Syracuse University

SURFACE

Dissertations - ALL   SURFACE

May 2016

Resource Management for Distributed Estimation via Sparsity-Promoting Regularization

Sijia Liu
Syracuse University

Follow this and additional works at: https://surface.syr.edu/etd

Part of the Engineering Commons

ABSTRACT

Recent advances in wireless communications and electronics have enabled the development of low-cost, low-power, multifunctional sensor nodes that are small in size and communicate untethered in a sensor network. These sensor nodes can sense, measure, and gather information from the environment and, based on some local processing, they transmit the sensed data to a fusion center that is responsible for making the global inference. Sensor networks are often tasked to perform parameter estimation; example applications include battlefield surveillance, medical monitoring, and navigation. However, under limited resources, such as limited communication bandwidth and sensor battery power, it is important to design an energy-efficient estimation architecture. The goal of this thesis is to provide a fundamental understanding and characterization of the optimal tradeoffs between estimation accuracy and resource usage in sensor networks.

In the thesis, two basic issues of resource management are studied, sensor selection/scheduling and sensor collaboration for distributed estimation, where the former refers to finding the best subset of sensors to activate for data acquisition in order to minimize the estimation error subject to a constraint on the number of activations, and the latter refers to seeking the optimal inter-sensor communication topology and energy allocation scheme for distributed estimation systems. Most research on resource management so far has been based on several key assumptions, a) independence of observation, b) strict resource constraints, and c) absence of inter-sensor communication, which lend analytical tractability to the problem but are often found lacking in practice. This thesis introduces novel techniques to relax these assumptions and provide new insights into addressing resource management problems.

The thesis analyzes how noise correlation affects solutions of sensor selection problems, and proposes both a convex relaxation approach and a greedy algorithm to find these solutions. Compared to the existing sensor selection approaches that are limited to the case of uncorrelated noise or weakly correlated noise, the methodology proposed in this thesis is valid for any arbitrary noise

correlation regime. Moreover, this thesis shows a correspondence between active sensors and the nonzero columns of an estimator gain matrix. Based on this association, a sparsity-promoting optimization framework is established, where the desire to reduce the number of selected sensors is characterized by a sparsity-promoting penalty term in the objective function. Instead of placing a hard constraint on sensor activations, the promotion of sparsity leads to trade-offs between estimation performance and the number of selected sensors. To account for the individual power constraint of each sensor, a novel sparsity-promoting penalty function is presented to avoid scenarios in which the same sensors are successively selected. For solving the proposed optimization problem, we employ the alternating direction method of multipliers (ADMM), which allows the optimization problem to be decomposed into subproblems that can be solved analytically to obtain exact solutions.

The problem of sensor collaboration arises when inter-sensor communication is incorporated in sensor networks, where sensors are allowed to update their measurements by taking a linear combination of the measurements of those they interact with prior to transmission to a fusion center. In this thesis, a sparsity-aware optimization framework is presented for the joint design of optimal sensor collaboration and selection schemes, where the cost of sensor collaboration is associated with the number of nonzero entries of a collaboration matrix, and the cost of sensor selection is characterized by the number of nonzero rows of the collaboration matrix. It is shown that a) the presence of sensor collaboration smooths out the observation noise, thereby improving the quality of the signal and eventual estimation performance, and b) there exists a trade-off between sensor selection and sensor collaboration. This thesis further addresses the problem of sensor collaboration for the estimation of time-varying parameters in dynamic networks that involve, for example, time-varying observation gains and channel gains. Impact of parameter correlation and temporal dynamics of sensor networks on estimation performance is illustrated from both theoretical and practical points of view. Last but not least, optimal energy allocation and storage control polices are designed in sensor networks with energy-harvesting nodes. We show that the resulting optimization problem can be solved as a special nonconvex problem, where the only source of

nonconvexity can be isolated to a constraint that contains the difference of convex functions. This specific problem structure enables the use of a convex-concave procedure to obtain a near-optimal solution.

To my family:

past, present, and future.

RESOURCE MANAGEMENT FOR DISTRIBUTED

ESTIMATION VIA SPARSITY-PROMOTING

REGULARIZATION

By

Sijia Liu

B. S., Xi'an Jiaotong University, Xi'an, China, 2008
M. S., Xi'an Jiaotong University, Xi'an, China, 2011

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering

Syracuse University
May 2016

ACKNOWLEDGMENTS

Pramod and Makan, thank you so much for your continued intellectual, motivational and financial support during my doctoral studies. It was a great pleasure to be your student.

Pramod, thank you for always supporting me and believing in me for the tremendous patience and a friendly supervision. It is not often that one finds an advisor that always finds the time for listening to the little problems and roadblocks that unavoidably crop up in the course of performing research. However, I found one whom never failed to inspire me and teach me in the past 5 years.

Makan, thank you for your unreserved help and guidance. Your words always inspire me and bring me to a higher level of thinking. What I learnt is not just how to survive in doctoral studies, but how to make my student career colorful and successful. Thank you for the amazing last 5 years.

I would like to take this opportunity to thank my doctoral committee: Prof. Yingbin Liang, Prof. Biao Chen, Prof. Lixin Shen and Prof. Vinod Sharma. It was also my pleasure to have Prof. Yanzhi Wang in my dissertation defense. I am greatly indebted to my Sensor Fusion Group colleagues and other friends, including Engin Masazade, Ruixin Niu, Arun Subramanian, Sid Nadendla, Aditya Vempaty, Swastik Brahma, Raghed Ahmad El Bardan, Thakshila Wimalajeewa, Prashant Khanduri, Pranay Sharma, Swatantra Kafle, Qunwei Li, Shan Zhang, Nianxia Cao, Bhavya Kailkhura, Hao He, Yujiao Zheng, Xiaojing Shen, Gang Li, and Sundeep Prabhakar Chepuri.

Lastly, I would like to thank my parents and grandparents for their unconditional love and care. Fangrong, thank you for your help during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xiv

CHAPTER 1

INTRODUCTION

Over the last decade, advances in micro-electro-mechanical systems and in low-power sensor devices have created technical capabilities to realize multi-functional wireless sensor networks (WSNs). WSNs have been widely used for environmental monitoring (e.g., temperature, pollution, precipitation sensing) [1], medical monitoring (e.g., monitoring patients both in the clinical setting and at home, body health monitoring using wearable sensors) [2], power networks (e.g., power system state estimation, electricity market forcasting) [3], localization and surveillance (e.g., target tracking, indoor navigation) [4], and so on. In many applications, the task of sensor networks is to estimate an unknown parameter/state of interest, such as field intensity and target location. This thesis focuses on distributed estimation using WSNs.

With pervasive sensors continuously collecting and storing massive amounts of information, there is no doubt this is an era of data deluge. However, running analytics on all the gathered data by central processors is infeasible, since the computing resource, capacity of storage units, communication bandwidth, and sensor battery power are limited in networks. Under limited resources, it is of paramount importance to accquire only the most informative data and design an energy-efficient architecture for estimation tasks. Therefore, the problem of resource management (e.g., sensor selection/scheduling, and energy allocation) arises in order to reach a desirable estimation accuracy in an energy-efficient manner. However, finding the optimal resource management strat-

egy is difficult for most scenarios. The problem becomes more involved when sensing is corrupted by correlated measurement noise, or when inter-sensor communication is established in sensor networks to improve estimation performance as well as conserve resources.

In this thesis, we consider two types of resource management problems in sensor networks: a) sensor selection/scheduling, and b) energy allocation for collaborative estimation. The resource management problems are addressed in the context of distributed estimation, where sensors perform some resource-conscious local processing before transmission to the fusion center (FC). The ultimate goal of our research is the integrated design of local signal processing operations and strategies to achieve an optimal trade-off between resource consumption and estimation accuracy.

1.1   Sensor selection and scheduling

Due to limited network resources, such as limited communication bandwidth and sensor battery power, it may not be desirable to have all the sensors report their measurements at all time instants. Therefore, the problem of sensor selection/scheduling has attracted attention to strike a balance between estimation accuracy and sensor activations over space and/or time. The importance of sensor selection/scheduling has been discussed extensively in the context of various applications, such as target tracking [5], bit allocation [6], field monitoring [7], optimal control [8], power allocation [9], optimal experiment design [10], and leader selection in consensus networks [11].

Many research efforts have focused on the problem of sensor selection (also known as myopic sensor scheduling) [12–16], where at every instant the search is for the best sensors to be activated at the next time step (as opposed to a longer time horizon). The myopic selection policy has some benefits. One benefit is that the problem of sensor scheduling for a single time step is easily formulated under certain performance criteria, such as mutual information [12], entropy [13], and estimation error [14, 15]. Another benefit is that many myopic scheduling methods are computationally efficient; examples of approaches include tree search [14], and convex relaxation [15, 16].

The problem of non-myopic scheduling, where sensor schedules are determined for multiple

future time steps, has received a fair amount of interest in recent years [17–20]. In [17], two branch-and-bound-based pruning algorithms via a tree structure were proposed to solve the non-myopic sensor scheduling problem. The authors in [18] presented a multi-step sensor selection strategy by using convex relaxations via a reformulation of the Kalman filter, which is able to address different performance metrics and constraints on available resources. In [19], the problem of non-myopic sensor scheduling is transformed into the maximization of a generalized information gain. In [20], an approximate dynamic programming based non-myopic sensor selection method is proposed where dynamic programming is performed efficiently for short time horizons.

However, if the length of the time horizon becomes large or infinite then finding an optimal non-myopic schedule would be difficult, since the number of sensor sequences grows intractably large as the time horizon grows. Therefore, some researchers have considered the problem of periodic sensor schedules on an infinite time horizon [21, 22]. In [8, 23], periodicity of the optimal sensor schedule was observed even for finite time horizon problems in which a periodic schedule was not assumed *a priori*. Furthermore, in [24] it was shown that the optimal sensor schedule for an infinite horizon problem can be approximated arbitrarily well by a periodic schedule with a finite period. Although periodicity makes infinite horizon sensor scheduling problems tractable, via the design of optimal sensor schedule over a finite time period, it brings in other challenges such as periodicity constraint in problem formulation and optimization compared to conventional sensor scheduling problems.

1.2   Sensor collaboration

In the context of distributed estimation, sensor collaboration refers to the act of sharing measurements with neighboring sensors prior to transmission to the FC; see an illustrative example in Fig. 1.1. Here sensors are allowed to update their measurements by taking a linear combination of the measurements of those they interact with.

In the absence of collaboration, the estimation architecture reduces to a classical distributed es-
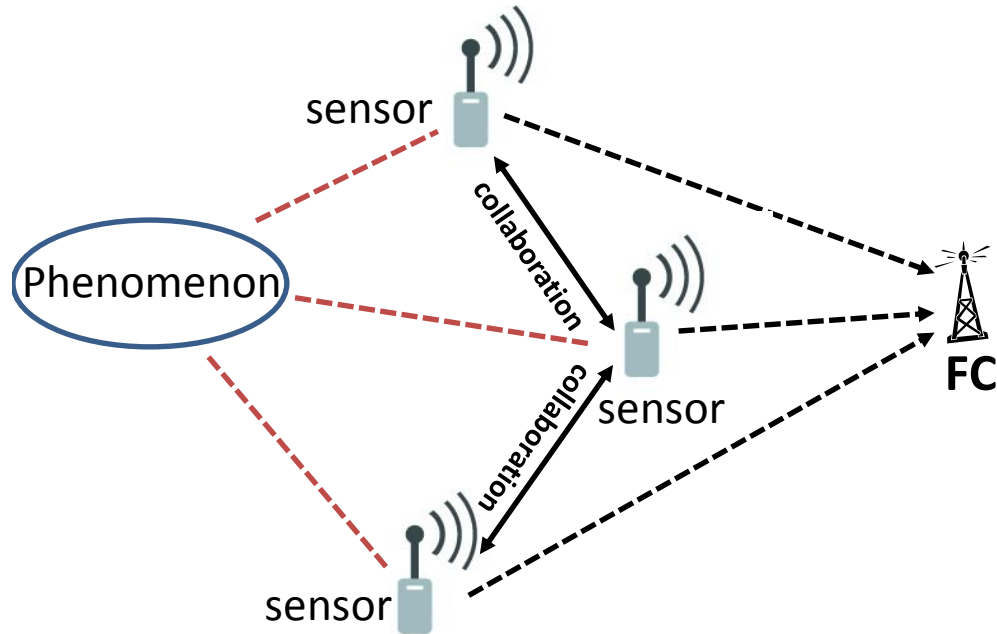
Fig. 1.1: Example of collaborative estimation system.

timation network, where scaled versions of sensor measurements are transmitted using an amplify-and-forward strategy [25]. In this setting, one of the key problems is to design the optimal power amplifying factors to reach certain design criteria for performance measures, such as estimation distortion and energy cost. Several variations of the conventional distributed estimation problem have been addressed in the literature depending on the quantity to be estimated (random parameter or process) [26, 27], the type of communication (analog-based or quantization-based) [28, 29], nature of transmission channels (coherent or orthogonal) [30, 31] and energy constraints [32]. The problem of sensor collaboration was first proposed in [33] by assuming an orthogonal multiple access channel (MAC) setting with a fully connected network, where all the sensors are allowed to collaborate. The presence of sensor collaboration smooths out the observation noise, thereby improving the quality of the signal and the eventual estimation performance.

1.3   Summary of contributions and outline of thesis

This thesis is organized into four parts. In the first part of the thesis (i.e., Chapter 2), the distributed signal processing framework is formally introduced. Also, we review important references that are closely related to our work, and review some important concepts in estimation and convex/nonconvex optimization theory. In the second part of the thesis (i.e., Chapters 3–5), theory and algorithms for optimal sensor selection/scheduling are discussed in depth. More specifically, we derive the solution to a) the problem of sensor selection with correlated measurement noise, b) the problem of optimal periodic sensor scheduling, and c) the problem of sensor scheduling from an energy balance point of view. In the third part of the thesis (i.e., Chapters 6–8), design of optimal sensor collaboration strategy with nonzero collaboration cost and unknown collaboration topologies is studied. More specifically, we derive the solution to a) the problem of sensor collaboration for estimation of static parameters, b) the problem of sensor collaboration for estimation of dynamic parameters, and c) the problem of sensor collaboration in networks with energy harvesting nodes. Finally, the thesis concludes with the fourth part (i.e., Chapter 9), where we pose some interesting problems for future research. The content of Chapters 3–8 has been published as papers, which include some additive subtopics that are not included in this thesis. The relationship between these chapters and the publications is shown in the next section.

Chapter 2 presents preliminaries required for the later chapters of the thesis, which include the system model, theory and algorithms for estimation and optimization. In particular, the alternating direction method of multipliers (ADMM) is elaborated on, which is well-suited for optimization problems that involve sparsity-inducing regularizers. Moreover, we present an extensive literature review of resource management problems that are closely related to our work.

Chapter 3 focuses on the problem of sensor selection with correlated measurement noise. In this chapter, we seek optimal sensor activations by formulating an optimization problem, in which the estimation error, given by the trace of the inverse of the Bayesian Fisher information matrix, is minimized subject to energy constraints. Fisher information has been widely used as an effective sensor selection criterion. However, existing information-based sensor selection methods are

limited to the case of uncorrelated noise or weakly correlated noise due to the use of approximate metrics. By contrast, here we derive the closed form of the Fisher information matrix with respect to sensor selection variables that is valid for any arbitrary noise correlation regime, and develop both a convex relaxation approach and a greedy algorithm to find near-optimal solutions. Furthermore, we generalize our framework of sensor selection to solve the problem of non-myopic sensor scheduling.

Chapter 4 focuses on the problem of optimal time-periodic sensor schedules for estimating the state of discrete-time dynamical systems. Here multiple sensors are deployed in a region of interest and sensors are subject to resource constraints, which limit the number of times each can be activated over one period of the periodic schedule. We seek an algorithm that strikes a balance between estimation accuracy and total sensor activations over one period. We make a correspondence between active sensors and the nonzero columns of the estimator gain. We formulate an optimization problem in which we minimize the trace of the error covariance with respect to the estimator gain while simultaneously penalizing the number of nonzero columns of the estimator gain. For solving the resulting optimization problem, we employ ADMM, which allows the problem to be decomposed into subproblems that can either be solved efficiently using iterative numerical methods or solved analytically to obtain exact solutions.

Chapter 5 presents a novel sparsity-promoting sensor scheduling framework, which has the advantage of discouraging the excessive use of the same sensors in the network. In the sparsity-promoting framework of sensor selection/scheduling, a sensor being off at a certain time instant is represented by the corresponding column of the estimator coefficient matrix being identically equal to zero. However, the existing sparsity-promoting techniques lead to scenarios in which the most informative sensors are successively selected. Successive selections would result in faster energy depletion of these sensors, which may render the network nonfunctional. In order to achieve a balance between activating the most informative sensors and uniformly allocating sensor energy, we propose a novel sparsity-promoting approach by adding a penalty function that avoids scenarios where the same sensors are successively selected.

Chapter 6 focuses on the problem of sensor collaboration while incorporating the cost of sensor collaboration and selection. To determine the optimal sensor collaboration and selection schemes, we associate (a) the cost of sensor collaboration with the number of nonzero entries of the collaboration matrix (i.e., its overall sparsity), and (b) the cost of sensor selection with the number of nonzero rows of the collaboration matrix (i.e., its row-sparsity). Based on these associations, we present a sparsity-inducing optimization framework that jointly designs the optimal sensor selection and collaboration schemes. It is shown that there exists a trade-off between sensor collaboration and sensor selection for a given estimation performance.

Chapter 7 studies the problem of sensor collaboration for estimation of time-varying parameters in dynamic sensor networks. Based on prior knowledge about parameter correlation, the resulting sensor collaboration problem is solved for the estimation of temporally uncorrelated and correlated parameters. In the case of temporally uncorrelated parameters, we show that the sensor collaboration problem can be cast as a special nonconvex optimization problem, where a difference of convex functions carries all the nonconvexity. By exploiting problem structure, we solve the problem by using a convex-concave procedure, which renders a near optimal solution evidenced by numerical results. In the case of correlated parameters, we show that the sensor collaboration problem can be converted into a semidefinite program together with a nonconvex rank-one constraint. Spurred by the problem structure, we employ a semidefinite programming based penalty convex-concave procedure to solve the sensor collaboration problem. In order to improve computational efficiency, we further propose a fast algorithm that scales gracefully with problem size via ADMM. We empirically show the impact of parameter correlation and temporal dynamics of sensor networks on the performance of distributed estimation with sensor collaboration.

Chapter 8 extends the theory developed in Chapters 6–7 to the collaboration problem with energy harvesting sensors. Here we propose optimal energy allocation and storage control polices at each time step by minimizing the estimation distortion subject to energy harvesting constraints.

Chapter 9 contains the conclusion and outlines a number of directions for future research.

1.4   Bibliographic notes

Most of the research work appearing in this thesis has either been already published or in different stages of publication at various venues. The relationship between the chapters and the publications is shown in Table 1.1, while the list of publications is provided as follows.

Publications related to the thesis

Journal papers

J1 S. Liu, S. Kar, M. Fardad and P. K. Varshney, "Optimized Sensor Collaboration for Estimation of Temporally Correlated Parameters," *IEEE Transactions on Signal Processing (Submitted)*, Mar. 2016

J2 S. Liu, S. P. Chepuri, M. Fardad, E. Masazade, G. Leus and P. K. Varshney, "Sensor Selection for Estimation with Correlated Measurement Noise," to appear in *IEEE Transactions on Signal Processing*, 2016

J3 S. Liu, S. Kar, M. Fardad and P. K. Varshney, "Sparsity-Aware Sensor Collaboration for Linear Coherent Estimation," *IEEE Transactions on Signal Processing*, vol.63, no.10, pp. 2582–2596, May 2015

J4 S. Liu, A. Vempaty, M. Fardad, E. Masazade and P. K. Varshney, "Energy-Aware Sensor Selection in Field Reconstruction," *IEEE Signal Processing Letters*, vol.21, no.12, pp.1476-1480, Dec. 2014

J5 S. Liu, M. Fardad, E. Masazade and P. K. Varshney, "Optimal Periodic Sensor Scheduling in Networks of Dynamical Systems," *IEEE Transactions on Signal Processing*, vol.62, no.12, pp.3055–3068, June 2014

Conference papers

C1 **S. Liu**, Y. Wang, M. Fardad and P. K. Varshney, "Optimal Energy Allocation and Storage Control for Distributed Estimation with Sensor Collaboration," in *Proc. the 50th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2016

C2 **S. Liu**, S. Kar, M. Fardad and P. K. Varshney, "On Optimal Sensor Collaboration for Distributed Estimation with Individual Power Constraints," in *Proc. Asilomar Conference on Signals, Systems and Computers (Asilomar)*, Nov. 2015

C3 **S. Liu**, F. Chen, A. Vempaty, M. Fardad, L. Shen and P. K. Varshney, "Sparsity-Promoting Sensor Management for Estimation: An Energy Balance Point of View," in *Proc. International Conference on Information Fusion (Fusion)*, July 2015

C4 **S. Liu**, E. Masazade, M. Fardad and P. K. Varshney, "Sensor Selection with Correlated Measurements for Target Tracking in Wireless Sensor Networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015

C5 **S. Liu**, M. Fardad, S. Kar and P. K. Varshney, "On Optimal Sensor Collaboration Topologies for Linear Coherent Estimation," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2014

C6 **S. Liu**, E. Masazade, M. Fardad and P. K. Varshney, "Sparsity-Aware Field Estimation via Ordinary Kriging," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014

C7 **S. Liu**, M. Fardad, E. Masazade and P. K. Varshney, "On Optimal Periodic Sensor Scheduling for Field Estimation in Wireless Sensor Networks," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2013

Other related contributions

Journal papers

J6 **S. Liu**, S. Kar, M. Fardad and P. K. Varshney, "Joint Design of Optimal Sensor Selection and Collaboration Strategies for Distributed Estimation," *IEEE ComSoc MMTC E-letter, Special Issue on "Energy efficiency management for distributed computation and applications in sensor networks"*, Mar. 2016

J7 B. Kailkhura, **S. Liu**, T. Wimalajeewa and P. K. Varshney, "Measurement Matrix Design for Compressive Detection with Secrecy Guarantees," *IEEE Wireless Communications Letters (submitted)*, Feb. 2016

J8 X. Shen, **S. Liu** and P. K. Varshney, "Sensor Selection for Nonlinear Systems in Large Sensor Networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol.50, no.4, Oct. 2014

Conference papers

C8 V. Gupta, B. Kailkhura, T. Wimalajeewa, **S. Liu** and P. K. Varshney "Joint Sparsity Pattern Recovery with 1-bit Compressive Sensing in Sensor Networks," in *Proc. Asilomar Conference on Signals, Systems and Computers (Asilomar)*, Nov. 2015

C9 V. S. S. Nadendla, **S. Liu** and P. K. Varshney, "On Enhancing Secrecy in Centralized Detection using Transmit-Beamforming with Artificial Noise," in *Proc. Annual Allerton Conference on Communication, Control and Computing (Allerton)*, Oct. 2015

C10 **S. Liu**, E. Masazade, X. Shen and P. K. Varshney, "Adaptive Non-Myopic Quantizer Design for Target Tracking in Wireless Sensor Networks," in *Proc. Asilomar Conference on Signals, Systems and Computers (Asilomar)*, Nov. 2013

C11 **S. Liu**, E. Masazade and P. K. Varshney, "Temporally Staggered Sensing for Field Estimation with Quantized Data in Wireless Sensor Networks," in *Proc. IEEE Statistical Signal*

Processing Workshop (SSP), Aug. 2012

	Chapter 3	Chapter 4	Chapter 5	Chapter 6	Chapter 7	Chapter 8
J1					●	
J2	●					
J3				●		
J4			●			
J5		●				
C1						●
C2					●	
C3			●			
C4	●					
C5				●		
C6			●			
C7		●				

Table 1.1: Connection between publications & chapters

CHAPTER 2

BACKGROUND: RESOURCE MANAGEMENT, ESTIMATION AND OPTIMIZATION

2.1 Introduction

Estimation in resource constrained sensor networks usually involves some local processing following certain resource-conscious management protocols. The goal of resource management is to determine the optimal resource allocation scheme, such as optimal sensor activation, sensor schedule, and energy allocation for local sensors. In this chapter, we begin by elaborating on differences between this thesis and the existing literature on sensor selection/scheduling and energy allocation. We then demonstrate the signal processing framework employed and review some relevant concepts and algorithms in estimation theory and convex/nonconvex optimization theory.

2.2   Literature review

Sensor selection with correlated noise

The problem of sensor selection with correlated noise was considered in several works [18, 19, 34, 35]. It was shown that the presence of correlated noise makes optimal sensor selection/scheduling problems more challenging, since Fisher information is a complicated non-linear function of the selection variables. This is in contrast to the scenario of sensor selection with uncorrelated noise, where each measurement contributes to Fisher information (equivalently, inverse of the Cramér-Rao bound on the error covariance matrix) in an additive manner [36]. In [34, 35], the problem of sensor selection with correlated noise was formulated so as to minimize the estimation error subject to an energy constraint and to minimize the energy consumption subject to an estimation constraint, respectively. In [19], a generalized information gain given by the trace of the Fisher information matrix was maximized while seeking optimal sensor schedules with correlated noise. In [18], a reformulation of the multi-step Kalman filter was introduced to schedule sensors for linear dynamical systems with correlated noise.

However, the sensor selection schemes of [19, 34, 35] considered an approximate formulation where the noise covariance matrix is assumed to be independent of the sensor selection variables. In contrast, in Chapter 3, we derive the closed form expression of the estimation error with respect to sensor selection variables under correlated measurement noise, which is valid for any arbitrary noise correlation matrix. This expression is optimized via a convex relaxation method to determine the optimal sensor selection scheme. We propose a greedy algorithm to solve the corresponding sensor selection problem, where we show that when an inactive sensor is made active, the increase in Fisher information yields an information gain in terms of a rank-one matrix. We also extend the proposed sensor selection approach to address the problem of non-myopic sensor scheduling, where the length of time horizon and energy constraints on individual sensors are taken into account. To the best of our knowledge, the sensor selection problem with correlated noise is solved for the first time in the thesis.

Periodic sensor scheduling

The design of optimal periodic sensor schedules has been recently studied in [21, 22, 37]. In [21], the authors construct the optimal periodic schedule only for two sensors. For a multiple sensor scenario, the work of [22] studied the problem of periodic sensor scheduling by assuming the process noise to be very small, which results in a linear matrix inequality (LMI) problem. As a consequence of the assumption that the process noise is negligible, the ordering of the measurements does not factor into the solution of this LMI problem. Clearly, a sensor schedule in which the order of sensor activations is irrelevant can not be optimal for sensor scheduling problems. For example, it was shown in [38] that temporally staggered sensor schedules constitute the optimal sensing policy. In [37], a lower bound on the performance of scheduling sensors over an infinite time horizon is obtained, and then an open-loop periodic switching policy is constructed by using a doubly substochastic matrix. The authors show that the presented switching policy achieves the best estimation performance as the period length goes to zero (and thus sensors are switched as fast as possible). In this thesis, we present a comparison of both the performance and the computational complexity of our methodology with other prominent work in the literature. We demonstrate that our method performs as well or significantly better than these works, and is computationally efficient for sensor scheduling in problems with large-scale dynamical systems.

The sensor scheduling framework presented in Chapter 4 relies on making a one-to-one correspondence between every sensor and a column of the estimator gain. Namely, a sensor being off at a certain time instant is equivalent to the corresponding column of the estimator gain being identically zero. This idea has been exploited in [39] on sparsity-promoting extended Kalman filtering, where sensors are scheduled only for the next time step and have no resource constraints involved. Different from [39], we consider a periodic sensor scheduling problem on an infinite time horizon, where measurement frequency constraints and periodicity place further restrictions on the number of nonzero columns of the time-periodic Kalman filter gain matrices.

Energy-aware sparse sensor management

A sparsity-promoting technique has been used for sensor selection in [16, 39, 40], where the constraint on the number of selected sensors is characterized by a sparsity-promoting penalty term in the objective function. This provides a tractable optimization framework to obtain the trade-off between the estimation performance and the number of selected sensors. Also, this sparsity-promoting framework facilitates the optimization procedure, where several efficient approaches can be employed, e.g., the alternating direction method of multipliers (ADMM) in [39, 40] and the projected subgradient method in [16].

The current sparsity-promoting techniques may lead to scenarios in which the most 'informative sensors' are successively selected because of certain performance criterion, e.g., larger mutual information with the target [41] or stronger correlation with the field point of interest [42]. This behavior would result in faster energy depletion of the most informative sensors. From the perspective of network lifetime [43], the death of the first sensor (or a percentage of sensors) can make the network nonfunctional. Therefore, it is desired to have a balanced use of sensors while discouraging the excessive use of any group of sensors in the network.

In Chapter 5, we propose a new sparsity-promoting penalty function, which penalizes successive selection of the same sensors. This framework generates sparse sensor schedules which achieve a trade-off between activating the most informative sensors and balancing the energy consumption in the network.

Sensor collaboration for linear coherent estimation

The problem of sensor collaboration over a coherent MAC was studied in [44, 45], where it was observed that even a partially connected network can yield performance close to that of a fully connected network, and the problem of sensor collaboration for a family of sparsely connected networks was investigated. Further, the problem of sensor collaboration for estimating a vector of random parameters is studied in [46]. The works [9, 33, 44–47] assumed that there is no cost associated with collaboration, the collaboration topologies are fixed and given in advance, and

the only unknowns are the collaboration weights used to combine sensor observations. A more relevant reference to the present work is [48], where the nonzero collaboration cost was taken into account for linear coherent estimation, and a greedy algorithm was developed for seeking the optimal collaboration topology in energy constrained sensor networks.

In Chapter 6, we present a non-convex optimization framework to solve the collaboration problem with nonzero collaboration cost. To elaborate, we describe collaboration through a collaboration matrix, in which the nonzero entries characterize the collaboration topology and the values of these entries characterize the collaboration weights. We introduce a formulation that simultaneously optimizes both the collaboration topology and the collaboration weights. In contrast, the optimization in [48] was performed in a sequential manner, where a sub-optimal collaboration topology was first obtained, and then the optimal collaboration weights were sought. The new formulation leads to a more efficient allocation of energy resources as evidenced by improved distortion performance in numerical results.

In the existing literature [9, 33, 44–48], sensor collaboration was studied in static networks, where sensors take a single snapshot of the static parameter, and then initiate sensor collaboration protocols designed for single-snapshot estimation. Moreover, the parameters to be estimated, such as daily temperature and precipitation in environmental monitoring [49, 50], are often temporally correlated. Therefore, development of sensor collaboration schemes for the estimation of temporally correlated parameters in dynamic sensor networks is an attractive and important problem.

In Chapter 7, we find the optimal sensor collaboration scheme at each time step by minimizing the estimation distortion over a finite time horizon subject to individual energy constraints of sensors. Due to the presence of (a) temporal dynamics in the system, (b) temporal correlation of parameters, and (c) energy constraints in time, the design of optimal sensor collaboration schemes at multiple time steps is coupled with each other, and thus poses many challenges in problem formulation and optimization compared to the previous work. For example, when parameters of interest are temporally correlated, expressing the estimation distortion in a succinct closed form (with respect to collaboration variables) becomes intractable. It should be pointed out that even for

uncorrelated parameters, finding the optimal collaboration scheme for each time step is nontrivial since energy constraints are temporally inseparable.

Collaborative estimation with energy harvesting sensors

Several research efforts have been devoted to the design of optimal energy allocation schemes in energy harvesting networks for distributed estimation without inter-sensor collaboration [51–54]. In [51], a single sensor equipped with an energy harvester was used for parameter estimation, where the optimal energy allocation strategy was designed based on both causal and non-causal side information of energy harvesting. In [52], the design of energy allocation was studied over an orthogonal multiple access channel (MAC), where the estimation distortion resulting from the best linear unbiased estimator was minimized subject to energy harvesting constraints. In [53], a stochastic control problem was formulated for power allocation over a finite or an infinite time horizon.

In Chapter 8, we present a unified optimization framework for the joint design of optimal energy allocation and storage control polices while incorporating the cost of sensor collaboration over a finite time horizon. We show that the resulting optimization problem is nonconvex. However, by identifying the special types of nonconvexities, the methods of convex relaxation and restriction can be effectively used to find locally optimal solutions. Extensive numerical results are provided to demonstrate the utility of our approach for energy allocation and storage control in collaborative estimation.

2.3   Preliminaries: signal processing framework

The task of many sensor networks is to estimate an unknown parameter, such as a target location or field intensity. In the system, sensors first accquire raw measurements about the parameter via a given sensing model. The raw measurements are then locally processed based on certain resource management protocols prior to transmission to the fusion center (FC). The FC determines the

global estimate of the parameter by using a certain estimator. The general distributed estimation architecture consists of several blocks as depicted in Fig. 2.1.
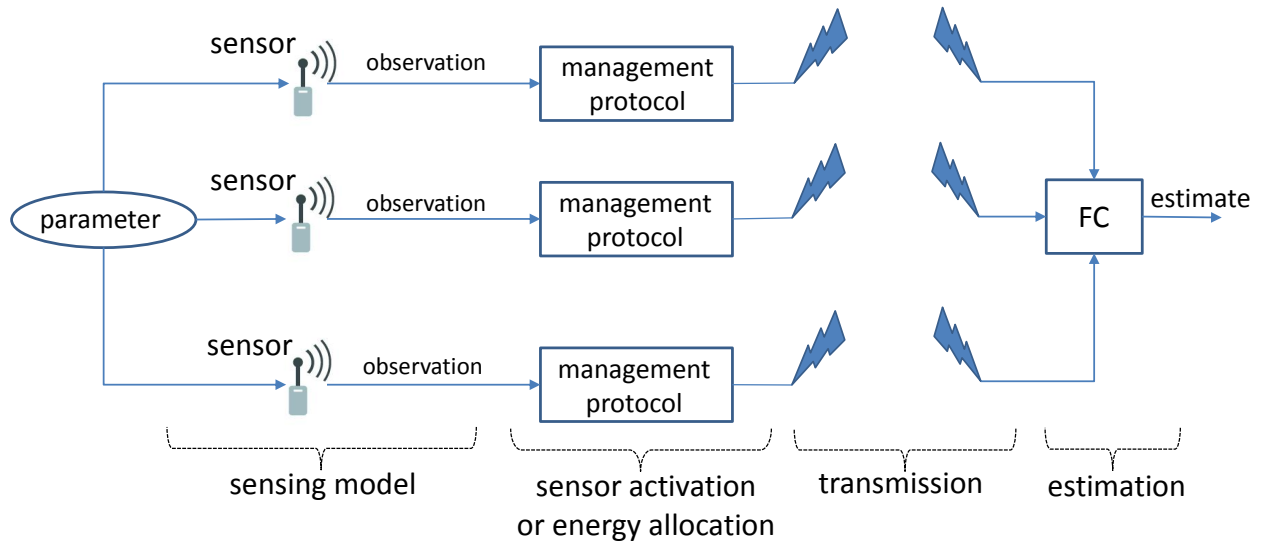
Fig. 2.1: System diagram.

Depending on the particular application, several models of signal, sensing and transmission are assumed in the thesis. The estimation application can either be single-snapshot or over a finite time horizon as well as static or dynamic. We specify each system block as follows.

Signal model

The signal of interest may either be an unknown deterministic parameter or a random parameter with certain prior knowledge about its statistics or distribution. In Chapter 3, the signal is assumed to be a random Gaussian variable. In Chapter 4, the signal is modeled as the state of a linear dynamical system. In Chapter 5, the signal is characterized as a spatially correlated field intensity. In Chapters 6–8, the signal is an arbitrary random parameter/process with prior knowledge about its second-order statistics.

Sensing model

A linear sensing model is used in the thesis. In Chapters 3–5, the observation gain is assumed to be constant. In Chapters 6–8, temporal dynamics of the observation gain are incorporated in the system model. We also remark that non-linear and quantization-based sensing models have been considered in our work [42, 55–57].

Management protocol

The thesis focuses on two types of resource management problems: a) sensor selection/scheduling in Chapters 3–5, and b) sensor collaboration in Chapters 6–8. The former aims to strike a balance between estimation accuracy and energy usage. And the latter is to find the optimal inter-sensor communication scheme subject to a certain information or energy constraint. In this thesis, design of the optimal resource management scheme is completed offline. An online resource management approach for target tracking has been reported in [55–57], which are not included in this thesis.

Transmission model

The digital/analog information is transmitted to the FC over a noise-corrupted communication channel. In Chapters 6–8, we consider a coherent multiple access channel (MAC), where sensors coherently form a beam into a common channel received at the FC.

Estimation

The FC chooses to implement a particular fusion rule, namely, estimator in the thesis. To estimate a random parameter, we adopt the linear minimum mean-squared error estimator (LMMSE) under the Bayesian setup. The resulting mean squared error (MSE) is then employed to evaluate the estimation performance of resource management. Given a linear Gaussian observation model, the LMMSE is the optimal estimator in the sense of minimum MSE. In Chapter 4, we use a Kalman filter to track the state of a random process.

2.4 Preliminaries: Bayesian estimation and Fisher information

Given the prior and the sensor measurements, the goal of the Bayesian approach is to estimate a particular realization of a random variable $\boldsymbol{\theta}$. The optimal Bayesian estimator that minimizes the MSE is given by the mean of the posterior PDF. However, the optimal Bayesian estimator is often difficult to determine in a closed form, which involves multidimensional integration. In this thesis, we foucs on the linear Baysian estimatior that minimizes the MSE, which is referred to as LMMSE.

