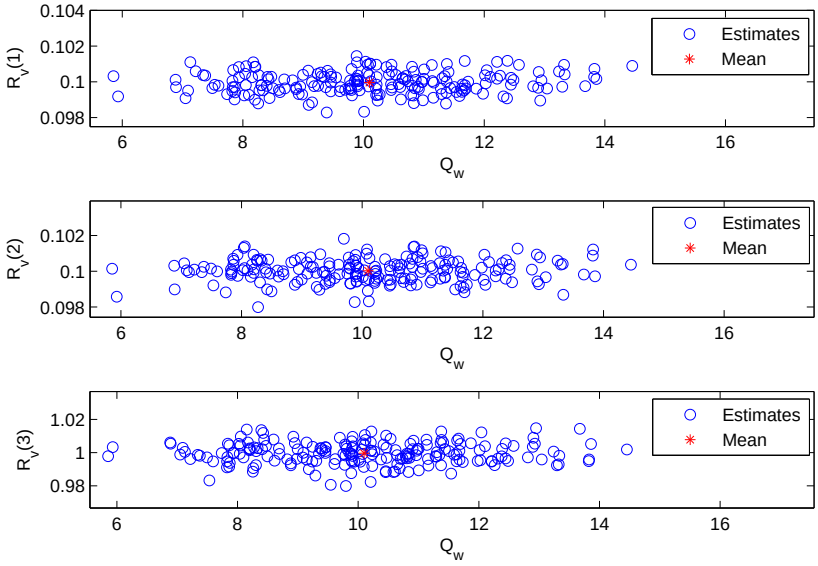


Figure 4.9: Covariance estimation in the single reach model. Third scenario. Each element of R_v is plotted against the Q_w

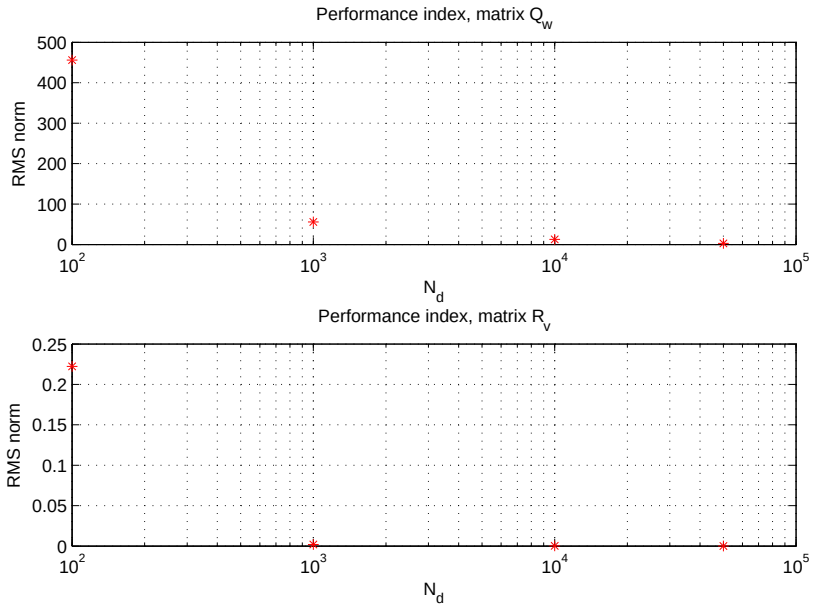


Figure 4.10: Root mean square norm (RMS) as the number of available data grows. Reduced ALS method applied to the single reach model.

can be noticed that in order to achieve small variance in the matrix Q_w the amount of data must be increased considerably.

Bibliography

[1] B. M. Åkesson, J. B. Jørgensen, and S. B. Jørgensen. "A Generalized Least-Squares Method for Covariance Estimation". *Proc. of the ACC*, New York City, USA, 2007.

[2] B. M. Åkesson, J. B. Jørgensen, N. K. Poulsen, and S. B. Jørgensen. "A Generalized Autocovariance Least-Squares Method for Kalman Filter Tuning". *Journal of Process Control*, **18**, pp: 769-779, 2008.

[3] A. Alessandri, M. Baglietto, and G. Battistelli. Receding-horizon estimation for discrete-time linear systems. *IEEE Trans. on Automatic Control*, 48(3):473 – 478, 2003.

[4] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: new stability results and approximation schemes. *Automatica*, 44:1753 – 1765, 2008.

