UNIVERSITY OF SÃO PAULO
SÃO CARLOS SCHOOL OF ENGINEERING

Kaio Douglas Teófilo Rocha

Robust Distributed Filtering for Sensor Networks under Parametric Uncertainties

São Carlos

2022

Kaio Douglas Teófilo Rocha

Robust Distributed Filtering for Sensor Networks under Parametric Uncertainties

Thesis submitted to the São Carlos School of Engineering at the University of São Paulo as requirement for a Doctor of Science degree - Graduate Program in Electrical Engineering.

Concentration area: Dynamical Systems

Advisor: Prof. Dr. Marco Henrique Terra

FINAL VERSION

Trata-se da versão corrigida da tese. A versão original se encontra disponível na EESC/USP que aloja o Programa de Pós-Graduação de Engenharia Elétrica.

São Carlos

2022

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

Rocha, Kaio Douglas Teófilo
R672r Robust Distributed Filtering for Sensor Networks under Parametric Uncertainties / Kaio Douglas Teófilo Rocha; orientador Marco Henrique Terra. São Carlos, 2022.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em Sistemas Dinâmicos -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2022.

1. Sensor networks. 2. Distributed filtering. 3. Consensus. 4. Uncertain systems. 5. Kalman filter. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE JULGAMENTO

Candidato: Engenheiro **KAIO DOUGLAS TEOFILO ROCHA.**

Título da tese: "Filtragem distribuída robusta para redes de sensores sujeitas à incertezas paramétricas".

Data da defesa: 16/12/2022.

Comissão Julgadora	Resultado
Prof. Titular **Marco Henrique Terra** **(Orientador)** (Escola de Engenharia de São Carlos/EESC-USP)	APROVADO
Prof. Associado **Daniel Ferreira Coutinho** (Universidade Federal de Santa Catarina/UFSC)	APROVADO
Prof. Associado **João Yoshiyuki Ishihara** (Universidade de Brasília/UnB)	APROVADO
Prof. Titular **José Cláudio Geromel** (Universidade de Campinas/UNICAMP)	APROVADO
Prof. Dr. **Tiago Roux Oliveira** (Universidade Estadual do Rio de Janeiro/UERJ)	APROVADO

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica:
Prof. Dr. **João Bosco Augusto London Junior**

Presidente da Comissão de Pós-Graduação:
Prof. Titular **Murilo Araujo Romero**

To my parents, Keila and Douglas.
To my grandparents, Geralda, Zilma, João, and Manuel.

Acknowledgments

I thank God for the gift of life.

I wholeheartedly thank my beloved parents, Keila and Douglas, for being my firm foundation and for all the unconditional support on my every goal.

I thank my advisor, Prof. Dr. Marco Terra, for the opportunity and trust placed in me to carry out this academic research and for the valuable ideas and motivation.

I am grateful for all the friendships that the Laboratory of Intelligent Systems (LASI) has given me. Thank you for making São Carlos feel more like home, for all the support and companionship, and for making all the lunches and dinners at "bandejão" much funnier.

I am also thankful to the University of São Paulo, the São Carlos School of Engineering, and LASI for the excellent infrastructure that made this research project possible.

Finally, I thank the São Paulo Research Foundation (FAPESP) for the financial support (process 2017/16346-4) during the whole duration of this project and for the opportunity to showcase its results at several national and international conferences.

"I wish there was a way to know you're in the good old days
before you've actually left them."
Andy Bernard (The Office - S9 E23)

Abstract

ROCHA, K. D. T. **Robust Distributed Filtering for Sensor Networks under Parametric Uncertainties**. 2022. 174p. Thesis (Doctor) - São Carlos School of Engineering, University of São Paulo, São Carlos, 2022.

In the past few years, we have witnessed the rapid popularization of networked cooperative multi-agent systems, which consistently move towards becoming ubiquitous in our society. As one of the most well-established examples of such systems, sensor networks have been applied to increasingly more complex systems, demanding even more robust, efficient, and reliable technologies. Distributed state estimation is the most fundamental task that one can accomplish with these networks. The main objective of this thesis is to develop robust distributed filtering strategies for sensor networks applied to linear discrete-time systems subject to model parametric uncertainties. Specifically, we deal with two types of uncertainties: norm-bounded and polytopic. To achieve this goal, we also address other related problems, divided into two categories. The first category of problems refers to the single-sensor state estimation task. Within this category, we consider the scenarios in which the underlying models are perfectly known and where they are subject to each of the two kinds of uncertainty. We propose nominal and robust filters for each situation. The second category concerns the networks with multiple sensors, considering the same three scenarios. For each one, we propose both centralized and distributed estimators. We use the average consensus algorithm to obtain the distributed filters, which approximate their centralized counterparts. The proposed filters are based on the celebrated Kalman filter and present a similar recursive and relatively simple structure. We evaluate the performance of the proposed estimators with application examples, in which we also compare them to existing strategies from the related literature.

Keywords: Sensor networks. Distributed filtering. Consensus. Uncertain systems. Kalman filter.

Resumo

ROCHA, K. D. T. **Filtragem Distribuída Robusta para Redes de Sensores Sujeitas a Incertezas Paramétricas**. 2022. 174p. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2022.

Nos últimos anos, tem-se testemunhado a rápida popularização de sistemas multiagentes cooperativos em rede, que consistentemente tendem a se tornar onipresentes em nossa sociedade. Sendo um dos exemplos mais bem estabelecidos de tais sistemas, as redes de sensores têm sido aplicadas a sistemas cada vez mais complexos, exigindo tecnologias cada vez mais robustas, eficientes e confiáveis. A estimação distribuída de estado é a tarefa mais fundamental que podemos realizar com essas redes. O principal objetivo desta tese é desenvolver estratégias robustas de filtragem distribuída para redes de sensores aplicadas a sistemas lineares em tempo discreto sujeitos a incertezas paramétricas. Especificamente, consideram-se dois tipos de incertezas: limitadas em norma e politópicas. Para atingir esse objetivo, outros problemas relacionados também são abordados, divididos em duas categorias. A primeira categoria de problemas refere-se à tarefa de estimativa de estado baseada em um único sensor. Dentro dessa categoria, considera-se o cenário em que os modelos são perfeitamente conhecidos, assim como os em que eles são sujeitos a cada um dos dois tipos de incerteza. São propostos filtros nominais e robustos para cada situação. A segunda categoria diz respeito às redes com múltiplos sensores, considerando os mesmos três cenários. Para cada um, são propostos estimadores centralizados e distribuídos. O algoritmo de consenso é utilizado para obter-se os filtros distribuídos, que aproximam suas versões centralizadas correspondentes. Os filtros propostos são baseados no célebre filtro de Kalman e apresentam uma estrutura recursiva semelhante e relativamente simples. O desempenho dos estimadores propostos é avaliado por meio de exemplos de aplicação, sendo também comparados com estratégias existentes na literatura relacionada.

Palavras-chave: Redes de sensores. Filtragem distribuída. Consenso. Sistemas com incertezas. Filtro de Kalman.

List of Figures

List of Tables

List of abbreviations and acronyms

BDU	Bounded Data Uncertainties
CE	Consensus on Estimates
CI	Consensus on Information
CM	Consensus on Measurements
CKF	Nominal Centralized Kalman Filter
DKCF	Nominal Distributed Kalman Consensus Filter
DKF	Distributed Kalman Filter
HCMCI	Hybrid Consensus on Measurements and Information
KCF	Kalman Consensus Filter
KF	Nominal Kalman Filter
LMI	Linear Matrix Inequality
MSE	Mean Squared Estimation Error
NCMAS	Networked Cooperative Multi-Agent System
PRCKF	Polytopic Robust Centralized Kalman Filter
PRDKCF	Polytopic Robust Distributed Kalman Consensus Filter
PRKF	Polytopic Robust Kalman Filter
RCKF	Robust Centralized Kalman Filter
RDKCF	Robust Distributed Kalman Consensus Filter

RKF	Robust Kalman Filter
UAV	Unmanned Aerial Vehicle

List of symbols

$\mathbb{Z}$	Set of integer numbers
$\mathbb{Z}_+$	Set of positive integer numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	Set of n-dimensional vectors with elements in $\mathbb{R}$
$\mathbb{R}^{n\times m}$	Set of $n \times m$ matrices with elements in $\mathbb{R}$
$\mathbb{C}$	Set of complex numbers
$\mathbb{C}^n$	Set of n-dimensional vectors with elements in $\mathbb{C}$
$\mathbb{C}^{n\times m}$	Set of $n \times m$ matrices with elements in $\mathbb{C}$
I_n	Identity matrix with dimensions $n \times n$
$\mathbf{1}_N$	Column vector of ones with dimensions $N \times 1$
A^T	Transpose of matrix A
A^{-1}	Inverse of matrix A
$A^\dagger$	Moore-Penrose inverse (pseudoinverse) of matrix A
$A^{1/2}$	Square-root factor of matrix A
$\rho(A)$	Spectral radius of A, given by $\max\{\lvert\lambda_i\rvert\}$, where λ_i are the distinct eigenvalues of A
$A \succ 0$	A is symmetric positive definite
$A \succeq 0$	A is symmetric positive semidefinite
$\mathbf{col}(\bullet)$	Operator that stacks its operands in a block-column matrix/vector

$\mathbf{diag}(\bullet)$	Operator that represents a block-diagonal matrix with its operands as diagonal elements
$\otimes$	Kronecker product
$\lvert a \rvert$	Absolute value of scalar a
$\lVert x \rVert$	Euclidean norm of vector x
$\lVert x \rVert_W^2$	Weighted squared Euclidean norm of vector x, given by $x^T W x$
$\lVert A \rVert$	Norm of matrix A, given by its maximum singular value, unless otherwise stated
$X^T P(\bullet)$	Simplified representation of $X^T P X$
$\boldsymbol{E}\{x\}$	Expected value of x
$\mathbb{G}$	Generic graph (directed or undirected)
$\mathbb{S}$	Node set of a graph
$\mathbb{E}$	Edge set of a graph
$\mathcal{N}_i$	Neighborhood of node i
N_i	Cardinality of $\mathcal{N}_i$
$\leftarrow$	Assignment (mapping) operator
$:=$	Definition operator

Contents

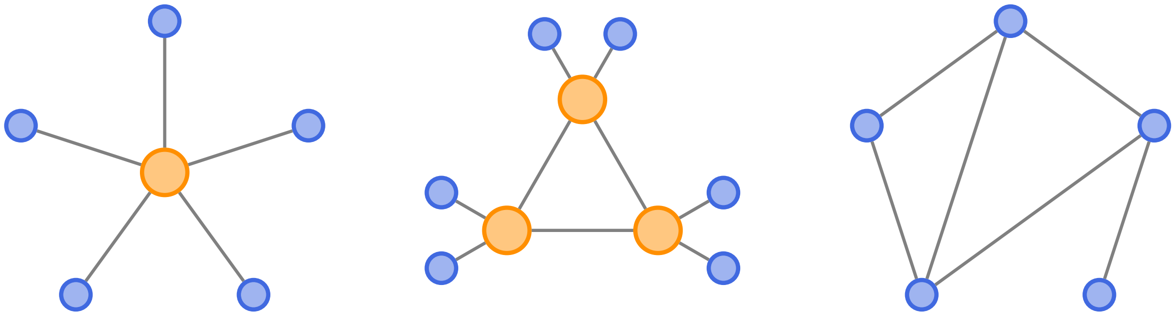
Chapter 1

Introduction

Over the past few decades, networked cooperative multi-agent systems (NCMASs) have received significant attention from researchers of a diverse spectrum of disciplines, including engineering, computer science, physics, biology, and economics, to name a few. Such diversity explains how NCMASs has evolved into an intrinsically multidisciplinary field. As the name suggests, NCMASs consist of multiple dynamical agents that can work together to achieve collective group behaviors or tasks. The agents may represent different entities, depending on the context. For instance, they can be robots, sensors, or computer processes. The interaction is in information exchange via some communication channel, hence the networked denomination.

The continuous and rapid advancements in computer miniaturization, communication, sensing, and actuation technologies have enabled the popularization of NCMASs in a broad range of civilian and military applications. In addition, instead of using a single complex agent to carry out a complicated task, using multiple simple agents can significantly improve operational effectiveness, reduce costs, and increase the reliability of the overall system. Moreover, teams of networked agents can carry out tasks that would be impossible for a single entity to accomplish alone, like surveillance of a large area. Some potential applications of NCMASs include satellite formation flying for space interferometry and surveillance (BEARD; LAWTON; HADAEGH, 2001; TILLERSON; INALHAN; HOW, 2002), formation control of unmanned aerial vehicles (UAVs) (DONG *et al.*, 2016), air and ground surveillance (GROCHOLSKY *et al.*, 2006), healthcare systems (ALEMDAR; ERSOY, 2010; SHAKSHUKI; REID, 2015), microgrid control in smart grids (BIDRAM *et al.*, 2013), and intelligent transportation systems (LEE; PARK, 2012; LU *et al.*, 2014). Another primary application of NCMASs is in sensor networks. They are present in environment monitoring (BAI *et al.*, 2018; OTHMAN; SHAZALI, 2012; MAINWARING *et al.*, 2002), target detection, classification and tracking (ARORA *et al.*, 2004), robotics (LI; SHEN, 2011), smart cities (ZANELLA *et al.*, 2014), industrial cyber-physical systems (DING *et al.*, 2019), and several other applications.

Figure 1 – Types of NCMASs architectures.



(a) Centralized. (b) Decentralized. (c) Distributed.

The connection architecture among the NCMASs can be classified into three types, as shown in Figure 1. In a *centralized* architecture, there is a central node/agent (in orange) with access to all other agents (in blue). One of the main drawbacks of this architecture is that it contains a single point of failure, i.e., if the central node is lost, the whole system is jeopardized. It also does not scale well as the number of agents increases since the central node needs to communicate and process information from all agents. The *decentralized* architecture divides the network into smaller, locally centralized units. The local centers (in orange) directly communicate and process data from a limited set of agents. They can access the rest of the network by interacting with the other local centers. This strategy is more robust than the centralized one since the failure of a single agent only affects its local unit. It is also more scalable, as it divides the information processing among the local centers. Finally, in the *distributed* architecture, there is no central node whatsoever. Every agent is independent and can only interact with a limited set of neighbors, thus reducing communication and processing costs. For this reason, it is the most robust, flexible, and scalable among the three (CHEN; REN, 2019). Given the advantages of the distributed architecture compared to the others, it is the strategy we chose to focus on in this work.

Many ideas that are now well-established tools in distributed filtering and control of NCMASs were inspired by nature. For instance, in flocks of birds, schools of fish, or swarms of bees, the agents exploit local interaction mechanisms in order to achieve collective group objectives that are essential for survival. These behaviors inspired works like the one by Reynolds (1987), which proposed three rules: collision avoidance, velocity matching, and flock centering, producing simulated flocking for computer graphics. Next, Vicsek *et al.* (1995) proposed a simple discrete-time model of autonomous agents (called particles) that move in a plane with the same speed but with different headings. At each time step, each particle updates its heading to its nearest neighbors' average direction of motion, with some added random perturbation. Applying this simple local interaction rule eventually leads all particles to align their headings, even when the neighborhood set of each particle changes over time. Afterwards, Jadbabaie, Lin and Morse (2003) formally addressed the alignment problem introduced by Vicsek *et al.* (1995), from a more theoretical viewpoint, based on graph theory and nonnegative matrices.

The alignment problem described above is a classic example of the so-called *consensus problem*, a fundamental concept in distributed filtering and control of NCMASs. In a consensus protocol, each agent in the network interacts with a limited set of neighbors to agree on a specific quantity. The idea of consensus and its formal study originated in management science and statistics in the 1960s (DEGROOT, 1974). However, consensus problems also have a long history in computer science, being the basis for distributed (LYNCH, 1996) and parallel (BERTSEKAS; TSITSIKLIS, 1989) computing. Later, the pioneering works by Jadbabaie, Lin and Morse (2003) and Olfati-Saber and Murray (2004) stimulated the ever-growing interest in consensus problems in the context of NCMASs, helping to pave the way for the subsequent development of a series of works on the topic. As aforementioned, Jadbabaie, Lin and Morse (2003) provided a theoretical treatment for the alignment behavior observed in the discrete-time model of Vicsek *et al.* (1995). On the other hand, Olfati-Saber and Murray (2004) established a general framework for consensus in networks of continuous-time single integrator agents, including the effects of switching communication topologies and time delays.

This thesis mainly focuses on applying sensor networks to perform consensus-based robust distributed filtering of systems subject to parametric uncertainties. However, we also address the robust estimation problem with a single sensor, which is the foundation of robust distributed filtering. The following two sections introduce these types of estimation, along with a brief literature review for each.

1.1 Robust Filtering

State estimation, also known as filtering, is paramount to many control systems (ANDERSON; MOORE, 1979). The problem consists of estimating the state of a dynamical system based on noisy measurement data. A broad range of applications employs filtering techniques, such as robotics, computer vision, communications, power systems, and economics, to name a few. Given its importance, state estimation has been extensively studied over the past decades, especially after the early 1960s, when the celebrated Kalman filter (KALMAN, 1960) was first introduced.

Due to its simplicity and practicality, the Kalman filter has been one of the most popular and widely used approaches since its inception. It operates by minimizing the estimation error variance. Nevertheless, one of its well-known shortcomings is assuming exact knowledge of both the target system and sensing models, which seldom holds in practice. Parametric uncertainties often arise from linearization, unmodeled dynamics, model reduction, or varying parameters, and they can appreciably degrade the estimation performance. Despite the numerous efforts towards alleviating such effects, this is still an active area of research.

There are several ways to model such parametric uncertainties. This work specifically addresses two of the possible uncertainty models: the norm-bounded and the polytopic uncertainty types, which we discuss in the following subsections.

1.1.1 Robust Filtering for Systems with Norm-Bounded Uncertainties

The most representative robust state estimation approaches for systems subject to norm-bounded parametric uncertainties found in the literature are $\mathcal{H}_\infty$ filtering, guaranteed cost designs, risk-sensitive filtering, and robust regularized least-squares strategies.

The objective in $\mathcal{H}_\infty$ filtering is to minimize the $\mathcal{H}_\infty$ norm of the mapping from the disturbances to the estimation error. Some important results are reported in Xie, Souza and Fu (1991), Geromel *et al.* (2000), Xie *et al.* (2004), and references therein. One drawback of this approach, especially in online systems, is that it is hard to guarantee that the $\mathcal{H}_\infty$ performance parameter γ will satisfy the filter existence conditions at every step. In robust guaranteed-cost approaches, the goal is to design an estimator such that the estimation error variance has a guaranteed upper bound for all admissible uncertainties (see, for instance, Xie, Soh and Souza (1994), Theodor and Shaked (1996), Petersen and McFarlane (1996), Zhu, Soh and Xie (2002), and Dong and You (2006)). The procedure usually involves Riccati equations, which depend on selecting one or more scaling parameters to guarantee a solution's existence at each time step. This selection, however, is often not straightforward and has a significant impact on the estimation performance.

On the other hand, risk-sensitive filtering has been recently applied to overcome some of the shortcomings of the previous strategies. The goal is to minimize the expected value of an exponential of the quadratic error function, ensuring a certain degree of robustness against model uncertainties (SPEYER; DEYST; JACOBSON, 1974). For instance, Levy and Zorzi (2016) proposed a block-update Kalman filter with clear and easily computed conditions for the convergence of the associated risk-sensitive Riccati equation. Additionally, Zorzi (2017) presented a robust Kalman filter whose gain is updated based on a time-varying risk-sensitive parameter, which characterizes a tolerance upon the divergence between the actual (uncertain) and nominal system models, guaranteeing a well-defined filter at each iteration.

The fourth approach consists of formulating the estimation problem as a robust regularized least-squares problem (SAYED; NASCIMENTO, 1999). The aim is to minimize the worst-possible regularized residual norm over the set of admissible uncertainties. In general, the resulting filters are recursive and resemble the classic Kalman filter, which is convenient for online applications. Sayed (2001) was the first work to employ this approach in robust filtering. It introduces the so-called bounded data uncertainties (BDU) filter, which considers that the system is subject to norm-bounded parametric uncertainties. However, Xu and Mannor (2009) point out that considering the worst-case effect of uncertainties in least-squares designs may lead to over-conservative filters. They propose an estimator that combines the Kalman and BDU filters to counteract this issue. Ishihara, Terra and Cerri (2015) presented a robust filter in a symmetric matrix arrangement. Their design does not depend on any parameter tuning and assumes that all parameter matrices of the target system and sensing models have uncertainties. However, it involves the inversion of a large matrix block at each time step. More recently, Abolhasani and Rahmani (2018) extended the BDU filter to deal with both norm-bounded and stochastic uncertainties in all parameter matrices. The filter results from the solution of an optimization problem subject to a linear matrix inequality (LMI), which minimizes the estimation error variance at each iteration. Nonetheless, LMI-based strategies often require excessive computational effort, which might be prohibitive for real-time systems.

Motivated by this discussion, we propose a robust Kalman filter for uncertain linear discrete-time systems in this thesis. Unlike most works, we assume that all matrices of the target system and sensing models are subject to norm-bounded parametric uncertainties. This way, it is also possible to handle systems with uncertain noise variance matrices, as shown in Dong and You (2006).

We adopt a purely deterministic interpretation of the robust estimation problem, as discussed in Bryson and Ho (1975) and Sayed (2001), and formulate a constrained regularized least-squares estimation problem with norm-bounded uncertainties. We further apply the penalty function method (LUENBERGER; YE, 2021) to transform it into a more convenient unconstrained problem. The solution to this problem ultimately yields the proposed robust filter, which we present in a recursive correction-prediction Kalman-like structure. As such, we avoid using numerical solvers to derive the analytical filter equations.

The robust Kalman filter presented in this thesis was introduced in paper number 2, shown in Section 1.3.

1.1.2 Robust Filtering for Systems with Polytopic Uncertainties

In the polytopic uncertainty model, we consider that the system parameters arbitrarily vary within a convex polyhedron centered at the nominal parameters (CHANG; PARK; TANG, 2015).

The last two decades have witnessed the rise of several robust filtering techniques for systems subject to polytopic uncertainties. The pioneering work by Geromel *et al.* (2000) presents both $\mathcal{H}_2$ and $\mathcal{H}_\infty$ filters that are robust to polytopic uncertainties in discrete-time systems. The proposed estimators are based on the quadratic stability concept, which uses a single Lyapunov matrix to evaluate the estimation error norm over the entire uncertainty domain. However, the quadratic stability assumption is rather conservative for time-invariant systems. Geromel, Oliveira and Bernussou (2002) addressed this drawback, proposing a new stability condition based on parameter-dependent Lyapunov matrices. The authors also provide a method to synthesize a robust $\mathcal{H}_2$ filter, obtained as the result of a linear problem constrained by LMIs.

There has been a continuous effort to improve performance and reduce conservativeness of robust $\mathcal{H}_2$ and $\mathcal{H}_\infty$ filters. A common strategy is to introduce additional slack variables into the underlying LMIs, as reported in Shaked, Xie and Soh (2001), Xie *et al.* (2004), Duan *et al.* (2006), Zhang, Xia and Shi (2009), and Chang, Park and Tang (2015). While providing extra dimensions to the optimization problem solution space, this approach also increases the computational burden required to solve the more complex LMIs. Moreover, optimal filters usually demand fine-tuning of project parameters. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ strategies are proposed, e.g., in Palhares and Peres (2001) and Gao *et al.* (2005). The goal is to minimize an upper bound of the $\mathcal{H}_2$ norm of the estimation error whilst guaranteeing a prescribed $\mathcal{H}_\infty$ attenuation level. Additionally, there is the set-membership approach (YANG; LI, 2011), which involves a recursive algorithm for calculating an ellipsoid that always contains the true system state. More recently, the Finsler lemma has been used to improve the performance of estimators for systems with state-multiplicative noise and polytopic uncertainties, for instance, in Gershon and Shaked (2015), Morais *et al.* (2017), and Gershon and Shaked (2020).

A common aspect among the strategies discussed above is their dependence on solving optimization problems subject to LMIs. Usually, each vertex of the uncertainty polytope provides one inequality constraint. Hence, the problem complexity increases with the number of vertices. Moreover, introducing extra free parameters, which often rely on additional optimization or manual tuning, increases the overall complexity. Nevertheless, LMI-based robust filters have consistently been considered an effective and valuable state estimation strategy. However, the numerical solvers involved in their solution often require computational resources that may not be available in some applications.

With these points in mind, we propose a robust filter for linear discrete-time systems subject to polytopic uncertainties. We also formulate the robust estimation problem from a deterministic viewpoint (BRYSON; HO, 1975), as a min-max optimization problem subject to linear equality constraints obtained from each polytope vertex. The proposed polytopic robust filter has a recursive correction-prediction structure that resembles the classic Kalman filter. Its main advantage compared to the aforementioned results is that it does not depend on the solution of LMI-based optimization problems, avoiding the use of computationally expensive numerical solvers.

The polytopic robust Kalman filter is also presented in paper 3, shown in Section 1.3.

1.2 Robust Distributed Filtering

One of the major applications of NCMASs is in distributed filtering over sensor networks and has thus been an important research field over the past few decades. These networks are composed of interconnected nodes with sensing, computing, and communication capabilities. In a distributed filtering setup, each sensor observes a target dynamical system and shares information with a limited set of neighboring sensors to collectively obtain the best estimate of the system state. The cooperation between sensors allows for improved estimation accuracy, flexibility, and reliability of the overall system. Furthermore, unlike centralized architectures, a single sensor's failure does not compromise the entire system.

The average consensus protocol is among the most successful strategies used in distributed filtering. Olfati-Saber (2005) is a pioneering work in this sense, combining consensus and the Kalman filter (KALMAN, 1960) to propose a distributed Kalman filter (DKF). The idea was to reproduce the result of a global centralized Kalman filter. However, rather than having access to all sensor nodes at once, it uses a distributed architecture where the sensors can only access their neighbors. Each sensor carries out so-called micro-Kalman filter iterations, similar to a Kalman filter in its information form. However, it uses consensus filters (OLFATI-SABER; SHAMMA, 2005) to fuse the local measurements and innovation matrices (inverse of the measurement variance matrix) among the sensor neighborhood. The literature now refers to this technique as consensus on measurements (CM). A limitation of the DKF is that it assumes all sensors have identical models, meaning that the target system has to be observable by every sensor.

Olfati-Saber (2007) further extended the DKF to accommodate heterogeneous sensing models. The new filter is called the Kalman consensus filter (KCF). It also introduces the consensus on estimates (CE) approach, in which the sensor nodes share their prior state estimates, significantly increasing estimation accuracy. The KCF is presented in both continuous-time and discrete-time versions. Later, Olfati-Saber (2009) showed that the discrete-time KCF is a suboptimal solution to the distributed Kalman filtering problem. It further proposed an optimal solution, which was shown to not be scalable for large networks, as it requires the computation of cross-covariance matrices between every pair of sensor nodes, a prohibitive process in terms of communication and computational efforts. Nevertheless, the author proposed a feasible alternative to the optimal filter by approximating the optimal consensus gain, assumed to be a constant scale of the prior error covariance matrix. Using a Lyapunov-based stability analysis, the author showed that this suboptimal KCF is stable.

Deshmukh, Kwon and Hwang (2017) proposed a different optimal KCF where both the Kalman and consensus gains are derived as the solution to an optimization problem in which the total network estimation error is minimized. However, it suffers from the same scalability problem of the simpler optimal KCF of Olfati-Saber (2009), requiring even higher communication and computational burdens in exchange for minor performance improvement.

Another consensus-based approach for distributed filtering was proposed in Battistelli and Chisci (2014), later referred to as the consensus on information (CI) approach. The strategy consists of performing consensus on the information pairs, composed of the posterior state estimates and information matrices (inverse of the error covariance matrix) of the sensor nodes. A similar approach was reported in Kamal, Farrell and Roy-Chowdhury (2013), which introduced the information-weighted consensus filter and presented its application in a distributed camera network. The set of cameras performs a target-tracking task, such that the filter should compensate for the fact that the target may not be visible to all cameras.

Finally, in Battistelli *et al.* (2015), the authors propose the hybrid consensus on measurements and information (HCMCI) approach, which combines the CM and CI strategies mentioned above, leveraging their complementary features. Based on this approach, they design a hybrid consensus filter applicable to linear and nonlinear systems. For a comprehensive review of distributed filtering over sensor networks using consensus-based strategies and alternatives such as diffusion-based and gossip-based approaches, check the compilations in He *et al.* (2020) and Modalavalasa *et al.* (2021).

Note that many works on distributed filtering in the literature are based on the Kalman filter. As such, they inherit the pitfall of requiring exact knowledge of the target system and sensing models. In practice, these models are often subject to parametric uncertainties, which can jeopardize the estimation performance. Therefore, dealing with these uncertainties in distributed filtering has stirred attention from researchers and is the primary motivation of this thesis. Analogous to the single-sensor robust estimation mentioned earlier, we also address two types of parametric uncertainties in robust distributed filtering: norm-bounded and polytopic uncertainties, which we discuss in the following subsections.

1.2.1 Robust Distributed Filtering for Systems with Norm-Bounded Uncertainties

Different from the large body of research dedicated to the single-sensor robust estimation problem for systems subject to norm-bounded uncertainties, as summarized in Section 1.1.1, this problem has not been as well-explored in the context of distributed estimation over sensor networks.

Most works concerning robust distributed filtering adopt a stochastic treatment of uncertainties. For instance, Ding *et al.* (2012) and Wang *et al.* (2018) proposed $\mathcal{H}_\infty$

distributed filters for systems with stochastic uncertainties. The former also deals with stochastic nonlinearities, and the latter further considers successive missing measurements. Feng, Wang and Zeng (2013), Tian, Sun and Li (2016), and Ding *et al.* (2017) presented recursive Kalman-like distributed filters. The first two additionally address auto- and cross-correlated noises, whereas the third assumes uniform quantization effects and compensates for deception attacks. Rastgar and Rahmani (2018) extended the work of Deshmukh, Kwon and Hwang (2017) to systems with stochastic uncertainties, proposing an optimal consensus-based distributed filter. Nonetheless, it relies on computing cross-covariance matrices between sensors to achieve optimality, which renders a computationally and communication-intensive filter. Later, in Rastgar and Rahmani (2020), the authors circumvented this issue and proposed a distributed estimator based on the HCMCI approach.

On the other hand, Shen, Wang and Hung (2010), Dong, Ding and Ren (2014), Hedayati and Rahmani (2020), and Han *et al.* (2021) proposed $\mathcal{H}_\infty$-consensus filters to handle norm-bounded uncertainties. The first three also deal with the missing measurements problem. The work by Hedayati and Rahmani (2020) further considers state time delays. However, they rely on the solution of complex LMIs and compute the filter gains all at once, requiring knowledge of the whole network. Zhang, He and Zhou (2018) proposed a robust recursive distributed filter for sensor networks with parameter and network topology uncertainties without assuming any particular structure. At each time step, it minimizes the trace of the estimation error covariance. As in the previous estimators, it also computes the filter gains altogether. Hence, the works above are not fully distributed strategies since they require network-wide information. In contrast, Duan *et al.* (2020) and Rocha and Terra (2020) proposed fully distributed robust recursive filters. While the former only considers norm-bounded uncertainties in the target system model, the latter also treats uncertainties in the sensing models. Moreover, both employ a single consensus iteration.

Considering the low number of works on the subject, we propose a robust fully distributed consensus-based filter for sensor networks estimating systems subject to norm-bounded parametric uncertainties. To derive this filter, we first propose a robust centralized Kalman filter, which generalizes the robust Kalman filter mentioned in Section 1.1.1 to the multiple sensor case. In this setup, a fusion center gathers the measurements from all the sensors. Then, through the HCMCI protocol, we arrive at the distributed formulation, which approximates the centralized estimator's behavior but considers only the local interactions between each sensor and its neighborhood.

A preliminary version of this distributed filter which employs a single iteration of the CI protocol was first presented in paper number 1, listed in Section 1.3, whereas paper number 4 features the final version of the robust distributed filter.

1.2.2 Robust Distributed Filtering for Systems with Polytopic Uncertainties

The literature on robust distributed estimation for systems with polytopic uncertainties is even more scarce than the norm-bounded uncertainty type. Shen, Wang and Hung (2010) and Souza, Coutinho and Kinnaert (2016) are among the few works that tackle this problem. The former proposed an $\mathcal{H}_\infty$-consensus filter and also considered the effects of missing measurements. The latter presented distributed filters with a more general structure, whose matrices are designed to minimize the mean squared estimation error. Two strategies are proposed: one based on the observability Gramian and the other based on the controllability Gramian. A common drawback of the aforementioned solutions is that they require knowledge of the whole network structure when computing the filter gains, which reduces their flexibility. Moreover, they depend on the solution of LMI-based optimization problems, whose complexity increases with the number of sensors in the network and the number of vertices of the uncertainty polytopes, which may be prohibitive for real-time systems.

To help fill the gap regarding this specific type of problem, in this thesis, we propose a robust and fully distributed consensus-based filter for estimating the state of systems subject to polytopic uncertainties using sensor networks. Similar to the norm-bounded uncertainty case discussed in Section 1.2.1, we first propose a centralized filter, extending the polytopic robust Kalman filter mentioned in Section 1.1.2 to the multiple-sensor scenario. Then, we also employ the HCMCI protocol to derive a distributed implementation of the centralized estimator, aiming to approximate its performance.

Paper number 5, shown in Section 1.3, also presents the polytopic robust distributed filter proposed here.

1.3 List of Publications

The following list contains all the published journal and conference papers concerning direct results of the research reported in this thesis:

1. ROCHA, K. D. T.; TERRA, M. H. Robust distributed consensus-based filtering for uncertain systems over sensor networks. **IFAC-PapersOnline**, v. 53, n. 2, p. 3571–3576, 21st IFAC World Congress, Berlin (Germany), 2020.

2. ROCHA, K. D. T.; TERRA, M. H. Robust Kalman filter for systems subject to parametric uncertainties. **Systems & Control Letters**, v. 57, 2021.

3. ROCHA, K. D. T.; BUENO, J. N. A. D.; MARCOS, L. B.; TERRA, M. H. Robust Kalman filtering for systems subject to polytopic uncertainties. *In:* **Proceedings of**

the 10th Mediterranean Conference on Control and Automation (MED), 2022, Vouliagmeni (Greece). 2022. p. 271–276.

4. ROCHA, K. D. T.; TERRA, M. H. Robust distributed Kalman consensus filter for sensor networks under parametric uncertainties. *In:* **Proceedings of the 20th European Control Conference (ECC), 2022, London (England)**. 2022. p. 2209–2215.

5. ROCHA, K. D. T.; BUENO, J. N. A. D.; MARCOS, L. B.; TERRA, M. H. Polytopic robust distributed Kalman consensus filter for sensor networks. *In:* **Proceedings of the 8th IFAC Symposium on Systems Structure and Control (SSSC), 2022, Montreal (Canada)**. 2022.

The author also collaborated in other related research topics which resulted in some publications, listed below:

6. ODORICO, E. K.; ROCHA, K. D. T.; TERRA, M. H. Recursive estimation for discrete-time Markovian jump singular systems with random state delays. **IFAC-PapersOnline**, v. 52, n. 18, p. 168–173, 15th IFAC Workshop on Time Delay Systems (TDS), Sinaia (Romania), 2019.

7. ODORICO, E. K.; ROCHA, K. D. T.; TERRA, M. H. Estimador robusto recursivo para sistemas singulares incertos com atraso invariante no estado via método de elevação. *In:* **Anais do 14º Simpósio Brasileiro de Automação Inteligente (SBAI), 2019, Ouro Preto (Brazil)**. 2019. p. 2384–2389.

8. BUENO, J. N. A. D.; ROCHA, K. D. T.; TERRA, M. H. Gain-scheduled robust recursive lateral control for autonomous ground vehicles subject to polytopic uncertainties. *In:* **Proceedings of the 2020 Latin American Robotics Symposium (LARS), 2020, Natal (Brazil)**. 2020.

9. CHÁVEZ-FUENTES, J. R.; COSTA, E. F.; TERRA, M. H.; ROCHA, K. D. T. The linear quadratic optimal control problem for discrete-time Markov jump linear singular systems. **Automatica**, v. 127, 2021.

10. BUENO, J. N. A. D.; ROCHA, K. D. T.; TERRA, M. H. Robust recursive regulator for systems subject to polytopic uncertainties. **IEEE Access**, v. 9, p. 139352–139360, 2021.

11. BUENO, J. N. A. D.; MARCOS, L. B.; ROCHA, K. D. T.; TERRA, M. H. Longitudinal control of an autonomous truck with unobserved gears. *In:* **Proceedings of the 2021 IEEE URUCON, 2021, Montevideo (Uruguay)**. 2021. p. 339–342.

12. BUENO, J. N. A. D.; MARCOS, L. B.; ROCHA, K. D. T.; TERRA, M. H. Regulation of Markov jump linear systems subject to polytopic uncertainties. **IEEE Transactions on Automatic Control**, v. 67, n. 11, p. 6279–6286, 2022.

13. BUENO, J. N. A. D.; ROCHA, K. D. T.; MARCOS, L. B.; TERRA, M. H. Mode-independent regulator for polytopic Markov jump linear systems. *In:* **Proceedings of the 10th Mediterranean Conference on Control and Automation (MED), 2022, Vouliagmeni (Greece)**. 2022. p. 277–282.

14. BUENO, J. N. A. D.; MARCOS, L. B.; ROCHA, K. D. T.; TERRA, M. H. Robust regulation of Markov jump linear systems with uncertain polytopic transition probabilities. *In:* **Proceedings of the 20th European Control Conference (ECC), 2022, London (England)**. 2022. p. 1373–1378.

15. MARCOS, L. B.; BUENO, J. N. A. D.; ROCHA, K. D. T.; TERRA, M. H. Longitudinal control of self-driving heavy-duty vehicles: a robust Markovian approach. *In:* **Proceedings of the 61st IEEE Conference on Decision and Control (CDC), 2022, Cancún (Mexico)**. 2022. p. 6966–6972.

1.4 Document Structure

This thesis is organized into five chapters and an appendix, described as follows:

- **Chapter 1:** Introduces the networked cooperative multi-agent systems, including the sensor networks, the central subject of this work. Moreover, it explains the different robust and distributed filtering problems to be addressed, featuring brief reviews of the related works in the literature.

- **Chapter 2:** Presents preliminary concepts that are fundamental to the development of this work, namely, the penalty function method, least-squares problems, notions of graph theory, and the average consensus algorithm.

- **Chapter 3:** Addresses the state estimation problem in the single-sensor scenario, explaining the deterministic approach to derive nominal and robust filters as the outcome of solving least-squares problems. Furthermore, it presents the two proposed robust Kalman filters for systems with norm-bounded and polytopic uncertainties.

- **Chapter 4:** Extends the results in Chapter 3 for applications using sensor networks. It presents the robust centralized and consensus-based distributed versions of the respective filters proposed in the previous chapter, considering the two distinct types of parametric uncertainties: norm-bounded and polytopic.

- **Chapter 5:** Presents the concluding remarks of this work and suggests possible directions to extend the proposed results further.
- **Appendix A:** Provides a collection of matrix analysis results used throughout this document.

Chapter 2

Preliminary Concepts

This chapter introduces preliminary concepts fundamental to this work's development. The proposed robust and distributed filters result from the solution of optimization problems subject to linear equality constraints. When dealing with these problems, using the penalty function method is a standard approach. It approximates the constrained problem with a sequence of unconstrained problems, which are more convenient to solve. The approximation includes the constraint equations into the objective function multiplied by a parameter that penalizes violations of the constraints.

In addition, the classic least-squares problem, as well as its weighted and regularized variants, are reviewed. We also present the regularized least-squares problem with uncertainties and an adapted version that sits at the foundation of the estimation algorithms proposed in this thesis.

A graph is a standard instrument to model the intercommunication among sensors in a network. Therefore, this chapter also presents some basic graph theory notions and results used in this work. We conclude the chapter with an introduction to the average consensus algorithm, a paramount technique in distributed estimation.

2.1 Penalty Function Method

This section presents the penalty function method, a technique used to solve optimization problems subject to linear equality constraints. The following results are extracted from Luenberger and Ye (2021).

Consider the constrained optimization problem

$$\begin{aligned} &\min_{z} \ f(z) \\ &\text{s.t.} \ \ h(z) = 0, \end{aligned} \tag{2.1}$$

where $z \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous objective function, and $h : \mathbb{R}^n \to \mathbb{R}$ is a linear equality constraint. The penalty function method consists of replacing the constrained problem (2.1) by an unconstrained problem of the form

$$\min_{z} q(z,\mu) = f(z) + \mu P(z), \tag{2.2}$$

in which $q(z,\mu)$ is a new objective function with a constant *penalty parameter* $\mu > 0$ and *penalty function* $P(z) = h(z)^T h(z)$. Observe that the penalty parameter is associated with the constraint term $h(z)$ in a way that the penalty function $P(z)$ is penalized when the constraint is violated, i.e., if $h(z) \neq 0$. This way, μ determines how close the solution to (2.2) is to the solution of the original problem (2.1).

In general, we apply the penalty function method as an iterative process. At each step $k \geq 0$, the penalty parameter μ_k is fixed and the optimal solution z_k of problem (2.2) is then obtained. At each iteration, μ_k is updated, such that an ascending sequence $\{\mu_k\}_{k=0}^{+\infty}$ is generated. As $\mu_k \to +\infty$, the penalty in function $P(z)$ increases and, in the limit, $P(z) \to 0$. Consequently, $h(z) \to 0$ and the unconstrained problem becomes equivalent to the constrained problem. Algorithm 2.1 summarizes the procedure for the penalty function method.

Algorithm 2.1 Penalty function method

Initialization: Set a desired precision $\epsilon > 0$ and $\mu_0 > 0$
  for $k = 0, 1, \ldots$ **do**
    Define the auxiliary objective function $q(z,\mu_k) = f(z) + \mu_k P(z)$
    Obtain $z_k = \arg\min_{z} q(z,\mu_k)$
    if $\mu_k P(z_k) < \epsilon$ **then**
      Stop and return the solution z_k
    else
      $\mu_k \leftarrow \mu_{k+1} > \mu_k$
      $k \leftarrow k + 1$
    end if
  end for

The following results concern the convergence of the penalty function method. The detailed proofs can be found in Luenberger and Ye (2021).

Lemma 2.1. *(LUENBERGER; YE, 2021) Consider an ascending sequence of penalty parameters $\{\mu_k\}_{k=0}^{+\infty}$ and the corresponding sequence of functions $\{q(z,\mu_k)\}_{k=0}^{+\infty}$, with $q(z,\mu_k) = f(z) + \mu_k P(z)$. The following properties hold:*

(i) $q(z_k,\mu_k) \leq q(z_{k+1},\mu_{k+1})$;

(ii) $P(z_k) \geq P(z_{k+1})$;

(iii) $f(z_k) \leq f(z_{k+1})$.

Lemma 2.2. *(LUENBERGER; YE, 2021) Let z^* be a solution to problem (2.1). Then, for each k,*

$$f(z^*) \geq q(z_k,\mu_k) \geq f(z_k).$$

Definition 2.1. *A point z is a* limit point *of sequence $\{z_k\}$ if there exists a subsequence of $\{z_k\}$ that converges to z. Equivalently, z is a limit point of $\{z_k\}$ if there exists a subset $\mathbb{K} \subset \mathbb{Z}_+$ such that $\{z_k\}_{k \in \mathbb{K}}$ converges to z.*

The next theorem establishes the global convergence of the penalty function method. More precisely, it states that the limit point of any sequence of solutions to unconstrained problem (2.2) corresponds to the solution to the original constrained problem (2.1). The result follows from the two previous lemmas.

Theorem 2.1. *(LUENBERGER; YE, 2021) Let $\{z_k\}$ be a sequence generated by the penalty function method. Then, any limit point of this sequence is a solution to problem (2.1).*

2.2 Least-Squares Problems

This section reviews the classic least-squares problem and some variations: the weighted least-squares, the regularized least-squares, and the regularized least-squares with uncertainties. Further details about the classic problems and their solutions can be found, for instance, in Kailath, Sayed and Hassibi (2000a).

2.2.1 Least-Squares Problem

Consider the quadratic optimization problem

$$\min_{z} J(z), \tag{2.3}$$

with objective function $J : \mathbb{R}^n \to \mathbb{R}$ given by

$$J(z) = \|\mathcal{A}z - b\|^2 = (\mathcal{A}z - b)^T(\mathcal{A}z - b), \tag{2.4}$$

in which $z \in \mathbb{R}^n$ is an unknown vector and $\mathcal{A} \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are known.

Lemma 2.3. *A vector z^* is a solution to problem (2.3)-(2.4) if, and only if, it satisfies the* normal equation

$$\mathcal{A}^T\mathcal{A}z^* = \mathcal{A}^T b. \tag{2.5}$$

The resulting minimum value of the objective function is then given by

$$J(z^*) = \|\mathcal{A}z^* - b\|^2 = \|b\|^2 - \|\mathcal{A}z^*\|^2.$$

If $\mathcal{A}$ has full column rank n, then $\mathcal{A}^T\mathcal{A}$ is nonsingular and there is a unique z^* satisfying (2.5) given by

$$z^* = \left(\mathcal{A}^T\mathcal{A}\right)^{-1}\mathcal{A}^T b.$$

Furthermore, the resulting minimum value of the objective function is

$$J(z^*) = \|\mathcal{A}z^* - b\|^2 = b^T\left(I_m - \mathcal{A}\left(\mathcal{A}^T\mathcal{A}\right)^{-1}\mathcal{A}^T\right)b.$$

2.2.2 Weighted Least-Squares Problem

Consider the optimization problem

$$\min_z J(z), \tag{2.6}$$

with objective function $J : \mathbb{R}^n \to \mathbb{R}$ given by

$$J(z) = \|\mathcal{A}z - b\|^2_{\mathcal{W}} = (\mathcal{A}z - b)^T \mathcal{W}(\mathcal{A}z - b), \tag{2.7}$$

where $z \in \mathbb{R}^n$ is an unknown vector, $\mathcal{W} \in \mathbb{R}^{m\times m}$ is a known symmetric positive definite *weighting matrix*, and $\mathcal{A} \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are also known. Notice that this is a generalization of the previous least-squares problem, for which $\mathcal{W} = I_m$.

Lemma 2.4. *A vector z^* is a solution to problem (2.6)-(2.7) if, and only if, it satisfies the normal equation*

$$\mathcal{A}^T \mathcal{W}\mathcal{A}z^* = \mathcal{A}^T \mathcal{W}b. \tag{2.8}$$

The minimum value of the objective function is thus given by

$$J(z^*) = \|\mathcal{A}z^* - b\|^2_{\mathcal{W}} = b^T \mathcal{W}b - b^T \mathcal{W}\mathcal{A}z^*.$$

If $\mathcal{A}$ has full column rank n, then $\mathcal{A}^T\mathcal{W}\mathcal{A}$ is nonsingular and there is a unique z^* satisfying (2.8) given by

$$z^* = \left(\mathcal{A}^T\mathcal{W}\mathcal{A}\right)^{-1}\mathcal{A}^T\mathcal{W}b.$$

Furthermore, the minimum value of the objective function is

$$J(z^*) = \|\mathcal{A}z^* - b\|^2_{\mathcal{W}} = b^T\left(\mathcal{W} - \mathcal{W}\mathcal{A}\left(\mathcal{A}^T\mathcal{W}\mathcal{A}\right)^{-1}\mathcal{A}^T\mathcal{W}\right)b.$$

2.2.3 Regularized Least-Squares Problem

Consider the optimization problem defined by

$$\min_z J(z), \tag{2.9}$$

with objective function $J : \mathbb{R}^n \to \mathbb{R}$ given by

$$J(z) = \|z\|^2_{\mathcal{Q}} + \|\mathcal{A}z - b\|^2_{\mathcal{W}} = z^T\mathcal{Q}z + (\mathcal{A}z - b)^T \mathcal{W}(\mathcal{A}z - b), \tag{2.10}$$

where $z \in \mathbb{R}^n$ is an unknown vector, $\mathcal{Q} \in \mathbb{R}^{n\times n}$ and $\mathcal{W} \in \mathbb{R}^{m\times m}$ are known symmetric weighting matrices, with $\mathcal{Q} \succ 0$ and $\mathcal{W} \succeq 0$, and $\mathcal{A} \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are also known. Unlike the weighted least-squares problem, this function has an additional *regularization term*.

Lemma 2.5. *The unique optimal solution* z^* *to problem* (2.9)-(2.10) *is given by*

$$z^* = \left(\mathcal{Q} + \mathcal{A}^T \mathcal{W} \mathcal{A}\right)^{-1} \mathcal{A}^T \mathcal{W} b \tag{2.11}$$

and the minimum value of the objective function is

$$J(z^*) = b^T\left(\mathcal{W} - \mathcal{W}\mathcal{A}\left(\mathcal{Q} + \mathcal{A}^T\mathcal{W}\mathcal{A}\right)^{-1}\mathcal{A}^T\mathcal{W}\right)b = b^T\left(\mathcal{W}^{-1} + \mathcal{A}\mathcal{Q}^{-1}\mathcal{A}^T\right)^{-1}b.$$

Note that, in Lemma 2.5, the uniqueness of the solution is guaranteed by requiring that $\mathcal{Q} \succ 0$. However, as we shall see in Section 3.1.2, we are rather interested in the case where $\mathcal{Q} \succeq 0$ and $\mathcal{W} \succ 0$, such that an additional condition is needed to ensure the uniqueness of the solution. Therefore, in the following lemma, we provide such conditions when dealing with this alternative problem.

Lemma 2.6. *Consider problem* (2.9)-(2.10) *with* $\mathcal{Q} = \begin{bmatrix} \mathcal{Q}_1 & 0 \\ 0 & \mathcal{Q}_2 \end{bmatrix} \succeq 0$*, in which* $\mathcal{Q}_1 \succeq 0$ *and* $\mathcal{Q}_2 \succ 0$*, and* $\mathcal{W} \succ 0$*. Define* $\mathcal{J}_2 := \begin{bmatrix} 0 & I_2 \end{bmatrix}$*, where the identity matrix* I_2 *has the same dimensions as* $\mathcal{Q}_2$*. If the block* $\begin{bmatrix} \mathcal{J}_2 \\ \mathcal{A} \end{bmatrix}$ *has full column rank* n*, the optimal solution* z^* *given in* (2.11) *is unique.*

Proof. The uniqueness of the optimal solution z^* in (2.11) is a consequence of the invertibility of the term $(\mathcal{Q} + \mathcal{A}^T\mathcal{W}\mathcal{A})$. Let us rewrite this term as

$$\mathcal{J}_1^T\mathcal{Q}_1\mathcal{J}_1 + \mathcal{J}_2^T\mathcal{Q}_2\mathcal{J}_2 + \mathcal{A}^T\mathcal{W}\mathcal{A} = \mathcal{J}_1^T\mathcal{Q}_1\mathcal{J}_1 + \begin{bmatrix} \mathcal{J}_2^T & \mathcal{A}^T \end{bmatrix} \begin{bmatrix} \mathcal{Q}_2 & 0 \\ 0 & \mathcal{W} \end{bmatrix} \begin{bmatrix} \mathcal{J}_2 \\ \mathcal{A} \end{bmatrix},$$

in which we define $\mathcal{J}_1 := \begin{bmatrix} I_1 & 0 \end{bmatrix}$ and $\mathcal{J}_2 := \begin{bmatrix} 0 & I_2 \end{bmatrix}$, where I_1 and I_2 are identity matrices with the same dimensions as $\mathcal{Q}_1$ and $\mathcal{Q}_2$, respectively. From Lemma A.5 (item **(i)**), since $\mathcal{Q}_1 \succeq 0$, the first term is positive semidefinite. The central block diagonal matrix in the last term is positive definite, thus, according to Lemma A.5 (item **(ii)**), if the block $\begin{bmatrix} \mathcal{J}_2 \\ \mathcal{A} \end{bmatrix}$ has full column rank, the last term is also positive definite, and so is the full expression. In consequence, we ensure its invertibility and the uniqueness of the optimal solution. □

2.2.4 Regularized Least-Squares Problem with Uncertainties

Consider now a regularized least-squares problem with parametric uncertainties defined by

$$\min_{z} \max_{\delta\mathcal{A}, \delta b} J(z, \delta\mathcal{A}, \delta b), \tag{2.12}$$

with objective function given by

$$\begin{aligned} J(z, \delta\mathcal{A}, \delta b) &= \|z\|_{\mathcal{Q}}^2 + \|(\mathcal{A} + \delta\mathcal{A})z - (b + \delta b)\|_{\mathcal{W}}^2 \\ &= z^T\mathcal{Q}z + \left[(\mathcal{A} + \delta\mathcal{A})z - (b + \delta b)\right]^T \mathcal{W} \left[(\mathcal{A} + \delta\mathcal{A})z - (b + \delta b)\right], \end{aligned} \tag{2.13}$$

in which $z \in \mathbb{R}^n$ is an unknown vector, $\mathcal{Q} \in \mathbb{R}^{n\times n}$ and $\mathcal{W} \in \mathbb{R}^{m\times m}$ are known symmetric weighting matrices, with $\mathcal{Q} \succ 0$ and $\mathcal{W} \succeq 0$. $\mathcal{A} \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are known and subject to parametric uncertainties $\delta\mathcal{A} \in \mathbb{R}^{m\times n}$ and $\delta b \in \mathbb{R}^m$, modeled as

$$\begin{bmatrix}\delta\mathcal{A} & \delta b\end{bmatrix} = M\Delta\begin{bmatrix}E_{\mathcal{A}} & E_b\end{bmatrix}, \tag{2.14}$$

where $M \neq 0$, $E_{\mathcal{A}}$ and E_b are known matrices with appropriate dimensions, and Δ is an arbitrary contraction matrix, such that $\|\Delta\| \leq 1$. Notice that the objective function should be minimized under the maximum influence of the parametric uncertainties.

The unique optimal solution to problem (2.12)-(2.14) is shown next. It was proposed in Sayed and Nascimento (1999), where the detailed proof can be found.

Lemma 2.7. *(SAYED; NASCIMENTO, 1999) Problem (2.12)-(2.14) admits a unique solution z^* given by*

$$z^* = \left(\widehat{\mathcal{Q}} + \mathcal{A}^T\widehat{\mathcal{W}}\mathcal{A}\right)^{-1}\left(\mathcal{A}^T\widehat{\mathcal{W}}b + \hat{\lambda}E_{\mathcal{A}}^T E_b\right), \tag{2.15}$$

where $\widehat{\mathcal{Q}}$ and $\widehat{\mathcal{W}}$ are modified weighting matrices defined as

$$\widehat{\mathcal{Q}} := \mathcal{Q} + \hat{\lambda}E_{\mathcal{A}}^T E_{\mathcal{A}},$$
$$\widehat{\mathcal{W}} := \mathcal{W} + \mathcal{W}M\left(\hat{\lambda}I - M^T\mathcal{W}M\right)^{\dagger}M^T\mathcal{W},$$

and $\hat{\lambda}$ is a nonnegative scalar parameter obtained as the solution to the optimization problem

$$\hat{\lambda} := \arg\min_{\lambda \geq \|M^T\mathcal{W}M\|} \Gamma(\lambda), \tag{2.16}$$

with objective function $\Gamma(\lambda)$ given by

$$\Gamma(\lambda) := \|z(\lambda)\|_{\mathcal{Q}}^2 + \lambda\|E_{\mathcal{A}}z(\lambda) - E_b\|^2 + \|\mathcal{A}z(\lambda) - b\|_{\mathcal{W}(\lambda)}^2, \tag{2.17}$$

in which

$$\mathcal{Q}(\lambda) := \mathcal{Q} + \lambda E_{\mathcal{A}}^T E_{\mathcal{A}},$$
$$\mathcal{W}(\lambda) := \mathcal{W} + \mathcal{W}M\left(\lambda I - M^T\mathcal{W}M\right)^{\dagger}M^T\mathcal{W},$$
$$z(\lambda) := \left(\mathcal{Q}(\lambda) + \mathcal{A}^T\mathcal{W}(\lambda)\mathcal{A}\right)^{-1}\left(\mathcal{A}^T\mathcal{W}(\lambda)b + \lambda E_{\mathcal{A}}^T E_b\right).$$

Remark 2.1. *(SAYED, 2001) In Lemma 2.7, if $\lambda > \|M^T\mathcal{W}M\|$ and $\mathcal{W}$ is positive definite, the term $\left(\lambda I - M^T\mathcal{W}M\right)$ also becomes positive definite. In this case, the pseudoinverse operations in $\widehat{\mathcal{W}}$ and $\mathcal{W}(\lambda)$ can thus be replaced by normal matrix inverse operations.*

Remark 2.2. *(SAYED, 2001) Instead of explicitly solving the auxiliary problem (2.16)-(2.17) in Lemma 2.7, we can reasonably approximate the optimal $\hat{\lambda}$ parameter as $\hat{\lambda} = (1+\xi)\|M^T\mathcal{W}M\|$, for some $\xi > 0$. This comes from the observation that the function $\Gamma(\lambda)$ tends to reach amplitudes close to its minimum for values of λ that are generally close to its lower bound $\|M^T\mathcal{W}M\|$, as reported in Sayed and Chen (2002).*

Similar to the nominal regularized least-squares problem (Section 2.2.3), by requiring that $\mathcal{Q} \succ 0$, we guarantee the uniqueness of the solution presented in Lemma 2.7. Nevertheless, as we will see in Section 3.2.2 and Section 3.3.2, we are interested in the problem where $\mathcal{Q} \succeq 0$ and $\mathcal{W} \succ 0$. Thus, to ensure the uniqueness of the solution in (2.15), we need an additional condition on the problem parameters, as the following lemma explains.

Lemma 2.8. *Consider problem* (2.12)-(2.14) *with* $\mathcal{Q} = \begin{bmatrix} \mathcal{Q}_1 & 0 \\ 0 & \mathcal{Q}_2 \end{bmatrix} \succeq 0$, *in which* $\mathcal{Q}_1 \succeq 0$ *and* $\mathcal{Q}_2 \succ 0$, *and* $\mathcal{W} \succ 0$. *Define* $\mathcal{I}_2 := \begin{bmatrix} 0 & I_2 \end{bmatrix}$, *where the identity matrix* I_2 *has the same dimensions as* $\mathcal{Q}_2$. *If the block* $\begin{bmatrix} \mathcal{I}_2 \\ \mathcal{A} \\ E_{\mathcal{A}} \end{bmatrix}$ *has full column rank* n, *the optimal solution* z^* *given in* (2.15) *is unique.*

Proof. The uniqueness of the optimal solution z^* in (2.15) is a consequence of the invertibility of the term $\left(\widehat{\mathcal{Q}} + \mathcal{A}^T \widehat{\mathcal{W}} \mathcal{A}\right)$. Let us rewrite this term as

$$\mathcal{I}_1^T \mathcal{Q}_1 \mathcal{I}_1 + \mathcal{I}_2^T \mathcal{Q}_2 \mathcal{I}_2 + \mathcal{A}^T \widehat{\mathcal{W}} \mathcal{A} + \hat{\lambda} E_{\mathcal{A}}^T E_{\mathcal{A}} = \mathcal{I}_1^T \mathcal{Q}_1 \mathcal{I}_1 + \begin{bmatrix} \mathcal{I}_2^T & \mathcal{A}^T E_{\mathcal{A}}^T \end{bmatrix} \begin{bmatrix} \mathcal{Q}_2 & 0 & 0 \\ 0 & \widehat{\mathcal{W}} & 0 \\ 0 & 0 & \hat{\lambda} I \end{bmatrix} \begin{bmatrix} \mathcal{I}_2 \\ \mathcal{A} \\ E_{\mathcal{A}} \end{bmatrix},$$

in which we define $\mathcal{I}_1 := \begin{bmatrix} I_1 & 0 \end{bmatrix}$ and $\mathcal{I}_2 := \begin{bmatrix} 0 & I_2 \end{bmatrix}$, where I_1 and I_2 are identity matrices with the same dimensions as $\mathcal{Q}_1$ and $\mathcal{Q}_2$, respectively. From Lemma A.5 (item **(i)**), since $\mathcal{Q}_1 \succeq 0$, the first term is positive semidefinite. As mentioned in Remark 2.1, given that $\mathcal{W} \succ 0$ and $\hat{\lambda} > \|M^T \mathcal{W} M\| > 0$, we have that $\widehat{\mathcal{W}} \succ 0$. Hence, the central block diagonal matrix in the last term is positive definite. According to Lemma A.5 (item **(ii)**), if the block $\begin{bmatrix} \mathcal{I}_2 \\ \mathcal{A} \\ E_{\mathcal{A}} \end{bmatrix}$ has full column rank, the last term is also positive definite, and so is the full expression. In consequence, we ensure its invertibility and the uniqueness of the optimal solution. □

2.3 Notions of Graph Theory

Graphs can naturally represent the structure of information exchange among agents in a multi-agent system. For this reason, this section introduces basic notions of graph theory that are essential to model the sensor networks considered in this work. The concepts are borrowed from the tutorial presented in Ren, Beard and Atkins (2007). For a more in-depth study on graph theory, see Diestel (2017) and Godsil and Royle (2001).

2.3.1 Basic Definitions

A *graph* is a pair $\mathbb{G} = (\mathbb{S}, \mathbb{E})$, in which $\mathbb{S} = \{v_1, \ldots, v_S\}$ is the nonempty finite *vertex* or *node set* and $\mathbb{E} \subseteq \mathbb{S} \times \mathbb{S}$ is the *edge set*. The elements of $\mathbb{E}$ are denoted (v_i, v_j), meaning in this case that nodes v_i and v_j are *adjacent*, or *neighbors*.

In a *directed graph* (or *digraph*), the edge (v_i, v_j) indicates that node v_j, called the *child*, receives information from node v_i, called the *parent*, but not necessarily vice versa. Moreover, v_i is an *(in-)neighbor* of v_j. The *neighborhood* of a node v_i is the set $\mathcal{N}_i = \{v_j \,|\, (v_j, v_i) \in \mathbb{E}\}$. The *in-degree* of v_i, denoted $|\mathcal{N}_i|$, is the cardinality of its neighborhood, i.e., the number of elements in $\mathcal{N}_i$. In contrast, in *undirected graphs*, the edges are bidirectional, such that $(v_i, v_j) \in \mathbb{E} \Rightarrow (v_j, v_i) \in \mathbb{E}, \forall i, j$, and edges can be treated as unordered pairs.

We can assign weights to the edges of a graph, yielding the so-called *weighted graphs*. For example, we assign weight a_{ij} to the edge (v_j, v_i). Note the order of the indices here. For weighted undirected graphs, we have that $a_{ij} = a_{ji}, \forall i, j$.

A *directed path* is a sequence of nodes $v_0, v_1, \ldots, v_r$ such that $(v_k, v_{k+1}) \in \mathbb{E}$, for $k = 0, 1, \ldots, r-1$. Node v_i is *connected* to node v_j if there is a directed path between them. A directed graph is *strongly connected* if there is a directed path between every pair of distinct nodes. Analogously, undirected graphs are said to be *connected* in this case. The qualifier "strongly" is omitted since if there is a directed path from v_i to v_j, there is also one from v_j to v_i in undirected graphs.

2.3.2 Algebraic Graph Theory

The structure and properties of a graph can be studied by analyzing the properties of certain matrices that can be associated with it. In this section, consider a weighted graph $\mathbb{G} = (\mathbb{S}, \mathbb{E})$ with node set $\mathbb{S} = \{v_1, \ldots, v_S\}$ and edge weights a_{ij}.