If the sensor measurements are described by the linear sensing model

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v}, \tag{2.1}$$

where $\mathbf{y}$ is a measurement vector, $\mathbf{H}$ is a known observation matrix, $\boldsymbol{\theta}$ is the parameter to be estiamted with mean $\mathbb{E}(\boldsymbol{\theta})$ and covariance matrix $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}$, and $\mathbf{v}$ is a random vector with zero mean and covariance matrix $\mathbf{C}_v$ and is uncorrelated with $\boldsymbol{\theta}$, then the LMMSE of $\boldsymbol{\theta}$ is [58]

$$\hat{\boldsymbol{\theta}} = \mathbb{E}(\boldsymbol{\theta}) + \mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_v)^{-1}(\mathbf{y} - \mathbf{H}\mathbb{E}(\boldsymbol{\theta})). \tag{2.2}$$

The performance of the estimator is measured by the error $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ whose mean is zero and whose covariance matrix is

$$\mathbf{P} = \mathbb{E}_{\mathbf{y},\theta}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = (C_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} + \mathbf{H}^T\mathbf{C}_v^{-1}\mathbf{H})^{-1}. \tag{2.3}$$

Suppose the parameter $\boldsymbol{\theta}$ has a Gaussian prior and the sensing model (2.1) contains Gausian noise, the LMMSE given by (2.2) is optimal in the sense of minimum MSE. And the inverse of $\mathbf{P}$

is known as the Fisher information matrix under the linear Gaussian measurement model

$$\mathbf{J} = \mathbf{P}^{-1} = C_{\boldsymbol{\theta\theta}}^{-1} + \mathbf{H}^T \mathbf{C}_v^{-1} \mathbf{H}.$$

2.5 Preliminaries: convex and nonconvex optimization

Epigraph

The graph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as [59] $\{(\mathbf{x}, f(\mathbf{x})) \mid \mathbf{x} \in \mathrm{dom}\, f\}$, which is a subset of $\mathbf{R}^{n+1}$. The epigraph of a function f is defined as

$$\mathrm{epi}\, f = \{(\mathbf{x}, t) \mid \mathbf{x} \in \mathrm{dom}\, f,\; f(\mathbf{x}) \le t\}.$$

Here 'Epi' means 'above' so epigraph means 'above the graph'.

The link between convex sets and convex functions is via the epigraph: A convex function is convex if and only if its epigraph is a convex set.

Indicator function of a constrained set

Let C be a convex/nonconvex set, and consider the function $\mathcal{I}_C$ with domain C and $\mathcal{I}_C(\mathbf{x}) = 0$ for all $\mathbf{x} \in C$. Its extended-value version is given by

$$\mathcal{I}_C(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in C \\ \infty & \mathbf{x} \notin C, \end{cases}$$

which is called the indicator function of the set C.

Semidefinite programming

A standard form semidefinite program (SDP) has linear equality constraints, and a matrix generalized inequality defined on a positive semidefinite cone,

$$
\begin{aligned}
\text{minimize} \quad & \operatorname{tr}(\mathbf{CX}) \\
\text{subject to} \quad & \operatorname{tr}(\mathbf{A}_i \mathbf{X}) = b_i, \ i = 1, 2, \dots, p \\
& \mathbf{X} \succeq 0,
\end{aligned}
$$

where $\mathbf{C}$, $\{\mathbf{A}_i\}$ and $\{b_i\}$ are known problem data, and $\succeq$ denotes a generalized inequality with respect to a positive semidefinite cone, $\mathbf{X} \succeq \mathbf{Y}$, which means that $\mathbf{Y} - \mathbf{X}$ is positive semidefinite.

Alternating direction method of multipliers

ADMM is an operator splitting method that solves convex problems of the form [60]

$$
\begin{aligned}
\text{minimize} \quad & f(\mathbf{x}) + g(\mathbf{y}) \\
\text{subject to} \quad & \mathbf{Ax} + \mathbf{By} = \mathbf{c},
\end{aligned}
\tag{2.4}
$$

where f and g may be nonsmooth or indicator functions, and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{c}$ are known coefficients.

The standard ADMM algorithm at the kth iteration is

$$
\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \quad \mathcal{L}(\mathbf{x}, \mathbf{y}^k, \boldsymbol{\mu}^k)
\tag{2.5}
$$

$$
\mathbf{y}^{k+1} = \arg\min_{\mathbf{y}} \quad \mathcal{L}(\mathbf{x}^{k+1}, \mathbf{y}, \boldsymbol{\mu}^k)
\tag{2.6}
$$

$$
\boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^k + \rho(\mathbf{x}^{k+1} - \mathbf{y}^{k+1}),
\tag{2.7}
$$

where $\rho > 0$ is a step size parameter, $\boldsymbol{\mu}$ is the dual variable associated with the constraint of problem (2.4), and $\mathcal{L}$ is the augmented Lagrangian function corresponding to problem (2.4)

$$
\mathcal{L}(\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}) = f(\mathbf{x}) + g(\mathbf{y}) + \boldsymbol{\mu}^T(\mathbf{Ax} + \mathbf{By} - \mathbf{c}) + (\rho/2)\|\mathbf{Ax} + \mathbf{By} - \mathbf{c}\|_2^2.
$$

The initial points $\mathbf{y}^0$ and $\boldsymbol{\mu}^0$ can be arbitrarily chosen, and under some mild conditions [60, Sec. 3.2], ADMM converges to the optimal solution of problem (2.4).

CHAPTER 3

SENSOR SELECTION WITH CORRELATED

MEASUREMENT NOISE

3.1 Introduction

In this chapter, we consider the problem of sensor selection for parameter estimation with correlated measurement noise. We seek optimal sensor activations by formulating an optimization problem, in which the estimation error, given by the trace of the inverse of the Bayesian Fisher information matrix, is minimized subject to energy constraints. We derive the closed form of the Fisher information matrix with respect to sensor selection variables that is valid for any arbitrary noise correlation regime, and develop both a convex relaxation approach and a greedy algorithm to find near-optimal solutions. We demonstrate that the commonly used formulations [19, 34, 35] for sensor selection are valid only when measurement noises are weakly correlated. In this scenario, we show that maximization of the trace of the Fisher information matrix used in [19] is equivalent to the problem of maximizing a convex quadratic function over a bounded polyhedron. The resutling problem structure enables the use of optimization methods with reduced computational complexity. We further extend our framework of sensor selection to solve the problem of sensor scheduling, where a greedy algorithm is proposed to determine non-myopic (multi-time step

ahead) sensor schedules.

The rest of the chapter is organized as follows. In Section 3.2, we formulate the problem of sensor selection with correlated noise. In Section 3.3, we present a convex relaxation approach and a greedy algorithm to solve the problem of sensor selection with an arbitrary noise correlation matrix. In Section 3.4, we reveal drawbacks of the existing formulations for sensor selection with correlated noise, and demonstrate their validity in only the weak noise correlation regime. In Section 3.5, we extend our framework to solve the problem of non-myopic sensor scheduling. In Section 3.6, we provide numerical results to illustrate the effectiveness of our approach, and to reveal the effect of noise correlation on estimation performance. We summarize our work in Section 3.7,

3.2   Problem statement

We wish to estimate a random vector $\mathbf{x} \in \mathbb{R}^n$ with a Gaussian prior probability density function (PDF) $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Observations of $\mathbf{x}$ from m sensors are corrupted by correlated measurement noise. To strike a balance between estimation accuracy and sensor activations, we formulate the problem of sensor selection, where the estimation error is minimized subject to a constraint on the total number of sensor activations.

Consider a linear system

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{3.1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector whose mth entry corresponds to a scalar observation from the mth sensor, $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the observation matrix, and $\mathbf{v} \in \mathbb{R}^m$ is the measurement noise vector that follows a Gaussian distribution with zero mean and an invertible covariance matrix $\mathbf{R}$. We assume that $\mathbf{x}$ and $\mathbf{v}$ are mutually independent random variables, and the noise covariance matrix is positive definite and thus invertible. We note that the noise covariance matrix is not restricted to being diagonal, so that the measurement noise could be *correlated* among the sensors.

The task of sensor selection is to determine the best subset of sensors to activate in order to minimize the estimation error, subject to a constraint on the number of activations. We introduce a sensor selection vector to represent the activation scheme

$$\mathbf{w} = [w_1, w_2, \ldots, w_m]^T, \ \ w_i \in \{0, 1\}, \tag{3.2}$$

where w_i indicates whether or not the ith sensor is selected. For example, if the ith sensor reports a measurement then $w_i = 1$, otherwise $w_i = 0$. In other words, the *active* sensor measurements can be compactly expressed as

$$\mathbf{y}_w = \mathbf{\Phi}_w \mathbf{y} = \mathbf{\Phi}_w \mathbf{H} \mathbf{x} + \mathbf{\Phi}_w \mathbf{v}, \tag{3.3}$$

where $\mathbf{y}_w \in \mathbb{R}^{\|\mathbf{w}\|_1}$ is the vector of measurements of selected sensors, $\|\mathbf{w}\|_1$ is the ℓ_1-norm of $\mathbf{w}$ which yields the total number of sensor activations, $\mathbf{\Phi}_w \in \{0, 1\}^{\|\mathbf{w}\|_1 \times m}$ is a submatrix of $\mathrm{diag}(\mathbf{w})$ after all rows corresponding to the unselected sensors have been removed, and $\mathrm{diag}(\mathbf{w})$ is a diagonal matrix whose diagonal entries are given by $\mathbf{w}$. Note that $\mathbf{\Phi}_w$ and $\mathbf{w}$ are linked as below

$$\mathbf{\Phi}_w \mathbf{\Phi}_w^T = \mathbf{I}_w \ \ \text{and} \ \ \mathbf{\Phi}_w^T \mathbf{\Phi}_w = \mathrm{diag}(\mathbf{w}), \tag{3.4}$$

where $\mathbf{I}_w$ denotes an identity matrix with dimension $\|\mathbf{w}\|_1$.

We employ the LMMSE to estimate the unknown parameter under the Bayesian setup. It is worth mentioning that the use of the Bayesian estimation framework ensures the validity of parameter estimation for an underdetermined system, in which the number of selected sensors is less than the dimension of the parameter to be estimated.

To evaluate the estimation performance, we consider the Bayesian Fisher information matrix $\mathbf{J}_w$ for a linear Gaussian measurement model with a Gaussian prior distribution (see Chapter 2.4)

$$\mathbf{J}_w = \mathbf{\Sigma}^{-1} + \mathbf{H}^T \mathbf{\Phi}_w^T \mathbf{R}_w^{-1} \mathbf{\Phi}_w \mathbf{H}, \quad \mathbf{R}_w = \mathbf{\Phi}_w \mathbf{R} \mathbf{\Phi}_w^T, \tag{3.5}$$

where the second term is related to the sensor selection scheme. It is clear from (3.5) that the dependence of $\mathbf{J}_w$ on $\mathbf{w}$ is through $\boldsymbol{\Phi}_w$. This dependency does not lend itself to easy optimization of scalar-valued functions of $\mathbf{J}_w$ with respect to $\mathbf{w}$. In what follows, we will rewrite $\mathbf{J}_w$ as an explicit function of the selection vector $\mathbf{w}$.

Closed form of Fisher information for sensor selection

The key idea of expressing (3.5) as an explicit function of $\mathbf{w}$ is to replace $\boldsymbol{\Phi}_w$ with $\mathbf{w}$ based on their relationship given by (3.4). Consider a decomposition of the noise covariance matrix [61]

$$\mathbf{R} = a\mathbf{I} + \mathbf{S}, \tag{3.6}$$

where a positive scalar a is chosen such that the matrix $\mathbf{S}$ is positive definite, and $\mathbf{I}$ is the identity matrix. We remark that the decomposition given in (3.6) is readily obtained through an eigenvalue decomposition of the positive definite matrix $\mathbf{R}$, and it helps us in deriving the closed form of the Fisher information matrix with respect to $\mathbf{w}$.

According to (3.6), we rewrite a part of the second term on the right hand side of (3.5) as

$$
\begin{aligned}
\boldsymbol{\Phi}_w^T \mathbf{R}_w^{-1} \boldsymbol{\Phi}_w &= \boldsymbol{\Phi}_w^T (a\mathbf{I}_w + \boldsymbol{\Phi}_w \mathbf{S} \boldsymbol{\Phi}_w^T)^{-1} \boldsymbol{\Phi}_w \\
&\overset{(1)}{=} \mathbf{S}^{-1} - \mathbf{S}^{-1}(\mathbf{S}^{-1} + a^{-1}\boldsymbol{\Phi}_w^T\boldsymbol{\Phi}_w)^{-1}\mathbf{S}^{-1} \\
&\overset{(2)}{=} \mathbf{S}^{-1} - \mathbf{S}^{-1}(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w}))^{-1}\mathbf{S}^{-1},
\end{aligned}
\tag{3.7}
$$

where step (1) is obtained from the matrix inversion lemma[1], and step (2) holds due to (3.4).

Substituting (3.7) into (3.5), the Fisher information matrix can be expressed as

$$\mathbf{J}_w = \boldsymbol{\Sigma}^{-1} + \mathbf{H}^T\mathbf{S}^{-1}\mathbf{H} - \mathbf{H}^T\mathbf{S}^{-1}(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w}))^{-1}\mathbf{S}^{-1}\mathbf{H}. \tag{3.8}$$

---

[1]For appropriate matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$, the matrix inversion lemma states that $(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$, which yields $\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D} = \mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1}\mathbf{A}$.

It is clear from (3.8) that the decomposition of $\mathbf{R}$ in (3.6) allows us to isolate the dependence of $\mathbf{J}_w$ on $\mathbf{w}$.

Formulation of optimal sensor selection

We now state the main optimization problem

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & \text{tr}(\mathbf{J}_w^{-1}) \\
\text{subject to} \quad & \mathbf{1}^T \mathbf{w} \leq s, \ \mathbf{w} \in \{0, 1\}^m,
\end{aligned}
\tag{P0}
$$

where $\mathbf{J}_w \in \mathbb{R}^n$ is given by (3.8), and $s \leq m$ is a prescribed energy budget given by the maximum number of sensors to be activated. (P0) is a nonconvex optimization problem due to the presence of Boolean selection variables. In what follows, we discuss two special cases for the formulations of the sensor selection problem under two different structures of the noise covariance matrix $\mathbf{R}$: a) $\mathbf{R}$ is diagonal, and b) $\mathbf{R}$ has small off-diagonal entries.

When measurement noises are uncorrelated, the noise covariance matrix $\mathbf{R}$ becomes diagonal. From (3.4) and (3.5), the Fisher information matrix in the objective function of (P0) simplifies to

$$
\mathbf{J}_w = \mathbf{\Sigma}^{-1} + \mathbf{H}^T \mathbf{\Phi}_w^T \mathbf{\Phi}_w \mathbf{R}^{-1} \mathbf{\Phi}_w^T \mathbf{\Phi}_w \mathbf{H} = \mathbf{\Sigma}^{-1} + \sum_{i=1}^{m} w_i R_{ii}^{-1} \mathbf{h}_i \mathbf{h}_i^T,
\tag{3.9}
$$

where $\mathbf{h}_i^T$ denotes the ith row of $\mathbf{H}$, R_{ii} denotes the ith diagonal entry of $\mathbf{R}$. It is clear from (3.9) that each sensor contributes to Fisher information in an additive manner. As demonstrated in [16] and [15], the linearity of the inverse mean squared error (Fisher information) with respect to $\mathbf{w}$ enables the use of convex optimization to solve the problem of sensor selection.

When measurement noises are weakly correlated (namely, $\mathbf{R}$ has small off-diagonal entries), it will be shown in Sec. 3.4 that the Fisher information matrix can be approximately expressed as

$$
\hat{\mathbf{J}}_w := \mathbf{\Sigma}^{-1} + \mathbf{H}^T (\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1}) \mathbf{H},
\tag{3.10}
$$

where $\circ$ stands for the Hadamard (elementwise) product. The problem of sensor selection with weakly correlated noise becomes

$$\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & \text{tr} \left(\mathbf{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})\mathbf{H} \right)^{-1} \\
\text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \ \mathbf{w} \in \{0,1\}^m.
\end{aligned} \quad \text{(P1)}$$

Compared to the generalized formulation (P0), the objective function of (P1) is convex with respect to the rank-one matrix $\mathbf{w}\mathbf{w}^T$. Such structure introduces computational benefits while solving (P1). We emphasize that (P1) has been formulated in [19, 34, 35] for sensor selection with correlated noise, however, using this formulation, without acknowledging that it is only valid when the correlation is weak, can lead to incorrect sensor selection results. We elaborate on the problem of sensor selection with weakly correlated noise in Section 3.4.

3.3 General case: proposed optimization methods for sensor selection

In this section, we present two methods to solve (P0): the first is based on convex relaxation techniques, and the second is based on a greedy algorithm. First, we show that after relaxing the Boolean constraints the selection problem can be cast as a standard semidefinite program (SDP). Given the solution of the relaxed (P0) we then use the randomization method to generate a near-optimal selection scheme. Next, we show that given a subset of sensors, activating a new sensor always improves the estimation performance. Motivated by this, we present a greedy algorithm that scales gracefully with the problem size to obtain locally optimal solutions of (P0).

Convex relaxation

Substituting the expression of Fisher information (3.8) into problem (P0), we obtain

$$
\begin{aligned}
& \underset{\mathbf{w}}{\text{minimize}} && \operatorname{tr}\left(\mathbf{C} - \mathbf{B}^T\left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1}\mathbf{B}\right)^{-1} \\
& \text{subject to} && \mathbf{1}^T\mathbf{w} \leq s, \ \mathbf{w} \in \{0,1\}^m,
\end{aligned}
\tag{3.11}
$$

where for notational simplicity we have defined $\mathbf{C} := \mathbf{\Sigma}^{-1} + \mathbf{H}^T\mathbf{S}^{-1}\mathbf{H}$ and $\mathbf{B} := \mathbf{S}^{-1}\mathbf{H}$.

In Proposition 3.1, we show that problem (3.11) can be further converted into an SDP with a (nonconvex) rank-one constraint.

Proposition 3.1. *Problem* (3.11) *is equivalent to*

$$
\begin{aligned}
& \underset{\mathbf{w},\mathbf{W},\mathbf{Z},\mathbf{V}}{\text{minimize}} && \operatorname{tr}(\mathbf{Z}) \\
& \text{subject to} && \begin{bmatrix} \mathbf{C} - \mathbf{V} & \mathbf{I} \\ \mathbf{I} & \mathbf{Z} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \mathbf{V} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w}) \end{bmatrix} \succeq 0 \\
& && \operatorname{tr}(\mathbf{W}) \leq s, \ \operatorname{diag}(\mathbf{W}) = \mathbf{w} \\
& && \mathbf{W} = \mathbf{w}\mathbf{w}^T,
\end{aligned}
\tag{3.12}
$$

where $\mathbf{w}, \mathbf{W} \in \mathbb{S}^m$, $\mathbf{Z} \in \mathbb{S}^n$, $\mathbf{V} \in \mathbb{S}^n$ *are optimization variables, and* $\mathbb{S}^n$ *represents the set of* $n \times n$ *symmetric matrices.*

Proof: See Appendix A.1. ∎

After relaxing the rank-one constraint of problem (3.12) to $\mathbf{W} \succeq \mathbf{w}\mathbf{w}^T$, we reach the SDP

$$
\begin{aligned}
& \underset{\mathbf{w},\mathbf{W},\mathbf{Z},\mathbf{V}}{\text{minimize}} && \operatorname{tr}(\mathbf{Z}) \\
& \text{subject to} && \text{constraints of (3.12) except } \mathbf{W} = \mathbf{w}\mathbf{w}^T \\
& && \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq 0,
\end{aligned}
\tag{3.13}
$$

where the last inequality is derived through the application of a Schur complement to $\mathbf{W} \succeq \mathbf{w}\mathbf{w}^T$.

We can use an interior-point algorithm to solve the SDP (3.13). In practice, if the dimension of the unknown parameter vector is much less than the number of sensors, the computational complexity of SDP is roughly given by $O(m^{4.5})$ [62]. Once the SDP (3.13) is solved, we employ a randomization method to generate a near-optimal sensor selection scheme, where the effectiveness of the randomization method has been shown in our extensive numerical experiments. We refer the readers to [63] for more details on the motivation and benefits of randomization used in SDR. The aforementioned procedure is summarized in Algorithm 3.1, which includes the randomization procedure described in Algorithm 3.2.

Algorithm 3.1 SDR with randomization for sensor selection

Require: prior information Σ, $\mathbf{R} = a\mathbf{I} + \mathbf{S}$ as in (3.6), observation matrix $\mathbf{H}$ and energy budget s
 1: solve the SDP (3.13) and obtain solution $(\mathbf{w}, \mathbf{W})$
 2: call Algorithm 3.2 for Boolean solution.

Algorithm 3.2 Randomization method [63]

Require: solution pair $(\mathbf{w}, \mathbf{W})$ from the SDP (3.13)
 1: **for** $l = 1, 2, \ldots, N$ **do**
 2:     pick a random number $\boldsymbol{\xi}^{(l)} \sim \mathcal{N}(\mathbf{w}, \mathbf{W} - \mathbf{w}\mathbf{w}^T)$
 3:     map $\boldsymbol{\xi}^{(l)}$ to a sub-optimal sensor selection scheme $\mathbf{w}^{(l)}$

$$w_j^{(l)} = \begin{cases} 1 & \xi_j^{(l)} \geq [\boldsymbol{\xi}^{(l)}]_s \\ 0 & \text{otherwise,} \end{cases} \quad j = 1, 2, \ldots, m,$$

    where $w_j^{(l)}$ is the jth element of $\mathbf{w}^{(l)}$, and $[\boldsymbol{\xi}^{(l)}]_s$ denotes the sth largest entry of $\boldsymbol{\xi}^{(l)}$
 4: **end for**
 5: choose a vector in $\{\mathbf{w}^{(l)}\}_{l=1}^N$ which yields the smallest objective value of (3.11).

Greedy algorithm

We begin by showing in Proposition 3.2 that even in the presence of correlated measurement noise, the Fisher information increases if an inactive sensor is made active.

Proposition 3.2. *If $\mathbf{w}$ and $\tilde{\mathbf{w}}$ represent two sensor selection vectors, where $w_i = \tilde{w}_i$ for $i \in \{1, 2, \ldots, m\} \setminus \{j\}$, $w_j = 0$ and $\tilde{w}_j = 1$, then the resulting Fisher information matrix satisfies*

$\mathbf{J}_{\tilde{w}} \succeq \mathbf{J}_w$. *More precisely,*

$$\mathbf{J}_{\tilde{w}} - \mathbf{J}_w = c_j \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^T, \quad \mathrm{tr}(\mathbf{J}_w^{-1}) - \mathrm{tr}(\mathbf{J}_{\tilde{w}}^{-1}) = \frac{c_j \boldsymbol{\alpha}_j^T \mathbf{J}_w^{-2} \boldsymbol{\alpha}_j}{1 + c_j \boldsymbol{\alpha}_j \mathbf{J}_w^{-1} \boldsymbol{\alpha}_j} \geq 0, \tag{3.14}$$

where

$$c_j = \begin{cases} R_{jj}^{-1} & \mathbf{w} = \mathbf{0} \\ (R_{jj} - \mathbf{r}_j^T \mathbf{R}_w^{-1} \mathbf{r}_j)^{-1} & otherwise, \end{cases} \qquad \boldsymbol{\alpha}_j = \begin{cases} \mathbf{h}_j & \mathbf{w} = \mathbf{0} \\ \mathbf{H}^T \boldsymbol{\Phi}_w^T \mathbf{R}_w^{-1} \mathbf{r}_j - \mathbf{h}_j & otherwise, \end{cases}$$

R_{jj} *is the jth diagonal entry of* $\mathbf{R}$, $\mathbf{r}_j$ *represents the covariance vector between the measurement noise of the jth sensor and that of the active sensors in* $\mathbf{w}$, $c_j > 0$, $\mathbf{h}_j^T$ *is the jth row of* $\mathbf{H}$, $\boldsymbol{\Phi}_w$ *and* $\mathbf{R}_w$ *are given by* (3.3) *and* (3.5), *respectively.*

Proof: See Appendix A.2. ∎

It is clear from (3.14) that when an inactive sensor is made active, the increase in Fisher information leads to an information gain in terms of the rank-one matrix. Since activating a new sensor does not degrade the estimation performance, the inequality (energy) constraint in (P0) can be reformulated as an equality constraint.

In a greedy algorithm, we iteratively select a new sensor which gives the largest performance improvement until the energy constraint is satisfied with equality. The greedy algorithm is attractive due to its simplicity, and has been employed in a variety of applications [11, 64, 65]. In particular, a greedy algorithm was proposed in [65] for sensor selection under the assumption of uncorrelated measurement noise. We generalize the framework of [65] by taking into account noise correlation. Clearly, in each iteration of the greedy algorithm, the newly activated sensor is the one that maximizes the performance improvement characterized by $\mathrm{tr}(\mathbf{J}_w^{-1}) - \mathrm{tr}(\mathbf{J}_{\tilde{w}}^{-1})$ in (3.14). We summarize the greedy algorithm in Algorithm 3.3.

In Step 2 of Algorithm 3.3, we search $O(m)$ sensors to achieve the largest performance improvement. In (3.14), the computation of $\mathbf{J}_w^{-1}$ incurs a complexity of $O(n^{2.373})$ [66]. Since Algorithm 3.3 terminates after s iterations, its overall complexity is given by $O(sm + sn^{2.373})$, where at

---

Algorithm 3.3 Greedy algorithm for sensor selection

Require: $\mathbf{w} = \mathbf{0}, \mathcal{I} = \{1, 2, \ldots, m\}$ and $\mathbf{J}_w = \mathbf{\Sigma}^{-1}$
1: **for** $l = 1, 2, \ldots, s$ **do**
2:      given $\mathbf{w}$, activate sensor $j \in \mathcal{I}$ such that $\mathrm{tr}(\mathbf{J}_w^{-1}) - \mathrm{tr}(\mathbf{J}_{\tilde{w}}^{-1})$ in (3.14) is maximized
3:      update $\mathbf{w}$ by setting $w_j = 1$, and update $\mathbf{J}_w$
4:      remove j from $\mathcal{I}$.
5: **end for**

---

each iteration, the calculation of $\mathbf{J}_w^{-1}$ is independent of the search for the new active sensor. If the dimension of $\mathbf{x}$ is much less than the number of sensors, the complexity of Algorithm 3.3 reduces to $O(sm)$. Our extensive numerical experiments show that the greedy algorithm is able to yield good locally optimal sensor selection schemes.

3.4   Special case: sensor selection with weak noise correlation

In this section, we establish that in contrast to the approach proposed in this thesis, the existing sensor selection model in [19,34,35] is only valid when measurement noises are weakly correlated. In this scenario, the proposed sensor selection problem given by (P0) would simplify to (P1). Moreover, if the trace of the Fisher information matrix (also known as information gain defined in [19]) is adopted as the performance measure for sensor selection, we show that the resulting optimization problem can be cast as a special problem of maximizing a convex quadratic function over a bounded polyhedron.

Validity of existing formulation: weak correlation

We consider the scenario of weakly correlated noise, in which the noise covariance matrix $\mathbf{R}$ has small off-diagonal entries, namely, noises are weakly correlated across the sensors. For ease of

representation, we express the noise covariance matrix as

$$\mathbf{R} = \mathbf{\Lambda} + \epsilon \mathbf{\Upsilon}, \tag{3.15}$$

where $\mathbf{\Lambda}$ is a diagonal matrix which consists of the diagonal entries of $\mathbf{R}$, $\epsilon \mathbf{\Upsilon}$ is a symmetric matrix whose diagonal entries are zero and off-diagonal entries correspond to those of $\mathbf{R}$, the parameter ϵ is introduced to govern the strength of noise correlation across the sensors, and $\mathbf{\Lambda}$ and $\mathbf{\Upsilon}$ are independent of ϵ. Clearly, the covariance of weakly correlated noises can be described by (3.15) for some small value of ϵ since $\mathbf{\Upsilon}$ is ϵ-independent. As $\epsilon \to 0$, the off-diagonal entries of $\mathbf{R}$ are forced to go to zero.

Proposition 3.3 below shows that the correct expression (3.5) of Fisher information is equal to the expression (3.10), as presented in [19, 34, 35], up to first order in ϵ as $\epsilon \to 0$.

Proposition 3.3. *If measurement noises are weakly correlated and* $\mathbf{R} = \mathbf{\Lambda} + \epsilon \mathbf{\Upsilon}$*, then the Fisher information matrix (3.5) can be expressed as*

$$\mathbf{J}_w = \hat{\mathbf{J}}_w + O(\epsilon^2) \quad \text{as } \epsilon \to 0,$$

where $\hat{\mathbf{J}}_w$ *is given by (3.10).*

Proof: See Appendix A.3. ■

It is clear from Proposition 3.3 that (P1) is valid only when the noise correlation is weak. Proceeding with the same logic as in Proposition 3.1 for the problem reformulation, we can relax (P1) to the SDP

$$
\begin{aligned}
\underset{\mathbf{w}, \mathbf{W}, \mathbf{Z}}{\text{minimize}} \quad & \text{tr}(\mathbf{Z}) \\
\text{subject to} \quad & \begin{bmatrix} \mathbf{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{W} \circ \mathbf{R}^{-1})\mathbf{H} & \mathbf{I} \\ \mathbf{I} & \mathbf{Z} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq 0 \\
& \text{tr}(\mathbf{W}) \le s, \ \text{diag}(\mathbf{W}) = \mathbf{w},
\end{aligned} \tag{3.16}
$$

where $\mathbf{Z} \in \mathbb{S}^n$ is an auxiliary optimization variable. Given the solution pair $(\mathbf{w}, \mathbf{W})$ of problem (3.16), we can use the randomization method in Algorithm 3.2 to construct a near-optimal sensor selection scheme. The computational complexity of solving problem (3.16) is close to that of solving the SDP (3.13). However, as will be evident later, the sensor selection problem with weakly correlated noise can be further simplified if the trace of the Fisher information matrix is used as the performance measure. In this scenario, the obtained problem structure enables the use of more computationally inexpensive algorithms, e.g., bilinear programing, to solve the sensor selection problem.

Sensor selection by maximizing trace of Fisher information

Instead of minimizing the estimation error, the trace of Fisher information (so-called T-optimality [67]) also has been used as a performance metric in problems of sensor selection [19, 22, 55]. According to [68, Lemma 1], the trace of Fisher information constitutes a lower bound to the trace of error covariance matrix given by $\mathbf{J}_w^{-1}$ in (3.5). That is, $\mathrm{tr}(\mathbf{J}_w^{-1}) \geq \frac{n^2}{\mathrm{tr}(\mathbf{J}_w)}$. Motivated by this, we propose to minimize the lower bound of the objective function in (P1), which leads to the problem [19]

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{maximize}} \quad & \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})\mathbf{H}\right) \\
\text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \ \mathbf{w} \in \{0,1\}^m.
\end{aligned}
\tag{P2}
$$

It is worth mentioning that the sensor selection scheme obtained from (P2) may not be optimal in the MMSE sense. However, the trace operator is linear and introduces computational benefits in optimization. Reference [19] has shown that (P2) is not convex even if Boolean selection variables are relaxed. However, there is no theoretical justification and analysis provided in [19] on the problem structure. In what follows, we demonstrate that the Boolean constraint in (P2) can be replaced by its convex hull $\mathbf{w} \in [0,1]^m$ without loss of performance, to obtain an equivalent optimization problem.

Proposition 3.4. (P2) *is equivalent to*

$$\underset{\mathbf{w}}{maximize} \quad \mathbf{w}^T \mathbf{\Omega} \mathbf{w}$$

$$subject\ to \quad \mathbf{1}^T \mathbf{w} \leq s, \ \mathbf{w} \in [0, 1]^m, \tag{3.17}$$

where $\mathbf{\Omega}$ is a positive semidefinite matrix given by $\mathbf{A}(\mathbf{R}^{-1} \otimes \mathbf{I}_n)\mathbf{A}^T$, $\otimes$ denotes the Kronecker product, $\mathbf{A} \in \mathbb{R}^{m \times mn}$ is a block-diagonal matrix whose diagonal blocks are given by $\{\mathbf{h}_i^T\}_{i=1}^m$, and $\mathbf{h}_i^T$ denotes the ith row of the measurement matrix $\mathbf{H}$.

Proof: See Appendix A.4. ∎

It is clear from Proposition 3.4 that (P2) eventually approaches the problem of maximizing a convex quadratic function over a bounded polyhedron. It is known [69] that finding a globally optimal solution of (3.17) is NP-hard. Therefore, we resort to local optimization methods, such as bilinear programming and SDR, to solve problem (3.17). To be specific, bilinear programming is a special case of alternating convex optimization, where at each iteration we solve two linear programs. Since bilinear programming is based on linear programming, it scales gracefully with problem size but with a possibility of only finding local optima. If we rewrite the constraints of problem (3.17) as quadratic forms in $\mathbf{w}$, (P2) can be further transformed into a nonconvex *homogeneous* quadratically constrained quadratic program (QCQP), which refers to a QCQP without involving linear terms of optimization variables. In this scenario, SDR can be applied to solve the problem. Compared to the application of SDR in (3.16), the homogeneous QCQP leads to an SDP with a smaller problem size. We refer the readers to [56, Sec. V] and [19, Sec. V] for more details on the application of bilinear programming and SDR.

3.5 Sensor scheduling for parameter tracking

In this section, we extend the sensor selection framework with correlated noise to the problem of non-myopic sensor scheduling, which determines sensor activations for multiple future time steps. Since the Fisher information matrices at consecutive time steps are coupled with each other,

expressing them in a closed form with respect to the sensor selection variables becomes intractable. Therefore, we employ a greedy algorithm to seek locally optimal solutions of the non-myopic sensor scheduling problem.

Consider a discrete-time dynamical system

$$\mathbf{x}_{t+1} = \mathbf{F}_t\mathbf{x}_t + \mathbf{u}_t \tag{3.18}$$

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t, \tag{3.19}$$

where $\mathbf{x}_t \in \mathbb{R}^n$ is the target state at time t, $\mathbf{y}_t \in \mathbb{R}^m$ is the measurement vector whose ith entry corresponds to a scalar observation from the ith sensor at time t, $\mathbf{F}_t$ is the state transition matrix from time t to time $t+1$, and $\mathbf{H}_t$ denotes the observation matrix at time t. The inputs $\mathbf{u}_t$ and $\mathbf{v}_t$ are white, Gaussian, zero-mean random vectors with covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively. We note that the covariance matrix $\mathbf{R}$ may not be diagonal, since the noises experienced by different sensors could be spatially correlated. We also remark that although the dynamical system (3.18)-(3.19) is assumed to be linear, it will be evident later that the proposed sensor scheduling framework is also applicable to non-linear dynamical systems.

The PDF of the initial state $\mathbf{x}_0$ at time step t_0 is assumed to be Gaussian with mean $\hat{\mathbf{x}}_0$ and covariance matrix $\hat{\mathbf{P}}_0$, where $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{P}}_0$ are estimates of the initial state and error covariance from the previous measurements obtained using filtering algorithms, such as a particle filter or a Kalman filter [70,71]. At time step t_0, we aim to find the optimal sensor schedule over the next τ time steps $t_0 + 1, t_0 + 2, \ldots, t_0 + \tau$. Hereafter, for notational simplicity, we assume $t_0 = 0$. The sensor schedule can be represented by a vector of binary variables

$$\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_\tau^T]^T \in \{0, 1\}^{\tau m}, \tag{3.20}$$

where $\mathbf{w}_t = [w_{t,1}, w_{t,2}, \ldots, w_{t,m}]^T$ characterizes the sensor schedule at time $1 \leq t \leq \tau$. In what follows, we assume that $\tau > 1$. If $\tau = 1$, the non-myopic sensor scheduling problem reduces to the sensor selection problem for one snapshot or the so-called myopic scheduling problem. This

case has been studied in the previous sections.

In the context of state tracking [36, 72], the Fisher information matrix has the following recursive form

$$\mathbf{J}_t = (\mathbf{Q} + \mathbf{F}_{t-1}\mathbf{J}_{t-1}^{-1}\mathbf{F}_{t-1}^T)^{-1} + \mathbf{G}_t \tag{3.21}$$

$$\mathbf{G}_t = \mathbf{H}_t^T \mathbf{\Phi}_{w_t}^T (\mathbf{\Phi}_{w_t} \mathbf{R} \mathbf{\Phi}_{w_t}^T)^{-1} \mathbf{\Phi}_{w_t} \mathbf{H}_t, \tag{3.22}$$

for $t = 1, 2, \ldots, \tau$, where $\mathbf{J}_t$ denotes the Fisher information at time t, $\mathbf{G}_t$ denotes the part of Fisher information matrix which incorporates the updated measurement, and $\mathbf{\Phi}_{w_t}$ is a submatrix of $\mathrm{diag}(\mathbf{w}_t)$ where all the rows corresponding to the unselected sensors are removed. It is clear from (3.7) that the term involving $\mathbf{\Phi}_{w_t}$ in (3.22) can be further expressed as an explicit form with respect to $\mathbf{w}_t$.

Remark 1. *In case of non-linear measurement models, the term $\mathbf{G}_t$ in the Fisher information matrix becomes*

$$\mathbf{G}_t = \mathbf{E}_{\mathbf{x}_t}[(\nabla_{\mathbf{x}_t^T}\mathbf{h})^T \mathbf{\Phi}_{w_t}^T (\mathbf{\Phi}_{w_t} \mathbf{R} \mathbf{\Phi}_{w_t}^T)^{-1} \mathbf{\Phi}_{w_t} (\nabla_{\mathbf{x}_t^T}\mathbf{h})],$$

where $\mathbf{h}(\cdot)$ is a nonlinear measurement function, and $\nabla_{\mathbf{x}_t^T}\mathbf{h}$ is the Jacobian matrix of $\mathbf{h}$ with respect to $\mathbf{x}_t$. In this equation, the expectation with respect to $\mathbf{x}_t$ is commonly calculated with the help of the prediction state $\hat{\mathbf{x}}_t := \mathbf{F}_{t-1}\mathbf{F}_{t-2}\cdots\mathbf{F}_0\hat{\mathbf{x}}_0$ [71, 73]. To be concrete, we approximate the PDF of $\mathbf{x}_t$ with $p(\mathbf{x}_t) = \delta(\mathbf{x}_t - \hat{\mathbf{x}}_t)$, where $\delta(\cdot)$ is a δ-function. The matrix $\mathbf{G}_t$ is then given by

$$\mathbf{G}_t = \hat{\mathbf{H}}_t^T \mathbf{\Phi}_{w_t}^T (\mathbf{\Phi}_{w_t} \mathbf{R} \mathbf{\Phi}_{w_t}^T)^{-1} \mathbf{\Phi}_{w_t} \hat{\mathbf{H}}_t, \tag{3.23}$$

where $\hat{\mathbf{H}}_t := \nabla_{\mathbf{x}_t^T}\mathbf{h}(\hat{\mathbf{x}}_t)$.

We note that the Fisher information matrices at consecutive time steps are coupled with each other due to the recursive structure in (3.21). Therefore, $\mathbf{J}_t$ is a function of all selection variables $\{\mathbf{w}_k\}_{k=1}^t$. The recursive structure makes the closed form of Fisher information intractable. This

is in sharp contrast with the problem of myopic sensor selection, where expressing the Fisher information matrix in a closed form is possible.