[5] A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli. A neural state estimator with bounded errors for nonlinear systems. *IEEE Trans. on Automatic Control*, 44(11):2028 – 2042, 1999.

[6] P. Alriksson and A. Rantzer. Distributed Kalman filtering using weighted averaging. In *Proc. of the 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, 2006.

[7] D. L. Alspach. A parallel filtering algorithm for linear systems with unknown time varying noise statistics. IEEE Trans. on Automatic Control, 19(5), pp:552-556, 1974.

[8] A. Averbuch, S. Itzikovitz, and T. Kapon. "Radar target tracking - Viterbi versus IMM". IEEE Trans. on Aerospace and Electronic Systems, 27(3), pp:550-563, 1991.

[9] P. R. Bélanger. "Estimation of noise covariance matrices for a linear time-varying stochastic process". Automatica, **10**, pp: 267-275, 1974.

[10] D. P. Bertsekas. *Linear network optimization: algorithms and codes*. MIT Press, 1991.

[11] T. Bohlin. "Four cases of identification of changing systems". In R.K. Mehra and D.G. Lainiotis (Eds.), System Identification: Advances and case studies (1st ed.). New York: Academic Press., 1976.

[12] B. Bollobás. *Modern graph theory*, volume 184 of *Graduate texts in mathematics*. Springer, New York, 1998.

[13] P. E. Caines and D. Q. Mayne. On the discrete-time matrix Riccati equation of optimal control. *International Journal of Control*, 12(5):785–794, 1970.

[14] G. C. Calafiore and F. Abrate. Distributed linear estimation over sensor networks. *International Journal of Control*, 82(5):868 – 882, 2009.

[15] B. Carew and P. R. Bélanger. "Identification of optimum filter steady-state gain for systems with unknown noise covariances". IEEE Trans. on Automatic Control, 18(6):582-587, 1973.

[16] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas In Communications*, (4):622 – 633, 2008.

[17] S. Dashkovskiy, B. S. Rüffer, and F.R. Wirth. An ISS small gain theorem for iss general networks. *Mathematics of Control, Signals, and Systems*, 19(2):93–122, 2007.

[18] L. Farina and S. Rinaldi. *Positive Linear Systems*. J. Wiley and Sons, 2000.

[19] G. Ferrari-Trecate, D. Mignone, and M. Morari. Moving horizon estimation for hybrid systems. *IEEE Trans. on Automatic Control*, 47(10):1663 – 1676, Oct. 2002.

[20] M. Farina, G. Ferrari-Trecate, and R. Scattolini. Distributed Moving Horizon Estimation for Linear Constrained Systems. *IEEE Trans. on Automatic Control*, 55(11):2462 – 2475, Nov. 2010.

[21] M. Farina, G. Ferrari-Trecate, and R. Scattolini. Distributed moving horizon estimation for sensor Networks. *IFAC Workshop on Estimation and Control of Networked Systems*, 126 – 131, Venice, 2009.

[22] G. C. Goodwin, M. M. Seron, and J. A. De Doná. *Constrained Control and Estimation*. Springer, New Jersey, 2005.

[23] H. R. Hashemipour, S. Roy, and A. J. Laub. Decentralized structures for parallel Kalman filtering. *IEEE Trans. on Automatic Control*, 33(1):88 – 94, Jan. 1988.

[24] H. Heffes. "The Effect of Erroneous Models on the Kalman Filter Response". *IEEE Trans. on Autom. Control*, **AC-11**, pp: 541-543, 1966.

[25] C. G. Hilborn and D. G. Lainiotis. "Optimal estimation in the presence of unknown parameters". IEEE Trans. on System Science and Cybernetics, 5(1), pp:38-43, 1969.

[26] G. M. Jenkins, and D. G. Watts. "Spectral Analysis and its Applications". San Francisco: Holden Day Publ., 1968.

[27] R. E. Kalman "New Methods and Results in Linear Prediction and Filter Theory". *Proc. Symp. on Engineering Applications on Random Function Theory and Probability*, New York: Wiley, 1961.

[28] R. L. Kashyap. "Maximum likelihood identification of stochastic linear systems". IEEE Trans. on Automatic Control, 15(1):25-34, 1970.

[29] M. Kamgarpour and C. Tomlin. Convergence properties of a decentralized Kalman filter. *Proc. 47th IEEE Conference on Decision and Control*, pages 3205 – 3210, 2008.

[30] H. K. Khalil. *Nonlinear systems - Third edition*. Prentice Hall, 2000.