The graph can be represented by an *adjacency matrix* $\mathcal{A} = [a_{ij}]$, in which $a_{ij} > 0$ if $(v_j, v_i) \in \mathbb{E}$ and $a_{ij} = 0$ otherwise. Recall that, for undirected graphs, $a_{ij} = a_{ji}, \forall i, j$, therefore, the adjacency matrix is symmetrical in this case. Moreover, its eigenvalues are all real. If weights are not relevant, simply set $a_{ij} = 1$ whenever $(v_j, v_i) \in \mathbb{E}$. A graph is said to be *balanced* if $\sum_{j=1}^{S} a_{ij} = \sum_{j=1}^{S} a_{ji}, \forall i$. This implies that all undirected graphs are balanced.

Another important matrix that can characterize a graph is the *Laplacian matrix* $\mathcal{L} = [\ell_{ij}]$, defined as

$$\ell_{ij} = \begin{cases} \displaystyle\sum_{j=1, i \neq j}^{S} a_{ij}, & \text{if } i = j, \\ -a_{ij}, & \text{if } i \neq j. \end{cases}$$

There is an equivalent definition of the Laplacian matrix, given by $\mathcal{L} = \mathcal{D} - \mathcal{A}$. Here, $\mathcal{D} = [d_{ij}]$ is the *weighted in-degree matrix*, where $d_{ij} = 0$ if $i \neq j$ and $d_{ii} = \sum_{j=1, i\neq j}^{S} a_{ij}$, $i = 1, \ldots, S$, and $\mathcal{A}$ is the adjacency matrix. Note that, for undirected graphs, the Laplacian matrix is symmetric, thus all of its eigenvalues are real.

2.4 Average Consensus

A consensus algorithm (or protocol) is an interaction rule that defines how information should be exchanged among agents in a network. We say that the agents reach a consensus when they agree on a certain numerical entity of interest (a scalar, vector, or matrix) that depends on the states of all agents (OLFATI-SABER; FAX; MURRAY, 2007). By design, consensus algorithms are distributed and rely only on the information exchange between each agent and its limited set of neighbors, i.e., no fusion center is required (REN; BEARD; ATKINS, 2007).

In particular, the average consensus is an algorithm to compute the arithmetic mean of a set of numerical entities. Consider a network described by a connected undirected graph $\mathbb{G}$ with S nodes. Each node $i \in \mathbb{S} := \{1, 2, \ldots, S\}$ can exchange data with a limited set $\mathcal{N}_i$ of neighbors at discrete instants of time. Suppose that each node is initialized with a state $\alpha_i(0)$. At each step ℓ, the nodes update their state using data from their neighborhood. As $\ell \to \infty$, the goal is to make the state $\alpha_i(\ell)$ of all nodes converge to the average value of their initial states. Through Algorithm 2.2, adapted from Ren, Beard and Atkins (2007), we can achieve average consensus in a distributed fashion.

Algorithm 2.2 Average consensus (each node $i \in \mathbb{S}$) (REN; BEARD; ATKINS, 2007)

Initialization: Set initial consensus state $\alpha_i(0)$.

for $\ell = 0, 1, \ldots, L-1$ **do**

1. Send the current state $\alpha_i(\ell)$ to all neighbors $j \in \mathcal{N}_i$.
2. Receive the current state $\alpha_j(\ell)$ from all neighbors $j \in \mathcal{N}_i$.
3. Update the consensus state

$$\alpha_i(\ell + 1) = \sum_{j=1}^{S} \pi_{ij} \alpha_j(\ell)$$

end for

Definition 2.2. *In step 3 of Algorithm 2.2, π_{ij}, $\forall i, j \in \mathbb{S}$, are the so-called* consensus weights*, for which we assume the following characteristics:*

(i) $\pi_{ij} > 0$, *if* $i = j$;

(ii) $\pi_{ij} > 0$, *if* $j \in \mathcal{N}_i$;

(iii) $\pi_{ij} = 0$, *otherwise;*

(iv) $\sum_{j=1}^{S} \pi_{ij} = 1$, $\forall i \in \mathbb{S}$.

Moreover, we define the weighted adjacency matrix $\Pi := [\pi_{ij}]$, *which describes the network communication topology.*

The following lemma outlines some properties of the weighted adjacency matrix Π defined above.

Lemma 2.9. *The weighted adjacency matrix* Π *(Definition 2.2) that describes the undirected connected graph* $\mathbb{G}$ *has the following properties:*

(i) Π *is a nonnegative doubly stochastic matrix;*

(ii) Π *is an irreducible matrix;*

(iii) *All eigenvalues of* Π *are real and* $\rho(\Pi) = 1$*;*

(iv) $\rho(\Pi) = 1$ *is an algebraically simple eigenvalue of* Π*;*

(v) Π *is a primitive matrix.*

Proof. Let us address each of one the properties:

(i) This property holds by the way we construct Π. Since $\pi_{ij} \geq 0$, Π is nonnegative (see Definition A.3) and, given that the underlying graph $\mathbb{G}$ is undirected, $\pi_{ij} = \pi_{ji}$, $\forall i, j \in \mathbb{S}$, then Π is symmetric. Moreover, as $\sum_{j=1}^{S} \pi_{ij} = 1$, $\forall i \in \mathbb{S}$, every row and column sum of Π is unitary, therefore, it is doubly stochastic (see Definition A.4).

(ii) The second property follows directly from Lemma A.6, as the underlying graph $\mathbb{G}$ is undirected and connected.

(iii) The eigenvalues of Π are all real because it is a real symmetric matrix. The unitary spectral radius property follows from the Geršgorin Disk theorem (Theorem A.1) and the fact that it is doubly stochastic.

(iv) This property holds according to the Perron-Frobenius theorem (Theorem A.2), since Π is an irreducible matrix and $\rho(\Pi) = 1$.

(v) Finally, since Π is irreducible and has a single nonzero eigenvalue of maximum modulus, the last property follows directly from the definition of a primitive matrix (Definition A.6). □

The next theorem establishes the main result about the convergence of Algorithm 2.2 and how it achieves average consensus.

Theorem 2.2. *Consider a network described by a connected undirected graph $\mathbb{G}$ with S nodes and associated weighted adjacency matrix Π having the properties outlined in Lemma 2.9. If each node $i \in \mathbb{S}$ performs Algorithm 2.2 with an infinite number of consensus iterations, i.e., $L \to \infty$, the states of all nodes converge asymptotically to the average value of their initial states, i.e.,*

$$\lim_{L\to\infty} \alpha_i(L) = \frac{1}{S}\sum_{i=1}^{S} \alpha_i(0), \ \ \forall i \in \mathbb{S}, \tag{2.18}$$

for any set of initial states.

Proof. Given the properties of the weighted adjacency matrix Π in Lemma 2.9, from its irreducibility and the fact that $\rho(\Pi) = 1$, the Perron-Frobenius theorem (Theorem A.2) states that there exist unique vectors $v \in \mathbb{R}^S$ and $w \in \mathbb{R}^S$ such that $\Pi v = v$ and $w^T\Pi = w^T$. Moreover, $v^T w = 1$. Since Π is doubly stochastic, $v = \mathbf{1}_S$, which implies that $w = (1/S)\mathbf{1}_S$, where $\mathbf{1}_S$ is a column vector of S ones. Then, as Π is primitive, Lemma A.7 indicates that

$$\lim_{\ell\to\infty} \Pi^\ell = vw^T = \frac{1}{S}\mathbf{1}_S\mathbf{1}_S^T. \tag{2.19}$$

To conclude, note that the collective dynamics of the network under Algorithm 2.2 can be written as $\boldsymbol{\alpha}(\ell) = \Pi^\ell\boldsymbol{\alpha}(0)$, where $\boldsymbol{\alpha}(\ell)$ is a column vector which stacks the states $\alpha_i(\ell)$ of all nodes. Therefore, substituting Π^ℓ by the result in (2.19), we obtain (2.18). □

Xiao, Boyd and Lall (2005) propose a possible choice of consensus weights π_{ij} that satisfy the necessary conditions discussed above, the so-called *Metropolis weights*:

$$\pi_{ij} = \begin{cases} \dfrac{1}{1+\max\{N_i, N_j\}}, & \text{if } i \neq j, \ \forall j \in \mathbb{N}_i, \\ 1 - \displaystyle\sum_{j\in\mathbb{N}_i} \pi_{ij}, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \tag{2.20}$$

where N_i and N_j are the number of neighbors of nodes i and j, respectively.

Chapter 3

Robust Kalman Filtering

This chapter addresses the problem of estimating the state of a dynamical system based on measurements obtained by a single sensor. It is divided into three sections. In the first section, we consider that the underlying target system and sensing models are nominal, i.e., perfectly known. However, exact models are seldom available in practice due to factors such as unmodeled dynamics, linearization, model reduction, and varying parameters. Therefore, in the second and third sections, we respectively assume that these models are subject to norm-bounded and polytopic parametric uncertainties. In each section, we propose a filtering strategy inspired by the celebrated Kalman filter (KALMAN, 1960), taking advantage of its efficiency and simplicity whilst overcoming one of its main weaknesses by compensating for the model uncertainties.

To develop the nominal and robust filters, we adopt a deterministic viewpoint (BRYSON; HO, 1975) and propose estimation problems constrained by each situation's specific target system and sensing models. Then, we develop a similar framework to solve them and obtain the filter expressions. The procedure fundamentally consists of applying the penalty function method (Section 2.1) to rewrite the estimation problems in the same form as one of the classic least-squares problems discussed in Section 2.2, depending on the presence and type of model uncertainties.

3.1 Nominal Kalman Filtering

In this section, we revisit the classic Kalman Filter (KALMAN, 1960), referring to it as the Nominal Kalman Filter (KF), to emphasize that all the parameter matrices that define the linear discrete-time target system and sensing models are perfectly known. Nevertheless, we assume a slightly more general model than usually found in the related works. The framework we develop here will serve as the foundation and reference to derive all the filters proposed in this work.

We formulate the nominal estimation problem from a deterministic viewpoint as a constrained regularized least-squares problem (Section 2.2.3), which we transform into an unconstrained equivalent by applying the penalty function method (Section 2.1). The solution to this modified problem then provides the recursive expressions of the Nominal Kalman Filter. We conclude the section with a stability analysis of the proposed estimator, assuming a time-invariant model.

3.1.1 Problem Formulation

3.1.1.1 System Model

Consider the following discrete-time state-space description of a linear dynamical system:

$$\begin{aligned} x_{k+1} &= F_k x_k + G_k u_k + H_k w_k, \\ y_k &= C_k x_k + D_k v_k, \end{aligned} \tag{3.1}$$

for $k = 0, 1, \ldots, N$, with state vector $x_k \in \mathbb{R}^n$, input vector $u_k \in \mathbb{R}^m$, system noise vector $w_k \in \mathbb{R}^p$, measurement vector $y_k \in \mathbb{R}^r$, and measurement noise vector $v_k \in \mathbb{R}^q$. $F_k \in \mathbb{R}^{n\times n}$, $G_k \in \mathbb{R}^{n\times m}$, $H_k \in \mathbb{R}^{n\times p}$, $C_k \in \mathbb{R}^{r\times n}$, and $D_k \in \mathbb{R}^{r\times q}$ are known nominal parameter matrices.

In a stochastic setting, it is usually assumed that x_0, w_k, and v_k are mutually independent zero-mean Gaussian random variables with respective variances

$$\boldsymbol{E}\Big\{x_0 x_0^T\Big\} = P_0 \succ 0, \quad \boldsymbol{E}\Big\{w_k w_l^T\Big\} = Q_k \delta_{kl} \succ 0, \quad \text{and} \quad \boldsymbol{E}\Big\{v_k v_l^T\Big\} = R_k \delta_{kl} \succ 0,$$

where δ_{kl} is the Kronecker delta function, such that $\delta_{kl} = 1$ if $k = l$, and $\delta_{kl} = 0$ otherwise. Nonetheless, as we shall see, the strategy we adopt to derive the filter does not require that these variables have any particular distribution.

3.1.1.2 Nominal Estimation Problem

Since the system state sequence $\{x_k\}$ is not perfectly observed, the problem consists of obtaining an estimate $\hat{x}_k$ of x_k leveraging all the information available up to time instant k, denoted

$$\boldsymbol{Y}_k = \{y_0, \ldots, y_k, u_0, \ldots, u_k\}. \tag{3.2}$$

In this context, we define two types of state estimates:

a) $\hat{x}_{k|k}$ denotes the *filtered* (or *posterior*) estimate of x_k, given $\boldsymbol{Y}_k$;

b) $\hat{x}_{k+1|k}$ denotes the *predicted* (or *predicted prior*) estimate of x_{k+1}, given $\boldsymbol{Y}_k$.

As discussed in Bryson and Ho (1975), stochastic estimation problems also admit a deterministic interpretation and can be formulated as least-squares problems. To avoid confusion, here we adopt the variables $\hat{x}_k$, $\hat{x}_{k+1}$, $\hat{w}_k$, and $\hat{v}_k$ as substitutes of the random variables x_k, x_{k+1}, w_k, and v_k in the stochastic model (3.1).

In the deterministic context, the estimation problem consists of obtaining $\hat{x}_k$, $\hat{x}_{k+1}$, $\hat{w}_k$, and $\hat{v}_k$ that best fit the model

$$\begin{aligned} \hat{x}_{k+1} &= F_k\hat{x}_k + G_k u_k + H_k\hat{w}_k, \\ y_k &= C_k\hat{x}_k + D_k\hat{v}_k, \end{aligned} \tag{3.3}$$

where $\hat{w}_k$ and $\hat{v}_k$ are interpreted as fitting errors, weighted by matrices $Q_k \succ 0$ and $R_k \succ 0$, respectively. In addition, we define the approximation errors $e_{k|k} = \hat{x}_k - \hat{x}_{k|k}$ and $e_{k+1|k} = \hat{x}_{k+1} - \hat{x}_{k+1|k}$, respectively weighted by matrices $P_{k|k} \succ 0$ and $P_{k+1|k} \succ 0$.

The goal is to formulate an optimization problem whose optimal solution ($\hat{x}_k^*$, $\hat{x}_{k+1}^*$, $\hat{w}_k^*, \hat{v}_k^*$) satisfy model (3.3), given the available observation set $\boldsymbol{Y}_k$ in (3.2). We can then relate this solution to the best estimates of the original random variables x_k, x_{k+1}, w_k, and v_k, according to the following definitions:

$$\hat{x}_{k|k} \coloneqq \hat{x}_k^*, \quad \hat{x}_{k+1|k} \coloneqq \hat{x}_{k+1}^*, \quad \hat{w}_{k|k} \coloneqq \hat{w}_k^*, \quad \text{and} \quad \hat{v}_{k|k} \coloneqq \hat{v}_k^*.$$

To fulfill this objective, based on Kailath, Sayed and Hassibi (2000a) and Sayed (2001), assuming that at each time step k, an *a priori* state estimate $\hat{x}_{k|k-1}$, a measurement y_k, and the input u_k are available, we formulate a constrained optimization problem with a one-step quadratic objective function, as follows:

$$\begin{aligned} \min_{\substack{\hat{x}_k, \hat{x}_{k+1}, \\ \hat{w}_k, \hat{v}_k}} \quad & J_k(\hat{x}_k, \hat{w}_k, \hat{v}_k) = \|\hat{x}_k - \hat{x}_{k|k-1}\|^2_{P^{-1}_{k|k-1}} + \|\hat{w}_k\|^2_{Q_k^{-1}} + \|\hat{v}_k\|^2_{R_k^{-1}}, \\ \text{subject to} \quad & \begin{cases} \hat{x}_{k+1} = F_k\hat{x}_k + G_k u_k + H_k\hat{w}_k, \\ \quad y_k = C_k\hat{x}_k + D_k\hat{v}_k, \end{cases} \end{aligned} \tag{3.4}$$

for $k = 0, 1, \ldots, N$. The solution to this problem recursively provides the filtered and predicted state estimates $\hat{x}_{k|k}$ and $\hat{x}_{k|k+1}$, respectively. Note that, from a stochastic viewpoint, matrices Q_k and R_k represent the variances of the random variables w_k and v_k. Nevertheless, in this more general deterministic setting, they are understood as weighting matrices. We refer to problem (3.4) as a *regularized least-squares estimation problem*, which we discuss in the next section.

3.1.2 Regularized Least-Squares Estimation Problem

Consider the general problem of obtaining an estimate $\hat{x}$ of an unknown vector x based on measurements y, related to x according to the linear system

$$y = Ax + Bw, \tag{3.5}$$

where w is a noise vector, also unknown, A and B are known matrices, and y is a known measurement vector. Furthermore, assume that an *a priori* estimate $\bar{x}$ of x is available as well.

From a deterministic viewpoint, we formulate the so-called regularized least-squares estimation problem as

$$\begin{aligned} \min_{x,\,w} \quad & J(x,w) = \|x-\bar{x}\|^2_{\bar{P}} + \|w\|^2_Q, \\ \text{subject to} \quad & y = Ax + Bw, \end{aligned} \tag{3.6}$$

where $\bar{P} \succeq 0$ and $Q \succ 0$ are given weighting matrices respectively associated with the *a priori* estimation error $x - \bar{x}$ and the model fitting error w.

The first step we take to solve the constrained problem (3.6) is to transform it into a more convenient unconstrained problem. Since the linear constraint (3.5) is quite general, it cannot be inserted into the objective function by direct substitution. Therefore, we apply the penalty function method presented in Section 2.1. The constraint is thus included into the objective function as a quadratic term multiplied by a penalty parameter $\mu > 0$. Violating the problem constraint will thus be penalized by this parameter. Hence, for a fixed $\mu > 0$, we rewrite problem (3.6) as

$$\min_{x,\,w} \; J^{\mu}(x,w), \tag{3.7}$$

with a new objective function

$$J^{\mu}(x,w) = \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix}^T \begin{bmatrix} \bar{P} & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix} + \left(\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix} - (y - A\bar{x}) \right)^T \mu I \left(\bullet \right). \tag{3.8}$$

Notice that problem (3.7)-(3.8) has the form of a regularized least-squares problem (Section 2.2.3), considering the following mappings between (2.10) and (3.8):

$$z \leftarrow \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix}, \quad \mathcal{Q} \leftarrow \begin{bmatrix} \bar{P} & 0 \\ 0 & Q \end{bmatrix}, \quad \mathcal{A} \leftarrow \begin{bmatrix} A & B \end{bmatrix}, \quad b \leftarrow y - A\bar{x}, \quad \text{and} \quad \mathcal{W} \leftarrow \mu I. \tag{3.9}$$

Therefore, to find a solution to problem (3.7)-(3.8), we use the results presented in Section 2.2.3. From the solution, we then extract the estimate $\hat{x}^{\mu}$ of x, conditioned by the penalty parameter μ, as the next lemma shows.

Lemma 3.1. *Consider problem (3.7)-(3.8), in which $\bar{P} \succeq 0$, $Q \succ 0$, and A has full column rank. The estimate $\hat{x}^{\mu}$ of x, conditioned by the penalty parameter $\mu > 0$, is given by*

$$\hat{x}^{\mu} = \left(\bar{P} + A^T \left(\mu^{-1} I + B Q^{-1} B^T \right)^{-1} A \right)^{-1} \left(\bar{P}\bar{x} + A^T \left(\mu^{-1} I + B Q^{-1} B^T \right)^{-1} y \right). \tag{3.10}$$

Proof. As previously mentioned, problem (3.7)-(3.8) is a regularized least-squares problem, considering the mappings in (3.9). Since $\bar{P} \succeq 0$, $Q \succ 0$, and $\mu > 0$, we have that $\mathcal{Q} \succeq 0$ and $\mathcal{W} \succ 0$. Therefore, we can use Lemma 2.6 to find the solution. Additionally, the block $\begin{bmatrix} 0 & I \\ A & B \end{bmatrix}$ should have full column rank, which is satisfied, since A is required to have full column rank. Thus, substituting the mappings (3.9) into the unique solution (2.11) yields

$$\begin{bmatrix} \hat{x}^{\mu} - \bar{x} \\ \hat{w}^{\mu} \end{bmatrix} = \begin{bmatrix} \bar{P} + \mu A^T A & \mu A^T B \\ \mu B^T A & Q + \mu B^T B \end{bmatrix}^{-1} \begin{bmatrix} \mu A^T \big(y - A\bar{x}\big) \\ \mu B^T \big(y - A\bar{x}\big) \end{bmatrix}.$$

Summing $\begin{bmatrix} \bar{x} \\ 0 \end{bmatrix}$ to both sides then gives

$$\begin{bmatrix} \hat{x}^{\mu} \\ \hat{w}^{\mu} \end{bmatrix} = \begin{bmatrix} \bar{P} + \mu A^T A & \mu A^T B \\ \mu B^T A & Q + \mu B^T B \end{bmatrix}^{-1} \begin{bmatrix} \bar{P}\bar{x} + \mu A^T y \\ \mu B^T y \end{bmatrix}.$$

The equation above represents a system of simultaneous equations. Hence, one can write the following set of equations:

$$\big(\bar{P} + \mu A^T A\big)\hat{x}^{\mu} + \mu A^T B \hat{w}^{\mu} = \bar{P}\bar{x} + \mu A^T y, \tag{3.11}$$

$$\mu B^T A \hat{x}^{\mu} + \big(Q + \mu B^T B\big)\hat{w}^{\mu} = \mu B^T y. \tag{3.12}$$

Isolating $\hat{w}^{\mu}$ in (3.12), we have

$$\hat{w}^{\mu} = \big(Q + \mu B^T B\big)^{-1} \mu B^T \big(y - A\hat{x}^{\mu}\big).$$

Substituting $\hat{w}^{\mu}$ back into (3.11) yields

$$\Big[\bar{P} + A^T\Big(\mu I - \mu B\big(Q + \mu B^T B\big)^{-1} B^T \mu\Big)A\Big]\hat{x}^{\mu} = \bar{P}\bar{x} + A^T\Big(\mu I - \mu B\big(Q + \mu B^T B\big)^{-1} B^T \mu\Big)y. \tag{3.13}$$

From Lemma A.1, we have that $\Big(\mu I - \mu B\big(Q + \mu B^T B\big)^{-1} B^T \mu\Big) = \big(\mu^{-1} I + BQ^{-1}B^T\big)^{-1}$. Then, substituting this term in (3.13) and isolating $\hat{x}^{\mu}$, we obtain the estimate shown in (3.10). □

Recall from Section 2.1 that according to the penalty function method, when the penalty parameter $\mu \to +\infty$, we have that problems (3.6) and (3.7)-(3.8) become equivalent. In this case, the solution to the unconstrained problem yields the optimal estimate $\hat{x}$, no longer conditioned by μ.

Corollary 3.1. *Consider the estimate $\hat{x}^{\mu}$ obtained in Lemma 3.1 as part of the solution to problem (3.7)-(3.8). If A has full column rank, B has full row rank, and we let $\mu \to +\infty$, then the optimal estimate $\hat{x}$ of x is given by*

$$\hat{x} = \Big(\bar{P} + A^T\big(BQ^{-1}B^T\big)^{-1}A\Big)^{-1}\Big(\bar{P}\bar{x} + A^T\big(BQ^{-1}B^T\big)^{-1}y\Big). \tag{3.14}$$

Proof. By letting $\mu \to +\infty$ in (3.10), we have that $\mu^{-1} \to 0$, such that the term $\big(\mu^{-1}I + BQ^{-1}B^T\big)^{-1}$ becomes $\big(BQ^{-1}B^T\big)^{-1}$ and we obtain the optimal estimate $\hat{x}$ in (3.14). In addition, as explained in Lemma A.5 (item **(ii)**), given that B has full row rank, we ensure invertibility of the term $\big(BQ^{-1}B^T\big)$, assuming that $Q \succ 0$. □

Finally, the next lemma shows that if we adopt a stochastic viewpoint of the estimation problem discussed here, we can further derive a variance matrix to associate with the estimation error $x - \hat{x}$.

Lemma 3.2. *Consider that in the linear system (3.5), the noise w and the prior estimation error $x-\bar{x}$ are mutually independent zero-mean Gaussian variables with respective variances $\boldsymbol{E}\left\{ww^T\right\} = Q^{-1}$ and $\boldsymbol{E}\left\{(x-\bar{x})(x-\bar{x})^T\right\} = \bar{P}^\dagger = \left(\bar{P}^T\bar{P}\right)^{-1}\bar{P}^T$. The variance matrix of the estimation error $x - \hat{x}$, for $\hat{x}$ as in (3.14), is given by*

$$\boldsymbol{E}\left\{(x-\hat{x})(x-\hat{x})^T\right\} = \left(\bar{P} + A^T\left(BQ^{-1}B^T\right)^{-1}A\right)^{-1}. \tag{3.15}$$

Proof. First, define $\hat{P} := \left(\bar{P} + A^T\left(BQ^{-1}B^T\right)^{-1}A\right)^{-1}$. Then, substituting y from (3.5) into (3.14) yields

$$\hat{x} = \hat{P}\left[\bar{P}\bar{x} + A^T\left(BQ^{-1}B^T\right)^{-1}\left(Ax + Bw\right)\right].$$

By adding $\hat{P}\bar{P}x$ to both sides of the equation above and performing some algebraic operations, we obtain the estimation error

$$x - \hat{x} = \hat{P}\left[\bar{P}\left(x - \bar{x}\right) - A^T\left(BQ^{-1}B^T\right)^{-1}Bw\right].$$

Then, we compute the estimation error variance matrix, as follows:

$$\boldsymbol{E}\left\{(x-\hat{x})(x-\hat{x})^T\right\} = \hat{P}\left[\bar{P}\boldsymbol{E}\left\{(x-\bar{x})(x-\bar{x})^T\right\}\bar{P}^T + A^T\left(BQ^{-1}B^T\right)^{-1}B\boldsymbol{E}\left\{ww^T\right\}B^T\left(BQ^{-1}B^T\right)^{-1}A\right]\hat{P}.$$

Since $\boldsymbol{E}\left\{(x-\bar{x})(x-\bar{x})^T\right\} = \bar{P}^\dagger = \left(\bar{P}^T\bar{P}\right)^{-1}\bar{P}^T$, with $\bar{P} = \bar{P}^T$, and $\boldsymbol{E}\left\{ww^T\right\} = Q^{-1}$, the equation above becomes

$$\boldsymbol{E}\left\{(x-\hat{x})(x-\hat{x})^T\right\} = \hat{P}\left(\bar{P} + A^T\left(BQ^{-1}B^T\right)^{-1}A\right)\hat{P} = \hat{P}\hat{P}^{-1}\hat{P} = \hat{P},$$

which corresponds to the result shown in (3.15). □

Remark 3.1. *In a deterministic context, the estimation error variance matrix found in Lemma 3.2 can be interpreted as a weighting matrix on the estimation error $x - \hat{x}$.*

3.1.3 Nominal Kalman Filter

We are now ready to use the results presented in Section 3.1.2 and ultimately obtain the Nominal Kalman Filter. As mentioned earlier, the deterministic estimation problem (3.4) is a special case of a regularized least-squares estimation problem, according to the following mappings between (3.4) and (3.6):

$$x \leftarrow \begin{bmatrix} \hat{x}_k \\ \hat{x}_{k+1} \end{bmatrix}, \quad \bar{x} \leftarrow \begin{bmatrix} \hat{x}_{k|k-1} \\ 0 \end{bmatrix}, \quad w \leftarrow \begin{bmatrix} \hat{w}_k \\ \hat{v}_k \end{bmatrix}, \quad \bar{P} \leftarrow \begin{bmatrix} P_{k|k-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad Q \leftarrow \begin{bmatrix} Q_k^{-1} & 0 \\ 0 & R_k^{-1} \end{bmatrix},$$
$$y \leftarrow \begin{bmatrix} -G_k u_k \\ y_k \end{bmatrix}, \quad A \leftarrow \begin{bmatrix} F_k & -I_n \\ C_k & 0 \end{bmatrix}, \quad \text{and} \quad B \leftarrow \begin{bmatrix} H_k & 0 \\ 0 & D_k \end{bmatrix}. \tag{3.16}$$

Note that, since $P_{k|k-1}^{-1} \succ 0$, we have that $\bar{P} \succeq 0$. Also, $Q_k^{-1} \succ 0$ and $R_k^{-1} \succ 0$, such that $Q \succ 0$. Therefore, by using the results in Corollary 3.1 and Lemma 3.2, we obtain the optimal filtered and predicted state estimates, $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, along with their corresponding error weighting matrices $P_{k|k}$ and $P_{k+1|k}$.

Theorem 3.1. *Consider the regularized least-squares estimation problem (3.4) with H_k and D_k full row rank and given initial conditions $\hat{x}_{0|-1}$, $P_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, and $R_k \succ 0$. For each $k = 0, 1, \ldots, N$, its solution recursively provides the filtered and predicted state estimates of system (3.1), $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, as well as their corresponding error weighting matrices, $P_{k|k}$ and $P_{k+1|k}$, according to the procedure outlined in Algorithm 3.1.*

Algorithm 3.1 Nominal Kalman Filter (KF)

Model: Assume the system model in (3.1).
Initialization: Set $\hat{x}_{0|-1}$, $P_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, and $R_k \succ 0$.
for $k = 0, 1, \ldots, N$ **do**

1. Obtain a measurement y_k.
2. Compute the auxiliary matrices:

$$\widehat{Q}_k = H_k Q_k H_k^T \qquad \widehat{R}_k = D_k R_k D_k^T$$

3. Correction step:
   3.1. Compute the posterior error weighting matrix:

$$P_{k|k} = \left(P_{k|k-1}^{-1} + C_k^T \widehat{R}_k^{-1} C_k\right)^{-1}$$

   3.2. Compute the filtered state estimate:

$$\hat{x}_{k|k} = P_{k|k}\left(P_{k|k-1}^{-1}\hat{x}_{k|k-1} + C_k^T \widehat{R}_k^{-1} y_k\right)$$

4. Prediction step:
   4.1. Update the predicted prior error weighting matrix:

$$P_{k+1|k} = F_k P_{k|k} F_k^T + \widehat{Q}_k$$

   4.2. Update the predicted prior state estimate:

$$\hat{x}_{k+1|k} = F_k \hat{x}_{k|k} + G_k u_k$$

end for

Proof. Since problem (3.4) is a regularized least-squares estimation problem, one can apply the result in Corollary 3.1 to obtain the optimal system state estimates $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$. Thus, substituting the mappings (3.16) into the optimal solution (3.14) gives

$$\begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \end{bmatrix} = \begin{bmatrix} P_{k|k-1}^{-1} + F_k^T \widehat{Q}_k^{-1} F_k + C_k^T \widehat{R}_k^{-1} C_k & -F_k^T \widehat{Q}_k^{-1} \\ -\widehat{Q}_k^{-1} F_k & \widehat{Q}_k^{-1} \end{bmatrix}^{-1} \times \begin{bmatrix} P_{k|k-1}^{-1} \hat{x}_{k|k-1} + C_k^T \widehat{R}_k^{-1} y_k - F_k^T \widehat{Q}_k^{-1} G_k u_k \\ \widehat{Q}_k^{-1} G_k u_k \end{bmatrix},$$

in which we define the auxiliary matrices $\widehat{Q}_k := H_k Q_k H_k^T$ and $\widehat{R}_k := D_k R_k D_k^T$, whose positive definiteness is guaranteed according to Lemma A.5 (item **(ii)**), since $Q_k \succ 0$, $R_k \succ 0$, and H_k and D_k have full row rank.

The equation above also represents a system of simultaneous equations. Therefore, we can write it as the following set of equations:

$$\left(P_{k|k-1}^{-1} + F_k^T \widehat{Q}_k^{-1} F_k + C_k^T \widehat{R}_k^{-1} C_k\right) \hat{x}_{k|k} - F_k^T \widehat{Q}_k^{-1} \hat{x}_{k+1|k} = P_{k|k-1}^{-1} \hat{x}_{k|k-1} + C_k^T \widehat{R}_k^{-1} y_k - F_k^T \widehat{Q}_k^{-1} G_k u_k, \tag{3.17}$$

$$-\widehat{Q}_k^{-1} F_k \hat{x}_{k|k} + \widehat{Q}_k^{-1} \hat{x}_{k+1|k} = \widehat{Q}_k^{-1} G_k u_k. \tag{3.18}$$

Isolating $\hat{x}_{k+1|k}$ in (3.18), we have

$$\hat{x}_{k+1|k} = F_k \hat{x}_{k|k} + G_k u_k,$$

which corresponds to the update equation of the predicted prior state estimate in step 4.2 of Algorithm 3.1. Then, substituting $\hat{x}_{k+1|k}$ back into (3.17) and isolating $\hat{x}_{k|k}$ yields

$$\hat{x}_{k|k} = \left(P_{k|k-1}^{-1} + C_k^T \widehat{R}_k^{-1} C_k\right)^{-1} \left(P_{k|k-1}^{-1} \hat{x}_{k|k-1} + C_k^T \widehat{R}_k^{-1} y_k\right),$$

which corresponds to the equation for computing the filtered state estimate in step 3.2 of Algorithm 3.1.

Now, to obtain the error weighting matrices associated with $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, we apply Lemma 3.2, assuming a deterministic context, as mentioned in Remark 3.1. Thus, substituting the mappings (3.16) into (3.15) gives[1]

$$\begin{bmatrix} P_{k|k} & * \\ * & P_{k+1|k} \end{bmatrix} = \underbrace{\begin{bmatrix} P_{k|k-1}^{-1} + F_k^T \widehat{Q}_k^{-1} F_k + C_k^T \widehat{R}_k^{-1} C_k & -F_k^T \widehat{Q}_k^{-1} \\ -\widehat{Q}_k^{-1} F_k & \widehat{Q}_k^{-1} \end{bmatrix}^{-1}}_{\mathcal{M}^{-1}} =: \begin{bmatrix} \mathcal{M}_1 & \mathcal{M}_2 \\ \mathcal{M}_2^T & \mathcal{M}_3 \end{bmatrix}^{-1},$$

where we define the partitioned matrix $\mathcal{M}$. To find its inverse, we use the Banachiewicz inversion formula (Lemma A.4, item **(ii)**). According to Lemma A.3, the Schur complement of $\mathcal{M}_3$ in $\mathcal{M}$ is

$$(\mathcal{M}/\mathcal{M}_3) = \mathcal{M}_1 - \mathcal{M}_2 \mathcal{M}_3^{-1} \mathcal{M}_2^T = P_{k|k-1}^{-1} + C_k^T \widehat{R}_k^{-1} C_k.$$

[1] The elements marked with $*$ are byproducts with no particular meaning in our context.

The posterior error weighting matrix in step 3.1 of Algorithm 3.1 is then obtained as follows:

$$P_{k|k} = (\mathcal{M}/\mathcal{M}_3)^{-1} = \left(P_{k|k-1}^{-1} + C_k^T \widehat{R}_k^{-1} C_k\right)^{-1}.$$

Finally, we obtain the predicted prior error weighting matrix

$$P_{k+1|k} = \mathcal{M}_3^{-1} + \mathcal{M}_3^{-1}\mathcal{M}_2^T(\mathcal{M}/\mathcal{M}_3)^{-1}\mathcal{M}_2\mathcal{M}_3^{-1} = F_k P_{k|k} F_k^T + \widehat{Q}_k,$$

as shown in step 4.1 of Algorithm 3.1. □

Notice that when $G_k = 0$, $H_k = I_n$, and $D_k = I_r$, Algorithm 3.1 collapses to the standard Kalman filter. In this case, if one chooses the noise vectors weighting matrices Q_k and R_k as the noise variance matrices, then the posterior and predicted prior estimation error weighting matrices $P_{k|k}$ and $P_{k+1|k}$ can be interpreted as error variance matrices. Hence, although a deterministic viewpoint was used to derive the filter equations, there is an equivalence with the stochastic viewpoint. Nevertheless, the former can handle more generic problems, since no assumptions on the distributions of w_k and v_k are necessary. Furthermore, we emphasize how the use of the penalty function approach enabled the inclusion of matrices H_k and D_k in the more general linear discrete-time target system model (3.1).

3.1.4 Stability Analysis

In this section, we conclude the discussion on nominal Kalman filtering by examining the steady-state behavior of the estimator described in Algorithm 3.1 when the system model parameters are constant and there is no input u_k. Thus, consider the following discrete-time state-space description of a linear system:

$$\begin{aligned} x_{k+1} &= F x_k + H w_k, \\ y_k &= C x_k + D v_k, \end{aligned} \tag{3.19}$$

for $k \geq 0$. We seek to establish conditions for the stability of the steady-state filter.

Let us consider the system model (3.19). Thus, the KF equations in steps 3 and 4 of Algorithm 3.1 become:

$$P_{k|k} = \left(P_{k|k-1}^{-1} + C^T \widehat{R}^{-1} C\right)^{-1}, \tag{3.20}$$

$$\hat{x}_{k|k} = P_{k|k}\left(P_{k|k-1}^{-1}\hat{x}_{k|k-1} + C^T \widehat{R}^{-1} y_k\right), \tag{3.21}$$

$$P_{k+1|k} = F P_{k|k} F^T + \widehat{Q}, \tag{3.22}$$

$$\hat{x}_{k+1|k} = F \hat{x}_{k|k}, \tag{3.23}$$

where $\widehat{Q} = HQH^T$ and $\widehat{R} = DRD^T$. Applying the matrix inversion lemma (Lemma A.1), we expand expression (3.20), as follows:

$$P_{k|k} = P_{k|k-1} - P_{k|k-1}C^T\left(\widehat{R} + CP_{k|k-1}C^T\right)^{-1}CP_{k|k-1}. \tag{3.24}$$

Then, combining (3.24) with (3.21) and substituting in (3.23) yields the steady-state predicted state estimate

$$\hat{x}_{k+1|k} = \widetilde{F}_k \hat{x}_{k|k-1} + \widetilde{F}_k P_{k|k-1} C^T \widehat{R}^{-1} y_k, \tag{3.25}$$

where

$$\widetilde{F}_k = F\Big(I_n - P_{k|k-1} C^T \big(\widehat{R} + C P_{k|k-1} C^T\big)^{-1} C\Big)$$

is the filter closed-loop matrix. Moreover, substituting $P_{k|k}$ from (3.24) into (3.22), we obtain the following expression for the predicted prior error weighting matrix:

$$P_{k+1|k} = F\Big(P_{k|k-1} - P_{k|k-1} C^T \big(\widehat{R} + C P_{k|k-1} C^T\big)^{-1} C P_{k|k-1}\Big) F^T + \widehat{Q}. \tag{3.26}$$

The following theorem establishes a result concerning the convergence of the filter to a stable steady-state filter.

Theorem 3.2. *Consider the linear system model* (3.19) *and the corresponding filter* (3.25)-(3.26). *Assume that* $\{F, C\}$ *is* detectable *and* $\{F, \widehat{Q}^{1/2}\}$ *is* controllable. *Then, for any initial condition* $P_{0|-1} \succ 0$, $P_{k+1|k}$ *converges to the* unique stabilizing solution $P \succ 0$ *of the algebraic Riccati equation*

$$P = F\Big(P - P C^T \big(\widehat{R} + C P C^T\big)^{-1} C P\Big) F^T + \widehat{Q}. \tag{3.27}$$

The solution P is stabilizing in the sense that the steady-state filter closed-loop matrix

$$\widetilde{F} = F\Big(I_n - P C^T \big(\widehat{R} + C P C^T\big)^{-1} C\Big) \tag{3.28}$$

is Schur stable.

Proof. As shown in Kailath, Sayed and Hassibi (2000b), detectability of $\{F, C\}$ and controllability of $\{F, \widehat{Q}^{1/2}\}$ ensure the convergence of $P_{k+1|k}$ in (3.26) to the unique stabilizing positive definite solution P of the algebraic Riccati equation (3.27) that stabilizes (3.28), which is the filter steady-state closed-loop matrix. □

3.2 Robust Kalman Filtering for Systems with Norm-Bounded Uncertainties

In this section, we present a robust version of the Nominal Kalman Filter introduced in Section 3.1. We address the special case where the underlying target system and sensing models are subject to norm-bounded parametric uncertainties.

We extend the framework used in Section 3.1 to derive a robust filter. Analogously, we formulate the robust estimation problem as a deterministic constrained regularized least-squares problem with uncertainties (Section 2.2.4). We then apply the penalty function method (Section 2.1) to transform it into a more convenient unconstrained problem, whose solution ultimately provides the recursive expressions of the Robust Kalman Filter (RKF). The estimator is presented as correction-prediction algorithm, similar to the Nominal Kalman Filter. We further study the stability properties of the proposed estimator and conclude the section with an illustrative example.

3.2.1 Problem Formulation

3.2.1.1 System Model

Consider the following discrete-time state-space description of a linear system with uncertainties:

$$\begin{aligned} x_{k+1} &= (F_k + \delta F_k)x_k + (G_k + \delta G_k)u_k + (H_k + \delta H_k)w_k, \\ y_k &= (C_k + \delta C_k)x_k + (D_k + \delta D_k)v_k, \end{aligned} \tag{3.29}$$

for $k = 0, 1, \ldots, N$, with state vector $x_k \in \mathbb{R}^n$, input vector $u_k \in \mathbb{R}^m$, system noise vector $w_k \in \mathbb{R}^p$, measurement vector $y_k \in \mathbb{R}^r$, and measurement noise vector $v_k \in \mathbb{R}^q$. $F_k \in \mathbb{R}^{n\times n}$, $G_k \in \mathbb{R}^{n\times m}$, $H_k \in \mathbb{R}^{n\times p}$, $C_k \in \mathbb{R}^{r\times n}$, and $D_k \in \mathbb{R}^{r\times q}$ are known nominal parameter matrices, whereas $\delta F_k \in \mathbb{R}^{n\times n}$, $\delta G_k \in \mathbb{R}^{n\times m}$, $\delta H_k \in \mathbb{R}^{n\times p}$, $\delta C_k \in \mathbb{R}^{r\times n}$, and $\delta D_k \in \mathbb{R}^{r\times q}$ are norm-bounded parametric uncertainties modeled as

$$\begin{aligned} \begin{bmatrix}\delta F_k & \delta G_k & \delta H_k\end{bmatrix} &= M_{1,k}\Delta_{1,k}\begin{bmatrix}E_{F_k} & E_{G_k} & E_{H_k}\end{bmatrix}, \quad \|\Delta_{1,k}\| \le 1, \\ \begin{bmatrix}\delta C_k & \delta D_k\end{bmatrix} &= M_{2,k}\Delta_{2,k}\begin{bmatrix}E_{C_k} & E_{D_k}\end{bmatrix}, \qquad \|\Delta_{2,k}\| \le 1, \end{aligned} \tag{3.30}$$

where $M_{1,k} \in \mathbb{R}^{n\times s_1}$ and $M_{2,k} \in \mathbb{R}^{r\times s_2}$ are known nonzero matrices, $E_{F_k} \in \mathbb{R}^{t_1\times n}$, $E_{G_k} \in \mathbb{R}^{t_1\times m}$, $E_{H_k} \in \mathbb{R}^{t_1\times p}$, $E_{C_k} \in \mathbb{R}^{t_2\times n}$ and $E_{D_k} \in \mathbb{R}^{t_2\times q}$ are also known, and $\Delta_{1,k} \in \mathbb{R}^{s_1\times t_1}$ and $\Delta_{2,k} \in \mathbb{R}^{s_2\times t_2}$ are arbitrary contraction matrices. Perturbations of the form (3.30) are useful when modeling tolerance specifications on the physical parameters of a system and are thus common in robust filtering and control (SAYED, 2001).

In a stochastic interpretation, we usually assume that x_0, w_k, and v_k are mutually independent zero-mean Gaussian random variables with respective variances

$$\boldsymbol{E}\left\{x_0x_0^T\right\} = P_0 \succ 0, \quad \boldsymbol{E}\left\{w_kw_l^T\right\} = Q_k\delta_{kl} \succ 0, \quad \text{and} \quad \boldsymbol{E}\left\{v_kv_l^T\right\} = R_k\delta_{kl} \succ 0,$$

where δ_{kl} is the Kronecker delta function, such that $\delta_{kl} = 1$ if $k = l$, and $\delta_{kl} = 0$ otherwise. Nevertheless, the strategy we adopt to derive the robust filter does not require that these variables have any particular distribution.

3.2.1.2 Robust Estimation Problem

The goal is to design a robust state estimator for the uncertain system (3.29)-(3.30). As the system state sequence $\{x_k\}$ is not perfectly observed, the problem consists of using all the information available up to time instant k, $\boldsymbol{Y}_k = \{y_0, \dots, y_k, u_0, \dots, u_k\}$, to obtain a so-called filtered state estimate $\hat{x}_{k|k}$ of x_k, as well as a predicted estimate $\hat{x}_{k+1|k}$ of x_{k+1}, despite the presence of model uncertainties $\delta_k \coloneqq \{\delta F_k,\, \delta G_k,\, \delta H_k,\, \delta C_k,\, \delta D_k\}$.

Following the procedure described in Section 3.1.1.2 for the Nominal Kalman Filter, we adopt a deterministic viewpoint (BRYSON; HO, 1975). Moreover, we introduce the variables $\hat{x}_k$, $\hat{x}_{k+1}$, $\hat{w}_k$, and $\hat{v}_k$ as substitutes for the random variables x_k, x_{k+1}, w_k, and v_k in the stochastic model (3.29). Then, based on Sayed (2001) and Ishihara, Terra and Cerri (2015), assuming that at each time step k, an *a priori* state estimate $\hat{x}_{k|k-1}$, a measurement y_k, and the input u_k are available, we formulate a min-max constrained optimization problem in which a one-step quadratic objective function should be minimized under the maximum influence of the model parametric uncertainties δ_k, i.e.,

$$\begin{aligned} \min_{\substack{\hat{x}_k, \hat{x}_{k+1}, \\ \hat{w}_k, \hat{v}_k}} \max_{\delta_k} \quad & J_k(\hat{x}_k, \hat{w}_k, \hat{v}_k) = \|\hat{x}_k - \hat{x}_{k|k-1}\|^2_{P^{-1}_{k|k-1}} + \|\hat{w}_k\|^2_{Q_k^{-1}} + \|\hat{v}_k\|^2_{R_k^{-1}}, \\ \text{subject to} \quad & \begin{cases} \hat{x}_{k+1} = (F_k + \delta F_k)\hat{x}_k + (G_k + \delta G_k)u_k + (H_k + \delta H_k)\hat{w}_k, \\ \quad y_k = (C_k + \delta C_k)\hat{x}_k + (D_k + \delta D_k)\hat{v}_k, \end{cases} \end{aligned} \tag{3.31}$$

for $k = 0, 1, \dots, N$ and uncertainties δ_k as defined in (3.30). Here, $\hat{w}_k$ and $\hat{v}_k$ are fitting errors weighted respectively by $Q_k \succ 0$ and $R_k \succ 0$, and $P_{k|k-1} \succ 0$ weights the *a priori* estimation error $x_k - \hat{x}_{k|k-1}$. Note that, from a stochastic viewpoint, matrices Q_k and R_k represent the variances of the random variables w_k and v_k. Nevertheless, in this more general deterministic framework, they are treated as weighting matrices.

The solution to this problem recursively provides the filtered and predicted robust state estimates $\hat{x}_{k|k}$ and $\hat{x}_{k|k+1}$. We refer to problem (3.31) as a *regularized least-squares estimation problem with norm-bounded uncertainties*, which is the topic of the next section.

3.2.2 Regularized Least-Squares Estimation Problem with Norm-Bounded Uncertainties

Consider the general problem of obtaining an estimate $\hat{x}$ of an unknown vector x based on measurements y, related to x according to the uncertain linear system

$$(y + \delta y) = (A + \delta A)x + (B + \delta B)w, \tag{3.32}$$

where w is a noise vector, also unknown, A and B are known matrices, and y is a known measurement vector. The parametric uncertainties $\delta y, \delta A$, and δB are norm-bounded, being modeled as

$$\begin{bmatrix} \delta y & \delta A & \delta B \end{bmatrix} = M\Delta \begin{bmatrix} E_y & E_A & E_B \end{bmatrix}, \quad \|\Delta\| \le 1, \tag{3.33}$$

in which M is a known nonzero matrix, E_y, E_A, and E_B are also known, and Δ is an arbitrary contraction matrix. Furthermore, assume that an *a priori* estimate $\bar{x}$ of x is available as well.

From a deterministic viewpoint, we formulate the so-called regularized least-squares estimation problem with norm-bounded uncertainty as

$$\begin{aligned} \min_{x,w} \max_{\delta y, \delta A, \delta B} \quad & J(x,w) = \|x-\bar{x}\|_{\bar{P}}^2 + \|w\|_Q^2, \\ \text{subject to} \quad & (y+\delta y) = (A+\delta A)x + (B+\delta B)w, \end{aligned} \tag{3.34}$$

where $\bar{P} \succeq 0$ and $Q \succ 0$ are given weighting matrices respectively associated with the *a priori* estimation error $x - \bar{x}$ and the model fitting error w. The objective function should thus be minimized under the maximum influence of the parametric uncertainties.

Similar to the procedure carried out in Section 3.1.2 for the regularized least-squares estimation problem, we first transform the constrained problem (3.34) into an unconstrained problem. The linear constraint (3.32) has a general form and cannot be directly inserted into the objective function by substitution. Therefore, we apply the penalty function method presented in Section 2.1, whereby the constraint is included in the objective function as a quadratic term multiplied by a penalty parameter $\mu > 0$, which penalizes constraint violations. Hence, for a fixed $\mu > 0$, we rewrite problem (3.34) as

$$\min_{x,w} \max_{\delta y, \delta A, \delta B} J^\mu(x, w, \delta y, \delta A, \delta B), \tag{3.35}$$

with a new objective function

$$\begin{aligned} J^\mu(x,w,\delta y,\delta A,\delta B) = & \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix}^T \begin{bmatrix} \bar{P} & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix} + \\ & \left\{ \left(\begin{bmatrix} A & B \end{bmatrix} + \begin{bmatrix} \delta A & \delta B \end{bmatrix} \right) \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix} - \begin{bmatrix} (y-A\bar{x}) + (\delta y - \delta A \bar{x}) \end{bmatrix} \right\}^T \mu I \left\{ \bullet \right\}. \end{aligned} \tag{3.36}$$

Problem (3.35)-(3.36) has the form of a regularized least-squares problem with uncertainties (Section 2.2.4), considering the following mappings between (2.13) and (3.36):

$$\begin{gathered} z \leftarrow \begin{bmatrix} x-\bar{x} \\ w \end{bmatrix}, \quad \mathcal{Q} \leftarrow \begin{bmatrix} \bar{P} & 0 \\ 0 & Q \end{bmatrix}, \quad \mathcal{A} \leftarrow \begin{bmatrix} A & B \end{bmatrix}, \quad b \leftarrow y - A\bar{x}, \quad \mathcal{W} \leftarrow \mu I, \\ \delta\mathcal{A} \leftarrow \begin{bmatrix} \delta A & \delta B \end{bmatrix}, \quad \text{and} \quad \delta b \leftarrow \delta y - \delta A \bar{x}. \end{gathered} \tag{3.37}$$

Moreover, the correspondence with the norm-bounded parametric uncertainty model in (2.14) is given by

$$\begin{bmatrix} \delta\mathcal{A} & \delta b \end{bmatrix} = M\Delta \begin{bmatrix} E_{\mathcal{A}} & E_b \end{bmatrix}, \quad \|\Delta\| \leq 0,$$

where

$$M \leftarrow M, \quad \Delta \leftarrow \Delta, \quad E_{\mathcal{A}} \leftarrow \begin{bmatrix} E_A & E_B \end{bmatrix}, \quad \text{and} \quad E_b \leftarrow E_y - E_A \bar{x}. \tag{3.38}$$

Therefore, to find a solution to problem (3.35)-(3.36), we use the results in Section 2.2.4. From the solution, we then extract the estimate $\hat{x}^\mu$ of x, conditioned by the penalty parameter μ, as shown in the next lemma.

Lemma 3.3. *Consider problem (3.35)-(3.36) with norm-bounded parametric uncertainties given by (3.33), in which* $\bar{P} \succeq 0$, $Q \succ 0$, *and* $\begin{bmatrix} A \\ E_A \end{bmatrix}$ *has full column rank. The estimate* $\hat{x}^\mu$ *of* x, *conditioned by the penalty parameter* $\mu > 0$, *is given by*

$$\hat{x}^\mu = \Big(\bar{P} + \widehat{A}^T \widehat{Q}^{-1} \widehat{A} + E_A^T \bar{Q}^{-1} E_A\Big)^{-1} \Big(\bar{P}\bar{x} + \widehat{A}^T \widehat{Q}^{-1} \hat{y} + E_A^T \bar{Q}^{-1} E_y\Big), \tag{3.39}$$

in which we define the auxiliary entities

$$\begin{aligned} &\Phi := \mu I + \mu M\Big(\hat{\lambda} I - \mu M^T M\Big)^{-1} M^T \mu = \Big(\mu^{-1} I - \hat{\lambda}^{-1} M M^T\Big)^{-1}, \\ &\bar{Q} := \hat{\lambda}^{-1} I + E_B Q^{-1} E_B^T, \qquad \widehat{Q} := \Phi^{-1} + B\Big(Q + \hat{\lambda} E_B^T E_B\Big)^{-1} B^T, \\ &\widehat{A} := A - BQ^{-1} E_B^T \bar{Q}^{-1} E_A, \quad \hat{y} := y - BQ^{-1} E_B^T \bar{Q}^{-1} E_y, \end{aligned} \tag{3.40}$$

where $\hat{\lambda}$ *is a nonnegative scalar parameter obtained from the auxiliary optimization problem*

$$\hat{\lambda} := \arg \min_{\lambda > \mu \|M^T M\|} \Gamma(\lambda), \tag{3.41}$$

with objective function $\Gamma(\lambda)$ *given by*

$$\Gamma(\lambda) := \|z(\lambda)\|_{\mathcal{Q}}^2 + \lambda \|E_{\mathcal{A}} z(\lambda) - E_b\|^2 + \|\mathcal{A} z(\lambda) - b\|_{\Phi(\lambda)}^2, \tag{3.42}$$

in which

$$\begin{aligned} &\Phi(\lambda) := \mu I + \mu M\Big(\lambda I - \mu M^T M\Big)^{-1} M^T \mu, \\ &z(\lambda) := \Big(\mathcal{Q} + \mathcal{A}^T \Phi(\lambda) \mathcal{A} + \lambda E_{\mathcal{A}}^T E_{\mathcal{A}}\Big)^{-1} \Big(\mathcal{A}^T \Phi(\lambda) b + \lambda E_{\mathcal{A}}^T E_b\Big), \end{aligned}$$

considering the definitions in (3.37) and (3.38).

Proof. As we mentioned previously, problem (3.35)-(3.36) is a regularized least-squares problem with uncertainties, considering the mappings in (3.37) and (3.38). Since $\bar{P} \succeq 0$, $Q \succ 0$, and $\mu > 0$, we have that $\mathcal{Q} \succeq 0$ and $\mathcal{W} \succ 0$. Therefore, we can use Lemma 2.8 to find the solution. Additionally, the block $\begin{bmatrix} 0 & I \\ A & B \\ E_A & E_B \end{bmatrix}$ should have full column rank, which is satisfied by the requirement of $\begin{bmatrix} A \\ E_A \end{bmatrix}$ having full column rank. Thus, substituting the mappings (3.37) and (3.38) into the unique solution (2.15) yields

$$\begin{bmatrix} \hat{x}^\mu - \bar{x} \\ \hat{w}^\mu \end{bmatrix} = \begin{bmatrix} \bar{P} + A^T \Phi A + \hat{\lambda} E_A^T E_A & A^T \Phi B + \hat{\lambda} E_A^T E_B \\ B^T \Phi A + \hat{\lambda} E_B^T E_A & Q + B^T \Phi B + \hat{\lambda} E_B^T E_B \end{bmatrix}^{-1} \times \begin{bmatrix} A^T \Phi (y - A\bar{x}) + \hat{\lambda} E_A^T (E_y - E_A \bar{x}) \\ B^T \Phi (y - A\bar{x}) + \hat{\lambda} E_B^T (E_y - E_A \bar{x}) \end{bmatrix},$$

where we define $\Phi := \mu I + \mu M\left(\hat{\lambda} I - \mu M^T M\right)^{-1} M^T \mu$. Adding $\begin{bmatrix} \bar{x} \\ 0 \end{bmatrix}$ to both sides of the equation above then gives

$$\begin{bmatrix} \hat{x}^\mu \\ \hat{w}^\mu \end{bmatrix} = \begin{bmatrix} \bar{P} + A^T \Phi A + \hat{\lambda} E_A^T E_A & A^T \Phi B + \hat{\lambda} E_A^T E_B \\ B^T \Phi A + \hat{\lambda} E_B^T E_A & Q + B^T \Phi B + \hat{\lambda} E_B^T E_B \end{bmatrix}^{-1} \begin{bmatrix} \bar{P}\bar{x} + A^T \Phi y + \hat{\lambda} E_A^T E_y \\ B^T \Phi y + \hat{\lambda} E_B^T E_y \end{bmatrix}.$$

This equation also represents a system of simultaneous equations. Hence, we can write it as the following set of equations:

$$\left(\bar{P} + A^T \Phi A + \hat{\lambda} E_A^T E_A\right)\hat{x}^\mu + \left(A^T \Phi B + \hat{\lambda} E_A^T E_B\right)\hat{w}^\mu = \bar{P}\bar{x} + A^T \Phi y + \hat{\lambda} E_A^T E_y, \tag{3.43}$$

$$\left(B^T \Phi A + \hat{\lambda} E_B^T E_A\right)\hat{x}^\mu + \left(Q + B^T \Phi B + \hat{\lambda} E_B^T E_B\right)\hat{w}^\mu = B^T \Phi y + \hat{\lambda} E_B^T E_y. \tag{3.44}$$

Isolating $\hat{w}^\mu$ in (3.44), we have

$$\hat{w}^\mu = \left(Q + B^T \Phi B + \hat{\lambda} E_B^T E_B\right)^{-1}\left(B^T \Phi y + \hat{\lambda} E_B^T E_y - \left(B^T \Phi A + \hat{\lambda} E_B^T E_A\right)\hat{x}^\mu\right).$$

Substituting $\hat{w}^\mu$ back into (3.43) then yields

$$\Big[\bar{P} + A^T \Phi A + \hat{\lambda} E_A^T E_A -$$
$$\left(A^T \Phi B + \hat{\lambda} E_A^T E_B\right)\left(Q + B^T \Phi B + \hat{\lambda} E_B^T E_B\right)^{-1}\left(B^T \Phi A + \hat{\lambda} E_B^T E_A\right)\Big]\hat{x}^\mu =$$
$$\bar{P}\bar{x} + A^T \Phi y + \hat{\lambda} E_A^T E_y - \left(A^T \Phi B + \hat{\lambda} E_A^T E_B\right)\left(Q + B^T \Phi B + \hat{\lambda} E_B^T E_B\right)^{-1}\left(B^T \Phi y + \hat{\lambda} E_B^T E_y\right) \tag{3.45}$$

Expanding the left-hand side of (3.45), we obtain

$$\Big[\bar{P} + A^T\left(\Phi - \Phi B\left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} B^T \Phi\right)A -$$
$$A^T \Phi B\left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1}\hat{\lambda} E_B^T E_A - \hat{\lambda} E_A^T E_B\left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} B^T \Phi A +$$
$$\hat{\lambda} E_A^T E_A - \hat{\lambda} E_A^T E_B\left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} E_B^T E_A \hat{\lambda}\Big]\hat{x}^\mu. \tag{3.46}$$

Applying Lemma A.1, we can simplify the second term of (3.46), as follows:

$$A^T\left(\Phi - \Phi B\left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} B^T \Phi\right)A =$$
$$A^T\Big(\underbrace{\Phi^{-1} + B\left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} B^T}_{\widehat{Q}}\Big)^{-1} A = A^T \widehat{Q}^{-1} A. \tag{3.47}$$

Now, we simplify the third term of (3.46) by applying Lemma A.2 twice:

$$A^T \Phi B\left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1}\hat{\lambda} E_B^T E_A =$$
$$A^T\Big(\underbrace{\Phi^{-1} + B\left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} B^T}_{\widehat{Q}}\Big)^{-1} B Q^{-1} E_B^T \Big(\underbrace{\hat{\lambda}^{-1} I + E_B Q^{-1} E_B^T}_{\bar{Q}}\Big)^{-1} E_A =$$
$$A^T \widehat{Q}^{-1} B Q^{-1} E_B^T \bar{Q}^{-1} E_A. \tag{3.48}$$

Applying the same procedure above for the fourth term of (3.46) yields

$$\hat{\lambda} E_A^T E_B \left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} B^T \Phi A = E_A^T \bar{Q}^{-1} E_B Q^{-1} B^T \widehat{Q}^{-1} A. \tag{3.49}$$

Next, we use Lemma A.1 and Lemma A.2 to expand the last two terms of (3.46), as follows:

$$\begin{aligned}
&\hat{\lambda} E_A^T E_A - \hat{\lambda} E_A^T E_B \left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} E_B^T E_A \hat{\lambda} = \\
&\hat{\lambda} E_A^T E_A - \hat{\lambda} E_A^T E_B \Big[\left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} - \\
&\left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} B^T \Big(\underbrace{\Phi^{-1} + B \left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} B^T}_{\widehat{Q}} \Big)^{-1} B \left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} \Big] E_B^T E_A \hat{\lambda} = \\
&E_A^T \Big(\hat{\lambda} I - \hat{\lambda} E_B \left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} E_B^T \hat{\lambda}\Big) E_A + \\
&E_A^T \Big(\underbrace{\hat{\lambda}^{-1} I + E_B Q^{-1} E_B^T}_{\bar{Q}} \Big)^{-1} E_B Q^{-1} B^T \widehat{Q}^{-1} B Q^{-1} E_B^T \Big(\underbrace{\hat{\lambda}^{-1} I + E_B Q^{-1} E_B^T}_{\bar{Q}} \Big)^{-1} E_A = \\
&E_A^T \Big(\underbrace{\hat{\lambda}^{-1} I + E_B Q^{-1} E_B^T}_{\bar{Q}} \Big)^{-1} E_A + E_A^T \bar{Q}^{-1} E_B Q^{-1} B^T \widehat{Q}^{-1} B Q^{-1} E_B^T \bar{Q}^{-1} E_A = \\
&E_A^T \bar{Q}^{-1} E_A + E_A^T \bar{Q}^{-1} E_B Q^{-1} B^T \widehat{Q}^{-1} B Q^{-1} E_B^T \bar{Q}^{-1} E_A.
\end{aligned} \tag{3.50}$$

Then, substituting (3.47), (3.48), (3.49), and (3.50) back into (3.46) leads to

$$\begin{aligned}
&\Big[\bar{P} + E_A^T \bar{Q}^{-1} E_A + A^T \widehat{Q}^{-1} \Big(\underbrace{A - B Q^{-1} E_B^T \bar{Q}^{-1} E_A}_{\widehat{A}} \Big) - \\
&E_A^T \bar{Q}^{-1} E_B Q^{-1} B^T \widehat{Q}^{-1} \Big(\underbrace{A - B Q^{-1} E_B^T \bar{Q}^{-1} E_A}_{\widehat{A}} \Big) \Big] \hat{x}^{\mu} = \Big[\bar{P} + E_A^T \bar{Q}^{-1} E_A + \\
&\Big(\underbrace{A^T - E_A^T \bar{Q}^{-1} E_B Q^{-1} B^T}_{\widehat{A}^T} \Big) \widehat{Q}^{-1} \widehat{A} \Big] \hat{x}^{\mu} = \left(\bar{P} + \widehat{A}^T \widehat{Q}^{-1} \widehat{A} + E_A^T \bar{Q}^{-1} E_A \right) \hat{x}^{\mu}.
\end{aligned} \tag{3.51}$$

Similarly, we expand the right-hand side of (3.45) to obtain

$$\begin{aligned}
&\bar{P} \bar{x} + A^T \Big(\Phi - \Phi B \left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} B^T \Phi \Big) y - \\
&A^T \Phi B \left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} \hat{\lambda} E_B^T E_y - \hat{\lambda} E_A^T E_B \left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} B^T \Phi y + \\
&\hat{\lambda} E_A^T E_y - \hat{\lambda} E_A^T E_B \left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} E_B^T E_y \hat{\lambda}.
\end{aligned} \tag{3.52}$$

We first simplify the second term of (3.52) using Lemma A.1:

$$\begin{aligned}
&A^T \Big(\Phi - \Phi B \left(Q + \hat{\lambda} E_B^T E_B + B^T \Phi B\right)^{-1} B^T \Phi \Big) y = \\
&\qquad A^T \Big(\underbrace{\Phi^{-1} + B \left(Q + \hat{\lambda} E_B^T E_B\right)^{-1} B^T}_{\widehat{Q}} \Big)^{-1} y = A^T \widehat{Q}^{-1} y.
\end{aligned} \tag{3.53}$$

Then, we apply Lemma A.2 twice to simplify the third term of (3.52):

$$\begin{aligned} &A^T\Phi B\Big(Q+\hat{\lambda}E_B^TE_B+B^T\Phi B\Big)^{-1}\hat{\lambda}E_B^TE_y = \\ &A^T\Big(\underbrace{\Phi^{-1}+B\Big(Q+\hat{\lambda}E_B^TE_B\Big)^{-1}B^T}_{\widehat{Q}}\Big)^{-1}BQ^{-1}E_B^T\Big(\underbrace{\hat{\lambda}^{-1}I+E_BQ^{-1}E_B^T}_{\bar{Q}}\Big)^{-1}E_y = \\ &A^T\widehat{Q}^{-1}BQ^{-1}E_B^T\bar{Q}^{-1}E_y. \end{aligned} \tag{3.54}$$

The same procedure is used to simplify the fourth term of (3.52), such that

$$\hat{\lambda}E_A^TE_B\Big(Q+\hat{\lambda}E_B^TE_B+B^T\Phi B\Big)^{-1}B^T\Phi y = E_A^T\bar{Q}^{-1}E_BQ^{-1}B^T\widehat{Q}^{-1}y. \tag{3.55}$$

Now, we expand the last two terms of (3.52) using Lemma A.1 and Lemma A.2:

$$\begin{aligned} &\hat{\lambda}E_A^TE_y-\hat{\lambda}E_A^TE_B\Big(Q+\hat{\lambda}E_B^TE_B+B^T\Phi B\Big)^{-1}E_B^TE_y\hat{\lambda} = \\ &\hat{\lambda}E_A^TE_y-\hat{\lambda}E_A^TE_B\Big[\Big(Q+\hat{\lambda}E_B^TE_B\Big)^{-1} - \\ &\Big(Q+\hat{\lambda}E_B^TE_B\Big)^{-1}B^T\Big(\underbrace{\Phi^{-1}+B\Big(Q+\hat{\lambda}E_B^TE_B\Big)^{-1}B^T}_{\widehat{Q}}\Big)^{-1}B\Big(Q+\hat{\lambda}E_B^TE_B\Big)^{-1}\Big]E_B^TE_y\hat{\lambda} = \\ &E_A^T\Big(\hat{\lambda}I-\hat{\lambda}E_B\Big(Q+\hat{\lambda}E_B^TE_B\Big)^{-1}E_B^T\hat{\lambda}\Big)E_y + \\ &E_A^T\Big(\underbrace{\hat{\lambda}^{-1}I+E_BQ^{-1}E_B^T}_{\bar{Q}}\Big)^{-1}E_BQ^{-1}B^T\widehat{Q}^{-1}BQ^{-1}E_B^T\Big(\underbrace{\hat{\lambda}^{-1}I+E_BQ^{-1}E_B^T}_{\bar{Q}}\Big)^{-1}E_y = \\ &E_A^T\Big(\underbrace{\hat{\lambda}^{-1}I+E_BQ^{-1}E_B^T}_{\bar{Q}}\Big)^{-1}E_y+E_A^T\bar{Q}^{-1}E_BQ^{-1}B^T\widehat{Q}^{-1}BQ^{-1}E_B^T\bar{Q}^{-1}E_y = \\ &E_A^T\bar{Q}^{-1}E_y+E_A^T\bar{Q}^{-1}E_BQ^{-1}B^T\widehat{Q}^{-1}BQ^{-1}E_B^T\bar{Q}^{-1}E_y. \end{aligned} \tag{3.56}$$

Hence, substituting (3.53), (3.54), (3.55), and (3.56) back into (3.52) gives

$$\begin{aligned} &\bar{P}\bar{x}+E_A^T\bar{Q}^{-1}E_y+\Big(\underbrace{A^T-E_A^T\bar{Q}^{-1}E_BQ^{-1}B^T}_{\widehat{A}^T}\Big)\widehat{Q}^{-1}y - \\ &\Big(\underbrace{A^T-E_A^T\bar{Q}^{-1}E_BQ^{-1}B^T}_{\widehat{A}^T}\Big)\widehat{Q}^{-1}BQ^{-1}E_B^T\bar{Q}^{-1}E_y = \bar{P}\bar{x}+E_A^T\bar{Q}^{-1}E_y + \\ &\widehat{A}^T\widehat{Q}^{-1}\Big(\underbrace{y-BQ^{-1}E_B^T\bar{Q}^{-1}E_y}_{\hat{y}}\Big) = \bar{P}\bar{x}+\widehat{A}^T\widehat{Q}^{-1}\hat{y}+E_A^T\bar{Q}^{-1}E_y. \end{aligned} \tag{3.57}$$

Lastly, we substitute the left- and right-hand sides of (3.45) respectively by (3.51) and (3.57) and isolate $\hat{x}^\mu$ to obtain the estimate in (3.39).