We now pose the non-myopic sensor scheduling problem

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{\tau} \sum_{t=1}^{\tau} \text{tr}(\mathbf{J}_t^{-1})$$

$$\text{subject to} \quad \mathbf{1}^T \mathbf{w} \leq s, \tag{3.24a}$$

$$\sum_{t=1}^{\tau} w_{t,i} \leq s_i, \ i = 1, 2, \ldots, m, \tag{3.24b}$$

$$\mathbf{w} \in \{0, 1\}^{m\tau},$$

where $\mathbf{J}_t$ is determined by (3.21)-(3.22), the *cumulative* energy constraint (3.24a) restricts the total number of activations for all sensors over the entire time horizon, and the *individual* energy constraint (3.24b) implies that the ith sensor can report at most s_i measurements over τ time steps.

To solve problem (3.24) in a numerically efficient manner, we employ a greedy algorithm that iteratively activates one sensor at a time until the energy constraints are satisfied with equality. The proposed greedy algorithm can be viewed as a generalization of Algorithm 3.3 by incorporating the length of the time horizon and individual energy constraints.

We elaborate on the greedy algorithm. In the initial step, we assume $\mathbf{w} = \mathbf{0}$ and split the set of indices of $\mathbf{w}$ into m subsets $\{\mathcal{I}_i\}_{i=1}^m$, where we use the entries of the set $\mathcal{I}_i$ to keep track of all the time instants at which the ith sensor is inactive. The set $\mathcal{I}_i$ is initially given by $\{i, i + m, \ldots, i + (\tau - 1)m\}$ for $i = 1, 2, \ldots, m$. There exists a one-to-one correspondence between an index $j \in \mathcal{I}_i$ and a time instant $t \in \{1, 2, \ldots, \tau\}$ at which the ith sensor can be scheduled, where $j = i + (t - 1)m$. At every iteration of the greedy optimization algorithm, we update $\mathcal{I}_i$ for $i = 1, 2, \ldots, m$ such that it only contains indices of zero entries of $\mathbf{w}$. The quantity $\tau - |\mathcal{I}_i|$ gives the number of times that the ith sensor has been used, where $|\cdot|$ denotes the cardinality of a set. The condition $\tau - |\mathcal{I}_i| \geq s_i$ indicates a violation of the individual energy constraint. Note that the union $\{\mathcal{I}_1 \cup \mathcal{I}_2 \cup \ldots \cup \mathcal{I}_m\}$ gives all the remaining time instants at which the sensors can be activated. We enumerate all the indices in the union to determine the index j^* such that the

objective function of (3.24) is minimized as $w_{j^*} = 1$. We summarize the greedy algorithm for non-myopic sensor scheduling in Algorithm 3.4.

Algorithm 3.4 Greedy algorithm for sensor scheduling

Require: $\mathbf{w} = \mathbf{0}$ and $\mathcal{I}_i = \{i, i+m, \ldots, i+(\tau-1)m\}$ for $i = 1, 2, \ldots, m$

1: **for** $l = 1, 2, \ldots, \min\{s, \sum_{i=1}^{m} s_i\}$ **do**
2:     if $\tau - |\mathcal{I}_i| \geq s_i$, then replace $\mathcal{I}_i$ with an empty set for $i = 1, 2, \ldots, m$,
3:     enumerate indices of $\mathbf{w}$ in $\{\mathcal{I}_1 \cup \mathcal{I}_2 \cup \ldots \cup \mathcal{I}_m\}$ to select j^* such that the objective function of (3.24) is minimized when $w_j = 1$,
4:     remove j from $\mathcal{I}_{i^*}$, where i^* is given by the remainder of $\frac{j}{m}$ for $i^* \neq m$, and $i^* = m$ if the remainder is 0.
5: **end for**

The computational complexity of Algorithm 3.4 is dominated by Step 3. Specifically, we evaluate the objective function of (3.24) using $O(\tau m)$ operations. And the computation of the Fisher information matrix requires a complexity of $O(\tau m^{2.373})$, where $O(\tau)$ accounts for the number of recursions, and $O(m^{2.373})$ is the complexity of matrix inversion in (3.23) [66]. We emphasize that different from Proposition 3.2, expressing the closed form of the performance improvement in a greedy manner becomes intractable, since the Fisher information matrices are coupled with each other over the time horizon. Therefore, the computation cost of Algorithm 3.4 is given by $O(\tau^2 m^{3.373})$ per iteration.

For additional perspective, we compare the computational complexity of Algorithm 3.4 with the method in [18], where a reweighted ℓ_1 based quadratic programming (QP) was used to obtain locally optimal sensor schedules under linear (or linearized) dynamical systems with correlated noise. It was shown in [18] that the computational complexity of QP was ideally given by $O(m^{2.5} \tau^5)$ for every reweighting ℓ_1 iteration. We note that the computational complexity of the greedy algorithm increases slightly in terms of the network size by a factor $m^{0.873}$, while it decreases significantly in terms of the length of the time horizon by a factor τ^3.

3.6 Numerical results

In this section, we demonstrate the effectiveness of the proposed approach for sensor selection with correlated measurement noise. In our numerical examples, we assume that the sensors are randomly deployed in a square region, where each of them provides the measurement of an unknown parameter or state. For parameter estimation, we use the LMMSE to estimate the unknown parameter. For state tracking, we use the extended Kalman filter [58, Sec. 13] to track the target state.

Sensor selection for parameter estimation

We consider a network with $m \in \{20, 50\}$ sensors to estimate the vector of parameters $\mathbf{x} \in \mathbb{R}^n$ with $n = 2$, where sensors are randomly deployed over a 50×50 lattice. The prior PDF of $\mathbf{x}$ is given by $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = [10, 10]^T$ and $\boldsymbol{\Sigma} = \mathbf{I}$. For simplicity, the row vectors of the measurement matrix $\mathbf{H}$ are chosen randomly, and independently, from the distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}/\sqrt{n})$ [15]. The covariance matrix of the measurement noise is set by an exponential model [74]

$$R_{ij} = \mathrm{cov}(v_i, v_j) = \sigma_v^2 \, e^{-\varrho \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|_2}, \tag{3.25}$$

for $i, j = 1, 2, \ldots, m$, where $\sigma_v^2 = 1$, $\boldsymbol{\beta}_i \in \mathbb{R}^2$ is the location of the ith sensor in the 2D plane, $\|\cdot\|_2$ denotes the Euclidean norm, and ϱ is the correlation parameter which governs the strength of spatial correlation, namely, a larger (or smaller) ϱ corresponds to a weaker (or stronger) correlation. We choose $N = 100$ while performing the randomization method. Also, we employ an exhaustive search that enumerates all possible sensor selection schemes to obtain the globally optimal solution of (P0). The estimation performance is measured through the empirical MSE, which is averaged over 1000 numerical trials.

In Fig. 3.1, we present the MSE as a function of the energy budget by solving (P0) with correlation parameter $\varrho = 0.1$. In Fig. 3.1-(a) for the tractability of exhaustive search, we consider a small network with $m = 20$ sensors. We compare the performance of the proposed greedy algo-

rithm and SDR with randomization to that of SDR without randomization and exhaustive search. In particular, the right plots of Fig. 3.1-(a) show the performance gaps for the obtained locally optimal solutions compared to the globally optimal solutions resulting from an exhaustive search. We observe that the SDR method with randomization outperforms the greedy algorithm and yields optimal solutions. The randomization method also significantly improves the performance of SDR in sensor selection. This is not surprising, and our numerical observations agree with the literature [63, 75] that demonstrate the power and utility of randomization in SDR.

In Fig. 3.1-(b), we present the MSE as a function of the energy budget for a relatively large network ($m = 50$). Similar to the results of Fig. 3.1-(a), the SDR method with randomization yields the lowest estimation error. We also observe that the MSE ceases to decrease significantly when $s \geq 20$. This indicates that a subset of sensors is sufficient to provide satisfactory estimation performance, since the presence of correlation among sensors introduces information redundancy and makes observations less diverse.

In Fig. 3.2, we solve the problem of sensor selection with weak noise correlation ($\varrho = 0.5$), and present the MSE as a function of the energy budget $s \in \{2, 3, \ldots, 50\}$. We compare the performance of three optimization approaches: SDR with randomization for solving (P1), bilinear programming (BP) for solving (P2), and SDR with randomization for solving (P2). We recall that in (P1) the goal is to minimize the trace of the error covariance matrix and in (P2) the goal is to maximize the trace of Fisher information. As we can see, approaches that maximize the trace of Fisher information yield worse estimation performance than those that minimize the estimation error. This is because (P2) ignores the contribution of prior information Σ in sensor selection. We also note that although BP (a linear programming based approach) has the lowest computational complexity, it leads to the worst optimization performance.

In Fig. 3.3, we present the MSE as a function of the correlation parameter ϱ, where $m = 50$ and $s \in \{7, 13\}$. We consider sensor selection schemes by using SDR with randomization to solve problems (P0) and (P1), respectively. For comparison, we also present the estimation performance when all the sensors are selected. As shown in Fig. 3.3, we consider two correlation regimes: weak

correlation and strong correlation. We observe that in the weak correlation regime, solutions of both (P0) and (P1) yield the same estimation performance. For very large values of ϱ (say $\varrho > 2$ in this example), the MSE gets saturated since measurement noises tend to be uncorrelated. This leads to the same sensor selection scheme and the same estimation error even if we continue to increase ϱ. In the strong correlation regime, solutions of (P1) lead to worse estimation performance for sensor selection. We also observe that the sensitivity to the sensor selection strategy reduces if the strength of correlation becomes extremely large, e.g., $\varrho \leq 0.05$. More interestingly, the estimation performance is improved as the correlation becomes stronger. A similar result was reported in [76], where the presence of correlation was shown to significantly improve the estimation performance. This is because for strongly correlated noise, noise cancellation could be achieved by subtracting one observation from the other [76]. Further if we fix the value of ϱ, the estimation error decreases when the energy budget increases, and the performance gap between solutions of (P0) and (P1) reduces.

Sensor scheduling for state tracking

In this example, we track a target with $m = 30$ sensors over 30 time steps. We assume that the target state is a 4×1 vector $\mathbf{x}_t = [x_{t,1}, x_{t,2}, x_{t,3}, x_{t,4}]^T$, where $(x_{t,1}, x_{t,2})$ and $(x_{t,3}, x_{t,4})$ denote the target location and velocity at time step t. The state equation (3.18) follows a white noise acceleration model [71]

$$
\mathbf{F}_t = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{Q} = q \begin{bmatrix} \frac{\Delta^3}{3} & 0 & \frac{\Delta^2}{2} & 0 \\ 0 & \frac{\Delta^3}{3} & 0 & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & 0 & \Delta & 0 \\ 0 & \frac{\Delta^2}{2} & 0 & \Delta \end{bmatrix},
$$

where Δ and q denote the sampling interval and the process noise parameter, respectively. In our simulations, we set $\Delta = 1$ and $q = 0.01$. The prior PDF of the initial state is assumed to be Gaussian with mean $\hat{\mathbf{x}}_0 = [1, 1, 0.5, 0.5]^T$ and covariance $\hat{\boldsymbol{\Sigma}}_0 = \text{diag}(1, 1, 0.1, 0.1)$.

The measurement equation follows a power attenuation model [77],

$$h_i(\mathbf{x}_t) = \sqrt{\frac{P_0}{1 + (x_{t,1} - \beta_{i,1})^2 + (x_{t,2} - \beta_{i,2})^2}} \tag{3.26}$$

for $i = 1, 2, \ldots, m$, where $P_0 = 10^4$ is the signal power of the source, and the pair $(\beta_{i,1}, \beta_{i,2})$ is the position of the ith sensor. The covariance matrix of the measurement noise is given by (3.25) with $\varrho = 0.035$.

In the sensor scheduling problem (3.24), we assume $s = \sum_{i=1}^m s_i$ and $s_1 = s_2 = \cdots = s_m$. In order to implement the proposed greedy algorithm and the existing method in [18], the nonlinear measurement function (3.26) is linearized at the prediction state $\hat{\mathbf{x}}_t = \mathbf{F}_{t-1}\mathbf{F}_{t-2}\cdots\mathbf{F}_0\hat{\mathbf{x}}_0$ as suggested in Remark 1. We determine sensor schedules for every $\tau = 6$ future time steps, and then update the estimate of the target state based on the selected measurements via an extended Kalman filter [78]. The estimation performance is measured through the empirical MSE, which is obtained by averaging the estimation error over 30 time steps and 1000 simulation trials.

In Fig. 3.4, we present the MSE as a function of the individual energy budget. We compare the performance of our proposed greedy algorithm with that of the sensor scheduling method in [18]. We remark that the method in [18] relies on a reformulation of linearized dynamical systems and an ℓ_1 relaxation in optimization. In contrast, the proposed greedy algorithm is independent of the dynamical system models and convex relaxations. We observe that the greedy algorithm outperforms the method in [18]. This result together with the previous results in Fig. 3.1 and 3.2 implies that the greedy algorithm could yield satisfactory estimation performance. Sensor schedules at time steps $t = 10$ and 24 are shown in Fig. 3.5. We observe that some sensors closest to the target are selected due to their high signal power. However, from the entire network point of view, the active sensors tend to be spatially distributed rather than aggregating in a small neighborhood around the target. This is because observations from neighboring sensors are strongly correlated in space and may lead to information redundancy in target tracking.

3.7 Summary

In this chapter, we addressed the problem of sensor selection with correlated measurement noise. We showed that the commonly used formulation for sensor selection is only valid for the special case of weak noise correlation. By contrast, here we proposed a tractable sensor selection framework that is valid for an arbitrary noise correlation matrix, and presented a suite of efficient optimization algorithms. Numerical results were provided to illustrate the effectiveness of our approach and the impact of noise correlation on the performance of sensor selection. In the following chapter, we will study the periodic sensor scheduling problem in networks of dynamical systems.
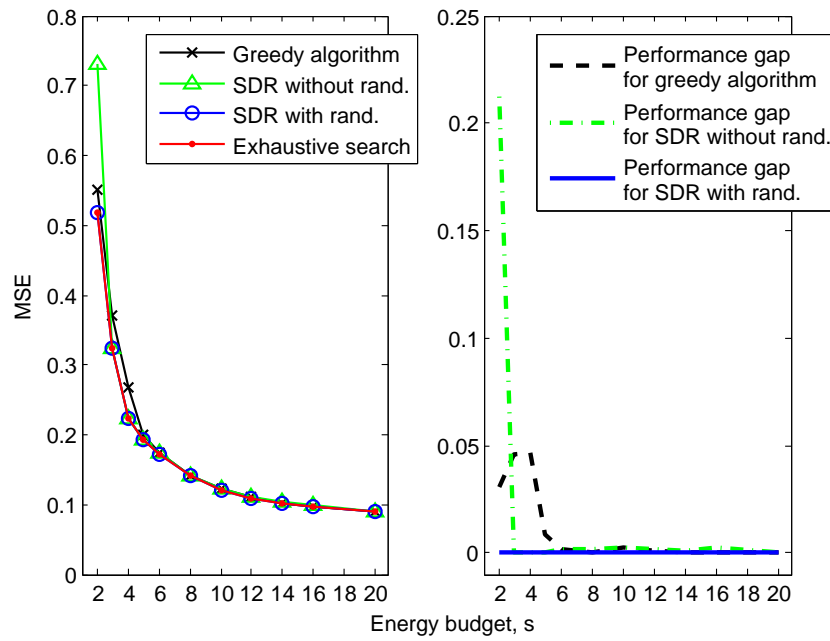
(a)
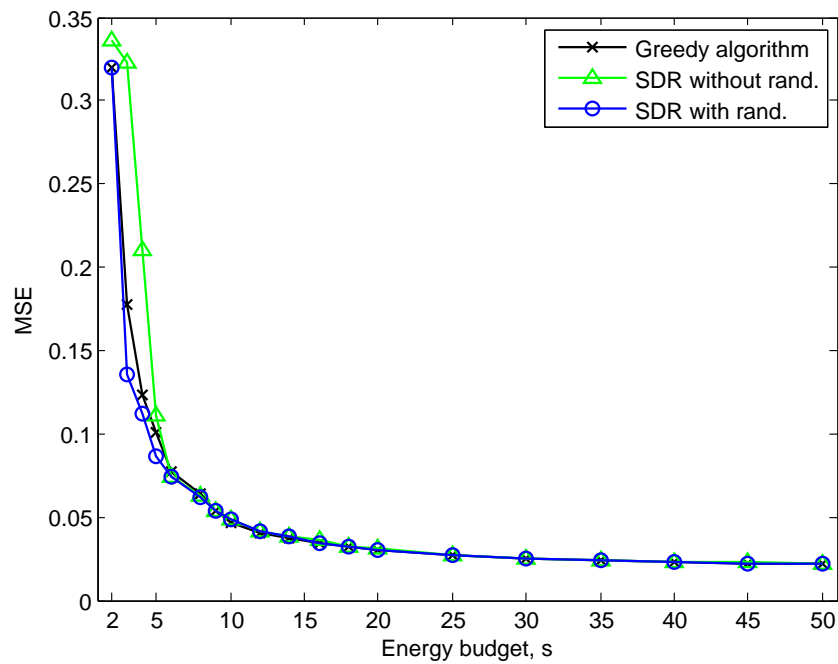
(b)

Fig. 3.1: MSE versus energy budget with correlated measurement noise.
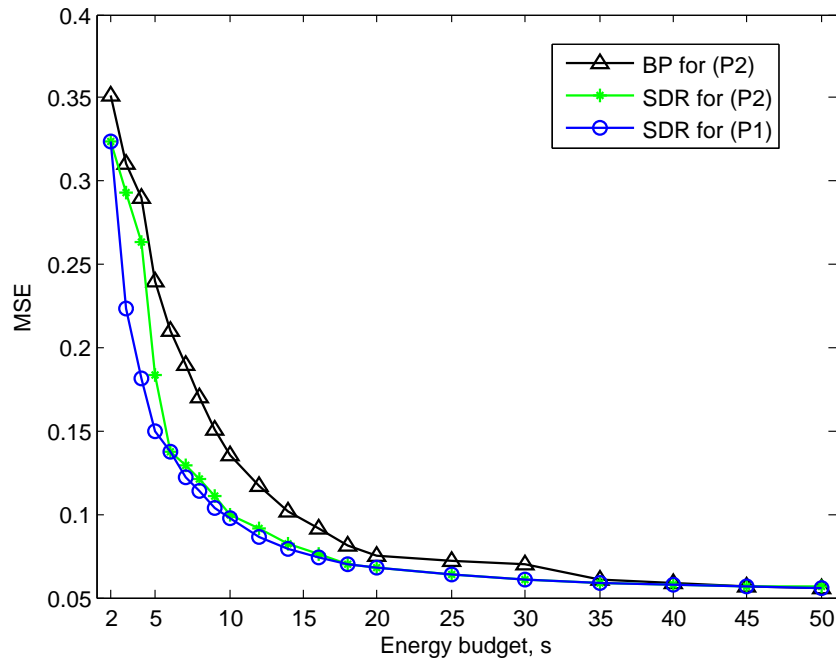
Fig. 3.2: MSE versus energy budget for sensor selection with weak noise correlation.
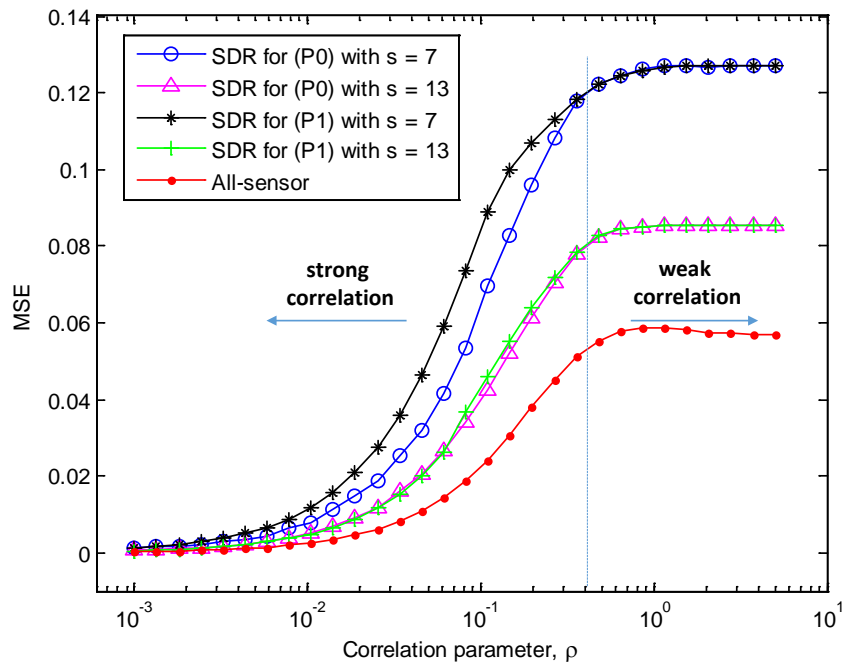
Fig. 3.3: MSE versus the strength of correlation.
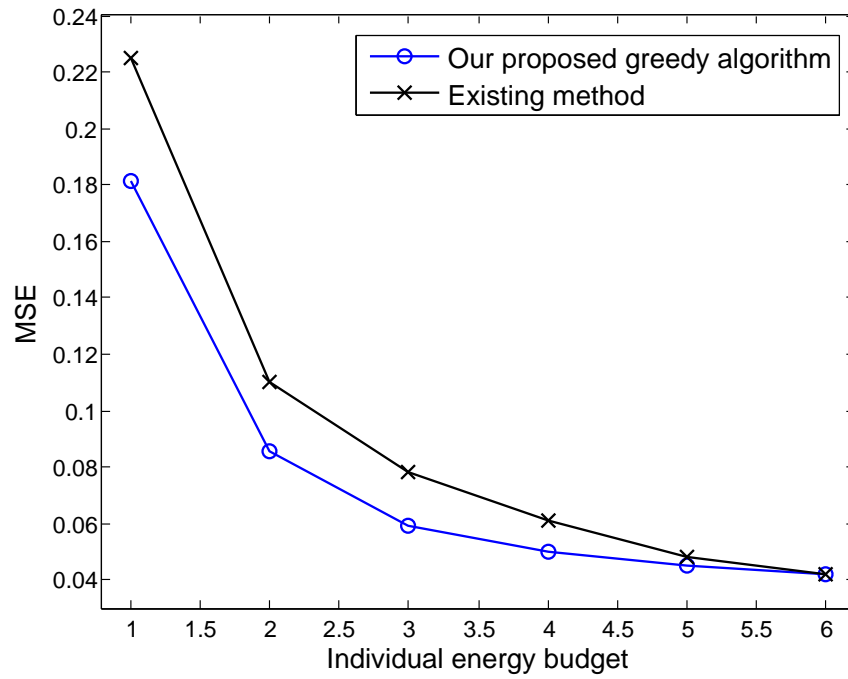
Fig. 3.4: MSE versus individual energy budget in target tracking.
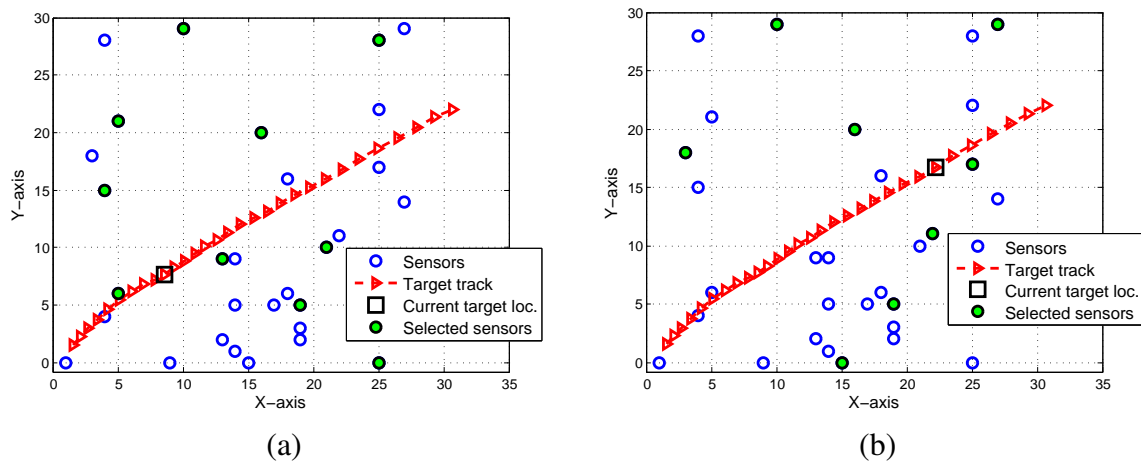
Fig. 3.5: Sensor schedules when $s_i = 2$: (a) $t = 10$, (b) $t = 24$.

CHAPTER 4

OPTIMAL PERIODIC SENSOR SCHEDULING VIA THE DESIGN OF SPARSE ESTIMATOR GAIN MATRICES

4.1 Introduction

Periodicity in the optimal sensor schedule was observed for both finite and infinite time horizon problems in which a periodic schedule was not assumed *a priori*. It was proved in [24], that the optimal sensor schedule for an infinite horizon problem can be approximated arbitrarily well by a periodic schedule with a finite period. We emphasize that the results in [24] are nonconstructive, in the sense that it is shown that the optimal sensor schedule is time-periodic but an algorithm for obtaining this schedule, or even the length of its period, is not provided. Although periodicity makes infinite horizon sensor scheduling problems tractable via the design of an optimal schedule over a finite period, it poses other challenges in problem formulation and optimization under the condition of periodicity.

In this chapter, we consider the problem of finding optimal time-periodic sensor schedules for estimating the state of discrete-time dynamical systems. We associate the activation of sensors

with the promotion of nonzero columns of an estimator gain matrix. Based on this association, we formulate an optimization problem in which we minimize the trace of the error covariance with respect to the estimator gain while simultaneously penalizing the number of nonzero columns of the estimator gain. We show that the desired estimator gain matrices satisfy a coupled sequence of periodic Lyapunov recursions, where a new block-cyclic representation is introduced to transform the coupled matrix recursions into algebraic matrix equations for ease of computation. Through application of ADMM, we uncover relationships between the frequency constraint parameter, the sparsity-promoting parameter, and the sensor schedule.

The remainder of the chapter is organized as follows. In Section 4.2, we motivate the problem of periodic sensor scheduling over an infinite time horizon. In Section 4.3, we formulate the sparsity-promoting periodic sensor scheduling problem. In Section 4.4, we invoke the ADMM method, which leads to a pair of efficiently solvable subproblems. In Section 4.5, we illustrate the effectiveness of our proposed approach through examples. We summarize our work in Section 4.6.

4.2   Periodicity of infinite horizon sensor scheduling

Consider a discrete-time linear dynamical system evolving according to the equations

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{w}_k, \tag{4.1}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k, \tag{4.2}$$

where $\mathbf{x}_k \in \mathbb{R}^N$ is the state vector at time k, $\mathbf{y}_k \in \mathbb{R}^M$ is the measurement vector whose mth entry corresponds to a scalar observation from sensor m, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrices of appropriate dimensions. The inputs $\mathbf{w}_k$ and $\mathbf{v}_k$ are white, Gaussian, zero-mean random vectors with covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively. Finally, we assume that $(\mathbf{A}, \mathbf{C})$ is detectable and $(\mathbf{A}, \boldsymbol{\Sigma})$ is stabilizable, where $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T = \mathbf{B}\mathbf{Q}\mathbf{B}^T$.

For ease of describing the sensor schedule, we introduce the auxiliary binary variables $\zeta_{k,m} \in \{0, 1\}$, to represent whether or not the mth sensor is activated at time k. The sensor schedule

over an infinite time horizon can then be denoted by $\boldsymbol{\mu}_\infty = [\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots]$, where the vector $\boldsymbol{\zeta}_k = [\zeta_{k,1}, \ldots, \zeta_{k,M}]^T$ indicates which sensors are active at time k. The performance of an infinite-horizon sensor schedule is then measured as follows [21, 22],

$$J(\boldsymbol{\mu}_\infty) \triangleq \lim_{\overline{K} \to \infty} \frac{1}{\overline{K}} \sum_{k=1}^{\overline{K}} \mathrm{tr}\,(\mathbf{P}_k) \tag{4.3}$$

where $\mathbf{P}_k$ is the estimation error covariance at time k under the sensor schedule $\boldsymbol{\mu}_\infty$. Due to the combinatorial nature of the problem, it is intractable to find the optimal sensor schedule that minimizes the cost (4.3) in general [22].

In [79], it was suggested that the optimal sensor schedule can be treated as a time-periodic schedule over the infinite time horizon if the system (4.1)-(4.2) is detectable and stabilizable. Furthermore, in [24] it was proved that the optimal sensor schedule for an infinite horizon problem can be approximated arbitrarily well by a periodic schedule with a finite period, and that the error covariance matrix converges to a unique limit cycle. In this case, the cost in (4.3) is given by

$$J(\boldsymbol{\mu}_K) = \frac{1}{K} \sum_{k=0}^{K-1} \mathrm{tr}\,(\mathbf{P}_k) \tag{4.4}$$

where K is the length of the period and $\mathbf{P}_k$ is the error covariance matrix at instant k of its limit cycle. In the thesis, we assume the length K of the period is given, and will focus on the optimization method to find the optimal periodic sensor schedule.

4.3 Problem statement

For the discrete-time linear dynamical system (4.1)–(4.2), we consider state estimators of the form

$$\hat{\mathbf{x}}_{k+1} = \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{L}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k) = (\mathbf{A} - \mathbf{L}_k\mathbf{C})\hat{\mathbf{x}}_k + \mathbf{L}_k\mathbf{y}_k,$$

where $\mathbf{L}_k$ is the estimator gain (also known as the observer gain [80]) at time k. In what follows, we aim to determine the matrices $\mathbf{L}_k$, $k = 0, 1, \ldots$, by solving an optimization problem that, in particular, promotes the column sparsity of $\mathbf{L}_k$. We define the estimation error covariance $\mathbf{P}_k$ as

$$\mathbf{P}_k = \mathcal{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T\},$$

where $\mathcal{E}$ is the expectation operator.[1] It is easy to show that $\mathbf{P}_k$ satisfies the Lyapunov recursion

$$\mathbf{P}_{k+1} = (\mathbf{A} - \mathbf{L}_k\mathbf{C})\mathbf{P}_k(\mathbf{A} - \mathbf{L}_k\mathbf{C})^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T + \mathbf{L}_k\mathbf{R}\mathbf{L}_k^T. \tag{4.5}$$

Finally, partitioning the matrices $\mathbf{L}_k$ and $\mathbf{C}$ into their respective columns and rows, we have

$$\mathbf{L}_k\mathbf{C} = \begin{bmatrix} \mathbf{L}_{k,1} & \mathbf{L}_{k,2} & \ldots & \mathbf{L}_{k,M} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_2^T \\ \vdots \\ \mathbf{C}_M^T \end{bmatrix} = \mathbf{L}_{k,1}\mathbf{C}_1^T + \mathbf{L}_{k,2}\mathbf{C}_2^T + \cdots + \mathbf{L}_{k,M}\mathbf{C}_M^T, \tag{4.6}$$

where each row of $\mathbf{C}$ characterizes the measurement of one sensor. Therefore, each column of the matrix $\mathbf{L}_k$ can be thought of as corresponding to the measurement of a particular sensor.

In estimation and inference problems using wireless sensor networks, minimizing the energy consumption of sensors is often desired. Therefore, we seek algorithms that schedule the turning on and off of the sensors in order to strike a balance between energy consumption and estimation performance. Suppose, for example, that at time step k only the νth sensor reports a measurement. In this case, it follows from (4.6) that $\mathbf{L}_k\mathbf{C} = \mathbf{L}_{k,\nu}\mathbf{C}_\nu^T$, where $\mathbf{C}_\nu^T$ is the νth row of $\mathbf{C}$. This can also be interpreted as having the column vectors $\mathbf{L}_{k,m}$ equal to zero for all $m \neq \nu$. Thus, hereafter we assume that the measurement matrix $\mathbf{C}$ is constant and the scheduling of the sensors is captured by the nonzero columns of the estimator gains $\mathbf{L}_k$, in the sense that if $\mathbf{L}_{k,m} = \mathbf{0}$ then at time k the

---

[1]In the system theory literature, $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ are often denoted by $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$; here we use $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ for simplicity of notation.

mth sensor is not making a measurement.

We aim to search for optimal *time-periodic* sensor schedules, i.e., we seek optimal sequences $\{\mathbf{L}_k\}_{k=0,1,\ldots,K-1}$ and $\{\mathbf{P}_k\}_{k=0,1,\ldots,K-1}$ that satisfy

$$\mathbf{L}_{k+K} = \mathbf{L}_k, \quad \mathbf{P}_{k+K} = \mathbf{P}_k, \tag{4.7}$$

where K is a given period. Note that the choice of K is not a part of the optimization problem considered in this thesis. As suggested in [24], one possible procedure for choosing K is to find the optimal sensor schedule for gradually-increasing values of K until the performance ceases to improve significantly. Furthermore, the condition on the periodicity of $\mathbf{P}_k$ assumes that the system and estimator with $\mathbf{L}_{k+K} = \mathbf{L}_k$ have been running for a long time so that $\mathbf{P}_k$ has reached its steady-state limit cycle [24]. In this thesis, we consider $k = -\infty$ as the initial time and without loss of generality consider the design of $\mathbf{L}_k$ over the period $k = 0, 1, \ldots, K-1$, when the system has settled into its periodic cycle.

To incorporate the energy constraints on individual sensors over a period of length K, we consider

$$\sum_{k=0}^{K-1} \mathrm{card}\big(\|\mathbf{L}_{k,m}\|_2\big) \leq \eta_m, \quad m = 1, 2, \ldots, M, \tag{4.8}$$

where η_m denotes the measurement frequency bound. This implies that the mth sensor can make and transmit at most η_m measurements over the period of length K. Placing a bound on the measurement frequency of each sensor avoids scenarios in which a set of 'informative sensors' are successively selected, which would result in their early energy depletion. For simplicity, we assume $\eta_1 = \eta_2 = \ldots = \eta_M = \eta$. We remark that the proposed sensor scheduling methodology applies equally well to the case where the η_i are not necessarily equal to each other.

We pose the optimal sensor scheduling problem as the optimization problem

$$
\begin{aligned}
\text{minimize} \quad & \sum_{k=0}^{K-1} \text{tr}(\mathbf{P}_k) + \gamma \sum_{k=0}^{K-1} g(\mathbf{L}_k) \\
\text{subject to} \quad & \begin{cases} \text{Lyapunov recursion (4.5) for } k = 0, 1, \ldots, K-1 \\[2mm] \text{periodicity condition (4.7)} \\[2mm] \text{measurement frequency constraints (4.8),} \end{cases}
\end{aligned}
\tag{4.9}
$$

where the matrices $\{\mathbf{L}_k\}_{k=0,\cdots,K-1}$ are the optimization variables, $\text{card}(\cdot)$ denotes the cardinality function which gives the number of nonzero elements of its (vector) argument, and

$$
g(\mathbf{L}_k) := \text{card}\Big(\begin{bmatrix} \|\mathbf{L}_{k,1}\|_2 & \|\mathbf{L}_{k,2}\|_2 & \cdots & \|\mathbf{L}_{k,M}\|_2 \end{bmatrix} \Big).
\tag{4.10}
$$

Therefore $g(\mathbf{L}_k)$ is equal to the number of nonzero columns of $\mathbf{L}_k$, also referred to as the column-cardinality of $\mathbf{L}_k$. The incorporation of the sparsity-promoting term $g(\cdot)$ in the objective function encourages the use of a small subset of sensors at each time instant. The positive scalar γ characterizes the relative importance of the two conflicting terms in the objective, namely the relative importance of achieving good estimation performance versus activating a small number of sensors.

4.4 Optimal periodic sensor scheduling using ADMM

In this section, we apply ADMM to the sensor scheduling problem (4.9). Our treatment uses ideas introduced in [81], where ADMM was used for the identification of optimal sparse state-feedback gains. We extend the framework of [81] to account for the time periodicity of the estimator gains, their sparsity across both space and time, and the addition of measurement frequency constraints on individual sensors.

We begin by reformulating the optimization problem in (4.9) in a way that lends itself to the

application of ADMM. For $\mathbf{P}_k$ that satisfies the Lyapunov recursion in (4.5), it is easy to show that

$$\mathbf{P}_k = \mathbf{B}\mathbf{Q}\mathbf{B}^T + \mathbf{L}_{k-1}\mathbf{R}\mathbf{L}_{k-1}^T + \sum_{n=k-1}^{-\infty}(\mathbf{A} - \mathbf{L}_{k-1}\mathbf{C})\cdots(\mathbf{A} - \mathbf{L}_n\mathbf{C})(\mathbf{B}\mathbf{Q}\mathbf{B}^T + \mathbf{L}_{n-1}\mathbf{R}\mathbf{L}_{n-1}^T)$$
$$\cdot (\mathbf{A} - \mathbf{L}_n\mathbf{C})^T \cdots (\mathbf{A} - \mathbf{L}_{k-1}\mathbf{C})^T.$$

Invoking the periodicity of $\mathbf{L}_k$, $\operatorname{tr}(\mathbf{P}_k)$ can be expressed as a function f_k of $\{\mathbf{L}_k\}_{k=0,\cdots,K-1}$ so that the optimization problem (4.9) can be rewritten as

$$\begin{aligned}
\text{minimize} \quad & \sum_{k=0}^{K-1} f_k(\mathbf{L}_0,\cdots,\mathbf{L}_{K-1}) + \gamma\sum_{k=0}^{K-1} g(\mathbf{L}_k) \\
\text{subject to} \quad & \sum_{k=0}^{K-1}\operatorname{card}\big(\|\mathbf{L}_{k,m}\|_2\big) \le \eta, \quad m = 1, 2, \ldots, M.
\end{aligned}$$

We next introduce the indicator function corresponding to the constraint set of the above optimization problem as

$$\mathcal{I}(\{\mathbf{L}_k\}) = \begin{cases} 0 & \text{if } \sum_{k=0}^{K-1}\operatorname{card}\big(\|\mathbf{L}_{k,m}\|_2\big) \le \eta \text{ for } m = 1, 2, \ldots, M, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.11}$$

where for notational simplicity we have used, and henceforth will continue to use, $\{\cdot\}$ instead of $\{\cdot\}_{k=0,\ldots,K-1}$. Incorporating the indicator function into the objective function, problem (4.9) is equivalent to the unconstrained optimization problem

$$\text{minimize} \quad \sum_{k=0}^{K-1} f_k(\{\mathbf{L}_k\}) + \gamma\sum_{k=0}^{K-1} g(\mathbf{L}_k) + \mathcal{I}(\{\mathbf{L}_k\}).$$

Finally, we introduce the new set of variables $\{\mathbf{G}_k\}$, together with the new set of constraints

$\mathbf{L}_k = \mathbf{G}_k$, $k = 0, 1, \ldots, K - 1$, and formulate

$$\begin{aligned}
\text{minimize} \quad & \sum_{k=0}^{K-1} f_k(\{\mathbf{L}_k\}) + \gamma \sum_{k=0}^{K-1} g(\mathbf{G}_k) + \mathcal{I}(\{\mathbf{G}_k\}) \\
\text{subject to} \quad & \mathbf{L}_k = \mathbf{G}_k, \quad k = 0, 1, \ldots, K - 1,
\end{aligned} \tag{4.12}$$

which is now in a form suitable for the application of ADMM.