[31] U. A. Khan and J. M. F. Moura. Distributing the Kalman filter for large-scale systems. *IEEE Trans. on Signal Processing*, 56(10):4919 – 4935, Oct. 2008.

[32] R. K. Mehra. "On the Identification of Variances and Adaptive Kalman Filtering". *IEEE Trans. on Autom. Control*, **AC-15**, No 2, pp: 175-184, 1970.

[33] R. K. Mehra. "Approaches to adaptive filtering". *IEEE Trans. on Autom. Control*, **17**, No 17, pp: 903-908, 1972.

[34] K. A. Myers and B. D. Tapley. "Adaptive sequencial estimation with unknown noise statistics". IEEE Trans. on Automatic Control, 19(5), pp:623-625, 1976.

[35] B. J. Odelson, M. R. Rajamani and J. B. Rawlings. "A New Autocovariance Least-Squares Method for Estimating Noise Covariances". *AUTOMATICA*, **42**, pp: 303-308, 2006.

[36] B. J. Odelson, A. Lutz, and J. B. Rawlings. "The Autocovariance Least-Squares Method for Estimating Convariances: Application to Model-Based Control of Chemical Reactors". *IEEE Trans. on Control Systems Technology*, **14**, No. 3, pp: 532-540, 2006.

[37] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. *Proc.* 44^{th} *IEEE Conference on Decision and Control - European Control Conference*, pages 8179 – 8184, 2005.

[38] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. *Proc.* 46^{th} *IEEE Conference on Decision and Control*, pages 5492 – 5498, 2007.

[39] R. Olfati-Saber and J. Shamma. Consensus filters for sensor networks and distributed sensor fusion. *Proc.* 44^{th} *IEEE Conference on Decision and Control - European Control Conference*, pages 6698 – 6703, 2005.

[40] F. Petrone. "Model Predictive Control of a Hidro Power Valley". M.Sc Thesis, Politecnico di Milano, 2009-2010.

[41] B. S. Rao and H. F. Durrant-Whyte. Fully decentralised algorithm for multisensor Kalman filtering. In *IEE Proc. on Control Theory and Applications, D*, volume 138, pages 413 – 420, Sept. 1991.

[42] C. V. Rao and J. B. Rawlings. Nonlinear moving horizon state estimation. *in F. Allgöwer and A. Zheng, editors*, Nonlinear Model Predictive Control, *Progress in Systems and Control Theory, Birkhauser*, pages 45–70, 2000.

[43] C. V. Rao, J. B. Rawlings, and J. H. Lee. Stability of constrained linear moving horizon estimation. *Proc. American Control Conference*, pages 3387 – 3391, 1999.

[44] C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained linear state estimation - a moving horizon approach. *Automatica*, 37:1619 – 1628, 2001.

[45] C. V. Rao, J. B. Rawlings, and D. Q Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Trans. on Automatic Control*, 48(2):246 – 258, 2003.

[46] M. R. Rajamani. "Data-based Techniques to Improve State Estimation in Model Predictive Control". Doctoral Dissertation, University of Wisconsin-Madison, Wisconsin, USA, 2007.

[47] M. R. Rajamani and J. B. Rawlings. “Estimation of Noise Covariances and Disturbance Structure from Data using Least-Squares with Optimal Weighting”. *AIChE Annual Meeting*, San Francisco, CA, USA, 2006.

[48] M. R. Rajamani and J. B. Rawlings. “Estimation of the disturbance structure from data using semidefinite programming and optimal weighting”. *AUTOMATICA*, **45**, pp: 142-148, 2009.

[49] C. Romani. “Hierarchical and distributed control methods for large scale systems”. Doctoral dissertation, Politecnico di Milano, 2010. Draft version.

[50] D. P. Spanos, R. Olfati-Saber, and R. M. Murray. Approximate distributed Kalman filtering in sensor networks with quantifiable performance. *Fourth International Symposium on Information Processing in Sensor Networks*, pages 133 – 139, 2005.

[51] A. Speranzon, C. Fischione, K. H. Johansson, and A. L. Sangiovanni-Vincentelli. A distributed minimum variance estimator for sensor networks. *IEEE Journal on Selected Areas in Communications*, 26(4):609–621, 2008.

[52] M. Verhaegen and P. Van Dooren. Numerical aspects of different Kalman filter implementations. *IEEE Trans. on Automatic Control*, 31(10):907 – 917, 1986.

[53] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, England, 1965.