Furthermore, we follow the procedure described in Lemma 2.7 to obtain the auxiliary parameter $\hat{\lambda}$, i.e., by solving the optimization problem (3.41)-(3.42). Note that, according to Remark 2.1, as we search for $\lambda > \mu\|M^TM\|$ in problem (3.41)-(3.42) and $\mu > 0$, the invertibility of Φ is ensured. □

At this point, it is important to analyze how the penalty parameter μ influences the solution presented in Lemma 3.3. As explained in Section 2.1, when applying the penalty function method with $\mu \to \infty$, the solution to the unconstrained problem (3.35)-(3.36) approaches the optimal solution to the original constrained problem (3.34) and we say that they are equivalent. Nevertheless, note that, in the process of obtaining a solution $\hat{x}^\mu$ to (3.35)-(3.36), we are faced with the auxiliary minimization problem (3.41)-(3.42), in which an optimal parameter $\hat{\lambda}$ is sought. Since $\lambda > \mu \|M^T M\|$, if one lets $\mu \to \infty$, we have that $\lambda \to \infty$ and $\Phi(\lambda) \to \infty$. As a consequence, the second and third terms in $\Gamma(\lambda)$ will be excessively penalized compared to the first term, leading to an unbalanced objective function.

Therefore, in this robust estimation context, unlike the suggestion in Ishihara, Terra and Cerri (2015), the penalty parameter μ should rather be understood as a robustness measure of the estimator, taking finite values instead of approaching infinity. In this sense, when the system model is subject to significant uncertainty, smaller values of μ will increase the robustness, which translates into better estimation performance. On the other hand, when we have more confidence in the system model, larger values of μ can be used. In the limit, when there are no uncertainties, the model is exact, meaning that one can let $\mu \to \infty$.

Furthermore, some works, such as Xu and Mannor (2009) and Liu and Zhou (2017), point out that robust least-squares estimators obtained by considering the worst-case influence of the model uncertainties may be over-conservative, which in turn leads to poor estimation performance. Tuning the penalty parameter μ is therefore a possible approach to counteract this effect.

Remark 3.2. *The solution outlined in Lemma 3.3 depends on the solution of the optimization problem* (3.41)-(3.42) *to compute the $\hat{\lambda}$ parameter. A constrained line search method can be used to obtain a solution, however, this requires additional computation time. Nevertheless, as Remark 2.2 points out, a practical and reasonable approximation for $\hat{\lambda}$ is to select $\hat{\lambda} = (1+\xi)\,\mu\,\|M^T M\|$, for some $\xi > 0$.*

To conclude this section, we propose a result equivalent to Lemma 3.2 and associate a weighting matrix $\hat{P}$ to the estimation error $x - \hat{x}^\mu$. Since the underlying model considered in the regularized least-squares estimation problem is subject to parametric uncertainties, unlike the nominal case (Section 3.1.2), we cannot refer to this weighting matrix as an error variance matrix, which, in fact, we cannot compute analytically (SAYED, 2001). Therefore, we rely on the deterministic understanding of the problem and associate the following error weighting matrix, considering the estimate $\hat{x}^\mu$ in (3.39):

$$\hat{P} = \left(\bar{P} + \widehat{A}^T \widehat{Q}^{-1} \widehat{A} + E_A^T \bar{Q}^{-1} E_A\right)^{-1}. \tag{3.58}$$

3.2.3 Robust Kalman Filter

At this point, we are ready to apply the results in Section 3.2.2 and obtain the Robust Kalman Filter. Recall that the deterministic estimation problem (3.31) is a special case of a regularized least-squares estimation problem with norm-bounded uncertainties, considering the following mappings between (3.31) and (3.34):

$$
\begin{aligned}
x \leftarrow \begin{bmatrix} \hat{x}_k \\ \hat{x}_{k+1} \end{bmatrix}, \quad \bar{x} \leftarrow \begin{bmatrix} \hat{x}_{k|k-1} \\ 0 \end{bmatrix}, \quad w \leftarrow \begin{bmatrix} \hat{w}_k \\ \hat{v}_k \end{bmatrix}, \quad \bar{P} \leftarrow \begin{bmatrix} P_{k|k-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad Q \leftarrow \begin{bmatrix} Q_k^{-1} & 0 \\ 0 & R_k^{-1} \end{bmatrix}, \\
y \leftarrow \begin{bmatrix} -G_k u_k \\ y_k \end{bmatrix}, \quad A \leftarrow \begin{bmatrix} F_k & -I_n \\ C_k & 0 \end{bmatrix}, \quad B \leftarrow \begin{bmatrix} H_k & 0 \\ 0 & D_k \end{bmatrix}, \\
\delta y \leftarrow \begin{bmatrix} -\delta G_k u_k \\ 0 \end{bmatrix}, \quad \delta A \leftarrow \begin{bmatrix} \delta F_k & 0 \\ \delta C_k & 0 \end{bmatrix}, \quad \text{and} \quad \delta B \leftarrow \begin{bmatrix} \delta H_k & 0 \\ 0 & \delta D_k \end{bmatrix}.
\end{aligned} \tag{3.59}
$$

Moreover, consider the following mappings between the uncertainty models (3.30) and (3.33):

$$
\begin{aligned}
M \leftarrow \begin{bmatrix} M_{1,k} & 0 \\ 0 & M_{2,k} \end{bmatrix}, \quad \Delta \leftarrow \begin{bmatrix} \Delta_{1,k} & 0 \\ 0 & \Delta_{2,k} \end{bmatrix}, \\
E_y \leftarrow \begin{bmatrix} -E_{G_k} u_k \\ 0 \end{bmatrix}, \quad E_A \leftarrow \begin{bmatrix} E_{F_k} & 0 \\ E_{C_k} & 0 \end{bmatrix}, \quad \text{and} \quad E_B \leftarrow \begin{bmatrix} E_{H_k} & 0 \\ 0 & E_{D_k} \end{bmatrix}.
\end{aligned} \tag{3.60}
$$

Since $P_{k|k-1}^{-1} \succ 0$, we have that $\bar{P} \succeq 0$. In addition, $Q_k^{-1} \succ 0$ and $R_k^{-1} \succ 0$, such that $Q \succ 0$. Therefore, by using the results in Lemma 3.3 and in equation (3.58), we obtain the filtered and predicted robust state estimates, $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, along with their corresponding error weighting matrices $P_{k|k}$ and $P_{k+1|k}$.

Theorem 3.3. *Consider the regularized least-squares estimation problem with norm-bounded uncertainties (3.31) with given initial conditions $\hat{x}_{0|-1}$, $P_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R_k \succ 0$, and fixed parameters $\mu > 0$ and $\xi > 0$. For each $k = 0, 1, \ldots, N$, its solution recursively provides the filtered and predicted robust state estimates of system (3.29)-(3.30), $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, as well as their corresponding error weighting matrices, $P_{k|k}$ and $P_{k+1|k}$, according to the procedure outlined in Algorithm 3.2.*

Proof. Problem (3.31) is a regularized least-squares estimation problem with norm-bounded uncertainties, hence we can apply the result in Lemma 3.3 to obtain the robust system state estimates $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$. Thus, we first substitute the mappings (3.59) and (3.60) into (3.40) to compute the modified system and sensing model matrices

$$
\Phi = \begin{bmatrix} \left(\mu^{-1} I_n - \hat{\lambda}_k^{-1} M_{1,k} M_{1,k}^T\right)^{-1} & 0 \\ 0 & \left(\mu^{-1} I_r - \hat{\lambda}_k^{-1} M_{2,k} M_{2,k}^T\right)^{-1} \end{bmatrix} =: \begin{bmatrix} \Phi_{1,k}^{-1} & 0 \\ 0 & \Phi_{2,k}^{-1} \end{bmatrix},
$$

Algorithm 3.2 Robust Kalman Filter (RKF)

Model: Assume the uncertain system model in (3.29)-(3.30).

Initialization: Set $\hat{x}_{0|-1}$, $P_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R_k \succ 0$, $\mu > 0$, and $\xi > 0$.

for $k = 0, 1, \ldots, N$ **do**

1. Obtain a measurement y_k.
2. Compute $\hat{\lambda}_k$ using the approximation:

$$\hat{\lambda}_k = (1+\xi)\,\mu\,\left\|\mathbf{diag}\left(M_{1,k}^T M_{1,k},\; M_{2,k}^T M_{2,k}\right)\right\|$$

3. Compute the modified system and sensing model matrices:

$$\begin{aligned}
\Phi_{1,k} &= \mu^{-1} I_n - \hat{\lambda}_k^{-1} M_{1,k} M_{1,k}^T & \widehat{Q}_k &= \Phi_{1,k} + H_k\left(Q_k^{-1} + \hat{\lambda}_k E_{H_k}^T E_{H_k}\right)^{-1} H_k^T \\
\Phi_{2,k} &= \mu^{-1} I_r - \hat{\lambda}_k^{-1} M_{2,k} M_{2,k}^T & \widehat{R}_k &= \Phi_{2,k} + D_k\left(R_k^{-1} + \hat{\lambda}_k E_{D_k}^T E_{D_k}\right)^{-1} D_k^T \\
\bar{Q}_k &= \hat{\lambda}_k^{-1} I_{t_1} + E_{H_k} Q_k E_{H_k}^T & \bar{R}_k &= \hat{\lambda}_k^{-1} I_{t_2} + E_{D_k} R_k E_{D_k}^T \\
\widehat{F}_k &= F_k - H_k Q_k E_{H_k}^T \bar{Q}_k^{-1} E_{F_k} & \widehat{C}_k &= C_k - D_k R_k E_{D_k}^T \bar{R}_k^{-1} E_{C_k} \\
\widehat{G}_k &= G_k - H_k Q_k E_{H_k}^T \bar{Q}_k^{-1} E_{G_k} & &
\end{aligned}$$

4. Correction step:
   4.1. Compute the posterior error weighting matrix:

$$P_{k|k} = \left(P_{k|k-1}^{-1} + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + E_{C_k}^T \bar{R}_k^{-1} E_{C_k} + E_{F_k}^T \bar{Q}_k^{-1} E_{F_k}\right)^{-1}$$

   4.2. Compute the filtered robust state estimate:

$$\hat{x}_{k|k} = P_{k|k}\left(P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - E_{F_k}^T \bar{R}_k^{-1} E_{G_k} u_k\right)$$

5. Prediction step:
   5.1. Update the predicted prior error weighting matrix:

$$P_{k+1|k} = \widehat{F}_k P_{k|k} \widehat{F}_k^T + \widehat{Q}_k$$

   5.2. Update the predicted prior robust state estimate:

$$\hat{x}_{k+1|k} = \widehat{F}_k \hat{x}_{k|k} + \widehat{G}_k u_k$$

end for

$$\begin{aligned}
\bar{Q} &= \begin{bmatrix} \hat{\lambda}_k^{-1} I_{t_1} + E_{H_k} Q_k E_{H_k}^T & 0 \\ 0 & \hat{\lambda}_k^{-1} I_{t_2} + E_{D_k} R_k E_{D_k}^T \end{bmatrix} =: \begin{bmatrix} \bar{Q}_k & 0 \\ 0 & \bar{R}_k \end{bmatrix}, \\
\widehat{Q} &= \begin{bmatrix} \Phi_{1,k} + H_k\left(Q_k^{-1} + \hat{\lambda}_k E_{H_k}^T E_{H_k}\right)^{-1} H_k^T & 0 \\ 0 & \Phi_{2,k} + D_k\left(R_k^{-1} + \hat{\lambda}_k E_{D_k}^T E_{D_k}\right)^{-1} D_k^T \end{bmatrix} =: \begin{bmatrix} \widehat{Q}_k & 0 \\ 0 & \widehat{R}_k \end{bmatrix}, \\
\widehat{A} &= \begin{bmatrix} F_k - H_k Q_k E_{H_k}^T \bar{Q}_k^{-1} E_{F_k} & -I_n \\ C_k - D_k R_k E_{D_k}^T \bar{R}_k^{-1} E_{C_k} & 0 \end{bmatrix} =: \begin{bmatrix} \widehat{F}_k & -I_n \\ \widehat{C}_k & 0 \end{bmatrix}, \\
\hat{y} &= \begin{bmatrix} -\left(G_k - H_k Q_k E_{H_k}^T \bar{Q}_k^{-1} E_{G_k}\right) u_k \\ y_k \end{bmatrix} =: \begin{bmatrix} -\widehat{G}_k u_k \\ y_k \end{bmatrix}.
\end{aligned} \tag{3.61}$$

Moreover, to compute the $\hat{\lambda}_k$ parameter, we consider the practical approximation discussed in Remark 3.2, such that

$$\hat{\lambda}_k = (1+\xi)\,\mu \left\| \mathbf{diag}\left(M_{1,k}^T M_{1,k},\; M_{2,k}^T M_{2,k}\right) \right\|,$$

for some $\xi > 0$. Now, we substitute the mappings (3.59) and (3.60), as well as the modified matrices (3.61) into the solution (3.39) to obtain

$$\begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \end{bmatrix} = \begin{bmatrix} P_{k|k-1}^{-1} + \widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{F}_k + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + E_{F_k}^T \bar{Q}_k^{-1} E_{F_k} + E_{C_k}^T \bar{R}_k^{-1} E_{C_k} & -\widehat{F}_k^T \widehat{Q}_k^{-1} \\ -\widehat{Q}_k^{-1} \widehat{F}_k & \widehat{Q}_k^{-1} \end{bmatrix}^{-1} \times$$
$$\begin{bmatrix} P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - \left(\widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{G}_k + E_{F_k}^T \bar{Q}_k^{-1} E_{G_k} \right) u_k \\ \widehat{Q}_k^{-1} \widehat{G}_k u_k \end{bmatrix}. \tag{3.62}$$

Note that (3.62) also represents a system of simultaneous equations. Therefore, we can write it as the following set of equations:

$$\left(P_{k|k-1}^{-1} + \widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{F}_k + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + E_{F_k}^T \bar{Q}_k^{-1} E_{F_k} + E_{C_k}^T \bar{R}_k^{-1} E_{C_k} \right) \hat{x}_{k|k} - \widehat{F}_k^T \widehat{Q}_k^{-1} \hat{x}_{k+1|k} =$$
$$P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - \left(\widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{G}_k + E_{F_k}^T \bar{Q}_k^{-1} E_{G_k} \right) u_k, \tag{3.63}$$
$$-\widehat{Q}_k^{-1} \widehat{F}_k \hat{x}_{k|k} + \widehat{Q}_k^{-1} \hat{x}_{k+1|k} = \widehat{Q}_k^{-1} \widehat{G}_k u_k. \tag{3.64}$$

Isolating $\hat{x}_{k+1|k}$ in (3.64) yields

$$\hat{x}_{k+1|k} = \widehat{F}_k \hat{x}_{k|k} + \widehat{G}_k u_k, \tag{3.65}$$

which is the update equation of the predicted prior robust state estimate in step 5.2 of Algorithm 3.2. Then, substituting $\hat{x}_{k+1|k}$ back into (3.63) and isolating $\hat{x}_{k|k}$ gives

$$\hat{x}_{k|k} = \left(P_{k|k-1}^{-1} + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + E_{C_k}^T \bar{R}_k^{-1} E_{C_k} + E_{F_k}^T \bar{Q}_k^{-1} E_{F_k} \right)^{-1} \times$$
$$\left(P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - E_{F_k}^T \bar{Q}_k^{-1} E_{G_k} u_k \right),$$

which corresponds to the equation for computing the filtered robust state estimate in step 4.2 of Algorithm 3.2.

Now, to obtain the error weighting matrices associated with $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, we use equation (3.58), assuming a deterministic context. Thus, substituting the mappings (3.59) and (3.60), and the modified matrices (3.61) into (3.58) yields[2]

$$\begin{bmatrix} P_{k|k} & * \\ * & P_{k+1|k} \end{bmatrix} =$$
$$\underbrace{\begin{bmatrix} P_{k|k-1}^{-1} + \widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{F}_k + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + E_{F_k}^T \bar{Q}_k^{-1} E_{F_k} + E_{C_k}^T \bar{R}_k^{-1} E_{C_k} & -\widehat{F}_k^T \widehat{Q}_k^{-1} \\ -\widehat{Q}_k^{-1} \widehat{F}_k & \widehat{Q}_k^{-1} \end{bmatrix}}_{\mathcal{M}^{-1}}^{-1} =: \begin{bmatrix} \mathcal{M}_1 & \mathcal{M}_2 \\ \mathcal{M}_2^T & \mathcal{M}_3 \end{bmatrix}^{-1},$$

2 The elements marked with $*$ are byproducts with no particular meaning in our context.

where we define the partitioned matrix $\mathcal{M}$. To find its inverse, we use the Banachiewicz inversion formula (Lemma A.4, item **(ii)**). According to Lemma A.3, the Schur complement of $\mathcal{M}_3$ in $\mathcal{M}$ is

$$(\mathcal{M}/\mathcal{M}_3) = \mathcal{M}_1 - \mathcal{M}_2\mathcal{M}_3^{-1}\mathcal{M}_2^T = P_{k|k-1}^{-1} + \widehat{C}_k^T\widehat{R}_k^{-1}\widehat{C}_k + E_{C_k}^T\bar{R}_k^{-1}E_{C_k} + E_{F_k}^T\bar{Q}_k^{-1}E_{F_k}.$$

The posterior error weighting matrix in step 4.1 of Algorithm 3.2 is then obtained as follows:

$$P_{k|k} = (\mathcal{M}/\mathcal{M}_3)^{-1} = \left(P_{k|k-1}^{-1} + \widehat{C}_k^T\widehat{R}_k^{-1}\widehat{C}_k + E_{C_k}^T\bar{R}_k^{-1}E_{C_k} + E_{F_k}^T\bar{Q}_k^{-1}E_{F_k}\right)^{-1}.$$

Finally, we obtain the predicted prior error weighting matrix

$$P_{k+1|k} = \mathcal{M}_3^{-1} + \mathcal{M}_3^{-1}\mathcal{M}_2^T(\mathcal{M}/\mathcal{M}_3)^{-1}\mathcal{M}_2\mathcal{M}_3^{-1} = \widehat{F}_kP_{k|k}\widehat{F}_k^T + \widehat{Q}_k,$$

as shown in step 5.1 of Algorithm 3.2. □

Notice that in Algorithm 3.2, we consider that the penalty parameter μ assumes a finite value, which, based on the discussion in Section 3.2.2, we can tune to increase the filter performance in terms of smaller estimation error. The filter also depends on the ξ parameter, used to approximate $\hat{\lambda}_k$, as Remark 3.2 states. In most cases, choosing a small value for ξ within the interval $(0, 1)$ generally leads to adequate results.

Remark 3.3. *The expressions for the Robust Kalman Filter outlined in Algorithm 3.2 resemble those of the Nominal Kalman Filter, as shown in Algorithm 3.1. In fact, if there are no uncertainties, i.e.,* $M_{1,k}$, $M_{2,k}$, E_{F_k}, E_{G_k}, E_{H_k}, E_{C_k}, *and* E_{D_k} *are all zero, and we let* $\mu \to \infty$, *we have that* $\widehat{Q}_k = H_kQ_kH_k^T$, $\widehat{R}_k = D_kR_kD_k^T$, $\widehat{F}_k = F_k$, $\widehat{G}_k = G_k$, *and* $\widehat{C}_k = C_k$. *This way, the expressions in steps 4 and 5 of Algorithm 3.2 collapse to the expressions in steps 3 and 4 of Algorithm 3.1. For this reason, we say that the proposed estimator is a* robust *Kalman filter.*

To conclude this section, we emphasize the importance of using the penalty function method when deriving the proposed robust filter. This strategy allowed for uniformly considering parametric uncertainties in all system matrices, as well as provided a parameter μ that can be conveniently used to adjust the filter estimation performance when necessary. Furthermore, much like the standard Kalman filter, the robust filter recursive expressions outlined in Algorithm 3.2 can be easily implemented in online applications.

3.2.4 Stability Analysis

In this section, we investigate the stability properties of the proposed Robust Kalman Filter, as well as the boundedness of its estimation error variance. Based on the procedure described in Section 3.1.4 and Sayed (2001), we examine the steady-state behavior of Algorithm 3.2 when the system model parameters are constant and there is no input u_k. Nonetheless, we still assume that the contraction matrices $\Delta_{1,k}$ and $\Delta_{2,k}$ are time-varying. Thus, consider the following discrete-time state-space description of an uncertain linear system:

$$\begin{aligned} x_{k+1} &= (F + \delta F_k)x_k + (H + \delta H_k)w_k, \\ y_k &= (C + \delta C_k)x_k + (D + \delta D_k)v_k, \end{aligned} \tag{3.66}$$

for $k \geq 0$, with time-varying norm-bounded parametric uncertainties

$$\begin{aligned} \begin{bmatrix} \delta F_k & \delta H_k \end{bmatrix} &= M_1 \Delta_{1,k} \begin{bmatrix} E_F & E_H \end{bmatrix}, \quad \|\Delta_{1,k}\| \leq 1, \\ \begin{bmatrix} \delta C_k & \delta D_k \end{bmatrix} &= M_2 \Delta_{2,k} \begin{bmatrix} E_C & E_D \end{bmatrix}, \quad \|\Delta_{2,k}\| \leq 1. \end{aligned} \tag{3.67}$$

Let us first study the stability of the RKF in Algorithm 3.2. Considering the uncertain system model (3.66)-(3.67), the robust filter equations in steps 4 and 5 of Algorithm 3.2 become:

$$P_{k|k} = \left(P_{k|k-1}^{-1} + \widehat{C}^T \widehat{R}^{-1} \widehat{C} + E_C^T \bar{R}^{-1} E_C + E_F^T \bar{Q}^{-1} E_F\right)^{-1}, \tag{3.68}$$

$$\hat{x}_{k|k} = P_{k|k}\left(P_{k|k-1}^{-1}\hat{x}_{k|k-1} + \widehat{C}^T \widehat{R}^{-1} y_k\right), \tag{3.69}$$

$$P_{k+1|k} = \widehat{F} P_{k|k} \widehat{F}^T + \widehat{Q}, \tag{3.70}$$

$$\hat{x}_{k+1|k} = \widehat{F}\hat{x}_{k|k}, \tag{3.71}$$

where the modified system and sensing model parameter matrices are given by the corresponding equations listed in step 3 of Algorithm 3.2, considering constant parameters. The constant $\hat{\lambda}$ parameter is analogously computed as in step 2. To simplify the analysis, we further define the augmented matrices

$$\widetilde{C} := \begin{bmatrix} \widehat{C} \\ E_C \\ E_F \end{bmatrix} \quad \text{and} \quad \widetilde{R} := \begin{bmatrix} \widehat{R} & 0 & 0 \\ 0 & \bar{R} & 0 \\ 0 & 0 & \bar{Q} \end{bmatrix},$$

such that $P_{k|k}$ in (3.68) can be written in a more compact way, as

$$P_{k|k} = \left(P_{k|k-1}^{-1} + \widetilde{C}^T \widetilde{R}^{-1} \widetilde{C}\right)^{-1}.$$

Applying Lemma A.1 to expand this expression, we obtain

$$P_{k|k} = P_{k|k-1} - P_{k|k-1}\widetilde{C}^T\left(\widetilde{R} + \widetilde{C}P_{k|k-1}\widetilde{C}^T\right)^{-1}\widetilde{C}P_{k|k-1}. \tag{3.72}$$

Now, combining (3.72) with (3.69) and substituting in (3.71) yields the steady-state predicted robust state estimate

$$\hat{x}_{k+1|k} = \widetilde{F}_k \hat{x}_{k|k-1} + \widetilde{F}_k P_{k|k-1} \widehat{C}^T \widehat{R}^{-1} y_k, \tag{3.73}$$

where

$$\widetilde{F}_k = \widehat{F}\Big(I_n - P_{k|k-1}\widetilde{C}^T\big(\widetilde{R} + \widetilde{C}P_{k|k-1}\widetilde{C}^T\big)^{-1}\widetilde{C}\Big)$$

is the filter closed-loop matrix. Moreover, substituting $P_{k|k}$ from (3.72) into (3.70), we obtain the expression for the predicted prior error weighting matrix:

$$P_{k+1|k} = \widehat{F}\Big(P_{k|k-1} - P_{k|k-1}\widetilde{C}^T\big(\widetilde{R} + \widetilde{C}P_{k|k-1}\widetilde{C}^T\big)^{-1}\widetilde{C}P_{k|k-1}\Big)\widehat{F}^T + \widehat{Q}. \tag{3.74}$$

The next theorem establishes a result concerning the convergence of the proposed robust filter to a stable steady-state filter.

Theorem 3.4. *Consider the linear system model* (3.66) *with norm-bounded uncertainties* (3.67) *and the corresponding robust filter* (3.73)-(3.74). *Assume that* $\{\widehat{F}, \widetilde{C}\}$ *is* detectable *and* $\{\widehat{F}, \widehat{Q}^{1/2}\}$ *is* controllable. *Then, for any initial condition* $P_{0|-1} \succ 0$, $\xi > 0$, *and* $\mu > 0$, $P_{k+1|k}$ *converges to the* unique stabilizing solution $P \succ 0$ *of the algebraic Riccati equation*

$$P = \widehat{F}\Big(P - P\widetilde{C}^T\big(\widetilde{R} + \widetilde{C}P\widetilde{C}^T\big)^{-1}\widetilde{C}P\Big)\widehat{F}^T + \widehat{Q}. \tag{3.75}$$

The solution P *is stabilizing in the sense that the steady-state filter closed-loop matrix*

$$\widetilde{F} = \widehat{F}\Big(I_n - P\widetilde{C}^T\big(\widetilde{R} + \widetilde{C}P\widetilde{C}^T\big)^{-1}\widetilde{C}\Big) \tag{3.76}$$

is Schur stable.

Proof. The conditions $\xi > 0$ and $\mu > 0$ imply that $\hat{\lambda} > 0$, ensuring that matrices $\widehat{F}$, $\widetilde{C}$, $\widetilde{R}$, and $\widehat{Q}$ are well-defined. According to Kailath, Sayed and Hassibi (2000b), detectability of $\{\widehat{F}, \widetilde{C}\}$ and controllability of $\{\widehat{F}, \widehat{Q}^{1/2}\}$ ensure the convergence of $P_{k+1|k}$ in (3.74) to the unique stabilizing positive definite solution P of the algebraic Riccati equation (3.75) that stabilizes (3.76), which is the robust filter steady-state closed-loop matrix. □

We now investigate the robust filter estimation error variance. Again, consider the uncertain linear discrete-time system model (3.66)-(3.67). Moreover, assume that w_k and v_k are uncorrelated zero-mean Gaussian noise processes with joint covariance matrix

$$\mathcal{Q} = \boldsymbol{E}\left\{\begin{bmatrix} w_k \\ v_k \end{bmatrix}\begin{bmatrix} w_k^T & v_k^T \end{bmatrix}\right\} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \succ 0. \tag{3.77}$$

Additionally, assume that there is no correlation between the parametric uncertainties and the system and measurement noises.

Definition 3.1. *(XIE; SOH; SOUZA, 1994) The uncertain system* (3.66)-(3.67) *is said to be* quadratically stable *if there exists a symmetric positive definite matrix* U *such that*

$$(F + M_1\Delta_{1,k}E_F)^T U (F + M_1\Delta_{1,k}E_F) - U \prec 0$$

for all admissible contractions $\Delta_{1,k}$.

Remark 3.4. *(PETERSEN; MCFARLANE, 1996) Conversely, the uncertain system (3.66)-(3.67) is quadratically stable if, and only if*

1. *F is Schur stable, i.e., all its eigenvalues lie inside the open unit disk;*
2. *The discrete-time $\mathcal{H}_\infty$ normal bound $\left\| E_F(zI_n - F)^{-1}M_1 \right\|_\infty < 1$[3] is satisfied.*

In order to show that the proposed robust filter presents a bounded steady-state estimation error variance, we make the following assumptions about the uncertain system and the filter itself.

Assumption 3.1. *The uncertain system (3.66)-(3.67) is quadratically stable, according to Definition 3.1.*

Assumption 3.2. *The conditions outlined in Theorem 3.4 are satisfied, such that the robust filter steady-state closed-loop matrix $\widetilde{F}$ is Schur stable.*

Under Assumption 3.1 and Assumption 3.2, we can show that the steady-state robust filter (3.73) is also quadratically stable. For a more compact notation, we define the so-called steady-state filter gain

$$\widetilde{K} := \widetilde{F} P \widehat{C}^T \widehat{R}^{-1},$$

where $\widetilde{F}$ is given by (3.76), with P being the stabilizing solution of the algebraic Riccati equation (3.75). This way, the steady-state robust filter equation can be rewritten as

$$\hat{x}_{k+1|k} = \widetilde{F}\hat{x}_{k|k-1} + \widetilde{K} y_k. \tag{3.78}$$

Now, substituting y_k from (3.66) into (3.78) yields

$$\hat{x}_{k+1|k} = \widetilde{F}\hat{x}_{k|k-1} + \widetilde{K}(C + \delta C_k)x_k + \widetilde{K}(D + \delta D_k)v_k. \tag{3.79}$$

In addition, we introduce the state estimation error vector $e_k := x_k - \hat{x}_{k|k-1}$. Then, subtracting (3.79) from x_{k+1} in (3.66) gives

$$e_{k+1} = \left[(F - \widetilde{F} - \widetilde{K}C) + (\delta F_k - \widetilde{K}\delta C_k)\right]x_k + \widetilde{F}e_k + (H + \delta H_k)w_k - \widetilde{K}(D + \delta D_k)v_k. \tag{3.80}$$

Consider now an augmented system composed of the target system state x_k and the estimation error e_k. Therefore, from (3.66), (3.67), and (3.80), this augmented system is described by

$$\begin{aligned} \zeta_{k+1} &= (\mathscr{F} + \delta\mathscr{F}_k)\zeta_k + (\mathscr{H} + \delta\mathscr{H}_k)\eta_k, \\ \begin{bmatrix} \delta\mathscr{F}_k & \delta\mathscr{H}_k \end{bmatrix} &= \mathscr{M}\,\Delta_k \begin{bmatrix} E_{\mathscr{F}} & E_{\mathscr{H}} \end{bmatrix}, \end{aligned} \tag{3.81}$$

[3] $\|\cdot\|_\infty$ denotes the maximum singular value of its argument for values of z on the unit disk.

where

$$\zeta_k := \begin{bmatrix} x_k \\ e_k \end{bmatrix}, \quad \eta_k := \begin{bmatrix} w_k \\ v_k \end{bmatrix}, \quad \mathscr{F} := \begin{bmatrix} F & 0 \\ F - \widetilde{F} - \widetilde{K}C & \widetilde{F} \end{bmatrix}, \quad \mathscr{H} := \begin{bmatrix} H & 0 \\ H & -\widetilde{K}D \end{bmatrix},$$

$$\mathscr{M} := \begin{bmatrix} M_1 & 0 \\ M_1 & -\widetilde{K}M_2 \end{bmatrix}, \quad \Delta_k := \begin{bmatrix} \Delta_{1,k} & 0 \\ 0 & \Delta_{2,k} \end{bmatrix}, \quad E_{\mathscr{F}} := \begin{bmatrix} E_F & 0 \\ E_C & 0 \end{bmatrix}, \quad E_{\mathscr{H}} := \begin{bmatrix} E_H & 0 \\ 0 & E_D \end{bmatrix}.$$

Lemma 3.4. *Given that Assumption 3.1 and Assumption 3.2 are satisfied, the augmented system (3.81) is quadratically stable.*

Proof. Observe that the augmented system matrix $\mathscr{F}$ is lower triangular with diagonal elements F and $\widetilde{F}$, which are both Schur stable. Therefore, $\mathscr{F}$ is also Schur stable. Moreover, we have that

$$\begin{aligned} E_{\mathscr{F}}(zI_{2n} - \mathscr{F})^{-1}\mathscr{M} &= \begin{bmatrix} E_F & 0 \\ E_C & 0 \end{bmatrix} \begin{bmatrix} zI_n - F & 0 \\ -(F - \widetilde{F} - \widetilde{K}C) & zI_n - \widetilde{F} \end{bmatrix}^{-1} \begin{bmatrix} M_1 & 0 \\ M_1 & -\widetilde{K}M_2 \end{bmatrix} \\ &= \begin{bmatrix} E_F(zI_n - F)^{-1}M_1 & 0 \\ E_C(zI_n - F)^{-1}M_1 & 0 \end{bmatrix} = \begin{bmatrix} E_F \\ E_C \end{bmatrix} (zI_n - F)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix}. \end{aligned}$$

In addition, note that

$$F + M_1\Delta_{1,k}E_F = F + \begin{bmatrix} M_1 & 0 \end{bmatrix} \begin{bmatrix} \Delta_{1,k} & 0 \\ 0 & \Delta_{2,k} \end{bmatrix} \begin{bmatrix} E_F \\ E_C \end{bmatrix}.$$

Since system (3.66)-(3.67) is quadratically stable, according to Remark 3.4, we have

$$\left\| \begin{bmatrix} E_F \\ E_C \end{bmatrix} (zI_n - F)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix} \right\|_\infty < 1,$$

for all admissible contractions $\Delta_{1,k}$ and $\Delta_{2,k}$. Therefore, $\left\| E_{\mathscr{F}}(zI_{2n} - \mathscr{F})^{-1}\mathscr{M} \right\|_\infty < 1$ and the augmented system (3.81) is also quadratically stable. □

We now define the covariance matrix of the augmented system state as $\mathscr{P}_k := \boldsymbol{E}\left\{\zeta_k\zeta_k^T\right\}$. Then, it follows from (3.81) that $\mathscr{P}_k$ satisfies the Lyapunov recursion

$$\mathscr{P}_{k+1} = (\mathscr{F} + \delta\mathscr{F}_k)\mathscr{P}_k(\mathscr{F} + \delta\mathscr{F}_k)^T + (\mathscr{H} + \delta\mathscr{H}_k)\mathscr{Q}(\mathscr{H} + \delta\mathscr{H}_k)^T, \tag{3.82}$$

where $\mathscr{Q}$ is defined in (3.77). The next theorem provides a result on the boundedness of the steady-state estimation error variance of the proposed robust filter.

Theorem 3.5. *Under Assumption 3.1 and Assumption 3.2, the state estimation error variance of the steady-state robust filter (3.78) satisfies*

$$\lim_{k\to\infty} \boldsymbol{E}\left\{e_k e_k^T\right\} \preceq \mathscr{V}_{22},$$

where $\mathcal{V}_{22}$ *is the* $(2,2)$ *block entry with the smallest trace among all* $(2,2)$ *block entries of matrices* $\mathcal{V} \succ 0$ *that satisfy the inequality*

$$(\mathcal{F} + \mathcal{M}\Delta E_{\mathcal{F}})\,\mathcal{V}\,(\mathcal{F} + \mathcal{M}\Delta E_{\mathcal{F}})^T + (\mathcal{H} + \mathcal{M}\Delta E_{\mathcal{H}})\,\mathcal{Q}\,(\mathcal{H} + \mathcal{M}\Delta E_{\mathcal{H}})^T - \mathcal{V} \preceq 0,$$

for all admissible contraction matrices Δ, *with* $\|\Delta\| \leq 1$.

Proof. From Lemma 3.4, the augmented system (3.81) is quadratically stable, therefore, according to Definition 3.1, there exists a matrix $\mathcal{U} \succ 0$ such that

$$(\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})\,\mathcal{U}\,(\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})^T - \mathcal{U} \prec 0,$$

for any admissible contraction matrix Δ_k. Based on the arguments developed in Petersen and McFarlane (1996) and Sayed (2001), the existence of matrix $\mathcal{U} \succ 0$ above guarantees the existence of a sufficiently large scaling parameter $\epsilon > 0$, such that we can find a matrix $\mathcal{V} = \epsilon\,\mathcal{U}$ that satisfies

$$(\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})\,\mathcal{V}\,(\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})^T + (\mathcal{H} + \mathcal{M}\Delta_k E_{\mathcal{H}})\,\mathcal{Q}\,(\mathcal{H} + \mathcal{M}\Delta_k E_{\mathcal{H}})^T \preceq \mathcal{V}.$$

Thus, subtracting the recursion for the augmented system covariance (3.82) from the inequality above yields

$$(\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})(\mathcal{V} - \mathcal{P}_k)(\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})^T \preceq \mathcal{V} - \mathcal{P}_{k+1},$$

or, equivalently,

$$\mathcal{V} - \mathcal{P}_{k+1} = (\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})(\mathcal{V} - \mathcal{P}_k)(\mathcal{F} + \mathcal{M}\Delta_k E_{\mathcal{F}})^T + \mathcal{W}_k,$$

for some $\mathcal{W}_k \succeq 0$. Finally, since the augmented system is quadratically stable, as $k \to \infty$, we have that $\mathcal{V} - \mathcal{P}_{k+1} \succeq 0$, or $\mathcal{P}_{k+1} \preceq \mathcal{V}$. The $(2,2)$ block entry of $\mathcal{P}_k$ corresponds to the estimation error variance, which is thus bounded. □

3.2.5 Illustrative Example

In this section, we assess the performance of the proposed Robust Kalman Filter with a numerical example. We further compare our results with other existing robust filtering strategies from the literature, as well as with the Nominal Kalman filter.

Consider a discrete-time linear system with norm-bounded uncertainties, as described in (3.29)-(3.30) with the following constant parameter matrices (adapted from Xie, Soh and Souza (1994)):

$$F_k = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix}, \quad G_k = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad H_k = \begin{bmatrix} -6 \\ 1 \end{bmatrix}, \quad C_k = \begin{bmatrix} -100 & 10 \end{bmatrix}, \quad D_k = 1,$$

$$M_{1,k} = \begin{bmatrix} 0 \\ 10 \end{bmatrix}, \quad M_{2,k} = 10, \quad E_{F_k} = E_{C_k} = \begin{bmatrix} 0.01 & 0.03 \end{bmatrix}, \quad E_{G_k} = 0, \quad E_{H_k} = E_{D_k} = 0.01.$$

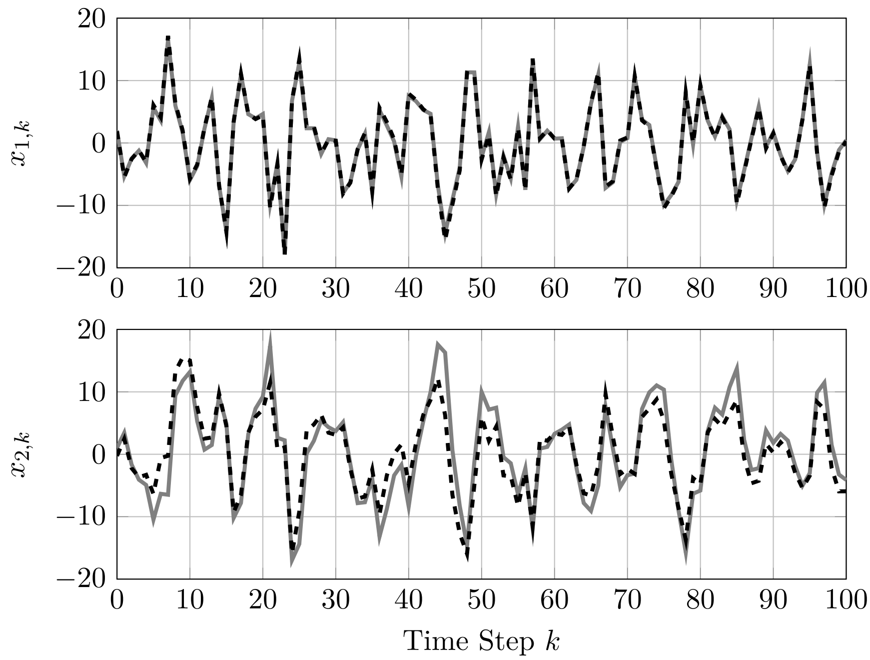
No input signal u_k is present and the system and measurement noises, w_k and v_k, are mutually independent zero-mean white Gaussian signals with variances $Q_k = 1$ and $R_k = 1$, respectively. The initial state is $x_0 = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$.

Then, we apply Algorithm 3.2 with the following initialization data:

$$\hat{x}_{0|-1} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \quad P_{0|-1} = I_2, \quad \mu = 1, \text{ and } \xi = 0.1.$$

Figure 2 shows the evolution of the actual system state along with the estimation performed by the proposed Robust Kalman Filter. At each time step, $\Delta_{1,k}$ and $\Delta_{2,k}$ are real numbers randomly chosen from a uniform distribution with interval $[-1, 1]$. The results show that the proposed RKF can successfully track the state of the target system, despite the norm-bounded parametric uncertainties, present in all matrices of the target system and sensing models.

Figure 2 – Actual (solid lines) and estimated (dashed lines) target system state obtained with the proposed RKF (Algorithm 3.2).



We further evaluate the proposed RKF by comparing its performance with that of some other existing robust filtering strategies. Namely, the optimal robust filter of Ishihara, Terra and Cerri (2015), the robust regularized bounded data uncertainties filter of Sayed (2001), the robust guaranteed cost filter proposed in Dong and You (2006), the robust risk-sensitive Kalman filter presented in Zorzi (2017), and the LMI-based robust Kalman filter of Abolhasani and Rahmani (2018). All filters also assume uncertainties in all parameter matrices, except for the one by Sayed (2001), which only takes into account the uncertainties in the target system model. Furthermore, we also compare the RKF with the Nominal Kalman Filter outlined in Algorithm 3.1.

The simulation consists of performing $M = 5000$ Monte Carlo experiments, each with time horizon $N = 1000$. At each time step k, we compute the mean squared estimation error (MSE), averaged over all experiments, as follows:

$$\mathrm{MSE}_k = \frac{1}{M}\sum_{e=1}^{M} \|x_k - \hat{x}_{k|k,e}\|^2.$$

Since one cannot analytically compute the actual estimation error variance due to the model uncertainties, we use this ensemble average as a reasonable approximation, as suggested in Sayed (2001).

The results are depicted in Figure 3 and summarized in Table 1, which reports the estimation performance of each simulated filter by listing the mean $\overline{\mathrm{MSE}}$ and standard deviation $\sigma(\mathrm{MSE})$ of their error variances, respectively computed as

$$\overline{\mathrm{MSE}} = \sum_{k=0}^{N} \frac{\mathrm{MSE}_k}{N+1} \quad \text{and} \quad \sigma^2(\mathrm{MSE}) = \sum_{k=0}^{N} \frac{(\mathrm{MSE}_k - \overline{\mathrm{MSE}})^2}{N+1},$$

as well as the average time each iteration takes to be executed, Δt_{iter}. The simulation was performed on a 2.3 GHz i7-12700H CPU with 32 GB of RAM using MATLAB R2022b, the YALMIP toolbox (LÖFBERG, 2004), and the SeDuMi solver (STURM, 1999).

Figure 3 – Estimation error variance curves of the robust filters.

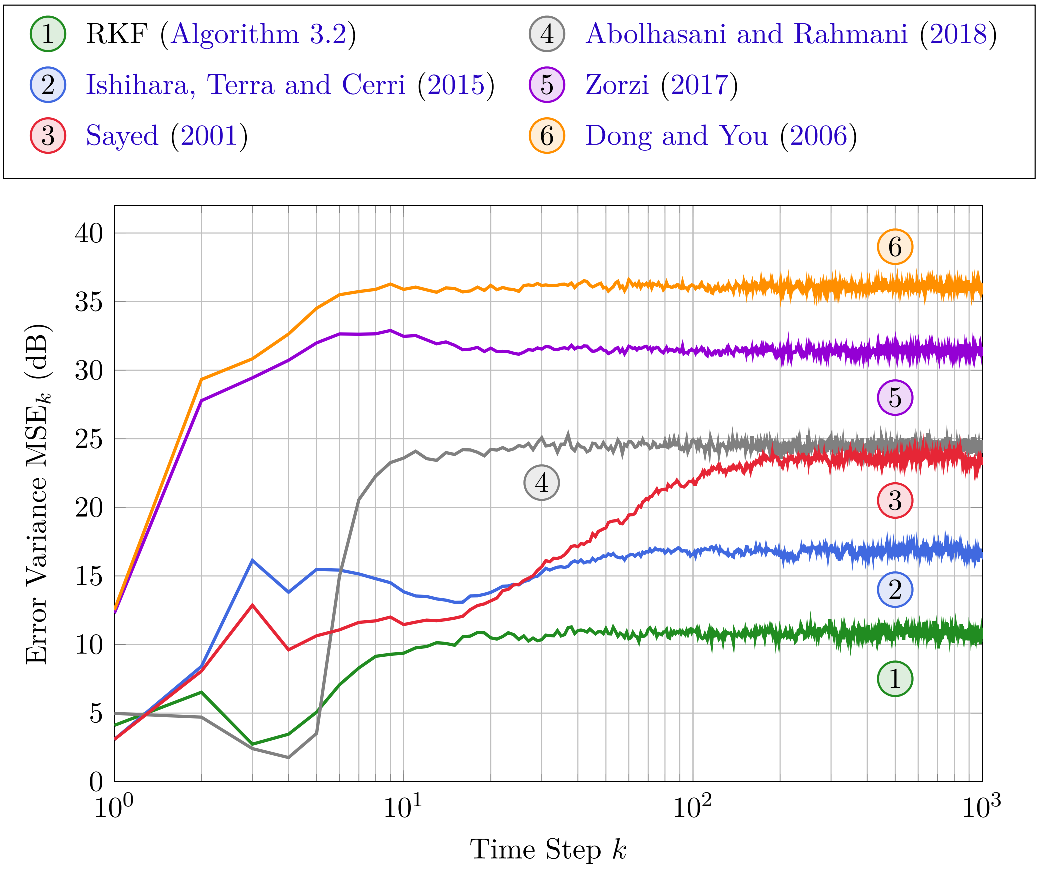


Table 1 – Estimation performance of each robust filter.

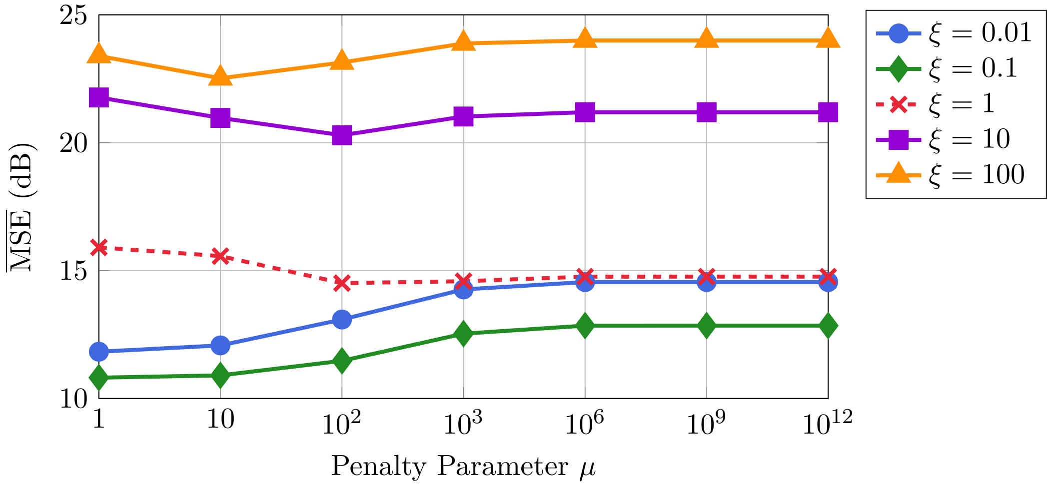
Filter	$\overline{\text{MSE}}$ (dB)	σ(MSE) (dB)	Δt_{iter} (ms)
① RKF (Algorithm 3.2)	**10.79**	**0.5694**	0.0325
② Ishihara, Terra and Cerri (2015)	16.70	0.7245	0.0512
③ Sayed (2001)	22.93	2.2207	0.0099
④ Abolhasani and Rahmani (2018)	24.34	1.5437	68.729
⑤ Zorzi (2017)	31.40	0.6517	0.5289
⑥ Dong and You (2006)	36.03	0.8235	0.0096
KF (Algorithm 3.1)	37.38	6.8568	**0.0085**

Bold numbers indicate the smallest values.

The proposed RKF outperforms all the other robust filtering strategies in terms of error variance. The KF, however, was not able to estimate the system state, presenting an exponentially increasing error variance. Hence, it is not shown in Figure 3. This emphasizes how the parametric uncertainties can severely degrade its performance. Comparing the RKF with the robust filter in Ishihara, Terra and Cerri (2015) corroborates how choosing a smaller value of the penalty parameter μ instead of letting $\mu \to \infty$ can increase the estimation performance. Moreover, using the algebraic expressions in Algorithm 3.2 rather than inverting a large matrix block also reduces execution time. The other robust filtering strategies exhibit significantly larger error variances compared to the RKF. Naturally, the KF has the largest mean error variance, as it assumes a nominal system model. The RKF also presents the smallest standard deviation. In terms of execution time, as expected, the KF takes the least time due to its simplicity. The RKF requires slightly more time than the filters by Dong and You (2006) and Sayed and Nascimento (1999), which is compensated by its superior estimation quality. In contrast, the robust LMI-based filter (ABOLHASANI; RAHMANI, 2018) demands significantly more time than the other strategies, since it depends on the solution of an LMI at each time step, which might be problematic in online applications. The risk-sensitive filter (ZORZI, 2017) also requires a relatively large amount of time for each iteration, mainly due to the computation of the risk-sensitive parameter. Overall, the proposed RKF features a satisfactory estimation performance at a reasonable computational cost, being therefore suitable for real-time applications.

Additionally, we take a closer look at how the two parameters of the proposed RKF, namely the penalty parameter μ and the approximation parameter ξ, influence the filter performance. Figure 4 compiles the results of a series of simulations with several combinations of the RKF parameters. For each combination, we compute the mean estimation error variance $\overline{\text{MSE}}$, as previously described. As pointed out in Section 3.2.3, we obtain better results when $\xi \in (0, 1)$. Furthermore, within this range, smaller values of μ lead to smaller mean error variances. Above this range, the filter performance significantly degrades.

Figure 4 – Effect of the RKF parameters μ and ξ on the mean error variance $\overline{\text{MSE}}$.



3.3 Robust Kalman Filtering for Systems with Polytopic Uncertainties

In this section, we propose another robust version of the Nominal Kalman Filter introduced in Section 3.1. This time, we address the case where the underlying system is subject to polytopic parametric uncertainties. In this specific model description, we consider that the system parameters arbitrarily vary within a convex polyhedron centered at the nominal parameters (CHANG; PARK; TANG, 2015).

We follow a similar procedure as the one outlined in Section 3.2 to propose a robust filter for linear discrete-time systems subject to polytopic uncertainties. From a deterministic viewpoint, we formulate the robust estimation problem as a constrained regularized least-squares estimation problem with uncertainties (Section 2.2.4). The linear equality constraints correspond to each vertex of the uncertainty polytope. We also use the penalty function method (Section 2.1) to transform this problem into an unconstrained equivalent, whose solution provides the recursive expressions of the Polytopic Robust Kalman Filter (PRKF). Like the previous estimators, we present the PRKF as a correction-prediction algorithm. Additionally, we analyze the stability properties of the proposed filter and conclude the section with an illustrative example.

3.3.1 Problem Formulation

3.3.1.1 System Model

Consider the following discrete-time state-space description of an uncertain linear system:

$$\begin{aligned} x_{k+1} &= (F_{0,k} + \delta F_k)x_k + (G_{0,k} + \delta G_k)u_k + (H_{0,k} + \delta H_k)w_k, \\ y_k &= (C_{0,k} + \delta C_k)x_k + (D_{0,k} + \delta D_k)v_k, \end{aligned} \tag{3.83}$$

for $k = 0, 1, \ldots, N$, where $x_k \in \mathbb{R}^n$ is a state vector, $u_k \in \mathbb{R}^m$ is an input vector, $w_k \in \mathbb{R}^p$ is the system noise, $y_k \in \mathbb{R}^r$ is a measurement vector, and $v_k \in \mathbb{R}^q$ the measurement noise. $F_{0,k} \in \mathbb{R}^{n\times n}$, $G_{0,k} \in \mathbb{R}^{n\times m}$, $H_{0,k} \in \mathbb{R}^{n\times p}$, $C_{0,k} \in \mathbb{R}^{r\times n}$, and $D_{0,k} \in \mathbb{R}^{r\times q}$ are known nominal parameter matrices, whereas $\delta F_k \in \mathbb{R}^{n\times n}$, $\delta G_k \in \mathbb{R}^{n\times m}$, $\delta H_k \in \mathbb{R}^{n\times p}$, $\delta C_k \in \mathbb{R}^{r\times n}$, and $\delta D_k \in \mathbb{R}^{r\times q}$ are unknown uncertainties bounded to a convex polyhedral domain described by V vertices,

$$\mathbb{V}_k := \left\{\left(\delta F_k,\, \delta G_k,\, \delta H_k,\, \delta C_k,\, \delta D_k\right) = \sum_{\nu=1}^{V} \alpha_{\nu,k}\left(F_{\nu,k},\, G_{\nu,k},\, H_{\nu,k},\, C_{\nu,k},\, D_{\nu,k}\right)\right\}, \tag{3.84}$$

where $F_{\nu,k} \in \mathbb{R}^{n\times n}$, $G_{\nu,k} \in \mathbb{R}^{n\times m}$, $H_{\nu,k} \in \mathbb{R}^{n\times p}$, $C_{\nu,k} \in \mathbb{R}^{r\times n}$ and $D_{\nu,k} \in \mathbb{R}^{r\times q}$ are known, and $\alpha_k := \begin{bmatrix}\alpha_{1,k} & \cdots & \alpha_{V,k}\end{bmatrix}^T$ belongs to the unit simplex

$$\Lambda_V := \left\{\alpha \in \mathbb{R}^V \;:\; \sum_{\nu=1}^{V} \alpha_\nu = 1,\, \alpha_\nu \geq 0\right\}. \tag{3.85}$$

When a stochastic interpretation is adopted, we usually assume that x_0, w_k, and v_k are mutually independent zero-mean Gaussian random variables with respective variances

$$\boldsymbol{E}\left\{x_0 x_0^T\right\} = P_0 \succ 0, \quad \boldsymbol{E}\left\{w_k w_l^T\right\} = Q_k \delta_{kl} \succ 0, \quad \text{and} \quad \boldsymbol{E}\left\{v_k v_l^T\right\} = R_k \delta_{kl} \succ 0,$$

where δ_{kl} is the Kronecker delta function, such that $\delta_{kl} = 1$ if $k = l$, and $\delta_{kl} = 0$ otherwise. However, the strategy we develop to derive the polytopic robust filter does not require that these variables have any particular distribution.

3.3.1.2 Robust Estimation Problem

The goal is to design a robust state estimator for the uncertain system (3.83)-(3.84). Since the system state sequence $\{x_k\}$ is not readily available nor is perfectly observed, the problem consists of using all the information available up to time instant k, $\boldsymbol{Y}_k = \{y_0, \ldots, y_k, u_0, \ldots, u_k\}$, to obtain a so-called filtered state estimate $\hat{x}_{k|k}$ of x_k, as well as a predicted estimate $\hat{x}_{k+1|k}$ of x_{k+1}, despite the presence of the polytopic model uncertainties $\delta_k := \{\delta F_k,\, \delta G_k,\, \delta H_k,\, \delta C_k,\, \delta D_k\}$.

Following the procedure described in Section 3.1.1.2 for the Nominal Kalman Filter, we adopt a deterministic viewpoint (BRYSON; HO, 1975). As such, we introduce the variables $\hat{x}_k$, $\hat{x}_{k+1}$, $\hat{w}_k$, and $\hat{v}_k$ as substitutes for the corresponding random variables x_k, x_{k+1}, w_k, and v_k in the stochastic model (3.83). Then, based on Sayed (2001) and Ishihara, Terra and Cerri (2015), assuming that at each time step k, an *a priori* state estimate $\hat{x}_{k|k-1}$, a measurement y_k, and the input u_k are available, we formulate a min-max constrained optimization problem in which a one-step quadratic objective function should be minimized under the maximum influence of the polytopic parametric uncertainties δ_k, i.e.,

$$\min_{\substack{\hat{x}_k,\hat{x}_{k+1},\\ \hat{w}_k,\hat{v}_k}} \max_{\delta_k} \; J_k(\hat{x}_k, \hat{w}_k, \hat{v}_k) = \|\hat{x}_k - \hat{x}_{k|k-1}\|^2_{P^{-1}_{k|k-1}} + \|\hat{w}_k\|^2_{Q_k^{-1}} + \|\hat{v}_k\|^2_{R_k^{-1}}, \tag{3.86}$$

subject to the set of constraints

$$\begin{bmatrix} I_n \\ \vdots \\ I_n \end{bmatrix} \hat{x}_{k+1} = \begin{bmatrix} F_{0,k} + \alpha_{1,k} V F_{1,k} \\ \vdots \\ F_{0,k} + \alpha_{V,k} V F_{V,k} \end{bmatrix} \hat{x}_k + \begin{bmatrix} G_{0,k} + \alpha_{1,k} V G_{1,k} \\ \vdots \\ G_{0,k} + \alpha_{V,k} V G_{V,k} \end{bmatrix} u_k + \begin{bmatrix} H_{0,k} + \alpha_{1,k} V H_{1,k} \\ \vdots \\ H_{0,k} + \alpha_{V,k} V H_{V,k} \end{bmatrix} \hat{w}_k, \tag{3.87a}$$

$$\begin{bmatrix} I_r \\ \vdots \\ I_r \end{bmatrix} y_k = \begin{bmatrix} C_{0,k} + \alpha_{1,k} V C_{1,k} \\ \vdots \\ C_{0,k} + \alpha_{V,k} V C_{V,k} \end{bmatrix} \hat{x}_k + \begin{bmatrix} D_{0,k} + \alpha_{1,k} V D_{1,k} \\ \vdots \\ D_{0,k} + \alpha_{V,k} V D_{V,k} \end{bmatrix} \hat{v}_k, \tag{3.87b}$$

for $k = 0, 1, \ldots, N$, where $\hat{w}_k$ and $\hat{v}_k$ are fitting errors weighted respectively by $Q_k \succ 0$ and $R_k \succ 0$, and $P_{k|k-1} \succ 0$ weights the *a priori* estimation error $x_k - \hat{x}_{k|k-1}$. Recall that, from a stochastic viewpoint, matrices Q_k and R_k represent the variances of the random variables w_k and v_k. Nonetheless, in this more general deterministic setting, they are rather understood as weighting matrices.