The augmented Lagrangian [81, 82] corresponding to optimization problem (4.12) is given by

$$\begin{aligned}
\mathscr{L}(\{\mathbf{L}_k\}, \{\mathbf{G}_k\}, \{\boldsymbol{\Lambda}_k\}) = & \sum_{k=0}^{K-1} f_k(\{\mathbf{L}_k\}) + \gamma \sum_{k=0}^{K-1} g(\mathbf{G}_k) + \mathcal{I}(\{\mathbf{G}_k\}) \\
& + \sum_{k=0}^{K-1} \mathrm{tr}[\boldsymbol{\Lambda}_k(\mathbf{L}_k - \mathbf{G}_k)] + \frac{\rho}{2} \sum_{k=0}^{K-1} ||\mathbf{L}_k - \mathbf{G}_k||_F^2,
\end{aligned} \tag{4.13}$$

where the matrices $\{\boldsymbol{\Lambda}_k\}$ are the Lagrange multipliers (also referred to as the dual variables), the scalar $\rho > 0$ is a penalty weight, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, $\|\mathbf{X}\|_F^2 = \mathrm{tr}(\mathbf{X}^T\mathbf{X})$. The ADMM algorithm can be described as follows; see Chapter 2.5. For $i = 0, 1, \ldots$, we iteratively execute the following three steps

$$\{\mathbf{L}_k^{i+1}\} := \underset{\{\mathbf{L}_k\}}{\arg\min} \ \mathscr{L}(\{\mathbf{L}_k\}, \{\mathbf{G}_k^i\}, \{\boldsymbol{\Lambda}_k^i\}), \tag{4.14}$$

$$\{\mathbf{G}_k^{i+1}\} := \underset{\{\mathbf{G}_k\}}{\arg\min} \ \mathscr{L}(\{\mathbf{L}_k^{i+1}\}, \{\mathbf{G}_k\}, \{\boldsymbol{\Lambda}_k^i\}), \tag{4.15}$$

$$\boldsymbol{\Lambda}_k^{i+1} := \boldsymbol{\Lambda}_k^i + \rho(\mathbf{L}_k^{i+1} - \mathbf{G}_k^{i+1}), k = 0, 1, \ldots, K-1, \tag{4.16}$$

until both of the conditions $\sum_{k=0}^{K-1} \|\mathbf{L}_k^{i+1} - \mathbf{G}_k^{i+1}\|_F \leq \epsilon$, and $\sum_{k=0}^{K-1} \|\mathbf{G}_k^{i+1} - \mathbf{G}_k^i\|_F \leq \epsilon$ are satisfied.

The rationale behind using ADMM can be described as follows [81]. The original nonconvex optimization problem (4.9) is difficult to solve due to the nondifferentiability of the sparsity-promoting function. By defining the new set of variables $\{\mathbf{G}_k\}$, we effectively separate the original problem into an "**L**-minimization" step (4.14) and a "**G**-minimization" step (4.15), of which the

former can be addressed using variational methods and descent algorithms and the latter can be solved analytically.

L-minimization using Anderson-Moore method

Completing the squares with respect to $\{\mathbf{L}_k\}$ in the augmented Lagrangian (4.13), the $\mathbf{L}$-minimization step in (4.14) can be expressed as [81, 82]

$$\text{minimize} \quad \sum_{k=0}^{K-1} f_k(\{\mathbf{L}_k\}) + \sum_{k=0}^{K-1} \frac{\rho}{2}||\mathbf{L}_k - \mathbf{U}_k^i||_F^2 \tag{4.17}$$

where $\mathbf{U}_k^i := \mathbf{G}_k^i - (1/\rho)\mathbf{\Lambda}_k^i$ for $k = 0, 1, \dots, K-1$. For notational simplicity, henceforth we will use $\mathbf{U}_k$ instead of $\mathbf{U}_k^i$, where i indicates the iteration index. We bring attention to the fact that, by defining the indicator function $\mathcal{I}$ in (6.22) and then splitting the optimization variables in (4.12), we have effectively removed both sparsity penalties and energy constraints from the variables $\{\mathbf{L}_k\}$ in the $\mathbf{L}$-minimization problem (4.17). This is a key advantage of applying ADMM to the sensor scheduling problem. Recalling the definition of f_k, problem (4.17) can be written as

$$\text{minimize} \quad \phi(\{\mathbf{L}_k\}) := \sum_{k=0}^{K-1} \text{tr}(\mathbf{P}_k) + \sum_{k=0}^{K-1} \frac{\rho}{2}||\mathbf{L}_k - \mathbf{U}_k||_F^2$$

$$\text{subject to} \quad \begin{cases} \text{Lyapunov recursion (4.5) for } k = 0, 1, \dots, K-1, \\ \\ \text{periodicity condition (4.7),} \end{cases}$$

where $\{\mathbf{L}_k\}$ are optimization variables.

Proposition 4.1. *The necessary conditions for the optimality of a sequence $\{\mathbf{L}_k\}$ can be expressed as the set of coupled matrix recursions*

$$\mathbf{P}_{k+1} = (\mathbf{A} - \mathbf{L}_k\mathbf{C})\mathbf{P}_k(\mathbf{A} - \mathbf{L}_k\mathbf{C})^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T + \mathbf{L}_k\mathbf{R}\mathbf{L}_k^T$$

$$\mathbf{V}_k = (\mathbf{A} - \mathbf{L}_k\mathbf{C})^T\mathbf{V}_{k+1}(\mathbf{A} - \mathbf{L}_k\mathbf{C}) + \mathbf{I}$$

$$0 = 2\mathbf{V}_{k+1}\mathbf{L}_k\mathbf{R} - 2\mathbf{V}_{k+1}(\mathbf{A} - \mathbf{L}_k\mathbf{C})\mathbf{P}_k\mathbf{C}^T + \rho(\mathbf{L}_k - \mathbf{U}_k)$$

for $k = 0, \ldots, K - 1$, *where* $\mathbf{U}_k := \mathbf{G}_k^i - (1/\rho)\mathbf{\Lambda}_k^i$ *and* $\mathbf{L}_K = \mathbf{L}_0$, $\mathbf{P}_K = \mathbf{P}_0$. *The expression on the right of the last equation is the gradient of* ϕ *with respect to* $\mathbf{L}_k$.

Proof: See Appendix A.5. ∎

It is a difficult exercise to solve the above set of matrix equations due to their coupling. We thus employ the Anderson-Moore method [81, 83], which is an efficient technique for iteratively solving systems of coupled Lyapunov and Sylvester *equations*. We note, however, that the set of matrix equations given in the proposition include (periodic) Lyapunov *recursions* rather than (time-independent) Lyapunov equations. We next apply what can be thought of as a *lifting* procedure [84] to take the periodicity out of these equations and place them in a form appropriate for the application of the Anderson-Moore method.

Let $\mathcal{T}$ denote the following permutation matrix in block-cyclic form

$$
\mathcal{T} := \begin{bmatrix} \mathbf{0} & & & \mathbf{I} \\ \mathbf{I} & \ddots & & \\ & \ddots & \ddots & \\ & & \mathbf{I} & \mathbf{0} \end{bmatrix}, \quad \text{and define } \mathcal{L} := \mathcal{T}\mathrm{diag}\{\mathbf{L}_k\} = \begin{bmatrix} \mathbf{0} & & & \mathbf{L}_{K-1} \\ \mathbf{L}_0 & & & \\ & \ddots & \ddots & \\ & & \mathbf{L}_{K-2} & \mathbf{0} \end{bmatrix},
$$

$$
\mathcal{P} := \mathrm{diag}\{\mathbf{P}_k\}, \quad \mathcal{V} := \mathrm{diag}\{\mathbf{V}_k\}, \quad \mathcal{U} := \mathcal{T}\mathrm{diag}\{\mathbf{U}_k\}, \quad \mathcal{Q} := \mathrm{diag}\{\mathbf{Q}\}, \quad \mathcal{R} := \mathrm{diag}\{\mathbf{R}\}
$$

$$
\mathcal{I} := \mathrm{diag}\{\mathbf{I}\}, \quad \mathcal{A} := \mathcal{T}\mathrm{diag}\{\mathbf{A}\}, \quad \mathcal{B} := \mathrm{diag}\{\mathbf{B}\}, \quad \mathcal{C} := \mathrm{diag}\{\mathbf{C}\}.
$$

In the sequel, we do not distinguish between the sequence $\{\mathbf{L}_k\}$ and its cyclic form $\mathcal{L}$, and will alternate between the two representations as needed.

The recursive equations in the statement of Proposition 4.1 can now be rewritten in the time-independent form

$$
\mathcal{P} = (\mathcal{A} - \mathcal{L}\mathcal{C})\mathcal{P}(\mathcal{A} - \mathcal{L}\mathcal{C})^T + \mathcal{B}\mathcal{Q}\mathcal{B}^T + \mathcal{L}\mathcal{R}\mathcal{L}^T \tag{4.18}
$$

$$
\mathcal{V} = (\mathcal{A} - \mathcal{L}\mathcal{C})^T\mathcal{V}(\mathcal{A} - \mathcal{L}\mathcal{C}) + \mathcal{I} \tag{4.19}
$$

$$
\mathbf{0} = 2\mathcal{V}\mathcal{L}\mathcal{R} - 2\mathcal{V}(\mathcal{A} - \mathcal{L}\mathcal{C})\mathcal{P}\mathcal{C}^T + \rho(\mathcal{L} - \mathcal{U}) \tag{4.20}
$$

Furthermore, defining $\nabla\Phi := \mathcal{T}\text{diag}\{\nabla_{\mathbf{L}_k}\phi\}$, it can be shown that

$$\nabla\Phi = 2\mathcal{V}\mathcal{L}\mathcal{R} - 2\mathcal{V}(\mathcal{A} - \mathcal{L}\mathcal{C})\mathcal{P}\mathcal{C}^{\mathcal{T}} + \rho(\mathcal{L} - \mathcal{U}), \tag{4.21}$$

i.e., the right side of (4.20) gives the gradient direction for $\mathcal{L}$, or equivalently the gradient direction for each $\mathbf{L}_k$, $k = 0, 1, \ldots, K - 1$.

We briefly describe the implementation of the Anderson-Moore method as follows. For each iteration of this method, we first keep the value of $\mathcal{L}$ fixed and solve (4.18) and (4.19) for $\mathcal{P}$ and $\mathcal{V}$, then keep $\mathcal{P}$ and $\mathcal{V}$ fixed and solve (4.20) for a new value $\mathcal{L}_{new}$ of $\mathcal{L}$. Proposition 4.2 shows that the difference $\tilde{\mathcal{L}} := \mathcal{L}_{new} - \mathcal{L}$ between the values of $\mathcal{L}$ over two consecutive iterations constitutes a descent direction for $\phi(\{\mathbf{L}_k\})$; see [81, 83] for related results. We employ a line search [59] to determine the step-size s in $\mathcal{L} + s\tilde{\mathcal{L}}$ in order to accelerate the convergence to a stationary point.

Proposition 4.2. *The difference $\tilde{\mathcal{L}} := \mathcal{L}_{new} - \mathcal{L}$ constitutes a descent direction for $\phi(\{\mathbf{L}_k\})$,*

$$\langle \nabla\Phi, \tilde{\mathcal{L}} \rangle < 0, \tag{4.22}$$

where $\langle \nabla\Phi, \tilde{\mathcal{L}} \rangle := \text{tr}(\nabla\Phi^T\tilde{\mathcal{L}})$. Moreover, $\langle \nabla\Phi(\mathcal{L}), \tilde{\mathcal{L}} \rangle = 0$ if and only if $\mathcal{L}$ is a stationary point of Φ, i.e., $\nabla\Phi(\mathcal{L}) = \mathbf{0}$.

Proof: The proof is similar to [85, Prop. 1] and omitted for brevity. ∎

We summarize the Anderson-Moore method for solving the **L**-minimization step in Algorithm 4.1. This algorithm calls on the Armijo rule [86], given in Algorithm 4.2, to update $\mathcal{L}$.

Algorithm 4.1 L-minimization step (4.14) using Anderson-Moore method

---

1: If $i = 0$, choose $\mathcal{L}^0$. If $i \geq 1$, set $\mathcal{L}^0$ equal to solution of (4.14) from previous ADMM iteration.
2: **for** $t = 0, 1, \ldots$ **do**
3:     Set $\mathcal{L} = \mathcal{L}^t$ and solve (4.18), (4.19) to find $\mathcal{P}^t$, $\mathcal{V}^t$.
4:     Set $\mathcal{V} = \mathcal{V}^t$, $\mathcal{P} = \mathcal{P}^t$ and solve (4.20) to find $\bar{\mathcal{L}}^t$.
5:     Compute $\tilde{\mathcal{L}}^t = \bar{\mathcal{L}}^t - \mathcal{L}^t$ and update $\mathcal{L}^{t+1} = \mathcal{L}^t + s^t\tilde{\mathcal{L}}^t$, where s^t is given by Algorithm 4.2.
6:     **until** $\|\nabla\Phi(\mathcal{L}^t)\| < \epsilon$.
7: **end for**

Algorithm 4.2 Armijo rule for choosing step-size s^t

1: Set $s^t = 1$ and choose $\alpha, \beta \in (0, 1)$.
2: **repeat**
3:     $s^t = \beta s^t$,
4: **until** $\phi(\boldsymbol{\mathcal{L}}^t + s^t \tilde{\boldsymbol{\mathcal{L}}}^t) < \phi(\boldsymbol{\mathcal{L}}^t) + \alpha\, s^t \operatorname{tr}\big(\nabla\Phi(\boldsymbol{\mathcal{L}}^t)^T \tilde{\boldsymbol{\mathcal{L}}}^t\big).$

G-minimization

In this section, we consider the **G**-minimization step (4.15) and demonstrate that it can be solved

analytically. In what follows, we extend the approach of [81] to account for the periodicity and

energy constraints in the sensor schedule.

Completing the squares with respect to $\{\mathbf{G}_k\}$ in the augmented Lagrangian (4.13), the **G**-

minimization step in (4.15) can be expressed as [81, 82]

$$
\begin{aligned}
\text{minimize} \quad & \gamma \sum_{k=0}^{K-1} g(\mathbf{G}_k) + \frac{\rho}{2}\sum_{k=0}^{K-1}||\mathbf{G}_k - \mathbf{S}_k^i||_F^2 \\
\text{subject to} \quad & \sum_{k=0}^{K-1} \operatorname{card}\big(||\mathbf{G}_{k,m}||_2\big) \leq \eta, \quad m = 1, 2, \ldots, M,
\end{aligned}
$$

where $\mathbf{S}_k^i := \mathbf{L}_k^{i+1} + (1/\rho)\boldsymbol{\Lambda}_k^i$ for $k = 0, 1, \ldots, K-1$. For notational simplicity, henceforth we

will use $\mathbf{S}_k$ instead of $\mathbf{S}_k^i$, where i indicates the iteration index. Recalling the definition of g from

(4.10), and replacing $||\mathbf{G}_k - \mathbf{S}_k||_F^2$ with $\sum_{m=1}^{M} ||\mathbf{G}_{k,m} - \mathbf{S}_{k,m}||_2^2$ yields the equivalent optimization

problem

$$
\begin{aligned}
\text{minimize} \quad & \psi(\{\mathbf{G}_k\}) := \sum_{m=1}^{M}\Bigg(\sum_{k=0}^{K-1}\gamma\operatorname{card}\big(||\mathbf{G}_{k,m}||_2\big) + \sum_{k=0}^{K-1}\frac{\rho}{2}||\mathbf{G}_{k,m} - \mathbf{S}_{k,m}||_2^2\Bigg) \\
\text{subject to} \quad & \sum_{k=0}^{K-1}\operatorname{card}\big(||\mathbf{G}_{k,m}||_2\big) \leq \eta, \quad m = 1, 2, \ldots, M,
\end{aligned}
$$

where we have exploited the *column-wise separability* of $g(\cdot)$ and that of the Frobenius norm.

We form the matrix $\boldsymbol{\mathcal{G}}_m$ by picking out the mth column from each of the matrices in the set

$\{\mathbf{G}_k\}$ and stacking them, $\boldsymbol{\mathcal{G}}_m := [\mathbf{G}_{0,m}, \mathbf{G}_{1,m}, \cdots, \mathbf{G}_{K-1,m}]$. Then the **G**-minimization problem

decomposes into the subproblems

$$\text{minimize} \quad \psi_m(\boldsymbol{\mathcal{G}}_m) := \sum_{k=0}^{K-1} \gamma \operatorname{card}(\|\mathbf{G}_{k,m}\|_2) + \sum_{k=0}^{K-1} \frac{\rho}{2} \|\mathbf{G}_{k,m} - \mathbf{S}_{k,m}\|_2^2$$

$$\text{subject to} \quad \sum_{k=0}^{K-1} \operatorname{card}(\|\mathbf{G}_{k,m}\|_2) \leq \eta, \tag{4.23}$$

which can be solved separately for $m = 1, 2, \ldots, M$.

To solve problem (4.23) we rewrite the feasible set F of (4.23), $F = \big\{ \boldsymbol{\mathcal{G}}_m : \sum_{k=0}^{K-1} \operatorname{card}(\|\mathbf{G}_{k,m}\|_2) \leq \eta \big\}$, as the union $F = F_0 \cup F_1 \cup \cdots \cup F_\eta$ of the smaller sets F_q, $q = 0, \ldots, \eta$,

$$F_q = \big\{ \boldsymbol{\mathcal{G}}_m : \sum_{k=0}^{K-1} \operatorname{card}(\|\mathbf{G}_{k,m}\|_2) = q \big\}.$$

Let $\boldsymbol{\mathcal{G}}_m^q$ denote a solution of

$$\text{minimize} \quad \psi_m(\boldsymbol{\mathcal{G}}_m)$$

$$\text{subject to} \quad \boldsymbol{\mathcal{G}}_m \in F_q. \tag{4.24}$$

Then a minimizer of (4.23) can be obtained by comparing $\psi_m(\boldsymbol{\mathcal{G}}_m^q)$ for $q = 0, \ldots, \eta$ and choosing the one with the least value. The above procedure, together with finding the solution of (4.24), is made precise by the following proposition.

Proposition 4.3. *The solution of (4.23) is obtained by solving the sequence of minimization problems (4.24) for $q = 0, 1, \ldots, \min\{\eta, \kappa\}$, $\kappa = \sum_{k=0}^{K-1} \operatorname{card}(\|\mathbf{S}_{k,m}\|_2)$. Furthermore, the solution of (4.24) is given by*

$$\mathbf{G}_{k,m} = \begin{cases} \mathbf{S}_{k,m} & \|\mathbf{S}_{k,m}\|_2 \geq \|[\boldsymbol{\mathcal{S}}_m]_q\|_2 \text{ and } q \neq 0, \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

for $k = 0, 1, \cdots, K-1$, where $\mathbf{S}_k := \mathbf{L}_k^{i+1} + (1/\rho)\boldsymbol{\Lambda}_k^i$, $\boldsymbol{\mathcal{S}}_m := [\mathbf{S}_{0,m}, \cdots, \mathbf{S}_{K-1,m}]$, $[\boldsymbol{\mathcal{S}}_m]_q$ denotes the qth largest column of $\boldsymbol{\mathcal{S}}_m$ in the 2-norm sense, and $\mathbf{G}_{k,m}$, $\mathbf{S}_{k,m}$ denote the mth columns of $\mathbf{G}_k$, $\mathbf{S}_k$, respectively.

Proof: See Appendix A.6. ∎

We note that problem (4.23) can be solved via a sequence of equality constrained problems (4.24) whose analytical solution is determined by Proposition 4.3. However, instead of solving $\min\{\eta, \kappa\} + 1$ equality constrained problems, it is shown in Proposition 4.4 that the solution of the G-minimization problem (4.23) is determined by the magnitude of the sparsity-promoting parameter γ.

Proposition 4.4. *The solution $\boldsymbol{\mathcal{G}}_m$ of (4.23) is determined by solving one subproblem (4.24) based on the value of γ,*

$$
\boldsymbol{\mathcal{G}}_m = \begin{cases}
\boldsymbol{\mathcal{G}}_m^0 & \frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_1\|_2^2 < \gamma \\
\boldsymbol{\mathcal{G}}_m^1 & \frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_2\|_2^2 < \gamma \leq \frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_1\|_2^2 \\
\vdots & \vdots \\
\boldsymbol{\mathcal{G}}_m^{\min\{\eta,\kappa\}} & \gamma \leq \frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_{\min\{\eta,\kappa\}}\|_2^2
\end{cases} \tag{4.25}
$$

where $\boldsymbol{\mathcal{G}}_m^q$ denotes a solution of (4.24) with $q = 0, 1, \ldots, \min\{\eta, \kappa\}$, and κ and $[\boldsymbol{\mathcal{S}}_m]_q$ are defined as in Proposition 4.3.

Proof: See Appendix A.7. ∎

It is clear from Proposition 4.4 that the parameter γ governs the column-sparsity of $\boldsymbol{\mathcal{G}}_m$. For example, $\boldsymbol{\mathcal{G}}_m$ becomes the zero matrix as $\gamma \to \infty$, which corresponds to the scenario in which all sensors are always inactive.

To reiterate, in order to solve the G-minimization problem (4.15), we first decompose it into the M subproblems (4.23) with separate optimization variables $\{\boldsymbol{\mathcal{G}}_m\}_{m=1,\ldots,M}$. Each inequality constrained subproblem (4.23) is then solved via Proposition 4.4, in which the solution of the equality constrained problem (4.24) is determined by Proposition 4.3. We summarize this procedure in Algorithm 4.3.

Algorithm 4.3 G-minimization step (4.15)

1: Given η and $\mathbf{S}_k = \mathbf{L}_k^{i+1} + 1/\rho\mathbf{\Lambda}_k^i$, set $\kappa = \sum_{k=0}^{K-1} \text{card}\left(\|\mathbf{S}_{k,m}\|_2\right)$.
2: **for** $m = 1, \ldots, M$ **do**
3:     Set $\boldsymbol{\mathcal{S}}_m = [\mathbf{S}_{0,m}, \cdots, \mathbf{S}_{K-1,m}]$.
4:     Solve (4.23) using Prop. 4.4 to obtain $\boldsymbol{\mathcal{G}}_m = \boldsymbol{\mathcal{G}}_m^q$, where $\boldsymbol{\mathcal{G}}_m^q$ is determined from Prop. 4.3.
5: **end for**
6: Use $\{\boldsymbol{\mathcal{G}}_m\}_{m=1,\ldots,M}$ to construct $\{\mathbf{G}_k\}_{k=0,1,\ldots,K-1}$.

Convergence & Initialization of ADMM-based periodic sensor scheduling

The solution of ADMM for a nonconvex problem generally yields a locally optimal point, and in general depends on the parameter ρ and the initial values of $\{L_k\}$ and $\{G_k\}$ [82]. In fact for a nonconvex problem, such as the one considered here, even the convergence of ADMM is not guaranteed [82]. Our numerical experiments and those in other works such as [81] demonstrate that ADMM indeed works well when the value of ρ is chosen to be large. However, *very* large values of ρ make the Frobenius norm dominate the augmented Lagrangian (4.13) and thus lead to less emphasis on minimizing the estimation error. In order to select an appropriate value of ρ, certain extensions (e.g., varying penalty parameter) of the classical ADMM algorithm have been explored. The reader is referred to [82, Sec. 3].

To initialize the estimator gain $\{\mathbf{L}_k\}$, we start with a feasible initializing sensor schedule. Such a schedule can be expressed in terms of the observation matrices over one period, namely, $\mathbf{C}(k) = [\zeta_{k,1}\mathbf{C}_1, \ldots, \zeta_{k,M}\mathbf{C}_M]^T$ for $k = 0, 1, \ldots, K-1$, where the binary variable $\zeta_{k,m}$ indicates whether or not the mth sensor is active at time k. Note that the periodic sensor schedule $\{\mathbf{C}(k)\}$ uniquely determines the limit cycle of the periodic error covariance matrix [24]. We express the periodic sensor schedule $\{\mathbf{C}(k)\}$ in cyclic form $\boldsymbol{\mathcal{C}}^0 := \boldsymbol{\mathcal{T}} \text{ diag}\{\mathbf{C}(k)\}$ and solve the following algebraic Riccati equation for the cyclic form of $\{\mathbf{P}_k\}$

$$\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{Q}} + \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}^T - \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{C}}^{0T}(\boldsymbol{\mathcal{C}}^0\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{C}}^{0T} + \boldsymbol{\mathcal{R}})^{-1}\boldsymbol{\mathcal{C}}^0\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}^{-1}, \tag{4.26}$$

where $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{A}$ and $\mathcal{R}$ have the same definitions as in Sec. 4.4. The Riccati equation (4.26) gives the optimal periodic estimator gain corresponding to a discrete-time system with *given* periodic observation matrices $\{\mathbf{C}(k)\}$. Once the solution of (4.26) is found, the corresponding estimator gain in cyclic form is given by [80]

$$\mathcal{L}^0 = \mathcal{A}\mathcal{P}\mathcal{C}^{0T}(\mathcal{C}^0\mathcal{P}\mathcal{C}^{0T} + \mathcal{R})^{-1}\mathcal{T}^0, \tag{4.27}$$

where $\mathcal{T}^0$ has the same block-cyclic form of $\mathcal{T}$ but is instead formed using $M{\times}M$ identity matrices.

It is not difficult to show that the matrix $\mathcal{L}^0$ in (4.27) has the same sparsity pattern as $\mathcal{C}^0$. Thus, the sequence $\{\mathbf{L}_k^0\}$ obtained from $\mathcal{L}^0$ respects the energy constraints and can be used to initialize ADMM. Furthermore, we assume that $(\mathcal{C}^0, \mathcal{A})$ is observable, which guarantees that the spectrum of $\mathcal{A} - \mathcal{L}^0\mathcal{C}^0$ is contained inside the open unit disk and thus the initializing estimator gains $\{\mathbf{L}_k^0\}$ will be stabilizing. Finally, for simplicity $\{\mathbf{G}_k\}$ is initialized to $\mathbf{G}_k = \mathbf{0}$, $k = 0, 1, \ldots, K - 1$.

Complexity analysis

It has been shown that ADMM typically takes a few tens of iterations to converge with modest accuracy for many applications [39, 81–83, 87]. The computational complexity of each iteration of ADMM is dominated by the $\mathbf{L}$-minimization step, since the analytical solution of the $\mathbf{G}$-minimization step can be directly obtained and the dual update is calculated by matrix addition. For the $\mathbf{L}$-minimization subproblem, the descent Anderson-Moore method requires the solutions of two Lyapunov equations (4.18)-(4.19) and one Sylvester equation (4.20) at each iteration. To solve them, the Bartels-Stewart method [88] yields the complexity $O(K^3N^3 + K^3M^3 + K^3MN^2 + K^3NM^2)$, where K is the length of the period, M is the number of sensors and N is the dimension of the state vector. We also note that the convergence of the Anderson-Moore method is guaranteed by Prop. 4.2, and it typically requires a small number of iterations because of the implementation of the Armijo rule.

For additional perspective, we compare the computational complexity of our proposed method-

ology to a periodic sensor scheduling problem that is solved by semidefinite programming (SDP), for example as done in [22]. The complexity of SDP is approximated by $O(a^2 b^{2.5} + ab^{3.5})$ [62], where a and b denote the number of optimization variables and the size of the semidefinite matrix, respectively. For the linear matrix inequality (LMI) problem proposed in [22], the computation complexity is determined by $a = N(N + 1)/2 + M$ and $b = (K + 1)N + M$. Thus, problems involving large-scale dynamical system with many state variables, result in large SDPs with computation complexity $O(N^{6.5})$. It can be seen that our approach reduces the computational complexity by a factor of $N^{3.5}$ compared to the LMI-based method of [22].

4.5   Numerical results

To demonstrate the effectiveness of our proposed periodic sensor scheduling algorithm, we consider the example of field monitoring in Fig. 4.1.
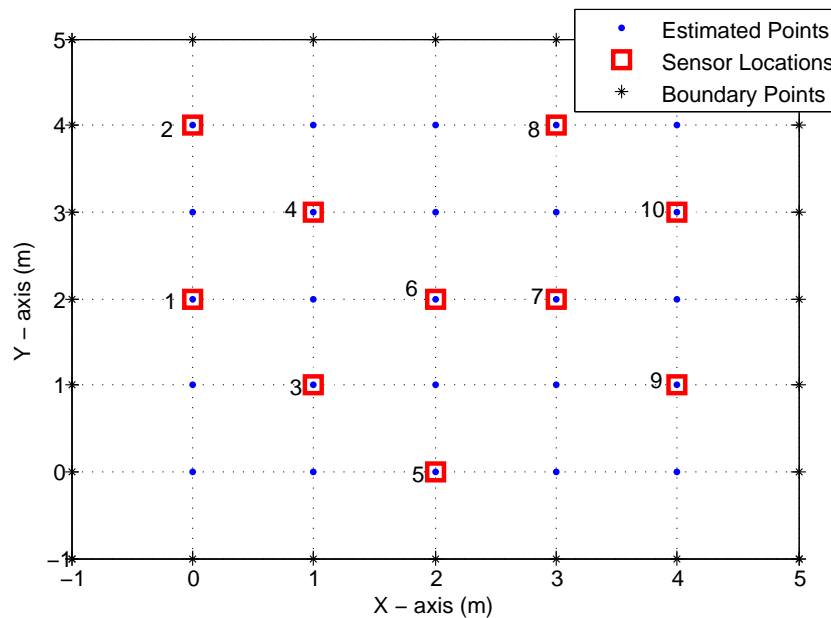
Fig. 4.1: $M = 10$ sensors deployed in a 6×6 region.

We assume that sensors are deployed on a rectangular region to estimate the state of a diffusion

process described by the partial differential equation [7, 18]

$$\frac{\partial \xi(\mathbf{s}, t)}{\partial t} = \nabla^2 \xi(\mathbf{s}, t) \tag{4.28a}$$

with Dirichlet boundary conditions

$$\xi(\mathbf{s}, \cdot) = 0 \quad \mathbf{s} \in \partial \mathcal{D} \tag{4.28b}$$

where $\xi(\mathbf{s}, t)$ denotes the field (or state) value at location $\mathbf{s}$ and time t, ∇^2 denotes the Laplace operator, and $\partial \mathcal{D}$ denotes the boundary of a rectangular region of interest $\mathcal{D}$.

We consider a spatially-discretized approximation of (4.28) and our aim is to estimate the state over the entire discrete lattice using a small number of sensors. With an abuse of notation, a simple discrete approximation of (4.28) can be generated by setting [18]

$$\nabla^2 \xi(\mathbf{s}, t)\big|_{s=(i,j)} \approx \frac{\xi(i+1, j, t) - 2\xi(i, j, t) + \xi(i-1, j, t)}{h^2}$$
$$+ \frac{\xi(i, j+1, t) - 2\xi(i, j, t) + \xi(i, j-1, t)}{h^2}, \tag{4.29}$$

for $i = 0, 1, \ldots, \ell_h$ and $j = 0, 1, \ldots, \ell_v$, where $\ell_h + 2$ and $\ell_v + 2$ are the width and length of a rectangular region, respectively In (4.29), h denotes the physical distance between the lattice points, and $\xi(-1, j, t) = \xi(\ell_h + 1, j, t) = \xi(i, -1, t) = \xi(i, \ell_v + 1, t) = 0$ for all indices i, j and time t.

From (4.28) and (4.29), we can obtain the evolution equations $\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}_\Delta x(t)$, where $\mathbf{x}(t) \in \mathbb{R}^N$, $N = (\ell_h + 1) \times (\ell_v + 1)$, denotes the state vector $\mathbf{x}(t) = [\xi(0, 0, t), \xi(0, 1, t), \ldots, \xi(\ell_h, \ell_v, t)]^T$, and $\mathbf{A}_\Delta$ can be directly computed from (4.29). Finally, applying a discretization in time and introducing process noise (i.e., a spatio-temporal random field) into the evolution yields

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{w}_k.$$

Here, $\mathbf{x}_k$ is the state vector, $\mathbf{w}_k$ is a white Gaussian process with zero mean and covariance matrix

$\mathbf{Q}$, $\mathbf{A}$ is the system transition matrix $\mathbf{A} = e^{\mathbf{A}_\triangle T}$, and T is the temporal sampling interval.

We assume that M sensors, $M < N$, are deployed and make measurements according to

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k,$$

where $\mathbf{y}_k \in \mathbb{R}^M$ is the measurement vector, $\mathbf{v}_k$ denotes the measurement noise which is a white Gaussian process with zero mean and covariance matrix $\mathbf{R}$, and $\mathbf{C}$ is the $M \times N$ observation matrix. For example, the case where the mth row of $\mathbf{C}$ contains only one nonzero entry equal to 1 corresponds to the scenario in which the mth entry of $\mathbf{y}_k$ represents measurements of the field at the location of the mth sensor.

We consider an instance in which $M = 10$ sensors are deployed to monitor $N = 25$ field points. We assume that each sensor can be selected at most η times, $\eta \in \{1, \ldots, 10\}$, during any period of length $K = 10$. Furthermore, we select $T = 0.5$, $\mathbf{Q} = 0.25\,\mathbf{I}$, and $\mathbf{R} = \mathbf{I}$. The ADMM stopping tolerance is $\epsilon = 10^{-3}$. In our computations, ADMM converges for $\rho \geq 10$ and the required number of iterations is around 50.

In Fig. 4.2, for our approach we present the estimation performance, namely the cumulative traces of error covariance matrices over one period, respectively as a function of the cumulative column-cardinality of $\{L_k\}$ shown by (4.10) and the measurement frequency bound η. In the left plot, we fix $\eta = 5$ and vary γ, which results in changes in the column-cardinality of $\{L_k\}$ and renders the trade-off curve between the conflicting objectives of good estimation performance and minimal sensor usage. Numerical results demonstrate that as the column-cardinality of $\{L_k\}$ increases and more sensors are activated, the estimation performance improves. In the right plot, we observe that the estimation performance is improved by increasing η. This is not surprising, as a larger value of η allows the (most informative) sensors to be active more frequently.

In Fig. 4.3, we compare the estimation performance of our approach to that of random scheduling, where the latter method refers to randomly selected sensor schedules that satisfy the measurement frequency constraint and have the same total number of active sensors over one period as the
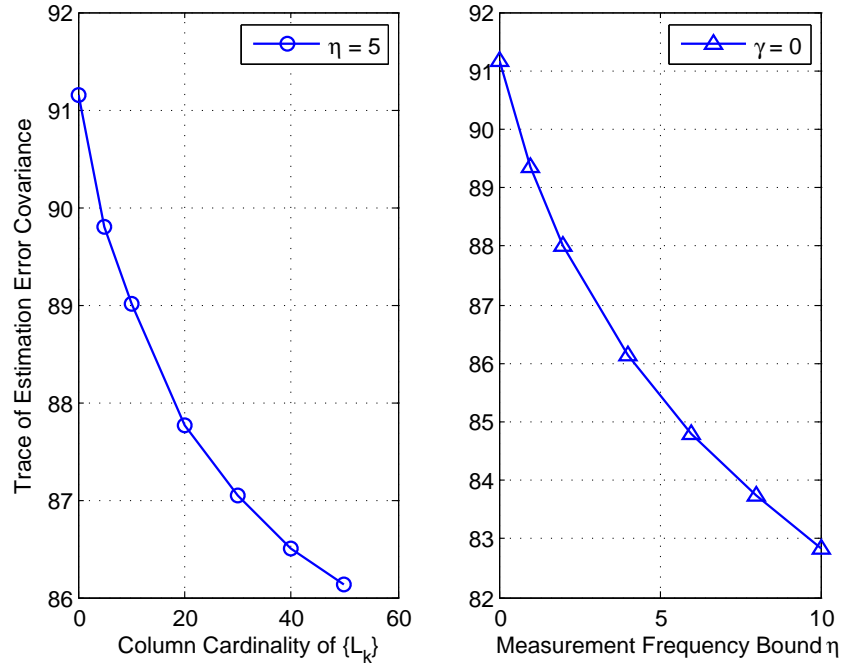
Fig. 4.2: Estimation performance. Left plot: Tradeoff between estimation performance and total number of sensors (in terms of column-cardinality of $\{L_k\}$) for a fixed $\eta = 5$; Right plot: Estimation performance as a function of measurement frequency bound η.

schedule obtained from our approach. The performance of the random strategy is taken to be the average of the traces of error covariance matrices over 500 simulation trials. Here the estimation performance is presented as a function of the measurement frequency bound η for three different values of the sparsity-promoting parameter $\gamma = 0, 0.1, 0.15$. Numerical results show that our approach significantly outperforms the random strategy for $\gamma = 0.1, 0.15$, as the former approach takes into account sensor activations over both time and space. For $\gamma = 0$ there is no penalty on sensor activations, and to achieve the best estimation performance every sensor is active η times per period (i.e., all sensors attain their measurement frequency bound). As a consequence, the performance gap between our approach and that of the random strategy is not as large for $\gamma = 0$ as it is for $\gamma > 0$. In our numerical experiments for smaller versions of this example, where exhaustive searches are feasible, we observed that our proposed method yields sensor schedules that are identical or close in performance to the globally optimal schedule found via an exhaustive search.

In Fig. 4.4, we compare the performance of our proposed sensor scheduling approach and the periodic switching policy of [37], where $M = 4$ and $Q = 0.25\mathbf{I}$. In Fig. 4.4-(a), each plot rep-
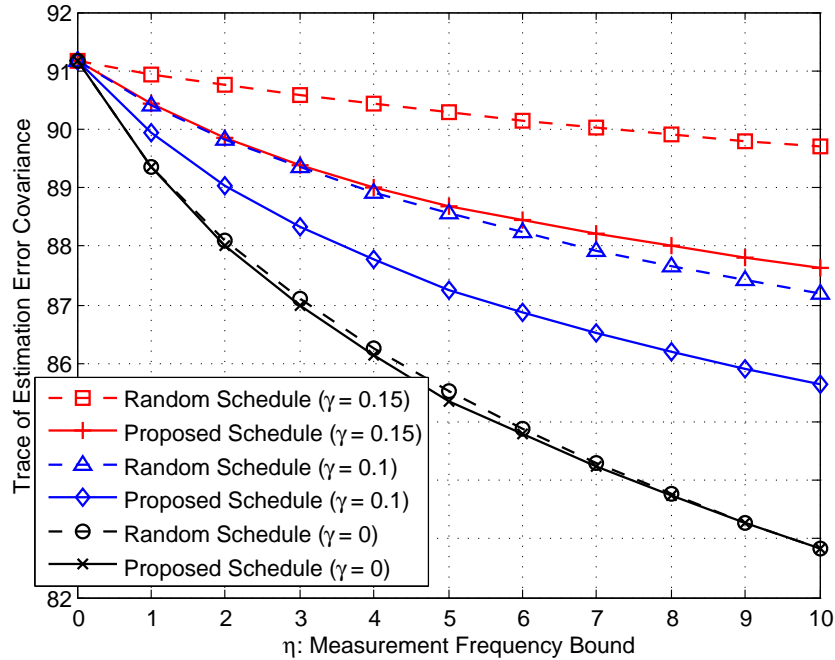
Fig. 4.3: Performance comparison of random schedules versus our proposed schedule.

resents the gap between the approach being considered and the globally optimal sensor schedule for different values of the sampling interval T, where $K = 4$, and T denotes the sampling time used to discretize the continuous-time system in (4.28). Since K is the period in discrete time, the period length in continuous time is given by $\epsilon = KT$. Simulation results show that both of the sensor scheduling methods achieve the performance of the globally optimal sensor schedule as $T \to 0$. This is due to the fact that $\epsilon \to 0$ while $T \to 0$. And it has been shown in [37] that the best estimation performance is attained by using a periodic switching policy as $\epsilon \to 0$ (and thus sensors are switched as fast as possible). However, as T increases, our approach outperforms the periodic switching policy significantly, which indicates that the sensor schedules obtained from the method of [37] are inappropriate for scheduling sensors for discrete-time systems with moderate sampling rates.

In Fig. 4.4-(b), we use the periodic switching policy of [37] to obtain the optimal sensor schedule and compare its performance with our approach for the fixed sampling interval $T = 0.5$ and different values of K. Fig. 4.4-(b) demonstrates that the periodic switching policy of [37] loses optimality as K increases. This is not surprising, since the optimality of the periodic switching
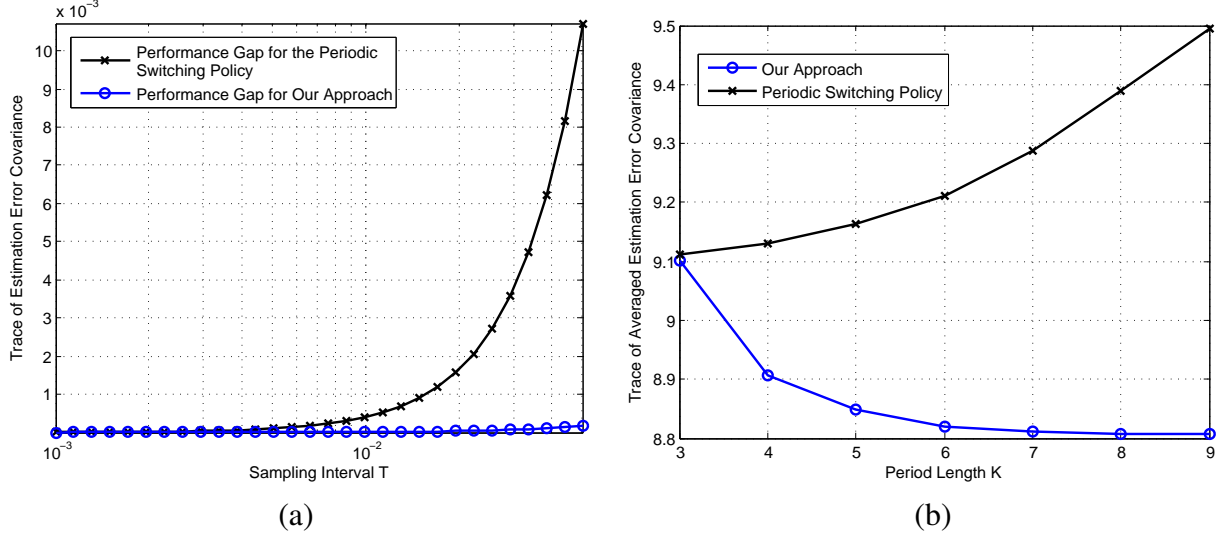
Fig. 4.4: Performance comparison with the periodic switching policy: a) Performance gaps of the periodic switching policy and our approach with respect to the optimal schedule; b) Performance comparison between our approach and the periodic switching policy for different values of period lengths.

policy is only guaranteed as the period length goes to zero. Therefore, to schedule sensors on a discrete-time system with a moderate sampling interval, such as $T = 0.5$ in the present example, our approach achieves better estimation performance than the periodic switching policy of [37].

In Fig. 4.5, we use ADMM to obtain the sensor schedule over a time period of length $K = 10$ for $\gamma \in \{0, 0.1, 0.15\}$ and $\eta \in \{1, 5, 8\}$; the subplots represent increasing values of γ from left to right and increasing values of η from top to bottom. In each subplot, the horizontal axis represents discrete time, the vertical axis represents sensor indices, and circles represent activated sensors. We also observe that sensors selected at two consecutive time instances tend to be spatially distant from each other. For example, at time instants $t = 1, 2, 3$, the active sensors are $6, 9, 4$, respectively.

In Figs. 4.5-(I-a), (I-b), and (I-c), we assume $\eta = 1$ and vary the magnitude of the sparsity-promoting parameter γ. As seen in Fig. 4.5-(I-a), for $\gamma = 0$ every sensor is selected exactly once over $K = 10$ time steps. Figs. 4.5-(I-b) and (I-c) demonstrate that fewer sensors are selected as γ is further increased. This is to be expected, as the value of γ in (4.9) determines our emphasis on the column-cardinality of $\{L_k\}$.

In Figs. 4.5-(I-c), (II-c), and (III-c) for $\gamma = 0.15$ we compare the optimal time-periodic sched-

ules for different values of the frequency bound $\eta = 1, 5, 8$. Numerical results show that for $\gamma = 0.15$ the 6th and 7th sensor are selected. To justify this selection, we note that these two sensors are located close to the center of the spatial region $\mathcal{D}$; see Fig. 4.1. Although we consider excitations that are in the form of a Gaussian random field, the states at the boundary $\partial \mathcal{D}$ are forced to take the value zero and the states closest to the center of $\mathcal{D}$ are subject to the largest uncertainty. Therefore, from the perspective of entropy, the measurements taken from the sensors 6 and 7 are the most informative for the purpose of field estimation. As we increase η, we allow such informative sensors to be active more frequently.

Moreover, the sensor schedule in Fig. 4.5-(II-c) verifies the optimality of the *uniform staggered sensing* schedule for two sensors, a sensing strategy whose optimality was proven in [38] and [21, Proposition 5.2]. In addition, although the periodicity of the sensor schedule was *a priori* fixed at the value $K = 10$, as η increases numerical results demonstrate repetitive patterns in the optimal sensor schedule. As seen in Figs. 4.5-(II-c) and (III-c), for $\eta = 5$ and $\eta = 8$ the sensor schedule repeats itself five times over 10 time steps and two times over 10 time steps, respectively. This indicates that the value of the sensing period K can be made smaller than 10.

4.6 Summary

In this chapter, we designed the optimal periodic sensor schedule for linear dynamical systems. In order to strike a balance between estimation accuracy and the number of sensor activations, the optimization problem was formulated to minimize the trace of the estimation error covariance matrices while penalizing the number of nonzero columns of the estimator gains. Through the application of ADMM, we split the optimization problem into subproblems that can either be solved efficiently using iterative numerical methods or solved analytically to obtain exact solutions. We showed that our approach outperforms previously available periodic sensor scheduling algorithms. In the next chapter, to account for the individual power constraint of each sensor, we will generalize the sparsity-promoting optimization framework by introducing a new sparsity-inducing penalty

function to avoid scenarios in which the same sensors are successively selected.
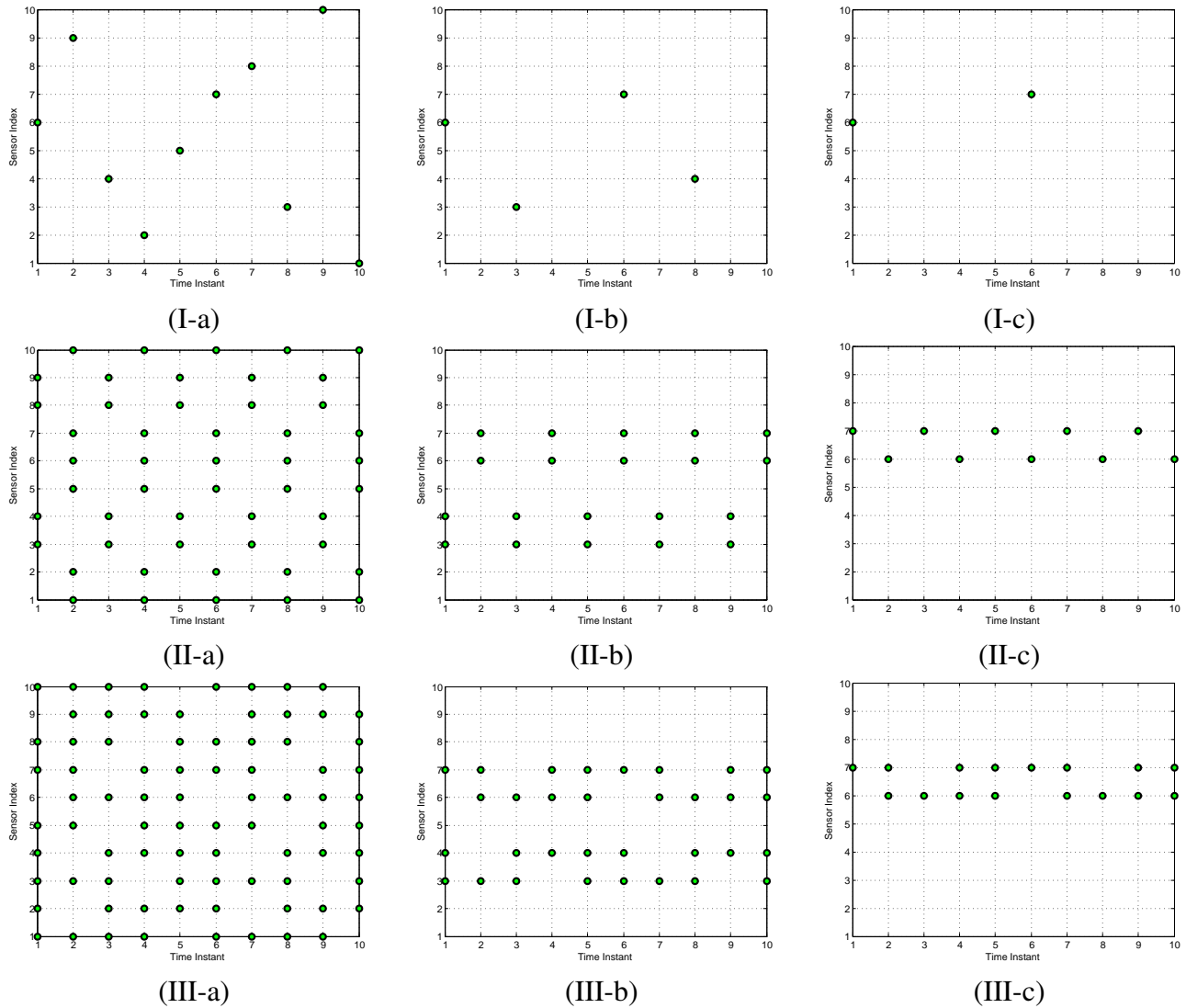
Fig. 4.5: I- Sensor scheduling schemes with measurement frequency bound $\eta = 1$: (I-a) $\gamma = 0$, (I-b) $\gamma = 0.1$, (I-c) $\gamma = 0.15$. II- Sensor scheduling schemes with measurement frequency bound $\eta = 5$: (II-a) $\gamma = 0$, (II-b) $\gamma = 0.1$, (II-c) $\gamma = 0.15$. III- Sensor scheduling schemes with measurement frequency bound $\eta = 8$: (III-a) $\gamma = 0$, (III-b) $\gamma = 0.1$, (III-c) $\gamma = 0.15$.

CHAPTER 5

ENERGY-AWARE SPARSE SENSOR MANAGEMENT

5.1 Introduction

Promoting sparsity in sensor management may lead to scenarios in which a fixed set of sensors, which we hereafter refer to as the most 'informative sensors', are successively selected, e.g., due to their larger mutual information with the target [41] or stronger correlation with the field point of interest [42]. This behavior would result in faster energy depletion of the most informative sensors. From the perspective of network life time [43], the death of a sensor (or a percentage of sensors) can make the network nonfunctional. Therefore, it is desired to have a balanced use of sensors while discouraging the excessive use of any group of sensors in the network.

In this chapter, we propose a new sparsity-promoting penalty function, which penalizes successive selection of the same group of sensors. This framework generates sparse sensor schedules which achieve a trade-off between activating the most informative sensors and balancing the energy consumption in the network. The remainder of the chapter is organized as follows. In Section 5.2, we motivate the importance of sensor scheduling from an energy balance point of view. In Section 5.3, we present a quadratic function, with respect to the number of times that each sensor is

selected over a time horizon, to penalize the overuse of the same group of sensors. In Section 5.4, we solve an ℓ_2-norm optimization problem to find sparse sensor schedules in an energy-balanced manner. In Section 5.5, a reweighted ℓ_1 norm is used to characterize the sparsity. In Section 5.6, we provide numerical results and comparison with other sensor scheduling algorithms in the literature to demonstrate the effectiveness of our approach. We summarize our work in Section 5.7.

5.2 Motivation

Consider a generic system where multiple sensors are deployed to monitor a spatio-temporally correlated random field. Measurements from these multiple sensors at different time instants are used to estimate the field intensity at an unobserved location over multiple time steps. Let $x(\mathbf{s}, t)$ denote the field intensity at location $\mathbf{s}$ and time instant t, and with mean μ and variance σ^2. The vector of field intensities to be estimated is given by

$$\mathbf{x} = [x(\varsigma, \tau_1), x(\varsigma, \tau_2), \ldots, x(\varsigma, \tau_N)]^T, \tag{5.1}$$

where ς is the location where the field intensity is to be estimated at time instants $\{\tau_n\}_{n=1,2,\ldots,N}$.

The measurement of the mth sensor at the sampling time t_k, denoted by $y_{k+(m-1)K}$, is given by

$$y_{k+(m-1)K} = x(\mathbf{s}_m, t_k) + v_{k+(m-1)K}, \; k = 1, 2, \ldots, K, \; m = 1, 2, \ldots, M, \tag{5.2}$$

where K is the number of samples per sensor, M is the number of sensors, $\mathbf{s}_m$ denotes the location of the mth sensor, and $v_{k+(m-1)K}$ is a zero-mean Gaussian noise with variance σ_v^2.

Based on sensor measurements, we employ the LMMSE to estimate $\mathbf{x}$

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{y} + \mathbf{a} = \sum_{m=1}^{M} \sum_{k=1}^{K} y_{k+(m-1)K} \mathbf{w}_{k+(m-1)K} + \mathbf{a}, \tag{5.3}$$

where $\mathbf{W} \in \mathbb{R}^{N \times KM}$ is an unknown coefficient matrix determined by the minimum mean square

error criterion, $\mathbf{a} = \mu(\mathbf{I} - \mathbf{W})\mathbf{1}$ to ensure unbiasedness, $\mathbf{y} = [y_1, y_2, \ldots, y_{KM}]^T$, and $\mathbf{w}_{k+(m-1)K}$ denotes the jth column of $\mathbf{W}$.

It is clear from (5.3) that the non-zero columns of $\mathbf{W}$ characterize the selected sensor measurements. For example, if only the ith sensor reports measurement at time t_j. It follows from (5.3) that $\mathbf{Wy} = y_{j+(i-1)K}\mathbf{w}_{j+(i-1)K}$. To seek an optimal tradeoff between estimation accuracy and sensor activations, a sparsity-promoting optimization problem has been proposed in [16, 40, 89]

$$\underset{\mathbf{w}}{\text{minimize}} \quad J(\mathbf{w}) + \gamma h(\mathbf{w}), \tag{5.4}$$

where $\mathbf{w}$ is the columnwise vector of $\mathbf{W}$, $J(\mathbf{w})$ denotes the estimation distortion with respect to the columnwise vector of $\mathbf{W}$, which is commonly given by a convex quadratic function for proper coefficients [90, Sec. II & Lemma 1]

$$J(\mathbf{w}) = (1/2)\mathbf{w}^T\mathbf{P}\mathbf{w} + \mathbf{q}^T\mathbf{w}, \tag{5.5}$$

and $h(\mathbf{w})$ gives the number of nonzero columns of $\mathbf{W}$

$$h(\mathbf{w}) = \sum_{m=1}^{M} \sum_{k=1}^{K} \|\|\mathbf{w}_{k,m}\|_p\|_0. \tag{5.6}$$

Here for notational simplicity, we replace $\mathbf{w}_{k+(m-1)K}$ with $\mathbf{w}_{k,m}$, $\|x\|_0 = 1$ if $x \neq 0$, and $\|x\|_0 = 0$ if $x = 0$, and typically $p = 1$ or 2. In problem (5.4), $\gamma \in \mathbb{R}_+$ is a sparsity-promoting parameter since sparser sensor schedules can be achieved by making γ larger. As will be evident later, our framework can be viewed as a generalization of (5.4).

The total number of selected sensors in the solution of problem (5.4) decreases when γ increases, thereby promoting sparsity in sensor activations. However, there exist scenarios in which a fixed subset of sensors (the most informative ones) will be successively selected, and thus result in unbalanced energy usage among sensors. We demonstrate this phenomenon in Fig. 5.1 (details provided in Section 5.6). As we can see, for $\gamma = 0.1$ all the sensors are selected over all the ten
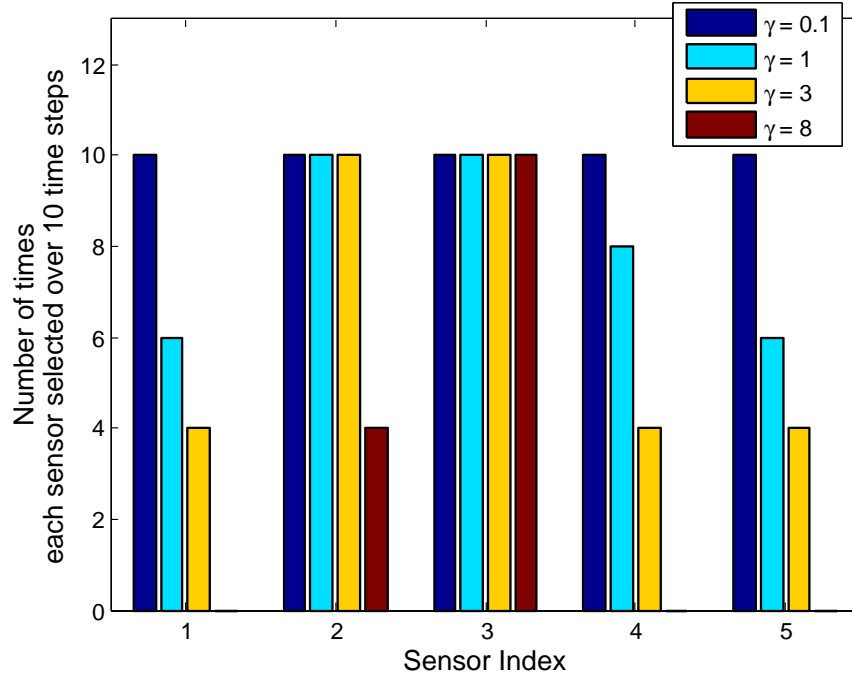
Fig. 5.1: Sensor schedules obtained by the conventional sparsity-promoting sensor management for different values of γ.

time steps while for $\gamma = 8$ only sensor 2 and sensor 3 are activated. The successive use of sensor 2 and sensor 3 leads to an unbalanced energy usage over the entire network. In what follows, we will present a new sparsity-promoting framework that achieves a balance between activating the most informative sensors and uniformly allocating sensor energy over the entire network.

5.3 Energy-aware sparsity-promoting regularization

In this section, we introduce a quadratic function, with respect to the number of times that each sensor is selected over a time horizon, to penalize the overuse of the same group of sensors. After relaxing the ℓ_0 norm, we propose a convex optimization problem to find sparse sensor schedules in an energy-balanced manner.

Consider a new sparsity-promoting function which characterizes the 'cost' of each sensor being

repeatedly selected,

$$g(\mathbf{w}) = \sum_{m=1}^{M} \kappa_m^2, \ \kappa_m = \sum_{k=1}^{K} \|\|\mathbf{w}_{k,m}\|_p\|_0 \tag{5.7}$$

where the quantity κ_m is the number of times the mth sensor is selected over K time steps. The rationale behind using a quadratic function κ_m^2 is its relatively fast growth as a function of κ_m. Consider a system with $M = 2$ sensors. It is clear from (5.7) that the penalty value of using the first sensor 4 times and the second sensor 0 times ($4^2 + 0^2 = 16$ units) is greater than the penalty of using each sensor 2 times ($2^2 + 2^2 = 8$ units).

Based on (5.7), we modify the formulation (5.4) by penalizing successive selections of sensors,

$$\underset{\mathbf{w}}{\text{minimize}} \quad J(\mathbf{w}) + \gamma h(\mathbf{w}) + \eta g(\mathbf{w}), \tag{5.8}$$

where $\gamma \in \mathbb{R}_+$ and $\eta \in \mathbb{R}_+$ are regularization parameters. In problem (5.8), sparser sensor schedules can be achieved by making either γ or η larger. For large γ, the resulting sparse sensor schedule may contain less total number of active sensors, but some sensors will be selected more frequently. Conversely, the sparse sensor schedule resulting from large η may include more number of total active sensors, but the set of selected sensors is more diverse, which leads to the balanced usage of sensor energy in the network. When $\eta = 0$, problem (5.8) reduces to problem (5.4).

Problem (5.8) is not convex due to the presence of the ℓ_0 norm. For tractability, we relax the ℓ_0 norm to an ℓ_1 norm [91],

$$\underset{\mathbf{w}}{\text{minimize}} \ J(\mathbf{w}) + \gamma \sum_{m=1}^{M} \sum_{k=1}^{K} \|\mathbf{w}_{k,m}\|_p + \eta \sum_{m=1}^{M} \left(\sum_{k=1}^{K} \|\mathbf{w}_{k,m}\|_p \right)^2. \tag{5.9}$$

where different choices of ℓ_p norm will lead to different approximations of problem (5.8).

5.4 ℓ_2 optimization: convexity and solution

In problem (5.9) with $p = 2$, an ℓ_2 norm is used to characterize the column-sparsity of $\mathbf{W}$,

$$\underset{\mathbf{w}}{\text{minimize}} \ J(\mathbf{w}) + \gamma \sum_{m=1}^{M} \sum_{k=1}^{K} \|\mathbf{w}_{k,m}\|_2 + \eta \sum_{m=1}^{M} \left(\sum_{k=1}^{K} \|\mathbf{w}_{k,m}\|_2 \right)^2, \tag{5.10}$$

where the use of ℓ_2 norm is well motivated by the problem of group Lasso [92].

We begin by showing the convexity of problem (5.10) in Proposition 5.1.

Proposition 5.1. *Problem* (5.10) *is a convex optimization problem.*

Proof: see Appendix A.8. ∎

Next, we develop an ADMM-based algorithm to find the optimal solution of the convex optimization problem (5.10). The major advantage of ADMM is that it allows us to split the optimization problem (5.10) into a sequence of subproblems, each of which can be solved analytically. We will elaborate on ADMM in the following.

We begin by reformulating the optimization problem (5.10) in a way that lends itself to the application of ADMM,

$$\underset{\mathbf{w},\mathbf{v},\mathbf{u}}{\text{minimize}} \ J(\mathbf{w}) + \gamma \sum_{m=1}^{M} \sum_{k=1}^{K} h_{k,m}(\mathbf{v}) + \eta \sum_{m=1}^{M} g_m(\mathbf{u})$$

$$\text{subject to} \quad \mathbf{w} = \mathbf{v}, \ \mathbf{w} = \mathbf{u}. \tag{5.11}$$

where $h_{k,m}(\mathbf{v}) := \|\mathbf{v}_{k,m}\|_2$, and $g_m(\mathbf{u}) := \left(\sum_{k=1}^{K} \|\mathbf{u}_{k,m}\|_2 \right)^2$. The augmented Lagrangian of (5.11) is given by

$$\mathcal{L}(\mathbf{w},\mathbf{v},\mathbf{u},\boldsymbol{\mu},\boldsymbol{\xi}) = J(\mathbf{w}) + \gamma \sum_{m=1}^{M} \sum_{k=1}^{K} h_{k,m}(\mathbf{v}) + \eta \sum_{m=1}^{M} g_m(\mathbf{u}) + \boldsymbol{\mu}^T(\mathbf{w} - \mathbf{v}) + \boldsymbol{\eta}^T(\mathbf{w} - \mathbf{u})$$

$$+ \frac{\rho}{2}\|\mathbf{w} - \mathbf{v}\|_2^2 + \frac{\rho}{2}\|\mathbf{w} - \mathbf{u}\|_2^2, \tag{5.12}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ are Lagrangian multipliers, and $\rho \in \mathbb{R}_+$ is a penalty weight.

The ADMM algorithm iteratively executes the following steps [82] for iteration $l = 1, 2, \ldots$

$$\mathbf{w}^{l+1} = \arg\min_{\mathbf{w}} \left\{ J(\mathbf{w}) + \frac{\rho}{2}\|\mathbf{w} - \mathbf{a}_1^l\|_2^2 + \frac{\rho}{2}\|\mathbf{w} - \mathbf{a}_2^l\|_2^2 \right\} \tag{5.13}$$

$$\mathbf{v}^{l+1} = \arg\min_{\mathbf{v}} \left\{ \gamma\sum_{m=1}^{M}\sum_{k=1}^{K} h_{k,m}(\mathbf{v}) + \frac{\rho}{2}\|\mathbf{v} - \mathbf{b}^l\|_2^2 \right\} \tag{5.14}$$

$$\mathbf{u}^{l+1} = \arg\min_{\mathbf{v}} \left\{ \eta\sum_{m=1}^{M} g_m(\mathbf{u}) + \frac{\rho}{2}\|\mathbf{u} - \mathbf{c}^l\|_2^2 \right\}, \tag{5.15}$$

$$\boldsymbol{\mu}^{l+1} = \boldsymbol{\mu}^l + \rho(\mathbf{w}^{l+1} - \mathbf{v}^{l+1}), \; \boldsymbol{\xi}^{l+1} = \boldsymbol{\xi}^l + \rho(\mathbf{w}^{l+1} - \mathbf{u}^{l+1}), \tag{5.16}$$

where $\mathbf{a}_1^l = \mathbf{v}^l - \frac{1}{\rho}\boldsymbol{\mu}^l$, $\mathbf{a}_2^l = \mathbf{v}^l - \frac{1}{\rho}\boldsymbol{\xi}^l$, $\mathbf{b}^l = \mathbf{w}^{l+1} + \frac{1}{\rho}\boldsymbol{\mu}^l$, and $\mathbf{c}^l = \mathbf{w}^{l+1} + \frac{1}{\rho}\boldsymbol{\xi}^t$. The ADMM algorithm terminates until

$$\|\mathbf{w}^{l+1} - \mathbf{v}^{l+1}\|_2^2 + \|\mathbf{w}^{l+1} - \mathbf{u}^{l+1}\|_2^2 \le \epsilon, \;\; \|\mathbf{v}^{l+1} - \mathbf{v}^l\|_2^2 + \|\mathbf{u}^{l+1} - \mathbf{u}^l\|_2^2 \le \epsilon,$$

where ϵ is a stopping tolerance. In what follows, we will derive the analytical solutions of '$\mathbf{w}$-minimization' problem (5.13), '$\mathbf{v}$-minimization' problem (5.14), and '$\mathbf{u}$-minimization' problem (5.15).

$\mathbf{w}$-*minimization*

Completing squares with respect to $\mathbf{w}$ in (5.12), the $\mathbf{w}$-minimization problem (5.13) is given by an unconstrained quadratic program (UQP)

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{2}\mathbf{w}^T(\mathbf{P} + 2\rho\mathbf{I})\mathbf{w} - (\rho\mathbf{a}_1 + \rho\mathbf{a}_2 - \mathbf{q})^T\mathbf{w}, \tag{5.17}$$

where $J(\mathbf{w})$ is given by (5.5), and for simplicity, we have omitted the ADMM iteration index l. The minimizer of problem (5.13) is given by $\mathbf{w} = (\mathbf{P} + 2\rho\mathbf{I})^{-1}(\rho\mathbf{a}_1 + \rho\mathbf{a}_2 - \mathbf{q})$.

v-*minimization*

Completing squares with respect to $\mathbf{v}$ in (5.12), the $\mathbf{v}$-minimization problem (5.14) becomes

$$\underset{\mathbf{v}}{\text{minimize}} \ \sum_{m=1}^{M} \sum_{k=1}^{K} \|\mathbf{v}_{k,m}\|_2 + \frac{1}{2\hat{\gamma}} \|\mathbf{v} - \mathbf{b}^l\|_2^2,$$

where $\hat{\gamma} = \gamma/\rho$. The solution of problem (5.14) is then given by a block soft thresholding operator [93]

$$\mathbf{v}_{k,m}^{l+1} = \begin{cases} (1 + \frac{\hat{\gamma}}{\|\mathbf{b}_{k,m}^l\|_2})\mathbf{b}_{k,m}^l & \|\mathbf{b}_{k,m}^l\|_2 \geq \hat{\gamma} \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{5.18}$$

for $k = 1, 2, \ldots, K$ and $m = 1, 2, \ldots, M$, where $\mathbf{b}_{k,m}^l$ is the $(k + mK - K)$th subvector of $\mathbf{b}^l$.

u-*minimization*

Completing squares with respect to $\mathbf{u}$ in (5.12), the $\mathbf{u}$-minimization problem (5.15) is

$$\underset{\mathbf{u}}{\text{minimize}} \ \eta \sum_{m=1}^{M} \left(\sum_{k=1}^{K} \|\mathbf{u}_{k,m}\|_2 \right)^2 + \frac{\rho}{2} \|\mathbf{u} - \mathbf{c}^l\|_2^2, \tag{5.19}$$

which can be decomposed into M subproblems

$$\underset{\{\mathbf{u}_{k,m}\}_{k=1,2,\ldots,K}}{\text{minimize}} \ \left(\sum_{k=1}^{K} \|\mathbf{u}_{k,m}\|_2 \right)^2 + \frac{1}{2\hat{\eta}} \sum_{k=1}^{K} \|\mathbf{u}_{k,m} - \mathbf{c}_{k,m}\|_2^2 \tag{5.20}$$

for $m = 1, 2, \ldots, M$, where $\hat{\eta} = \eta/\rho$, and we use $\mathbf{c}_{k,m}$ instead of $\mathbf{c}_{k,m}^l$ for ease of notation. The quadratic function with respect to the sum of ℓ_2 norms makes finding the minimizer of problem (5.20) more challenging, since it is a nonlinear composition of nonsmooth functions. However, Proposition 5.2 will demonstrate that the solution of (5.20) is achievable via quadratic programming (QP).

Proposition 5.2. *The minimizer of problem* (5.20) *is given by*

$$\mathbf{u}^*_{k,m} = r^*_k \frac{\mathbf{c}^T_{k,m}}{\|\mathbf{c}^T_{k,m}\|_2}, \quad k = 1, 2, \ldots, K. \tag{5.21}$$

In (5.23), $\mathbf{r}^* := [r^*_1, r^*_2, \ldots, r^*_K]^T$ *is the minimizer of QP*

$$\begin{aligned}
\underset{\mathbf{r}}{minimize} \quad & \mathbf{r}^T \mathbf{1}\mathbf{1}^T \mathbf{r} + \tfrac{1}{2\hat{\eta}}\|\mathbf{r} - \mathbf{f}\|_2^2, \\
subject\ to \quad & \mathbf{r} \geq \mathbf{0},
\end{aligned} \tag{5.22}$$

where $\mathbf{f} = [\|\mathbf{c}^T_{1,m}\|_2, \|\mathbf{c}^T_{2,m}\|_2, \ldots, \|\mathbf{c}^T_{K,m}\|_2]^T$.

   Proof: See Appendix A.9. ∎

   Proposition 5.2 indicates that the solution of the $\mathbf{u}$-minimization problem can be found through M QPs, each of which has complexity $O(K^{3.5})$ [62]. However, instead of using QP, we will show that the analytical solution of (5.22) is also tractable. We first present an important feature of problem (5.22) in Proposition 5.3.

Proposition 5.3. *If the entries of* $\mathbf{f}$ *satisfy* $f_1 \leq f_2 \leq \ldots \leq f_K$, *then the solution of* (5.22) *yields* $r^*_1 \leq r^*_2 \leq \ldots \leq r^*_K$.

   Proof: The proof is straightforward, proceeding by contradiction. ∎

   The rationale behind Proposition 5.3 is that the analytical solution of problem (5.22) is tractable by sorting $\mathbf{f}$ in an ascending order. Let $\mathcal{I}$ denote the index set that describes the rearrangement of $\{f_i\}_{i=1,2,\ldots,K}$ in an ascending order. And let $\mathbf{r}^*_{\mathcal{I}}$ be the solution of (5.22) rearranged by $\mathcal{I}$. For instance, if $\mathbf{f} = [5, 4]^T$, then $\mathcal{I} = \{2, 1\}$ and $\mathbf{r}^*_{\mathcal{I}} = [r^*_2, r^*_1]^T$. The closed-form of $\mathbf{r}^*_{\mathcal{I}}$ is given in Proposition 5.4. Together with $\mathcal{I}$, we then obtain the solution of (5.22).

Proposition 5.4. *The minimizer of problem* (5.22), *with an index rearrangement* $\mathcal{I}$, *is given by*

$$r^*_{\mathcal{I},i} = \begin{cases} 0 & 1 \leq i \leq \iota - 1,\, i \in \mathbb{N} \\ f_i - \dfrac{2\hat{\eta}\sum_{k=\iota}^K f_k}{1 + 2\hat{\eta}(K - \iota + 1)} & \iota \leq i \leq K,\, i \in \mathbb{N}, \end{cases} \tag{5.23}$$

for $i = 1, 2, \ldots, K$, where $\mathcal{I}$ is the index set that describes the rearrangement of $\{f_i\}_{i=1,2,\ldots,K}$ in an ascending order, f_i is the ith entry of $\mathbf{f}$ given in (5.22), ι is the index of the first positive element in the set of numbers $\left\{ f_i - \frac{2\hat{\eta} \sum_{k=i}^{K} f_k}{1 + 2\hat{\eta}(K - i + 1)} \right\}_{i=1,2,\ldots,K}$ and $\iota = K + 1$ if no positive element exists.

Proof: See Appendix A.10. ∎

According to Proposition 5.4, the complexity of solving problem (5.22) is given by $O(K \log K)$ owing to the sorting of the entries of $\mathbf{f}$ in an ascending order. This is much lower than the complexity of QP with $O(K^{3.5})$. Combining (5.17), (5.18) and (5.23), the ADMM algorithm follows from steps (5.14)–(5.16), each of which yields the analytical solution.

5.5 Weighted ℓ_1 optimization: convexity and solution

In problem (5.9) with $p = 1$, we use a reweighted ℓ_1 minimization method to relax the ℓ_0 norm,

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{2}J(\mathbf{w}) + \gamma \sum_{m=1}^{M} \sum_{k=1}^{K} \alpha_{k,m} \|\mathbf{w}_{k,m}\|_1 + \eta \sum_{m=1}^{M} \left(\sum_{k=1}^{K} \alpha_{k,m} \|\mathbf{w}_{k,m}\|_1 \right)^2, \qquad (5.24)$$

where we denote its solution by $\{\mathbf{w}_{k,m}^*\}$ for $k = 1, 2, \ldots, K$ and $m = 1, 2, \ldots, M$. Based on $\{\mathbf{w}_{k,m}^*\}$, the weights $\{\alpha_{k,m}\}$ with respect to optimization variables $\mathbf{w}_{k,m}$ are updated as $\alpha_{k,m} \leftarrow 1/(\|\mathbf{w}_{k,m}^*\|_1 + \kappa)$ for the next iteration, where the parameter $\kappa > 0$ is a small number to ensure that the denominator does not become zero.

Proposition 5.5 shows that problem (5.24) can be solved by transformation to an QP.

Proposition 5.5. *The optimal solution of problem (5.24) can be obtained by solving an equivalent problem*

$$\underset{\mathbf{w},\mathbf{u}}{\text{minimize}} \quad \frac{1}{2}J(\mathbf{w}) + \gamma \sum_{m=1}^{M} \sum_{k=1}^{K} \alpha_{k,m} \mathbf{1}^T \mathbf{u}_{k,m} + \eta \sum_{m=1}^{M} \left(\sum_{k=1}^{K} \alpha_{k,m} \mathbf{1}^T \mathbf{u}_{k,m} \right)^2,$$

$$\text{subject to} \quad -\mathbf{u}_{k,m} \leq \mathbf{w}_{k,m} \leq \mathbf{u}_{k,m}, \ k = 1, 2, \ldots, K, \ m = 1, 2, \ldots, M, \qquad (5.25)$$

where $\mathbf{u}_{k,m} \in \mathbb{R}^N$ is a vector of introduced auxiliary variables, and the inequality is defined in

element-wise fashion.