Remark 3.5. *The constraints* (3.87) *of problem* (3.86) *are derived from* (3.83)-(3.84) *by individually considering each vertex of the polytope. The equivalence between them can be easily shown by summing all the correspondent state and measurement equations in* (3.87)*, as follows:*

$$\begin{aligned} \hat{x}_{k+1} &= \left(F_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} F_{\nu,k} \right) \hat{x}_k + \left(G_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} G_{\nu,k} \right) u_k + \left(H_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} H_{\nu,k} \right) \hat{w}_k, \\ y_k &= \left(C_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} C_{\nu,k} \right) \hat{x}_k + \left(D_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} D_{\nu,k} \right) \hat{v}_k, \end{aligned}$$

which correspond to the same equations in (3.83)-(3.84)*, considering the deterministic variables.*

To simplify the notation, we rewrite the constraints in (3.87) in the more compact form

$$\begin{aligned} \boldsymbol{I}_n \hat{x}_{k+1} &= (\boldsymbol{F}_{0,k} + \boldsymbol{\delta F}_k)\hat{x}_k + (\boldsymbol{G}_{0,k} + \boldsymbol{\delta G}_k) u_k + (\boldsymbol{H}_{0,k} + \boldsymbol{\delta H}_k)\hat{w}_k, \\ \boldsymbol{I}_r y_k &= (\boldsymbol{C}_{0,k} + \boldsymbol{\delta C}_k)\hat{x}_k + (\boldsymbol{D}_{0,k} + \boldsymbol{\delta D}_k)\hat{v}_k, \end{aligned} \tag{3.88}$$

in which we define

$$\begin{aligned} \boldsymbol{I}_n &:= \mathbf{1}_V \otimes I_n, \quad \boldsymbol{F}_{0,k} := \mathbf{1}_V \otimes F_{0,k}, \quad \boldsymbol{G}_{0,k} := \mathbf{1}_V \otimes G_{0,k}, \quad \boldsymbol{H}_{0,k} := \mathbf{1}_V \otimes H_{0,k}, \\ \boldsymbol{I}_r &:= \mathbf{1}_V \otimes I_r, \quad \boldsymbol{C}_{0,k} := \mathbf{1}_V \otimes C_{0,k}, \quad \text{and} \quad \boldsymbol{D}_{0,k} := \mathbf{1}_V \otimes D_{0,k}, \end{aligned} \tag{3.89}$$

where $\mathbf{1}_V := \begin{bmatrix} 1 \cdots 1 \end{bmatrix}^T \in \mathbb{R}^V$ and $\otimes$ denotes the Kronecker product. Moreover, the uncertainties are given by

$$\begin{aligned} \begin{bmatrix} \boldsymbol{\delta F}_k & \boldsymbol{\delta G}_k & \boldsymbol{\delta H}_k \end{bmatrix} &= \bar{\boldsymbol{\alpha}}_{1,k} V \begin{bmatrix} \bar{\boldsymbol{F}}_k & \bar{\boldsymbol{G}}_k & \bar{\boldsymbol{H}}_k \end{bmatrix}, \\ \begin{bmatrix} \boldsymbol{\delta C}_k & \boldsymbol{\delta D}_k \end{bmatrix} &= \bar{\boldsymbol{\alpha}}_{2,k} V \begin{bmatrix} \bar{\boldsymbol{C}}_k & \bar{\boldsymbol{D}}_k \end{bmatrix}, \end{aligned} \tag{3.90}$$

where

$$\bar{\boldsymbol{\alpha}}_{1,k} := \begin{bmatrix} \alpha_{1,k} I_n & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_{V,k} I_n \end{bmatrix}, \quad \bar{\boldsymbol{F}}_k := \begin{bmatrix} F_{1,k} \\ \vdots \\ F_{V,k} \end{bmatrix}, \quad \bar{\boldsymbol{G}}_k := \begin{bmatrix} G_{1,k} \\ \vdots \\ G_{V,k} \end{bmatrix}, \quad \bar{\boldsymbol{H}}_k := \begin{bmatrix} H_{1,k} \\ \vdots \\ H_{V,k} \end{bmatrix},$$
$$\bar{\boldsymbol{\alpha}}_{2,k} := \begin{bmatrix} \alpha_{1,k} I_r & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_{V,k} I_r \end{bmatrix}, \quad \bar{\boldsymbol{C}}_k := \begin{bmatrix} C_{1,k} \\ \vdots \\ C_{V,k} \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{D}}_k := \begin{bmatrix} D_{1,k} \\ \vdots \\ D_{V,k} \end{bmatrix}. \tag{3.91}$$

The solution to problem (3.86)-(3.87) recursively provides the filtered and predicted robust state estimates $\hat{x}_{k|k}$ and $\hat{x}_{k|k+1}$, which compose the Polytopic Robust Kalman Filter. In addition, we refer to this problem as a *regularized least-squares estimation problem with polytopic uncertainties*, whose details we discuss in the next section.

3.3.2 Regularized Least-Squares Estimation Problem with Polytopic Uncertainties

Consider the general problem of obtaining an estimate $\hat{x}$ of an unknown vector x based on measurements y_0, related to x according to the uncertain linear system

$$(y_0 + \delta y) = (A_0 + \delta A)x + (B_0 + \delta B)w, \tag{3.92}$$

where w is a noise vector, also unknown, A_0 and B_0 are known matrices, and y_0 is a known measurement vector. The parametric uncertainties $\delta y, \delta A$, and δB are unknown but bounded to a convex polyhedral domain described by V vertices,

$$\mathbb{V} := \left\{ \Big(\delta y,\, \delta A,\, \delta B\Big) = \sum_{\nu=1}^{V} \alpha_\nu \Big(y_\nu,\, A_\nu,\, B_\nu\Big) \right\}, \tag{3.93}$$

in which y_ν, A_ν, and B_ν are known and $\alpha := \begin{bmatrix} \alpha_1 & \cdots & \alpha_V \end{bmatrix}^T$ belongs to the unit simplex

$$\Lambda_V := \left\{ \alpha \in \mathbb{R}^V \;:\; \sum_{\nu=1}^{V} \alpha_\nu = 1,\, \alpha_\nu \geq 0 \right\}. \tag{3.94}$$

Moreover, assume that an *a priori* estimate $\bar{x}$ of x is available as well.

Adopting a deterministic viewpoint, we formulate the regularized least-squares estimation problem with polytopic uncertainty as

$$\min_{x,\,w} \max_{\delta y,\, \delta A,\, \delta B} \; J(x, w) = \|x - \bar{x}\|_P^2 + \|w\|_Q^2, \tag{3.95}$$

subject to the set of constraints

$$\begin{bmatrix} y_0 + \alpha_1 V y_1 \\ \vdots \\ y_0 + \alpha_V V y_V \end{bmatrix} = \begin{bmatrix} A_0 + \alpha_1 V A_1 \\ \vdots \\ A_0 + \alpha_V V A_V \end{bmatrix} x + \begin{bmatrix} B_0 + \alpha_1 V B_1 \\ \vdots \\ B_0 + \alpha_V V B_V \end{bmatrix} w. \tag{3.96}$$

In the objective function of problem (3.95), $\bar{P} \succeq 0$ and $Q \succ 0$ are given weighting matrices respectively associated with the *a priori* estimation error $x - \bar{x}$ and the model fitting error w. Therefore, we minimize the objective function under the maximum influence of the parametric polytopic uncertainties.

Remark 3.6. *We derive the constraints* (3.96) *of problem* (3.95) *from* (3.92)-(3.93) *by individually considering each vertex of the polytope. The equivalence between them can be shown by pre-multiplying both sides of* (3.96) *by* $\begin{bmatrix} I & \cdots & I \end{bmatrix} = \mathbf{1}_V^T \otimes I$*, such that*

$$y_0 + \sum_{\nu=1}^{V} \alpha_\nu y_\nu = \left(A_0 + \sum_{\nu=1}^{V} \alpha_\nu A_\nu \right) x + \left(B_0 + \sum_{\nu=1}^{V} \alpha_\nu B_\nu \right) w$$

which corresponds to the same equation in (3.92)-(3.93).

To use a simpler notation, we rewrite the constraints in (3.96) in a more compact form, as follows:

$$\boldsymbol{y}_0 + \boldsymbol{\delta y} = (\boldsymbol{A}_0 + \boldsymbol{\delta A})x + (\boldsymbol{B}_0 + \boldsymbol{\delta B})w, \tag{3.97}$$

where we define

$$\boldsymbol{y}_0 := \mathbf{1}_V \otimes y_0, \quad \boldsymbol{A}_0 := \mathbf{1}_V \otimes A_0, \quad \text{and} \quad \boldsymbol{B}_0 := \mathbf{1}_V \otimes B_0, \tag{3.98}$$

and the uncertainties are given by

$$\begin{bmatrix} \boldsymbol{\delta y} & \boldsymbol{\delta A} & \boldsymbol{\delta B} \end{bmatrix} = \bar{\boldsymbol{\alpha}}\, V \begin{bmatrix} \bar{\boldsymbol{y}} & \bar{\boldsymbol{A}} & \bar{\boldsymbol{B}} \end{bmatrix}, \tag{3.99}$$

in which

$$\bar{\boldsymbol{\alpha}} := \begin{bmatrix} \alpha_1 I & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_V I \end{bmatrix}, \quad \bar{\boldsymbol{y}} := \begin{bmatrix} y_1 \\ \vdots \\ y_V \end{bmatrix}, \quad \bar{\boldsymbol{A}} := \begin{bmatrix} A_1 \\ \vdots \\ A_V \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{B}} := \begin{bmatrix} B_1 \\ \vdots \\ B_V \end{bmatrix}. \tag{3.100}$$

The first step to solve the constrained problem (3.95)-(3.96) is transforming it into a more convenient unconstrained problem. Since the linear constraints in (3.96) cannot be inserted into the objective function by direct substitution, we rely on the penalty function method presented in Section 2.1. This way, we include the redefined constraints (3.97) in the objective function as a quadratic term multiplied by a penalty parameter $\mu > 0$, which penalizes constraint violations. Therefore, for a fixed $\mu > 0$, we rewrite problem (3.95)-(3.96) as

$$\min_{x,w} \max_{\boldsymbol{\delta y}, \boldsymbol{\delta A}, \boldsymbol{\delta B}} J^\mu(x, w, \boldsymbol{\delta y}, \boldsymbol{\delta A}, \boldsymbol{\delta B}), \tag{3.101}$$

with a new objective function

$$J^\mu(x, w, \boldsymbol{\delta y}, \boldsymbol{\delta A}, \boldsymbol{\delta B}) = \begin{bmatrix} x - \bar{x} \\ w \end{bmatrix}^T \begin{bmatrix} \bar{P} & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} x - \bar{x} \\ w \end{bmatrix} +$$

$$\left\{ \left(\begin{bmatrix} \boldsymbol{A}_0 & \mathbf{B}_0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\delta A} & \delta\mathbf{B} \end{bmatrix} \right) \begin{bmatrix} x - \bar{x} \\ w \end{bmatrix} - \left[(\boldsymbol{y}_0 - \boldsymbol{A}_0 \bar{x}) + (\boldsymbol{\delta y} - \boldsymbol{\delta A} \bar{x}) \right] \right\}^T \mu I \left\{ \bullet \right\}, \tag{3.102}$$

considering the definitions in (3.97) through (3.100). Moreover, note that since $\alpha = \begin{bmatrix}\alpha_1 \cdots \alpha_V\end{bmatrix}^T$ belongs to the unit simplex Λ_V in (3.94), we have that $\|\bar{\boldsymbol{\alpha}}\| \leq 1$ in (3.99).

Problem (3.101)-(3.102) thus has the form of a regularized least-squares problem with uncertainties (Section 2.2.4), considering the following mappings between (2.13) and (3.102):

$$\begin{gathered} z \leftarrow \begin{bmatrix} x - \bar{x} \\ w \end{bmatrix}, \quad \mathcal{Q} \leftarrow \begin{bmatrix} \bar{P} & 0 \\ 0 & Q \end{bmatrix}, \quad \mathcal{A} \leftarrow \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \end{bmatrix}, \quad b \leftarrow \boldsymbol{y}_0 - \boldsymbol{A}_0\bar{x}, \quad \mathcal{W} \leftarrow \mu I, \\ \delta\mathcal{A} \leftarrow \begin{bmatrix} \boldsymbol{\delta A} & \boldsymbol{\delta B} \end{bmatrix}, \quad \text{and} \quad \delta b \leftarrow \boldsymbol{\delta y} - \boldsymbol{\delta A}\bar{x}. \end{gathered} \tag{3.103}$$

In addition, the correspondence between the parametric uncertainty model in (2.14) and the polytopic uncertainty model in (3.99) is given by

$$\begin{bmatrix} \delta\mathcal{A} & \delta b \end{bmatrix} = M\Delta \begin{bmatrix} E_{\mathcal{A}} & E_b \end{bmatrix}, \quad \|\Delta\| \leq 0,$$

where

$$M \leftarrow I, \quad \Delta \leftarrow \bar{\boldsymbol{\alpha}} \otimes I, \quad E_{\mathcal{A}} \leftarrow V \begin{bmatrix} \bar{\boldsymbol{A}} & \bar{\boldsymbol{B}} \end{bmatrix}, \quad \text{and} \quad E_b \leftarrow V\Big(\bar{\boldsymbol{y}} - \bar{\boldsymbol{A}}\bar{x}\Big). \tag{3.104}$$

Hence, to solve problem (3.101)-(3.102), we use the results in Section 2.2.4. From the solution, we then extract the estimate $\hat{x}^\mu$ of x, which is conditioned by the penalty parameter μ, as we shown in the following lemma.

Lemma 3.5. *Consider problem* (3.101)-(3.102) *with polytopic parametric uncertainties given by* (3.99)*, in which* $\bar{P} \succeq 0$*,* $Q \succ 0$*, and* $\begin{bmatrix} A_0 \\ \bar{\boldsymbol{A}} \end{bmatrix}$ *has full column rank. The estimate* $\hat{x}^\mu$ *of* x*, conditioned by the penalty parameter* $\mu > 0$*, is given by*

$$\hat{x}^\mu = \Big(\bar{P} + \widehat{A}^T\widehat{Q}^{-1}\widehat{A} + \bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{A}}\Big)^{-1}\Big(\bar{P}\bar{x} + \widehat{A}^T\widehat{Q}^{-1}\hat{y} + \bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}\Big), \tag{3.105}$$

in which we define the auxiliary entities

$$\begin{aligned} \Phi &:= \hat{\lambda}\,\mu\, V\Big(\hat{\lambda} - \mu\Big)^{-1} I, & \varphi &:= \hat{\lambda}V^2, \\ \bar{Q} &:= \varphi^{-1}I + \bar{\boldsymbol{B}}Q^{-1}\bar{\boldsymbol{B}}^T, & \widehat{Q} &:= \Phi^{-1} + B_0\Big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\Big)^{-1}B_0^T, \\ \widehat{A} &:= A_0 - B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{A}}, & \hat{y} &:= y_0 - B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}, \end{aligned} \tag{3.106}$$

where $\hat{\lambda}$ *is a nonnegative scalar parameter obtained from the auxiliary optimization problem*

$$\hat{\lambda} := \arg\min_{\lambda > \mu} \Gamma(\lambda), \tag{3.107}$$

with objective function $\Gamma(\lambda)$ *given by*

$$\Gamma(\lambda) := \|z(\lambda)\|_{\mathcal{Q}}^2 + \lambda\|E_{\mathcal{A}}z(\lambda) - E_b\|^2 + \|\mathcal{A}z(\lambda) - b\|_{\mathcal{W}(\lambda)}^2, \tag{3.108}$$

in which

$$\mathcal{W}(\lambda) := \lambda\,\mu\big(\lambda-\mu\big)^{-1} I$$
$$z(\lambda) := \big(\mathcal{Q}+\mathcal{A}^T\mathcal{W}(\lambda)\mathcal{A}+\lambda E_{\mathcal{A}}^T E_{\mathcal{A}}\big)^{-1}\big(\mathcal{A}^T\mathcal{W}(\lambda)b+\lambda E_{\mathcal{A}}^T E_b\big),$$

considering the definitions in (3.103) *and* (3.104).

Proof. Problem (3.101)-(3.102) is a special case of a regularized least-squares problem with uncertainties, considering the mappings in (3.103) and (3.104). Moreover, as $\bar{P}\succeq 0$, $Q\succ 0$, and $\mu>0$, we have that $\mathcal{Q}\succeq 0$ and $\mathcal{W}\succ 0$, such that we can use Lemma 2.8 to find the solution. It is further required that the block $\begin{bmatrix} 0 & I \\ \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \bar{\boldsymbol{A}} & \bar{\boldsymbol{B}}\end{bmatrix}$ should have full column rank, which is satisfied by conditioning $\begin{bmatrix} A_0 \\ \bar{\boldsymbol{A}}\end{bmatrix}$ to have full column rank. Therefore, by substituting the mappings (3.103) and (3.104) into the unique solution (2.15), we obtain

$$\begin{bmatrix}\hat{x}^\mu-\bar{x}\\ \hat{w}^\mu\end{bmatrix} = \begin{bmatrix}\bar{P}+A_0^T\Phi A_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}} & A_0^T\Phi B_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\\ B_0^T\Phi A_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}} & Q+B_0^T\Phi B_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}}\end{bmatrix}^{-1}\times$$
$$\begin{bmatrix}A_0^T\Phi(y_0-A_0\bar{x})+\varphi\bar{\boldsymbol{A}}^T(\bar{\boldsymbol{y}}-\bar{\boldsymbol{A}}\bar{x})\\ B_0^T\Phi(y_0-A_0\bar{x})+\varphi\bar{\boldsymbol{B}}^T(\bar{\boldsymbol{y}}-\bar{\boldsymbol{A}}\bar{x})\end{bmatrix},$$

in which we define $\Phi := \hat{\lambda}\,\mu\,V\big(\hat{\lambda}-\mu\big)^{-1}I$ and $\varphi := \hat{\lambda}V^2$. Then, summing $\begin{bmatrix}\bar{x}\\0\end{bmatrix}$ to both sides of the equation above yields

$$\begin{bmatrix}\hat{x}^\mu\\ \hat{w}^\mu\end{bmatrix} = \begin{bmatrix}\bar{P}+A_0^T\Phi A_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}} & A_0^T\Phi B_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\\ B_0^T\Phi A_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}} & Q+B_0^T\Phi B_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\end{bmatrix}^{-1}\begin{bmatrix}\bar{P}\bar{x}+A_0^T\Phi y_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{y}}\\ B_0^T\Phi y_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{y}}\end{bmatrix}.$$

Note that the equation above represents a system of simultaneous equations, such that we can write it as the following set equations:

$$\big(\bar{P}+A_0^T\Phi A_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}}\big)\hat{x}^\mu+\big(A_0^T\Phi B_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\big)\hat{w}^\mu = \bar{P}\bar{x}+A_0^T\Phi y_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{y}}, \quad (3.109)$$
$$\big(B_0^T\Phi A_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}}\big)\hat{x}^\mu+\big(Q+B_0^T\Phi B_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)\hat{w}^\mu = B_0^T\Phi y_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{y}}. \quad (3.110)$$

Then, isolating $\hat{w}^\mu$ in (3.110) gives

$$\hat{w}^\mu = \big(Q+B_0^T\Phi B_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}\Big(B_0^T\Phi y_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{y}}-\big(B_0^T\Phi A_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}}\big)\hat{x}^\mu\Big).$$

Substituting $\hat{w}^\mu$ back into (3.109) thus yields

$$\Big[\bar{P}+A_0^T\Phi A_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}}-$$
$$\big(A_0^T\Phi B_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\big)\big(Q+B_0^T\Phi B_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}\big(B_0^T\Phi A_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}}\big)\Big]\,\hat{x}^\mu =$$
$$\bar{P}\bar{x}+A_0^T\Phi y_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{y}}-\big(A_0^T\Phi B_0+\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\big)\big(Q+B_0^T\Phi B_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}\big(B_0^T\Phi y_0+\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{y}}\big)$$
$$(3.111)$$

Now, we expand the left-hand side of (3.111) and obtain

$$\Big[\bar{P} + A_0^T\Big(\Phi - \Phi B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}B_0^T\Phi\Big)A_0 - A_0^T\Phi B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}} - \varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}B_0^T\Phi A_0 + \varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}} - \varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}}\varphi\Big]\hat{x}^{\mu}. \tag{3.112}$$

We can simplify the second term of (3.112) by applying Lemma A.1, as follows:

$$A_0^T\Big(\Phi - \Phi B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}B_0^T\Phi\Big)A_0 = A_0^T\Big(\underbrace{\Phi^{-1} + B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}B_0^T}_{\widehat{Q}}\Big)^{-1}A_0 = A_0^T\widehat{Q}^{-1}A_0. \tag{3.113}$$

Then, we simplify the third term of (3.112) using Lemma A.2 twice:

$$A_0^T\Phi B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}\varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}} = A_0^T\Big(\underbrace{\Phi^{-1} + B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}B_0^T}_{\widehat{Q}}\Big)^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\Big(\underbrace{\varphi^{-1}I + \bar{\boldsymbol{B}}Q^{-1}\bar{\boldsymbol{B}}^T}_{\bar{Q}}\Big)^{-1}\bar{\boldsymbol{A}} = A_0^T\widehat{Q}^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{A}}. \tag{3.114}$$

Applying the same procedure above for the fourth term of (3.112) gives

$$\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}B_0^T\Phi A_0 = \bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T\widehat{Q}^{-1}A_0. \tag{3.115}$$

Now, we expand the last two terms of (3.112) using both Lemma A.1 and Lemma A.2, as follows:

$$\begin{aligned}
&\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}} - \varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}} + B_0^T\Phi B_0\big)^{-1}\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}}\varphi = \\
&\varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}} - \varphi\bar{\boldsymbol{A}}^T\bar{\boldsymbol{B}}\Big[\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1} - \\
&\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}B_0^T\Big(\underbrace{\Phi^{-1} + B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}B_0^T}_{\widehat{Q}}\Big)^{-1}B_0\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}\Big]\bar{\boldsymbol{B}}^T\bar{\boldsymbol{A}}\varphi = \\
&\bar{\boldsymbol{A}}^T\Big(\varphi I - \varphi\bar{\boldsymbol{B}}\big(Q + \varphi\bar{\boldsymbol{B}}^T\bar{\boldsymbol{B}}\big)^{-1}\bar{\boldsymbol{B}}^T\varphi\Big)\bar{\boldsymbol{A}} + \\
&\bar{\boldsymbol{A}}^T\Big(\underbrace{\varphi^{-1}I + \bar{\boldsymbol{B}}Q^{-1}\bar{\boldsymbol{B}}^T}_{\bar{Q}}\Big)^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T\widehat{Q}^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\Big(\underbrace{\varphi^{-1}I + \bar{\boldsymbol{B}}Q^{-1}\bar{\boldsymbol{B}}^T}_{\bar{Q}}\Big)^{-1}\bar{\boldsymbol{A}} = \\
&\bar{\boldsymbol{A}}^T\Big(\underbrace{\varphi^{-1}I + \bar{\boldsymbol{B}}Q^{-1}\bar{\boldsymbol{B}}^T}_{\bar{Q}}\Big)^{-1}\bar{\boldsymbol{A}} + \bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T\widehat{Q}^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{A}} = \\
&\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{A}} + \bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T\widehat{Q}^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{A}}.
\end{aligned} \tag{3.116}$$

Then we substitute (3.113), (3.114), (3.115), and (3.116) back into (3.112) to obtain

$$\Big[\bar{P} + \bar{\boldsymbol{A}}^T \bar{Q}^{-1} \bar{\boldsymbol{A}} + A_0^T \widehat{Q}^{-1} \big(\underbrace{A_0 - B_0 Q^{-1} \bar{\boldsymbol{B}}^T \bar{Q}^{-1} \bar{\boldsymbol{A}}}_{\widehat{A}}\big) -$$
$$\bar{\boldsymbol{A}}^T \bar{Q}^{-1} \bar{\boldsymbol{B}} Q^{-1} B_0^T \widehat{Q}^{-1} \big(\underbrace{A_0 - B_0 Q^{-1} \bar{\boldsymbol{B}}^T \bar{Q}^{-1} \bar{\boldsymbol{A}}}_{\widehat{A}}\big)\Big] \hat{x}^\mu = \Big[\bar{P} + \bar{\boldsymbol{A}}^T \bar{Q}^{-1} \bar{\boldsymbol{A}} +$$
$$\big(\underbrace{A_0^T - \bar{\boldsymbol{A}}^T \bar{Q}^{-1} \bar{\boldsymbol{B}} Q^{-1} B_0^T}_{\widehat{A}^T}\big) \widehat{Q}^{-1} \widehat{A}\Big] \hat{x}^\mu = \big(\bar{P} + \widehat{A}^T \widehat{Q}^{-1} \widehat{A} + \bar{\boldsymbol{A}}^T \bar{Q}^{-1} \bar{\boldsymbol{A}}\big) \hat{x}^\mu. \tag{3.117}$$

In a similar fashion, we expand the right-hand side of (3.111), as follows:

$$\bar{P}\bar{x} + A_0^T \Big(\Phi - \Phi B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} B_0^T \Phi\Big) y_0 -$$
$$A_0^T \Phi B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{y}} - \varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{B}} \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} B_0^T \Phi y_0 +$$
$$\varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{y}} - \varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{B}} \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} \bar{\boldsymbol{B}}^T \bar{\boldsymbol{y}} \varphi. \tag{3.118}$$

First, we apply Lemma A.1 to simplify the second term of (3.118):

$$A_0^T \Big(\Phi - \Phi B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} B_0^T \Phi\Big) y_0 =$$
$$A_0^T \Big(\underbrace{\Phi^{-1} + B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}}\big)^{-1} B_0^T}_{\widehat{Q}}\Big)^{-1} y_0 = A_0^T \widehat{Q}^{-1} y_0. \tag{3.119}$$

Next, we simplify the third term of (3.118) using Lemma A.2 twice:

$$A_0^T \Phi B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{y}} =$$
$$A_0^T \Big(\underbrace{\Phi^{-1} + B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}}\big)^{-1} B_0^T}_{\widehat{Q}}\Big)^{-1} B_0 Q^{-1} \bar{\boldsymbol{B}}^T \Big(\underbrace{\varphi^{-1} I + \bar{\boldsymbol{B}} Q^{-1} \bar{\boldsymbol{B}}^T}_{\bar{Q}}\Big)^{-1} \bar{\boldsymbol{y}} =$$
$$A_0^T \widehat{Q}^{-1} B_0 Q^{-1} \bar{\boldsymbol{B}}^T \bar{Q}^{-1} \bar{\boldsymbol{y}}. \tag{3.120}$$

The same procedure is used to simplify the fourth term of (3.118), such that

$$\varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{B}} \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} B_0^T \Phi y_0 = \bar{\boldsymbol{A}}^T \bar{Q}^{-1} \bar{\boldsymbol{B}} Q^{-1} B_0^T \widehat{Q}^{-1} y_0. \tag{3.121}$$

Then, we apply Lemma A.1 and Lemma A.2 to expand the last two terms of (3.118):

$$\varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{y}} - \varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{B}} \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}} + B_0^T \Phi B_0\big)^{-1} \bar{\boldsymbol{B}}^T \bar{\boldsymbol{y}} \varphi =$$
$$\varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{y}} - \varphi \bar{\boldsymbol{A}}^T \bar{\boldsymbol{B}} \Big[\big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}}\big)^{-1} -$$
$$\big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}}\big)^{-1} B_0^T \Big(\underbrace{\Phi^{-1} + B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}}\big)^{-1} B^T}_{\widehat{Q}}\Big)^{-1} B_0 \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}}\big)^{-1}\Big] \bar{\boldsymbol{B}}^T \bar{\boldsymbol{y}} \varphi =$$
$$\bar{\boldsymbol{A}}^T \Big(\varphi I - \varphi \bar{\boldsymbol{B}} \big(Q + \varphi \bar{\boldsymbol{B}}^T \bar{\boldsymbol{B}}\big)^{-1} \bar{\boldsymbol{B}}^T \varphi\Big) \bar{\boldsymbol{y}} +$$
$$\bar{\boldsymbol{A}}^T \Big(\underbrace{\varphi^{-1} I + \bar{\boldsymbol{B}} Q^{-1} \bar{\boldsymbol{B}}^T}_{\bar{Q}}\Big)^{-1} \bar{\boldsymbol{B}} Q^{-1} B_0^T \widehat{Q}^{-1} B_0 Q^{-1} \bar{\boldsymbol{B}}^T \Big(\underbrace{\varphi^{-1} I + \bar{\boldsymbol{B}} Q^{-1} \bar{\boldsymbol{B}}^T}_{\bar{Q}}\Big)^{-1} \bar{\boldsymbol{y}} =$$

$$\bar{\boldsymbol{A}}^T\big(\underbrace{\varphi^{-1}I+\bar{\boldsymbol{B}}Q^{-1}\bar{\boldsymbol{B}}^T}_{\bar{Q}}\big)^{-1}\bar{\boldsymbol{y}}+\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T\widehat{Q}^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}=$$
$$\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}+\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T\widehat{Q}^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}. \tag{3.122}$$

Thus, substituting (3.119), (3.120), (3.121), and (3.122) back into (3.118) yields

$$\bar{P}\bar{x}+\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}+\big(\underbrace{A_0^T-\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T}_{\widehat{A}^T}\big)\widehat{Q}^{-1}y_0-$$
$$\big(\underbrace{A_0^T-\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{B}}Q^{-1}B_0^T}_{\widehat{A}^T}\big)\widehat{Q}^{-1}B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}=\bar{P}\bar{x}+\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}+$$
$$\widehat{A}^T\widehat{Q}^{-1}\big(\underbrace{y_0-B_0Q^{-1}\bar{\boldsymbol{B}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}}_{\hat{y}}\big)=\bar{P}\bar{x}+\widehat{A}^T\widehat{Q}^{-1}\hat{y}+\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{y}}. \tag{3.123}$$

Finally, we substitute the left- and right-hand sides of (3.111) respectively by (3.117) and (3.123) and isolate $\hat{x}^\mu$ to obtain the estimate in (3.105).

Moreover, the procedure described in Lemma 2.7 is followed to obtain the parameter $\hat{\lambda}$, i.e., by solving the auxiliary optimization problem (3.107)-(3.108). Notice that, since we search for $\lambda > \mu > 0$ in problem (3.107)-(3.108), according to Remark 2.1, the invertibility of Φ is ensured. □

Remark 3.7. *The solution in Lemma 3.5 depends on the optimal parameter $\hat{\lambda}$, which results from solving the optimization problem (3.107)-(3.108). While a constrained line search method can be used to obtain a solution, this requires additional computation time. Therefore, we rather adopt the practical approximation $\hat{\lambda} = (1+\xi)\,\mu$, for some $\xi > 0$, as explained in Remark 2.2.*

Remark 3.8. *As discussed in the end of Section 3.2.2, in a robust estimation context, the penalty parameter μ can be understood as a robustness measure of the estimator. In this sense, when the system model is subject to significant uncertainties, smaller values of μ increase the estimator robustness. Conversely, for mild uncertainties, larger values of μ can be used.*

To conclude the section, we further associate a weighting matrix $\hat{P}$ to the estimation error $x - \hat{x}^\mu$. Recall that, since the underlying model contains parametric polytopic uncertainties, we cannot refer to this weighting matrix as an error variance matrix, as we cannot compute it analytically. Hence, based on the result in Lemma 3.2 and relying on the deterministic view of the robust estimation problem, we associate the weighting matrix

$$\hat{P}=\left(\bar{P}+\widehat{A}^T\widehat{Q}^{-1}\widehat{A}+\bar{\boldsymbol{A}}^T\bar{Q}^{-1}\bar{\boldsymbol{A}}\right)^{-1} \tag{3.124}$$

to the estimation error $x - \hat{x}^\mu$, considering the estimate $\hat{x}^\mu$ in (3.105).

3.3.3 Polytopic Robust Kalman Filter

In this section, we apply the results in Section 3.3.2 to obtain the so-called Polytopic Robust Kalman Filter. As aforementioned, the deterministic estimation problem (3.86)-(3.87) is a special case of a regularized least-squares estimation problem with polytopic uncertainties, considering the following mappings between (3.86)-(3.88) and (3.95)-(3.97):

$$x \leftarrow \begin{bmatrix} \hat{x}_k \\ \hat{x}_{k+1} \end{bmatrix}, \quad \bar{x} \leftarrow \begin{bmatrix} \hat{x}_{k|k-1} \\ 0 \end{bmatrix}, \quad w \leftarrow \begin{bmatrix} \hat{w}_k \\ \hat{v}_k \end{bmatrix}, \quad \bar{P} \leftarrow \begin{bmatrix} P_{k|k-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad Q \leftarrow \begin{bmatrix} Q_k^{-1} & 0 \\ 0 & R_k^{-1} \end{bmatrix},$$
$$\boldsymbol{y}_0 \leftarrow \begin{bmatrix} -\boldsymbol{G}_{0,k} u_k \\ \boldsymbol{I}_r y_k \end{bmatrix}, \quad \boldsymbol{A}_0 \leftarrow \begin{bmatrix} \boldsymbol{F}_{0,k} & -\boldsymbol{I}_n \\ \boldsymbol{C}_{0,k} & 0 \end{bmatrix}, \quad \boldsymbol{B}_0 \leftarrow \begin{bmatrix} \boldsymbol{H}_{0,k} & 0 \\ 0 & \boldsymbol{D}_{0,k} \end{bmatrix},$$
$$\boldsymbol{\delta y} \leftarrow \begin{bmatrix} -\boldsymbol{\delta G}_k u_k \\ 0 \end{bmatrix}, \quad \boldsymbol{\delta A} \leftarrow \begin{bmatrix} \boldsymbol{\delta F}_k & 0 \\ \boldsymbol{\delta C}_k & 0 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\delta B} \leftarrow \begin{bmatrix} \boldsymbol{\delta H}_k & 0 \\ 0 & \boldsymbol{\delta D}_k \end{bmatrix}. \tag{3.125}$$

Moreover, consider the following mappings between the uncertainty models (3.90) and (3.99):

$$\bar{\boldsymbol{\alpha}} \leftarrow \begin{bmatrix} \bar{\boldsymbol{\alpha}}_{1,k} & 0 \\ 0 & \bar{\boldsymbol{\alpha}}_{2,k} \end{bmatrix}, \quad \bar{\boldsymbol{y}} \leftarrow \begin{bmatrix} -\bar{\boldsymbol{G}}_k u_k \\ 0 \end{bmatrix}, \quad \bar{\boldsymbol{A}} \leftarrow \begin{bmatrix} \bar{\boldsymbol{F}}_k & 0 \\ \bar{\boldsymbol{C}}_k & 0 \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{B}} \leftarrow \begin{bmatrix} \bar{\boldsymbol{H}}_k & 0 \\ 0 & \bar{\boldsymbol{D}}_k \end{bmatrix}. \tag{3.126}$$

Notice that $P_{k|k-1}^{-1} \succ 0$, thus $\bar{P} \succeq 0$. In addition, $Q_k^{-1} \succ 0$ and $R_k^{-1} \succ 0$, such that $Q \succ 0$. Therefore, by using the results in Lemma 3.5 and in equation (3.124), we obtain the filtered and predicted robust state estimates, $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, as well as their corresponding error weighting matrices $P_{k|k}$ and $P_{k+1|k}$.

Theorem 3.6. *Consider the regularized least-squares estimation problem with polytopic uncertainties* (3.86)-(3.87) *with given initial conditions $\hat{x}_{0|-1}$, $P_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R_k \succ 0$, and fixed parameters $\mu > 0$ and $\xi > 0$. For each $k = 0, 1, \ldots, N$, its solution recursively provides the filtered and predicted robust state estimates of system* (3.83)-(3.84)*, $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$, as well as their corresponding error weighting matrices, $P_{k|k}$ and $P_{k+1|k}$, according to the procedure outlined in Algorithm 3.3.*

Proof. Since problem (3.86)-(3.87) is a regularized least-squares estimation problem with polytopic uncertainties, we can leverage the result in Lemma 3.5 to obtain the robust system state estimates $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$. Therefore, recalling the definitions in (3.89) and (3.91) we first substitute the mappings (3.125) and (3.126) into (3.106) to compute the modified system and sensing model matrices

$$\varphi = \hat{\lambda} V^2 = (1+\xi)\mu V^2,$$
$$\Phi = \begin{bmatrix} \hat{\lambda}\,\mu\, V\big(\hat{\lambda}-\mu\big)^{-1} I_n & 0 \\ 0 & \hat{\lambda}\,\mu\, V\big(\hat{\lambda}-\mu\big)^{-1} I_r \end{bmatrix} = \begin{bmatrix} \varphi(\xi V)^{-1} I_n & 0 \\ 0 & \varphi(\xi V)^{-1} I_r \end{bmatrix} =: \begin{bmatrix} \Phi_1^{-1} & 0 \\ 0 & \Phi_2^{-1} \end{bmatrix},$$

Algorithm 3.3 Polytopic Robust Kalman Filter (PRKF)

Model: Assume the uncertain system model in (3.83)-(3.84).

Initialization: Set $\hat{x}_{0|-1}$, $P_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R_k \succ 0$, $\mu > 0$, and $\xi > 0$.

for $k = 0, 1, \ldots, N$ **do**

1. Obtain a measurement y_k.
2. Compute φ using the approximation for $\hat{\lambda}$:

$$\varphi = \hat{\lambda} V^2 = (1 + \xi)\, \mu V^2$$

3. Compute the modified target system and sensing model matrices:

$$\Phi_{1,k} = \xi V \varphi^{-1} I_n \qquad \widehat{Q}_k = \Phi_1 + H_{0,k}\left(Q_k^{-1} + \varphi \bar{\boldsymbol{H}}_k^T \bar{\boldsymbol{H}}_k\right)^{-1} H_{0,k}^T$$

$$\Phi_{2,k} = \xi V \varphi^{-1} I_r \qquad \widehat{R}_k = \Phi_2 + D_{0,k}\left(R_k^{-1} + \varphi \bar{\boldsymbol{D}}_k^T \bar{\boldsymbol{D}}_k\right)^{-1} D_{0,k}^T$$

$$\bar{Q}_k = \varphi^{-1} I_{nV} + \bar{\boldsymbol{H}}_k Q_k \bar{\boldsymbol{H}}_k^T \qquad \bar{R}_k = \varphi^{-1} I_{rV} + \bar{\boldsymbol{D}}_k R_k \bar{\boldsymbol{D}}_k^T$$

$$\widehat{F}_k = F_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k \qquad \widehat{C}_k = C_{0,k} - D_{0,k} R_k \bar{\boldsymbol{D}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k$$

$$\widehat{G}_k = G_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k$$

4. Correction step:
   4.1. Compute the posterior error weighting matrix:

$$P_{k|k} = \left(P_{k|k-1}^{-1} + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + \bar{\boldsymbol{C}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k\right)^{-1}$$

   4.2. Compute the filtered robust state estimate:

$$\hat{x}_{k|k} = P_{k|k}\left(P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - \bar{\boldsymbol{F}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{G}}_k u_k\right)$$

5. Prediction step:
   5.1. Update the predicted prior error weighting matrix:

$$P_{k+1|k} = \widehat{F}_k P_{k|k} \widehat{F}_k^T + \widehat{Q}_k$$

   5.2. Update the predicted prior robust state estimate:

$$\hat{x}_{k+1|k} = \widehat{F}_k \hat{x}_{k|k} + \widehat{G}_k u_k$$

end for

$$\begin{aligned}
\bar{Q} &= \begin{bmatrix} \varphi^{-1} I_{nV} + \bar{\boldsymbol{H}}_k Q_k \bar{\boldsymbol{H}}_k^T & 0 \\ 0 & \varphi I_{rV} + \bar{\boldsymbol{D}}_k R_k \bar{\boldsymbol{D}}_k^T \end{bmatrix} =: \begin{bmatrix} \bar{Q}_k & 0 \\ 0 & \bar{R}_k \end{bmatrix}, \\
\widehat{Q} &= \begin{bmatrix} \Phi_1 + H_{0,k}\left(Q_k^{-1} + \varphi \bar{\boldsymbol{H}}_k^T \bar{\boldsymbol{H}}_k\right)^{-1} H_{0,k}^T & 0 \\ 0 & \Phi_2 + D_{0,k}\left(R_k^{-1} + \varphi \bar{\boldsymbol{D}}_k^T \bar{\boldsymbol{D}}_k\right)^{-1} D_{0,k}^T \end{bmatrix} =: \begin{bmatrix} \widehat{Q}_k & 0 \\ 0 & \widehat{R}_k \end{bmatrix}, \\
\widehat{A} &= \begin{bmatrix} F_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k & -I_n \\ C_{0,k} - D_{0,k} R_k \bar{\boldsymbol{D}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k & 0 \end{bmatrix} =: \begin{bmatrix} \widehat{F}_k & -I_n \\ \widehat{C}_k & 0 \end{bmatrix}, \\
\hat{y} &= \begin{bmatrix} -\left(G_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k\right) u_k \\ y_k \end{bmatrix} =: \begin{bmatrix} -\widehat{G}_k u_k \\ y_k \end{bmatrix}.
\end{aligned} \tag{3.127}$$

Notice that we consider Remark 3.7 to approximate the parameter $\hat{\lambda} = (1+\xi)\,\mu$, for some $\xi > 0$. Now, we substitute the mappings (3.125) and (3.126), as well as the modified matrices (3.127) into the solution (3.105), which yields

$$\begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \end{bmatrix} = \begin{bmatrix} P_{k|k-1}^{-1} + \widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{F}_k + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k + \bar{\boldsymbol{C}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k & -\widehat{F}_k^T \widehat{Q}_k^{-1} \\ -\widehat{Q}_k^{-1} \widehat{F}_k & \widehat{Q}_k^{-1} \end{bmatrix}^{-1} \times \begin{bmatrix} P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - \left(\widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{G}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k \right) u_k \\ \widehat{Q}_k^{-1} \widehat{G}_k u_k \end{bmatrix}. \tag{3.128}$$

Note that (3.128) also represents a system of simultaneous equations, such that we can write it as the following set of equations:

$$\left(P_{k|k-1}^{-1} + \widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{F}_k + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k + \bar{\boldsymbol{C}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k \right) \hat{x}_{k|k} - \widehat{F}_k^T \widehat{Q}_k^{-1} \hat{x}_{k+1|k} = P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - \left(\widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{G}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k \right) u_k, \tag{3.129}$$

$$- \widehat{Q}_k^{-1} \widehat{F}_k \hat{x}_{k|k} + \widehat{Q}_k^{-1} \hat{x}_{k+1|k} = \widehat{Q}_k^{-1} \widehat{G}_k u_k. \tag{3.130}$$

Isolating $\hat{x}_{k+1|k}$ in (3.64) then gives

$$\hat{x}_{k+1|k} = \widehat{F}_k \hat{x}_{k|k} + \widehat{G}_k u_k, \tag{3.131}$$

which is the update equation of the predicted prior robust state estimate in step 5.2 of Algorithm 3.3. Then, we substitute $\hat{x}_{k+1|k}$ back into (3.129) and isolate $\hat{x}_{k|k}$ to obtain

$$\hat{x}_{k|k} = \left(P_{k|k-1}^{-1} + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + \bar{\boldsymbol{C}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k \right)^{-1} \times \left(P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}_k^T \widehat{R}_k^{-1} y_k - \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k u_k \right),$$

which is the equation for computing the filtered robust state estimate in step 4.2 of Algorithm 3.3.

Lastly, assuming a deterministic context, we use equation (3.58) to obtain the error weighting matrices associated with $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$. Thus, substituting the mappings (3.125) and (3.126), and the modified matrices (3.127) into (3.124) gives[4]

$$\begin{bmatrix} P_{k|k} & * \\ * & P_{k+1|k} \end{bmatrix} = \underbrace{\begin{bmatrix} P_{k|k-1}^{-1} + \widehat{F}_k^T \widehat{Q}_k^{-1} \widehat{F}_k + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k + \bar{\boldsymbol{C}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k & -\widehat{F}_k^T \widehat{Q}_k^{-1} \\ -\widehat{Q}_k^{-1} \widehat{F}_k & \widehat{Q}_k^{-1} \end{bmatrix}^{-1}}_{\mathcal{M}^{-1}} =: \begin{bmatrix} \mathcal{M}_1 & \mathcal{M}_2 \\ \mathcal{M}_2^T & \mathcal{M}_3 \end{bmatrix}^{-1},$$

where we define the partitioned matrix $\mathcal{M}$. To find its inverse, we use the Banachiewicz inversion formula (Lemma A.4, item **(ii)**). According to Lemma A.3, the Schur complement of $\mathcal{M}_3$ in $\mathcal{M}$ is

$$(\mathcal{M}/\mathcal{M}_3) = \mathcal{M}_1 - \mathcal{M}_2 \mathcal{M}_3^{-1} \mathcal{M}_2^T = P_{k|k-1}^{-1} + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + \bar{\boldsymbol{C}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k.$$

4 The elements marked with $*$ are byproducts with no particular meaning in our context.

Thus, the posterior error weighting matrix in step 4.1 of Algorithm 3.3 is obtained as follows:

$$P_{k|k} = (\mathcal{M}/\mathcal{M}_3)^{-1} = \left(P_{k|k-1}^{-1} + \widehat{C}_k^T \widehat{R}_k^{-1} \widehat{C}_k + \bar{\boldsymbol{C}}_k^T \bar{R}_k^{-1} \bar{\boldsymbol{C}}_k + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k\right)^{-1}.$$

Finally, we obtain the predicted prior error weighting matrix

$$P_{k+1|k} = \mathcal{M}_3^{-1} + \mathcal{M}_3^{-1}\mathcal{M}_2^T(\mathcal{M}/\mathcal{M}_3)^{-1}\mathcal{M}_2\mathcal{M}_3^{-1} = \widehat{F}_k P_{k|k} \widehat{F}_k^T + \widehat{Q}_k,$$

as shown in step 5.1 of Algorithm 3.3. □

As in the Robust Kalman Filter for systems subject to norm-bounded uncertainties (Section 3.2.3), here we consider that the penalty parameter μ assumes a finite value, which we can tune to adjust the filter estimation performance, as explained in Remark 3.8. Due to the approximation of the $\hat{\lambda}$ parameter, the Polytopic Robust Kalman Filter also depends on the ξ parameter, which is usually chosen as a value within the interval $(0, 1)$.

Remark 3.9. *The expressions for the Polytopic Robust Kalman Filter in Algorithm 3.3 resemble those of the Nominal Kalman Filter outlined in Algorithm 3.1. In fact, if there are no uncertainties, i.e.,* $\bar{\boldsymbol{F}}_k$, $\bar{\boldsymbol{G}}_k$, $\bar{\boldsymbol{H}}_k$, $\bar{\boldsymbol{C}}_k$, *and* $\bar{\boldsymbol{D}}_k$ *are all zero, and we let* $\mu \to \infty$, *we have that* $\widehat{Q}_k = H_{0,k} Q_k H_{0,k}^T$, $\widehat{R}_k = D_{0,k} R_k D_{0,k}^T$, $\widehat{F}_k = F_{0,k}$, $\widehat{G}_k = G_{0,k}$, *and* $\widehat{C}_k = C_{0,k}$. *This way, the expressions in steps 4 and 5 of Algorithm 3.3 collapse to the same expressions in steps 3 and 4 of Algorithm 3.1.*

To conclude this section, we reiterate the importance of the penalty function method in the development of the proposed robust filter. This strategy enabled us to consider polytopic uncertainties in all parameter matrices of the target system and sensing models, as well as to collectively weight all the vertices with a single parameter. Moreover, the penalty parameter can be carefully adjusted to improve the filter estimation performance, according to the level of uncertainty. Finally, the PRKF is a recursive estimator and does not depend on the solution of LMIs or the use of numerical solvers, which is a valuable computational advantage in online applications.

3.3.4 Stability Analysis

This section addresses the stability properties and estimation error variance boundedness of the proposed Polytopic Robust Kalman Filter. Following a procedure similar to the one shown in Section 3.2.4, we study the steady-state behavior of Algorithm 3.3, considering that the system model parameters are time-invariant and there is no input u_k. Nevertheless, we maintain the assumption that the polytope coefficients α_k are time-varying. Thus, consider the following discrete-time uncertain linear system:

$$\begin{aligned} x_{k+1} &= (F_0 + \delta F_k)x_k + (H_0 + \delta H_k)w_k, \\ y_k &= (C_0 + \delta C_k)x_k + (D_0 + \delta D_k)v_k, \end{aligned} \tag{3.132}$$

for $k \geq 0$, with time-varying parametric uncertainties bounded by the convex polyhedron

$$\mathbb{V}_k = \left\{ \left(\delta F_k,\, \delta H_k,\, \delta C_k,\, \delta D_k \right) = \sum_{\nu=1}^{V} \alpha_{\nu,k} \left(F_\nu,\, H_\nu,\, C_\nu,\, D_\nu \right) \right\}, \tag{3.133}$$

where $\alpha_k = \left[\alpha_{1,k} \cdots \alpha_{V,k} \right]^T$ belongs to the unit simplex Λ_V in (3.85), with V vertices.

First, let us investigate the stability conditions of the PRKF in Algorithm 3.3. Considering the uncertain system model (3.132)-(3.133), the polytopic robust filter equations in steps 4 and 5 of Algorithm 3.3 become:

$$P_{k|k} = \left(P_{k|k-1}^{-1} + \widehat{C}^T \widehat{R}^{-1} \widehat{C} + \bar{\boldsymbol{C}}^T \bar{R}^{-1} \bar{\boldsymbol{C}} + \bar{\boldsymbol{F}}^T \bar{Q}^{-1} \bar{\boldsymbol{F}} \right)^{-1}, \tag{3.134}$$

$$\hat{x}_{k|k} = P_{k|k} \left(P_{k|k-1}^{-1} \hat{x}_{k|k-1} + \widehat{C}^T \widehat{R}^{-1} y_k \right), \tag{3.135}$$

$$P_{k+1|k} = \widehat{F} P_{k|k} \widehat{F}^T + \widehat{Q}, \tag{3.136}$$

$$\hat{x}_{k+1|k} = \widehat{F} \hat{x}_{k|k}, \tag{3.137}$$

in which the modified model parameter matrices are given by the corresponding equations listed in step 3 of Algorithm 3.3, assuming constant parameters. To simplify the analysis, we also define the augmented matrices

$$\widetilde{C} := \begin{bmatrix} \widehat{C} \\ \bar{\boldsymbol{C}} \\ \bar{\boldsymbol{F}} \end{bmatrix} \quad \text{and} \quad \widetilde{R} := \begin{bmatrix} \widehat{R} & 0 & 0 \\ 0 & \bar{R} & 0 \\ 0 & 0 & \bar{Q} \end{bmatrix}.$$

Then, we can rewrite $P_{k|k}$ in (3.134) in a more compact way, as follows:

$$P_{k|k} = \left(P_{k|k-1}^{-1} + \widetilde{C}^T \widetilde{R}^{-1} \widetilde{C} \right)^{-1} = P_{k|k-1} - P_{k|k-1} \widetilde{C}^T \left(\widetilde{R} + \widetilde{C} P_{k|k-1} \widetilde{C}^T \right)^{-1} \widetilde{C} P_{k|k-1}, \tag{3.138}$$

where we applied Lemma A.1 to further expand the expression. Now, combining (3.138) with (3.135) and substituting back into (3.137), we obtain the steady-state predicted robust state estimate

$$\hat{x}_{k+1|k} = \widetilde{F}_k \hat{x}_{k|k-1} + \widetilde{F}_k P_{k|k-1} \widehat{C}^T \widehat{R}^{-1} y_k, \tag{3.139}$$

where

$$\widetilde{F}_k = \widehat{F} \left(I_n - P_{k|k-1} \widetilde{C}^T \left(\widetilde{R} + \widetilde{C} P_{k|k-1} \widetilde{C}^T \right)^{-1} \widetilde{C} \right)$$

is the filter closed-loop matrix. Then, we substitute $P_{k|k}$ from (3.138) into (3.136) to obtain the expression for the predicted prior error weighting matrix:

$$P_{k+1|k} = \widehat{F} \left(P_{k|k-1} - P_{k|k-1} \widetilde{C}^T \left(\widetilde{R} + \widetilde{C} P_{k|k-1} \widetilde{C}^T \right)^{-1} \widetilde{C} P_{k|k-1} \right) \widehat{F}^T + \widehat{Q}. \tag{3.140}$$

In the following theorem, we establish a result about the conditions for convergence of the proposed robust filter to a stable steady-state filter.

Theorem 3.7. *Consider the linear system model* (3.132) *with polytopic uncertainties* (3.133) *and the corresponding robust filter* (3.139)-(3.140)*. Assume that* $\{\widehat{F}, \widetilde{C}\}$ *is* detectable *and* $\{\widehat{F}, \widehat{Q}^{1/2}\}$ *is* controllable*. Then, for any initial condition* $P_{0|-1} \succ 0$, $\xi > 0$, *and* $\mu > 0$, $P_{k+1|k}$ *converges to the* unique stabilizing solution $P \succ 0$ *of the algebraic Riccati equation*

$$P = \widehat{F}\Big(P - P\widetilde{C}^T\big(\widetilde{R} + \widetilde{C}P\widetilde{C}^T\big)^{-1}\widetilde{C}P\Big)\widehat{F}^T + \widehat{Q}. \tag{3.141}$$

The solution P *is stabilizing in the sense that the steady-state filter closed-loop matrix*

$$\widetilde{F} = \widehat{F}\Big(I_n - P\widetilde{C}^T\big(\widetilde{R} + \widetilde{C}P\widetilde{C}^T\big)^{-1}\widetilde{C}\Big) \tag{3.142}$$

is Schur stable.

Proof. The conditions $\xi > 0$ and $\mu > 0$ imply that $\varphi > 0$, such that matrices $\widehat{F}$, $\widetilde{C}$, $\widetilde{R}$, and $\widehat{Q}$ are well-defined. Moreover, from Kailath, Sayed and Hassibi (2000b), we have that detectability of $\{\widehat{F}, \widetilde{C}\}$ and controllability of $\{\widehat{F}, \widehat{Q}^{1/2}\}$ guarantee the convergence of $P_{k+1|k}$ in (3.140) to the unique stabilizing solution $P \succ 0$ of the algebraic Riccati equation (3.141) that stabilizes (3.142), which is the polytopic robust filter steady-state closed-loop matrix. □

Now, let us establish the conditions for the boundedness of the estimation error variance of the proposed robust filter. Again, consider the uncertain linear discrete-time system model (3.132)-(3.133). Note that we can write the polytopic uncertainties described in (3.133) alternatively as

$$\begin{aligned}
\begin{bmatrix}\delta F_k & \delta H_k\end{bmatrix} &= \begin{bmatrix}I_n & \cdots & I_n\end{bmatrix}\begin{bmatrix}\alpha_{1,k}I_n & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \alpha_{V,k}I_n\end{bmatrix}\begin{bmatrix}F_1 & H_1\\ \vdots & \vdots\\ F_V & H_V\end{bmatrix} =: M_1\,\bar{\boldsymbol{\alpha}}_{1,k}\begin{bmatrix}\bar{\boldsymbol{F}} & \bar{\boldsymbol{H}}\end{bmatrix},\\
\begin{bmatrix}\delta C_k & \delta D_k\end{bmatrix} &= \begin{bmatrix}I_r & \cdots & I_r\end{bmatrix}\begin{bmatrix}\alpha_{1,k}I_r & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \alpha_{V,k}I_r\end{bmatrix}\begin{bmatrix}C_1 & D_1\\ \vdots & \vdots\\ C_V & D_V\end{bmatrix} =: M_2\,\bar{\boldsymbol{\alpha}}_{2,k}\begin{bmatrix}\bar{\boldsymbol{C}} & \bar{\boldsymbol{D}}\end{bmatrix},
\end{aligned} \tag{3.143}$$

in which, since $\alpha_k = \begin{bmatrix}\alpha_{1,k}\cdots\alpha_{V,k}\end{bmatrix}^T$ belongs to the unit simplex Λ_V in (3.85), we have that $\|\bar{\boldsymbol{\alpha}}_{1,k}\| \le 1$ and $\|\bar{\boldsymbol{\alpha}}_{2,k}\| \le 1$. Moreover, we assume that w_k and v_k are uncorrelated zero-mean Gaussian noise processes with joint covariance matrix

$$\mathscr{Q} = \boldsymbol{E}\left\{\begin{bmatrix}w_k\\ v_k\end{bmatrix}\begin{bmatrix}w_k^T & v_k^T\end{bmatrix}\right\} = \begin{bmatrix}Q & 0\\ 0 & R\end{bmatrix} \succ 0. \tag{3.144}$$

In addition, assume that there is no correlation between the parametric uncertainties and the system and measurement noises.

Recall Definition 3.1 of quadratic stability for systems with norm-bounded uncertainties. In the following, we make an adaptation of this definition, considering the alternative representation of the polytopic uncertainties in (3.143).

Definition 3.2. *The uncertain system* (3.132)-(3.143) *is* quadratically stable *if there exists a symmetric positive definite matrix* U *such that*

$$\left(F_0 + M_1\bar{\boldsymbol{\alpha}}_{1,k}\bar{\boldsymbol{F}}\right)^T U\left(F_0 + M_1\bar{\boldsymbol{\alpha}}_{1,k}\bar{\boldsymbol{F}}\right) - U \prec 0$$

for all admissible $\bar{\boldsymbol{\alpha}}_{1,k}$.

Remark 3.10. *Similar to Remark 3.4, we can also conversely say that the uncertain system* (3.132)-(3.143) *is quadratically stable if, and only if*

1. F_0 *is Schur stable;*
2. *The discrete-time* $\mathcal{H}_\infty$ *normal bound* $\left\|\bar{\boldsymbol{F}}(zI_n - F_0)^{-1}M_1\right\|_\infty < 1$[5] *is satisfied.*

Now, we make the following assumptions about the uncertain system and the robust filter to show that it presents a bounded steady-state estimation error variance.

Assumption 3.3. *The uncertain system* (3.132)-(3.143) *is quadratically stable, according to Definition 3.2.*

Assumption 3.4. *The conditions of Theorem 3.7 are satisfied, meaning that the polytopic robust filter steady-state closed-loop matrix* $\tilde{F}$ *is Schur stable.*

First, we show that if Assumption 3.3 and Assumption 3.4 are satisfied, the steady-state polytopic robust filter (3.139) is also quadratically stable. To simplify the notation, we define the steady-state filter gain, as follows:

$$\widetilde{K} \coloneqq \widetilde{F}P\widehat{C}^T\widehat{R}^{-1},$$

where $\widetilde{F}$ is given by (3.142) and P is the stabilizing solution of the algebraic Riccati equation (3.141). Thus, the steady-state robust filter equation becomes

$$\hat{x}_{k+1|k} = \widetilde{F}\hat{x}_{k|k-1} + \widetilde{K}y_k. \tag{3.145}$$

Now, we substitute y_k from (3.132) into (3.145), such that

$$\hat{x}_{k+1|k} = \widetilde{F}\hat{x}_{k|k-1} + \widetilde{K}(C_0 + \delta C_k)x_k + \widetilde{K}(D_0 + \delta D_k)v_k. \tag{3.146}$$

In addition, we define the state estimation error vector $e_k \coloneqq x_k - \hat{x}_{k|k-1}$. Then, subtracting (3.146) from x_{k+1} in (3.132) yields

$$e_{k+1} = \left[(F_0 - \widetilde{F} - \widetilde{K}C_0) + (\delta F_k - \widetilde{K}\delta C_k)\right]x_k + \widetilde{F}e_k + (H_0 + \delta H_k)w_k - \widetilde{K}(D_0 + \delta D_k)v_k. \tag{3.147}$$

[5] $\|\cdot\|_\infty$ denotes the maximum singular value of its argument for values of z on the unit disk.

We also introduce an augmented system comprised of the system state x_k and the estimation error e_k. Hence, from (3.132), (3.143), and (3.147), this augmented system is given by

$$\begin{aligned} \zeta_{k+1} &= (\mathscr{F} + \delta\mathscr{F}_k)\zeta_k + (\mathscr{H} + \delta\mathscr{H}_k)\eta_k, \\ \begin{bmatrix}\delta\mathscr{F}_k & \delta\mathscr{H}_k\end{bmatrix} &= \mathscr{M}\,\bar{\boldsymbol{\alpha}}_k \begin{bmatrix}\bar{\mathscr{F}} & \bar{\mathscr{H}}\end{bmatrix}, \end{aligned} \tag{3.148}$$

where

$$\zeta_k := \begin{bmatrix} x_k \\ e_k \end{bmatrix}, \quad \eta_k := \begin{bmatrix} w_k \\ v_k \end{bmatrix}, \quad \mathscr{F} := \begin{bmatrix} F_0 & 0 \\ F_0 - \widetilde{F} - \widetilde{K}C_0 & \widetilde{F} \end{bmatrix}, \quad \mathscr{H} := \begin{bmatrix} H_0 & 0 \\ H_0 & -\widetilde{K}D_0 \end{bmatrix},$$
$$\mathscr{M} := \begin{bmatrix} M_1 & 0 \\ M_1 & -\widetilde{K}M_2 \end{bmatrix}, \quad \bar{\boldsymbol{\alpha}}_k := \begin{bmatrix} \bar{\boldsymbol{\alpha}}_{1,k} & 0 \\ 0 & \bar{\boldsymbol{\alpha}}_{2,k} \end{bmatrix}, \quad \bar{\mathscr{F}} := \begin{bmatrix} \bar{\boldsymbol{F}} & 0 \\ \bar{\boldsymbol{C}} & 0 \end{bmatrix}, \quad \bar{\mathscr{H}} := \begin{bmatrix} \bar{\boldsymbol{H}} & 0 \\ 0 & \bar{\boldsymbol{D}} \end{bmatrix}.$$

Lemma 3.6. *If Assumption 3.3 and Assumption 3.4 are satisfied, then the augmented system* (3.148) *is quadratically stable.*

Proof. The augmented system matrix $\mathscr{F}$ is lower triangular with diagonal elements F_0 and $\widetilde{F}$, which are both Schur stable. Hence, $\mathscr{F}$ is also Schur stable. In addition, we have that

$$\begin{aligned} \bar{\mathscr{F}}(zI_{2n} - \mathscr{F})^{-1}\mathscr{M} &= \begin{bmatrix} \bar{\boldsymbol{F}} & 0 \\ \bar{\boldsymbol{C}} & 0 \end{bmatrix} \begin{bmatrix} zI_n - F_0 & 0 \\ -(F_0 - \widetilde{F} - \widetilde{K}C_0) & zI_n - \widetilde{F} \end{bmatrix}^{-1} \begin{bmatrix} M_1 & 0 \\ M_1 & -\widetilde{K}M_2 \end{bmatrix} \\ &= \begin{bmatrix} \bar{\boldsymbol{F}}(zI_n - F_0)^{-1}M_1 & 0 \\ \bar{\boldsymbol{C}}(zI_n - F_0)^{-1}M_1 & 0 \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{F}} \\ \bar{\boldsymbol{C}} \end{bmatrix} (zI_n - F_0)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix}. \end{aligned}$$

Also, note that

$$F_0 + M_1\bar{\boldsymbol{\alpha}}_{1,k}\bar{\boldsymbol{F}} = F_0 + \begin{bmatrix} M_1 & 0 \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\alpha}}_{1,k} & 0 \\ 0 & \bar{\boldsymbol{\alpha}}_{2,k} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{F}} \\ \bar{\boldsymbol{C}} \end{bmatrix}.$$

Given that system (3.132)-(3.143) is quadratically stable, according to Remark 3.10, we have

$$\left\| \begin{bmatrix} \bar{\boldsymbol{F}} \\ \bar{\boldsymbol{C}} \end{bmatrix} (zI_n - F_0)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix} \right\|_\infty < 1,$$

for all admissible contractions $\bar{\boldsymbol{\alpha}}_{1,k}$ and $\bar{\boldsymbol{\alpha}}_{2,k}$. In consequence, $\left\|\bar{\mathscr{F}}(zI_{2n} - \mathscr{F})^{-1}\mathscr{M}\right\|_\infty < 1$ and the augmented system (3.148) is also quadratically stable. □

Now, define the covariance matrix of the augmented system state as $\mathscr{P}_k := \boldsymbol{E}\left\{\zeta_k\zeta_k^T\right\}$. Then, it follows from (3.148) that $\mathscr{P}_k$ satisfies the Lyapunov recursion

$$\mathscr{P}_{k+1} = (\mathscr{F} + \delta\mathscr{F}_k)\mathscr{P}_k(\mathscr{F} + \delta\mathscr{F}_k)^T + (\mathscr{H} + \delta\mathscr{H}_k)\mathscr{Q}(\mathscr{H} + \delta\mathscr{H}_k)^T, \tag{3.149}$$

with $\mathscr{Q}$ as defined in (3.144). In the following theorem, we provide a result on the boundedness of the steady-state estimation error variance of the proposed polytopic robust filter.