Proof: See Appendix A.11. ∎

We further rewrite problem (5.25) as

$$
\begin{aligned}
\underset{\mathbf{w},\mathbf{u}}{\text{minimize}} \quad & (1/2)\mathbf{w}^T\mathbf{P}\mathbf{w} + \mathbf{q}^T\mathbf{w} + \gamma\mathbf{u}^T\mathbf{r} + \eta\mathbf{u}^T\mathbf{R}\mathbf{u} \\
\text{subject to} \quad & -\mathbf{u} \le \mathbf{w} \le \mathbf{u},
\end{aligned}
\tag{5.26}
$$

where $\mathbf{r} := [\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \ldots, \boldsymbol{\alpha}_M^T]^T$, $\boldsymbol{\alpha}_m := [\alpha_{1,m}\mathbf{1}^T, \alpha_{2,m}\mathbf{1}^T, \ldots, \alpha_{K,m}\mathbf{1}^T]^T$ for $m = 1, 2, \ldots, M$, and $\mathbf{R}$ is a block diagonal matrix formed by $\{\boldsymbol{\alpha}_m\boldsymbol{\alpha}_m^T\}_{m=1,2,\ldots,M}$.

Let $\mathbf{z} := \mathbf{C}[\mathbf{w}^T, \mathbf{u}^T]^T$, problem (5.26) simplifies to

$$
\begin{aligned}
\underset{\mathbf{z}}{\text{minimize}} \quad & (1/2)\mathbf{z}^T\mathbf{H}\mathbf{z} - \mathbf{z}^T\mathbf{h} \\
\text{subject to} \quad & \mathbf{z} \le \mathbf{0},
\end{aligned}
\tag{5.27}
$$

where $\mathbf{C} = [\mathbf{I}, -\mathbf{I}; -\mathbf{I}, -\mathbf{I}]$, $\mathbf{A} := [\mathbf{I}, \mathbf{0}]$, $\mathbf{B} := [\mathbf{0}, \mathbf{I}]$, $\mathbf{H} := (\mathbf{A}\mathbf{C}^{-1})^T\mathbf{P}\mathbf{A}\mathbf{C}^{-1} + 2\eta(\mathbf{B}\mathbf{C}^{-1})^T\mathbf{R}\mathbf{B}\mathbf{C}^{-1}$, and $\mathbf{h} := -(\mathbf{A}\mathbf{C}^{-1})^T\mathbf{q} - \gamma(\mathbf{B}\mathbf{C}^{-1})^T\mathbf{r}$.

Problem (5.27) is a convex QP. Convexity holds since $\mathbf{P}$ and $\mathbf{R}$ are positive semidefinite. To solve such an QP, the complexity of using standard solvers such as CVX [94] is roughly $O(L^3)$ per Newton iteration [62], where L denotes the dimension of the optimization variable $\mathbf{z}$. Instead of using an QP solver, we employ an accelerated proximal gradient method, which yields much lower computational complexity per iteration.

We express problem (5.27) as

$$
\underset{\mathbf{z}}{\text{minimize}} \quad \phi(\mathbf{z}) + \psi(\mathbf{z}),
\tag{5.28}
$$

where $\phi(\mathbf{z}) := \frac{1}{2}\mathbf{z}^T\mathbf{H}\mathbf{z} - \mathbf{z}^T\mathbf{h}$, and $\psi(\mathbf{z})$ is an indicator function with respect to the convex set $\{\mathbf{z} \mid \mathbf{z} \le \mathbf{0}\}$

Problems in the form of (5.28) whose objective function consists of a differentiable convex function ϕ and a nonsmooth convex function ψ can be efficiently solved via the accelerated prox-

imal gradient method (APGM) [93]. The advantage of APGM is its low complexity at each iteration. We summarize APGM in Algorithm 5.1.

Algorithm 5.1 APGM-based algorithm for solving (5.28)

Require: Given $\mathbf{z}^{-1} = \mathbf{z}^0$, λ^0, $\varrho = 0.5$ and $\epsilon_{\text{prox}} > 0$.

1: **for** $i = 1, 2, \ldots$ **do**
2:     Let $\lambda := \lambda^{i-1}$ and $\mathbf{s}^i := \mathbf{z}^i + \frac{i}{i+3}(\mathbf{z}^i - \mathbf{z}^{i-1})$.
3:     **repeat**
4:       $\boldsymbol{\chi} := \mathbf{prox}_{\lambda\psi}\left(\mathbf{s}^i - \lambda\nabla\phi(\mathbf{s}^i)\right) = \left(\mathbf{s}^i - \lambda\nabla\phi(\mathbf{s}^i)\right)_-$.
5:     **until** if $\phi(\boldsymbol{\chi}) \leq \hat{\phi}_{\lambda,\mathbf{z}^i}(\boldsymbol{\chi})$ where $\hat{\phi}_{\lambda,\mathbf{z}^i}(\boldsymbol{\chi})$ is given by (5.30); else $\lambda = \varrho\lambda$, go to step 3.
6:     Let $\lambda^i := \lambda$ and $\mathbf{z}^{i+1} := \boldsymbol{\chi}$.
7:     **until** $\|\phi(\mathbf{z}^{i+1}) - \phi(\mathbf{z}^i)\| < \epsilon_{\text{prox}}$.
8: **end for**

In Step 4, the proximal operator $\mathbf{prox}_{\lambda\psi}(\cdot)$ of ψ with parameter λ is given by

$$\mathbf{prox}_{\lambda\psi}(\cdot) = \arg\min_{\mathbf{z}} \left(\psi(\mathbf{z}) + \frac{1}{2\lambda}\|\mathbf{z} - \cdot\|_2^2\right). \tag{5.29}$$

By recalling the definition of $\psi(\mathbf{z})$, the proximal operator (5.29) precisely defines a Euclidean projection onto the halfspace $\{\mathbf{z}|\mathbf{z} \leq 0\}$. Namely, $\mathbf{prox}_{\lambda\psi}(\cdot) = (\cdot)_-$, where $(\mathbf{z})_-$ takes non-positive elements of a vector $\mathbf{z}$ and sets 0 for its positive elements. In Step 5, the function $\hat{\phi}_{\lambda,\mathbf{z}^i}(\boldsymbol{\chi})$ is defined by

$$\hat{\phi}_{\lambda,\mathbf{z}^i}(\boldsymbol{\chi}) := \phi(\mathbf{z}^i) + \nabla\phi(\mathbf{z}^i)^T(\boldsymbol{\chi} - \mathbf{z}^i) + \frac{1}{2\lambda}\|\boldsymbol{\chi} - \mathbf{z}^i\|_2, \tag{5.30}$$

where $\nabla\phi(\mathbf{z}^i) = \tilde{\mathbf{H}}\mathbf{z}^i - \tilde{\mathbf{h}}$. In (5.30), $\hat{\phi}_{\lambda,\mathbf{z}^i}(\boldsymbol{\chi})$ can be interpreted as a quadratic approximation of $\phi(\boldsymbol{\chi})$ [95]. Steps 3-5 constitute a procedure of backtracking line search [59], which is used to determine an appropriate step size λ at iteration i.

At each iteration of Algorithm 5.1, the computation cost is dominated by matrix-vector multiplication, which has $O(L^2)$ complexity [96]. The total number of iterations required for APGM is approximated by $O(1/\sqrt{\epsilon})$ [93]. In our implementation, it takes around 100 iterations as $\epsilon_{\text{prox}} = 10^{-4}$.

5.6 Numerical results

We consider a sensor network with $M \in \{5, 10\}$ sensors on a 20×20 grid, where each sensor takes $K = 10$ measurements at the sampling time $0.2k$ for $k = 1, 2, \ldots, K$. The task of the sensor network is to reconstruct the field intensities at the coordinate $\varsigma = (10, 10)$ over time slots $\{\tau_n\}_{n=1,2,\ldots,N}$, where $\tau_n = 0.2n + 0.1$ and $N = 9$. We assume that the correlation model of the random field is given by $\mathrm{cov}(x(\mathbf{s}, t), x(\mathbf{s}', t')) = \sigma^2 \exp\{-c_s \|\mathbf{s} - \mathbf{s}'\|_2 - c_t(t - t')^2\}$, where $\sigma^2 = 1$, $c_s = 0.1$, and $c_t = 0.1$. We select the ADMM parameters as $\rho = 100$ and $\epsilon = 10^{-3}$.

In what follows, we define a factor to measure the energy imbalance in the usage of each sensor due to successive selections. Let $\boldsymbol{\omega} = [\omega_1, \omega_2, \ldots, \omega_M] \in \mathbb{N}^M$ denote the sensor activation scheme of M sensors over K time steps, where $\omega_i \in \{0, 1, \ldots, K\}$ represents the number of times the ith sensor is selected over K time steps, and $\mathbf{1}^T \boldsymbol{\omega}$ gives the total number of sensor activations. For example, we consider a specific sensor schedule which satisfies $\omega_1 \geq \omega_2 \geq \ldots \geq \omega_M$. From the perspective of energy balance, we expect to select sensors as uniformly as possible over K time steps. The most balanced schedule is given by $\bar{\boldsymbol{\omega}} := [\omega_0 + \bar{a}_1, \omega_0 + \bar{a}_2, \ldots, \omega_0 + \bar{a}_M]^T$, where $\omega_0 = \lfloor \frac{\mathbf{1}^T \boldsymbol{\omega}}{M} \rfloor$ yields the maximum number of times each sensor is uniformly scheduled given the total number of sensor activations, and

$$
\bar{a}_i = \begin{cases} 1 & \text{if } i = 1, 2, \ldots, (\mathbf{1}^T \boldsymbol{\omega} - M\omega_0) \\ 0 & \text{otherwise,} \end{cases}
$$

which ensures that $\mathbf{1}^T \bar{\boldsymbol{\omega}} = \mathbf{1}^T \boldsymbol{\omega}$. Based on $\boldsymbol{\omega}$ and $\bar{\boldsymbol{\omega}}$, we adopt the distance $d_{\mathrm{im}} = \|\boldsymbol{\omega} - \bar{\boldsymbol{\omega}}\|_2$ to measure the energy imbalance for the schedule $\boldsymbol{\omega}$ with respect to the balanced schedule $\bar{\boldsymbol{\omega}}$. In our numerical examples, a normalized value of d_{im} is used for multiple simulation trials. Specifically, let $d_{\mathrm{im}}^{(n)}$ denote the distance between $\boldsymbol{\omega}^{(n)}$ and $\bar{\boldsymbol{\omega}}^{(n)}$ associated with the sensor schedule for the nth simulation, where $n \in \{1, 2, \ldots, N_{\mathrm{sim}}\}$. We can then define an *energy imbalance measure (EIM)*

as follows,

$$\rho_{\text{im}}^{(n)} = \frac{d_{\text{im}}^{(n)} - \underline{d}}{\bar{d} - \underline{d}} \in [0, 1], \qquad (5.31)$$

where $\underline{d}$ and $\bar{d}$ denote the minimum and maximum value of $\{d_{\text{im}}^{(n)}\}_{n=1,2,\ldots,N_{\text{sim}}}$, respectively. Clearly, if $\rho_{\text{im}}^{(n)} < \rho_{\text{im}}^{(m)}$, the sensor schedule $\boldsymbol{\omega}^{(n)}$ yields a more balanced energy usage.
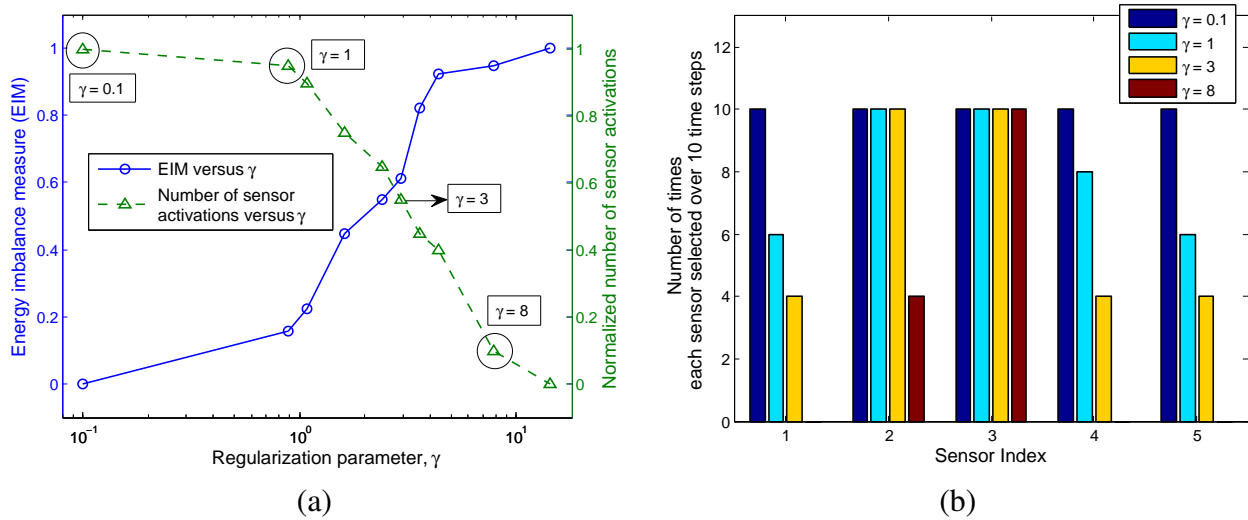
Fig. 5.2: Conventional sparsity-promoting sensor management by varying γ: (a) EIM and total number of sensor activations; (b) Sensor schedules.
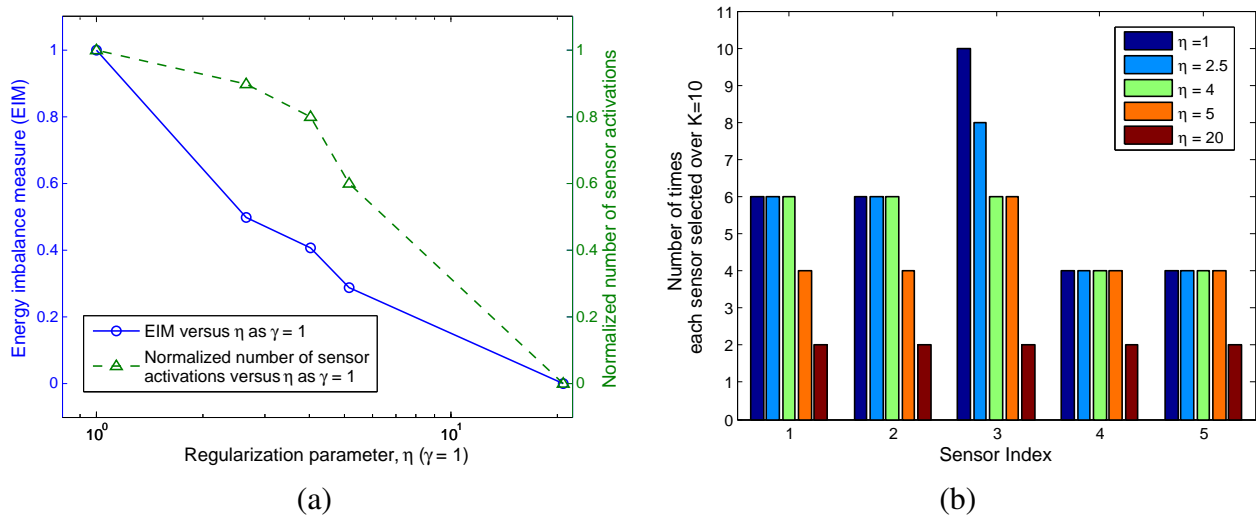
Fig. 5.3: Sparsity-promoting sensor management from an energy balanced point of view by varying η: (a) EIM and total number of sensor activations; (b) Sensor schedules.

In Fig. 5.2, we address the conventional sensor scheduling problem (5.4) and present the resulting sensor schedules for a small network with $M = 5$ sensors. Fig. 5.2-(a) presents EIM and the total number of sensor activations as functions of regularization parameter γ. Here the total number of sensor activations is normalized over $N_{\text{sim}} = 10$ numerical trials, each of which yields the solution of (5.4) for a given value of γ. We observe that when γ increases, EIM increases although less sensors are activated. This is not surprising since sensors with strong correlation to the field point of interest are successively selected. The specific sensor activation schemes that correspond to the marked values of γ are shown in Fig. 5.2-(b). For an extreme case of $\gamma = 8$, only the 2nd and 3rd sensors are used, but the 3rd sensor is activated at every time step.
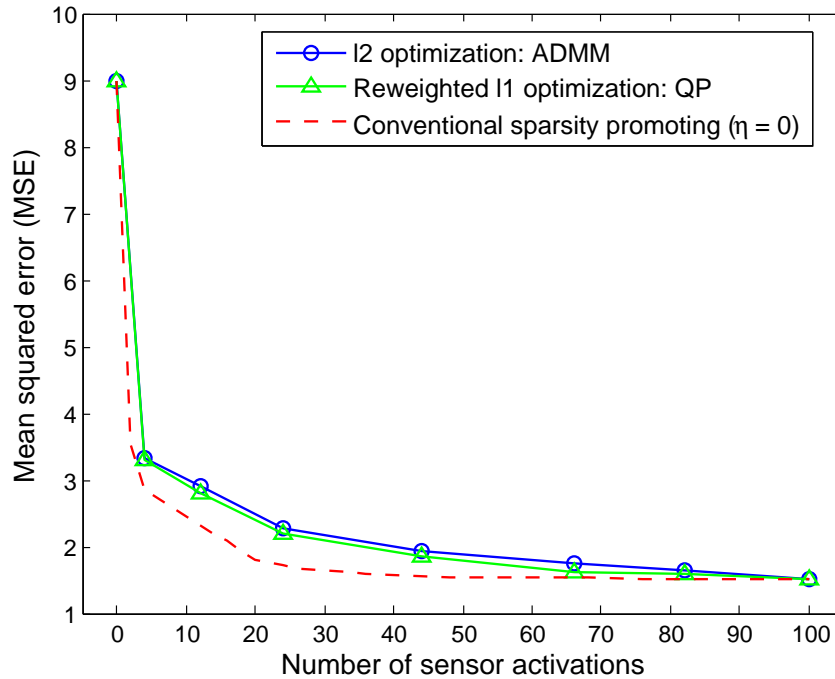
Fig. 5.4: Estimation performance comparison for different sensor scheduling algorithms.

In Fig. 5.3, we present solutions of our proposed sensor scheduling problem (5.10) for different values of η at $\gamma = 1$. In Fig. 5.3-(a), we observe that both EIM and the number of sensor activations decrease when η increases. This is in contrast to the result shown in Fig. 5.2, where EIM increased when the sparsity of sensor schedules was promoted. The energy-balanced sensor schedule is obtained since the new sparsity-promoting penalty function (5.7) enforces sensor energy to be

consumed in a balanced way. In Fig. 5.3-(b), we present the specific sensor schedules for different values of η. As we can see, when η becomes larger than γ, sensors are selected as uniformly as possible. This is because a larger η places more emphasis on minimizing the new sparsity-promoting function, which discourages the successive selection of the most informative sensors.

In Fig. 5.4, we present the mean squared error (MSE) of field estimation based on the new formulation as a function of the number of sensor activations. In this example, we consider a network with $M = 10$ sensors. For sensor scheduling, we employ three different methods: ℓ_2 optimization via ADMM with $\gamma = 0$, reweighted ℓ_1 optimization via QP with $\gamma = 0$, and the conventional method to solve problem (5.10) with $\eta = 0$. First, we observe that when the new penalty on energy imbalance is considered, our approach yields a higher MSE compared to the conventional sparsity-promoting method. This is because the most informative sensors can be successively selected when the energy-balance concern is ignored. Second, we observe that the solution of ℓ_2 optimization problem leads to a slightly higher MSE than that of the reweighted ℓ_1 method. This is due to the fact that a better proxy of the cardinality function (i.e., the reweighted ℓ_1 norm [97]) is used to solve the ℓ_0 optimization problem (5.8) compared to an unweighted ℓ_2 formulation.

5.7   Summary

In this chapter, a novel sparsity-promoting function was introduced to penalize successive selections of the same group of sensors. With the aid of the proposed penalty function, sensor scheduling was performed in an energy-balanced manner. We demonstrated that the resulting ℓ_2 and ℓ_1 optimization problems are convex, and they can be efficiently solved via ADMM and APGM, respectively. Furthermore, an energy imbalance measure was defined to illustrate the effectiveness of our approach, and comparisons with other sensor scheduling algorithms in the literature were provided. In the next chapter, we will take into account the cost of inter-sensor collaboration in the design of resource management protocols.

CHAPTER 6

SENSOR COLLABORATION FOR ESTIMATION OF STATIC PARAMETERS

6.1 Introduction

Recently, the problem of distributed estimation with inter-sensor collaboration has attracted significant attention, where collaboration refers to the act of sharing measurements with neighboring sensors prior to transmission to a fusion center (FC) [9, 33, 44–46]. The presence of sensor collaboration smooths out the observation noise, thereby improving the quality of the signal and the eventual estimation performance. Most literature on sensor collaboration assumed that there is no cost associated with collaboration, the collaboration topologies are fixed and given in advance, and the only unknowns are the collaboration weights used to combine sensor observations.

In this chapter, we present a tractable optimization framework to solve the collaboration problem with non-zero collaboration cost and unknown collaboration topologies for single-snapshot estimation. We also incorporate energy costs associated with selected sensors while determining the optimal subset of sensors that communicate with the FC. For the joint design of optimal sensor collaboration and selection schemes, we describe collaboration through a collaboration matrix, and associate (a) the cost of sensor collaboration with the number of nonzero entries of a collaboration

matrix (i.e., its overall sparsity), and b) the cost of sensor selection with the number of nonzero rows of the collaboration matrix (i.e., its row-sparsity). We show that there exists a trade-off between sensor selection and sensor collaboration for a given estimation performance.

The rest of the chapter is organized as follows. In Section 6.3, we formulate the information and energy constrained sensor collaboration problems with non-zero collaboration cost. In Section 6.4, we develop an efficient approach to solve the information constrained collaboration problem. In Section 6.5, we study the energy constrained collaboration problem. In Section 6.6, we investigate the issue of joint sensor selection and collaboration. In Section 6.7, we demonstrate the effectiveness of our proposed framework through numerical examples. Finally, in Section 6.8 we summarize our work.

6.2 Preliminaries: model for sensor collaboration

In this section, we introduce a distributed estimation system that involves inter-sensor collaboration. We assume that the task of the sensor network is to estimate a random parameter θ, which has zero mean and variance η^2. In the estimation system, sensors first acquire their raw measurements via a linear sensing model. Then, individual sensors can update their observations through spatial collaboration, which refers to (linearly) combining observations from other sensors. The updated measurements are transmitted through a coherent MAC. Finally, the FC determines a global estimate of θ by using a linear estimator. We show the collaborative estimation system in Fig. 6.1, and in what follows we elaborate on each of its parts.

The linear sensing model is given by

$$\mathbf{x} = \tilde{\mathbf{h}}\theta + \boldsymbol{\epsilon}, \tag{6.1}$$

where $\mathbf{x} = [x_1, \ldots, x_N]^T$ denotes the vector of measurements from N sensors, $\tilde{\mathbf{h}}$ is the vector of observation gains with known second order statistics $\mathbb{E}[\tilde{\mathbf{h}}] = \mathbf{h}$ and $\mathrm{cov}(\tilde{\mathbf{h}}) = \boldsymbol{\Sigma}_h$, and $\boldsymbol{\epsilon}$ represents the vector of zero-mean Gaussian noise with $\mathrm{cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}_\epsilon$.
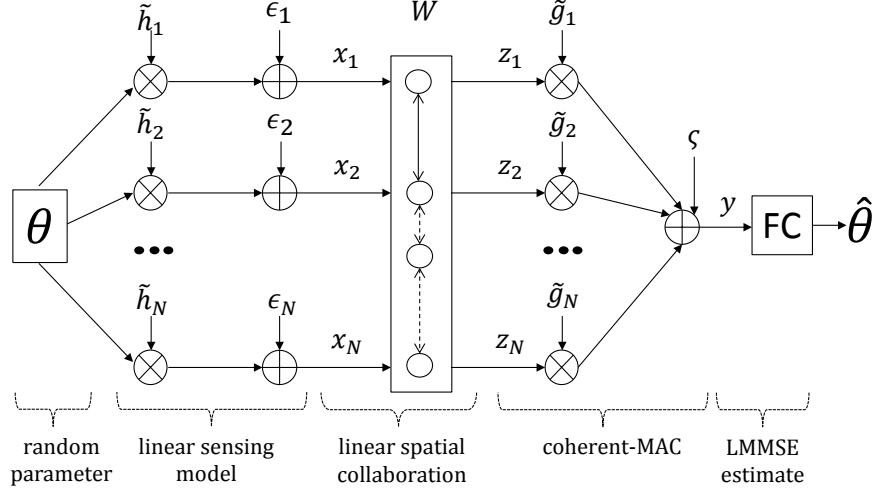
Fig. 6.1: Collaborative estimation architecture showing the sensor measurements, transmitted signals, and the received signal at FC.

The *sensor collaboration* process is described by

$$\mathbf{z} = \mathbf{W}\mathbf{x}, \tag{6.2}$$

where $\mathbf{z} \in \mathbb{R}^N$ denotes the message after collaboration, and $\mathbf{W} \in \mathbb{R}^{N \times N}$ is the collaboration matrix that contains weights used to combine sensor measurements. In (6.2), we assume that sharing of an observation is realized through a reliable (noiseless) communication link that consumes power C_{mn}, regardless of its implementation. And the matrix $\mathbf{C} \in \mathbb{R}^{N \times N}$ describing all the collaboration costs among various sensors is assumed to be known in advance.

After sensor collaboration, the message $\mathbf{z}$ is transmitted to the FC through a coherent MAC, so that the received signal is a coherent sum [30]

$$y = \tilde{\mathbf{g}}^T \mathbf{z} + \varsigma, \tag{6.3}$$

where $\tilde{\mathbf{g}}$ is the vector of channel gains with known second order statistics $\mathbb{E}[\tilde{\mathbf{g}}] = \mathbf{g}$ and $\mathrm{cov}(\tilde{\mathbf{g}}) = \Sigma_g$, and ς is a zero-mean Gaussian noise with variance ξ^2. The transmission cost is given by the

energy required for transmitting the message $\mathbf{z}$ in (6.2), namely,

$$T_{\mathbf{W}} = \mathbb{E}_{\theta,\tilde{\mathbf{h}},\epsilon}[\mathbf{z}^T\mathbf{z}] = \mathbb{E}[\mathbf{x}^T\mathbf{W}^T\mathbf{W}\mathbf{x}] = \mathrm{tr}[\mathbf{W}\boldsymbol{\Sigma}_x\mathbf{W}^T], \tag{6.4}$$

where $\boldsymbol{\Sigma}_x := \mathbb{E}_{\theta,\tilde{\mathbf{h}},\epsilon}[\mathbf{x}\mathbf{x}^T] = \boldsymbol{\Sigma}_\epsilon + \eta^2(\mathbf{h}\mathbf{h}^T + \boldsymbol{\Sigma}_h)$.

We assume that the FC knows the *second-order statistics* of the observation gain, information gain, and additive noises, and that the corresponding variance and covariance matrices are invertible. To estimate the random parameter θ, we consider the LMMSE [58]

$$\hat{\theta} = a_{\mathrm{LMMSE}}y, \tag{6.5}$$

where a_{LMMSE} is determined by the minimum mean square error criterion. From the theory of linear Bayesian estimators [58], we can readily obtain a_{LMMSE} and the corresponding estimation distortion

$$a_{\mathrm{LMMSE}} = \arg\min_a \mathbb{E}[(\theta - ay)^2] = \frac{\mathbb{E}[y\theta]}{\mathbb{E}[y^2]}, \quad \text{and} \tag{6.6a}$$

$$D_{\mathbf{W}} = \mathbb{E}[(\theta - a_{\mathrm{LMMSE}}y)^2] = \eta^2 - \frac{(\mathbb{E}[y\theta])^2}{\mathbb{E}[y^2]}. \tag{6.6b}$$

In (6.6), substituting (6.2) and (6.3), we obtain

$$\mathbb{E}[y^2] = \mathbb{E}[\tilde{\mathbf{g}}^T\mathbf{W}\mathbf{x}\mathbf{x}^T\mathbf{W}^T\tilde{\mathbf{g}}] + \xi^2 = \mathbb{E}[\mathrm{tr}(\tilde{\mathbf{g}}\tilde{\mathbf{g}}^T\mathbf{W}\mathbf{x}\mathbf{x}^T\mathbf{W}^T)] = \mathrm{tr}[\boldsymbol{\Sigma}_{\tilde{g}}\mathbf{W}\boldsymbol{\Sigma}_x\mathbf{W}^T], \tag{6.7}$$

where $\boldsymbol{\Sigma}_{\tilde{g}} := \mathbb{E}[\tilde{\mathbf{g}}\tilde{\mathbf{g}}^T] = \mathbf{g}\mathbf{g}^T + \boldsymbol{\Sigma}_g$, and $\boldsymbol{\Sigma}_x$ was defined in (6.4). Moreover, it is easy to show that

$$\mathbb{E}[y\theta] = \eta^2\mathbf{g}^T\mathbf{W}\mathbf{h}. \tag{6.8}$$

Now, the coefficient of the LMMSE a_{LMMSE} and the corresponding estimation distortion $D_{\mathbf{W}}$ are determined according to (6.6), (6.7) and (6.8).

We define an *equivalent* Fisher information $J_{\mathbf{W}}$ which is monotonically related to $D_{\mathbf{W}}$,

$$J_{\mathbf{W}} := \frac{1}{D_{\mathbf{W}}} - \frac{1}{\eta^2} = \frac{(\mathbf{g}^T\mathbf{W}\mathbf{h})^2}{\text{tr}[\mathbf{\Sigma}_{\tilde{g}}\mathbf{W}\mathbf{\Sigma}_x\mathbf{W}^T] - \eta^2(\mathbf{g}^T\mathbf{W}\mathbf{h})^2 + \xi^2}, \tag{6.9}$$

For convenience, we often express the estimation distortion (6.6b) as a function of the Fisher information

$$D_{\mathbf{W}} := \frac{\eta^2}{1 + \eta^2 J_{\mathbf{W}}}. \tag{6.10}$$

6.3 Optimal sparse sensor collaboration

In this section, we begin by making an association between the collaboration topology and the sparsity structure of the collaboration matrix $\mathbf{W}$. We then define the collaboration cost and sensor selection cost with the help of the cardinality function (also known as the ℓ_0 norm). For simplicity of presentation, we concatenate the entries of $\mathbf{W}$ into a vector, and present the main optimization problems for sensor collaboration.

Recalling the collaboration matrix $\mathbf{W}$ in (6.2), we note that the nonzero entries of $\mathbf{W}$ correspond to the active collaboration links among sensors. For example, $W_{mn} = 0$ indicates the absence of a collaboration link from the nth sensor to the mth sensor, where W_{mn} is the (m,n)th entry of $\mathbf{W}$. Conversely, $W_{mn} \neq 0$ signifies that the nth sensor shares its observation with the mth sensor. Thus, the *sparsity structure* of $\mathbf{W}$ characterizes the *collaboration topology*.

For a given collaboration topology, the *collaboration cost* is given by

$$Q_{\mathbf{W}} = \sum_{m=1}^{N}\sum_{n=1}^{N} C_{mn}\text{card}(W_{mn}). \tag{6.11}$$

where C_{mn} is the cost of sharing an observation from the nth sensor with the mth sensor. Note that $C_{nn} = 0$, since each node can collaborate with itself at no cost. To account for an active

collaboration link, we use the cardinality function

$$\text{card}(W_{mn}) = \begin{cases} 0 & W_{mn} = 0 \\ 1 & W_{mn} \neq 0. \end{cases} \tag{6.12}$$

Next, we define the sensor selection/activation cost. Partitioning the matrix $\mathbf{W}$ rowwise, the linear spatial collaboration process (6.2) can be written as

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix} = \mathbf{W}\mathbf{x} = \begin{bmatrix} \boldsymbol{\omega}_1^T \\ \boldsymbol{\omega}_2^T \\ \vdots \\ \boldsymbol{\omega}_N^T \end{bmatrix} \mathbf{x}, \tag{6.13}$$

where $\boldsymbol{\omega}_n^T$ is the nth row vector of $\mathbf{W}$. It is clear from (6.13) that the *non-zero rows* of $\mathbf{W}$ characterize the selected sensors that communicate with the FC. Suppose, for example, that only the nth sensor communicates with the FC. In this case, it follows from (6.13) that $z_n = \boldsymbol{\omega}_n^T \mathbf{x}$ and $\boldsymbol{\omega}_m = 0$ for $m \neq n$. The goal of sensor selection is to find the best *subset* of sensors to communicate with the FC. This is in contrast to the existing work [44, 45, 48], where the communicating sensors are selected *a priori*.

The sensor *selection cost* can be defined through the row-wise cardinality of $\mathbf{W}$

$$S_{\mathbf{W}} = \sum_{n=1}^{N} d_n \, \text{card}(\|\boldsymbol{\omega}_n\|_2), \tag{6.14}$$

where $\mathbf{d} = [d_1, d_2, \ldots, d_N]^T$ is a given vector of sensor selection cost, and $\| \cdot \|_2$ denotes the ℓ_2 norm (or Euclidean norm). In (6.14), use of the ℓ_2 norm is motivated by the problem of group Lasso [92], which uses the ℓ_2 norm to promote the group-sparsity of a vector.

Note that both the expressions of transmission cost (6.4) and Fisher information (6.9) involve a quadratic matrix function [98]. For simplicity of presentation, we convert the quadratic matrix function by concatenating the entries of $\mathbf{W}$ into a row vector $\mathbf{w}$. Specifically, the vector $\mathbf{w}$ is given

by

$$\mathbf{w} = [w_1, w_2, \ldots, w_L]^T, \ \ w_l = W_{m_l n_l}, \tag{6.15}$$

where $L = N^2$, $m_l = \lceil \frac{l}{N} \rceil$, $n_l = l - (\lceil \frac{l}{N} \rceil - 1)N$ and $\lceil x \rceil$ is the ceiling function that yields the smallest integer not less than x.

As shown in Appendix A.12, the expressions of transmission cost (6.4), Fisher information (6.9), collaboration cost (6.11), and selection cost (6.14) can be converted into functions of the collaboration vector $\mathbf{w}$,

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{\Omega}_{\text{JN}} \mathbf{w}}{\mathbf{w}^T \mathbf{\Omega}_{\text{JD}} \mathbf{w} + \xi^2}, \ \ T(\mathbf{w}) = \mathbf{w}^T \mathbf{\Omega}_{\text{T}} \mathbf{w}, \ \ Q(\mathbf{w}) = \sum_{l=1}^{L} c_l \operatorname{card}(w_l)$$

$$P(\mathbf{w}) := T(\mathbf{w}) + Q(\mathbf{w}), \ \ S(\mathbf{w}) = \sum_{n=1}^{N} d_n \operatorname{card}(\|\mathbf{w}_{\text{G}_n}\|_2), \tag{6.16}$$

where the matrices $\mathbf{\Omega}_{\text{T}}, \mathbf{\Omega}_{\text{JN}}, \mathbf{\Omega}_{\text{JD}}$ are all symmetric positive semidefinite and defined in Appendix A.12, c_l is the lth entry of $\mathbf{c}$ that is the row-wise vector of the known cost matrix $\mathbf{C}$, and G_n is an index set such that $\mathbf{w}_{\text{G}_n} = [w_{(n-1)N+1}, w_{(n-1)N+2}, \ldots, w_{nN}]^T = \boldsymbol{\omega}_n$ in (6.13) for $n = 1, 2, \ldots, N$. Clearly, the row-sparsity of $\mathbf{W}$ is precisely characterized by the group-sparsity of $\mathbf{w}$ over the index sets $\{\text{G}_n\}_{n=1,2,\ldots,N}$.

Based on the transmission cost $T(\mathbf{w})$, collaboration cost $Q(\mathbf{w})$, and performance measure $J(\mathbf{w})$, we first pose two sensor collaboration problems by disregarding the cost of sensor selection and assuming that all N sensors are active.

- *Information constrained* sensor collaboration

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & P(\mathbf{w}) \\ \text{subject to} \quad & J(\mathbf{w}) \geq \check{J}, \end{aligned} \tag{6.17}$$

where $\check{J} > 0$ is a given information threshold.

- *Energy constrained* sensor collaboration

$$\begin{aligned}
\underset{\mathbf{w}}{\text{maximize}} \quad & J(\mathbf{w}) \\
\text{subject to} \quad & P(\mathbf{w}) \leq \hat{P},
\end{aligned} \tag{6.18}$$

where $\hat{P} > 0$ is a given energy budget.

Next, we incorporate the sensor selection cost $S(\mathbf{w})$, and pose the optimization problem for the joint design of optimal sensor selection and collaboration schemes.

- *Joint* sensor selection and collaboration

$$\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & P(\mathbf{w}) + S(\mathbf{w}) \\
\text{subject to} \quad & J(\mathbf{w}) \geq \check{J}.
\end{aligned} \tag{6.19}$$

In (6.19), we minimize the total energy cost subject to an information constraint. This formulation is motivated by scenarios where saving energy is the major goal in the context of sensor selection [99]. Problem (6.19) is similar to (6.17) except for the incorporation of sensor selection cost. However, we will show that the presence of sensor selection cost makes finding the solution of (6.19) more challenging; see Section 6.6 for details.

In (6.17)-(6.19), the cardinality function, which appears in $Q(\mathbf{w})$ and $S(\mathbf{w})$, promotes the sparsity of $\mathbf{w}$, and therefore the sparsity of $\mathbf{W}$. Thus, we refer to (6.17)-(6.19) as sparsity-aware sensor collaboration problems. We also note that (6.17)-(6.19) are nonconvex optimization problems due to the presence of the cardinality function and the nonconvexity of the expression for Fisher information (see Remark 2). In the following sections, we will elaborate on the optimization approaches for solving (6.17)-(6.19).

6.4 Information constrained sensor collaboration

In this section, we relax the original information constrained problem (6.17) by using an iterative reweighted ℓ_1 minimization method. This results in an ℓ_1 optimization problem, which can be efficiently solved by ADMM.

Due to the presence of the cardinality function, problem (6.17) is combinatorial in nature. A method for solving (6.17) is to replace the cardinality function with a weighted ℓ_1 norm [100]. This leads to the following optimization problem

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \boldsymbol{\Omega}_{\text{T}} \mathbf{w} + \|\boldsymbol{\Omega}_{\text{C}} \mathbf{w}\|_1 \\
\text{subject to} \quad & \mathbf{w}^T (\check{J} \boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}) \mathbf{w} + \check{J} \xi^2 \le 0,
\end{aligned}
\tag{6.20}
$$

where $\boldsymbol{\Omega}_{\text{C}} = \text{diag}(\alpha_1 c_1, \alpha_2 c_2, \ldots, \alpha_L c_L)$, $\{\alpha_l\}_{l=1,2,\ldots,L}$ denote the weights that are iteratively updated based on $\alpha_l \leftarrow 1/|w_l| + \varepsilon$, in order to make $\|\boldsymbol{\Omega}_{\text{C}} \mathbf{w}\|_1$ a good approximation for $Q(\mathbf{w})$. Here ε is a small positive number which insures that the denominator is always nonzero. To elaborate, if $\alpha_l = 1$ for all $l \in \{1, 2, \ldots, L\}$, we recover the standard unweighted ℓ_1 norm. Since the ℓ_0 norm only counts the number of nonzero entries of a vector, using the ℓ_1 norm to approximate the ℓ_0 norm has the disadvantage that the amplitudes of the nonzero entries come into play. To compensate for the amplitude of nonzero entries, we iteratively normalize the entries of the argument of the ℓ_1 norm, to make this norm a better proxy for the ℓ_0 norm.

Remark 2. *In (6.20), the constraint is equivalent to the information inequality in (6.17). According to Lemma 2 in Appendix A.13, we obtain that the matrix $\check{J} \boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}$ is not positive semidefinite. Indeed, if the matrix $\check{J} \boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}$ was positive semidefinite, problem (6.20) would have an empty feasible set.*

Given $\{\alpha_l\}_{l=1,\ldots,L}$, problem (6.20) is a nonconvex optimization problem, and its objective function is not differentiable. In what follows, we will employ ADMM to find its locally optimal solutions.

We begin by reformulating the optimization problem (6.20) in a way that lends itself to the

application of ADMM,

$$\underset{\mathbf{w},\mathbf{v}}{\text{minimize}} \quad \mathbf{w}^T\mathbf{\Omega}_{\text{T}}\mathbf{w} + \|\mathbf{\Omega}_{\text{C}}\mathbf{v}\|_1 + \mathcal{I}(\mathbf{w}) \tag{6.21}$$
$$\text{subject to} \quad \mathbf{w} = \mathbf{v},$$

where we have introduced the indicator function $\mathcal{I}(\mathbf{w})$

$$\mathcal{I}(\mathbf{w}) = \begin{cases} 0 & \text{if } \mathbf{w}^T(\check{J}\mathbf{\Omega}_{\text{JD}} - \mathbf{\Omega}_{\text{JN}})\mathbf{w} + J\xi^2 \leq 0 \\ \infty & \text{otherwise.} \end{cases} \tag{6.22}$$

The augmented Lagrangian of (6.21) is given by

$$\mathcal{L}(\mathbf{w}, \mathbf{v}, \boldsymbol{\chi}) = \mathbf{w}^T\mathbf{\Omega}_{\text{T}}\mathbf{w} + \|\mathbf{\Omega}_{\text{C}}\mathbf{v}\|_1 + \mathcal{I}(\mathbf{w}) + \boldsymbol{\chi}^T(\mathbf{w} - \mathbf{v}) + \frac{\rho}{2}\|\mathbf{w} - \mathbf{v}\|_2^2, \tag{6.23}$$

where the vector $\boldsymbol{\chi}$ is the Lagrangian multiplier, and the scalar $\rho > 0$ is a penalty weight. The ADMM algorithm iteratively executes the following three steps for $k = 1, 2, \ldots$

$$\mathbf{w}^{k+1} = \arg\min_{\mathbf{w}} \ \mathcal{L}(\mathbf{w}, \mathbf{v}^k, \boldsymbol{\chi}^k), \tag{6.24}$$

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \ \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}, \boldsymbol{\chi}^k), \tag{6.25}$$

$$\boldsymbol{\chi}^{k+1} = \boldsymbol{\chi}^k + \rho(\mathbf{w}^{k+1} - \mathbf{v}^{k+1}), \tag{6.26}$$

until $\|\mathbf{w}^{k+1} - \mathbf{v}^{k+1}\|_2 \leq \epsilon_{\text{ad}}$ and $\|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2 \leq \epsilon_{\text{ad}}$, where ϵ_{ad} is a stopping tolerance.

It is clear from ADMM steps (6.24)-(6.25) that the original nondifferentiable problem can be effectively separated into a '$\mathbf{w}$-minimization' subproblem (6.24) and a '$\mathbf{v}$-minimization' subproblem (6.25), of which the former can be treated as a nonconvex quadratic program with only one quadratic constraint (QP1QC) and the latter admits an analytical solution. In the subsections that follow, we will elaborate on the execution of the minimization problems (6.24) and (6.25).

w-minimization step

Completing the squares with respect to $\mathbf{w}$ in (6.23), the $\mathbf{w}$-minimization step (6.24) is given by

$$\underset{\mathbf{w}}{\text{minimize}} \quad \mathbf{w}^T \mathbf{\Omega}_{\text{T}} \mathbf{w} + \frac{\rho}{2} \|\mathbf{w} - \mathbf{a}\|_2^2$$
$$\text{subject to} \quad \mathbf{w}^T (\check{J} \mathbf{\Omega}_{\text{JD}} - \mathbf{\Omega}_{\text{JN}}) \mathbf{w} + \check{J}\xi^2 \leq 0, \tag{6.27}$$

where we have applied the definition of $\mathcal{I}(\mathbf{w})$ in (6.22), and $\mathbf{a} := \mathbf{v}^k - 1/\rho\boldsymbol{\chi}^k$. Problem (6.27) is a nonconvex QP1QC. To seek the global minimizer of a nonconvex QP1QC, an SDP-based approach is commonly used [59]. However, computing solutions to SDPs becomes inefficient for problems with hundreds or thousands of variables. Therefore, we develop a faster approach by exploiting the KKT conditions of (6.27). This is presented in Proposition 6.1.

Proposition 6.1. *The KKT-based solution of problem (6.27) is given by*

$$\mathbf{w}^{k+1} = \tilde{\mathbf{\Omega}}_T^{-\frac{1}{2}} \mathbf{U}\mathbf{u},$$

where $\tilde{\mathbf{\Omega}}_T := \mathbf{\Omega}_T + \frac{\rho}{2}\mathbf{I}$, $\mathbf{U}$ is an orthogonal matrix that satisfies the eigenvalue decomposition

$$\frac{1}{\check{J}\xi^2} \tilde{\mathbf{\Omega}}_T^{-\frac{1}{2}} (\check{J}\mathbf{\Omega}_{\text{JD}} - \mathbf{\Omega}_{\text{JN}}) \tilde{\mathbf{\Omega}}_T^{-\frac{1}{2}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \tag{6.28}$$

and $\mathbf{u}$ is given by

$$\begin{cases} \mathbf{u} = -\mathbf{g} & \text{if } \mathbf{g}^T \mathbf{\Lambda}\mathbf{g} + 1 \leq 0 \\ \mathbf{u} = -(\mathbf{I} + \mu_0 \mathbf{\Lambda})^{-1} \mathbf{g} & \text{otherwise}. \end{cases} \tag{6.29}$$

In (6.29), $\mathbf{g} := -\frac{\rho}{2} \mathbf{U}^T \tilde{\mathbf{\Omega}}_T^{-\frac{1}{2}} \mathbf{a}$, and μ_0 is a positive root of the equation in μ

$$f(\mu) := \sum_{l=1}^{L} \frac{\lambda_l g_l^2}{(\mu\lambda_l + 1)^2} + 1 = 0, \tag{6.30}$$

where g_l is the lth entry of $\mathbf{g}$, and λ_l is the lth diagonal entry of $\mathbf{\Lambda}$.

Proof: See Appendix A.14. ∎

The rationale behind deriving the eigenvalue decomposition (6.28) is that by introducing $\mathbf{u} = \mathbf{U}^T \tilde{\Omega}_1^{\frac{1}{2}} \mathbf{w}$, problem (6.27) can be transformed to

$$
\begin{aligned}
\underset{\mathbf{u}}{\text{minimize}} \quad & \mathbf{u}^T\mathbf{u} + 2\mathbf{u}^T\mathbf{g} \\
\text{subject to} \quad & \mathbf{u}^T\Lambda\mathbf{u} + 1 \leq 0.
\end{aligned}
\tag{6.31}
$$

The benefit of this reformulation is that the KKT conditions (6.29)-(6.30) for problem (6.31) are more compact and more easily solved than the SDP formulation, since Λ is a diagonal matrix and its inversion is tractable.

In general, Eq. (6.30) is a high-order polynomial function and it is difficult to obtain all of its positive roots. However, we have observed that numerical searches over small targeted intervals yield satisfactory results. One such interval is given by Lemma 3, and the others are demonstrated in Remark 3 below. Solutions of (6.30) over small targeted intervals can be found by using the MATLAB function fminbnd or by using Newton's method. If we find multiple positive roots, we select the one corresponding to the lowest objective value of the nonconvex QP1QC (6.27).

Proposition 6.2. *The function $f(\mu)$ is monotonically decreasing on the interval $(0, -\frac{1}{\lambda_1})$ and the positive root of $f(\mu) = 0$ is unique when $f(0) > 0$, where λ_1 represents the unique negative eigenvalue in $\{\lambda_l\}_{l=1,2,\dots,L}$.*

Proof: See Appendix A.15. ∎

Remark 3. *Motivated by Proposition 6.2, one may inquire about the monotonicity of $f(\mu)$ over the interval $\mu \in (-\frac{1}{\lambda_1}, \infty)$. In Appendix A.15, we show that the sign of the first-order derivative of $f(\mu)$ is difficult to determine from (A.35) and (A.36). And our numerical results show that there may exist other positive roots over the interval $(-\frac{1}{\lambda_1}, \infty)$. To address this issue, we are thus motivated to consider multiple subintervals of $(-\frac{1}{\lambda_1}, \infty)$. Although we cannot guarantee global optimality for solutions of (6.27) found through KKT [59], our extensive numerical results show that numerical search over several small intervals of $(-\frac{1}{\lambda_1}, \infty)$ together with $(0, -\frac{1}{\lambda_1})$ works*

effectively for finding the positive roots of Eq. (6.30), and the ADMM algorithm always converges to a near-optimal solution of the information constrained collaboration problem.

v-minimization step

Completing the squares with respect to $\mathbf{v}$ in (6.23), the $\mathbf{v}$-minimization step (6.25) becomes

$$\underset{\mathbf{v}}{\text{minimize}} \quad \|\mathbf{\Omega}_c\mathbf{v}\|_1 + \frac{\rho}{2}\|\mathbf{v} - \mathbf{b}\|_2^2, \tag{6.32}$$

where $\mathbf{b} := \frac{1}{\rho}\boldsymbol{\chi}^k + \mathbf{w}^{k+1}$. The solution of (6.32) is given by soft thresholding

$$v_l = \begin{cases} (1 - \frac{\alpha_l^t c_l}{\rho|b_l|})b_l & |b_l| > \frac{\alpha_l^t c_l}{\rho} \\ 0 & |b_l| \leq \frac{\alpha_l^t c_l}{\rho} \end{cases} \tag{6.33}$$

for $l = 1, 2, \ldots, L$, where v_l denotes the lth entry of $\mathbf{v}$.

Initialization

To initialize ADMM we require a feasible vector. It was shown in [48, Theorem 1] that the optimal collaboration vector for a fully-connected network with an information threshold $\check{J}$ is a feasible vector for (6.20); see Appendix A.13. Thus, we choose $\mathbf{v}^0 = \mathbf{w}^0$ and $\mathbf{w}^0 = \tilde{\mathbf{w}}$, where $\tilde{\mathbf{w}}$ is given by (A.29a).

Complexity analysis

To solve the information constrained collaboration problem (6.17), the iterative reweighted ℓ_1 method is used as the outer loop, and the ADMM algorithm constitutes the inner loop. It is often the case that the iterative reweighted ℓ_1 method converges within a few iterations [100]. Moreover, it has been shown in [60] that the ADMM algorithm typically requires a few tens of iterations for converging with modest accuracy. At each iteration, the major cost is associated with solving

the KKT conditions in the w-minimization step. The complexity of obtaining a KKT-based solution is given by $O(L^3)$ [101], since the complexity of the eigenvalue decomposition dominates that of Newton's method. By contrast, using SDP to solve problem (6.17) yields a much higher complexity of $O(L^{4.5})$.

6.5 Energy constrained sensor collaboration

In this section, we first explore the correspondence between the energy constrained collaboration problem and the information constrained problem. With the help of this correspondence, we propose a bisection algorithm to solve the energy constrained problem.

The energy constrained sensor collaboration problem (6.18) can be written as

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{maximize}} \quad & \frac{\mathbf{w}^T \mathbf{\Omega}_{\text{JN}} \mathbf{w}}{\mathbf{w}^T \mathbf{\Omega}_{\text{JD}} \mathbf{w} + \xi^2} \\
\text{subject to} \quad & \mathbf{w}^T \mathbf{\Omega}_{\text{T}} \mathbf{w} + \sum_{l=1}^{L} c_l \operatorname{card}(w_l) \leq \hat{P}.
\end{aligned}
$$

Compared to the information constrained problem (6.17), problem (6.18) is more involved due to the nonconvex objective function and the cardinality function in the inequality constraint. Even if we replace the cardinality function with its ℓ_1 norm relaxation, the resulting ℓ_1 optimization problem is still difficult, since the feasibility of the relaxed constraint does not guarantee the feasibility of the original problem (6.18).

However, if the collaboration topology is given, the collaboration cost $\sum_{l=1}^{L} c_l \operatorname{card}(w_l)$ is a constant and the constraint in (6.18) becomes a homogeneous quadratic constraint, in the sense that inequality no longer contain terms linear in $\mathbf{w}$. Therefore, problem (6.18) can be solved by [48, Theorem 1].

In Proposition 6.3, we present the relationship between the energy constrained problem (6.18) and the information constrained problem (6.17). Motivated by this relationship, we then take advantage of the solution of (6.17) to obtain the collaboration topology for (6.18). This idea will be elaborated on later.

Proposition 6.3. *Consider the two problems (6.17) and (6.18)*

$$
\begin{array}{ll}
\underset{\mathbf{w}}{minimize} \quad P(\mathbf{w}) & \underset{\mathbf{w}}{maximize} \quad J(\mathbf{w}) \\
subject\ to \quad J(\mathbf{w}) \geq \check{J} \quad and \quad subject\ to \quad P(\mathbf{w}) \leq \hat{P},
\end{array}
$$

where the optimal solutions are denoted by $\mathbf{w}_1$ *and* $\mathbf{w}_2$, *respectively. If* $\check{J} = J(\mathbf{w}_2)$, *then* $\mathbf{w}_1 = \mathbf{w}_2$; *If* $\hat{P} = P(\mathbf{w}_1)$, *then* $\mathbf{w}_2 = \mathbf{w}_1$.

Proof: see Appendix A.16. ∎

Proposition. 6.3 implies that the solution of the energy constrained problem (6.18) can be obtained by seeking the global minimizer of the information constrained problem (6.17), if the information threshold in (6.17) is set by using the optimal value of (6.18). However, this methodology is intractable in practice since the optimal value of (6.18) is unknown in advance, and the globally optimal solution of problem (6.17) may not be found using reweighted ℓ_1-based methods.

Instead of deriving the solution of (6.18) from (6.17), we can infer the collaboration topology of the energy constrained problem (6.18) from the sparsity structure of the solution to the information constrained problem (6.17) using a bisection algorithm. The objective function of (6.18) (in terms of Fisher information) is bounded over the interval $[0, J_0)$, where J_0 is given in Lemma 1 of Appendix A.13. And there is a one-to-one correspondence between the value of Fisher information evaluated at the optimal solution of (6.18) and the energy budget $\hat{P}$. Therefore, we perform a bisection algorithm on the interval, and then solve the information constrained problem to obtain the resulting energy cost and collaboration topology. The procedure terminates if the resulting energy cost is close to the energy budget $\hat{P}$. We summarize this method in Algorithm 6.1.

We remark that the collaboration topology obtained in Step 2 of Algorithm 6.1 is not globally optimal. However, we have observed that for the information constrained sensor collaboration (6.17), the value of energy cost P is monotonically related to the value of desired estimation distortion. Therefore, the proposed bisection algorithm converges in practice and at most requires $\lceil \log_2(\overline{J}/\epsilon_{\mathrm{bi}}) \rceil$ iterations. Once the bisection procedure terminates, we obtain a locally optimal collaboration topology for (6.18). Given this topology, the energy constrained problem (6.18)

Algorithm 6.1 Bisection algorithm for seeking the optimal collaboration topology of (6.18)

Require: given $\epsilon_{\text{bi}} > 0$, $\underline{J} = 0$ and $\overline{J} = J_0$

  1: **repeat** $\check{J} = \frac{\underline{J} + \overline{J}}{2}$

  2:     for a given $\check{J}$, solve (6.17) to obtain the collaboration topology and the resulting energy cost P

  3:     **if** $P < \hat{P}$ **then** $\underline{J} = \check{J}$

  4:     **else** $\overline{J} = \check{J}$

  5:     **end if**

  6: **until** $\overline{J} - \underline{J} < \epsilon_{\text{bi}}$ or $|\hat{P} - P| < \epsilon_{\text{bi}}$

---

becomes a problem with a quadratic constraint and an objective that is a ratio of homogeneous quadratic functions, whose analytical solution is given by [48, Theorem 1]. Through the aforementioned procedure, we obtain a locally optimal solution to problem (6.18).

6.6 Joint sensor selection and collaboration

In this section, we study the problem of the joint design of optimal sensor selection and collaboration schemes as formulated in (6.19). Similar to solving the information constrained collaboration problem (6.17), we first relax the original problem to a nonconvex ℓ_1 optimization problem. However, in contrast to Section 6.4, we observe that ADMM fails to converge (a possible reason is explored later). To circumvent this, we adopt an iterative method to solve the nonconvex ℓ_1 optimization problem.

Using the reweighted ℓ_1 minimization method, we iteratively replace the cardinality function with the weighted ℓ_1 norm, which yields at every iteration

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \mathbf{\Omega}_{\text{T}} \mathbf{w} + \|\tilde{\mathbf{\Omega}}_{\text{C}} \mathbf{w}\|_1 + \sum_{n=1}^{N} \tilde{d}_n \|\mathbf{w}_{\text{G}_n}\|_2 \\
\text{subject to} \quad & \mathbf{w}^T (\check{J} \mathbf{\Omega}_{\text{JD}} - \mathbf{\Omega}_{\text{JN}}) \mathbf{w} + \check{J} \xi^2 \leq 0,
\end{aligned}
\tag{6.34}
$$

where $\tilde{\mathbf{\Omega}}_{\text{C}} := \text{diag}(\tau_1 c_1, \tau_2 c_2, \ldots, \tau_L c_L)$, $\tilde{d}_n := \delta_n d_n$, and τ_l and δ_n are positive weights that are iteratively updated in order to make $\|\tilde{\mathbf{\Omega}}_{\text{C}} \mathbf{w}\|_1$ and $\sum_{n=1}^{N} \tilde{d}_n \|\mathbf{w}_{\text{G}_n}\|_2$ good approximations for $Q(\mathbf{w})$ and $S(\mathbf{w})$. Let the solution of (6.34) be $\mathbf{w}$, then the weights τ_l and δ_n for the next reweighting

iteration are updated as [100]

$$\tau_l \leftarrow \frac{1}{|w_l| + \varepsilon}, \quad \delta_n \leftarrow \frac{1}{\|\mathbf{w}_{\mathrm{G}_n}\|_2 + \varepsilon},$$

where ε is a small positive number that insures a nonzero denominator.

Convex restriction

Problem (6.34) is a nonconvex optimization problem. One can use ADMM to split (6.34) into a nonconvex QP1QC ($\mathbf{w}$-minimization step) and an unconstrained optimization problem with an objective function composed of ℓ_1 and ℓ_2 norms ($\mathbf{v}$-minimization step), where the latter can be solved analytically. However, our numerical examples show that the resulting ADMM algorithm fails to converge. Note that in the $\mathbf{w}$-minimization step, the sensor selection cost is excluded and each sensor collaborates with itself at no cost. Therefore, to achieve an information threshold, there exist scenarios in which the collaboration matrix yields nonzero diagonal entries. This implies that the $\mathbf{w}$-minimization step does not produce group-sparse solutions (i.e., row-wise sparse collaboration matrices). However, the $\mathbf{v}$-minimization step always leads to group-sparse solutions. The mismatched sparsity structures of solutions in the subproblems of ADMM are responsible for lack of convergence, which we circumvent by using a linearization method to convexify the optimization problem.

A linearization method is introduced in [102] for solving the nonconvex quadratically constrained quadratic program (QCQP) by *linearizing* the nonconvex parts of quadratic constraints, thus rendering a convex QCQP. In (6.34), the nonconvex constraint is given by

$$\mathbf{w}^T \check{J} \mathbf{\Omega}_{\mathrm{JD}} \mathbf{w} + \check{J} \xi^2 \leq \mathbf{w}^T \mathbf{\Omega}_{\mathrm{JN}} \mathbf{w}, \tag{6.35}$$

where we recall that $\mathbf{\Omega}_{\mathrm{JD}}$ and $\mathbf{\Omega}_{\mathrm{JN}}$ are positive semidefinite.

In (6.35), the only source of nonconvexity is the difference of convex (DC) function $\mathbf{w}^T \check{J} \mathbf{\Omega}_{\mathrm{JD}} \mathbf{w} -$

$\mathbf{w}^T \mathbf{\Omega}_{\text{JN}} \mathbf{w}$. We linearize the right hand side of (6.35) around a feasible point $\boldsymbol{\beta}$

$$\mathbf{w}^T \check{J} \mathbf{\Omega}_{\text{JD}} \mathbf{w} + \check{J}\xi^2 \leq \boldsymbol{\beta}^T \mathbf{\Omega}_{\text{JN}} \boldsymbol{\beta} + 2\boldsymbol{\beta}^T \mathbf{\Omega}_{\text{JN}}(\mathbf{w} - \boldsymbol{\beta}). \tag{6.36}$$

Note that the right hand side of (6.36) is an affine *lower bound* on the convex function $\mathbf{w}^T \mathbf{\Omega}_{\text{JN}} \mathbf{w}$. This implies that the set of $\mathbf{w}$ that satisfy (6.36) is a strict subset of the set of $\mathbf{w}$ that satisfy (6.35).

By replacing (6.35) with (6.36), we obtain a 'restricted' convex version of problem (6.34)

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \varphi(\mathbf{w}) := \mathbf{w}^T \mathbf{\Omega}_{\text{T}} \mathbf{w} + \|\tilde{\mathbf{\Omega}}_{\text{C}} \mathbf{w}\|_1 + \sum_{n=1}^{N} \tilde{d}_n \|\mathbf{w}_{\text{G}_n}\|_2 \\ \text{subject to} \quad & \mathbf{w}^T \tilde{\mathbf{\Omega}}_{\text{JD}} \mathbf{w} - 2\tilde{\boldsymbol{\beta}}^T \mathbf{w} + \tilde{\gamma} \leq 0, \end{aligned} \tag{6.37}$$

where $\tilde{\mathbf{\Omega}}_{\text{JD}} := \check{J} \mathbf{\Omega}_{\text{JD}}$, $\tilde{\boldsymbol{\beta}} := \mathbf{\Omega}_{\text{JN}} \boldsymbol{\beta}$, and $\tilde{\gamma} := \boldsymbol{\beta}^T \mathbf{\Omega}_{\text{JN}} \boldsymbol{\beta} + \check{J}\xi^2$. Different from (6.35), the inequality constraint in (6.37) no longer represents the information inequality but becomes a convex quadratic constraint. And the optimal value of (6.37) yields an upper bound to problem (6.34).

Algorithm 6.2 CCP for solving problem (6.34)

Require: given $\epsilon_{\text{li}} > 0$ and $\mathbf{w}^0 = \tilde{\mathbf{w}}$.
1: **for** iteration $s = 1, 2, \ldots$ **do**
2:      set $\boldsymbol{\beta} = \mathbf{w}^{s-1}$.
3:      solve (6.37) for the solution $\mathbf{w}^s$ by using ADMM.
4:      **until** $\|\varphi(\mathbf{w}^s) - \varphi(\mathbf{w}^{s-1})\| < \epsilon_{\text{li}}$.
5: **end for**

The use of linearization to convexify nonconvex problems with DC type functions is known as the convex-concave procedure (CCP) [102]. We summarize the application of CCP in Algorithm 6.2. In the following subsection, we will elaborate on the implementation of ADMM in Step 3 of Algorithm 6.2. We remark that the convergence of Algorithm 6.2 is guaranteed [103], since Algorithm 6.2 starts from a feasible point $\mathbf{w}^0$ that satisfies (6.35) and at each iteration, we solve a linearized convex problem with a smaller feasible set which contains the linearization point (i.e., the solution at the previous iteration). In other words, for a given linearization point $\boldsymbol{\beta} = \mathbf{w}^{s-1}$, we always obtain a new feasible point $\mathbf{w}^s$ with a lower or equal objective value at each

iteration.

Solution via ADMM

Similar to (6.21) in Section 6.4, we introduce the auxiliary variable $\mathbf{v}$ to replace $\mathbf{w}$ in the ℓ_1 and ℓ_2 norms in (6.37) while adding the constraint $\mathbf{w} = \mathbf{v}$, and split problem (6.37) into a sequence of subproblems as in (6.24)-(6.26). However, compared to Section 6.4, the current ADMM algorithm yields different subproblems due to the presence of the ℓ_2 norm in the objective function and the convexification in the constraint.

Completing sqaures with respect to $\mathbf{w}$ in the augmented Lagrangian function corresponding to problem (6.37), the $\mathbf{w}$-minimization step is given by

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \tilde{\boldsymbol{\Omega}}_{\mathrm{T}} \mathbf{w} - \rho \mathbf{a}^T \mathbf{w} \\
\text{subject to} \quad & \mathbf{w}^T \tilde{\boldsymbol{\Omega}}_{\mathrm{JD}} \mathbf{w} - 2 \tilde{\boldsymbol{\beta}}^T \mathbf{w} + \tilde{\gamma} \leq 0,
\end{aligned}
\tag{6.38}
$$

where $\tilde{\boldsymbol{\Omega}}_{\mathrm{T}} = \boldsymbol{\Omega}_{\mathrm{T}} + \frac{\rho}{2}\mathbf{I}$, $\mathbf{a} = \mathbf{v}^k - 1/\rho \boldsymbol{\chi}^k$, $\mathbf{v}^k$ and $\boldsymbol{\chi}^k$ denote the value of $\mathbf{v}$ and $\boldsymbol{\chi}$ at the kth iteration of ADMM, and $\boldsymbol{\chi}$ is the dual variable.

To solve the convex QCQP (6.38), the complexity of using interior-point method in standard solvers is roughly $O(L^{3.5})$ [104]. To reduce the computational complexity, we can derive the KKT-based solution. Since problem (6.38) is convex, KKT conditions are both necessary and sufficient for optimality. This is summarized in Proposition 6.4.

Proposition 6.4. *The optimal solution of problem (6.38) is given by*

$$
\mathbf{w}^{k+1} = \tilde{\boldsymbol{\Omega}}_T^{-\frac{1}{2}} \mathbf{U} \mathbf{u},
$$

where $\mathbf{U}$ is given by the following eigenvalue decomposition

$$
\frac{1}{\tilde{\gamma}} \tilde{\boldsymbol{\Omega}}_T^{-\frac{1}{2}} \tilde{\boldsymbol{\Omega}}_{JD} \tilde{\boldsymbol{\Omega}}_T^{-\frac{1}{2}} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T,
$$

and **u** *is given by*

$$\begin{cases} \mathbf{u} = -\mathbf{g} & \text{if } \mathbf{g}^T \mathbf{\Lambda} \mathbf{g} + 2\mathbf{g}^T \mathbf{e} + 1 \leq 0 \\ \mathbf{u} = -(\mathbf{I} + \mu_1 \mathbf{\Lambda})^{-1}(\mathbf{g} + \mu_1 \mathbf{e}) & \text{otherwise.} \end{cases} \tag{6.39}$$

In (6.39), $\mathbf{g} = -\rho \mathbf{U}^T \tilde{\mathbf{\Omega}}_\tau^{-\frac{1}{2}} \mathbf{a}/2$, $\mathbf{e} = -\mathbf{U}^T \tilde{\mathbf{\Omega}}_\tau^{-\frac{1}{2}} \tilde{\boldsymbol{\beta}}/\tilde{\gamma}$, μ_1 *is the positive root of the equation in* μ

$$\sum_{l=1}^{L} \frac{(\lambda_l g_l - e_l)^2}{\lambda_l(1 + \mu \lambda_l)^2} - \sum_{l=1}^{L} \frac{e_l^2}{\lambda_l} + 1 = 0,$$

e_l *and* g_l *are the lth elements of* **e** *and* **g***, respectively, and* λ_l *is the lth diagonal entry of* $\mathbf{\Lambda}$*.*

Proof: See Appendix A.14. ∎

Completing sqaures with respect to **v** in the augmented Lagrangian function corresponding to problem (6.37), the **v**-minimization step is given by

$$\underset{\mathbf{v}}{\text{minimize}} \quad \|\tilde{\mathbf{\Omega}}_\mathrm{c} \mathbf{v}\|_1 + \sum_{n=1}^{N} \tilde{d}_n \|\mathbf{v}_{\mathrm{G}_n}\|_2 + \frac{\rho}{2}\|\mathbf{v} - \mathbf{b}\|_2^2, \tag{6.40}$$

where $\mathbf{b} = \mathbf{w}^{k+1} + 1/\rho \boldsymbol{\chi}^k$.

We recall that $\tilde{\mathbf{\Omega}}_\mathrm{c}$ defined in (6.34) is a diagonal matrix. Let the vector **f** be composed of the diagonal entries of $\tilde{\mathbf{\Omega}}_\mathrm{c}$. We then define a sequence of diagonal matrices $\mathbf{F}_n := \mathrm{diag}(\mathbf{f}_{\mathrm{G}_n})$ for $n = 1, 2, \ldots, N$, where $\mathbf{f}_{\mathrm{G}_n}$ is a vector composed of those entries of **f** whose indices belong to the set G_n. Since the index sets $\{\mathrm{G}_n\}_{j=1,2,\ldots,N}$ are disjoint, problem (6.40) can be decomposed into a sequence of subproblems for $n = 1, 2, \ldots, N$,

$$\underset{\mathbf{v}_{\mathrm{G}_n}}{\text{minimize}} \quad \|\mathbf{F}_n \mathbf{v}_{\mathrm{G}_n}\|_1 + \tilde{d}_n \|\mathbf{v}_{\mathrm{G}_n}\|_2 + \frac{\rho}{2}\|\mathbf{v}_{\mathrm{G}_n} - \mathbf{b}_{\mathrm{G}_n}\|_2^2. \tag{6.41}$$

Problem (6.41) can be solved analytically via the following Proposition.

Proposition 6.5. *The minimizer of (6.41) is given by*

$$
\mathbf{v}_{G_n} = \begin{cases} (1 - \frac{\tilde{d}_n}{\rho \|\boldsymbol{\nu}\|_2}) \boldsymbol{\nu} & \|\boldsymbol{\nu}\|_2 \geq \frac{\tilde{d}_n}{\rho} \\[2mm] 0 & \|\boldsymbol{\nu}\|_2 < \frac{\tilde{d}_n}{\rho}, \end{cases} \tag{6.42}
$$

where $\boldsymbol{\nu} = \mathrm{sgn}(\mathbf{b}_{G_n}) \odot \max(|\mathbf{b}_{G_n}| - \frac{1}{\rho}\mathbf{f}_{G_n}, 0)$, the operator $\mathrm{sgn}(\cdot)$ is defined in a componentwise fashion as

$$
\mathrm{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0, \end{cases}
$$

$\odot$ denotes the point-wise product, and the operator $\max(\mathbf{x}, \mathbf{y})$ returns a vector whose entries are the pointwise maximum of the entries of $\mathbf{x}$ and $\mathbf{y}$.

Proof: See Appendix A.17. ∎

In Algorithm 6.3, we present our proposed ADMM algorithm for solving (6.37).

Algorithm 6.3 Solving problem (6.37) via ADMM

Require: given ρ, ϵ_{ad}, $\boldsymbol{\chi}^0 = \mathbf{0}$ and $\mathbf{w}^0 = \mathbf{v}^0 = \tilde{\mathbf{w}}$.
1: **for** $k = 0, 1, \ldots$ **do**
2:      obtain $\mathbf{w}^{k+1}$ from a standard QCQP solver or Prop. 6.4.
3:      obtain $\mathbf{v}^{k+1} = [(\mathbf{v}_{G_1}^{k+1})^T, \ldots, (\mathbf{v}_{G_N}^{k+1})^T]^T$ from Prop. 6.5.
4:      update dual variable $\boldsymbol{\chi}^{k+1} = \boldsymbol{\chi}^k + \rho(\mathbf{w}^{k+1} - \mathbf{v}^{k+1})$.
5:      **until** $\|\mathbf{w}^{k+1} - \mathbf{v}^{k+1}\|_2 \leq \epsilon_{\mathrm{ad}}$, $\|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2 \leq \epsilon_{\mathrm{ad}}$.
6: **end for**

To summarize, for solving the original problem (6.19) we first replace the cardinality function with the weighted ℓ_1 norm, which yields the nonconvex problem (6.34). We then use the linearization method to convexify (6.34). The resulting convex problem (6.37) is solved by ADMM as outlined in Algorithm 6.3.

6.7 Numerical results

In this section, we will illustrate the performance of our proposed sparsity-aware sensor collaboration methods through numerical examples. The estimation system considered here is shown in Fig.6.1, where for simplicity, we assume that the channel gain and uncertainties are such that the network is homogeneous and equicorrelated. As in [48, Example 3], we denote the expected observation and channel gains by h_0 and g_0, the observation and channel gain uncertainties by α_h and α_g, the measurement noise variance and correlation by ζ^2 and ρ_{corr}, and thereby assume

$$
\begin{cases}
\mathbf{h} = h_0\sqrt{\alpha_h}\mathbf{1}, \ \boldsymbol{\Sigma}_h = h_0^2(1 - \alpha_h)\mathbf{I}, \\
\boldsymbol{\Sigma}_\epsilon = \zeta^2[(1 - \rho_{\text{corr}})\mathbf{I} + \rho_{\text{corr}}\mathbf{1}\mathbf{1}^T], \\
\mathbf{g} = g_0\sqrt{\alpha_g}\mathbf{1}, \ \boldsymbol{\Sigma}_g = g_0^2(1 - \alpha_g)\mathbf{I}.
\end{cases}
\tag{6.43}
$$

Note that channel gains can also be calculated based on path loss models [105] but we chose the homogeneous model for the sake of simplicity. The collaboration cost matrix $\mathbf{C}$ is given by

$$
C_{mn} = \alpha_{\text{c}}\|\mathbf{s}_m - \mathbf{s}_n\|_2
\tag{6.44}
$$

for $m, n = 1, 2, \ldots, N$, where α_{c} is a positive parameter and $\mathbf{s}_n$ denotes the location of sensor n. The vector of sensor selection cost $\mathbf{d}$ is give by

$$
d_n = \alpha_{\text{s}}\|\mathbf{s}_n - \mathbf{s}_{\text{fc}}\|_2
\tag{6.45}
$$

for $n = 1, 2, \ldots, N$, where α_{s} is a positive parameter and $\mathbf{s}_{\text{fc}}$ denotes the location of the FC.

In our experiments, unless specified otherwise, we shall assume that $h_0 = g_0 = 1$, $\alpha_h = \alpha_g = 0.7$, $\rho_{\text{corr}} = 0.5$, $\xi^2 = \zeta^2 = 1$, $\eta^2 = 0.1$, and $\alpha_{\text{c}} = \alpha_{\text{s}} = 0.01$. The FC and N sensors are randomly deployed on a 10×10 grid, where the value of N will be specified in different examples. While employing the proposed optimization methods, we select $\rho \geq 20$ in ADMM, $\varepsilon = 10^{-3}$ in the reweighted ℓ_1 method and $\epsilon_{\text{li}} = \epsilon_{\text{ad}} = 10^{-3}$ for the stopping tolerance. In our numerical examples,

the reweighted ℓ_1 method and the linearization method converge within 10 iterations. For ADMM, the required number of iterations is less than 100.

For a better depiction of estimation performance, since $D_0 < D_{\mathbf{w}} \leq \eta^2$ (see Lemma 1 in Appendix A.13), we display the *normalized distortion*

$$D_{\mathrm{norm}} := \frac{D(\mathbf{w}) - D_0}{\eta^2 - D_0} \in (0, 1], \tag{6.46}$$

where $D(\mathbf{w})$ defined in (6.10) is monotonically related to the value of Fisher information, and D_0 is the minimum estimation distortion given by Lemma 1. Further to characterize the number of established collaboration links, we define the percentage of collaboration links

$$\mathrm{Per}_{\mathbf{w}} := \frac{\sum_{l=1}^{L} \mathrm{card}(w_l) - N}{L - N} \times 100 \ \ (\%), \tag{6.47}$$

where $L = N^2$ is the dimension of $\mathbf{w}$, and $\mathrm{Per}_{\mathbf{w}}$ belongs to $[0, 100\%]$. When $\mathrm{Per}_{\mathbf{w}} = 0$, the network operates in a distributed manner (i.e., only the diagonal entries of $\mathbf{W}$ are nonzero). When $\mathrm{Per}_{\mathbf{w}} = 100\%$, the network is fully-connected (i.e., $\mathbf{W}$ has no zero entries).

In Fig. 6.2, we present results when we apply the reweighted ℓ_1-based ADMM algorithm to solve the information constrained problem (6.17). For comparison, we also show the results of using an exhaustive search that enumerates all possible sensor collaboration schemes, where for the tractability of an exhaustive search, we consider a small sized sensor network with $N = 5$. In the top subplot of Fig. 6.2, we present the minimum energy cost as a function of D_{norm}. We can see that the energy cost and estimation distortion is monotonically related, and the proposed approach assures near optimal performance compared to the results of exhaustive search. In the bottom subplot, we show the number of active collaboration links as a function of normalized distortion. Note that a larger estimation distortion corresponds to fewer collaboration links.