Theorem 3.8. *Consider that Assumption 3.3 and Assumption 3.4 are satisfied. Then, the state estimation error variance of the steady-state polytopic robust filter* (3.145) *satisfies*

$$\lim_{k\to\infty} \boldsymbol{E}\Big\{e_k e_k^T\Big\} \preceq \mathscr{V}_{22},$$

where $\mathscr{V}_{22}$ *is the* $(2,2)$ *block entry with the smallest trace among all* $(2,2)$ *block entries of matrices* $\mathscr{V} \succ 0$ *that satisfy the inequality*

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{F}})\,\mathscr{V}\,(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{F}})^T + (\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{H}})\,\mathscr{Q}\,(\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{H}})^T - \mathscr{V} \preceq 0,$$

for all admissible $\bar{\boldsymbol{\alpha}}$, *with* $\|\bar{\boldsymbol{\alpha}}\| \leq 1$.

Proof. From Lemma 3.6, we have that the augmented system (3.148) is quadratically stable, then, according to Definition 3.2, there exists a matrix $\mathscr{U} \succ 0$ such that

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})\,\mathscr{U}\,(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T - \mathscr{U} \prec 0,$$

for any admissible $\bar{\boldsymbol{\alpha}}_k$. Based on Petersen and McFarlane (1996) and Sayed (2001), the existence of such a matrix $\mathscr{U} \succ 0$ guarantees the existence of a sufficiently large scaling parameter $\epsilon > 0$, such that one can find a matrix $\mathscr{V} = \epsilon\,\mathscr{U}$ that satisfies

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})\,\mathscr{V}\,(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T + (\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{H}})\,\mathscr{Q}\,(\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{H}})^T \preceq \mathscr{V}.$$

Subtracting the recursion for the augmented system covariance (3.149) from the above inequality then gives

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})(\mathscr{V} - \mathscr{P}_k)(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T \preceq \mathscr{V} - \mathscr{P}_{k+1},$$

or, equivalently,

$$\mathscr{V} - \mathscr{P}_{k+1} = (\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})(\mathscr{V} - \mathscr{P}_k)(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T + \mathscr{W}_k,$$

for some $\mathscr{W}_k \succeq 0$. To conclude, since the augmented system is quadratically stable, as $k \to \infty$, we have that $\mathscr{V} - \mathscr{P}_{k+1} \succeq 0$, or $\mathscr{P}_{k+1} \preceq \mathscr{V}$. The $(2,2)$ block entry of $\mathscr{P}_k$ corresponds to the estimation error variance, which is therefore bounded. □

3.3.5 Illustrative Example

In this section, we evaluate the performance of the proposed Polytopic Robust Kalman Filter with a numerical example. We also compare the results with other existing polytopic robust filter from the literature, as well as with the Nominal Kalman filter.

Consider a discrete-time linear system with polytopic uncertainties, as described in (3.83)-(3.84), with the following constant nominal parameter matrices (adapted from Xie, Soh and Souza (1994)):

$$F_{0,k} = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix}, \quad G_{0,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad H_{0,k} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}, \quad C_{0,k} = \begin{bmatrix} -100 & 10 \end{bmatrix}, \quad D_{0,k} = 1$$

and uncertainties bounded by a 2-vertex polytope given by

$$F_{1,k} = \begin{bmatrix} 0 & 0 \\ 0.1 & 0.3 \end{bmatrix}, \quad G_{1,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad H_{1,k} = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}, \quad C_{1,k} = \begin{bmatrix} 0.1 & 0.3 \end{bmatrix}, \quad D_{1,k} = 0.1,$$
$$\left(F_{2,k},\, G_{2,k},\, H_{2,k},\, C_{2,k},\, D_{2,k}\right) = -\left(F_{1,k},\, G_{1,k},\, H_{1,k},\, C_{1,k},\, D_{1,k}\right).$$

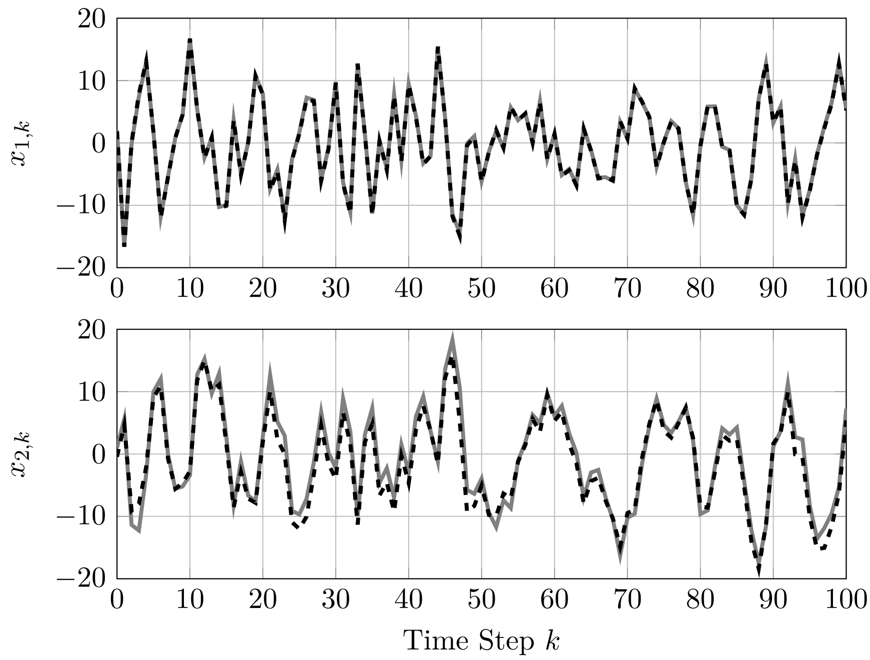
There is no input signal u_k and the system and measurement noises, w_k and v_k, are mutually independent zero-mean white Gaussian signals with variances $Q_k = 1$ and $R_k = 1$. The initial state is $x_0 = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$.

Then, we apply the PRKF (Algorithm 3.3) with the following initialization data:

$$\hat{x}_{0|-1} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \quad P_{0|-1} = I_2, \quad \mu = 1, \text{ and } \xi = 0.01.$$

Figure 5 depicts the evolution of the true target system state along with the estimation performed by the PRKF. At each time step, the coefficients $\alpha_k \in \Lambda_2$ (see (3.85)) are randomly selected. The results show that, despite the presence of polytopic uncertainties in both the target system and sensing models, the proposed PRKF can successfully estimate the state of the system.

Figure 5 – Actual (solid lines) and estimated (dashed lines) target system state obtained with the proposed PRKF (Algorithm 3.3).



We further evaluate the proposed PRKF by comparing its performance with some other existing polytopic robust filtering strategies. Namely, the robust $\mathcal{H}_\infty$ filters proposed by Chang, Park and Tang (2015), Gershon and Shaked (2015), Morais *et al.* (2017), and Gershon and Shaked (2020), as well as the $\mathcal{H}_2$ filter from Gershon and Shaked (2020). We also consider the Nominal Kalman filter (Algorithm 3.1) as a baseline. Like the proposed

PRKF, all of the robust filters in consideration also assume polytopic uncertainties in all parameter matrices. However, unlike the PRKF, they assume time-invariant polytope vertices.

The simulation consists of performing $M = 5000$ Monte Carlo experiments, each with time horizon $N = 1000$. At each time step k, we compute the mean squared estimation error (MSE), averaged over all experiments, as follows:

$$\mathrm{MSE}_k = \frac{1}{M}\sum_{e=1}^{M} \|x_k - \hat{x}_{k|k,e}\|^2,$$

which, as commented in Sayed (2001) and in Section 3.2.5, is a reasonable approximation of the estimation error variance, as it cannot be analytically computed due to the model uncertainties.

The simulation results are presented in Figure 6 and are also summarized in Table 2, which shows the mean $\overline{\mathrm{MSE}}$ and standard deviation $\sigma(\mathrm{MSE})$ of the estimation error variances, respectively computed as

$$\overline{\mathrm{MSE}} = \sum_{k=0}^{N} \frac{\mathrm{MSE}_k}{N+1} \quad \text{and} \quad \sigma^2(\mathrm{MSE}) = \sum_{k=0}^{N} \frac{(\mathrm{MSE}_k - \overline{\mathrm{MSE}})^2}{N+1}.$$

Figure 6 – Estimation error variance curves of the polytopic robust filters.

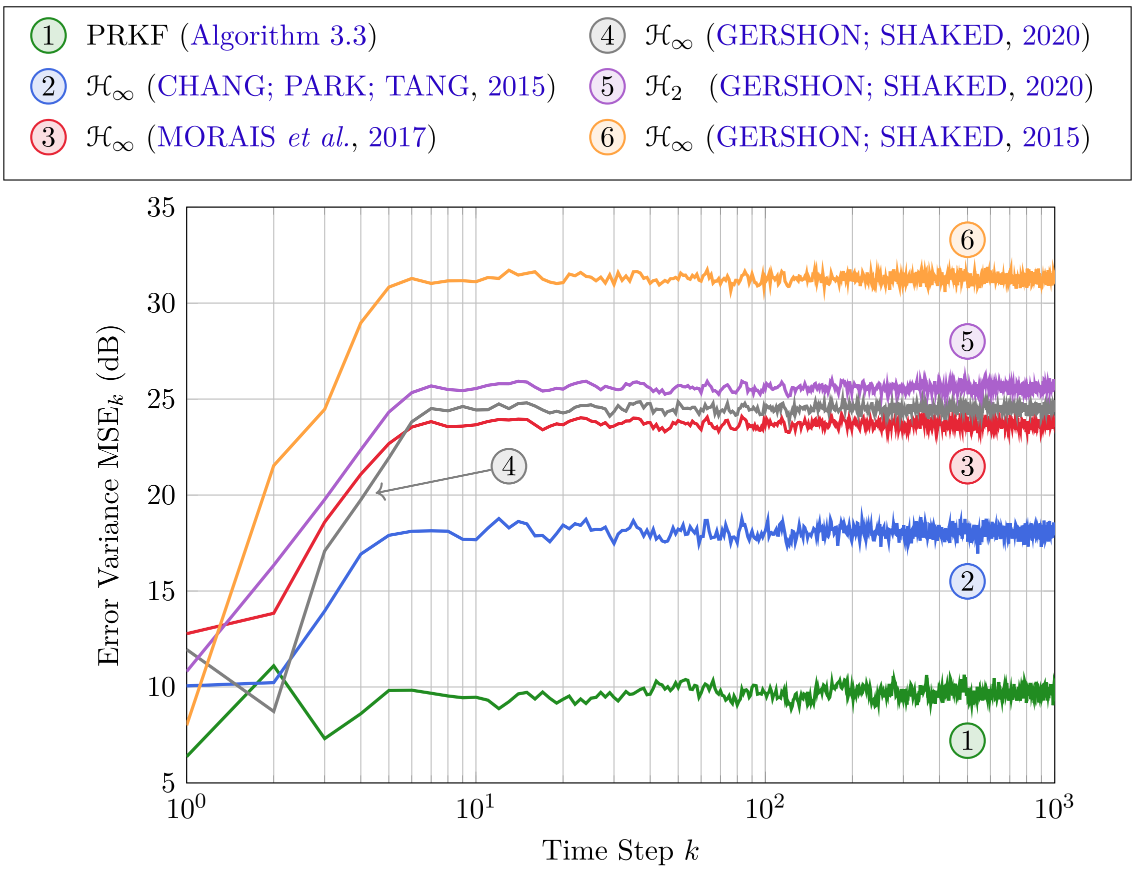


Table 2 – Estimation performance of each polytopic robust filter.

Filter	$\overline{\text{MSE}}$ (dB)	σ(MSE) (dB)	Δt_{comp} (ms)
① PRKF (Algorithm 3.3)	**9.705**	**0.3154**	0.8914
② $\mathcal{H}_\infty$ (CHANG; PARK; TANG, 2015)	18.04	0.4615	106.28
③ $\mathcal{H}_\infty$ (MORAIS *et al.*, 2017)	23.68	0.5307	78.591
④ $\mathcal{H}_\infty$ (GERSHON; SHAKED, 2020)	24.47	0.7211	68.729
⑤ $\mathcal{H}_2$ (GERSHON; SHAKED, 2020)	25.57	0.6152	168.99
⑥ $\mathcal{H}_\infty$ (GERSHON; SHAKED, 2015)	31.23	1.1380	66.360
KF (Algorithm 3.1)	34.17	6.8900	**0.4201**

Bold numbers indicate the smallest values.

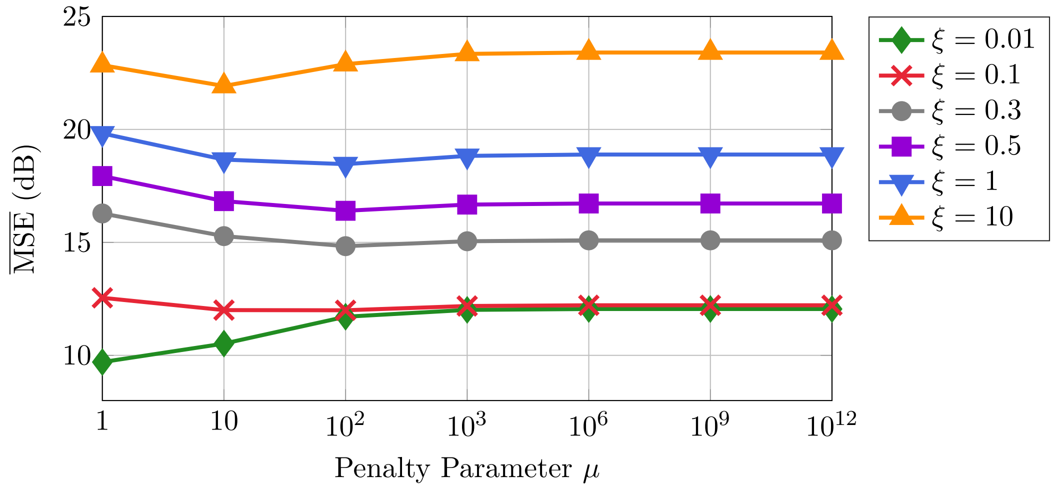
Since all of the simulated robust $\mathcal{H}_2$ and $\mathcal{H}_\infty$ filters have a similar structure, with constant design matrices computed offline, we measured the time each one demands to obtain these matrices, which result from the solution of LMI-based optimization problems. For comparison purposes, we also measure the time needed to compute the constant parts of the PRKF and KF, as the example system model parameter matrices are time-invariant. The results are shown in the column Δt_{comp} in Table 2. The simulation was performed on a 2.3 GHz i7-12700H CPU with 32 GB of RAM using MATLAB R2022b, the YALMIP toolbox (LÖFBERG, 2004), and the SeDuMi solver (STURM, 1999).

The simulation results in Figure 6 and Table 2 indicate that the proposed PRKF outperforms all the other robust filtering strategies in terms of error variance. The Nominal Kalman Filter, however, was unable to estimate the system state, presenting an exponentially increasing error variance. For this reason, it is not shown in Figure 6. This corroborates the fact that parametric uncertainties can indeed significantly degrade its performance. The $\mathcal{H}_2$ and $\mathcal{H}_\infty$ filters exhibit considerably larger error variances compared to the PRKF. For instance, the robust $\mathcal{H}_\infty$ filter of Chang, Park and Tang (2015) presents a mean error variance twice as large as that of the PRKF. The $\mathcal{H}_\infty$ filters of Morais *et al.* (2017) and Gershon and Shaked (2020) show similar results, closely followed by the $\mathcal{H}_2$ filter of Gershon and Shaked (2020). The $\mathcal{H}_\infty$ filter in Gershon and Shaked (2015) exhibits the largest error variance among the estimators. In terms of standard deviation, all of the robust filters present similar results, with the PRKF being the smallest among them.

Naturally, the Nominal Kalman Filter requires the least amount of computation time to obtain its results, but it is closely followed by the PRKF, as Table 2 shows. On the other hand, the other robust filtering approaches depend on the solution of optimization problems subject to LMIs, usually one for each vertex of the polytope, which requires more computational effort. This explains their significantly larger computation times. Therefore, the recursive and analytic expressions of the PRKF yield a satisfactory trade-off between estimation performance and computational cost, being thus suitable for online applications.

To conclude the evaluation, we further study how the two parameters of the PRKF, namely the penalty parameter μ and the approximation parameter ξ, influence the filter performance. Figure 7 compiles the results of a series of simulations with several combinations of these parameters. For each combination, we compute the mean estimation error variance $\overline{\text{MSE}}$, as previously described. As commented in Section 3.3.3, choosing $0 < \xi < 1$ generally yields better results. In addition, within this range, we found that smaller values of μ lead to smaller mean error variances. Above this range, the filter performance experiences some degradation.

Figure 7 – Effect of the PRKF parameters μ and ξ on the mean error variance $\overline{\text{MSE}}$.



Chapter 4

Robust Distributed Kalman Filtering

In this chapter, we discuss the distributed filtering problem for discrete-time linear systems in the context of sensor networks. As in the previous chapter, we also divide it into three sections. In the first section, we assume exact knowledge of the underlying target system and sensor models. This simpler setting will lay the foundation to the other two sections, in which we address the cases where the models are subject to norm-bounded and polytopic parametric uncertainties. As we previously pointed out, these uncertainties are usually unavoidable in practical systems, arising due to factors like unmodeled dynamics, linearization, model reduction, and varying parameters. Moreover, they can appreciably degrade the estimation performance if not taken into consideration.

Sensor networks are composed of nodes that have sensing, computing, and communication capabilities. In the distributed filtering context, these sensors observe a target system and exchange information to estimate the target system state. In general, the use of multiple sensors can significantly improve the estimation accuracy. Furthermore, it provides more flexibility and reliability to the overall system. Many distributed filtering strategies in the literature are based on the combination of the Kalman filter (KALMAN, 1960) with the average consensus protocol (Section 2.4). We also take advantage of this successful combination in this chapter, overcoming one of the main shortcomings of the Kalman filter by compensating for model parametric uncertainties with robust estimators.

In each of the upcoming sections, we obtain centralized and distributed versions of the filters proposed in the previous chapter. First, we formulate the corresponding centralized estimation problems, assuming access to all sensors in the network at once. These centralized estimation problems are also built from a deterministic point of view (BRYSON; HO, 1975), as regularized least-squares estimation problems, thoroughly discussed in Chapter 3. Then, we employ the hybrid consensus on measurement and information (HCMCI) approach (BATTISTELLI *et al.*, 2015) to derive distributed implementations of the corresponding centralized estimators. With enough consensus iterations, these distributed filters become reasonable approximations of their centralized counterparts.

4.1 Nominal Distributed Kalman Filtering

In this section, we extend the Nominal Kalman Filter introduced in Section 3.1 to the multiple sensor case. As the nominal denomination suggests, we assume perfect target system and sensing models. Before dealing with the distributed estimation problem, we first tackle the centralized scenario, in which a fusion center has access to measurements from all sensors in the network. Following the framework developed in Section 3.1, we formulate the centralized estimation problem in a deterministic manner, as a regularized least-squares estimation problem (Section 3.1.2), whose solution yields the so-called Nominal Centralized Kalman Filter (CKF). Then, by taking advantage of the HCMCI protocol (BATTISTELLI *et al.*, 2015), we derive a distributed variant of the CKF, called Nominal Distributed Kalman Consensus Filter (DKCF). We show that, for a large enough number of consensus steps, the DKCF approaches the behavior of the CKF. Since the centralized and distributed filters are based on the Nominal Kalman Filter, we present both as recursive correction-prediction algorithms. The section concludes with a stability analysis of both proposed estimators, assuming a time-invariant model.

4.1.1 Problem Formulation

4.1.1.1 System Model

Consider a sensor network featuring S sensors. The communication among them is represented by the undirected graph $\mathbb{G} = (\mathbb{S}, \mathbb{E})$, with node set $\mathbb{S} = \{1, 2, \ldots, S\}$ and edge set $\mathbb{E} \subseteq \mathbb{S} \times \mathbb{S}$. The neighborhood of a sensor i is denoted by $\mathcal{N}_i = \{j \in \mathbb{S} \,|\, (i, j) \in \mathbb{E}\}$ and has cardinality N_i (see Section 2.3 for an introduction on graph theory).

Assumption 4.1. *The undirected graph $\mathbb{G}$ has a fixed topology and is connected, i.e., there is a path between every pair of nodes.*

Consider the following discrete-time state-space description of a linear target dynamical system:

$$x_{k+1} = F_k x_k + G_k u_k + H_k w_k, \tag{4.1}$$

which is observed by the set of S sensors $\mathbb{S} = \{1, 2, \ldots, S\}$, each described by the model

$$y_k^i = C_k^i x_k + D_k^i v_k^i, \quad \forall i \in \mathbb{S}, \tag{4.2}$$

for $k = 0, 1, \ldots, N$, with state vector $x_k \in \mathbb{R}^n$, input vector $u_k \in \mathbb{R}^m$, and system noise vector $w_k \in \mathbb{R}^p$. For each sensor $i \in \mathbb{S}$, $y_k^i \in \mathbb{R}^r$ is the measurement vector and $v_k^i \in \mathbb{R}^q$ is the measurement noise. $F_k \in \mathbb{R}^{n\times n}$, $G_k \in \mathbb{R}^{n\times m}$, $H_k \in \mathbb{R}^{n\times p}$, $C_k^i \in \mathbb{R}^{r\times n}$, and $D_k^i \in \mathbb{R}^{r\times q}$ are known nominal parameter matrices.

In a stochastic setting, we usually assume that x_0, w_k, and $\{v_k^i\}_{i=1}^S$ are mutually independent zero-mean Gaussian random variables with respective variances

$$\boldsymbol{E}\Big\{x_0 x_0^T\Big\} = P_0 \succ 0, \quad \boldsymbol{E}\Big\{w_k w_l^T\Big\} = Q_k \delta_{kl} \succ 0, \quad \text{and} \quad \boldsymbol{E}\Big\{v_k^i (v_l^j)^T\Big\} = R_k^i \delta_{kl} \delta_{ij} \succ 0,$$

where δ_{ab} is the Kronecker delta function, such that $\delta_{ab} = 1$ if $a = b$, and $\delta_{ab} = 0$ otherwise. Nonetheless, as we shall see, the strategy we adopt does not require that these variables have any particular distribution.

4.1.1.2 Nominal Centralized Estimation Problem

Before addressing the distributed estimation problem, we first design a centralized state estimator for the system (4.1)-(4.2). In a centralized setup, we assume that the measurements obtained from all sensors in the network are available to a central estimator. As the system state sequence $\{x_k\}$ is not perfectly observed, the goal is thus to leverage all the information available up to time instant k,

$$\boldsymbol{Y}_k = \Big\{\{y_0^i\}_{i=1}^S, \ldots, \{y_k^i\}_{i=1}^S, u_0, \ldots, u_k\Big\},$$

to compute a so-called filtered state estimate $\hat{x}_{k|k}^c$ of x_k, as well as a predicted estimate $\hat{x}_{k+1|k}^c$ of x_{k+1}. Here, we use the superscript c to indicate the centralized entities.

We follow the procedure reported in Section 3.1.1.2 for the Nominal Kalman Filter, where a deterministic interpretation is assumed to the stochastic estimation problem. In this context, we introduce the variables $\hat{x}_k$, $\hat{x}_{k+1}$, $\hat{w}_k$, and $\{\hat{v}_k^i\}_{i=1}^S$ as substitutes for the random variables x_k, x_{k+1}, w_k, and $\{v_k^i\}_{i=1}^S$ in the stochastic model (4.1)-(4.2). Then, assuming that at each time step k, an *a priori* state estimate $\hat{x}_{k|k-1}^c$, a set of measurements $\{y_k^i\}_{i=1}^S$, and the input u_k are available, we formulate the constrained optimization problem with a one-step quadratic objective function, as follows:

$$\begin{aligned} \min_{\substack{\hat{x}_k, \hat{x}_{k+1}, \\ \hat{w}_k, \hat{\boldsymbol{v}}_k}} \quad & J_k(\hat{x}_k, \hat{w}_k, \hat{\boldsymbol{v}}_k) = \|\hat{x}_k - \hat{x}_{k|k-1}^c\|^2_{(P_{k|k-1}^c)^{-1}} + \|\hat{w}_k\|^2_{Q_k^{-1}} + \|\hat{\boldsymbol{v}}_k\|^2_{\boldsymbol{\mathfrak{R}}_k^{-1}}, \\ \text{subject to} \quad & \begin{cases} \hat{x}_{k+1} = F_k \hat{x}_k + G_k u_k + H_k \hat{w}_k, \\ \boldsymbol{\mathfrak{Y}}_k = \boldsymbol{\mathfrak{C}}_k \hat{x}_k + \boldsymbol{\mathfrak{D}}_k \hat{\boldsymbol{v}}_k, \end{cases} \end{aligned} \tag{4.3}$$

for $k = 0, 1, \ldots, N$, where we define the aggregated vectors and matrices

$$\boldsymbol{\mathfrak{Y}}_k := \begin{bmatrix} y_k^1 \\ \vdots \\ y_k^S \end{bmatrix}, \ \hat{\boldsymbol{v}}_k := \begin{bmatrix} \hat{v}_k^1 \\ \vdots \\ \hat{v}_k^S \end{bmatrix}, \ \boldsymbol{\mathfrak{C}}_k := \begin{bmatrix} C_k^1 \\ \vdots \\ C_k^S \end{bmatrix}, \ \boldsymbol{\mathfrak{D}}_k := \begin{bmatrix} D_k^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_k^S \end{bmatrix}, \text{ and } \boldsymbol{\mathfrak{R}}_k := \begin{bmatrix} R_k^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_k^S \end{bmatrix}. \tag{4.4}$$

Note that, $\hat{w}_k$ and $\{\hat{v}_k^i\}_{i=1}^S$ are fitting errors weighted respectively by $Q_k \succ 0$ and $R_k^i \succ 0$, $\forall i \in \mathbb{S}$, and $P_{k|k-1}^c \succ 0$ weights the *a priori* estimation error $x_k - \hat{x}_{k|k-1}^c$. From a stochastic viewpoint, matrices Q_k and R_k^i represent the variances of the random variables w_k and

$\{v_k^i\}_{i=1}^S$. However, in this deterministic setting, they are rather understood as general weighting matrices.

Problem (4.3) is the special case of a regularized least-squares estimation problem, as described in Section 3.1.2. Its solution recursively provides the filtered and predicted central state estimates $\hat{x}_{k|k}^c$ and $\hat{x}_{k|k+1}^c$, respectively.

4.1.1.3 Nominal Distributed Estimation Problem

In the distributed estimation case, there is no central estimator, such that each sensor in the network should estimate the state of the target system using only its own data and information gathered from its neighbors. Hence, the goal of each sensor node $i \in \mathbb{S}$ is to obtain the best estimates $\hat{x}_{k|k}^i$ of x_k and $\hat{x}_{k+1|k}^i$ of x_{k+1}, referred to as filtered and predicted state estimates, in a distributed rather than centralized fashion.

To achieve this objective, we can leverage the distributed feature of the average consensus algorithm (Algorithm 2.2) to approximate the results of a centralized estimator. Similar strategies are applied, e.g., in Kamal, Farrell and Roy-Chowdhury (2013) and Battistelli *et al.* (2015).

4.1.2 Nominal Centralized Kalman Filter

In this section, we present the solution to problem (4.3) and ultimately propose the Nominal Centralized Kalman Filter (CKF). As aforementioned, problem (4.3) is a special case of the regularized least-squares estimation problem (Section 3.1.2). Thus, consider the following mappings between (4.3) and (3.6):

$$x \leftarrow \begin{bmatrix} \hat{x}_k \\ \hat{x}_{k+1} \end{bmatrix}, \quad \bar{x} \leftarrow \begin{bmatrix} \hat{x}_{k|k-1}^c \\ 0 \end{bmatrix}, \quad w \leftarrow \begin{bmatrix} \hat{w}_k \\ \hat{\boldsymbol{v}}_k \end{bmatrix}, \quad \bar{P} \leftarrow \begin{bmatrix} (P_{k|k-1}^c)^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad Q \leftarrow \begin{bmatrix} Q_k^{-1} & 0 \\ 0 & \boldsymbol{\mathcal{R}}_k^{-1} \end{bmatrix},$$
$$y \leftarrow \begin{bmatrix} -G_k u_k \\ \boldsymbol{\mathcal{Y}}_k \end{bmatrix}, \quad A \leftarrow \begin{bmatrix} F_k & -I_n \\ \boldsymbol{\mathcal{C}}_k & 0 \end{bmatrix}, \quad \text{and} \quad B \leftarrow \begin{bmatrix} H_k & 0 \\ 0 & \boldsymbol{\mathcal{D}}_k \end{bmatrix}, \tag{4.5}$$

with the bold aggregated vectors and matrices as defined in (4.4). Note that since $(P_{k|k-1}^c)^{-1} \succ 0$, we have that $\bar{P} \succeq 0$. Also, $Q_k^{-1} \succ 0$ and $\boldsymbol{\mathcal{R}}_k^{-1} \succ 0$, such that $Q \succ 0$. Therefore, we can use the results in Corollary 3.1 and Lemma 3.2 to obtain the optimal filtered and predicted central state estimates, $\hat{x}_{k|k}^c$ and $\hat{x}_{k+1|k}^c$, along with their corresponding error weighting matrices $P_{k|k}^c$ and $P_{k+1|k}^c$, as stated in the following theorem.

Theorem 4.1. *Consider the regularized least-squares centralized estimation problem* (4.3) *with H_k and $\boldsymbol{\mathcal{D}}_k$ full row rank and given initial conditions $\hat{x}_{0|-1}^c$, $P_{0|-1}^c = P_0 \succ 0$, $Q_k \succ 0$, and $R_k^i \succ 0$, $\forall i \in \mathbb{S}$. For each $k = 0, 1, \ldots, N$, its solution recursively provides the filtered and predicted central state estimates of system* (4.1)-(4.2)*, $\hat{x}_{k|k}^c$ and $\hat{x}_{k+1|k}^c$, as well as their respective error weighting matrices, $P_{k|k}^c$ and $P_{k+1|k}^c$, according to the procedure described in Algorithm 4.1.*

Algorithm 4.1 Nominal Centralized Kalman Filter (CKF)

Model: Assume the system model in (4.1)-(4.2).
Initialization: Set $\hat{x}^c_{0|-1}$, $P^c_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, and $R^i_k \succ 0$, $\forall i \in \mathbb{S}$.
for $k = 0, 1, \ldots, N$ **do**

1. Obtain measurements y^i_k from all sensors $i \in \mathbb{S}$.
2. Compute the auxiliary matrices, for all $i \in \mathbb{S}$:

$$\widehat{Q}_k = H_k Q_k H_k^T \qquad \widehat{R}^i_k = D^i_k R^i_k (D^i_k)^T$$

3. Correction step:
   3.1. Compute the posterior error weighting matrix:

$$P^c_{k|k} = \left[(P^c_{k|k-1})^{-1} + \sum_{i=1}^{S} (C^i_k)^T (\widehat{R}^i_k)^{-1} C^i_k \right]^{-1}$$

   3.2. Compute the filtered central state estimate:

$$\hat{x}^c_{k|k} = P^c_{k|k} \left[(P^c_{k|k-1})^{-1} \hat{x}^c_{k|k-1} + \sum_{i=1}^{S} (C^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k \right]$$

4. Prediction step:
   4.1. Update the predicted prior error weighting matrix:

$$P^c_{k+1|k} = F_k P^c_{k|k} F_k^T + \widehat{Q}_k$$

   4.2. Update the predicted prior central state estimate:

$$\hat{x}^c_{k+1|k} = F_k \hat{x}^c_{k|k} + G_k u_k$$

end for

Proof. As previously mentioned, problem (4.3) is a regularized least-squares estimation problem. Then, we apply the result in Corollary 3.1 to obtain the central state estimates $\hat{x}^c_{k|k}$ and $\hat{x}^c_{k+1|k}$. This is achieved by substituting the mappings (4.5) into the solution (3.14). The algebraic details are quite similar to the steps described in the proof of Theorem 3.1 and are thus omitted here for brevity. The main difference is the presence of the summation terms present in step 3 of Algorithm 4.1, which appear due to the aggregate vectors and matrices defined in (4.4), which account for all the sensors in the network. Given their block column and diagonal structures, we have that

$$\mathfrak{C}_k^T \widehat{\mathfrak{R}}_k^{-1} \mathfrak{C}_k = \sum_{i=1}^{S} (C^i_k)^T (\widehat{R}^i_k)^{-1} C^i_k \quad \text{and} \quad \mathfrak{C}_k^T \widehat{\mathfrak{R}}_k^{-1} \mathfrak{y}_k = \sum_{i=1}^{S} (C^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k,$$

where $\widehat{\mathfrak{R}}_k := \mathbf{diag}\left(\widehat{R}^1_k, \ldots, \widehat{R}^S_k\right)$, with each $\widehat{R}^i_k$ as defined in step 2 of Algorithm 4.1. Analogously, we use Lemma 3.2 to obtain the corresponding estimation error weighting matrices $P^c_{k|k}$ and $P^c_{k+1|k}$, as also shown in the proof of Theorem 3.1. To conclude, note that, by requiring H_k and $\mathfrak{D}_k$ to have full row rank, we ensure that $\widehat{Q}_k \succ 0$ and $\widehat{R}^i_k \succ 0$, $\forall i \in \mathbb{S}$. □

4.1.3 Nominal Distributed Kalman Consensus Filter

As we mentioned earlier, the CKF is our benchmark for a distributed formulation. In this section, we address the problem proposed in Section 4.1.1.3 and show how the average consensus strategy can be employed to derive a distributed approximation of the CKF presented in Algorithm 4.1.

We assume that each sensor $i \in \mathbb{S}$ is initialized with the same prior state estimate $\hat{x}^i_{0|-1}$ and prior error weighting matrix $P^i_{0|-1} \succ 0$. Then, by communicating with its neighbors $j \in \mathcal{N}_i$, each sensor $i \in \mathbb{S}$ can receive their specific data and use it to improve its estimation performance. By adopting the hybrid consensus on measurements and information (HCMCI) approach proposed in Battistelli *et al.* (2015) each sensor is able to attain an approximation of the filtered and predicted prior central state estimates in a distributed fashion.

The proposed Nominal Distributed Kalman Consensus Filter (DKCF) is thus shown in Algorithm 4.2. The HCMCI strategy consists of simultaneously performing the average consensus protocol (Algorithm 2.2) for each sensor's so-called prior information and innovations pairs, denoted $\left(\Omega^i_k,\, \omega^i_k\right)$ and $\left(\delta\Omega^i_k,\, \delta\omega^i_k\right)$, respectively. This is done in steps 4 and 5 of Algorithm 4.2. In step 5.3, the consensus weights π_{ij} should satisfy the conditions established in Definition 2.2, such that the consensus states of each node $i \in \mathbb{S}$ are updated with a convex combination of the corresponding states within its inclusive neighborhood. One possible choice for these weights is the Metropolis weights (XIAO; BOYD; LALL, 2005), shown in (2.20), and is the one we will adopt here. Then, based on the outcome of the consensus step and on step 3 of the CKF (Algorithm 4.1), we perform the correction stage shown in step 6 of Algorithm 4.2. Note that we introduce a corrective scalar weight ρ^i_k to compensate for the scaling effect of the average consensus process (more details on this later). Finally, the prediction stage in step 7 of Algorithm 4.2 is the same as step 4 of the centralized filter (Algorithm 4.1).

The following theorem shows that, considering enough consensus iterations, the proposed distributed filter approaches the same result as the centralized estimator.

Theorem 4.2. *Consider the Nominal Distributed Kalman Consensus Filter in Algorithm 4.2 and that Assumption 4.1 is satisfied. Assume that the consensus weights* π_{ij} *are chosen according to Definition 2.2, the number of consensus iterations* $L \to \infty$ *in step 5, and that* $\rho^i_k = S$ *in step 6. Then, the filtered and predicted prior state estimates,* $\hat{x}^i_{k|k}$ *and* $\hat{x}^i_{k+1|k}$*, and their respective error weighting matrices,* $P^i_{k|k}$ *and* $P^i_{k+1|k}$*, obtained by each sensor* $i \in \mathbb{S}$ *converge to the corresponding central state estimates* $\hat{x}^c_{k|k}$ *and* $\hat{x}^c_{k+1|k}$*, and error weighting matrices* $P^c_{k|k}$ *and* $P^c_{k+1|k}$ *computed via the Nominal Centralized Kalman Filter in Algorithm 4.1.*

Algorithm 4.2 Nominal Distributed Kalman Consensus Filter (DKCF) (each sensor i)

Model: Assume the system model in (4.1)-(4.2).

Initialization: Set $\hat{x}^i_{0|-1}$, $P^i_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R^i_k \succ 0$, and $L \geq 1$.

for $k = 0, 1, \ldots, N$ **do**

1. Obtain a measurement y^i_k.
2. Compute the auxiliary matrices:

$$\widehat{Q}_k = H_k Q_k H_k^T \qquad \widehat{R}^i_k = D^i_k R^i_k (D^i_k)^T$$

4. Initialize the consensus states:

$$\Omega^i_k(0) = (P^i_{k|k-1})^{-1} \qquad \delta\Omega^i_k(0) = (C^i_k)^T (\widehat{R}^i_k)^{-1} C^i_k$$
$$\omega^i_k(0) = (P^i_{k|k-1})^{-1} \hat{x}^i_{k|k-1} \qquad \delta\omega^i_k(0) = (C^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k$$

5. Consensus step:

for $\ell = 0, 1, \ldots, L-1$ **do**

5.1. Send $\left\{\Omega^i_k(\ell),\ \omega^i_k(\ell),\ \delta\Omega^i_k(\ell),\ \delta\omega^i_k(\ell)\right\}$ to all neighbors $j \in \mathcal{N}_i$.

5.2. Receive $\left\{\Omega^j_k(\ell),\ \omega^j_k(\ell),\ \delta\Omega^j_k(\ell),\ \delta\omega^j_k(\ell)\right\}$ from all neighbors $j \in \mathcal{N}_i$.

5.3. Update the consensus states:

$$\Omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \Omega^j_k(\ell) \qquad \delta\Omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \delta\Omega^j_k(\ell)$$
$$\omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \omega^j_k(\ell) \qquad \delta\omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \delta\omega^j_k(\ell)$$

end for

6. Correction step:

6.1. Compute the posterior error weighting matrix:

$$P^i_{k|k} = \left[\Omega^i_k(L) + \rho^i_k\, \delta\Omega^i_k(L)\right]^{-1}$$

6.2. Compute the filtered state estimate:

$$\hat{x}^i_{k|k} = P^i_{k|k}\left[\omega^i_k(L) + \rho^i_k\, \delta\omega^i_k(L)\right]$$

7. Prediction step:

7.1. Update the predicted prior error weighting matrix:

$$P^i_{k+1|k} = F_k P^i_{k|k} F_k^T + \widehat{Q}_k$$

7.2. Update the predicted prior state estimate:

$$\hat{x}^i_{k+1|k} = F_k \hat{x}^i_{k|k} + G_k u_k$$

end for

Proof. Given that the undirected graph $\mathbb{G}$ describing the sensor network is connected and that the consensus weights π_{ij} are properly selected, for instance, the Metropolis weights are chosen, as previously discussed, the associated weighted adjacency matrix Π has the properties listed in Lemma 2.9. Since we assume that the number of consensus iterations $L \to \infty$, the convergence of the average consensus algorithm is guaranteed according to Theorem 2.2. Let us then prove that the DKCF described in Algorithm 4.2 approaches the CKF in Algorithm 4.1 through induction.

Assume that the CKF is initialized with $\hat{x}^c_{0|-1} = \hat{x}_0$ and $P^c_{0|-1} = P_0 \succ 0$, whereas all sensors $i \in \mathbb{S}$ initialize a DKCF with $\hat{x}^i_{0|-1} = \hat{x}_0$ and $P^i_{0|-1} = P_0 \succ 0$. According to (2.18), as $L \to \infty$, after the consensus step 5 of the DKCF, the information and innovation pairs of all sensors converge in the following manner:

$$\Omega^i_0(L) \to \frac{1}{S}\sum_{j=1}^{S} P_0^{-1} = P_0^{-1}, \qquad \delta\Omega^i_0(L) \to \frac{1}{S}\sum_{j=1}^{S}(C^j_0)^T(\widehat{R}^j_0)^{-1}C^j_0,$$
$$\omega^i_0(L) \to \frac{1}{S}\sum_{j=1}^{S} P_0^{-1}\hat{x}_0 = P_0^{-1}\hat{x}_0, \qquad \delta\omega^i_0(L) \to \frac{1}{S}\sum_{j=1}^{S}(C^j_0)^T(\widehat{R}^j_0)^{-1}y^j_0.$$

Then, substituting these consensus outcomes into the equations in step 6 and considering that the corrective scalar weight $\rho^i_0 = S$, we get

$$P^i_{0|0} \to \left[P_0^{-1} + S\frac{1}{S}\sum_{j=1}^{S}(C^j_0)^T(\widehat{R}^j_0)^{-1}C^j_0\right]^{-1} = \left[P_0^{-1} + \sum_{j=1}^{S}(C^j_0)^T(\widehat{R}^j_0)^{-1}C^j_0\right]^{-1} = P^c_{0|0},$$
$$\hat{x}^i_{0|0} \to P^i_{0|0}\left[P_0^{-1}\hat{x}_0 + S\frac{1}{S}\sum_{j=1}^{S}(C^j_0)^T(\widehat{R}^j_0)^{-1}y^j_0\right] = P^c_{0|0}\left[P_0^{-1}\hat{x}_0 + \sum_{j=1}^{S}(C^j_0)^T(\widehat{R}^j_0)^{-1}y^j_0\right] = \hat{x}^c_{0|0},$$

for all sensors $i \in \mathbb{S}$. Note here the importance of the scalar weight ρ^i_0, which compensates for the $1/S$ factor that appears in the outcome of the innovation pair $\left(\delta\Omega^i_0,\, \delta\omega^i_0\right)$ due to the averaging process. The convergence above thus implies that in step 7, we have $P^i_{1|0} \to P^c_{1|0}$ and $\hat{x}^i_{1|0} \to \hat{x}^c_{1|0}$. Therefore, for $k = 0$, the results of the DKCF do converge to those of the CKF.

Now, assume that at time step $k-1$, we have that $P^i_{k-1|k-1} \to P^c_{k-1|k-1}$, $\hat{x}^i_{k-1|k-1} \to \hat{x}^c_{k-1|k-1}$, $P^i_{k|k-1} \to P^c_{k|k-1}$, and $\hat{x}^i_{k|k-1} \to \hat{x}^c_{k|k-1}$, $\forall i \in \mathbb{S}$. Then, at time step k, we achieve the following outcome after performing step 5 of the DKCF:

$$\Omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}(P^c_{k|k-1})^{-1} = (P^c_{k|k-1})^{-1}, \qquad \delta\Omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}(C^j_k)^T(\widehat{R}^j_k)^{-1}C^j_k,$$
$$\omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} = (P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1}, \quad \delta\omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}(C^j_k)^T(\widehat{R}^j_k)^{-1}y^j_k.$$

Thus, substituting these outcomes into the equations in step 6 of the DKCF, considering $\rho_k^i = S$, yields

$$P_{k|k}^i \to \left[(P_{k|k-1}^c)^{-1} + S\frac{1}{S}\sum_{j=1}^{S}(C_k^j)^T(\widehat{R}_k^j)^{-1}C_k^j \right]^{-1} = \left[(P_{k|k-1}^c)^{-1} + \sum_{j=1}^{S}(C_k^j)^T(\widehat{R}_k^j)^{-1}C_k^j \right]^{-1} = P_{k|k}^c,$$

$$\hat{x}_{k|k}^i \to P_{k|k}^i \left[(P_{k|k-1}^c)^{-1}\hat{x}_{k|k-1}^c + S\frac{1}{S}\sum_{j=1}^{S}(C_k^j)^T(\widehat{R}_k^j)^{-1}y_k^j \right] =$$
$$P_{k|k}^c \left[(P_{k|k-1}^c)^{-1}\hat{x}_{k|k-1}^c + \sum_{j=1}^{S}(C_k^j)^T(\widehat{R}_k^j)^{-1}y_k^j \right] = \hat{x}_{k|k}^c,$$

for all sensors $i \in \mathbb{S}$. Plugging the results above into the equations in step 7 of the DKCF gives us that $P_{k+1|k}^i \to P_{k+1|k}^c$ and $\hat{x}_{k+1|k}^i \to \hat{x}_{k+1|k}^c$. Hence, given the aforementioned conditions, by induction, we have that for $k = 0, 1, \ldots, N$, the DKCF in Algorithm 4.2 converges to the CKF in Algorithm 4.1. □

The result in Theorem 4.2 shows how powerful the average consensus protocol is when applied to the context of distributed estimation over sensor networks. However, it is a theoretical outcome since, in practice, only a finite number of consensus iterations L is possible. Nevertheless, for a sufficiently large L, the performance of the distributed and centralized approaches can still be quite similar.

Remark 4.1. *In step 6 of Algorithm 4.2, we multiply the consensus outcome of the innovation pair* $\left(\delta\Omega_k^i(L),\, \delta\omega_k^i(L)\right)$ *by a corrective scalar weight* ρ_k^i*. The reason for this is to avoid the underweighting of the innovation pair due to scaling from the average consensus procedure. This actually turns Algorithm 4.2 into a family of distributed filters, depending on the choice of this weight.*

Remark 4.2. *As Theorem 4.2 states, to correctly approximate the centralized estimator, ideally one should have* $\rho_k^i = S$*. However, the total number of sensors* S *is usually not available to each sensor in the network. Nonetheless, according to Garin and Schenato (2010), we can also use average consensus to compute* S *in a distributed fashion, as follows. Initialize the consensus state of the sensors as* $\alpha_1(0) = 1$ *and* $\alpha_i(0) = 0$, $i = 2, \ldots, S$*. Each sensor then performs Algorithm 2.2, such that* $\alpha_i(L) \to 1/S$*. Therefore, we can use* $\rho_k^i = 1/\alpha_i(L)$*, if* $\alpha_i(L) > 0$*, or* $\rho_k^i = 1$ *otherwise. Note that this consensus procedure can be performed along with steps 4 and 5 in Algorithm 4.2.*

4.1.4 Stability Analysis

This section discusses the stability properties of both the proposed Nominal Centralized Kalman Filter and the Nominal Distributed Kalman Consensus Filter. To this end, we examine the steady-state behavior of Algorithm 4.1 and Algorithm 4.2 when the target system and sensing model parameters are constant and there is no input u_k. Thus, for $k \geq 0$, equations (4.1)-(4.2) take the form

$$\begin{aligned} x_{k+1} &= F x_k + H w_k, \\ y_k^i &= C^i x_k + D^i v_k^i, \quad \forall i \in \mathbb{S}. \end{aligned} \tag{4.6}$$

Based on the strategy adopted in Kamal, Farrell and Roy-Chowdhury (2013), we conduct our analysis under the assumptions described in Theorem 4.2. This way, we can assume that the DKCF converges to the CKF. This, in turn, allows us to extend the stability properties of the centralized filter to its distributed counterpart.

Therefore, let us first study the stability of the CKF described in Algorithm 4.1. Consider the time-invariant system model (4.6). Thus, the CKF equations in steps 3 and 4 of Algorithm 4.1 become:

$$P_{k|k}^c = \Big[(P_{k|k-1}^c)^{-1} + \mathfrak{C}^T \widehat{\mathfrak{R}}^{-1} \mathfrak{C}\Big]^{-1}, \tag{4.7}$$

$$\hat{x}_{k|k}^c = P_{k|k}^c \Big[(P_{k|k-1}^c)^{-1} \hat{x}_{k|k-1}^c + \mathfrak{C}^T \widehat{\mathfrak{R}}^{-1} \mathfrak{y}_k\Big], \tag{4.8}$$

$$P_{k+1|k}^c = F P_{k|k}^c F^T + \widehat{Q}, \tag{4.9}$$

$$\hat{x}_{k+1|k}^c = F \hat{x}_{k|k}^c, \tag{4.10}$$

where

$$\mathfrak{y}_k = \begin{bmatrix} y_k^1 \\ \vdots \\ y_k^S \end{bmatrix}, \; \mathfrak{C} = \begin{bmatrix} C^1 \\ \vdots \\ C^S \end{bmatrix}, \; \widehat{\mathfrak{R}} = \begin{bmatrix} \widehat{R}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widehat{R}^S \end{bmatrix}, \; \widehat{R}^i = D^i R^i (D^i)^T, \text{ and } \widehat{Q} = H Q H^T.$$

Then, we expand expression (4.7) using the matrix inversion lemma (Lemma A.1), as follows:

$$P_{k|k}^c = P_{k|k-1}^c - P_{k|k-1}^c \mathfrak{C}^T \Big(\widehat{\mathfrak{R}} + \mathfrak{C} P_{k|k-1}^c \mathfrak{C}^T\Big)^{-1} \mathfrak{C} P_{k|k-1}^c. \tag{4.11}$$

Combining (4.11) with (4.8) and substituting in (4.10) yields the steady-state predicted state estimate

$$\hat{x}_{k+1|k}^c = \widetilde{\mathfrak{F}}_k \hat{x}_{k|k-1}^c + \widetilde{\mathfrak{F}}_k P_{k|k-1}^c \mathfrak{C}^T \widehat{\mathfrak{R}}^{-1} \mathfrak{y}_k, \tag{4.12}$$

in which

$$\widetilde{\mathfrak{F}}_k = F\Big(I_n - P_{k|k-1}^c \mathfrak{C}^T \Big(\widehat{\mathfrak{R}} + \mathfrak{C} P_{k|k-1}^c \mathfrak{C}^T\Big)^{-1} \mathfrak{C}\Big)$$

is the centralized filter closed-loop matrix. In addition, substituting $P_{k|k}^c$ from (4.11) into (4.9), we obtain the expression for the predicted prior error weighting matrix:

$$P_{k+1|k}^c = F\Big(P_{k|k-1}^c - P_{k|k-1}^c \mathfrak{C}^T \Big(\widehat{\mathfrak{R}} + \mathfrak{C} P_{k|k-1}^c \mathfrak{C}^T\Big)^{-1} \mathfrak{C} P_{k|k-1}^c\Big) F^T + \widehat{Q}. \tag{4.13}$$

Theorem 4.3. *Consider the linear system model* (4.6) *and the corresponding centralized filter* (4.12)-(4.13)*. Assume that* $\{F, \mathfrak{C}\}$ *is* detectable *and* $\{F, \widehat{Q}^{1/2}\}$ *is* controllable*. Then, for any initial condition* $P^c_{0|-1} \succ 0$*,* $P^c_{k+1|k}$ *converges to the* unique stabilizing solution $P^c \succ 0$ *of the algebraic Riccati equation*

$$P^c = F\Big(P^c - P^c\mathfrak{C}^T\big(\widehat{\mathfrak{R}} + \mathfrak{C}P^c\mathfrak{C}^T\big)^{-1}\mathfrak{C}P^c\Big)F^T + \widehat{Q}. \tag{4.14}$$

The solution P^c *is stabilizing in the sense that the steady-state filter closed-loop matrix*

$$\widetilde{\mathfrak{F}} = F\Big(I_n - P^c\mathfrak{C}^T\big(\widehat{\mathfrak{R}} + \mathfrak{C}P^c\mathfrak{C}^T\big)^{-1}\mathfrak{C}\Big) \tag{4.15}$$

is Schur stable.

Proof. From Kailath, Sayed and Hassibi (2000b), detectability of $\{F, \mathfrak{C}\}$ and controllability of $\{F, \widehat{Q}^{1/2}\}$ ensure the convergence of $P^c_{k+1|k}$ in (4.13) to the unique stabilizing positive definite solution P^c of the algebraic Riccati equation (4.14) that stabilizes (4.15), which is the centralized filter steady-state closed-loop matrix. □

Corollary 4.1. *Given that the assumptions in Theorem 4.2 hold, the DKCF in Algorithm 4.2 converges to the CKF in Algorithm 4.1 and thus shares its stability properties, according to Theorem 4.3.*

4.2 Robust Distributed Kalman Filtering for Systems with Norm-Bounded Uncertainties

In this section, we present robust versions of the nominal centralized and distributed Kalman filters developed in Section 4.1. We specifically address the case in which both the underlying target system and sensing models are subject to norm-bounded parametric uncertainties.

Based on the procedure reported in Section 4.1, we start by dealing with the scenario in which measurements from all sensors in the network are available to a central estimator. This centralized estimation problem is formulated as a regularized least-squares estimation problem with norm-bounded uncertainties (Section 3.2.2) and its solution provides the Robust Centralized Kalman Filter (RCKF). Then, similar to the nominal case in the previous section, we leverage the HCMCI protocol (BATTISTELLI *et al.*, 2015) to implement the RCKF in a distributed fashion, leading to the Robust Distributed Kalman Consensus Filter (RDKCF), which approaches the RCKF if enough consensus iterations are carried out. Both the RCKF and RDKCF are presented as recursive correction-prediction algorithms, which resemble the single-sensor Robust Kalman Filter (Algorithm 3.2). Furthermore, we evaluate the stability properties of both estimators and conclude the section with an illustrative example.

4.2.1 Problem Formulation

4.2.1.1 System Model

Consider a sensor network composed of S sensors. The communication among them is represented by the undirected graph $\mathbb{G} = (\mathbb{S}, \mathbb{E})$, with node set $\mathbb{S} = \{1, 2, \ldots, S\}$ and edge set $\mathbb{E} \subseteq \mathbb{S} \times \mathbb{S}$. The neighborhood of a sensor i is denoted by $\mathbb{N}_i = \{j \in \mathbb{S} \,|\, (i,j) \in \mathbb{E}\}$ and has cardinality N_i (see Section 2.3 for an introduction on graph theory).

Assumption 4.2. *The undirected graph $\mathbb{G}$ has a fixed topology and is connected, i.e., there is a path between every pair of nodes.*

Consider the following discrete-time state-space description of a linear target system subject to uncertainties:

$$x_{k+1} = (F_k + \delta F_k)x_k + (G_k + \delta G_k)u_k + (H_k + \delta H_k)w_k, \tag{4.16}$$

which is observed by the set of S sensors $\mathbb{S} = \{1, 2, \ldots, S\}$, each described by the uncertain model

$$y_k^i = (C_k^i + \delta C_k^i)x_k + (D_k^i + \delta D_k^i)v_k^i, \quad \forall i \in \mathbb{S}, \tag{4.17}$$

for $k = 0, 1, \ldots, N$, with state vector $x_k \in \mathbb{R}^n$, input vector $u_k \in \mathbb{R}^m$, and system noise vector $w_k \in \mathbb{R}^p$. For each sensor $i \in \mathbb{S}$, $y_k^i \in \mathbb{R}^r$ is the measurement vector and $v_k^i \in \mathbb{R}^q$ is the measurement noise. $F_k \in \mathbb{R}^{n\times n}$, $G_k \in \mathbb{R}^{n\times m}$, $H_k \in \mathbb{R}^{n\times p}$, $C_k^i \in \mathbb{R}^{r\times n}$, and $D_k^i \in \mathbb{R}^{r\times q}$ are known nominal parameter matrices, whereas $\delta F_k \in \mathbb{R}^{n\times n}$, $\delta G_k \in \mathbb{R}^{n\times m}$, $\delta H_k \in \mathbb{R}^{n\times p}$, $\delta C_k^i \in \mathbb{R}^{r\times n}$, and $\delta D_k^i \in \mathbb{R}^{r\times q}$ are norm-bounded parametric uncertainties modeled as

$$\begin{aligned} \begin{bmatrix}\delta F_k & \delta G_k & \delta H_k\end{bmatrix} &= M_{1,k}\Delta_{1,k}\begin{bmatrix}E_{F_k} & E_{G_k} & E_{H_k}\end{bmatrix}, \quad \|\Delta_{1,k}\| \leq 1, \\ \begin{bmatrix}\delta C_k^i & \delta D_k^i\end{bmatrix} &= M_{2,k}^i\Delta_{2,k}^i\begin{bmatrix}E_{C_k}^i & E_{D_k}^i\end{bmatrix}, \qquad \|\Delta_{2,k}^i\| \leq 1, \end{aligned} \tag{4.18}$$

where $M_{1,k} \in \mathbb{R}^{n\times s_1}$ and $M_{2,k}^i \in \mathbb{R}^{r\times s_2}$ are known nonzero matrices, $E_{F_k} \in \mathbb{R}^{t_1\times n}$, $E_{G_k} \in \mathbb{R}^{t_1\times m}$, $E_{H_k} \in \mathbb{R}^{t_1\times p}$, $E_{C_k}^i \in \mathbb{R}^{t_2\times n}$ and $E_{D_k}^i \in \mathbb{R}^{t_2\times q}$ are also known, and $\Delta_{1,k} \in \mathbb{R}^{s_1\times t_1}$ and $\Delta_{2,k}^i \in \mathbb{R}^{s_2\times t_2}$ are arbitrary contraction matrices. Perturbations of this form are useful when modeling tolerance specifications on the physical parameters of a system and are common in robust filtering and control (SAYED, 2001).

In a stochastic setting, it is usually assumed that x_0, w_k, and $\{v_k^i\}_{i=1}^S$ are mutually independent zero-mean Gaussian random variables with respective variances

$$\boldsymbol{E}\Big\{x_0x_0^T\Big\} = P_0 \succ 0, \quad \boldsymbol{E}\Big\{w_kw_l^T\Big\} = Q_k\delta_{kl} \succ 0, \quad \text{and} \quad \boldsymbol{E}\Big\{v_k^i(v_l^j)^T\Big\} = R_k^i\delta_{kl}\delta_{ij} \succ 0,$$

where δ_{ab} is the Kronecker delta function, such that $\delta_{ab} = 1$ if $a = b$, and $\delta_{ab} = 0$ otherwise. However, we adopt a strategy which does not require that these variables have any particular distribution.

4.2.1.2 Robust Centralized Estimation Problem

As aforementioned, prior to tackling the robust distributed estimation problem, we first derive a centralized estimator for system (4.16)-(4.17). In the centralized case, we assume that a central estimator has access to all measurements obtained by the sensors in the network. Since the target system state sequence $\{x_k\}$ is not perfectly observed, nor is readily available, the goal is thus to use all the information available up to time instant k, $\boldsymbol{Y}_k = \left\{\{y_0^i\}_{i=1}^S, \ldots, \{y_k^i\}_{i=1}^S, u_0, \ldots, u_k\right\}$, to compute a so-called filtered robust central state estimate $\hat{x}_{k|k}^c$ of x_k, as well as a predicted robust central estimate $\hat{x}_{k+1|k}^c$ of x_{k+1}, despite the presence of the model uncertainties $\delta_k := \left\{\delta F_k, \delta G_k, \delta H_k, \{\delta C_k^i\}_{i=1}^S, \{\delta D_k^i\}_{i=1}^S\right\}$. Note that we use the superscript c to indicate the centralized entities.

We build upon the procedure described in Section 4.1.1.2 for the Nominal Centralized Kalman Filter, in which we adopt a deterministic interpretation of the centralized estimation problem (BRYSON; HO, 1975). Moreover, we introduce the variables $\hat{x}_k$, $\hat{x}_{k+1}$, $\hat{w}_k$, and $\{\hat{v}_k^i\}_{i=1}^S$ as substitutes for the random variables x_k, x_{k+1}, w_k, and $\{v_k^i\}_{i=1}^S$ in the stochastic model (4.16)-(4.17). Then, based on Sayed (2001) and Ishihara, Terra and Cerri (2015), assuming that at each time step k, an *a priori* state estimate $\hat{x}_{k|k-1}^c$, a set of measurements $\{y_k^i\}_{i=1}^S$, and the input u_k are available, we formulate a min-max constrained optimization problem in which a one-step quadratic objective function should be minimized under the maximum influence of the parametric uncertainties δ_k defined in (4.18), i.e.,

$$\begin{aligned} &\min_{\substack{\hat{x}_k, \hat{x}_{k+1}, \\ \hat{w}_k, \hat{\boldsymbol{v}}_k}} \max_{\delta_k} \quad J_k(\hat{x}_k, \hat{w}_k, \hat{\boldsymbol{v}}_k) = \|\hat{x}_k - \hat{x}_{k|k-1}^c\|^2_{(P_{k|k-1}^c)^{-1}} + \|\hat{w}_k\|^2_{Q_k^{-1}} + \|\hat{\boldsymbol{v}}_k\|^2_{\boldsymbol{\mathcal{R}}_k^{-1}}, \\ &\text{subject to} \quad \begin{cases} \hat{x}_{k+1} = (F_k + \delta F_k)\hat{x}_k + (G_k + \delta G_k)u_k + (H_k + \delta H_k)\hat{w}_k, \\ \boldsymbol{\mathcal{Y}}_k = (\boldsymbol{\mathcal{C}}_k + \boldsymbol{\delta\mathcal{C}}_k)\hat{x}_k + (\boldsymbol{\mathcal{D}}_k + \boldsymbol{\delta\mathcal{D}}_k)\hat{\boldsymbol{v}}_k, \end{cases} \end{aligned} \tag{4.19}$$

for $k = 0, 1, \ldots, N$, in which we define the aggregated vectors and matrices

$$\boldsymbol{\mathcal{Y}}_k := \begin{bmatrix} y_k^1 \\ \vdots \\ y_k^S \end{bmatrix}, \ \hat{\boldsymbol{v}}_k := \begin{bmatrix} \hat{v}_k^1 \\ \vdots \\ \hat{v}_k^S \end{bmatrix}, \ \boldsymbol{\mathcal{C}}_k := \begin{bmatrix} C_k^1 \\ \vdots \\ C_k^S \end{bmatrix}, \ \boldsymbol{\mathcal{D}}_k := \begin{bmatrix} D_k^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_k^S \end{bmatrix}, \ \boldsymbol{\mathcal{R}}_k := \begin{bmatrix} R_k^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_k^S \end{bmatrix}, \tag{4.20}$$

and, based on (4.18), we also define the aggregated sensing uncertainty model

$$\begin{bmatrix} \boldsymbol{\delta\mathcal{C}}_k & \boldsymbol{\delta\mathcal{D}}_k \end{bmatrix} = \boldsymbol{\mathcal{M}}_{2,k}\boldsymbol{\Delta}_{2,k} \begin{bmatrix} \boldsymbol{E}_{\mathcal{C}_k} & \boldsymbol{E}_{\mathcal{D}_k} \end{bmatrix}, \quad \|\boldsymbol{\Delta}_{2,k}\| \leq 1, \tag{4.21}$$

where

$$\begin{aligned} &\boldsymbol{\delta\mathcal{C}}_k := \begin{bmatrix} \delta C_k^1 \\ \vdots \\ \delta C_k^S \end{bmatrix}, \quad \boldsymbol{\delta\mathcal{D}}_k := \begin{bmatrix} \delta D_k^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \delta D_k^S \end{bmatrix}, \quad \boldsymbol{\mathcal{M}}_{2,k} := \begin{bmatrix} M_{2,k}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & M_{2,k}^S \end{bmatrix}, \\ &\boldsymbol{\Delta}_{2,k} := \begin{bmatrix} \Delta_{2,k}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Delta_{2,k}^S \end{bmatrix}, \quad \boldsymbol{E}_{\mathcal{C}_k} := \begin{bmatrix} E_{C_k}^1 \\ \vdots \\ E_{C_k}^S \end{bmatrix}, \text{ and } \boldsymbol{E}_{\mathcal{D}_k} := \begin{bmatrix} E_{D_k}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & E_{D_k}^S \end{bmatrix}. \end{aligned} \tag{4.22}$$

Notice that in (4.19), $\hat{w}_k$ and $\{\hat{v}_k^i\}_{i=1}^S$ are fitting errors weighted respectively by $Q_k \succ 0$ and $R_k^i \succ 0$, $\forall i \in \mathbb{S}$, whereas $P_{k|k-1}^c \succ 0$ weights the *a priori* estimation error $x_k - \hat{x}_{k|k-1}^c$. In a stochastic interpretation, matrices Q_k and R_k^i represent the variances of the random variables w_k and $\{v_k^i\}_{i=1}^S$. However, in this deterministic framework, they are treated as general weighting matrices.

Problem (4.19) is a special case of a regularized least-squares estimation problem with norm-bounded uncertainties, as discussed in Section 3.2.2. Its solution recursively yields the filtered and predicted robust central state estimates $\hat{x}_{k|k}^c$ and $\hat{x}_{k|k+1}^c$, respectively.

4.2.1.3 Robust Distributed Estimation Problem

Regarding the robust distributed estimation case, there is no central estimator, such that the goal of each sensor in the network is to estimate the state of the target system using only its own information and data gathered from its neighbors. Hence, each sensor node $i \in \mathbb{S}$ should obtain the best estimates $\hat{x}_{k|k}^i$ of x_k and $\hat{x}_{k+1|k}^i$ of x_{k+1}, referred to as filtered and predicted robust state estimates, in a distributed rather than centralized fashion.

To attain this goal, we leverage the distributed characteristic of the average consensus algorithm (Algorithm 2.2) to approximate the results of a centralized estimator, as we did in Section 4.1. This strategy has also been applied, for instance, in Kamal, Farrell and Roy-Chowdhury (2013) and Battistelli *et al.* (2015).

4.2.2 Robust Centralized Kalman Filter

This section presents the Robust Centralized Kalman Filter (RCKF), obtained as the outcome of the solution to problem (4.19). As previously mentioned, problem (4.19) has the form of a regularized least-squares estimation problem with norm-bounded uncertainties, as discussed in Section 3.2.2. Therefore, consider the following mappings between (4.19) and (3.34):

$$x \leftarrow \begin{bmatrix} \hat{x}_k \\ \hat{x}_{k+1} \end{bmatrix}, \quad \bar{x} \leftarrow \begin{bmatrix} \hat{x}_{k|k-1} \\ 0 \end{bmatrix}, \quad w \leftarrow \begin{bmatrix} \hat{w}_k \\ \hat{\boldsymbol{v}}_k \end{bmatrix}, \quad \bar{P} \leftarrow \begin{bmatrix} (P_{k|k-1}^c)^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad Q \leftarrow \begin{bmatrix} Q_k^{-1} & 0 \\ 0 & \boldsymbol{\mathfrak{R}}_k^{-1} \end{bmatrix},$$
$$y \leftarrow \begin{bmatrix} -G_k u_k \\ \boldsymbol{\mathfrak{Y}}_k \end{bmatrix}, \quad A \leftarrow \begin{bmatrix} F_k & -I_n \\ \boldsymbol{\mathfrak{C}}_k & 0 \end{bmatrix}, \quad B \leftarrow \begin{bmatrix} H_k & 0 \\ 0 & \boldsymbol{\mathfrak{D}}_k \end{bmatrix},$$
$$\delta y \leftarrow \begin{bmatrix} -\delta G_k u_k \\ 0 \end{bmatrix}, \quad \delta A \leftarrow \begin{bmatrix} \delta F_k & 0 \\ \boldsymbol{\delta\mathfrak{C}}_k & 0 \end{bmatrix}, \quad \text{and} \quad \delta B \leftarrow \begin{bmatrix} \delta H_k & 0 \\ 0 & \boldsymbol{\delta\mathfrak{D}}_k \end{bmatrix}, \tag{4.23}$$

with the bold aggregated vectors and matrices as defined in (4.20). Moreover, consider the following mappings between the uncertainty models (4.18)-(4.21) and (3.33):

$$
\begin{aligned}
M \leftarrow \begin{bmatrix} M_{1,k} & 0 \\ 0 & \boldsymbol{\mathfrak{M}}_{2,k} \end{bmatrix}, \quad \Delta \leftarrow \begin{bmatrix} \Delta_{1,k} & 0 \\ 0 & \boldsymbol{\Delta}_{2,k} \end{bmatrix}, \\
E_y \leftarrow \begin{bmatrix} -E_{G_k} u_k \\ 0 \end{bmatrix}, \quad E_A \leftarrow \begin{bmatrix} E_{F_k} & 0 \\ \boldsymbol{E}_{\mathfrak{C}_k} & 0 \end{bmatrix}, \quad \text{and} \quad E_B \leftarrow \begin{bmatrix} E_{H_k} & 0 \\ 0 & \boldsymbol{E}_{\mathfrak{D}_k} \end{bmatrix},
\end{aligned} \tag{4.24}
$$

with the uncertainty aggregated vectors and matrices as defined in (4.22).

Since $(P^c_{k|k-1})^{-1} \succ 0$, we have that $\bar{P} \succeq 0$. In addition, $Q_k^{-1} \succ 0$ and $\boldsymbol{\mathfrak{R}}_k^{-1} \succ 0$, such that $Q \succ 0$. Therefore, by using the results in Lemma 3.3 and in equation (3.58), we can obtain the filtered and predicted robust central state estimates, $\hat{x}^c_{k|k}$ and $\hat{x}^c_{k+1|k}$, along with their respective error weighting matrices $P^c_{k|k}$ and $P^c_{k+1|k}$, as the following theorem states.

Theorem 4.4. *Consider the regularized least-squares centralized estimation problem with norm-bounded uncertainties (4.19) with given initial conditions $\hat{x}^c_{0|-1}$, $P^c_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R^i_k \succ 0$, $\forall i \in \mathbb{S}$, and fixed parameters $\mu > 0$ and $\xi > 0$. For each $k = 0, 1, \ldots, N$, its solution recursively provides the filtered and predicted robust central state estimates of system (4.16)-(4.17), $\hat{x}^c_{k|k}$ and $\hat{x}^c_{k+1|k}$, along with their corresponding error weighting matrices, $P^c_{k|k}$ and $P^c_{k+1|k}$, according to the procedure outlined in Algorithm 4.3.*

Proof. Problem (4.19) is a regularized least-squares estimation problem with norm-bounded uncertainties, hence one can apply the result in Lemma 3.3 to obtain the robust central state estimates $\hat{x}^c_{k|k}$ and $\hat{x}^c_{k+1|k}$. Then, we substitute the mappings (4.23) and (4.24) into (3.40) to compute the modified system and sensing model matrices and then into the solution (3.39). The algebraic procedure closely follows the one described in the proof of Theorem 3.3 and we thus omit it for brevity. The main difference is in the summation terms present in step 4 of Algorithm 4.3, which appear due to the aggregate vectors and matrices defined in (4.20) and (4.22), which account for all the sensors in the network. Given their block column and diagonal structures, we have that

$$
\begin{aligned}
\widehat{\boldsymbol{\mathfrak{C}}}_k^T \widehat{\boldsymbol{\mathfrak{R}}}_k^{-1} \widehat{\boldsymbol{\mathfrak{C}}}_k + \boldsymbol{E}_{\mathfrak{C}_k}^T \bar{\boldsymbol{\mathfrak{R}}}_k^{-1} \boldsymbol{E}_{\mathfrak{C}_k} &= \sum_{i=1}^{S} \left[(\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} \widehat{C}^i_k + (E^i_{C_k})^T (\bar{R}^i_k)^{-1} E^i_{C_k} \right], \\
\widehat{\boldsymbol{\mathfrak{C}}}_k^T \widehat{\boldsymbol{\mathfrak{R}}}_k^{-1} \boldsymbol{\mathfrak{y}}_k &= \sum_{i=1}^{S} (\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k,
\end{aligned}
$$

where $\widehat{\boldsymbol{\mathfrak{C}}}_k := \mathbf{col}\left(\widehat{C}^1_k, \ldots, \widehat{C}^S_k\right)$, $\boldsymbol{E}_{\mathfrak{C}_k} := \mathbf{col}\left(E^1_{C_k}, \ldots, E^S_{C_k}\right)$, $\widehat{\boldsymbol{\mathfrak{R}}}_k := \mathbf{diag}\left(\widehat{R}^1_k, \ldots, \widehat{R}^S_k\right)$, and $\bar{\boldsymbol{\mathfrak{R}}}_k := \mathbf{diag}\left(\bar{R}^1_k, \ldots, \bar{R}^S_k\right)$, with each $\widehat{C}^i_k$, $\widehat{R}^i_k$, and $\bar{R}^i_k$ as defined in step 3 of Algorithm 4.3. Similarly, we use (3.58) to obtain the corresponding estimation error weighting matrices $P^c_{k|k}$ and $P^c_{k+1|k}$, as also shown in the proof of Theorem 3.3. Moreover, note that to compute the $\hat{\lambda}_k$ parameter, we consider the practical approximation discussed in Remark 3.2. □

Algorithm 4.3 Robust Centralized Kalman Filter (RCKF)

Model: Assume the uncertain system model in (4.16)-(4.17).

Initialization: Set $\hat{x}^c_{0|-1}$, $P^c_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R^i_k \succ 0$, $\forall i \in \mathbb{S}$, $\mu > 0$, and $\xi > 0$.

for $k = 0, 1, \ldots, N$ **do**

1. Obtain measurements y^i_k from all sensors $i \in \mathbb{S}$.
2. Compute $\hat{\lambda}_k$ using the approximation:

$$\hat{\lambda}_k = (1+\xi)\,\mu\,\left\|\mathbf{diag}\left(M^T_{1,k}M_{1,k},\, \mathcal{M}^T_{2,k}\mathcal{M}_{2,k}\right)\right\|$$

3. Compute the modified system and sensing model matrices, for all $i \in \mathbb{S}$:

$$\Phi_{1,k} = \mu^{-1}I_n - \hat{\lambda}_k^{-1}M_{1,k}M^T_{1,k} \qquad \widehat{Q}_k = \Phi_{1,k} + H_k\left(Q_k^{-1} + \hat{\lambda}_k E^T_{H_k}E_{H_k}\right)^{-1}H^T_k$$

$$\Phi^i_{2,k} = \mu^{-1}I_r - \hat{\lambda}_k^{-1}M^i_{2,k}(M^i_{2,k})^T \qquad \widehat{R}^i_k = \Phi^i_{2,k} + D^i_k\left[(R^i_k)^{-1} + \hat{\lambda}_k(E^i_{D_k})^T E^i_{D_k}\right]^{-1}(D^i_k)^T$$

$$\bar{Q}_k = \hat{\lambda}_k^{-1}I_{t_1} + E_{H_k}Q_kE^T_{H_k} \qquad \bar{R}^i_k = \hat{\lambda}_k^{-1}I_{t_2} + E^i_{D_k}R^i_k(E^i_{D_k})^T$$

$$\widehat{F}_k = F_k - H_kQ_kE^T_{H_k}\bar{Q}_k^{-1}E_{F_k} \qquad \widehat{C}^i_k = C^i_k - D^i_kR^i_k(E^i_{D_k})^T(\bar{R}^i_k)^{-1}E^i_{C_k}$$

$$\widehat{G}_k = G_k - H_kQ_kE^T_{H_k}\bar{Q}_k^{-1}E_{G_k}$$

4. Correction step:
   4.1. Compute the posterior error weighting matrix:

$$P^c_{k|k} = \left[(P^c_{k|k-1})^{-1} + \sum_{i=1}^{S}\left[(\widehat{C}^i_k)^T(\widehat{R}^i_k)^{-1}\widehat{C}^i_k + (E^i_{C_k})^T(\bar{R}^i_k)^{-1}E^i_{C_k}\right] + E^T_{F_k}\bar{Q}_k^{-1}E_{F_k}\right]^{-1}$$

   4.2. Compute the filtered robust central state estimate:

$$\hat{x}^c_{k|k} = P^c_{k|k}\left[(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} + \sum_{i=1}^{S}(\widehat{C}^i_k)^T(\widehat{R}^i_k)^{-1}y^i_k - E^T_{F_k}\bar{Q}_k^{-1}E_{G_k}u_k\right]$$

5. Prediction step:
   5.1. Update the predicted prior error weighting matrix:

$$P^c_{k+1|k} = \widehat{F}_kP^c_{k|k}\widehat{F}^T_k + \widehat{Q}_k$$

   5.2. Update the predicted prior robust central state estimate:

$$\hat{x}^c_{k+1|k} = \widehat{F}_k\hat{x}^c_{k|k} + \widehat{G}_ku_k$$

end for

Remark 4.3. *Analogous to the RKF in Algorithm 3.2, the RCKF proposed in Algorithm 4.3 also depends on the penalty parameter μ and the approximation parameter ξ. As also discussed in Section 3.2.3, one can tune μ based on the severity of uncertainties. For significant perturbations, smaller values of μ are recommended, otherwise, it can be increased for mild uncertainties. Regarding the ξ parameter, choosing values within the* $(0, 1)$ *interval generally yields satisfactory results.*

4.2.3 Robust Distributed Kalman Consensus Filter

In this section, we address the distributed estimation problem described in Section 4.2.1.3. As established earlier, we use the average consensus algorithm (Section 2.4) to derive a fully distributed approximation of the Robust Centralized Kalman Filter presented in Algorithm 4.3, which serves as our benchmark.

We assume that each sensor $i \in \mathbb{S}$ is initialized with the same prior state estimate $\hat{x}^i_{0|-1}$ and prior error weighting matrix $P^i_{0|-1} \succ 0$. By adopting the hybrid consensus on measurements and information (HCMCI) approach (BATTISTELLI *et al.*, 2015), each sensor $i \in \mathbb{S}$ exchanges information with its neighbors $j \in \mathcal{N}_i$ to ultimately obtain approximations of the filtered and predicted prior robust central state estimates in a distributed fashion.

In order to approximate the RCKF in Algorithm 4.3, note that the sensors first need to compute the $\hat{\lambda}_k$ parameter, which depends on $\mathcal{M}_{2,k}$, composed of all matrices $M^i_{2,k}$, $\forall i \in \mathbb{S}$. However, each sensor only has access to its own matrix and the matrices of its neighbors. Nevertheless, given the block diagonal structure of $\mathcal{M}_{2,k}$, one can apply a variant of the average consensus algorithm to compute $\hat{\lambda}_k$ in a distributed manner, as shown in Algorithm 4.4.

Algorithm 4.4 Distributed computation of $\hat{\lambda}_k$ (each node $i \in \mathbb{S}$)

Initialization: Set the initial consensus state

$$\hat{\lambda}^i_k(0) = (1+\xi)\,\mu \left\| \mathbf{diag}\left(M^T_{1,k} M_{1,k},\, (M^i_{2,k})^T M^i_{2,k}\right) \right\|$$

for $\ell = 0, 1, \ldots, L-1$ **do**

1. Send the current $\hat{\lambda}^i_k(\ell)$ to all neighbors $j \in \mathcal{N}_i$.
2. Receive the current $\hat{\lambda}^j_k(\ell)$ from all neighbors $j \in \mathcal{N}_i$.
3. Update the consensus state

$$\hat{\lambda}^i_k(\ell+1) = \max\left\{\hat{\lambda}^i_k(\ell),\, \hat{\lambda}^j_k(\ell)\right\},\ \forall j \in \mathcal{N}_i$$

end for

Output: $\hat{\lambda}^i_k(L) = \hat{\lambda}_k$

Algorithm 4.5 displays the proposed Robust Distributed Kalman Consensus Filter (RDKCF). Following the HCMCI protocol, in steps 4 and 5, we simultaneously perform the average consensus algorithm (Algorithm 2.2) for each sensor's prior information and innovation pairs, denoted $\left(\Omega^i_k,\, \omega^i_k\right)$ and $\left(\delta\Omega^i_k,\, \delta\omega^i_k\right)$, respectively. Notice that, in step 5.3, the consensus weights π_{ij} should satisfy the conditions listed in Definition 2.2, such that the consensus states of each node i are updated with a convex combination of the corresponding states within its inclusive neighborhood. The Metropolis weights (XIAO; BOYD; LALL, 2005), shown in (2.20), satisfy these conditions and are thus adopted here. Considering

Algorithm 4.5 Robust Distributed Kalman Consensus Filter (RDKCF) (each sensor i)

Model: Assume the uncertain system model in (4.16)-(4.17).

Initialization: Set $\hat{x}^i_{0|-1}$, $P^i_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R^i_k \succ 0$, $\mu > 0$, $\xi > 0$, and $L \geq 1$.

for $k = 0, 1, \ldots, N$ **do**

1. Obtain a measurement y^i_k.
2. Compute $\hat{\lambda}_k$ using Algorithm 4.4.
3. Compute the modified system and sensing model matrices:

$$\Phi_{1,k} = \mu^{-1} I_n - \hat{\lambda}_k^{-1} M_{1,k} M_{1,k}^T \qquad \widehat{Q}_k = \Phi_{1,k} + H_k \Big(Q_k^{-1} + \hat{\lambda}_k E_{H_k}^T E_{H_k}\Big)^{-1} H_k^T$$

$$\Phi^i_{2,k} = \mu^{-1} I_r - \hat{\lambda}_k^{-1} M^i_{2,k} (M^i_{2,k})^T \qquad \widehat{R}^i_k = \Phi^i_{2,k} + D^i_k \Big[(R^i_k)^{-1} + \hat{\lambda}_k (E^i_{D_k})^T E^i_{D_k}\Big]^{-1} (D^i_k)^T$$

$$\bar{Q}_k = \hat{\lambda}_k^{-1} I_{t_1} + E_{H_k} Q_k E_{H_k}^T \qquad \bar{R}^i_k = \hat{\lambda}_k^{-1} I_{t_2} + E^i_{D_k} R^i_k (E^i_{D_k})^T$$

$$\widehat{F}_k = F_k - H_k Q_k E_{H_k}^T \bar{Q}_k^{-1} E_{F_k} \qquad \widehat{C}^i_k = C^i_k - D^i_k R^i_k (E^i_{D_k})^T (\bar{R}^i_k)^{-1} E^i_{C_k}$$

$$\widehat{G}_k = G_k - H_k Q_k E_{H_k}^T \bar{Q}_k^{-1} E_{G_k}$$

4. Initialize the consensus states:

$$\Omega^i_k(0) = (P^i_{k|k-1})^{-1} \qquad \delta\Omega^i_k(0) = (\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} \widehat{C}^i_k + (E^i_{C_k})^T (\bar{R}^i_k)^{-1} E^i_{C_k}$$

$$\omega^i_k(0) = (P^i_{k|k-1})^{-1} \hat{x}^i_{k|k-1} \qquad \delta\omega^i_k(0) = (\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k$$

5. Consensus step:

for $\ell = 0, 1, \ldots, L-1$ **do**

5.1. Send $\Big\{\Omega^i_k(\ell),\ \omega^i_k(\ell),\ \delta\Omega^i_k(\ell),\ \delta\omega^i_k(\ell)\Big\}$ to all neighbors $j \in \mathcal{N}_i$.

5.2. Receive $\Big\{\Omega^j_k(\ell),\ \omega^j_k(\ell),\ \delta\Omega^j_k(\ell),\ \delta\omega^j_k(\ell)\Big\}$ from all neighbors $j \in \mathcal{N}_i$.

5.3. Update the consensus states:

$$\Omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \Omega^j_k(\ell) \qquad \delta\Omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \delta\Omega^j_k(\ell)$$

$$\omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \omega^j_k(\ell) \qquad \delta\omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \delta\omega^j_k(\ell)$$

end for

6. Correction step:

6.1. Compute the posterior error weighting matrix:

$$P^i_{k|k} = \Big[\Omega^i_k(L) + \rho^i_k\, \delta\Omega^i_k(L) + E_{F_k}^T \bar{Q}_k^{-1} E_{F_k}\Big]^{-1}$$

6.2. Compute the filtered state estimate:

$$\hat{x}^i_{k|k} = P^i_{k|k} \Big[\omega^i_k(L) + \rho^i_k\, \delta\omega^i_k(L) - E_{F_k}^T \bar{Q}_k^{-1} E_{G_k} u_k\Big]$$

7. Prediction step:

7.1. Update the predicted prior error weighting matrix:

$$P^i_{k+1|k} = \widehat{F}_k P^i_{k|k} \widehat{F}_k^T + \widehat{Q}_k$$

7.2. Update the predicted prior state estimate:

$$\hat{x}^i_{k+1|k} = \widehat{F}_k \hat{x}^i_{k|k} + \widehat{G}_k u_k$$

end for

the outcome of the consensus step and based on step 4 of the RCKF (Algorithm 4.3), we perform the correction stage outlined in step 6 of Algorithm 4.5. Note that, as in the nominal distributed case (Algorithm 4.2), we also introduce the corrective scalar weight ρ_k^i to compensate for the possible underweighting of the innovation pair $\left(\delta\Omega_k^i,\, \delta\omega_k^i\right)$ due to the average consensus process. Then, we conclude the algorithm with the prediction stage in step 7, which corresponds to step 4 of the RCKF (Algorithm 4.3).

The next theorem shows that for a large number of consensus iterations, the proposed robust distributed filter attains the same result as its centralized counterpart.

Theorem 4.5. *Consider the Robust Distributed Kalman Consensus Filter in Algorithm 4.5 and that Assumption 4.1 is satisfied. Assume that the consensus weights π_{ij} are chosen according to Definition 2.2, the number of consensus iterations $L \to \infty$ in step 5, and that $\rho_k^i = S$ in step 6. Then, the filtered and predicted prior robust state estimates, $\hat{x}_{k|k}^i$ and $\hat{x}_{k+1|k}^i$, and their respective error weighting matrices, $P_{k|k}^i$ and $P_{k+1|k}^i$, obtained by each sensor $i \in \mathbb{S}$ converge to the corresponding robust central state estimates $\hat{x}_{k|k}^c$ and $\hat{x}_{k+1|k}^c$, and error weighting matrices $P_{k|k}^c$ and $P_{k+1|k}^c$ computed via the Robust Centralized Kalman Filter in Algorithm 4.3.*

Proof. Since the undirected graph $\mathbb{G}$ describing the sensor network is connected and the consensus weights π_{ij} are properly selected, the associated weighted adjacency matrix Π has the properties listed in Lemma 2.9. Assuming that the number of consensus iterations $L \to \infty$, the convergence of the average consensus algorithm is guaranteed (Theorem 2.2).

Through induction, we now prove that the RDKCF detailed in Algorithm 4.5 converges to the RCKF in Algorithm 4.3. At time step $k = 0$, assume that the RCKF is initialized with $\hat{x}_{0|-1}^c = \hat{x}_0$ and $P_{0|-1}^c = P_0 \succ 0$, whereas all sensors $i \in \mathbb{S}$ initialize the RDKCF with $\hat{x}_{0|-1}^i = \hat{x}_0$ and $P_{0|-1}^i = P_0 \succ 0$. Then, Theorem 2.2 indicates that after the consensus step 5 of the RDKCF, the information and innovation pairs of all the sensors converge in the following way:

$$\Omega_0^i(L) \to \frac{1}{S}\sum_{j=1}^{S} P_0^{-1} = P_0^{-1}, \qquad \delta\Omega_0^i(L) \to \frac{1}{S}\sum_{j=1}^{S}\Big[(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}\widehat{C}_0^j + (E_{C_0}^j)^T(\bar{R}_0^j)^{-1}E_{C_0}^j\Big],$$

$$\omega_0^i(L) \to \frac{1}{S}\sum_{j=1}^{S} P_0^{-1}\hat{x}_0 = P_0^{-1}\hat{x}_0, \quad \delta\omega_0^i(L) \to \frac{1}{S}\sum_{j=1}^{S}(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}y_0^j.$$

Then, substituting these consensus outcomes into the equations in step 6 and considering that the corrective scalar weight $\rho_0^i = S$, we obtain

$$P_{0|0}^i \to \left[P_0^{-1} + S\frac{1}{S}\sum_{j=1}^{S}\Big[(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}\widehat{C}_0^j + (E_{C_0}^j)^T(\bar{R}_0^j)^{-1}E_{C_0}^j\Big] + E_{F_0}^T\bar{Q}_0^{-1}E_{F_0}\right]^{-1} =$$

$$\left[P_0^{-1} + \sum_{j=1}^{S}\Big[(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}\widehat{C}_0^j + (E_{C_0}^j)^T(\bar{R}_0^j)^{-1}E_{C_0}^j\Big] + E_{F_0}^T\bar{Q}_0^{-1}E_{F_0}\right]^{-1} = P_{0|0}^c,$$

$$\hat{x}^i_{0|0} \to P^i_{0|0}\left[P_0^{-1}\hat{x}_0 + S\frac{1}{S}\sum_{j=1}^{S}(\widehat{C}^j_0)^T(\widehat{R}^j_0)^{-1}y^j_0 - E^T_{F_0}\bar{R}_0^{-1}E_{G_0}u_0\right] = \\ P^c_{0|0}\left[P_0^{-1}\hat{x}_0 + \sum_{j=1}^{S}(\widehat{C}^j_0)^T(\widehat{R}^j_0)^{-1}y^j_0 - E^T_{F_0}\bar{Q}_0^{-1}E_{G_0}u_0\right] = \hat{x}^c_{0|0},$$

for all sensors $i \in \mathbb{S}$. Notice how the choice of scalar weight $\rho^i_0 = S$ correctly compensates for the $1/S$ factor that appears in the outcome of the innovation pair $\left(\delta\Omega^i_0,\, \delta\omega^i_0\right)$ due to the averaging process. The convergence above implies that in step 7, $P^i_{1|0} \to P^c_{1|0}$ and $\hat{x}^i_{1|0} \to \hat{x}^c_{1|0}$. Hence, for $k = 0$, we have that the RDKCF indeed converges to the RCKF.

Now, let us assume that at time step $k-1$, one has $P^i_{k-1|k-1} \to P^c_{k-1|k-1}$, $\hat{x}^i_{k-1|k-1} \to \hat{x}^c_{k-1|k-1}$, $P^i_{k|k-1} \to P^c_{k|k-1}$, and $\hat{x}^i_{k|k-1} \to \hat{x}^c_{k|k-1}$, $\forall i \in \mathbb{S}$. Then, based on Theorem 2.2, at time step k, we achieve the following consensus outcome after performing step 5 of the RDKCF:

$$\Omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}(P^c_{k|k-1})^{-1} = (P^c_{k|k-1})^{-1},$$
$$\delta\Omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}\left[(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}\widehat{C}^j_k + (E^j_{C_k})^T(\bar{R}^j_k)^{-1}E^j_{C_k}\right],$$
$$\omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} = (P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1},$$
$$\delta\omega^i_k(L) \to \frac{1}{S}\sum_{j=1}^{S}(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}y^j_k.$$

Thus, substituting these outcomes back into the equations in step 6 of the RDKCF, considering $\rho^i_k = S$, yields

$$P^i_{k|k} \to \left[(P^c_{k|k-1})^{-1} + S\frac{1}{S}\sum_{j=1}^{S}\left[(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}\widehat{C}^j_k + (E^j_{C_k})^T(\bar{R}^j_k)^{-1}E^j_{C_k}\right] + E^T_{F_k}\bar{Q}_k^{-1}E_{F_k}\right]^{-1} = \\ \left[(P^c_{k|k-1})^{-1} + \sum_{j=1}^{S}\left[(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}\widehat{C}^j_k + (E^j_{C_k})^T(\bar{R}^j_k)^{-1}E^j_{C_k}\right] + E^T_{F_k}\bar{Q}_k^{-1}E_{F_k}\right]^{-1} = P^c_{k|k},$$

$$\hat{x}^i_{k|k} \to P^i_{k|k}\left[(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} + S\frac{1}{S}\sum_{j=1}^{S}(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}y^j_k - E^T_{F_k}\bar{Q}_k^{-1}E_{G_k}u_k\right] = \\ P^c_{k|k}\left[(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} + \sum_{j=1}^{S}(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}y^j_k - E^T_{F_k}\bar{Q}_k^{-1}E_{G_k}u_k\right] = \hat{x}^c_{k|k},$$

for all sensors $i \in \mathbb{S}$. Finally, plugging the results above back into the equations in step 7 of the RDKCF gives us that $P^i_{k+1|k} \to P^c_{k+1|k}$ and $\hat{x}^i_{k+1|k} \to \hat{x}^c_{k+1|k}$. Hence, under the aforementioned conditions, by induction, we have that for $k = 0, 1, \ldots, N$, the RDKCF in Algorithm 4.5 converges to the RCKF in Algorithm 4.3. □

The robust distributed filter proposed in Algorithm 4.5 depends on the parameters μ and ξ. Remark 4.3 presents some guidelines on how to tune them based on the level of parametric uncertainties. As commented in the nominal distributed case, we reiterate how the HCMCI protocol combined with the average consensus algorithm allowed us to design a robust estimator that, despite being distributed, is able to approach the performance of a centralized estimator. Of course, true convergence cannot be achieved in practice, as only a finite number of consensus iterations L is possible. Nonetheless, we show with an illustrative example that with a sufficiently large L it is still possible to closely approximate the performance of the centralized filter using a distributed setup.

Remark 4.4. *As in the Nominal Distributed Kalman Consensus Filter presented in Algorithm 4.2, we also introduce a corrective scalar weight ρ_k^i in step 6 of Algorithm 4.5 to avoid the possible underweighting of the innovation pair $\left(\delta\Omega_k^i(L),\, \delta\omega_k^i(L)\right)$ due to scaling from the average consensus procedure. This turns Algorithm 4.5 into a family of robust distributed filters, depending on the choice of ρ_k^i. As explained in Remark 4.2, ideally, we should have $\rho_k^i = S$ to correctly approximate the centralized performance. The remark also provides a distributed procedure to compute S in case it is not available to the sensors.*

4.2.4 Stability Analysis

In this section, we discuss the stability properties of both the proposed Robust Centralized Kalman Filter and the Robust Distributed Kalman Consensus Filter, as well as the boundedness of their estimation error variance. To this end, we examine the steady-state behavior of Algorithm 4.3 and Algorithm 4.5 when the target system and sensing model parameters are constant and there is no input u_k. Nonetheless, we still assume that the contraction matrices $\Delta_{1,k}$ and $\Delta_{2,k}^i$ are time-varying, $\forall i \in \mathbb{S}$. Thus, for $k \geq 0$, equations (4.16)-(4.17) take the form

$$\begin{aligned} x_{k+1} &= (F + \delta F_k)x_k + (H + \delta H_k)w_k, \\ y_k^i &= (C^i + \delta C_k^i)x_k + (D^i + \delta D_k^i)v_k^i, \quad \forall i \in \mathbb{S}, \end{aligned} \tag{4.25}$$

with time-varying norm-bounded parametric uncertainties

$$\begin{aligned} \begin{bmatrix} \delta F_k & \delta H_k \end{bmatrix} &= M_1 \Delta_{1,k} \begin{bmatrix} E_F & E_H \end{bmatrix}, \quad \|\Delta_{1,k}\| \leq 1, \\ \begin{bmatrix} \delta C_k^i & \delta D_k^i \end{bmatrix} &= M_2^i \Delta_{2,k}^i \begin{bmatrix} E_C^i & E_D^i \end{bmatrix}, \quad \|\Delta_{2,k}^i\| \leq 1, \quad \forall i \in \mathbb{S}. \end{aligned} \tag{4.26}$$

Based on the strategy adopted in Kamal, Farrell and Roy-Chowdhury (2013) and in Section 4.1.4, we conduct our analysis under the assumptions described in Theorem 4.5, i.e., assuming that the RDKCF converges to the RCKF. This allows us to extend the stability properties of the robust centralized filter to its distributed counterpart.

Let us first study the stability of the robust centralized filter presented in Algorithm 4.3. Consider the time-invariant system model (4.25)-(4.26). Then, the RCKF equations in steps 4 and 5 of Algorithm 4.3 become:

$$P^c_{k|k} = \Big[(P^c_{k|k-1})^{-1} + \widehat{\mathfrak{C}}^T \widehat{\mathfrak{R}}^{-1} \widehat{\mathfrak{C}} + \boldsymbol{E}_{\mathfrak{C}}^T \bar{\mathfrak{R}}^{-1} \boldsymbol{E}_{\mathfrak{C}} + E_F^T \bar{Q}^{-1} E_F\Big]^{-1}, \tag{4.27}$$

$$\hat{x}^c_{k|k} = P^c_{k|k}\Big[(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} + \widehat{\mathfrak{C}}^T \widehat{\mathfrak{R}}^{-1} \mathfrak{Y}_k\Big], \tag{4.28}$$

$$P^c_{k+1|k} = \widehat{F} P^c_{k|k} \widehat{F}^T + \widehat{Q}, \tag{4.29}$$

$$\hat{x}^c_{k+1|k} = \widehat{F} \hat{x}^c_{k|k}, \tag{4.30}$$

where

$$\mathfrak{Y}_k = \begin{bmatrix} y^1_k \\ \vdots \\ y^S_k \end{bmatrix}, \ \widehat{\mathfrak{C}} = \begin{bmatrix} \widehat{C}^1 \\ \vdots \\ \widehat{C}^S \end{bmatrix}, \ \boldsymbol{E}_{\mathfrak{C}} = \begin{bmatrix} E^1_C \\ \vdots \\ E^S_C \end{bmatrix}, \ \widehat{\mathfrak{R}} = \begin{bmatrix} \widehat{R}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widehat{R}^S \end{bmatrix}, \text{ and } \bar{\mathfrak{R}} = \begin{bmatrix} \bar{R}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \bar{R}^S \end{bmatrix},$$

with each $\widehat{C}^i$, $\widehat{R}^i$, and $\bar{R}^i$, $\forall i \in \mathbb{S}$, as well as $\widehat{F}$, $\widehat{Q}$, and $\bar{Q}$ given by the corresponding equations listed in step 3 of Algorithm 4.3, considering constant parameters. The constant $\hat{\lambda}$ parameter is analogously computed as in step 2. To simplify the notation, we further define the augmented matrices

$$\widetilde{\mathfrak{C}} := \begin{bmatrix} \widehat{\mathfrak{C}} \\ \boldsymbol{E}_{\mathfrak{C}} \\ E_F \end{bmatrix} \quad \text{and} \quad \widetilde{\mathfrak{R}} := \begin{bmatrix} \widehat{\mathfrak{R}} & 0 & 0 \\ 0 & \bar{\mathfrak{R}} & 0 \\ 0 & 0 & \bar{Q} \end{bmatrix},$$

such that one can rewrite $P^c_{k|k}$ in (4.27) in a more compact way, as

$$P^c_{k|k} = \Big[(P^c_{k|k-1})^{-1} + \widetilde{\mathfrak{C}}^T \widetilde{\mathfrak{R}}^{-1} \widetilde{\mathfrak{C}}\Big]^{-1}.$$

Then, we apply Lemma A.1 to expand this expression and obtain

$$P^c_{k|k} = P^c_{k|k-1} - P^c_{k|k-1}\widetilde{\mathfrak{C}}^T \Big(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c_{k|k-1} \widetilde{\mathfrak{C}}^T\Big)^{-1} \widetilde{\mathfrak{C}} P^c_{k|k-1}. \tag{4.31}$$

Combining (4.31) with (4.28) and substituting back into (4.30) gives us the steady-state predicted robust central state estimate

$$\hat{x}^c_{k+1|k} = \widetilde{\mathfrak{F}}_k \hat{x}^c_{k|k-1} + \widetilde{\mathfrak{F}}_k P^c_{k|k-1} \widehat{\mathfrak{C}}^T \widehat{\mathfrak{R}}^{-1} \mathfrak{Y}_k, \tag{4.32}$$

in which

$$\widetilde{\mathfrak{F}}_k = \widehat{F}\Big(I_n - P^c_{k|k-1}\widetilde{\mathfrak{C}}^T \Big(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c_{k|k-1} \widetilde{\mathfrak{C}}^T\Big)^{-1} \widetilde{\mathfrak{C}}\Big)$$

is the robust centralized filter closed-loop matrix. Moreover, substituting $P^c_{k|k}$ from (4.31) into (4.29) yields the expression for the predicted prior error weighting matrix:

$$P^c_{k+1|k} = \widehat{F}\Big(P^c_{k|k-1} - P^c_{k|k-1}\widetilde{\mathfrak{C}}^T \Big(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c_{k|k-1} \widetilde{\mathfrak{C}}^T\Big)^{-1} \widetilde{\mathfrak{C}} P^c_{k|k-1}\Big) \widehat{F}^T + \widehat{Q}. \tag{4.33}$$

Theorem 4.6. *Consider the linear system model* (4.25) *with norm-bounded uncertainties* (4.25) *and the corresponding robust centralized filter* (4.32)-(4.33). *Assume that* $\{\widehat{F}, \widetilde{\mathfrak{C}}\}$ *is* detectable *and* $\{\widehat{F}, \widehat{Q}^{1/2}\}$ *is* controllable. *Then, for any initial condition* $P^c_{0|-1} \succ 0$, $\xi > 0$, *and* $\mu > 0$, $P^c_{k+1|k}$ *converges to the* unique stabilizing solution $P^c \succ 0$ *of the algebraic Riccati equation*

$$P^c = \widehat{F}\Big(P^c - P^c\widetilde{\mathfrak{C}}^T\Big(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}}P^c\widetilde{\mathfrak{C}}^T\Big)^{-1}\widetilde{\mathfrak{C}}P^c\Big)\widehat{F}^T + \widehat{Q}. \tag{4.34}$$

The solution P^c *is stabilizing in the sense that the steady-state filter closed-loop matrix*

$$\widetilde{\mathfrak{F}} = \widehat{F}\Big(I_n - P^c\widetilde{\mathfrak{C}}^T\Big(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}}P^c\widetilde{\mathfrak{C}}^T\Big)^{-1}\widetilde{\mathfrak{C}}\Big) \tag{4.35}$$

is Schur stable.

Proof. The conditions $\xi > 0$ and $\mu > 0$ imply that $\hat{\lambda} > 0$, ensuring that matrices $\widehat{F}$, $\widetilde{\mathfrak{C}}$, $\widetilde{\mathfrak{R}}$, and $\widehat{Q}$ are well-defined. According to Kailath, Sayed and Hassibi (2000b), detectability of $\{\widehat{F}, \widetilde{\mathfrak{C}}\}$ and controllability of $\{\widehat{F}, \widehat{Q}^{1/2}\}$ ensure the convergence of $P^c_{k+1|k}$ in (4.33) to the unique stabilizing positive definite solution P^c of the algebraic Riccati equation (4.34) that stabilizes (4.35), which is the robust centralized filter steady-state closed-loop matrix. □

We now investigate the conditions for the boundedness of the estimation error variance of the RCKF. Again, consider the uncertain linear discrete-time system model (4.25)-(4.26). In addition, assume that w_k and $\{v_k\}_{i=1}^S$ are mutually uncorrelated zero-mean Gaussian noise processes with joint covariance matrix

$$\mathcal{Q} = \boldsymbol{E}\left\{\begin{bmatrix} w_k \\ \boldsymbol{v}_k \end{bmatrix}\begin{bmatrix} w_k^T & \boldsymbol{v}_k^T \end{bmatrix}\right\} = \begin{bmatrix} Q & 0 \\ 0 & \mathfrak{R} \end{bmatrix} \succ 0, \tag{4.36}$$

in which $\boldsymbol{v}_k = \mathbf{col}\left(v_k^1, \ldots, v_k^S\right)$ and $\mathfrak{R} = \mathbf{diag}\left(R^1, \ldots, R^S\right)$. Moreover, assume that there is no correlation between the parametric uncertainties and the system and measurement noises. Finally, consider the following assumptions about the uncertain system and the robust centralized filter.

Assumption 4.3. *The uncertain system* (4.25)-(4.26) *is quadratically stable, according to Definition 3.1.*

Assumption 4.4. *The conditions of Theorem 4.6 are satisfied, such that the robust centralized filter steady-state closed-loop matrix* $\widetilde{\mathfrak{F}}$ *is Schur stable.*

Now, we show that under Assumption 3.1 and Assumption 3.2, the steady-state robust centralized filter (4.32) is also quadratically stable. To simplify the notation, we also define the following steady-state filter gain

$$\widetilde{\mathfrak{K}} := \widetilde{\mathfrak{F}}P^c\widehat{\mathfrak{C}}^T\widehat{\mathfrak{R}}^{-1},$$

with $\widetilde{\mathfrak{F}}$ given by (4.35), in which P^c is the stabilizing solution of the algebraic Riccati equation (4.34). Then, the steady-state robust centralized filter equation can be rewritten as

$$\hat{x}^c_{k+1|k} = \widetilde{\mathfrak{F}}\hat{x}^c_{k|k-1} + \widetilde{\mathfrak{K}}\boldsymbol{\mathfrak{Y}}_k, \tag{4.37}$$

where $\boldsymbol{\mathfrak{Y}}_k = \mathbf{col}\left(y^1_k, \ldots, y^S_k\right)$. Now, substituting each y^i_k from (4.25) into (4.37) yields

$$\hat{x}^c_{k+1|k} = \widetilde{\mathfrak{F}}\hat{x}^c_{k|k-1} + \widetilde{\mathfrak{K}}(\mathfrak{C} + \boldsymbol{\delta}\mathfrak{C}_k)x_k + \widetilde{\mathfrak{K}}(\mathfrak{D} + \boldsymbol{\delta}\mathfrak{D}_k)\boldsymbol{v}_k, \tag{4.38}$$

in which the aggregate matrices $\mathfrak{C}$, $\boldsymbol{\delta}\mathfrak{C}_k$, $\mathfrak{D}$, and $\boldsymbol{\delta}\mathfrak{D}_k$ are defined in (4.20), (4.21), and (4.22). Furthermore, we introduce the central state estimation error vector $e^c_k := x_k - \hat{x}^c_{k|k-1}$. Then, we subtract (4.38) from x_{k+1} in (4.25) to obtain

$$e^c_{k+1} = \left[(F - \widetilde{\mathfrak{F}} - \widetilde{\mathfrak{K}}\mathfrak{C}) + (\delta F_k - \widetilde{\mathfrak{K}}\boldsymbol{\delta}\mathfrak{C}_k)\right]x_k + \widetilde{\mathfrak{F}}e^c_k + (H + \delta H_k)w_k - \widetilde{\mathfrak{K}}(\mathfrak{D} + \boldsymbol{\delta}\mathfrak{D}_k)\boldsymbol{v}_k. \tag{4.39}$$

Consider now the augmented system composed of the target system state x_k and the central estimation error e^c_k. Thus, from (4.25), (4.26), and (4.39), this augmented system is described by

$$\begin{aligned} \zeta^c_{k+1} &= (\mathscr{F} + \boldsymbol{\delta}\mathscr{F}_k)\zeta^c_k + (\mathscr{H} + \boldsymbol{\delta}\mathscr{H}_k)\eta^c_k, \\ \begin{bmatrix}\boldsymbol{\delta}\mathscr{F}_k & \boldsymbol{\delta}\mathscr{H}_k\end{bmatrix} &= \mathscr{M}\,\boldsymbol{\Delta}_k \begin{bmatrix}\boldsymbol{E}_{\mathscr{F}} & \boldsymbol{E}_{\mathscr{H}}\end{bmatrix}, \end{aligned} \tag{4.40}$$

where

$$\zeta^c_k := \begin{bmatrix} x_k \\ e^c_k \end{bmatrix}, \quad \eta^c_k := \begin{bmatrix} w_k \\ \boldsymbol{v}_k \end{bmatrix}, \quad \mathscr{F} := \begin{bmatrix} F & 0 \\ F - \widetilde{\mathfrak{F}} - \widetilde{\mathfrak{K}}\mathfrak{C} & \widetilde{\mathfrak{F}} \end{bmatrix}, \quad \mathscr{H} := \begin{bmatrix} H & 0 \\ H & -\widetilde{\mathfrak{K}}\mathfrak{D} \end{bmatrix},$$

$$\mathscr{M} := \begin{bmatrix} M_1 & 0 \\ M_1 & -\widetilde{\mathfrak{K}}\mathfrak{M}_2 \end{bmatrix}, \quad \boldsymbol{\Delta}_k := \begin{bmatrix} \Delta_{1,k} & 0 \\ 0 & \boldsymbol{\Delta}_{2,k} \end{bmatrix}, \quad \boldsymbol{E}_{\mathscr{F}} := \begin{bmatrix} E_F & 0 \\ \boldsymbol{E}_{\mathfrak{C}} & 0 \end{bmatrix}, \quad \boldsymbol{E}_{\mathscr{H}} := \begin{bmatrix} E_H & 0 \\ 0 & \boldsymbol{E}_{\mathfrak{D}} \end{bmatrix},$$

in which the aggregate matrix definitions can be found in (4.20), (4.21), and (4.22).

Lemma 4.1. *Given that Assumption 4.3 and Assumption 4.4 hold, the augmented system (4.40) is quadratically stable.*

Proof. Since the augmented system matrix $\mathscr{F}$ is lower triangular with diagonal elements F and $\widetilde{\mathfrak{F}}$, which are both Schur stable, we have that $\mathscr{F}$ is also Schur stable. Moreover, note that

$$\begin{aligned} \boldsymbol{E}_{\mathscr{F}}(zI_{2n} - \mathscr{F})^{-1}\mathscr{M} &= \begin{bmatrix} E_F & 0 \\ \boldsymbol{E}_{\mathfrak{C}} & 0 \end{bmatrix} \begin{bmatrix} zI_n - F & 0 \\ -(F - \widetilde{\mathfrak{F}} - \widetilde{\mathfrak{K}}\mathfrak{C}) & zI_n - \widetilde{\mathfrak{F}} \end{bmatrix}^{-1} \begin{bmatrix} M_1 & 0 \\ M_1 & -\widetilde{\mathfrak{K}}\mathfrak{M}_2 \end{bmatrix} \\ &= \begin{bmatrix} E_F(zI_n - F)^{-1}M_1 & 0 \\ \boldsymbol{E}_{\mathfrak{C}}(zI_n - F)^{-1}M_1 & 0 \end{bmatrix} = \begin{bmatrix} E_F \\ \boldsymbol{E}_{\mathfrak{C}} \end{bmatrix} (zI_n - F)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix}. \end{aligned}$$

In addition, one has

$$F + M_1\Delta_{1,k}E_F = F + \begin{bmatrix} M_1 & 0 \end{bmatrix} \begin{bmatrix} \Delta_{1,k} & 0 \\ 0 & \boldsymbol{\Delta}_{2,k} \end{bmatrix} \begin{bmatrix} E_F \\ \boldsymbol{E}_{\mathfrak{C}} \end{bmatrix}.$$

Since system (4.25)-(4.26) is quadratically stable, according to Remark 3.4, we have

$$\left\| \begin{bmatrix} E_F \\ \boldsymbol{E}_{\mathcal{C}} \end{bmatrix} (zI_n - F)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix} \right\|_{\infty} < 1,$$

for all admissible contractions $\Delta_{1,k}$ and $\boldsymbol{\Delta}_{2,k}$. Therefore, $\left\| \boldsymbol{E}_{\mathscr{F}} (zI_{2n} - \mathscr{F})^{-1} \mathscr{M} \right\|_{\infty} < 1$ and the augmented system (4.40) is also quadratically stable. □

Now, let us define the covariance matrix of the augmented system state as $\mathscr{P}_k^c := \boldsymbol{E}\left\{ \zeta_k^c (\zeta_k^c)^T \right\}$. Then, it follows from (4.40) that $\mathscr{P}_k^c$ satisfies the Lyapunov recursion

$$\mathscr{P}_{k+1}^c = (\mathscr{F} + \boldsymbol{\delta}\mathscr{F}_k)\mathscr{P}_k^c(\mathscr{F} + \boldsymbol{\delta}\mathscr{F}_k)^T + (\mathscr{H} + \boldsymbol{\delta}\mathscr{H}_k)\mathscr{Q}(\mathscr{H} + \boldsymbol{\delta}\mathscr{H}_k)^T, \tag{4.41}$$

where $\mathscr{Q}$ is defined in (4.36).

Theorem 4.7. *Given that Assumption 4.3 and Assumption 4.4 hold, the state estimation error variance of the steady-state robust centralized filter* (4.37) *satisfies*

$$\lim_{k \to \infty} \boldsymbol{E}\left\{ e_k^c (e_k^c)^T \right\} \preceq \mathscr{V}_{22},$$

where $\mathscr{V}_{22}$ *is the* $(2,2)$ *block entry with the smallest trace among all* $(2,2)$ *block entries of matrices* $\mathscr{V} \succ 0$ *that satisfy the inequality*

$$(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}\boldsymbol{E}_{\mathscr{F}})\,\mathscr{V}\,(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}\boldsymbol{E}_{\mathscr{F}})^T + (\mathscr{H} + \mathscr{M}\boldsymbol{\Delta}\boldsymbol{E}_{\mathscr{H}})\,\mathscr{Q}\,(\mathscr{H} + \mathscr{M}\boldsymbol{\Delta}\boldsymbol{E}_{\mathscr{H}})^T - \mathscr{V} \preceq 0,$$

for all admissible contraction matrices $\boldsymbol{\Delta}$, *with* $\|\boldsymbol{\Delta}\| \leq 1$.

Proof. Lemma 4.1 indicates that the augmented system (4.40) is quadratically stable. Thus, from Definition 3.1, there exists a matrix $\mathscr{U} \succ 0$ such that

$$(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})\,\mathscr{U}\,(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})^T - \mathscr{U} \prec 0,$$

for any admissible contraction matrix $\boldsymbol{\Delta}_k$. As discussed in Petersen and McFarlane (1996) and Sayed (2001), the existence of matrix $\mathscr{U} \succ 0$ guarantees the existence of a sufficiently large scaling parameter $\epsilon > 0$, such that one can find a matrix $\mathscr{V} = \epsilon\,\mathscr{U}$ satisfying

$$(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})\,\mathscr{V}\,(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})^T + (\mathscr{H} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{H}})\,\mathscr{Q}\,(\mathscr{H} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{H}})^T \preceq \mathscr{V}.$$

Subtracting the recursion for the augmented system covariance (4.41) from the inequality above thus yields

$$(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})(\mathscr{V} - \mathscr{P}_k^c)(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})^T \preceq \mathscr{V} - \mathscr{P}_{k+1}^c,$$

or, equivalently,

$$\mathscr{V} - \mathscr{P}_{k+1}^c = (\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})(\mathscr{V} - \mathscr{P}_k^c)(\mathscr{F} + \mathscr{M}\boldsymbol{\Delta}_k\boldsymbol{E}_{\mathscr{F}})^T + \mathscr{W}_k,$$

for some $\mathscr{W}_k \succeq 0$. To conclude, since the augmented system is quadratically stable, as $k \to \infty$, $\mathscr{V} - \mathscr{P}_{k+1}^c \succeq 0$, or $\mathscr{P}_{k+1}^c \preceq \mathscr{V}$. The $(2,2)$ block entry of $\mathscr{P}_k^c$ corresponds to the estimation error variance, which is thus bounded. □

Corollary 4.2. *Given that the assumptions in Theorem 4.5 hold, as well as Assumptions 4.3 and 4.4, the RDKCF in Algorithm 4.5 converges to the RCKF in Algorithm 4.3 and thus shares its stability and bounded estimation error variance properties, according to Theorems 4.6 and 4.7.*

4.2.5 Illustrative Example

In this section, we assess the performance of the proposed RDKCF with an example adapted from Xie, Soh and Souza (1994) and Section 3.2.5. In addition, we also evaluate the RCKF, considered a benchmark for the distributed strategy. Furthermore, we compare our results with other existing robust distributed filtering strategies from the literature. For completeness, to establish a baseline, we further test the nominal centralized and distributed filters developed in Section 4.1.

Consider a linear discrete-time target-system with norm-bounded uncertainties, as described in (4.16)-(4.18) with the following constant parameter matrices:

$$F_k = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix}, \quad G_k = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad H_k = \begin{bmatrix} -6 \\ 1 \end{bmatrix},$$

$$M_{1,k} = \begin{bmatrix} 0 \\ 10 \end{bmatrix}, \quad E_{F_k} = \begin{bmatrix} 0.01 & 0.03 \end{bmatrix}, \quad E_{G_k} = 0, \quad E_{H_k} = 0.01.$$

There is no input signal u_k and w_k is a zero-mean white Gaussian noise signal with variance $Q_k = 1$. The initial state is $x_0 = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$.

A set of $S = 25$ sensors arranged in a random geometric undirected network, shown in Figure 8, measure the target system. The sensing model is described as in (4.17)-(4.18), with v_k^i as zero-mean white Gaussian noise signals with variances R_k^i. Two distinct types of sensors are considered. Sensors with odd number, i.e., $i = 1, 3, \dots, 25$, are of the first type, having constant parameter matrices

$$C_k^i = \begin{bmatrix} -100 & 9 \end{bmatrix}, \quad D_k^i = 1, \quad M_{2,k}^i = 10, \quad E_{C_k}^i = \begin{bmatrix} 0.01 & 0.03 \end{bmatrix}, \quad E_{D_k}^i = 0.01, \quad R_k^i = 1.$$

Sensors with even number, i.e., $i = 2, 4, \dots, 24$, are of the second type, with matrices

$$C_k^i = \begin{bmatrix} -50 & 12 \end{bmatrix}, \quad D_k^i = 1, \quad M_{2,k}^i = 15, \quad E_{C_k}^i = \begin{bmatrix} 0.01 & 0.03 \end{bmatrix}, \quad E_{D_k}^i = 0.02, \quad R_k^i = 0.8.$$

Then, we apply the proposed RDKCF in Algorithm 4.5 with the following initialization data for all sensors:

$$\hat{x}_{0|-1}^i = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \quad P_{0|-1}^i = I_2, \quad \mu = 0.01, \quad \xi = 0.01, \ \text{and} \ L = 10,$$

where μ and ξ are chosen according to the guidelines in Remark 4.3. In addition, for the consensus iterations, we adopt the Metropolis weights shown in (2.20).

Figure 8 – Sensor network with 25 nodes and 81 edges.

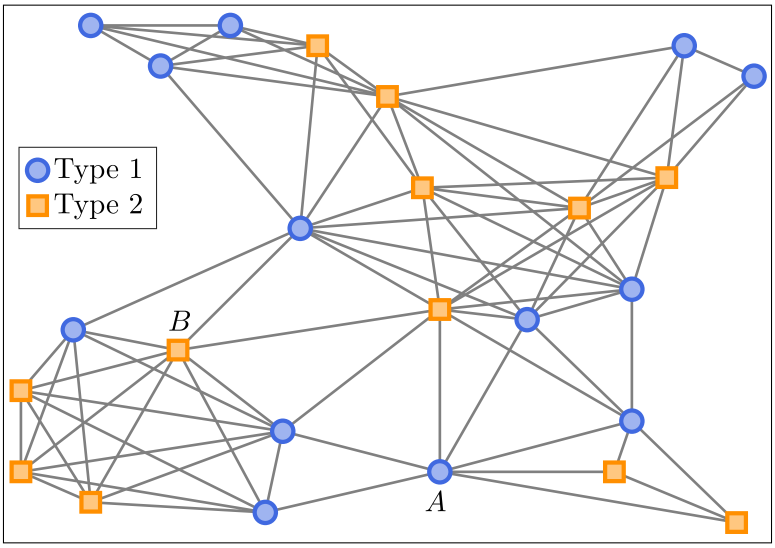


Figure 9 shows the evolution of the actual target system state along with the estimation performed by sensors A (Type 1) and B (Type 2), identified in Figure 8, using the proposed RDKCF. At each time step, $\Delta_{1,k}$ and $\Delta_{2,k}$ are real numbers randomly selected from a uniform distribution with interval $[-1, 1]$. The results show that both sensors were able to successfully track the state of the target system, irrespective of the norm-bounded parametric uncertainties. Moreover, their estimates are practically indistinguishable, which indicates that they reach consensus at each time step.

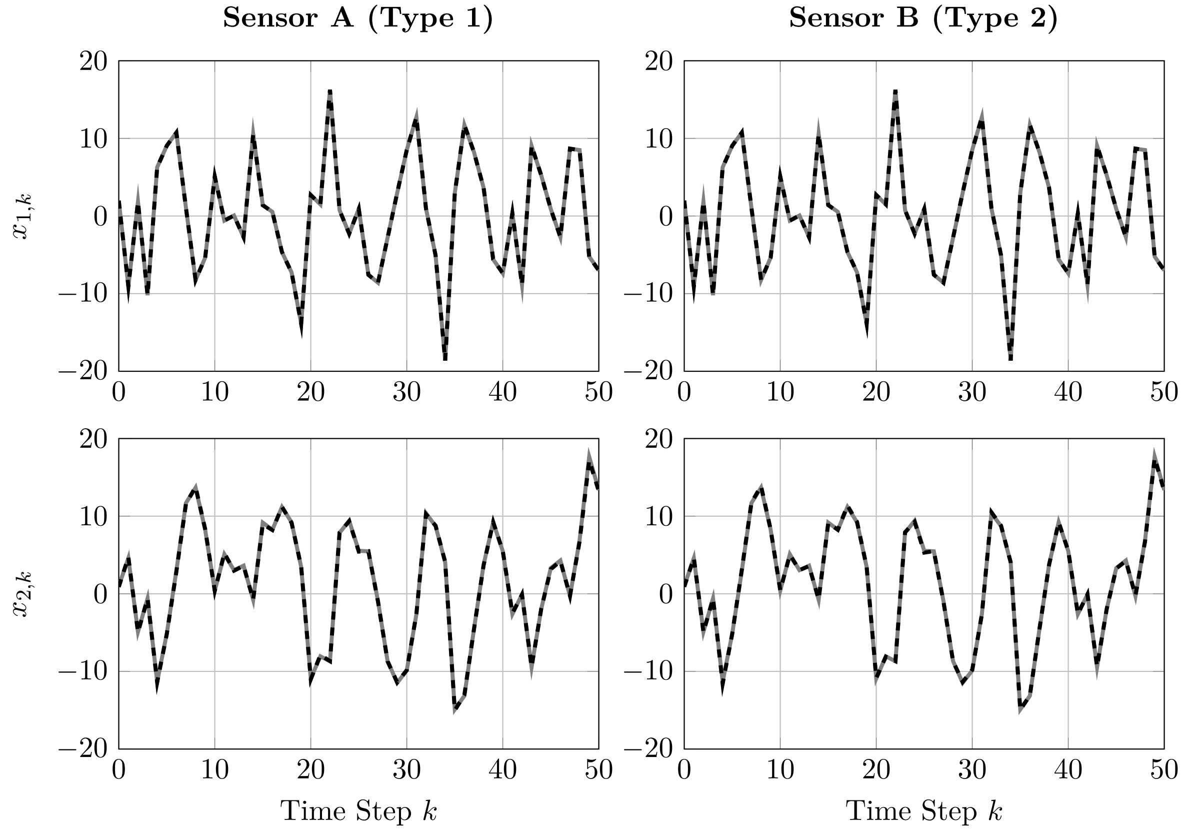
In order to further evaluate the proposed RDKCF, we carry out some comparisons. Moreover, we consider two distinct scenarios. In the first one, named RDKCF-1, we assume that the total number S of sensors in the network is available to every node, such that in step 6 of Algorithm 4.5, we choose $\rho_k^i = S$. In the second, RDKCF-2, the number S is estimated using the strategy in Remark 4.4. Then, we compare the results with those obtained with the robust centralized benchmark RCKF (Algorithm 4.3). To establish a baseline, we also apply the nominal centralized and distributed filters presented in Section 4.1, respectively CKF (Algorithm 4.1) and DKCF (Algorithm 4.2). In addition, we also compare the RDKCF with other robust distributed estimators from the literature, namely the recursive filters proposed in Rocha and Terra (2020) and Duan *et al.* (2020), as well as the $\mathcal{H}_\infty$-consensus filter presented by Shen, Wang and Hung (2010).

The simulation consists of performing $M = 1000$ Monte Carlo experiments, each with time horizon $N = 100$. At each time step k, we compute the mean squared estimation error (MSE), averaged over all experiments and sensors in the network, as follows:

$$\text{MSE}_k = \frac{1}{SM} \sum_{i=1}^{S} \sum_{e=1}^{M} \|x_k - \hat{x}_{k|k,e}^i\|^2,$$

which provides an approximation of the estimation error variance since, due to the parametric uncertainties, we cannot compute its actual value, as discussed in Sayed (2001).

Figure 9 – Actual (solid lines) and estimated (dashed lines) target system state obtained by sensors A and B with the proposed RDKCF (Algorithm 4.5).

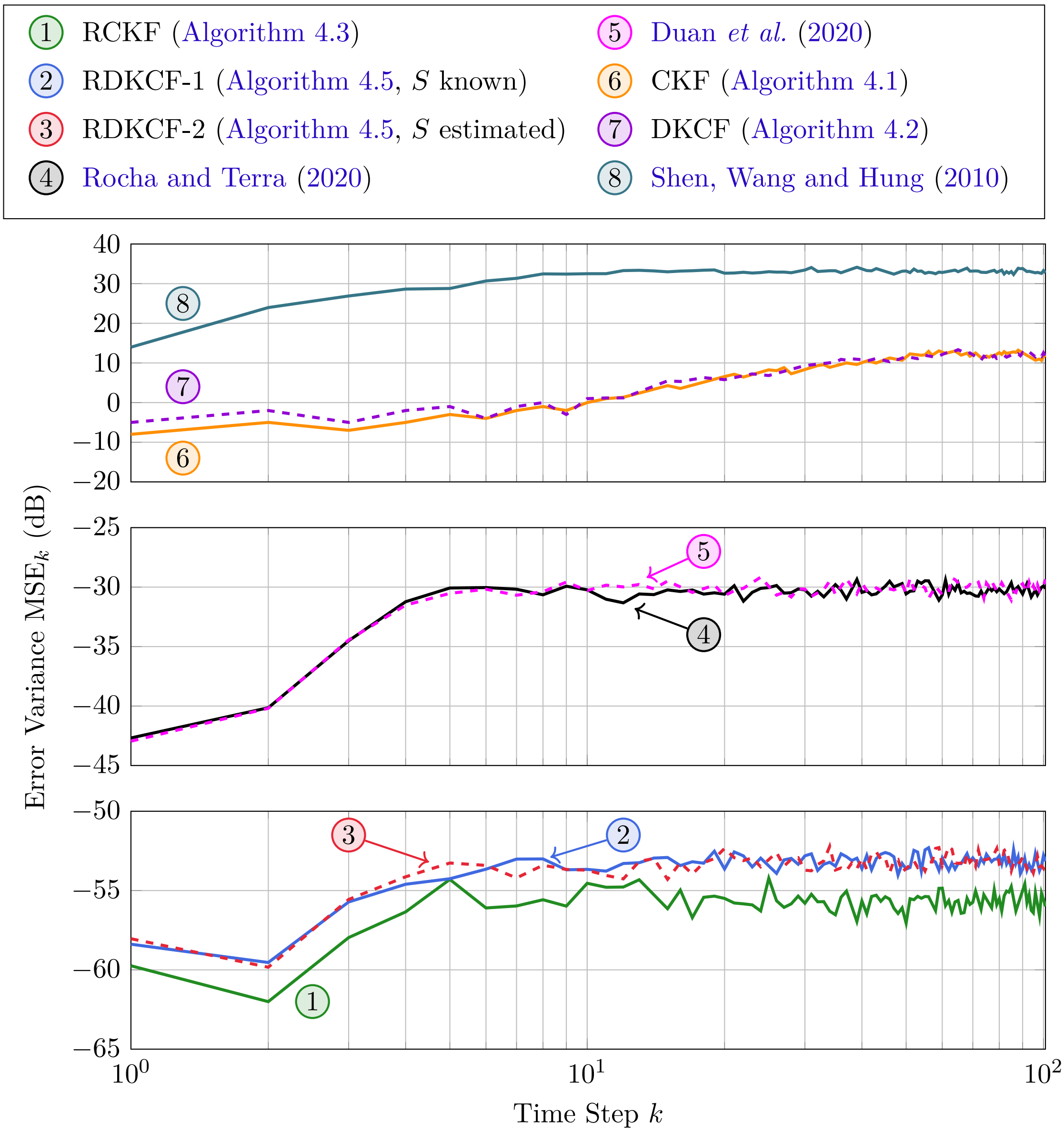


The results are presented in Figure 10 and Table 3, which summarizes the mean $\overline{\text{MSE}}$ and standard deviation $\sigma(\text{MSE})$ of the estimation error variances, calculated as

$$\overline{\text{MSE}} = \sum_{k=0}^{N} \frac{\text{MSE}_k}{N+1} \quad \text{and} \quad \sigma^2(\text{MSE}) = \sum_{k=0}^{N} \frac{(\text{MSE}_k - \overline{\text{MSE}})^2}{N+1}.$$

Since the RCKF has access to all sensors at once, it naturally presents the best possible performance, thus being considered a benchmark. Among the distributed approaches, the RDKCF exhibits the smallest estimation error. As expected, the RDKCF-1 (S is known) presents a slightly better performance compared to the RDKCF-2 (S is estimated). Moreover, they closely follow the RCKF, fulfilling their goal. Next, we have the robust distributed filters proposed by Rocha and Terra (2020) and Duan *et al.* (2020), which present a similar performance. The former considers uncertainties in all parameter matrices and performs a single consensus on information step, whereas the latter only assumes uncertainties in matrix F_k and does not use the consensus protocol. Then, we have the nominal centralized and distributed nominal estimators, CKF (Algorithm 4.1) and DKCF (Algorithm 4.2). They achieve similar results, but with significantly larger error variance, compared to the previous estimators, which was expected, as they do not compensate for model uncertainties. Finally, the robust $\mathcal{H}_\infty$-consensus filter by Shen, Wang and Hung (2010) obtained the highest error variance. It assumes uncertainties in matrices F_k and C_k^i and depends on the solution of an optimization problem subject to complex LMIs.

Figure 10 – Estimation error variance curves of the robust distributed filters.



Moreover, it is not fully distributed, as the gains are collectively computed, which may not scale well as the number of sensors increases. In terms of standard deviation, notice that both versions of the RDKCF exhibit smaller values compared to the RCKF. The nominal approaches present the largest deviations, again, due to the uncompensated uncertainties.

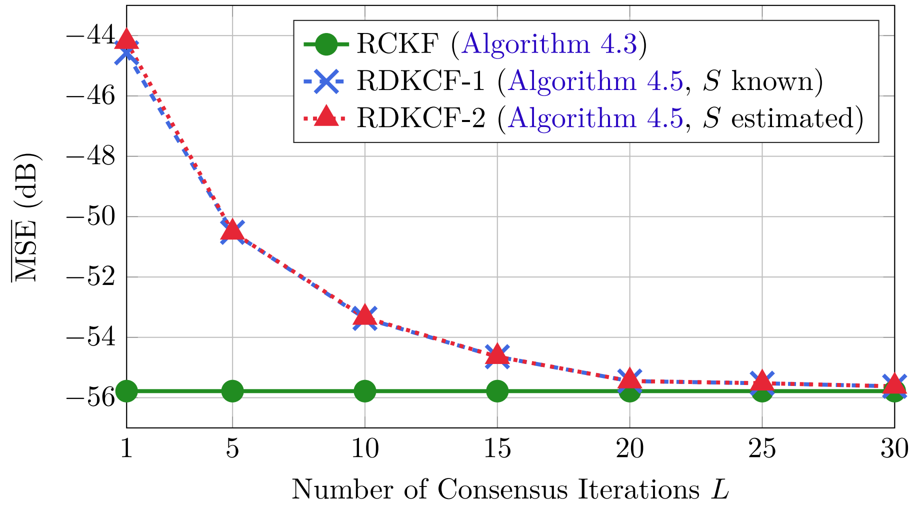
Furthermore, we investigate how the number of consensus iterations L affects the RDKCF performance. Figure 11 compiles a series of simulations with several values of L, considering both scenarios of the RDKCF. For each value, we compute the mean estimation error $\overline{\text{MSE}}$ over the whole time horizon, as previously described. The RCKF is also presented for comparison purposes. Note that, as the value of L increases, the distributed filters indeed approach the result of the centralized filter. This, however, requires more computation time, indicating a performance trade-off. In addition, the results show that the impact of knowing S beforehand or estimating it online is not significant, as both versions of the RDKCF present very similar results.

Table 3 – Estimation performance of each robust distributed filter.

Filter	$\overline{\text{MSE}}$ (dB)	σ(MSE) (dB)
① RCKF (Algorithm 4.3)	**-55.78**	0.9655
② RDKCF-1 (Algorithm 4.5, S known)	-53.37	**0.9461**
③ RDKCF-2 (Algorithm 4.5, S estimated)	-53.34	0.9558
④ Rocha and Terra (2020)	-30.55	1.6681
⑤ Duan *et al.* (2020)	-30.38	1.7212
⑥ CKF (Algorithm 4.1)	7.399	11.055
⑦ DKCF (Algorithm 4.2)	7.497	10.849
⑧ Shen, Wang and Hung (2010)	32.62	2.2911

Bold numbers indicate the smallest values.

Figure 11 – Effect of the number of consensus iterations L on the RDKCF (Algorithm 4.5).



4.3 Robust Distributed Kalman Filtering for Systems with Polytopic Uncertainties

In this section, we propose a second kind of robust alternatives to the nominal centralized and distributed Kalman filters presented in Section 4.1. We address the case where the underlying target system and sensing models are subject to polytopic parametric uncertainties, i.e., the parameters arbitrarily vary within a convex polyhedron centered at the nominal parameters (CHANG; PARK; TANG, 2015).

We follow a similar strategy as the one adopted in the previous norm-bounded case (Section 4.2). As such, we start by addressing the centralized estimation problem, assuming availability of measurements from all sensors in the network to a fusion center. Using a deterministic interpretation, we formulate this centralized estimation problem as a regularized least-squares estimation problem with polytopic uncertainties (Section 3.3.2).

From its solution, we extract the Polytopic Robust Centralized Kalman Filter (PRCKF), which is then modified to tackle the robust distributed estimation problem. This is achieved through the HCMCI protocol (BATTISTELLI *et al.*, 2015), yielding the Polytopic Robust Distributed Kalman Consensus Filter (PRDKCF). With sufficiently many consensus steps, both filters achieve similar results. As in the previous sections, we present the robust centralized and distributed estimators as recursive correction-prediction algorithms, which, in this case, resemble the Polytopic Robust Kalman Filter (Algorithm 3.3). Additionally, we study the stability properties of both filters and conclude the section with an illustrative example.

4.3.1 Problem Formulation

4.3.1.1 System Model

Consider a sensor network featuring S sensors. The communication among them is represented by the undirected graph $\mathbb{G} = (\mathbb{S}, \mathbb{E})$, with node set $\mathbb{S} = \{1, 2, \ldots, S\}$ and edge set $\mathbb{E} \subseteq \mathbb{S} \times \mathbb{S}$. The neighborhood of a sensor i is denoted by $\mathcal{N}_i = \{j \in \mathbb{S} \,|\, (i, j) \in \mathbb{E}\}$ and has cardinality N_i (see Section 2.3 for an introduction on graph theory).

Assumption 4.5. *The undirected graph* $\mathbb{G}$ *has a fixed topology and is connected, i.e., there is a path between every pair of nodes.*

Consider the following discrete-time state-space description of a linear target system subject to uncertainties:

$$x_{k+1} = (F_{0,k} + \delta F_k)x_k + (G_{0,k} + \delta G_k)u_k + (H_{0,k} + \delta H_k)w_k, \tag{4.42}$$

which is observed by the set of S sensors $\mathbb{S} = \{1, 2, \ldots, S\}$, each described by the uncertain model

$$y^i_k = (C^i_{0,k} + \delta C^i_k)x_k + (D^i_{0,k} + \delta D^i_k)v^i_k, \quad \forall i \in \mathbb{S}, \tag{4.43}$$

for $k = 0, 1, \ldots, N$, where $x_k \in \mathbb{R}^n$ is a state vector, $u_k \in \mathbb{R}^m$ is an input vector, and $w_k \in \mathbb{R}^p$ is the system noise. For each sensor $i \in \mathbb{S}$, $y^i_k \in \mathbb{R}^r$ is the measurement vector and $v^i_k \in \mathbb{R}^q$ is the measurement noise. $F_{0,k} \in \mathbb{R}^{n\times n}$, $G_{0,k} \in \mathbb{R}^{n\times m}$, $H_{0,k} \in \mathbb{R}^{n\times p}$, $C^i_{0,k} \in \mathbb{R}^{r\times n}$, and $D^i_{0,k} \in \mathbb{R}^{r\times q}$ are known nominal parameter matrices, whereas $\delta F_k \in \mathbb{R}^{n\times n}$, $\delta G_k \in \mathbb{R}^{n\times m}$, $\delta H_k \in \mathbb{R}^{n\times p}$, $\delta C^i_k \in \mathbb{R}^{r\times n}$, and $\delta D^i_k \in \mathbb{R}^{r\times q}$ are unknown uncertainties bounded to a convex polyhedral domain described by V vertices,

$$\mathbb{V}_k := \left\{ \left(\delta F_k,\, \delta G_k,\, \delta H_k,\, \delta C^i_k,\, \delta D^i_k\right) = \sum_{\nu=1}^{V} \alpha_{\nu,k} \left(F_{\nu,k},\, G_{\nu,k},\, H_{\nu,k},\, C^i_{\nu,k},\, D^i_{\nu,k}\right) \right\}, \tag{4.44}$$

where $F_{\nu,k} \in \mathbb{R}^{n\times n}$, $G_{\nu,k} \in \mathbb{R}^{n\times m}$, $H_{\nu,k} \in \mathbb{R}^{n\times p}$, $C^i_{\nu,k} \in \mathbb{R}^{r\times n}$ and $D^i_{\nu,k} \in \mathbb{R}^{r\times q}$ are known, and $\alpha_k := \begin{bmatrix}\alpha_{1,k} & \cdots & \alpha_{V,k}\end{bmatrix}^T$ belongs to the unit simplex

$$\Lambda_V := \left\{ \alpha \in \mathbb{R}^V \,:\, \sum_{\nu=1}^{V} \alpha_\nu = 1,\, \alpha_\nu \geq 0 \right\}. \tag{4.45}$$

In a stochastic setting, we usually assume that x_0, w_k, and $\{v_k^i\}_{i=1}^S$ are mutually independent zero-mean Gaussian random variables with respective variances

$$\boldsymbol{E}\Big\{x_0 x_0^T\Big\} = P_0 \succ 0, \quad \boldsymbol{E}\Big\{w_k w_l^T\Big\} = Q_k \delta_{kl} \succ 0, \quad \text{and} \quad \boldsymbol{E}\Big\{v_k^i (v_l^j)^T\Big\} = R_k^i \delta_{kl}\delta_{ij} \succ 0,$$

where δ_{ab} is the Kronecker delta function, such that $\delta_{ab} = 1$ if $a = b$, and $\delta_{ab} = 0$ otherwise. Nonetheless, the strategy we develop here does not require that these variables have any particular distribution.

4.3.1.2 Robust Centralized Estimation Problem

Before addressing the robust distributed estimation problem, we first design a centralized estimator for system (4.42)-(4.43). In the centralized problem, there is a central estimator with access to measurements obtained by all the sensors in the network. As the target system state sequence $\{x_k\}$ is not perfectly observed, the goal is to use all the information available up to time instant k, $\boldsymbol{Y}_k = \Big\{\{y_0^i\}_{i=1}^S, \ldots, \{y_k^i\}_{i=1}^S, u_0, \ldots, u_k\Big\}$, to compute a so-called filtered robust central state estimate $\hat{x}_{k|k}^c$ of x_k, as well as a predicted robust central estimate $\hat{x}_{k+1|k}^c$ of x_{k+1}, despite the presence of the polytopic model uncertainties $\delta_k := \Big\{\delta F_k, \delta G_k, \delta H_k, \{\delta C_k^i\}_{i=1}^S, \{\delta D_k^i\}_{i=1}^S\Big\}$. Here, the superscript c indicates the centralized entities.

Following the procedure reported in 4.1.1.2 for the Nominal Centralized Kalman Filter, we adopt a deterministic interpretation of the centralized estimation problem (BRYSON; HO, 1975). Then, to avoid confusion, we introduce the variables $\hat{x}_k$, $\hat{x}_{k+1}$, $\hat{w}_k$, and $\{\hat{v}_k^i\}_{i=1}^S$ as substitutes for the random variables x_k, x_{k+1}, w_k, and $\{v_k^i\}_{i=1}^S$ in the stochastic model (4.42)-(4.43). Based on Sayed (2001) and Ishihara, Terra and Cerri (2015), assuming that at each time step k, an *a priori* state estimate $\hat{x}_{k|k-1}^c$, a set of measurements $\{y_k^i\}_{i=1}^S$, and the input u_k are available, we formulate a min-max constrained optimization problem in which a one-step quadratic objective function should be minimized under the maximum influence of the polytopic parametric uncertainties δ_k, i.e.,

$$\min_{\substack{\hat{x}_k, \hat{x}_{k+1}, \\ \hat{w}_k, \hat{\boldsymbol{v}}_k}} \max_{\delta_k} J_k(\hat{x}_k, \hat{w}_k, \hat{\boldsymbol{v}}_k) = \|\hat{x}_k - \hat{x}_{k|k-1}^c\|^2_{(P_{k|k-1}^c)^{-1}} + \|\hat{w}_k\|^2_{Q_k^{-1}} + \|\hat{\boldsymbol{v}}_k\|^2_{\boldsymbol{\mathcal{R}}_k^{-1}}, \tag{4.46}$$

subject to the set of constraints

$$\begin{bmatrix} I_n \\ \vdots \\ I_n \end{bmatrix} \hat{x}_{k+1} = \begin{bmatrix} F_{0,k} + \alpha_{1,k} V F_{1,k} \\ \vdots \\ F_{0,k} + \alpha_{V,k} V F_{V,k} \end{bmatrix} \hat{x}_k + \begin{bmatrix} G_{0,k} + \alpha_{1,k} V G_{1,k} \\ \vdots \\ G_{0,k} + \alpha_{V,k} V G_{V,k} \end{bmatrix} u_k + \begin{bmatrix} H_{0,k} + \alpha_{1,k} V H_{1,k} \\ \vdots \\ H_{0,k} + \alpha_{V,k} V H_{V,k} \end{bmatrix} \hat{w}_k, \tag{4.47a}$$

$$\begin{bmatrix} I_r \\ \vdots \\ I_r \end{bmatrix} y_k^i = \begin{bmatrix} C_{0,k}^i + \alpha_{1,k} V C_{1,k}^i \\ \vdots \\ C_{0,k}^i + \alpha_{V,k} V C_{V,k}^i \end{bmatrix} \hat{x}_k + \begin{bmatrix} D_{0,k}^i + \alpha_{1,k} V D_{1,k}^i \\ \vdots \\ D_{0,k}^i + \alpha_{V,k} V D_{V,k}^i \end{bmatrix} \hat{v}_k^i, \quad \forall i \in \mathbb{S}, \tag{4.47b}$$

for $k = 0, 1, \ldots, N$, in which we define $\hat{\boldsymbol{v}}_k := \mathbf{col}\left(v_k^1, \ldots, v_k^S\right)$ and $\boldsymbol{\mathcal{R}}_k := \mathbf{diag}\left(R_k^1, \ldots, R_k^S\right)$. In the objective function, $\hat{w}_k$ and $\{\hat{v}_k^i\}_{i=1}^S$ are fitting errors weighted respectively by $Q_k \succ 0$ and $R_k^i \succ 0$, $\forall i \in \mathbb{S}$, and $P_{k|k-1}^c \succ 0$ weights the *a priori* estimation error $x_k - \hat{x}_{k|k-1}^c$. Recall that, from a stochastic viewpoint, matrices Q_k and R_k^i represent the variances of the random variables w_k and $\{v_k^i\}_{i=1}^S$. Nonetheless, in this general deterministic framework, they are rather understood as weighting matrices.

Remark 4.5. *We derive the constraints* (4.47) *of problem* (4.46) *from* (4.42), (4.43), *and* (4.44) *by individually considering each vertex of the polytope. The equivalence between them can be easily shown by summing all the correspondent target system and sensing model equations in* (4.47), *as follows:*

$$\begin{aligned}\hat{x}_{k+1} &= \left(F_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} F_{\nu,k}\right)\hat{x}_k + \left(G_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} G_{\nu,k}\right)u_k + \left(H_{0,k} + \sum_{\nu=1}^{V} \alpha_{\nu,k} H_{\nu,k}\right)\hat{w}_k,\\ y_k^i &= \left(C_{0,k}^i + \sum_{\nu=1}^{V} \alpha_{\nu,k} C_{\nu,k}^i\right)\hat{x}_k + \left(D_{0,k}^i + \sum_{\nu=1}^{V} \alpha_{\nu,k} D_{\nu,k}^i\right)\hat{v}_k^i, \quad \forall i \in \mathbb{S},\end{aligned}$$

which correspond to the same equations in (4.42), (4.43), *and* (4.44), *considering the deterministic variables.*

To simplify the notation, we rewrite the constraints in (4.47) in the following more compact form:

$$\boldsymbol{I}_n \hat{x}_{k+1} = (\boldsymbol{F}_{0,k} + \boldsymbol{\delta F}_k)\hat{x}_k + (\boldsymbol{G}_{0,k} + \boldsymbol{\delta G}_k)u_k + (\boldsymbol{H}_{0,k} + \boldsymbol{\delta H}_k)\hat{w}_k, \tag{4.48a}$$

$$\boldsymbol{I}_r y_k^i = (\boldsymbol{C}_{0,k}^i + \boldsymbol{\delta C}_k^i)\hat{x}_k + (\boldsymbol{D}_{0,k}^i + \boldsymbol{\delta D}_k^i)\hat{v}_k^i, \quad \forall i \in \mathbb{S}, \tag{4.48b}$$

in which we define

$$\begin{aligned}\boldsymbol{I}_n &:= \mathbf{1}_V \otimes I_n, \quad \boldsymbol{F}_{0,k} := \mathbf{1}_V \otimes F_{0,k}, \quad \boldsymbol{G}_{0,k} := \mathbf{1}_V \otimes G_{0,k}, \quad \boldsymbol{H}_{0,k} := \mathbf{1}_V \otimes H_{0,k},\\ \boldsymbol{I}_r &:= \mathbf{1}_V \otimes I_r, \quad \boldsymbol{C}_{0,k}^i := \mathbf{1}_V \otimes C_{0,k}^i, \quad \text{and} \quad \boldsymbol{D}_{0,k}^i := \mathbf{1}_V \otimes D_{0,k}^i,\end{aligned} \tag{4.49}$$

where $\mathbf{1}_V := \begin{bmatrix}1 \cdots 1\end{bmatrix}^T \in \mathbb{R}^V$ and $\otimes$ denotes the Kronecker product. Moreover, the uncertainties in (4.48) are given by

$$\begin{bmatrix}\boldsymbol{\delta F}_k & \boldsymbol{\delta G}_k & \boldsymbol{\delta H}_k\end{bmatrix} = \bar{\boldsymbol{\alpha}}_{1,k} V \begin{bmatrix}\bar{\boldsymbol{F}}_k & \bar{\boldsymbol{G}}_k & \bar{\boldsymbol{H}}_k\end{bmatrix}, \tag{4.50a}$$

$$\begin{bmatrix}\boldsymbol{\delta C}_k^i & \boldsymbol{\delta D}_k^i\end{bmatrix} = \bar{\boldsymbol{\alpha}}_{2,k} V \begin{bmatrix}\bar{\boldsymbol{C}}_k^i & \bar{\boldsymbol{D}}_k^i\end{bmatrix}, \quad \forall i \in \mathbb{S}, \tag{4.50b}$$

where

$$\begin{aligned}\bar{\boldsymbol{\alpha}}_{1,k} &:= \begin{bmatrix}\alpha_{1,k} I_n & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \alpha_{V,k} I_n\end{bmatrix}, \quad \bar{\boldsymbol{F}}_k := \begin{bmatrix}F_{1,k}\\ \vdots\\ F_{V,k}\end{bmatrix}, \quad \bar{\boldsymbol{G}}_k := \begin{bmatrix}G_{1,k}\\ \vdots\\ G_{V,k}\end{bmatrix}, \quad \bar{\boldsymbol{H}}_k := \begin{bmatrix}H_{1,k}\\ \vdots\\ H_{V,k}\end{bmatrix},\\ \bar{\boldsymbol{\alpha}}_{2,k} &:= \begin{bmatrix}\alpha_{1,k} I_r & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \alpha_{V,k} I_r\end{bmatrix}, \quad \bar{\boldsymbol{C}}_k^i := \begin{bmatrix}C_{1,k}^i\\ \vdots\\ C_{V,k}^i\end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{D}}_k^i := \begin{bmatrix}D_{1,k}^i\\ \vdots\\ D_{V,k}^i\end{bmatrix}.\end{aligned} \tag{4.51}$$

We can further aggregate the equations for each sensor i in (4.48b) into a single expression, as follows:

$$\mathcal{Y}_k = (\mathfrak{C}_{0,k} + \boldsymbol{\delta}\mathfrak{C}_k)\hat{x}_k + (\mathfrak{D}_{0,k} + \boldsymbol{\delta}\mathfrak{D}_k)\hat{\boldsymbol{v}}_k, \tag{4.52}$$

in which we define the aggregated vectors and matrices

$$\mathcal{Y}_k := \begin{bmatrix} \boldsymbol{I}_r y_k^1 \\ \vdots \\ \boldsymbol{I}_r y_k^S \end{bmatrix}, \ \hat{\boldsymbol{v}}_k := \begin{bmatrix} \hat{v}_k^1 \\ \vdots \\ \hat{v}_k^S \end{bmatrix}, \ \mathfrak{C}_{0,k} := \begin{bmatrix} \boldsymbol{C}_{0,k}^1 \\ \vdots \\ \boldsymbol{C}_{0,k}^S \end{bmatrix}, \ \text{and } \ \mathfrak{D}_{0,k} := \begin{bmatrix} \boldsymbol{D}_{0,k}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{D}_{0,k}^S \end{bmatrix}, \tag{4.53}$$

considering the definitions in (4.49). Similarly, we aggregate the sensing uncertainty models in (4.50b) into the single expression

$$\begin{bmatrix} \boldsymbol{\delta}\mathfrak{C}_k & \boldsymbol{\delta}\mathfrak{D}_k \end{bmatrix} = (I_S \otimes \bar{\boldsymbol{\alpha}}_{2,k})\, V \begin{bmatrix} \bar{\mathfrak{C}}_k & \bar{\mathfrak{D}}_k \end{bmatrix}, \tag{4.54}$$

with

$$\boldsymbol{\delta}\mathfrak{C}_k := \begin{bmatrix} \boldsymbol{\delta C}_k^1 \\ \vdots \\ \boldsymbol{\delta C}_k^S \end{bmatrix}, \ \boldsymbol{\delta}\mathfrak{D}_k := \begin{bmatrix} \boldsymbol{\delta D}_k^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\delta D}_k^S \end{bmatrix}, \ \bar{\mathfrak{C}}_k := \begin{bmatrix} \bar{\boldsymbol{C}}_k^1 \\ \vdots \\ \bar{\boldsymbol{C}}_k^S \end{bmatrix}, \ \text{and } \ \bar{\mathfrak{D}}_k := \begin{bmatrix} \bar{\boldsymbol{D}}_k^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \bar{\boldsymbol{D}}_k^S \end{bmatrix}, \tag{4.55}$$

according to the definitions in (4.51).

Problem (4.46)-(4.47) has the form of a regularized least-squares estimation problem with polytopic uncertainties, as presented in Section 3.3.2. Therefore, by solving it, we obtain the filtered and predicted robust central state estimates $\hat{x}^c_{k|k}$ and $\hat{x}^c_{k|k+1}$, respectively.

4.3.1.3 Robust Distributed Estimation Problem

In the distributed estimation setting, the goal of each sensor node $i \in \mathbb{S}$ is to obtain the best estimates $\hat{x}^i_{k|k}$ of x_k and $\hat{x}^i_{k+1|k}$ of x_{k+1}, referred to as filtered and predicted robust state estimates, irrespective of the polytopic uncertainties.

Moreover, these estimates should be computed in a distributed rather than centralized fashion. Therefore, each sensor node only has access to its own data and to information provided by its neighbors. We achieve this objective by taking advantage of the distributed characteristic of the average consensus protocol (Algorithm 2.2) to approximate the results of a robust centralized estimator. This strategy is similar to the one we applied in Section 4.1 and Section 4.2, as well as in Kamal, Farrell and Roy-Chowdhury (2013) and Battistelli *et al.* (2015).

4.3.2 Polytopic Robust Centralized Kalman Filter

In this section, we present the Polytopic Robust Centralized Kalman Filter (PRCKF), obtained as part of the solution to problem (4.46)-(4.47). As aforementioned, problem (4.46)-(4.47) is a special case of regularized least-squares problem with polytopic uncertainties (Section 3.3.2). This can be verified by mapping the objective function in (4.46) with (3.95), and the compact aggregated constraints (4.48a)-(4.52) with (3.97), as follows:

$$x \leftarrow \begin{bmatrix} \hat{x}_k^c \\ \hat{x}_{k+1}^c \end{bmatrix}, \quad \bar{x} \leftarrow \begin{bmatrix} \hat{x}_{k|k-1}^c \\ 0 \end{bmatrix}, \quad w \leftarrow \begin{bmatrix} \hat{w}_k \\ \hat{\boldsymbol{v}}_k \end{bmatrix}, \quad \bar{P} \leftarrow \begin{bmatrix} (P_{k|k-1}^c)^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad Q \leftarrow \begin{bmatrix} Q_k^{-1} & 0 \\ 0 & \boldsymbol{\mathcal{R}}_k^{-1} \end{bmatrix},$$
$$\boldsymbol{y}_0 \leftarrow \begin{bmatrix} -\boldsymbol{G}_{0,k} u_k \\ \boldsymbol{\mathcal{Y}}_k \end{bmatrix}, \quad \boldsymbol{A}_0 \leftarrow \begin{bmatrix} \boldsymbol{F}_{0,k} & -\boldsymbol{I}_n \\ \boldsymbol{\mathcal{C}}_{0,k} & 0 \end{bmatrix}, \quad \boldsymbol{B}_0 \leftarrow \begin{bmatrix} \boldsymbol{H}_{0,k} & 0 \\ 0 & \boldsymbol{\mathcal{D}}_{0,k} \end{bmatrix},$$
$$\boldsymbol{\delta y} \leftarrow \begin{bmatrix} -\boldsymbol{\delta G}_k u_k \\ 0 \end{bmatrix}, \quad \boldsymbol{\delta A} \leftarrow \begin{bmatrix} \boldsymbol{\delta F}_k & 0 \\ \boldsymbol{\delta \mathcal{C}}_k & 0 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\delta B} \leftarrow \begin{bmatrix} \boldsymbol{\delta H}_k & 0 \\ 0 & \boldsymbol{\delta \mathcal{D}}_k \end{bmatrix}, \tag{4.56}$$

in which the definitions in (4.49) and (4.53) are taken into consideration. In addition, consider the following mappings between the compact aggregated uncertainty models (4.50a)-(4.54) and (3.99):

$$\bar{\boldsymbol{\alpha}} \leftarrow \begin{bmatrix} \bar{\boldsymbol{\alpha}}_{1,k} & 0 \\ 0 & I_S \otimes \bar{\boldsymbol{\alpha}}_{2,k} \end{bmatrix}, \ \bar{\boldsymbol{y}} \leftarrow \begin{bmatrix} -\bar{\boldsymbol{G}}_k u_k \\ 0 \end{bmatrix}, \ \bar{\boldsymbol{A}} \leftarrow \begin{bmatrix} \bar{\boldsymbol{F}}_k & 0 \\ \bar{\boldsymbol{\mathcal{C}}}_k & 0 \end{bmatrix}, \ \text{and} \ \bar{\boldsymbol{B}} \leftarrow \begin{bmatrix} \bar{\boldsymbol{H}}_k & 0 \\ 0 & \bar{\boldsymbol{\mathcal{C}}}_k \end{bmatrix}, \tag{4.57}$$

where the definitions in (4.51) and (4.55) are considered.

Given that $(P_{k|k-1}^c)^{-1} \succ 0$, one has $\bar{P} \succeq 0$. Also, $Q_k^{-1} \succ 0$ and $\boldsymbol{\mathcal{R}}_k^{-1} \succ 0$ imply that $Q \succ 0$. Hence, we use the results in Lemma 3.5 and in equation (3.124) to obtain the filtered and predicted robust central state estimates, $\hat{x}_{k|k}^c$ and $\hat{x}_{k+1|k}^c$, as well as their respective error weighting matrices $P_{k|k}^c$ and $P_{k+1|k}^c$, as stated in the following theorem.

Theorem 4.8. *Consider the regularized least-squares centralized estimation problem with polytopic uncertainties (4.46)-(4.47) with given initial conditions $\hat{x}_{0|-1}^c$, $P_{0|-1}^c = P_0 \succ 0$, $Q_k \succ 0$, $R_k^i \succ 0$, $\forall i \in \mathbb{S}$, and fixed parameters $\mu > 0$ and $\xi > 0$. For each $k = 0, 1, \ldots, N$, its solution recursively provides the filtered and predicted robust central state estimates of system (4.42)-(4.43), $\hat{x}_{k|k}^c$ and $\hat{x}_{k+1|k}^c$, along with their respective error weighting matrices, $P_{k|k}^c$ and $P_{k+1|k}^c$, according to the procedure described in Algorithm 4.6.*

Proof. Since problem (4.46)-(4.47) is a regularized least-squares estimation problem with polytopic uncertainties, we can use the result in Lemma 3.5 to obtain the robust central state estimates $\hat{x}_{k|k}^c$ and $\hat{x}_{k+1|k}^c$. Thus, we substitute the mappings (4.56) and (4.57) into (3.106) to compute the modified target system and sensing model matrices. Next, we plug the mappings into the solution (3.105). The algebraic procedure is similar to the one described in the proof of Theorem 3.6 and we thus omit it for brevity. The main difference is the presence of the summation terms in step 4 of Algorithm 4.6, which appear due to the aggregate vectors and matrices defined in (4.53) and (4.55), which account for all the

---

Algorithm 4.6 Polytopic Robust Centralized Kalman Filter (PRCKF)

Model: Assume the uncertain system model in (4.42)-(4.43).
Initialization: Set $\hat{x}^c_{0|-1}$, $P^c_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R^i_k \succ 0$, $\forall i \in \mathbb{S}$, $\mu > 0$, and $\xi > 0$.
for $k = 0, 1, \ldots, N$ **do**

1. Obtain measurements y^i_k from all sensors $i \in \mathbb{S}$.
2. Compute φ using the approximation for $\hat{\lambda}$:

$$\varphi = \hat{\lambda} V^2 = (1+\xi)\, \mu V^2$$

3. Compute the modified target system and sensing model matrices, for all $i \in \mathbb{S}$:

$$\Phi_{1,k} = \xi V \varphi^{-1} I_n \qquad \widehat{Q}_k = \Phi_{1,k} + H_{0,k}\Big(Q_k^{-1} + \varphi \bar{\boldsymbol{H}}_k^T \bar{\boldsymbol{H}}_k\Big)^{-1} H_{0,k}^T$$

$$\Phi_{2,k} = \xi V \varphi^{-1} I_r \qquad \widehat{R}^i_k = \Phi_{2,k} + D^i_{0,k}\Big[(R^i_k)^{-1} + \varphi (\bar{\boldsymbol{D}}^i_k)^T \bar{\boldsymbol{D}}^i_k\Big]^{-1} (D^i_{0,k})^T$$

$$\bar{Q}_k = \varphi^{-1} I_{nV} + \bar{\boldsymbol{H}}_k Q_k \bar{\boldsymbol{H}}_k^T \qquad \bar{R}^i_k = \varphi^{-1} I_{rV} + \bar{\boldsymbol{D}}^i_k R^i_k (\bar{\boldsymbol{D}}^i_k)^T$$

$$\widehat{F}_k = F_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k \qquad \widehat{C}^i_k = C^i_{0,k} - D^i_{0,k} R^i_k (\bar{\boldsymbol{D}}^i_k)^T (\bar{R}^i_k)^{-1} \bar{\boldsymbol{C}}^i_k$$

$$\widehat{G}_k = G_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k$$

4. Correction step:
   4.1. Compute the posterior error weighting matrix:

$$P^c_{k|k} = \left[(P^c_{k|k-1})^{-1} + \sum_{i=1}^{S} \Big[(\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} \widehat{C}^i_k + (\bar{\boldsymbol{C}}^i_k)^T (\bar{R}^i_k)^{-1} \bar{\boldsymbol{C}}^i_k\Big] + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k\right]^{-1}$$

   4.2. Compute the filtered robust central state estimate:

$$\hat{x}^c_{k|k} = P^c_{k|k}\left[(P^c_{k|k-1})^{-1} \hat{x}^c_{k|k-1} + \sum_{i=1}^{S} (\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k - \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k u_k\right]$$

5. Prediction step:
   5.1. Update the predicted prior error weighting matrix:

$$P^c_{k+1|k} = \widehat{F}_k P^c_{k|k} \widehat{F}_k^T + \widehat{Q}_k$$

   5.2. Update the predicted prior robust central state estimate:

$$\hat{x}^c_{k+1|k} = \widehat{F}_k \hat{x}^c_{k|k} + \widehat{G}_k u_k$$

end for

---

sensors in the network. Given their block column and diagonal structures, we have that

$$\widehat{\mathfrak{C}}_k^T \widehat{\mathfrak{R}}_k^{-1} \widehat{\mathfrak{C}}_k + \bar{\mathfrak{C}}_k^T \bar{\mathfrak{R}}_k^{-1} \bar{\mathfrak{C}}_k = \sum_{i=1}^{S} \Big[(\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} \widehat{C}^i_k + (\bar{\boldsymbol{C}}^i_k)^T (\bar{R}^i_k)^{-1} \bar{\boldsymbol{C}}^i_k\Big],$$

$$\widehat{\mathfrak{C}}_k^T \widehat{\mathfrak{R}}_k^{-1} \boldsymbol{y}_k = \sum_{i=1}^{S} (\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k,$$

where $\boldsymbol{y}_k := \mathbf{col}\left(y^1_k, \ldots, y^S_k\right)$, $\widehat{\mathfrak{C}}_k := \mathbf{col}\left(\widehat{C}^1_k, \ldots, \widehat{C}^S_k\right)$, $\bar{\mathfrak{C}}_k := \mathbf{col}\left(\bar{\boldsymbol{C}}^1_k, \ldots, \bar{\boldsymbol{C}}^S_k\right)$, $\widehat{\mathfrak{R}}_k :=$

$\mathbf{diag}\left(\widehat{R}_k^1, \ldots, \widehat{R}_k^S\right)$, and $\bar{\mathcal{R}}_k := \mathbf{diag}\left(\bar{R}_k^1, \ldots, \bar{R}_k^S\right)$, with each $\widehat{C}_k^i$, $\widehat{R}_k^i$, and $\bar{R}_k^i$ as defined in step 3 of Algorithm 4.6. Analogously, we use (3.124) to obtain the corresponding estimation error weighting matrices $P_{k|k}^c$ and $P_{k+1|k}^c$, as also shown in the proof of Theorem 3.6. To conclude, note that to compute the $\hat{\lambda}$ parameter, we consider the practical approximation explained in Remark 3.7. □

Remark 4.6. *The proposed Polytopic Robust Centralized Kalman Filter (Algorithm 4.6) depends on the penalty and approximation parameters μ and ξ, respectively. They influence the PRCKF performance in a similar way to what is observed in the PRKF (Algorithm 3.3). Therefore, as discussed in Section 3.3.3, we tune μ based on the level of model uncertainties. The more severe the perturbations, the smaller the value of μ. As for the approximation parameter ξ, it is recommended to select a small value within the $(0, 1)$ interval.*

4.3.3 Polytopic Robust Distributed Kalman Concensus Filter

This section tackles the distributed estimation problem proposed in Section 4.3.1.3. As previously mentioned, in order to solve it, we leverage the average consensus algorithm (Section 2.4) to develop a distributed implementation that can approximate the Polytopic Robust Centralized Kalman Filter presented in Algorithm 4.6, which is considered a benchmark.

In the distributed estimation context, we assume that each sensor $i \in \mathbb{S}$ is initialized with the same prior state estimate $\hat{x}_{0|-1}^i$ and prior error weighting matrix $P_{0|-1}^i \succ 0$. We adopt the hybrid consensus on measurements and information (HCMCI) approach (BATTISTELLI *et al.*, 2015), such that each sensor $i \in \mathbb{S}$ exchanges information with its neighbors $j \in \mathcal{N}_i$ to ultimately obtain an approximation of the filtered and predicted prior robust central state estimates in a distributed fashion.

We propose the so-called Polytopic Robust Distributed Kalman Consensus Filter (PRDKCF) shown in Algorithm 4.7. In accordance with the HCMCI protocol, in steps 4 and 5, we simultaneously perform the average consensus algorithm (Algorithm 2.2) to each sensor's prior information and innovation pairs, denoted $\left(\Omega_k^i,\, \omega_k^i\right)$ and $\left(\delta\Omega_k^i,\, \delta\omega_k^i\right)$, respectively. Moreover, in step 5.3, the consensus weights π_{ij} should satisfy the conditions outlined in Definition 2.2, such that the consensus states of each node i are updated with a convex combination of the corresponding states within its inclusive neighborhood. Here, we adopt the Metropolis weights (XIAO; BOYD; LALL, 2005), shown in (2.20), for which these conditions hold. Then, based on step 4 of the PRCKF (Algorithm 4.6), we use the outcome of the consensus step in the correction stage shown in step 6 of Algorithm 4.7. Analogous to the nominal distributed case (Algorithm 4.2), we include the corrective scalar ρ_k^i to compensate for the possible underweighting of the innovation pair $\left(\delta\Omega_k^i,\, \delta\omega_k^i\right)$ due to the average consensus process. Concluding the algorithm, in step 7, we perform the prediction step, which is based on step 5 of the PRCKF (Algorithm 4.6).

Algorithm 4.7 Polytopic Robust Distributed Kalman Consensus Filter (PRDKCF) (each sensor i)

Model: Assume the uncertain system model in (4.42)-(4.43).

Initialization: Set $\hat{x}^i_{0|-1}$, $P^i_{0|-1} = P_0 \succ 0$, $Q_k \succ 0$, $R^i_k \succ 0$, $\mu > 0$, $\xi > 0$, and $L \geq 1$.

for $k = 0, 1, \ldots, N$ **do**

1. Obtain a measurement y^i_k.
2. Compute φ using the approximation for $\hat{\lambda}$: $\varphi = \hat{\lambda} V^2 = (1+\xi)\,\mu V^2$
3. Compute the modified system and sensing model matrices:

$$\Phi_{1,k} = \xi V \varphi^{-1} I_n \qquad \widehat{Q}_k = \Phi_{1,k} + H_{0,k}\Big(Q_k^{-1} + \varphi \bar{\boldsymbol{H}}_k^T \bar{\boldsymbol{H}}_k\Big)^{-1} H_{0,k}^T$$

$$\Phi_{2,k} = \xi V \varphi^{-1} I_r \qquad \widehat{R}^i_k = \Phi_{2,k} + D^i_{0,k}\Big[(R^i_k)^{-1} + \varphi(\bar{\boldsymbol{D}}^i_k)^T \bar{\boldsymbol{D}}^i_k\Big]^{-1} (D^i_{0,k})^T$$

$$\bar{Q}_k = \varphi^{-1} I_{nV} + \bar{\boldsymbol{H}}_k Q_k \bar{\boldsymbol{H}}_k^T \qquad \bar{R}^i_k = \varphi^{-1} I_{rV} + \bar{\boldsymbol{D}}^i_k R^i_k (\bar{\boldsymbol{D}}^i_k)^T$$

$$\widehat{F}_k = F_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k \qquad \widehat{C}^i_k = C^i_{0,k} - D^i_{0,k} R^i_k (\bar{\boldsymbol{D}}^i_k)^T (\bar{R}^i_k)^{-1} \bar{\boldsymbol{C}}^i_k$$

$$\widehat{G}_k = G_{0,k} - H_{0,k} Q_k \bar{\boldsymbol{H}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k$$

4. Initialize the consensus states:

$$\Omega^i_k(0) = (P^i_{k|k-1})^{-1} \qquad \delta\Omega^i_k(0) = (\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} \widehat{C}^i_k + (\bar{\boldsymbol{C}}^i_k)^T (\bar{R}^i_k)^{-1} \bar{\boldsymbol{C}}^i_k$$

$$\omega^i_k(0) = (P^i_{k|k-1})^{-1} \hat{x}^i_{k|k-1} \qquad \delta\omega^i_k(0) = (\widehat{C}^i_k)^T (\widehat{R}^i_k)^{-1} y^i_k$$

5. Consensus step:

for $\ell = 0, 1, \ldots, L-1$ **do**

5.1. Send $\Big\{\Omega^i_k(\ell),\ \omega^i_k(\ell),\ \delta\Omega^i_k(\ell),\ \delta\omega^i_k(\ell)\Big\}$ to all neighbors $j \in \mathcal{N}_i$.

5.2. Receive $\Big\{\Omega^j_k(\ell),\ \omega^j_k(\ell),\ \delta\Omega^j_k(\ell),\ \delta\omega^j_k(\ell)\Big\}$ from all neighbors $j \in \mathcal{N}_i$.

5.3. Update the consensus states:

$$\Omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \Omega^j_k(\ell) \qquad \delta\Omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \delta\Omega^j_k(\ell)$$

$$\omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \omega^j_k(\ell) \qquad \delta\omega^i_k(\ell+1) = \sum_{j=1}^{S} \pi_{ij}\, \delta\omega^j_k(\ell)$$

end for

6. Correction step:

6.1. Compute the posterior error weighting matrix:

$$P^i_{k|k} = \Big[\Omega^i_k(L) + \rho^i_k\, \delta\Omega^i_k(L) + \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{F}}_k\Big]^{-1}$$

6.2. Compute the filtered state estimate:

$$\hat{x}^i_{k|k} = P^i_{k|k}\Big[\omega^i_k(L) + \rho^i_k\, \delta\omega^i_k(L) - \bar{\boldsymbol{F}}_k^T \bar{Q}_k^{-1} \bar{\boldsymbol{G}}_k u_k\Big]$$

7. Prediction step:

7.1. Update the predicted prior error weighting matrix:

$$P^i_{k+1|k} = \widehat{F}_k P^i_{k|k} \widehat{F}_k^T + \widehat{Q}_k$$

7.2. Update the predicted prior state estimate:

$$\hat{x}^i_{k+1|k} = \widehat{F}_k \hat{x}^i_{k|k} + \widehat{G}_k u_k$$

end for

Theorem 4.9. *Consider the Polytopic Robust Distributed Kalman Consensus Filter in Algorithm 4.7 and that Assumption 4.1 is satisfied. In addition, assume that the consensus weights π_{ij} are chosen according to Definition 2.2, the number of consensus iterations $L \to \infty$ in step 5, and that $\rho_k^i = S$ in step 6. Then, the filtered and predicted prior robust state estimates, $\hat{x}_{k|k}^i$ and $\hat{x}_{k+1|k}^i$, and their respective error weighting matrices, $P_{k|k}^i$ and $P_{k+1|k}^i$, obtained by each sensor $i \in \mathbb{S}$ converge to the corresponding robust central state estimates $\hat{x}_{k|k}^c$ and $\hat{x}_{k+1|k}^c$, and error weighting matrices $P_{k|k}^c$ and $P_{k+1|k}^c$ obtained using the Polytopic Robust Centralized Kalman Filter in Algorithm 4.6.*

Proof. Since the undirected graph $\mathbb{G}$ describing the sensor network is connected and the consensus weights π_{ij} are properly selected, we have that the associated weighted adjacency matrix Π exhibit the properties listed in Lemma 2.9. Moreover, as the number of consensus iterations $L \to \infty$, Theorem 2.2 guarantees the convergence of the average consensus algorithm. Let us then prove through induction that the PRDKCF in Algorithm 4.7 converges to the PRCKF in Algorithm 4.6.

At time step $k = 0$, consider that the PRCKF is initialized with $\hat{x}_{0|-1}^c = \hat{x}_0$ and $P_{0|-1}^c = P_0 \succ 0$, whereas all sensors $i \in \mathbb{S}$ initialize the PRDKCF with $\hat{x}_{0|-1}^i = \hat{x}_0$ and $P_{0|-1}^i = P_0 \succ 0$. Thus, according to Theorem 2.2, after performing the consensus step of the PRDKCF, the information and innovation pairs of all the sensors converge as follows:

$$
\begin{aligned}
\Omega_0^i(L) &\to \frac{1}{S}\sum_{j=1}^{S} P_0^{-1} = P_0^{-1}, & \delta\Omega_0^i(L) &\to \frac{1}{S}\sum_{j=1}^{S}\Big[(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}\widehat{C}_0^j + (\bar{\boldsymbol{C}}_0^j)^T(\bar{R}_0^j)^{-1}\bar{\boldsymbol{C}}_0^j\Big],\\
\omega_0^i(L) &\to \frac{1}{S}\sum_{j=1}^{S} P_0^{-1}\hat{x}_0 = P_0^{-1}\hat{x}_0, & \delta\omega_0^i(L) &\to \frac{1}{S}\sum_{j=1}^{S}(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}y_0^j.
\end{aligned}
$$

We then substitute these consensus outcomes into the equations in step 6 of Algorithm 4.7, considering that $\rho_0^i = S$, such that

$$
\begin{aligned}
P_{0|0}^i \to &\left[P_0^{-1} + S\frac{1}{S}\sum_{j=1}^{S}\Big[(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}\widehat{C}_0^j + (\bar{\boldsymbol{C}}_0^j)^T(\bar{R}_0^j)^{-1}\bar{\boldsymbol{C}}_0^j\Big] + \bar{\boldsymbol{F}}_0^T\bar{Q}_0^{-1}\bar{\boldsymbol{F}}_0\right]^{-1} =\\
&\left[P_0^{-1} + \sum_{j=1}^{S}\Big[(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}\widehat{C}_0^j + (\bar{\boldsymbol{C}}_0^j)^T(\bar{R}_0^j)^{-1}\bar{\boldsymbol{C}}_0^j\Big] + \bar{\boldsymbol{F}}_0^T\bar{Q}_0^{-1}\bar{\boldsymbol{F}}_0\right]^{-1} = P_{0|0}^c,\\
\hat{x}_{0|0}^i \to P_{0|0}^i &\left[P_0^{-1}\hat{x}_0 + S\frac{1}{S}\sum_{j=1}^{S}(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}y_0^j - \bar{\boldsymbol{F}}_0^T\bar{R}_0^{-1}\bar{\boldsymbol{G}}_0u_0\right] =\\
&P_{0|0}^c\left[P_0^{-1}\hat{x}_0 + \sum_{j=1}^{S}(\widehat{C}_0^j)^T(\widehat{R}_0^j)^{-1}y_0^j - \bar{\boldsymbol{F}}_0^T\bar{Q}_0^{-1}\bar{\boldsymbol{G}}_0u_0\right] = \hat{x}_{0|0}^c,
\end{aligned}
$$

for all sensors $i \in \mathbb{S}$. Note how the choice of scalar weight $\rho_0^i = S$ is important to correctly compensate for the $1/S$ factor that appears in the outcome of the innovation pair $\left(\delta\Omega_0^i,\, \delta\omega_0^i\right)$ after the averaging process. Furthermore, the convergence above implies that, in step 7, $P^i_{1|0} \to P^c_{1|0}$ and $\hat{x}^i_{1|0} \to \hat{x}^c_{1|0}$. Therefore, for $k = 0$, we proved that the PRDKCF converges to the PRCKF.

Now, let us assume that at time step $k-1$, we have $P^i_{k-1|k-1} \to P^c_{k-1|k-1}$, $\hat{x}^i_{k-1|k-1} \to \hat{x}^c_{k-1|k-1}$, $P^i_{k|k-1} \to P^c_{k|k-1}$, and $\hat{x}^i_{k|k-1} \to \hat{x}^c_{k|k-1}$, $\forall i \in \mathbb{S}$. Then, according to Theorem 2.2, at time step k, we achieve the following consensus outcome after step 5 of the PRDKCF:

$$\begin{aligned}
\Omega^i_k(L) &\to \frac{1}{S}\sum_{j=1}^{S}(P^c_{k|k-1})^{-1} = (P^c_{k|k-1})^{-1},\\
\delta\Omega^i_k(L) &\to \frac{1}{S}\sum_{j=1}^{S}\Big[(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}\widehat{C}^j_k + (\bar{\boldsymbol{C}}^j_k)^T(\bar{R}^j_k)^{-1}\bar{\boldsymbol{C}}^j_k\Big],\\
\omega^i_k(L) &\to \frac{1}{S}\sum_{j=1}^{S}(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} = (P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1},\\
\delta\omega^i_k(L) &\to \frac{1}{S}\sum_{j=1}^{S}(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}y^j_k.
\end{aligned}$$

Substituting these outcomes into the equations in step 6 of the PRDKCF, assuming that $\rho^i_k = S$, then yields

$$\begin{aligned}
P^i_{k|k} \to &\left[(P^c_{k|k-1})^{-1} + S\frac{1}{S}\sum_{j=1}^{S}\Big[(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}\widehat{C}^j_k + (\bar{\boldsymbol{C}}^j_k)^T(\bar{R}^j_k)^{-1}\bar{\boldsymbol{C}}^j_k\Big] + \bar{\boldsymbol{F}}^T_k\bar{Q}^{-1}_k\bar{\boldsymbol{F}}_k\right]^{-1} =\\
&\left[(P^c_{k|k-1})^{-1} + \sum_{j=1}^{S}\Big[(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}\widehat{C}^j_k + (\bar{\boldsymbol{C}}^j_k)^T(\bar{R}^j_k)^{-1}\bar{\boldsymbol{C}}^j_k\Big] + \bar{\boldsymbol{F}}^T_k\bar{Q}^{-1}_k\bar{\boldsymbol{F}}_k\right]^{-1} = P^c_{k|k},\\
\hat{x}^i_{k|k} \to P^i_{k|k}&\left[(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} + S\frac{1}{S}\sum_{j=1}^{S}(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}y^j_k - \bar{\boldsymbol{F}}^T_k\bar{Q}^{-1}_k\bar{\boldsymbol{G}}_k u_k\right] =\\
&P^c_{k|k}\left[(P^c_{k|k-1})^{-1}\hat{x}^c_{k|k-1} + \sum_{j=1}^{S}(\widehat{C}^j_k)^T(\widehat{R}^j_k)^{-1}y^j_k - \bar{\boldsymbol{F}}^T_k\bar{Q}^{-1}_k\bar{\boldsymbol{G}}_k u_k\right] = \hat{x}^c_{k|k},
\end{aligned}$$

for all sensors $i \in \mathbb{S}$. Then, plugging the results above into the equations in step 7 of the PRDKCF leads to $P^i_{k+1|k} \to P^c_{k+1|k}$ and $\hat{x}^i_{k+1|k} \to \hat{x}^c_{k+1|k}$. Therefore, under the established conditions, by induction, we have that, for $k = 0, 1, \ldots, N$, the PRDKCF in Algorithm 4.7 converges to the PRCKF in Algorithm 4.6. □

Since the proposed PRDKCF is derived from the PRCKF (Algorithm 4.6), it also depends on the μ and ξ parameters. As such, Remark 4.6 provides guidelines on how to select their values. In conclusion, we emphasize how the combination of the HCMCI protocol with the average consensus algorithm enabled the derivation of a polytopic robust distributed estimator that approaches the performance of its centralized counterpart. However, it is also important to note that this convergence is theoretical, since it requires an infinite number of consensus iterations L, which is not possible in practice. Nonetheless, an illustrative example will show that, for a sufficiently large and finite L, the distributed filter closely follows the performance of the corresponding centralized filter.

Remark 4.7. *As in the Nominal Distributed Kalman Consensus Filter presented in Algorithm 4.2, we include a corrective scalar weight ρ_k^i in step 6 of Algorithm 4.7. Its purpose is to avoid the possible underweighting of the innovation pair $\big(\delta\Omega_k^i(L),\, \delta\omega_k^i(L)\big)$ due to scaling from the average consensus procedure. As a consequence, Algorithm 4.7 actually represents a family of polytopic robust distributed filters, depending on the choice of ρ_k^i. As discussed in Remark 4.2, ideally, one should have $\rho_k^i = S$ to reach the centralized performance. Since the value of S may not be available to the sensors in the network, Remark 4.2 also provides a procedure to estimate it in a distributed fashion.*

4.3.4 Stability Analysis

This section addresses the stability properties and boundedness of the estimation error variance of the proposed polytopic robust centralized and distributed filters. We study the steady-state behavior of Algorithm 4.6 and Algorithm 4.7, considering that the target system and sensing model parameters are time-invariant and there is no input u_k. Nevertheless, we still assume that the polytope coefficients α_k are time-varying. Thus, consider the following discrete-time uncertain linear system:

$$x_{k+1} = (F_0 + \delta F_k)x_k + (H_0 + \delta H_k)w_k, \tag{4.58a}$$

$$y_k^i = (C_0^i + \delta C_k^i)x_k + (D_0^i + \delta D_k^i)v_k^i, \quad \forall i \in \mathbb{S}, \tag{4.58b}$$

for $k \geq 0$, with time-varying parametric uncertainties bounded by the convex polyhedron

$$\mathbb{V}_k = \left\{ \Big(\delta F_k,\, \delta H_k,\, \delta C_k^i,\, \delta D_k^i\Big) = \sum_{\nu=1}^{V} \alpha_{\nu,k} \Big(F_\nu,\, H_\nu,\, C_\nu^i,\, D_\nu^i\Big) \right\}, \tag{4.59}$$

where $\alpha_k = \Big[\alpha_{1,k} \cdots \alpha_{V,k}\Big]^T$ belongs to the unit simplex Λ_V in (4.45), with V vertices.

Following the strategy carried out in Section 4.1.4, Section 4.2.4, as well as in Kamal, Farrell and Roy-Chowdhury (2013), we perform the analysis under the assumptions described in Theorem 4.9, i.e., considering that the PRDKCF converges to the PRCKF. This way, the stability properties of the polytopic robust centralized filter can be extended to its distributed implementation.

We start by establishing the stability conditions of the PRCKF in Algorithm 4.6. Considering the time-invariant uncertain system model (4.58)-(4.59), the PRCKF equations in steps 4 and 5 of Algorithm 3.3 become:

$$P_{k|k}^c = \Big[(P_{k|k-1}^c)^{-1} + \widehat{\boldsymbol{\mathfrak{C}}}^T \widehat{\boldsymbol{\mathfrak{R}}}^{-1} \widehat{\boldsymbol{\mathfrak{C}}} + \bar{\boldsymbol{\mathfrak{C}}}^T \bar{\boldsymbol{\mathfrak{R}}}^{-1} \bar{\boldsymbol{\mathfrak{C}}} + \bar{\boldsymbol{F}}^T \bar{Q}^{-1} \bar{\boldsymbol{F}}\Big]^{-1}, \tag{4.60}$$

$$\hat{x}_{k|k}^c = P_{k|k}^c \Big[(P_{k|k-1}^c)^{-1} \hat{x}_{k|k-1}^c + \widehat{\boldsymbol{\mathfrak{C}}}^T \widehat{\boldsymbol{\mathfrak{R}}}^{-1} \boldsymbol{y}_k\Big], \tag{4.61}$$

$$P_{k+1|k}^c = \widehat{F} P_{k|k}^c \widehat{F}^T + \widehat{Q}, \tag{4.62}$$

$$\hat{x}_{k+1|k}^c = \widehat{F} \hat{x}_{k|k}^c, \tag{4.63}$$

where

$$\boldsymbol{y}_k = \begin{bmatrix} y_k^1 \\ \vdots \\ y_k^S \end{bmatrix},\ \widehat{\mathfrak{C}} = \begin{bmatrix} \widehat{C}^1 \\ \vdots \\ \widehat{C}^S \end{bmatrix},\ \bar{\mathfrak{C}} = \begin{bmatrix} \bar{\boldsymbol{C}}^1 \\ \vdots \\ \bar{\boldsymbol{C}}^S \end{bmatrix},\ \widehat{\mathfrak{R}} = \begin{bmatrix} \widehat{R}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widehat{R}^S \end{bmatrix},\ \text{and}\ \bar{\mathfrak{R}} = \begin{bmatrix} \bar{R}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \bar{R}^S \end{bmatrix},$$

with each $\widehat{C}^i$, $\widehat{R}^i$, and $\bar{R}^i$, $\forall i \in \mathbb{S}$, as well as $\widehat{F}$, $\widehat{Q}$, and $\bar{Q}$ given by the corresponding equations in step 3 of Algorithm 4.6, assuming constant parameters. For a simpler notation, we also define the augmented matrices

$$\widetilde{\mathfrak{C}} := \begin{bmatrix} \widehat{\mathfrak{C}} \\ \bar{\mathfrak{C}} \\ \bar{\boldsymbol{F}} \end{bmatrix} \quad \text{and} \quad \widetilde{\mathfrak{R}} := \begin{bmatrix} \widehat{\mathfrak{R}} & 0 & 0 \\ 0 & \bar{\mathfrak{R}} & 0 \\ 0 & 0 & \bar{Q} \end{bmatrix}.$$

Then, we can rewrite (4.60) more compactly, as follows:

$$P^c_{k|k} = \left[(P^c_{k|k-1})^{-1} + \widetilde{\mathfrak{C}}^T \widetilde{\mathfrak{R}}^{-1} \widetilde{\mathfrak{C}}\right]^{-1},$$

which we expand by applying Lemma A.1, such that

$$P^c_{k|k} = P^c_{k|k-1} - P^c_{k|k-1}\widetilde{\mathfrak{C}}^T \left(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c_{k|k-1}\widetilde{\mathfrak{C}}^T\right)^{-1}\widetilde{\mathfrak{C}} P^c_{k|k-1}. \tag{4.64}$$

We then combine (4.64) with (4.61) and substitute the result into (4.63) to obtain the steady-state predicted robust central state estimate

$$\hat{x}^c_{k+1|k} = \widetilde{\mathfrak{F}}_k \hat{x}^c_{k|k-1} + \widetilde{\mathfrak{F}}_k P^c_{k|k-1} \widehat{\mathfrak{C}}^T \widehat{\mathfrak{R}}^{-1} \boldsymbol{y}_k, \tag{4.65}$$

in which

$$\widetilde{\mathfrak{F}}_k = \widehat{F}\Big(I_n - P^c_{k|k-1}\widetilde{\mathfrak{C}}^T \left(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c_{k|k-1}\widetilde{\mathfrak{C}}^T\right)^{-1}\widetilde{\mathfrak{C}}\Big)$$

is the polytopic robust centralized filter closed-loop matrix. Furthermore, we substitute $P^c_{k|k}$ from (4.64) back into (4.62) to obtain the following expression for the predicted prior error weighting matrix:

$$P^c_{k+1|k} = \widehat{F}\Big(P^c_{k|k-1} - P^c_{k|k-1}\widetilde{\mathfrak{C}}^T \left(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c_{k|k-1}\widetilde{\mathfrak{C}}^T\right)^{-1}\widetilde{\mathfrak{C}} P^c_{k|k-1}\Big)\widehat{F}^T + \widehat{Q}. \tag{4.66}$$

Theorem 4.10. *Consider the linear system model* (4.58) *with polytopic uncertainties* (4.59) *and the corresponding robust centralized filter* (4.65)-(4.66)*. Assume that* $\{\widehat{F}, \widetilde{\mathfrak{C}}\}$ *is* detectable *and* $\{\widehat{F}, \widehat{Q}^{1/2}\}$ *is* controllable. *Then, for any initial condition* $P^c_{0|-1} \succ 0$, $\xi > 0$, *and* $\mu > 0$, $P^c_{k+1|k}$ *converges to the* unique stabilizing solution $P^c \succ 0$ *of the algebraic Riccati equation*

$$P^c = \widehat{F}\Big(P^c - P^c\widetilde{\mathfrak{C}}^T \left(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c\widetilde{\mathfrak{C}}^T\right)^{-1}\widetilde{\mathfrak{C}} P^c\Big)\widehat{F}^T + \widehat{Q}. \tag{4.67}$$

The solution P^c *is stabilizing in the sense that the steady-state filter closed-loop matrix*

$$\widetilde{\mathfrak{F}} = \widehat{F}\Big(I_n - P^c\widetilde{\mathfrak{C}}^T \left(\widetilde{\mathfrak{R}} + \widetilde{\mathfrak{C}} P^c\widetilde{\mathfrak{C}}^T\right)^{-1}\widetilde{\mathfrak{C}}\Big) \tag{4.68}$$

is Schur stable.

Proof. The conditions $\xi > 0$ and $\mu > 0$ imply that $\varphi > 0$, ensuring that matrices $\widehat{F}$, $\widetilde{\mathfrak{C}}$, $\widetilde{\mathfrak{R}}$, and $\widehat{Q}$ are well-defined. From Kailath, Sayed and Hassibi (2000b), we have that detectability of $\{\widehat{F}, \widetilde{\mathfrak{C}}\}$ and controllability of $\{\widehat{F}, \widehat{Q}^{1/2}\}$ ensure the convergence of $P^c_{k+1|k}$ in (4.66) to the unique stabilizing positive definite solution P^c of the algebraic Riccati equation (4.67) that stabilizes (4.68), which is the polytopic robust centralized filter steady-state closed-loop matrix. □

Now, let us establish the conditions for the boundedness of the estimation error variance of the proposed PRCKF. Thus, consider the uncertain linear discrete-time system model (4.58)-(4.59). Note that we can write the polytopic uncertainties described in (4.59) alternatively as

$$\begin{bmatrix}\delta F_k & \delta H_k\end{bmatrix} = \begin{bmatrix}I_n & \cdots & I_n\end{bmatrix}\begin{bmatrix}\alpha_{1,k}I_n & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \alpha_{V,k}I_n\end{bmatrix}\begin{bmatrix}F_1 & H_1\\ \vdots & \vdots\\ F_V & H_V\end{bmatrix} =: M_1\,\bar{\boldsymbol{\alpha}}_{1,k}\begin{bmatrix}\bar{\boldsymbol{F}} & \bar{\boldsymbol{H}}\end{bmatrix}, \tag{4.69a}$$

$$\begin{bmatrix}\delta C^i_k & \delta D^i_k\end{bmatrix} = \begin{bmatrix}I_r & \cdots & I_r\end{bmatrix}\begin{bmatrix}\alpha_{1,k}I_r & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \alpha_{V,k}I_r\end{bmatrix}\begin{bmatrix}C^i_1 & D^i_1\\ \vdots & \vdots\\ C^i_V & D^i_V\end{bmatrix} =: M_2\,\bar{\boldsymbol{\alpha}}_{2,k}\begin{bmatrix}\bar{\boldsymbol{C}}^i & \bar{\boldsymbol{D}}^i\end{bmatrix},\ \forall i \in \mathbb{S}, \tag{4.69b}$$

in which, since $\alpha_k = \begin{bmatrix}\alpha_{1,k} \cdots \alpha_{V,k}\end{bmatrix}^T$ belongs to the unit simplex Λ_V in (3.85), we have that $\|\bar{\boldsymbol{\alpha}}_{1,k}\| \leq 1$ and $\|\bar{\boldsymbol{\alpha}}_{2,k}\| \leq 1$.

We can further aggregate the equations for each sensor i in (4.58b) and (4.69b) into the compact expressions

$$\begin{aligned}\boldsymbol{y}_k &= (\mathfrak{C}_0 + \boldsymbol{\delta}\mathfrak{C}_k)x_k + (\mathfrak{D}_0 + \boldsymbol{\delta}\mathfrak{D}_k)\boldsymbol{v}_k,\\ \begin{bmatrix}\boldsymbol{\delta}\mathfrak{C}_k & \boldsymbol{\delta}\mathfrak{D}_k\end{bmatrix} &= \mathfrak{M}_2\left(I_S \otimes \bar{\boldsymbol{\alpha}}_{2,k}\right)\begin{bmatrix}\bar{\mathfrak{C}} & \bar{\mathfrak{D}}\end{bmatrix},\end{aligned} \tag{4.70}$$

where

$$\begin{aligned}&\boldsymbol{y}_k := \begin{bmatrix}y^1_k\\ \vdots\\ y^S_k\end{bmatrix},\quad \boldsymbol{v}_k := \begin{bmatrix}v^1_k\\ \vdots\\ v^S_k\end{bmatrix},\quad \mathfrak{C}_0 := \begin{bmatrix}C^1_0\\ \vdots\\ C^S_0\end{bmatrix},\quad \mathfrak{D}_0 := \begin{bmatrix}D^1_0 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & D^S_0\end{bmatrix},\\ &\mathfrak{M}_2 := \begin{bmatrix}M_2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & M_2\end{bmatrix},\quad \bar{\mathfrak{C}} := \begin{bmatrix}\bar{\boldsymbol{C}}^1\\ \vdots\\ \bar{\boldsymbol{C}}^S\end{bmatrix},\quad \text{and}\quad \bar{\mathfrak{D}} := \begin{bmatrix}\bar{\boldsymbol{D}}^1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \bar{\boldsymbol{D}}^S\end{bmatrix}.\end{aligned} \tag{4.71}$$

Moreover, we assume that w_k and $\{v_k\}_{i=1}^S$ are uncorrelated zero-mean Gaussian noise processes with joint covariance matrix

$$\mathfrak{Q} = \boldsymbol{E}\left\{\begin{bmatrix}w_k\\ \boldsymbol{v}_k\end{bmatrix}\begin{bmatrix}w_k^T & \boldsymbol{v}_k^T\end{bmatrix}\right\} = \begin{bmatrix}Q & 0\\ 0 & \mathfrak{R}\end{bmatrix} \succ 0, \tag{4.72}$$

in which $\boldsymbol{\mathfrak{R}} = \mathbf{diag}\left(R^1, \ldots, R^S\right)$. In addition, assume that there is no correlation between the parametric uncertainties and the system and measurement noises. Finally, consider the following assumptions about the uncertain system and the robust centralized filter.

Assumption 4.6. *The uncertain system* (4.58a)-(4.69b) *is quadratically stable, according to Definition 3.2.*

Assumption 4.7. *The conditions of Theorem 4.10 are satisfied, such that the polytopic robust centralized filter steady-state closed-loop matrix* $\widetilde{\boldsymbol{\mathfrak{F}}}$ *is Schur stable.*

Under Assumption 3.3 and Assumption 3.4, we now show that the steady-state robust centralized filter (4.65) is also quadratically stable. To simplify the notation, we define the following steady-state filter gain

$$\widetilde{\boldsymbol{\mathfrak{K}}} := \widetilde{\boldsymbol{\mathfrak{F}}} P^c \widehat{\boldsymbol{\mathfrak{C}}}^T \widehat{\boldsymbol{\mathfrak{R}}}^{-1},$$

where $\widetilde{\boldsymbol{\mathfrak{F}}}$ is given by (4.68), in which P^c is the stabilizing solution of the algebraic Riccati equation (4.67). Hence, the steady-state polytopic robust centralized filter equation can be rewritten as

$$\hat{x}^c_{k+1|k} = \widetilde{\boldsymbol{\mathfrak{F}}} \hat{x}^c_{k|k-1} + \widetilde{\boldsymbol{\mathfrak{K}}} \boldsymbol{y}_k. \tag{4.73}$$

Now, substituting $\boldsymbol{y}_k$ from (4.70) into (4.73) gives

$$\hat{x}^c_{k+1|k} = \widetilde{\boldsymbol{\mathfrak{F}}} \hat{x}^c_{k|k-1} + \widetilde{\boldsymbol{\mathfrak{K}}}(\boldsymbol{\mathfrak{C}}_0 + \boldsymbol{\delta\mathfrak{C}}_k) x_k + \widetilde{\boldsymbol{\mathfrak{K}}}(\boldsymbol{\mathfrak{D}}_0 + \boldsymbol{\delta\mathfrak{D}}_k) \boldsymbol{v}_k, \tag{4.74}$$

with aggregate matrices as defined in (4.71). Additionally, we define the central state estimation error vector $e^c_k := x_k - \hat{x}^c_{k|k-1}$. Then, subtracting (4.74) from x_{k+1} in (4.58a) yields

$$e^c_{k+1} = \left[(F_0 - \widetilde{\boldsymbol{\mathfrak{F}}} - \widetilde{\boldsymbol{\mathfrak{K}}}\boldsymbol{\mathfrak{C}}_0) + (\delta F_k - \widetilde{\boldsymbol{\mathfrak{K}}}\boldsymbol{\delta\mathfrak{C}}_k)\right] x_k + \widetilde{\boldsymbol{\mathfrak{F}}} e^c_k + (H_0 + \delta H_k) w_k - \widetilde{\boldsymbol{\mathfrak{K}}}(\boldsymbol{\mathfrak{D}}_0 + \boldsymbol{\delta\mathfrak{D}}_k)\boldsymbol{v}_k. \tag{4.75}$$

Furthermore, we introduce the augmented system composed of the target system state x_k and the central estimation error e^c_k. Then, from (4.58a), (4.69a), (4.70), and (4.75), this augmented system is described by

$$\begin{gathered} \zeta^c_{k+1} = (\mathscr{F} + \boldsymbol{\delta}\mathscr{F}_k)\zeta^c_k + (\mathscr{H} + \boldsymbol{\delta}\mathscr{H}_k)\eta^c_k, \\ \left[\boldsymbol{\delta}\mathscr{F}_k \;\; \boldsymbol{\delta}\mathscr{H}_k\right] = \mathscr{M}\, \bar{\boldsymbol{\alpha}}_k \left[\bar{\mathscr{F}} \;\; \bar{\mathscr{H}}\right], \end{gathered} \tag{4.76}$$

where

$$\zeta^c_k := \begin{bmatrix} x_k \\ e^c_k \end{bmatrix}, \quad \eta^c_k := \begin{bmatrix} w_k \\ \boldsymbol{v}_k \end{bmatrix}, \quad \mathscr{F} := \begin{bmatrix} F_0 & 0 \\ F_0 - \widetilde{\boldsymbol{\mathfrak{F}}} - \widetilde{\boldsymbol{\mathfrak{K}}}\boldsymbol{\mathfrak{C}} & \widetilde{\boldsymbol{\mathfrak{F}}} \end{bmatrix}, \quad \mathscr{H} := \begin{bmatrix} H_0 & 0 \\ H_0 & -\widetilde{\boldsymbol{\mathfrak{K}}}\boldsymbol{\mathfrak{D}}_0 \end{bmatrix},$$

$$\mathscr{M} := \begin{bmatrix} M_1 & 0 \\ M_1 & -\widetilde{\boldsymbol{\mathfrak{K}}}\boldsymbol{\mathfrak{M}}_2 \end{bmatrix}, \quad \bar{\boldsymbol{\alpha}}_k := \begin{bmatrix} \bar{\boldsymbol{\alpha}}_{1,k} & 0 \\ 0 & I_S \otimes \bar{\boldsymbol{\alpha}}_{2,k} \end{bmatrix}, \quad \bar{\mathscr{F}} := \begin{bmatrix} \bar{\boldsymbol{F}} & 0 \\ \bar{\boldsymbol{\mathfrak{C}}} & 0 \end{bmatrix}, \quad \bar{\mathscr{H}} := \begin{bmatrix} \bar{\boldsymbol{H}} & 0 \\ 0 & \bar{\boldsymbol{\mathfrak{D}}} \end{bmatrix}.$$

Lemma 4.2. *If Assumption 4.6 and Assumption 4.7 are satisfied, then the augmented system (4.76) is quadratically stable.*

Proof. Note that the augmented system matrix $\mathscr{F}$ is lower triangular with diagonal elements F_0 and $\boldsymbol{\mathfrak{F}}$, which are both Schur stable, which implies that $\mathscr{F}$ is also Schur stable. In addition, we have that

$$\bar{\mathscr{F}}(zI_{2n}-\mathscr{F})^{-1}\mathscr{M} = \begin{bmatrix} \bar{\boldsymbol{F}} & 0 \\ \bar{\boldsymbol{\mathfrak{C}}} & 0 \end{bmatrix} \begin{bmatrix} zI_n - F_0 & 0 \\ -(F_0 - \tilde{\boldsymbol{\mathfrak{F}}} - \tilde{\boldsymbol{\mathfrak{K}}}\boldsymbol{\mathfrak{C}}_0) & zI_n - \tilde{\boldsymbol{\mathfrak{F}}} \end{bmatrix}^{-1} \begin{bmatrix} M_1 & 0 \\ M_1 & -\tilde{\boldsymbol{\mathfrak{K}}}\boldsymbol{\mathfrak{M}}_2 \end{bmatrix}$$
$$= \begin{bmatrix} \bar{\boldsymbol{F}}(zI_n - F_0)^{-1}M_1 & 0 \\ \bar{\boldsymbol{\mathfrak{C}}}(zI_n - F_0)^{-1}M_1 & 0 \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{F}} \\ \bar{\boldsymbol{\mathfrak{C}}} \end{bmatrix} (zI_n - F_0)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix}.$$

Moreover,

$$F_0 + M_1 \boldsymbol{\alpha}_{1,k} \bar{\boldsymbol{F}} = F_0 + \begin{bmatrix} M_1 & 0 \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\alpha}}_{1,k} & 0 \\ 0 & I_S \otimes \bar{\boldsymbol{\alpha}}_{2,k} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{F}} \\ \bar{\boldsymbol{\mathfrak{C}}} \end{bmatrix}.$$

Since system (4.58a)-(4.69b) is quadratically stable, according to Remark 3.10, one has

$$\left\| \begin{bmatrix} \bar{\boldsymbol{F}} \\ \bar{\boldsymbol{\mathfrak{C}}} \end{bmatrix} (zI_n - F_0)^{-1} \begin{bmatrix} M_1 & 0 \end{bmatrix} \right\|_\infty < 1,$$

for all admissible contractions for all admissible contractions $\bar{\boldsymbol{\alpha}}_{1,k}$ and $\bar{\boldsymbol{\alpha}}_{2,k}$. As a consequence, $\left\|\bar{\mathscr{F}}(zI_{2n}-\mathscr{F})^{-1}\mathscr{M}\right\|_\infty < 1$ and the augmented system (4.76) is also quadratically stable. □

Next, we define the covariance matrix of the augmented system state as $\mathscr{P}_k^c := \boldsymbol{E}\left\{\zeta_k^c(\zeta_k^c)^T\right\}$. Then, from (4.76), $\mathscr{P}_k^c$ satisfies the Lyapunov recursion

$$\mathscr{P}_{k+1}^c = (\mathscr{F} + \boldsymbol{\delta}\mathscr{F}_k)\mathscr{P}_k^c(\mathscr{F} + \boldsymbol{\delta}\mathscr{F}_k)^T + (\mathscr{H} + \boldsymbol{\delta}\mathscr{H}_k)\mathscr{Q}(\mathscr{H} + \boldsymbol{\delta}\mathscr{H}_k)^T, \tag{4.77}$$

with $\mathscr{Q}$ as defined in (4.72).

Theorem 4.11. *Given that Assumption 4.6 and Assumption 4.7 hold, the state estimation error variance of the steady-state polytopic robust centralized filter (4.73) satisfies*

$$\lim_{k\to\infty} \boldsymbol{E}\left\{e_k e_k^T\right\} \preceq \mathscr{V}_{22},$$

where $\mathscr{V}_{22}$ is the $(2,2)$ block entry with the smallest trace among all $(2,2)$ block entries of matrices $\mathscr{V} \succ 0$ that satisfy the inequality

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{F}})\mathscr{V}(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{F}})^T + (\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{H}})\mathscr{Q}(\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}\bar{\mathscr{H}})^T - \mathscr{V} \preceq 0,$$

for all admissible $\bar{\boldsymbol{\alpha}}$, with $\|\bar{\boldsymbol{\alpha}}\| \leq 1$.

Proof. According to Lemma 4.2, the augmented system (4.76) is quadratically stable, then Definition 3.2 implies that there exists a matrix $\mathscr{U} \succ 0$ such that

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})\,\mathscr{U}\,(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T - \mathscr{U} \prec 0,$$

for all admissible $\bar{\boldsymbol{\alpha}}_k$. Based on Petersen and McFarlane (1996) and Sayed (2001), the existence of such a matrix $\mathscr{U} \succ 0$ indicates that there exists a sufficiently large scaling parameter $\epsilon > 0$, such that one can find a matrix $\mathscr{V} = \epsilon\,\mathscr{U}$ satisfying

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})\,\mathscr{V}\,(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T + (\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{H}})\,\mathscr{Q}\,(\mathscr{H} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{H}})^T \preceq \mathscr{V}.$$

By subtracting the recursion for the augmented system covariance in (4.77) from the inequality above, we obtain

$$(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})(\mathscr{V} - \mathscr{P}_k^c)(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T \preceq \mathscr{V} - \mathscr{P}_{k+1}^c,$$

or, equivalently,

$$\mathscr{V} - \mathscr{P}_{k+1}^c = (\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})(\mathscr{V} - \mathscr{P}_k^c)(\mathscr{F} + \mathscr{M}\bar{\boldsymbol{\alpha}}_k\bar{\mathscr{F}})^T + \mathscr{W}_k,$$

for some $\mathscr{W}_k \succeq 0$. Finally, since the augmented system is quadratically stable, as $k \to \infty$, we have that $\mathscr{V} - \mathscr{P}_{k+1}^c \succeq 0$, or $\mathscr{P}_{k+1}^c \preceq \mathscr{V}$. The $(2, 2)$ block entry of $\mathscr{P}_k^c$ corresponds to the estimation error variance, which is therefore bounded. □

Corollary 4.3. *If the assumptions in Theorem 4.9 are satisfied, as well as Assumptions 4.6 and 4.7, the PRDKCF in Algorithm 4.7 converges to the PRCKF in Algorithm 4.6 and thus shares its stability and bounded estimation error variance properties, according to Theorems 4.10 and 4.11.*

4.3.5 Illustrative Example

In this section, we study the performance of the proposed Polytopic Robust Distributed Kalman Consensus Filter with an example adapted from Xie, Soh and Souza (1994) and Section 3.3.5. We also evaluate the centralized counterpart, PRCKF, considered the benchmark for the distributed strategy. We further compare our results with those of other polytopic robust distributed filtering approaches from the literature. Additionally, to establish a baseline, we also assess the results of the nominal centralized and distributed filters presented in Section 4.1.

Consider a linear discrete-time target-system with polytopic uncertainties, as described in (4.42)-(4.44), with the following constant nominal parameter matrices and uncertainties bounded to a 2-vertex polytope:

$$F_{0,k} = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix},\ G_{0,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ H_{0,k} = \begin{bmatrix} -6 \\ 1 \end{bmatrix},\ F_{1,k} = \begin{bmatrix} 0 & 0 \\ 0.1 & 0.3 \end{bmatrix},\ G_{1,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ H_{1,k} = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix},$$
$$\big(F_{2,k},\, G_{2,k},\, H_{2,k}\big) = -\big(F_{1,k},\, G_{1,k},\, H_{1,k}\big).$$

No input signal u_k is present and w_k is a zero-mean white Gaussian noise signal with variance $Q_k = 1$. The initial state is $x_0 = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$.

A set of $S = 25$ sensors arranged in a random geometric undirected network, shown in Figure 8, measure the target system. The sensing model is described as in (4.43)-(4.44), with v_k^i as zero-mean white Gaussian noise signals with variances R_k^i. Two different types of sensors are considered. Sensors with odd number, i.e., $i = 1, 3, \ldots, 25$, are of the first type, with constant parameter matrices

$$C_{0,k}^i = \begin{bmatrix} -100 & 9 \end{bmatrix}, \quad D_{0,k}^i = 1, \quad C_{1,k}^i = \begin{bmatrix} 0.1 & 0.3 \end{bmatrix}, \quad D_{1,k}^i = 0.1, \quad R_k^i = 1,$$
$$\left(C_{2,k}^i,\, D_{2,k}^i\right) = -\left(C_{1,k}^i,\, D_{1,k}^i\right).$$

Sensors with even number, i.e., $i = 2, 4, \ldots, 24$, are of the second type, with matrices

$$C_{0,k}^i = \begin{bmatrix} -50 & 12 \end{bmatrix}, \quad D_{0,k}^i = 1, \quad C_{1,k}^i = \begin{bmatrix} 0.15 & 0.45 \end{bmatrix}, \quad D_{1,k}^i = 0.3, \quad R_k^i = 0.8,$$
$$\left(C_{2,k}^i,\, D_{2,k}^i\right) = -\left(C_{1,k}^i,\, D_{1,k}^i\right).$$

Then, we apply the proposed PRDKCF (Algorithm 4.7) with the following initialization data for all sensors:

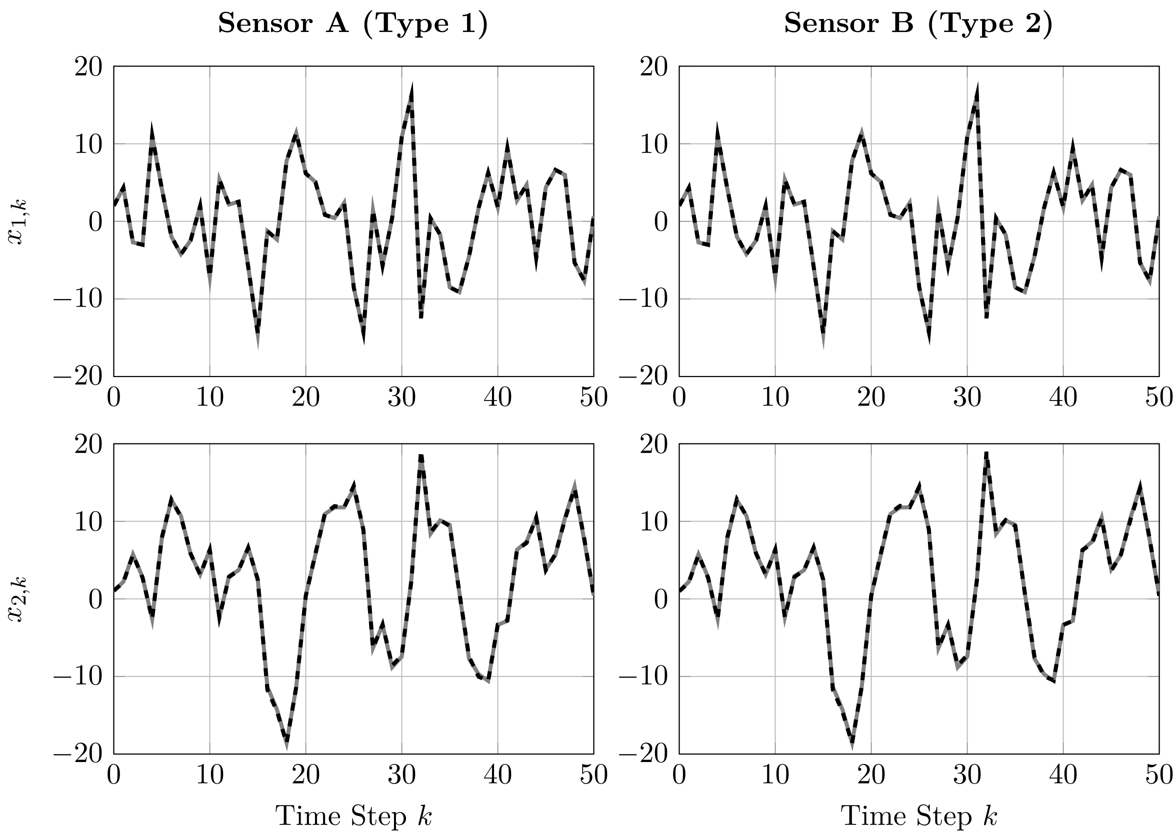
$$\hat{x}_{0|-1}^i = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \quad P_{0|-1}^i = I_2, \quad \mu = 0.01, \quad \xi = 0.01, \text{ and } L = 10,$$

with parameters μ and ξ selected according to the guidelines in Remark 4.6. For the consensus iterations, we adopt the Metropolis weights shown in (2.20). Figure 12 depicts the evolution of the actual target system state along with the estimation performed by sensors A (Type 1) and B (Type 2), identified in Figure 8, using the proposed PRDKCF. At each time step, the coefficients $\alpha_k \in \Lambda_2$ (see (4.45)) are randomly selected. According to the results, both sensors were able to successfully track the state of the target system, despite the polytopic model uncertainties. Moreover, their estimates are similar, indicating that they reach consensus at each time step.

We continue our analysis of the PRDKCF with some comparisons. Analogous to Section 4.2.5, we consider two versions of the PRDKCF. In the first, RDKCF-1, the number of sensors in the network S is known to each sensor, such that, in step 6 of Algorithm 4.7, we choose $\rho_k^i = S$. In the the second version, PRDKCF-2, S is estimated according to Remark 4.2. We compare the distributed filter results with those from the PRCKF (Algorithm 4.6), taken as a benchmark. In addition, we also simulate the nominal centralized and distributed filters, CKF (Algorithm 4.1) and DKCF (Algorithm 4.2), respectively. Furthermore, we compare the PRDKCF with other polytopic robust distributed estimators from the literature, namely the $\mathcal{H}_\infty$-consensus filter by Shen, Wang and Hung (2010) and the mean square state estimator of Souza, Coutinho and Kinnaert (2016).

The simulation consists of performing $M = 1000$ Monte Carlo experiments, each with time horizon $N = 100$. At each time step k, we compute the mean squared estimation

Figure 12 – Actual (solid lines) and estimated (dashed lines) target system state obtained by sensors A and B with the proposed PRDKCF (Algorithm 4.7).



error (MSE), averaged over all experiments and sensors in the network, as follows:

$$\mathrm{MSE}_k = \frac{1}{SM}\sum_{i=1}^{S}\sum_{e=1}^{M}\|x_k - \hat{x}^i_{k|k,e}\|^2,$$

which provides a reasonable approximation of the estimation error variance, as we cannot compute it analytically due to the parametric uncertainties, as discussed in Sayed (2001).

The results are presented in Figure 13 and Table 4, which summarizes the mean $\overline{\mathrm{MSE}}$ and standard deviation $\sigma(\mathrm{MSE})$ of the estimation error variances, respectively computed as

$$\overline{\mathrm{MSE}} = \sum_{k=0}^{N}\frac{\mathrm{MSE}_k}{N+1} \quad \text{and} \quad \sigma^2(\mathrm{MSE}) = \sum_{k=0}^{N}\frac{(\mathrm{MSE}_k - \overline{\mathrm{MSE}})^2}{N+1}.$$

As expected, since the PRCKF gathers information from all the sensors in the network, it achieves the best performance. Nevertheless, both versions of the proposed PRDKCF present a very similar performance, exhibiting the smallest error variance among the distributed approaches. When S is known (PRDKCF-1), we achieve a slightly smaller error variance compared to when we estimate it (PRDKCF-2), which was also anticipated. These filters also show the smallest standard deviation. The nominal centralized and distributed estimators, CKF (Algorithm 4.1) and DKCF (Algorithm 4.2), obtained similar

Figure 13 – Estimation error variance curves of the polytopic robust distributed filters.

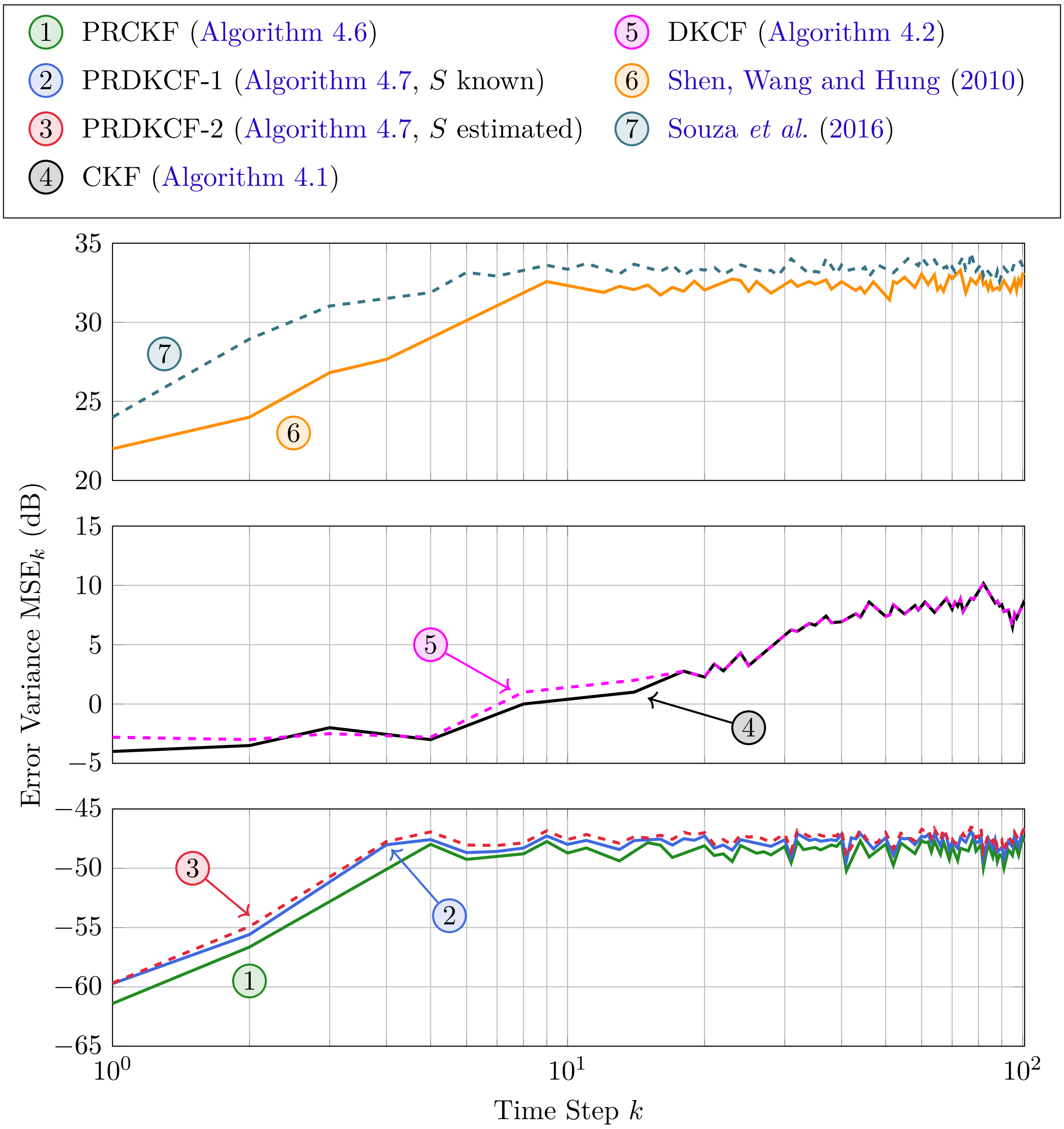


Table 4 – Estimation performance of each polytopic robust distributed filter.

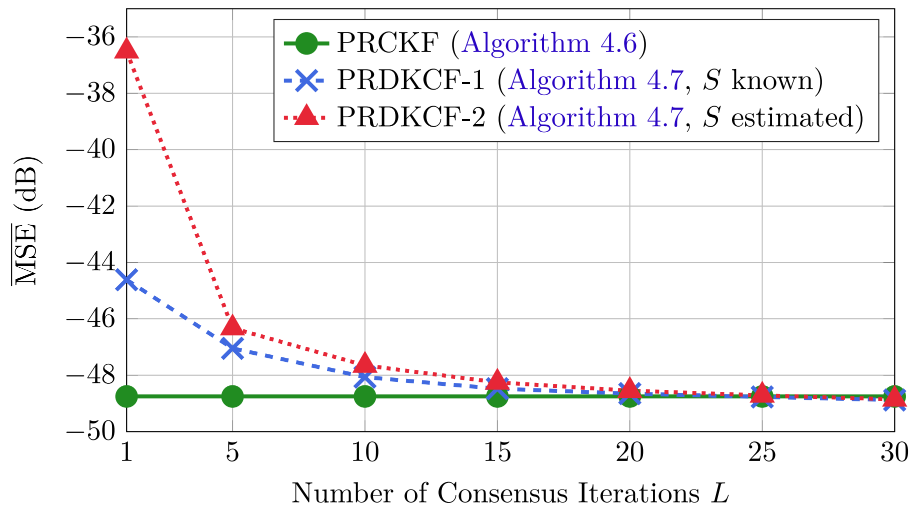
Filter	$\overline{\text{MSE}}$ (dB)	σ(MSE) (dB)
(1) PRCKF (Algorithm 4.6)	**-48.75**	1.6582
(2) PRDKCF-1 (Algorithm 4.7, S known)	-48.07	**1.5331**
(3) PRDKCF-2 (Algorithm 4.7, S estimated)	-47.65	1.5398
(4) CKF (Algorithm 4.1)	3.809	11.007
(5) DKCF (Algorithm 4.2)	3.836	10.902
(6) Shen, Wang and Hung (2010)	31.96	2.2239
(7) Souza, Coutinho and Kinnaert (2016)	33.14	2.0296

Bold numbers indicate the smallest values.

results, with significant larger error variance and standard deviation, compared to the previous estimators. This is explained by their lack of uncertainty compensation. The $\mathcal{H}_\infty$-consensus filter by Shen, Wang and Hung (2010) and the mean square state estimator of Souza, Coutinho and Kinnaert (2016) present the highest error variances, despite considering the polytopic model uncertainties. Moreover, both depend on the solution of LMI-based optimization problems and are not fully distributed, since the gains of all sensors are computed in a batch, which may be infeasible for larger networks.

We conclude our analysis with an evaluation of how the number of consensus iterations L affects the PRDKCF performance. Figure 14 compiles a series of simulations with several values of L, considering both scenarios of the PRDKCF. For each value of L, we compute the mean estimation error $\overline{\text{MSE}}$ over the entire time horizon, as previously described. The PRCKF is also shown for comparison purposes. The results show that, as we increase the value of L, the distributed filters approach the result of the centralized filter. This, however, requires more computation time, such that we have a performance trade-off. Moreover, note that except for when $L = 1$, both versions of the PRDKCF exhibit similar results, meaning that the impact of knowing S beforehand or estimating it online is not very significant.

Figure 14 – Effect of the number of consensus iterations L on the PRDKCF (Algorithm 4.7).



Chapter 5

Conclusion

In this thesis, we addressed several linear discrete-time state estimation problems under different conditions. Since these problems fundamentally consist of estimating the state of a dynamical system based on measurements obtained from some sensing device, we divided them into two main categories: single- and multiple-sensor state estimation. Furthermore, within each of these categories, we considered three situations. First, we assumed that the available target system and sensing models were perfectly known. In the other two situations, we dealt with the more realistic scenario where these models are subject to parametric uncertainties, which we considered to be of the norm-bounded or polytopic kind. For each category and scenario, we proposed filtering strategies inspired by the simplicity and efficiency of the Kalman filter (KALMAN, 1960).

We developed a core framework for the nominal single-sensor state estimation case. We adopt a deterministic interpretation of the estimation task and formulate it as a constrained regularized least-squares problem, as discussed in Bryson and Ho (1975). The constraints are the equations that define the target system and sensing models. Rather than solving the constrained problem, we used the penalty function method (LUENBERGER; YE, 2021) to transform it into a more convenient unconstrained equivalent, whose solution provided the so-called Nominal Kalman Filter (KF). As such, the proposed estimator inherits the recursive and analytical nature of the standard Kalman filter, which we presented as a simple correction-prediction algorithm.

Then, based on the works by Sayed (2001) and Ishihara, Terra and Cerri (2015), we extended this framework to deal with the cases where the underlying target system and sensing models are subject to norm-bounded or polytopic parametric uncertainties. We formulate these robust estimation problems as constrained regularized least-squares problems with uncertainties and apply the penalty function method, which conveniently provides a parameter we can adjust to improve the estimation accuracy. Using this methodology, we proposed the Robust Kalman Filter (RKF) and the Polytopic Robust Kalman Filter (PRKF), presenting both as recursive correction-prediction algorithms that resemble the nominal version. We further established conditions for stability and bounded estimation error variance of each filter. Finally, the performance of the proposed filters was accessed with numerical examples, which illustrated their advantage compared to other existing approaches found in the specialized literature. In summary, the proposed estimators have a simple structure. They do not depend on numerical solvers to deal with complex LMI-based optimization problems, such that they require a relatively low computational cost and reasonable estimation quality. Therefore, they are promising solutions for real-time applications, even with low-cost hardware setups.

The multiple-sensor state estimation solutions are the main contributions of this work. We addressed each of the three scenarios mentioned above, considering a sensor network context. As such, we first extended the single-sensor framework commented above to encompass the centralized estimation problem, i.e., assuming that there is a fusion center with access to measurements from all of the sensors in the network. As a result, we proposed the so-called Nominal Centralized Kalman Filter (CKF), the Robust Centralized Kalman Filter (RCKF), and the Polytopic Robust Kalman Filter (PRCKF), each extending the capabilities of their single-sensor versions.

Our main objective, however, was to solve the distributed variant of the problem, meaning that there is no central estimator, and the sensors work as independent units. Nevertheless, each sensor can communicate with a limited set of neighbors and exchange information to improve their estimation accuracy. Based on the strategy presented by Battistelli *et al.* (2015), we combined the average consensus algorithm (REN; BEARD; ATKINS, 2007) with the hybrid consensus on measurements and information (HCMCI) protocol to derive fully distributed versions of the centralized estimators. Then, we further proposed the Nominal Distributed Kalman Consensus Filter (DKCF), the Robust Distributed Kalman Consensus Filter (RDKCF), and the Polytopic Robust Distributed Kalman Consensus Filter (PRDKCF). Moreover, we showed that these distributed filters converge to their centralized counterparts for a sufficiently large number of consensus iterations. Since the proposed centralized and distributed estimators derive from the single-sensor solutions, they also inherit their recursive and relatively simple structures and are presented as correction-prediction algorithms. Furthermore, we also established the necessary conditions for stability and bounded estimation error variance of each filter. Finally, we evaluated the performance of the proposed centralized and distributed filters with illustrative examples. The results showed that the proposed strategies outperformed other approaches present in the relatively scarce related literature, which usually rely on the solution of LMI-based optimization problems that become increasingly complex as the network gets larger and require offline computation of the filter gains. In contrast, the proposed filters feature a good performance versus computational burden trade-off due to their simpler recursive and analytical structure, being suitable for online systems.

We emphasize how the development of the aforementioned robust distributed state estimators was carried out as an effort to reduce the gap in the literature on distributed filtering for systems with norm-bounded and polytopic uncertainties over sensor networks. Nonetheless, we further suggest some directions in which this work could be extended:

- Application of the proposed robust distributed filters to real-world systems.
- Extension of the proposed distributed estimators to deal with directed networks and time-varying communication topologies.
- Addressing the robust distributed estimation problem for nonlinear systems with norm-bounded and polytopic uncertainties.
- Taking network-induced effects such as time-delays and packet dropouts into consideration.

References

ABOLHASANI, M.; RAHMANI, M. Robust Kalman filtering for discrete-time time-varying systems with stochastic and norm-bounded uncertainties. **Journal of Dynamic Systems, Measurement, and Control**, v. 140, n. 3, 2018.

ALEMDAR, H.; ERSOY, C. Wireless sensor networks for healthcare: A survey. **Computer Networks**, v. 54, n. 15, p. 2688–2710, 2010.

ANDERSON, B. D. O.; MOORE, J. B. **Optimal Filtering**. Englewood Cliffs: Prentice Hall, 1979.

ARORA, A. *et al.* A line in the sand: A wireless sensor network for target detection, classification, and tracking. **Computer Networks**, v. 46, n. 5, p. 605–634, 2004.

BAI, X. *et al.* Collaborative fusion estimation over wireless sensor networks for monitoring CO2 concentration in a greenhouse. **Information Fusion**, v. 42, p. 119–126, 2018.

BATTISTELLI, G.; CHISCI, L. Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. **Automatica**, v. 50, n. 3, p. 707–718, 2014.

BATTISTELLI, G. *et al.* Consensus-based linear and nonlinear filtering. **IEEE Transactions on Automatic Control**, v. 60, n. 5, p. 1410–1415, 2015.

BEARD, R. W.; LAWTON, J.; HADAEGH, F. Y. A coordination architecture for spacecraft formation control. **IEEE Transactions on Control Systems Technology**, v. 9, n. 6, p. 777–790, 2001.

BERTSEKAS, D. P.; TSITSIKLIS, J. **Parallel and Distributed Computation: Numerical Methods**. Upper Saddle River: Prentice-Hall, 1989.

BIDRAM, A. *et al.* Secondary control of microgrids based on distributed cooperative control of multi-agent systems. **IET Generation, Transmission Distribution**, v. 7, n. 8, p. 822–831, 2013.

BRYSON, A. E.; HO, Y. Optimal filtering and prediction. *In*: ______. **Applied Optimal Control: Optimization, Estimation, and Control**. New York: Taylor & Francis, 1975. chap. 12, p. 348–389.

CHANG, X.-H.; PARK, J. H.; TANG, Z. New approach to $\mathcal{H}_\infty$ filtering for discrete-time systems with polytopic uncertainties. **Signal Processing**, v. 113, p. 147–158, 2015.

CHEN, F.; REN, W. On the control of multi-agent systems: A survey. **Foundations and Trends in Systems and Control**, v. 6, n. 4, p. 339–499, 2019.

DEGROOT, M. H. Reaching a consensus. **Journal of the American Statistical Association**, v. 69, n. 345, p. 118–121, 1974.

DESHMUKH, R.; KWON, C.; HWANG, I. Optimal discrete-time Kalman consensus filter. *In*: **Proceedings of the 2017 American Control Conference (ACC), 2017, Seattle (USA).** Piscataway: IEEE, 2017. p. 5801–5806.

DIESTEL, R. **Graph Theory**. 5th. ed. Berlin: Springer, 2017.

DING, D. *et al.* A survey on model-based distributed control and filtering for industrial cyber-physical systems. **IEEE Transactions on Industrial Informatics**, v. 15, n. 5, p. 2483–2499, 2019.

DING, D. *et al.* Distributed state estimation with stochastic parameters and nonlinearities through sensor networks: The finite-horizon case. **Automatica**, v. 48, n. 8, p. 1575–1585, 2012.

DING, D. *et al.* Distributed recursive filtering for stochastic systems under uniform quantizations and deception attacks through sensor networks. **Automatica**, v. 78, p. 231–240, 2017.

DONG, H.; DING, S. X.; REN, W. Distributed filtering with randomly occurring uncertainties over sensor networks: the channel fading case. **International Journal of General Systems**, v. 43, n. 3–4, p. 254–266, 2014.

DONG, X. *et al.* Time-varying formation control for unmanned aerial vehicles with switching interaction topologies. **Control Engineering Practice**, v. 46, p. 26–36, 2016.

DONG, Z.; YOU, Z. Finite-horizon robust Kalman filtering for uncertain discrete time-varying systems with uncertain-covariance white noises. **IEEE Signal Processing Letters**, v. 13, n. 8, p. 493–496, 2006.

DUAN, P. *et al.* Distributed state estimation for uncertain linear systems: A regularized least-squares approach. **Automatica**, v. 117, p. 109007, 2020.

DUAN, Z. *et al.* Robust $\mathcal{H}_2$ and $\mathcal{H}_\infty$ filtering for uncertain linear systems. **Automatica**, v. 42, n. 11, p. 1919–1926, 2006.

FENG, J.; WANG, Z.; ZENG, M. Distributed weighted robust Kalman filter fusion for uncertain systems with autocorrelated and cross-correlated noises. **Information Fusion**, v. 14, n. 1, p. 78–86, 2013.

GAO, H. *et al.* New approach to mixed $\mathcal{H}_2/\mathcal{H}_\infty$ filtering for polytopic discrete-time systems. **IEEE Transactions on Signal Processing**, v. 53, n. 8, p. 3183–3192, 2005.

GARIN, F.; SCHENATO, L. A survey on distributed estimation and control applications using linear consensus algorithms. *In*: **Networked Control Systems**. London: Springer, 2010. p. 75–107.

GEROMEL, J. C. *et al.* $\mathcal{H}_2$ and $\mathcal{H}_\infty$ robust filtering for discrete-time linear systems. **SIAM Journal on Control and Optimization**, v. 38, n. 5, p. 1353–1368, 2000.

GEROMEL, J. C.; OLIVEIRA, M. C. de; BERNUSSOU, J. Robust filtering of discrete-time linear systems with parameter dependent Lyapunov functions. **SIAM Journal on Control and Optimization**, v. 41, n. 3, p. 700–711, 2002.

GERSHON, E.; SHAKED, U. Vertex-dependent approach to robust $\mathcal{H}_\infty$ control and estimation of stochastic discrete-time systems. **IFAC-PapersOnLine**, v. 48, n. 11, p. 949–953, 2015.

GERSHON, E.; SHAKED, U. Vertex-dependent filtering of retarded uncertain state-multiplicative noisy systems. **International Journal of Robust and Nonlinear Control**, v. 30, n. 16, p. 6978–6993, 2020.

GODSIL, C.; ROYLE, G. **Algebraic Graph Theory**. New York: Springer, 2001.

GROCHOLSKY, B. *et al.* Cooperative air and ground surveillance. **IEEE Robotics Automation Magazine**, v. 13, n. 3, p. 16–25, 2006.

HAN, F. *et al.* Scalable consensus filtering for uncertain systems over sensor networks with Round-Robin protocol. **International Journal of Robust and Nonlinear Control**, v. 31, n. 3, p. 1051–1066, 2021.

HE, S. *et al.* Distributed estimation over a low-cost sensor network: A review of state-of-the-art. **Information Fusion**, v. 54, p. 21–43, 2020.

HEDAYATI, M.; RAHMANI, M. Robust distributed $\mathcal{H}_\infty$ filtering over an uncertain sensor network with multiple fading measurements and varying sensor delays. **International Journal of Robust and Nonlinear Control**, v. 30, n. 2, p. 538–566, 2020.

HORN, R. A.; JOHNSON, C. R. **Matrix Analysis**. 2nd. ed. New York: Cambridge University Press, 2013.

ISHIHARA, J. Y.; TERRA, M. H.; CERRI, J. P. Optimal robust filtering for systems subject to uncertainties. **Automatica**, v. 52, p. 111–117, 2015.

JADBABAIE, A.; LIN, J.; MORSE, A. S. Coordination of groups of mobile autonomous agents using nearest neighbor rules. **IEEE Transactions on Automatic Control**, v. 48, n. 6, p. 988–1001, 2003.

KAILATH, T.; SAYED, A. H.; HASSIBI, B. Deterministic least-squares problems. *In*: _____. **Linear Estimation**. Upper Saddle River: Prentice Hall, 2000a. chap. 2, p. 40–77.

KAILATH, T.; SAYED, A. H.; HASSIBI, B. Algebraic Riccati equations. *In*: _____. **Linear Estimation**. Upper Saddle River: Prentice Hall, 2000b. chap. E, p. 773–806.

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, v. 82, n. 1, p. 35–45, 1960.

KAMAL, A. T.; FARRELL, J. A.; ROY-CHOWDHURY, A. K. Information weighted consensus filters and their application in distributed camera networks. **IEEE Transactions on Automatic Control**, v. 58, n. 12, p. 3112–3125, 2013.

LEE, J.; PARK, B. Development and evaluation of a cooperative vehicle intersection control algorithm under the connected vehicles environment. **IEEE Transactions on Intelligent Transportation Systems**, v. 13, n. 1, p. 81–90, 2012.

LEVY, B. C.; ZORZI, M. A contraction analysis of the convergence of risk-sensitive filters. **SIAM Journal on Control and Optimization**, v. 54, n. 4, p. 2154–2173, 2016.

LI, W.; SHEN, W. Swarm behavior control of mobile multi-robots with wireless sensor networks. **Journal of Network and Computer Applications**, v. 34, n. 4, p. 1398–1407, 2011.

LIU, H.; ZHOU, T. Robust state estimation for uncertain linear systems with random parametric uncertainties. **Science China Information Sciences**, v. 60, n. 1, p. 1–13, 2017.

LÖFBERG, J. YALMIP : A toolbox for modeling and optimization in MATLAB. *In*: **Proceedings of the 2004 IEEE International Symposium on Computer Aided Control Systems Design (CACSD), 2004, Taipei (Taiwan)**. Piscataway: IEEE, 2004. p. 284–289.

LU, N. *et al.* Connected vehicles: Solutions and challenges. **IEEE Internet of Things Journal**, v. 1, n. 4, p. 289–299, 2014.

LUENBERGER, D. G.; YE, Y. Penalty and barrier methods. *In*: ______. **Linear and Nonlinear Programming**. 5th. ed. Cham: Springer International Publishing, 2021. chap. 13, p. 455–486.

LYNCH, N. A. **Distributed Algorithms**. San Francisco: Morgan Kaufmann, 1996.

MAINWARING, A. *et al.* Wireless sensor networks for habitat monitoring. *In*: **Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, 2002, Atlanta (USA).** New York: ACM, 2002. p. 88–97.

MODALAVALASA, S. *et al.* A review of robust distributed estimation strategies over wireless sensor networks. **Signal Processing**, v. 188, 2021.

MORAIS, C. F. *et al.* $\mathcal{H}_\infty$ filtering and control of discrete-time polytopic systems with state multiplicative noise. *In*: **Proceedings of the 56th IEEE Conference on Decision and Control (CDC), 2017, Melbourne (Australia).** Piscataway: IEEE, 2017.

OLFATI-SABER, R. Distributed Kalman filter with embedded consensus filters. *In*: **Proceedings of the 44th IEEE Conference on Decision and Control (CDC), 2005, Seville (Spain).** Piscataway: IEEE, 2005. p. 8179–8184.

OLFATI-SABER, R. Distributed Kalman filtering for sensor networks. *In*: **Proceedings of the 46th IEEE Conference on Decision and Control (CDC), 2007, New Orleans (USA).** Piscataway: IEEE, 2007. p. 5492–5498.

OLFATI-SABER, R. Kalman-consensus filter : Optimality, stability, and performance. *In*: **Proceedings of the 48th IEEE Conference on Decision and Control (CDC) and 28th Chinese Control Conference, 2009, Shangai (China).** Piscataway: IEEE, 2009. p. 7036–7042.

OLFATI-SABER, R.; FAX, J. A.; MURRAY, R. M. Consensus and cooperation in networked multi-agent systems. **Proceedings of the IEEE**, v. 95, n. 1, p. 215–233, 2007.

OLFATI-SABER, R.; MURRAY, R. M. Consensus problems in networks of agents with switching topology and time-delays. **IEEE Transactions on Automatic Control**, v. 49, n. 9, p. 1520–1533, 2004.

OLFATI-SABER, R.; SHAMMA, J. S. Consensus filters for sensor networks and distributed sensor fusion. *In*: **Proceedings of the 44th IEEE Conference on Decision and Control (CDC), 2005, Seville (Spain).** Piscataway: IEEE, 2005. p. 6698–6703.

OTHMAN, M. F.; SHAZALI, K. Wireless sensor network applications: A study in environment monitoring system. **Procedia Engineering**, v. 41, p. 1204–1210, 2012.

PALHARES, R. M.; PERES, P. L. D. LMI approach to the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ filtering design for discrete-time uncertain systems. **IEEE Transactions on Aerospace and Electronic Systems**, v. 37, n. 1, p. 292–296, 2001.

PETERSEN, I. R.; MCFARLANE, D. C. Optimal guaranteed cost filtering for uncertain discrete-time linear systems. **International Journal of Robust and Nonlinear Control**, v. 6, n. 4, p. 267–280, 1996.

RASTGAR, F.; RAHMANI, M. Consensus-based distributed robust filtering for multisensor systems with stochastic uncertainties. **IEEE Sensors Journal**, v. 18, n. 18, p. 7611–7618, 2018.

RASTGAR, F.; RAHMANI, M. Distributed robust filtering with hybrid consensus strategy for sensor networks. **IET Wireless Sensor Systems**, v. 10, n. 1, p. 37–46, 2020.

REN, W.; BEARD, R. W.; ATKINS, E. M. Information consensus in multivehicle cooperative control. **IEEE Control Systems Magazine**, v. 27, n. 2, p. 71–82, 2007.

REYNOLDS, C. W. Flocks, herds and schools: A distributed behavioral model. **Computer Graphics**, v. 21, n. 4, p. 25–34, 1987.

ROCHA, K. D. T.; TERRA, M. H. Robust distributed consensus-based filtering for uncertain systems over sensor networks. **IFAC-PapersOnLine**, v. 53, n. 2, p. 3571–3576, 2020.

SAYED, A. H. A framework for state-space estimation with uncertain models. **IEEE Transactions on Automatic Control**, v. 46, n. 7, p. 998–1013, 2001.

SAYED, A. H.; CHEN, H. A uniqueness result concerning a robust regularized least-squares solution. **Systems & Control Letters**, v. 46, n. 5, p. 361–369, 2002.

SAYED, A. H.; NASCIMENTO, V. H. Design criteria for uncertain models with structured and unstructured uncertainties. *In*: GARULI, A.; TESI, A.; VICINO, A. (ed.). **Robustness in identification and control**. London: Springer, 1999. p. 159–173.

SHAKED, U.; XIE, L.; SOH, Y. C. New approaches to robust minimum variance filter design. **IEEE Transactions on Signal Processing**, v. 49, n. 11, p. 2620–2629, 2001.

SHAKSHUKI, E.; REID, M. Multi-agent system applications in healthcare: Current technology and future roadmap. **Procedia Computer Science**, v. 52, p. 252–261, 2015.

SHEN, B.; WANG, Z.; HUNG, Y. Distributed $\mathcal{H}_\infty$-consensus filtering in sensor networks with multiple missing measurements: The finite-horizon case. **Automatica**, v. 46, n. 10, p. 1682–1688, 2010.

SOUZA, C. E. de; COUTINHO, D.; KINNAERT, M. Mean square state estimation for sensor networks. **Automatica**, v. 72, p. 108–114, 2016.

SPEYER, J.; DEYST, J.; JACOBSON, D. Optimization of stochastic linear systems with additive measurement and process noise using exponential performance criteria. **IEEE Transactions on Automatic Control**, v. 19, n. 4, p. 358–366, 1974.

STURM, J. F. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. **Optimization Methods and Software**, v. 11, n. 1-4, p. 625–653, 1999.

THEODOR, Y.; SHAKED, U. Robust discrete-time minimum-variance filtering. **IEEE Transactions on Signal Processing**, v. 44, n. 2, p. 181–189, 1996.

TIAN, T.; SUN, S.; LI, N. Multi-sensor information fusion estimators for stochastic uncertain systems with correlated noises. **Information Fusion**, v. 27, p. 126–137, 2016.

TILLERSON, M.; INALHAN, G.; HOW, J. P. Coordination and control of distributed spacecraft systems using convex optimization techniques. **International Journal of Robust and Nonlinear Control**, v. 12, n. 2, p. 207–242, 2002.

VICSEK, T. *et al.* Novel type of phase transition in a system of self-driven particles. **Physical Review Letters**, v. 75, n. 6, p. 1226–1229, 1995.

WANG, L. *et al.* Event-based variance-constrained $\mathcal{H}_\infty$ filtering for stochastic parameter systems over sensor networks with successive missing measurements. **IEEE Transactions on Cybernetics**, v. 48, n. 3, p. 1007–1017, 2018.

XIAO, L.; BOYD, S.; LALL, S. A scheme for robust distributed sensor fusion based on average consensus. *In*: **Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN), 2005.** Piscataway: IEEE, 2005. p. 63–70.

XIE, L. *et al.* Improved robust $\mathcal{H}_2$ and $\mathcal{H}_\infty$ filtering for uncertain discrete-time systems. **Automatica**, v. 40, n. 5, p. 873–880, 2004.

XIE, L.; SOH, Y. C.; SOUZA, C. E. de. Robust Kalman filtering for uncertain discrete-time systems. **IEEE Transactions on Automatic Control**, v. 39, n. 6, p. 1310–1314, 1994.

XIE, L.; SOUZA, C. E. de; FU, M. $\mathcal{H}_\infty$ estimation for discrete-time linear uncertain systems. **International Journal of Robust and Nonlinear Control**, v. 1, n. 2, p. 111–123, 1991.

XU, H.; MANNOR, S. A Kalman filter design based on the performance/robustness tradeoff. **IEEE Transactions on Automatic Control**, v. 54, n. 5, p. 1171–1175, 2009.

YANG, F.; LI, Y. Set-membership filtering for polytopic uncertain discrete-time systems. **International Journal of General Systems**, v. 40, n. 7, p. 741–754, 2011.

ZANELLA, A. *et al.* Internet of things for smart cities. **IEEE Internet of Things Journal**, v. 1, n. 1, p. 22–32, 2014.

ZHANG, J.; HE, X.; ZHOU, D. Distributed filtering over wireless sensor networks with parameter and topology uncertainties. **International Journal of Control**, v. 93, n. 4, p. 910–921, 2018.

ZHANG, J.; XIA, Y.; SHI, P. Parameter-dependent robust $\mathcal{H}_\infty$ filtering for uncertain discrete-time systems. **Automatica**, v. 45, n. 2, p. 560–565, 2009.

ZHU, X.; SOH, Y. C.; XIE, L. Design and analysis of discrete-time robust Kalman filters. **Automatica**, v. 38, n. 6, p. 1069–1077, 2002.

ZORZI, M. Robust Kalman filtering under model perturbations. **IEEE Transactions on Automatic Control**, v. 62, n. 6, p. 2902–2907, 2017.

APPENDIX

APPENDIX A

Matrix Analysis

This appendix contains a collection of results in matrix analysis that are used in this work. Some familiarity with basic linear algebra concepts is assumed. All of the definitions and results are extracted from the reference book by Horn and Johnson (2013), where the proofs, omitted here for brevity, can be found.

A.1 Matrix Inversion Lemmas

Lemma A.1. *(HORN; JOHNSON, 2013, **Sherman-Morrison-Woodbury Formula**) Consider matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{m\times m}$, and $D \in \mathbb{R}^{m\times n}$. If A, C, $(A+BCD)$, and $C^{-1} + DA^{-1}B$ are nonsingular, then*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

Lemma A.2. *(HORN; JOHNSON, 2013) As a consequence of Lemma A.1, if A, C, $A + BC^{-1}D$, and $(C + DA^{-1}B)$ are invertible, then*

$$(A + BC^{-1}D)^{-1}BC^{-1} = A^{-1}B(C + DA^{-1}B)^{-1}.$$

A.2 Partitioned Matrices and Schur Complement

Lemma A.3. *(HORN; JOHNSON, 2013, **Schur Complement**) Consider matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{m\times n}$, and $D \in \mathbb{R}^{m\times m}$. Let $M \in \mathbb{R}^{(n+m)\times(n+m)}$ be the partitioned matrix defined as*

$$M := \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

If A is nonsingular, we define the Schur complement of A in M *as*

$$M/A := D - CA^{-1}B.$$

Similarly, if D is nonsingular, we define the Schur complement of D in M *as*

$$M/D := A - BD^{-1}C.$$

Lemma A.4. *(HORN; JOHNSON, 2013,* ***Banachiewicz Inversion Formula****) Consider matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{m\times n}$, and $D \in \mathbb{R}^{m\times m}$. Let $M \in \mathbb{R}^{(n+m)\times(n+m)}$ be a partitioned matrix given by*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

(i) *Suppose A and M are nonsingluar, such that M/A is also nonsingular. Therefore,*

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(M/A)^{-1}CA^{-1} & -A^{-1}B(M/A)^{-1} \\ -(M/A)^{-1}CA^{-1} & (M/A)^{-1} \end{bmatrix}.$$

(ii) *Suppose D and M are nonsingluar, such that M/D is also nonsingular. Therefore,*

$$M^{-1} = \begin{bmatrix} (M/D)^{-1} & -(M/D)^{-1}BD^{-1} \\ -D^{-1}C(M/D)^{-1} & D^{-1} + D^{-1}C(M/D)^{-1}BD^{-1} \end{bmatrix}.$$

A.3 Positive Definite and Semidefinite Matrices

Definition A.1. *(HORN; JOHNSON, 2013) A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is* positive (negative) definite, *denoted as $A \succ 0$ ($A \prec 0$), if $x^TAx > 0$ ($x^TAx < 0$), $\forall x \in \mathbb{R}^n$, and $x^TAx = 0$ if $x = 0$.*

Definition A.2. *(HORN; JOHNSON, 2013) A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is* positive (negative) semidefinite, *denoted as $A \succeq 0$ ($A \preceq 0$), if $x^TAx \geq 0$ ($x^TAx \leq 0$), $\forall x \in \mathbb{R}^n$.*

Lemma A.5. *(HORN; JOHNSON, 2013) Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $B \in \mathbb{R}^{n\times m}$.*

(i) *Suppose that A is semidefinite. Then, B^TAB is also semidefinite and* $\operatorname{rank}(B^TAB) = \operatorname{rank}(AB)$.

(ii) *Suppose that A is definite. Then,* $\operatorname{rank}(B^TAB) = \operatorname{rank}(B)$. *Therefore, B^TAB is definite if, and only if, B has full column rank, i.e.,* $\operatorname{rank}(B) = m$.

A.4 Nonnegative Matrices

Definition A.3. *(HORN; JOHNSON, 2013) A matrix $A = [a_{ij}]$ is said to be* nonnegative *if all of its entries $a_{ij} \geq 0$. Analogously, matrix A is* positive *if all of its entries $a_{ij} > 0$.*

Definition A.4. *(HORN; JOHNSON, 2013) A square nonnegative matrix A is* row (column) stochastic *if all of its row (column) sums are unitary. If both row and column sums are all unitary, A is* doubly stochastic*.*

Definition A.5. *(HORN; JOHNSON, 2013) A matrix A is* reducible *if there exists a permutation matrix P such that $P^T AP$ is a block upper triangular matrix. Otherwise, the matrix is said to be* irreducible*.*

Definition A.6. *(HORN; JOHNSON, 2013) A square nonnegative matrix A is* primitive *if it is irreducible and has exactly one nonzero eigenvalue of maximum absolute value.*

Lemma A.6. *(HORN; JOHNSON, 2013) A matrix $A \in \mathbb{R}^{n\times n}$ is irreducible if, and only if, the directed graph associated with it is strongly connected. Equivalently, if A is symmetric, it is irreducible if, and only if, the associated undirected graph is connected.*

Theorem A.1. *(HORN; JOHNSON, 2013,* ***Geršgorin Disk Theorem****) Let $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ and let*

$$R'_i(A) = \sum_{j=1, i\neq j}^{n} |a_{ij}|, \; i = 1, \ldots, n,$$

denote the deleted absolute row sums of A. Then all eigenvalues of A are located in the union of n Geršgorin disks

$$G(A) = \bigcup_{i=1}^{n} \{z \in \mathbb{C} \, : \, |z - a_{ii}| \leq R'_i(A)\} \, .$$

Furthermore, if a union of k of these n disks forms a connected region that is disjoint from all of the remaining $n-k$ disks, then there are precisely k eigenvalues of A in this region.

Theorem A.2. *(HORN; JOHNSON, 2013,* ***Perron-Frobenius Theorem****) Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative irreducible matrix. Then,*

(i) *$\rho(A) > 0$;*

(ii) *$\rho(A)$ is an algebraically simple eigenvalue of A;*

(iii) *There is a unique vector $v \in \mathbb{R}^n$ such that $Av = \rho(A)v$;*

(iv) *There is a unique vector $w \in \mathbb{R}^n$ such that $w^T A = w^T \rho(A)$ and $v^T w = 1$.*

In the above, $\rho(A)$ denotes the spectral radius *of matrix A, given by $\rho(A) = \max\{|\lambda_i|\}$, where λ_i are the distinct eigenvalues of A.*

Lemma A.7. *(HORN; JOHNSON, 2013) Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative primitive matrix with right and left Perron vectors v and w, respectively. Then,*

$$\lim_{m\to\infty} \left[\rho(A)^{-1} A\right]^m = vw^T.$$

A.5 Kronecker Products

Definition A.7. *(HORN; JOHNSON, 2013) The Kronecker product of matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{p\times q}$ is defined as*

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{nm}B \end{bmatrix},$$

which satisfies the following properties, assuming compatible dimensions:

(i) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$;

(ii) $A \otimes (B + C) = A \otimes B + A \otimes C$;

(iii) $(A \otimes B)^T = A^T \otimes B^T$;

(iv) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, *provided that both A and B are nonsingular.*