In Fig. 6.3, we present the results for the information constrained problem for a relatively large network with $N = 10$ nodes, and present the number of collaboration links and the required transmission cost as a function of the collaboration cost parameter α_c for different values of estimation
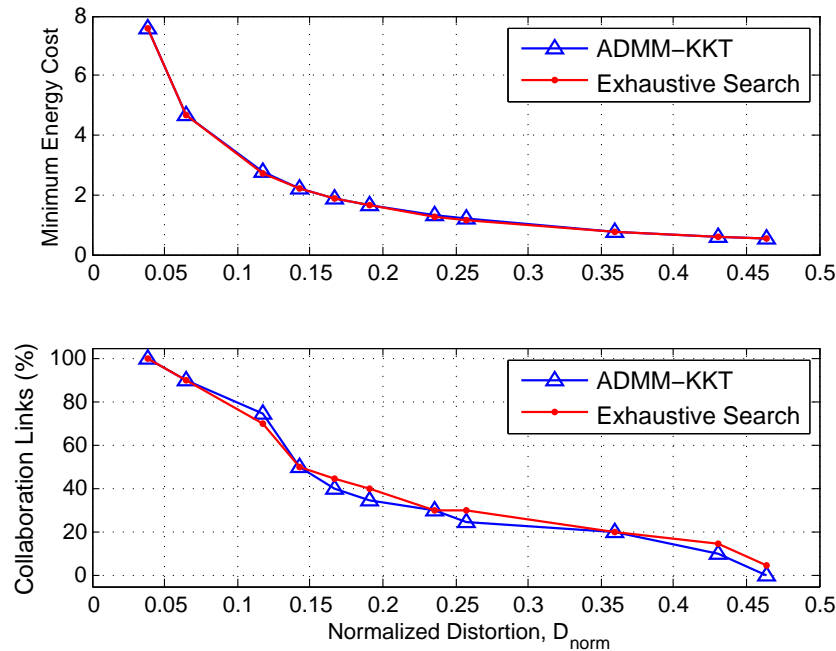
Fig. 6.2: Performance evaluation for information constrained sensor collaboration.

distortion $D_{\mathrm{norm}} \in \{0.05, 0.1, 0.2\}$. Fig. 6.3-(a) shows that the number of collaboration links increases as α_{c} decreases. This is expected, since a smaller value of α_{c} corresponds to a smaller cost of sensor collaboration, and thus encourages a larger number of collaboration links to be established. If we fix the value of α_{c}, we also observe that the number of collaboration links increases as D_{norm} decreases. This is consistent with the results in the bottom subplot of Fig 6.2. We show the specific collaboration topologies that correspond to the marked values of α_c in Fig. 6.4. These will be discussed in detail later.

Fig. 6.3-(b) shows that the transmission cost increases as α_{c} increases for a given estimation distortion. Note that a larger value of α_{c} indicates a higher cost of sensor collaboration. Therefore, to achieve a certain estimation performance, more transmission cost would be consumed instead of sensor collaboration. This implies that the transmission cost and collaboration cost are two conflicting terms. As we continue to increase α_{c}, the transmission cost converges to a fixed value for a given D_{norm}. This is because the network topology cannot be changed any further (converges to the distributed network), where the transmission cost is deterministic for the given topology and distortion.
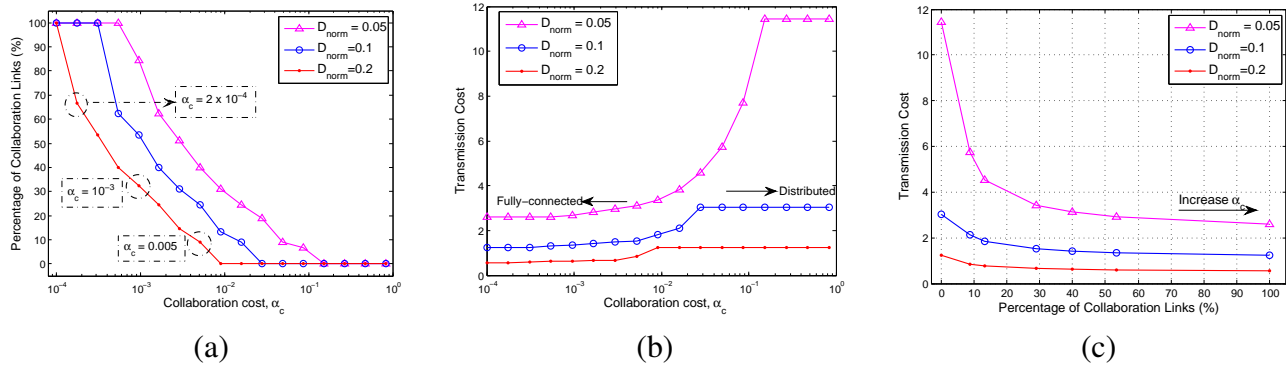
Fig. 6.3: Information constrained collaboration for different values of collaboration cost parameter α_c as $D_{norm} \in \{0.05, 0.1, 0.2\}$: (a) Percentage of collaboration links; (b) Transmission cost; (c) Trade-off between collaboration links and transmission cost.

Fig. 6.3-(c) shows the trade-off between the number of collaboration links and the consumed transmission cost by varying the parameter α_c. One interesting observation is that the transmission cost ceases to decrease significantly when over 50% collaboration links are established. The reason is that the transmission cost is characterized by the magnitude of nonzero entries in $\mathbf{w}$, which has very small increment as the number of active links is relatively large.
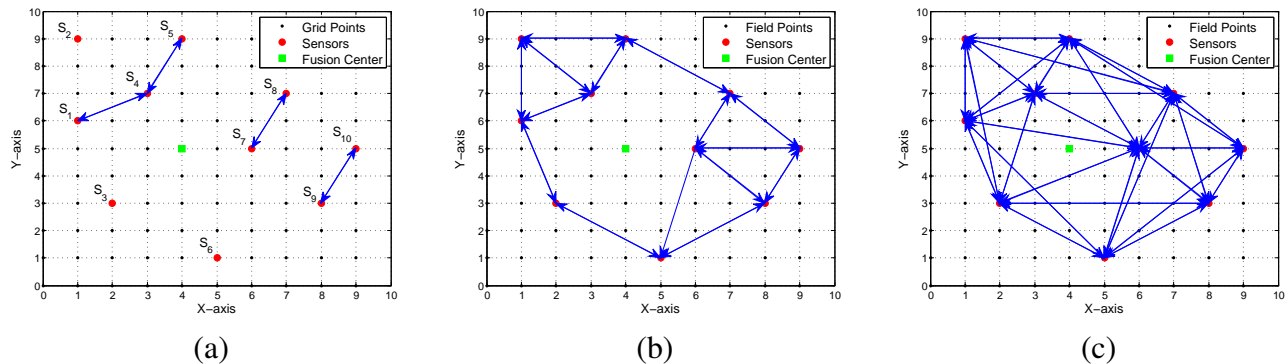
Fig. 6.4: Collaboration topologies: (a) $\alpha_c = 5 \times 10^{-3}$, $\mathrm{card}(\mathbf{w}) = 18$; (b) $\alpha_c = 10^{-3}$, $\mathrm{card}(\mathbf{w}) = 39$; (c) $\alpha_c = 2 \times 10^{-4}$, $\mathrm{card}(\mathbf{w}) = 70$.

In Fig. 6.4, we present the collaboration topologies obtained from solutions of the information constrained problem (with $D_{norm} = 0.2$) by varying the parameter of collaboration cost α_c; see the labeled points in Fig. 6.3-(a). In each subplot, the solid lines with arrows represent the collaboration links among local sensors. For example in Fig. 6.4-(a), the line between sensor 1 and sensor 4 indicates that these two sensors share measurements with each other. Fig. 6.4-(a) shows

that the nearest neighboring sensors collaborate initially because of the lower collaboration cost. We continue to decrease α_c, Fig. 6.4-(c) and (d) show that more collaboration links are established, and sensors tend to collaborate over the entire spatial field rather than aggregating in a small neighborhood.
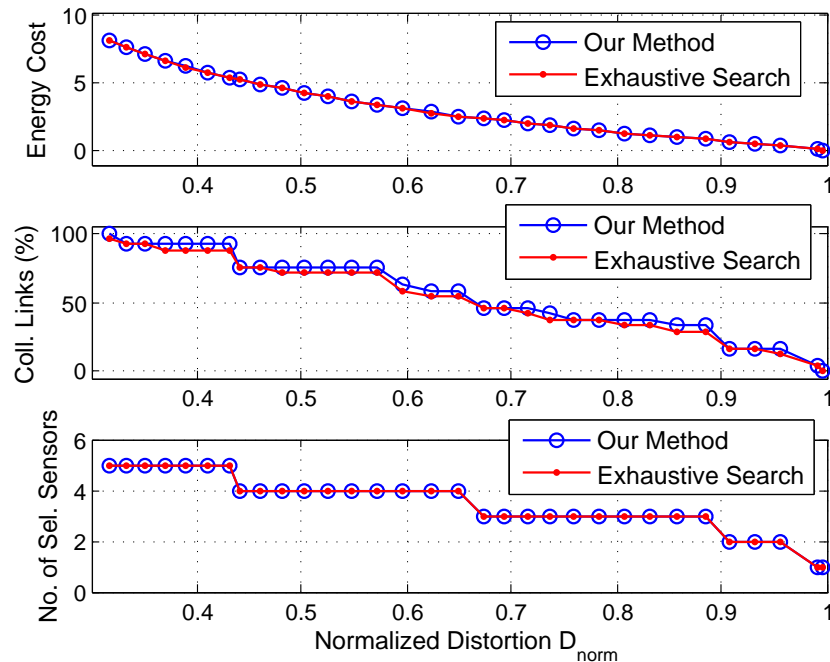
Fig. 6.5: Performance evaluation of sensor selection and collaboration.

In Fig. 6.5, we employ the convex restriction based ADMM method to solve problem (6.19) for joint sensor selection and collaboration. We show the resulting energy cost, number of collaboration links and selected sensors as functions of estimation distortion D_{norm}. For comparison, we also present the optimal results obtained from an exhaustive search, where $N = 5$ sensors are assumed in this example. We observe that the proposed approach assures near optimal performance for all values of D_{norm}. Moreover, the energy cost, number of collaboration links and selected sensors increases as D_{norm} decreases, since a smaller estimation distortion enforces more collaboration links and activated sensors.

In Fig. 6.6, we present the trade-offs between the established collaboration links and selected sensors that communicate with the FC. These trade-offs are achieved by fixing $\alpha_c = 0.1$ and
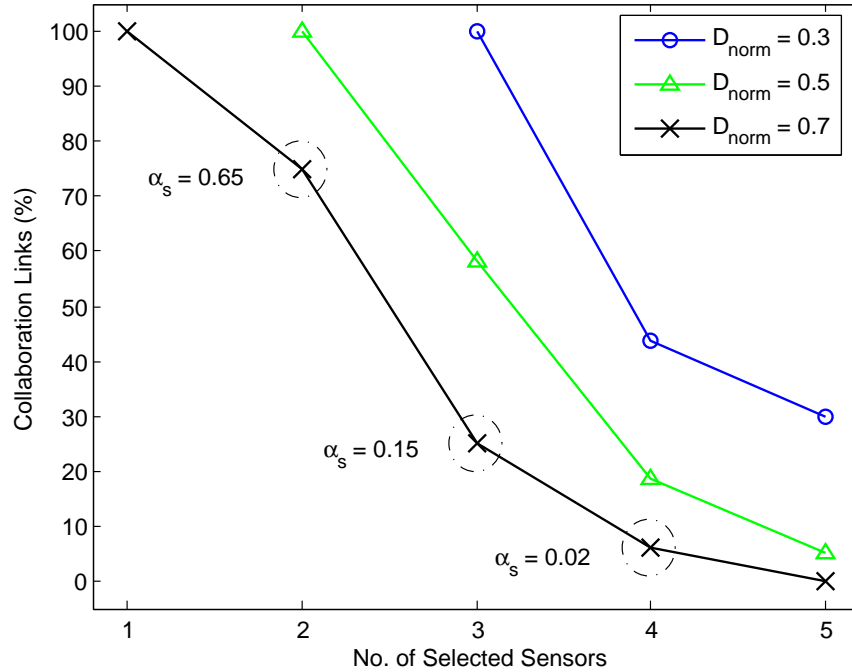
Fig. 6.6: Trade-off between collaboration links and selected sensors.

varying the parameter of sensor selection cost α_{s} for $D_{\mathrm{norm}} = 0.3$, 0.5 and 0.7. We fix D_{norm} and decrease α_{s}, which leads to an increase in the number of selected sensors, meanwhile, the number of collaboration links decreases. That is because to achieve a given estimation distortion, less collaboration links are required if more sensors are selected to communicate with the FC. If we fix the number of collaboration links, the number of selected sensors increases as D_{norm} decreases, since a smaller D_{norm} enforces more activated sensors. For the marked points as $D_{\mathrm{norm}} = 0.7$, we show the specific sensor collaboration and selection schemes in Fig. 6.7, where the solid line
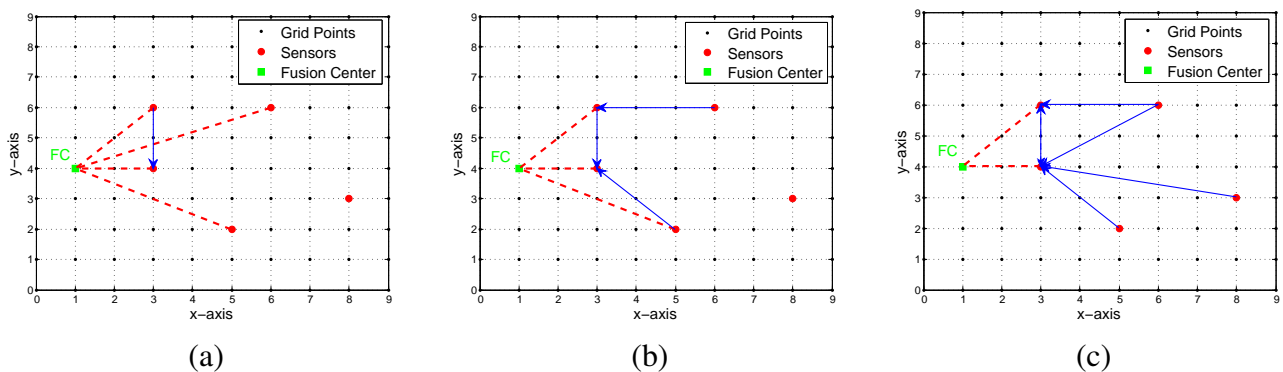
Fig. 6.7: Network topologies for $D_{\mathrm{norm}} = 0.7$ and $\alpha_{\mathrm{s}} = 0.02$, 0.15 and 0.65.

with an arrow represents the collaboration link between two sensors, and the dashed line from one sensor to the FC signifies that this sensor is selected to communicate with the FC. Clearly, more collaboration links are established as fewer sensors are selected to communicate with the FC.

6.8   Summary

In this chapter, we addressed the problem of sensor collaboration with nonzero collaboration and selection costs for distributed estimation over a coherent MAC. We showed that optimal sensor collaboration and selection schemes can be jointly designed by promoting the elementwise and rowwise sparsity of the collaboration matrix. The formulated sparsity-aware optimization problem is nonconvex in nature, and we convexified the problem by using a reweighted ℓ_1 norm and the convex-concave procedure, and solved the resulting convex program via ADMM. It was empirically shown that there exists a trade-off between sensor collaboration and sensor selection for a given estimation performance. In the next chapter, we will explore the problem of sensor collaboration for the estimation of time-varying parameters in dynamic networks.

CHAPTER 7

SENSOR COLLABORATION FOR

ESTIMATION OF DYNAMIC PARAMETERS

7.1 Introduction

In the previous chapter, sensor collaboration was studied in *static* networks, where sensors take a single snapshot of the static parameter, and then initiate sensor collaboration protocols designed in the setting of single-snapshot estimation. Here we study the problem of sensor collaboration for the estimation of *time-varying* parameters in *dynamic* networks that involve, for example, time-varying observation and channel gains. Moreover, the parameter to be estimated, such as daily temperature and precipitation in environmental monitoring [49,50], is often temporally correlated. Therefore, development of sensor collaboration schemes for the estimation of temporally correlated parameters is an important task.

In this chapter, we aim to find the optimal sensor collaboration scheme at each time step by minimizing the estimation distortion over a finite time horizon subject to individual energy constraints of sensors. Due to the presence of (a) temporal dynamics in system, (b) temporal correlation of parameter, and (c) energy constraints in time, the design of optimal sensor collaboration schemes at multiple time steps are coupled with each other, and thus poses many challenges in problem

formulation and optimization compared to the previous work. For example, when parameters of interest are temporally correlated, expressing the estimation distortion in a succinct closed form (with respect to collaboration variables) becomes intractable. It should be pointed out that even for uncorrelated parameters, finding the optimal collaboration scheme for each time step is nontrivial since energy constraints are temporally inseparable.

The rest of the chapter is organized as follows. In Section 7.2, we present the general formulation for optimal sensor collaboration in dynamic network. In Section 7.3, we discuss two types of sensor collaboration problems for the estimation of temporally uncorrelated and correlated parameters. In Section 7.4, we study the sensor collaboration problem with uncorrelated parameters. In Section 7.5, we propose efficient optimization methods to solve the sensor collaboration problem with correlated parameters. In Section 7.6, we demonstrate the effectiveness of our approach through numerical examples. Finally, in Section 7.7 we summarize our work and discuss future research directions.

7.2   Problem statement

The task here is to estimate a time-varying parameter θ_k over a time horizon of length K. As shown in Fig. 6.1, sensors first accquire their raw measurements via a linear sensing model, and then update their observations through spatial collaboration, where collaboration refers to the act of sharing measurements with neighboring sensors. The collaborative signals are then transmitted through a coherent MAC to the FC, which finally determines a global estimate of θ_k for $k \in [K]$.

The vector of measurements from N sensors at time k is given by the linear sensing model

$$\mathbf{x}_k = \mathbf{h}_k \theta_k + \boldsymbol{\epsilon}_k, \ k \in [K], \tag{7.1}$$

where for notational simplicity, let $[K]$ denote the integer set $\{1, 2, \ldots, K\}$, $\mathbf{x}_k = [x_{k,1}, \ldots, x_{k,N}]^T$ is the vector of measurements, $\mathbf{h}_k = [h_{k,1}, \ldots, h_{k,N}]^T$ is the vector of observation gains, θ_k is a random process with zero mean and variance σ_θ^2, $\boldsymbol{\epsilon}_k = [\epsilon_{k,1}, \ldots, \epsilon_{k,N}]^T$ is the vector of Gaussian

noises with i.i.d variables $\epsilon_{k,n} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ for $k \in [K]$ and $n \in [N]$.

After linear sensing, each sensor may pass its observation to another sensor for collaboration prior to transmission to the FC. With a relabelling of sensors, we assume that the first M sensors (out of a total of N sensor nodes) communicate with the FC. Collaboration among sensors is represented by a known adjacency matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ with zero-one entries, namely, $A_{mn} \in \{0, 1\}$ for $m \in [M]$ and $n \in [N]$. Here $A_{mn} = 1$ signifies that the nth sensor shares its observation with the mth sensor. Conversely, $A_{mn} = 0$ indicates the absence of a collaboration link between the nth and mth sensors.

Based on the adjacency matrix, the sensor collaboration process at time k is given by

$$\mathbf{z}_k = \mathbf{W}_k \mathbf{x}_k, \quad \mathbf{W}_k \circ (\mathbf{1}_M \mathbf{1}_N^T - \mathbf{A}) = \mathbf{0}, \ k \in [K], \tag{7.2}$$

where $\mathbf{z}_k = [z_{k,1}, z_{k,2}, \ldots, z_{k,M}]^T$, $z_{k,m}$ is the signal after collaboration at sensor m and time k, $\mathbf{W}_k \in \mathbb{R}^{M \times N}$ is the collaboration matrix that contains collaboration weights (based on the energy allocated) used to combine sensor measurements at time k, $\circ$ denotes the elementwise product, $\mathbf{1}_N$ is the $N \times 1$ vector of all ones, and $\mathbf{0}$ is the $M \times N$ matrix of all zeros. In what follows, while refering to vectors of all ones and all zeros, their dimensions will be omitted for simplicity but can be inferred from the context. In (7.2), we assume that sharing of an observation is realized through an ideal (noise-less and cost-free) communication link. The proposed ideal collaboration model enables us to obtain explicit expressions for transmission cost and estimation distortion.

After sensor collaboration, the message $\mathbf{z}_k$ is transmitted through a coherent MAC so that the received signal y_k at the FC is a coherent sum

$$y_k = \mathbf{g}_k^T \mathbf{z}_k + \varsigma_k, \ k \in [K], \tag{7.3}$$

where $\mathbf{g}_k = [g_{k,1}, g_{k,2}, \ldots, g_{k,M}]^T$ is the vector of channel gains, and ς_k is temporally white Gaussian noise with zero mean and variance σ_ς^2.

We next define the transmission cost of the mth sensor at time k, which refers to the energy

consumption of transmitting the collaborative message $\mathbf{z}_k$ to the FC. That is,

$$T_m(\mathbf{W}_k) = \mathbb{E}_{\theta_k,\epsilon_k}[z_{k,m}^2] = \mathbf{e}_m^T \mathbf{W}_k (\sigma_\theta^2 \mathbf{h}_k \mathbf{h}_k^T + \sigma_\epsilon^2 \mathbf{I}_N) \mathbf{W}_k^T \mathbf{e}_m, \tag{7.4}$$

for $m \in [M]$ and $k \in [K]$, where $\mathbf{e}_m \in \mathbb{R}^M$ is a basis vector with 1 at the mth coordinate and 0s elsewhere, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. In what follows, while refering to basis vector and identity matrix, their dimensions will be omitted for simplicity but can be inferred from the context.

From (7.1) – (7.3), the vector of received signals at the FC can be expressed as a linear function of time-varying parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^T$,

$$\mathbf{y} = \mathbf{D}_W \mathbf{D}_h \boldsymbol{\theta} + \boldsymbol{\nu}, \ \mathbf{D}_W := \text{blkdiag}\{\mathbf{g}_k^T \mathbf{W}_k\}_{k=1}^K, \tag{7.5}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_K]^T$, $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_K]^T$, $\nu_k := \mathbf{g}_k^T \mathbf{W}_k \boldsymbol{\epsilon}_k + \varsigma_k$, $\mathbf{D}_h := \text{blkdiag}\{\mathbf{h}_k\}_{k=1}^K$, and $\text{blkdiag}\{\mathbf{X}_i\}_{i=1}^n$ denotes the block-diagonal matrix with diagonal blocks $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$.

At the FC, we employ the LMMSE to estimate $\boldsymbol{\theta}$, where we assume that the FC knows the observation gains, channel gains, and the second-order statistics of the parameters of interest and additive noises. The corresponding estimator and estimation error covariance are given by [58, Theorem 10.3]

$$\begin{cases} \hat{\boldsymbol{\theta}}_W = (\boldsymbol{\Sigma}_\theta^{-1} + \mathbf{D}_h^T \mathbf{D}_W^T \mathbf{D}_\nu^{-1} \mathbf{D}_W \mathbf{D}_h)^{-1} \mathbf{D}_h^T \mathbf{D}_W^T \mathbf{D}_v^{-1} \mathbf{y} \\ \mathbf{P}_W = (\boldsymbol{\Sigma}_\theta^{-1} + \mathbf{D}_h^T \mathbf{D}_W^T \mathbf{D}_\nu^{-1} \mathbf{D}_W \mathbf{D}_h)^{-1}, \end{cases} \tag{7.6}$$

where $\boldsymbol{\Sigma}_\theta$ represents the prior knowledge about the parameter correlation, particularly $\boldsymbol{\Sigma}_\theta = \sigma_\theta^2 \mathbf{I}$ for temporally uncorrelated parameters, and $\mathbf{D}_\nu := \sigma_\epsilon^2 \mathbf{D}_W \mathbf{D}_W^T + \sigma_\varsigma^2 \mathbf{I}$. It is clear from (7.6) that both the LMMSE and the estimation error covariance matrix are functions of collaboration matrices $\{\mathbf{W}_k\}$, and their dependence on $\{\mathbf{W}_k\}$ is through $\mathbf{D}_W$. This dependency does not lend itself to easy optimization of scalar-valued functions of $\mathbf{P}_W$ for design of the optimal sensor collaboration

scheme. More insights into the LMMSE (7.6) will be provided later.

We now state the main optimization problem considered in this work for sensor collaboration

$$
\begin{aligned}
& \text{minimize} \quad \text{tr}\left(\mathbf{P}_W\right) \\
& \text{subject to} \quad \sum_{k=1}^{K} T_m(\mathbf{W}_k) \leq E_m, \ m \in [M] \\
& \qquad\qquad \mathbf{W}_k \circ (\mathbf{1}_M \mathbf{1}_N^T - \mathbf{A}) = \mathbf{0}, \ k \in [K],
\end{aligned}
\tag{7.7}
$$

where $\mathbf{W}_k$ is the optimization variable for $k \in [K]$, $\text{tr}(\mathbf{P}_W)$ denotes the estimation distortion of using the LMMSE, which has a special form shown in the next section if parameters are uncorrelated or the correlation prior is not available, $T_m(\mathbf{W}_k)$ is the transmission cost given by (7.4), E_m is a prescribed energy budget of the mth sensor, and $\mathbf{A}$ is the adjacency matrix to characterize the network topology. Although sensor collaboration is performed with respect to a time-invariant (fixed) topology matrix $\mathbf{A}$, energy allocation in terms of the magnitude of nonzero entries in $\mathbf{W}_t$ is time varying in the presence of temporal dynamics (e.g., time-varying observation and channel gains) of the sensor network. As will be evident later, the proposed sensor collaboration approach is also applicable to the problem with time-varying topologies. The problem structure and the solution of (7.7) will be elaborated on in the following sections.

7.3 Reformulation and simplification

In what follows, we simplify problem (7.7) by exploiting the sparsity structure of the adjacency matrix and concatenating the nonzero entries of a collaboration matrix into a collaboration vector. We show that the resulting optimization problem involves special types of nonconvexities.

In problem (7.7), the only unknowns are the nonzero entries of collaboration matrices. Motivated by that, we concatenate these nonzero entries (columnwise) into a collaboration vector

$$
\mathbf{w}_k = [w_{k,1}, w_{k,2}, \ldots, w_{k,L}]^T,
\tag{7.8}
$$

where $w_{k,l}$ denotes the lth entry of $\mathbf{w}_k$, and L is the number of nonzero entries of the adjacency matrix $\mathbf{A}$. We note that given $w_{k,l}$, there exists a row index m_l and a column index n_l such that $w_{k,l} = [\mathbf{W}_k]_{m_l n_l}$, where $[\mathbf{X}]_{mn}$ (or X_{mn}) denotes the (m, n)th entry of a matrix $\mathbf{X}$. We demonstrate the vectorization of $\mathbf{W}_k$ through an example in Fig. 7.1, where we consider $N = 3$ sensor nodes, $M = 3$ communicating nodes, and 2 collaboration links.
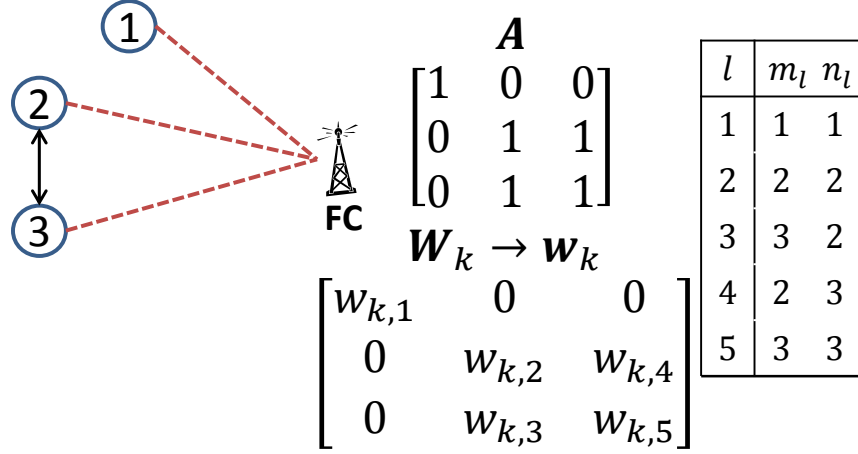
Fig. 7.1: Example of vectorization of $\mathbf{W}_k$.

Collaboration problem for estimation of uncorrelated parameters

When the parameters of interest are uncorrelated, the estimation error covariance matrix (7.6) simplifies to

$$
\begin{aligned}
\mathbf{P}_W &= \left(\sigma_\theta^{-2} \mathbf{I} + \mathbf{D}_h^T \mathbf{D}_w^T (\sigma_\epsilon^2 \mathbf{D}_w \mathbf{D}_w^T + \sigma_\varsigma^2 \mathbf{I})^{-1} \mathbf{D}_w \mathbf{D}_h \right)^{-1} \\
&= \left(\sigma_\theta^{-2} \mathbf{I} + \operatorname{diag} \left\{ \frac{\mathbf{g}_k^T \mathbf{W}_k \mathbf{h}_k \mathbf{h}_k^T \mathbf{W}_k^T \mathbf{g}_k}{\sigma_\epsilon^2 \mathbf{g}_t^T \mathbf{W}_k \mathbf{W}_k^T \mathbf{g}_t + \sigma_\varsigma^2} \right\}_{k=1}^K \right)^{-1} \\
&= \operatorname{diag} \left\{ \frac{\sigma_\theta^2 \sigma_\epsilon^2 \mathbf{g}_k^T \mathbf{W}_k \mathbf{W}_k^T \mathbf{g}_k + \sigma_\theta^2 \sigma_\varsigma^2}{\sigma_\theta^2 \mathbf{g}_k^T \mathbf{W}_k \mathbf{h}_k \mathbf{h}_k^T \mathbf{W}_k^T \mathbf{g}_k + \sigma_\epsilon^2 \mathbf{g}_k^T \mathbf{W}_k \mathbf{W}_k^T \mathbf{g}_k + \sigma_\varsigma^2} \right\}_{k=1}^K,
\end{aligned}
\tag{7.9}
$$

where $\operatorname{diag}\{a_k\}_{k=1}^K$ denotes a diagonal matrix with diagonal entries $a_1, a_2, \ldots, a_K$.

It is clear from (7.9) and (7.4) both the estimation error covariance matrix and the transmission cost contain quadratic matrix functions [98], which can be converted into quadratic vector

functions according to the relationship between $\mathbf{W}_k$ and $\mathbf{w}_k$ stated in Proposition 7.1.

Proposition 7.1. *Let* $\mathbf{w} \in \mathbb{R}^L$ *be the vector of stacking the nonzero entries of* $\mathbf{W} \in \mathbb{R}^{M \times N}$ *columnwise, the expression* $\mathbf{b}^T\mathbf{W}$ *can be written as a function of* $\mathbf{w}$,

$$\mathbf{b}^T\mathbf{W} = \mathbf{w}^T\mathbf{B}, \tag{7.10}$$

where $\mathbf{b} \in \mathbb{R}^N$ *is a coefficient vector,* $\mathbf{B}$ *is an* $L \times N$ *matrix whose* (l, n)*th entry is given by*

$$B_{ln} = \begin{cases} b_{m_l} & n = n_l \\ 0 & \text{otherwise}, \end{cases} \tag{7.11}$$

and the indices m_l *and* n_l *satisfy that* $w_l = W_{m_l n_l}$ *for* $l \in [L]$.

Proof: The proof follows from [48, Sec. III-A]. ∎

From (7.9) and Proposition 7.1, the objective function of problem (7.7) can be rewritten as

$$\phi(\mathbf{w}) := \text{tr}(\mathbf{P}_W) = \sum_{k=1}^{K} \frac{\sigma_\theta^2 \sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\theta^2 \sigma_\varsigma^2}{\mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\varsigma^2}, \tag{7.12}$$

where we use the fact that $\mathbf{g}_k^T \mathbf{W}_k = \mathbf{w}_k^T \mathbf{G}_k$, $\mathbf{G}_k$ is an $L \times N$ matrix defined as (7.11), $\mathbf{R}_k :=$ $\mathbf{G}_k \mathbf{G}_k^T$, and $\mathbf{S}_k := \mathbf{G}_k(\sigma_\theta^2 \mathbf{h}_k \mathbf{h}_k^T + \sigma_\epsilon^2 \mathbf{I})\mathbf{G}_k^T$. Clearly, both $\mathbf{R}_k$ and $\mathbf{S}_k$ are positive semidefinite matrices.

Moreover, the transmission cost (7.4) can be rewritten as

$$T_m(\mathbf{w}_k) := \mathbf{w}_k^T \mathbf{Q}_{k,m} \mathbf{w}_k, \tag{7.13}$$

where $\mathbf{Q}_{k,m} := \mathbf{E}_m(\sigma_{\theta,k}^2 \mathbf{h}_k \mathbf{h}_k^T + \sigma_\epsilon^2 \mathbf{I})\mathbf{E}_m^T$, and $\mathbf{E}_m$ is defined as (7.11) such that $\mathbf{e}_m^T \mathbf{W}_k = \mathbf{w}_k^T \mathbf{E}_m$. We remark that $\mathbf{Q}_{k,m}$ is positive semidefinite for $k \in [K]$ and $m \in [M]$.

From (7.12) and (7.13), the sensor collaboration problem for the estimation of temporally un-

correlated parameters becomes

$$
\begin{aligned}
\text{minimize} \quad & \phi(\mathbf{w}) \\
\text{subject to} \quad & \sum_{k=1}^{K} \mathbf{w}_k^T \mathbf{Q}_{k,m} \mathbf{w}_k \leq E_m, \ m \in [M],
\end{aligned}
\tag{7.14}
$$

where $\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_K^T]$ is the optimization variable, and $\phi(\mathbf{w})$ is the estimation distortion given by (7.12). Note that (7.14) cannot be decomposed in time since sensor energy constraints are temporally inseparable.

Compared to problem (7.7), the topology constraint in terms of the adjacency matrix is eliminated without loss of performance in (7.14) since the sparsity structure of the adjacency matrix has been taken into account while constructing the collaboration vector. (7.14) is a nonconvex optimization problem since its objective function is given by a sum of rational functions [106]. In the case of single-snapshot estimation (namely, $K = 1$), the objective function of (7.14) simplifies to a single quadratic ratio. It has been shown in [48] that such a nonconvex problem can be readily solved via convex programming. In contrast, the presence of the sum of quadratic ratios makes solving (7.14) more challenging. We present the solution to (7.14) in Section 7.4.

Collaboration problem for estimation of correlated parameters

When parameters are temporally correlated, the covariance matrix $\boldsymbol{\Sigma}_\theta$ is no longer diagonal. As a result, expressing the estimation error in a succinct form as in (7.12) becomes intractable. We recall from (7.6) that the dependence of the estimation error covariance on collaboration matrices is through $\mathbf{D}_W$. According to the matrix inversion lemma, we are able to simplify (7.6) via the relationship

$$
\mathbf{D}_W^T(\sigma_\epsilon^2 \mathbf{D}_W \mathbf{D}_W^T + \sigma_\varsigma^2 \mathbf{I})^{-1} \mathbf{D}_w = \sigma_\epsilon^{-2} \mathbf{I} - (\sigma_\epsilon^2 \mathbf{I} + \sigma_\epsilon^4 \sigma_\varsigma^{-2} \mathbf{D}_W^T \mathbf{D}_W)^{-1}.
\tag{7.15}
$$

Substituting (7.15) into (7.6), we obtain

$$\mathbf{P}_W = \left(\mathbf{C} - \sigma_\epsilon^{-2}\mathbf{D}_h^T(\mathbf{I} + \sigma_\epsilon^2\sigma_\varsigma^{-2}\mathbf{D}_W^T\mathbf{D}_W)^{-1}\mathbf{D}_h\right)^{-1} \tag{7.16}$$

where $\mathbf{C} := \boldsymbol{\Sigma}_\theta^{-1} + \sigma_\epsilon^{-2}\mathbf{D}_h^T\mathbf{D}_h$. According to the definition of $\mathbf{D}_W$ in (A.55), we obtain

$$\mathbf{D}_W^T\mathbf{D}_W = \mathrm{blkdiag}\{\mathbf{W}_k^T\mathbf{g}_k\mathbf{g}_k^T\mathbf{W}_k\}_{k=1}^K = \mathrm{blkdiag}\{\mathbf{G}_k^T\mathbf{w}_k\mathbf{w}_k^T\mathbf{G}_k\}_{k=1}^K, \tag{7.17}$$

where $\mathbf{G}_k$ has been introduced in (7.12).

Combining (7.16) and (7.17), we can rewrite the estimation error covariance as a function of the collaboration vector

$$\mathbf{P}_w := \left(\mathbf{C} - \sigma_\epsilon^{-2}\,\mathrm{diag}\left\{\mathbf{h}_k^T\left(\mathbf{I} + \sigma_\epsilon^2\sigma_\varsigma^{-2}\mathbf{G}_k^T\mathbf{w}_k\mathbf{w}_k^T\mathbf{G}_k\right)^{-1}\mathbf{h}_k\right\}_{k=1}^K\right)^{-1}. \tag{7.18}$$

From (7.18), the sensor collaboration problem for the estimation of temporally correlated parameters becomes

$$\begin{aligned}
&\text{minimize} \quad \mathrm{tr}(\mathbf{P}_w) \\
&\text{subject to} \quad \sum_{k=1}^K \mathbf{w}_k^T\mathbf{Q}_{k,m}\mathbf{w}_k \le E_m,\ m \in [M],
\end{aligned} \tag{7.19}$$

where $\mathbf{w}$ is the optimization variable, and the matrix $\boldsymbol{\Omega}_m$ is defined in (7.14). Note that (7.19) is not a convex optimization problem due to the presence of the rank-one matrix $\mathbf{w}_k\mathbf{w}_k^T$ in (7.18). However, such a nonconvexity can be effectively handled via proper convexification techniques. We will present the solution to (7.19) in Section 7.5.

We finally remark that the proposed sensor collaboration methodologies in this paper apply equally well to the case of sensor collaboration with respect to time-varying topologies, namely, $\mathbf{A}_t$ for $t \in [T]$. The only difference from sensor collaboration with a fixed topology is that the collaboration vector $\mathbf{w}_t$ would be constructed by concatenating the nonzero entries of $\mathbf{W}_t$ according

to $\mathbf{A}_t$ rather than $\mathbf{A}$ at each time step.

7.4 Sensor collaboration for estimation of uncorrelated parameters

In this section, we show that (7.14) can be transformed into a special nonconvex optimization problem, where the difference of convex (DC) functions carries all the nonconvexity. Spurred by the problem structure, we employ a convex-concave procedure (CCP) to solve (7.14).

Equivalent optimization problem

We express (7.14) in its epigraph form

$$\text{minimize} \quad \mathbf{1}^T \mathbf{u} \tag{7.20a}$$

$$\text{subject to} \quad \frac{\sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\varsigma^2}{\mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\varsigma^2} \leq u_k, \ k \in [K] \tag{7.20b}$$

$$\mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \ m \in [M], \tag{7.20c}$$

where $\mathbf{u} = [u_1, u_2, \ldots, u_K]^T$ is the vector of newly introduced optimization variables, and $\mathbf{Q}_m :=$ blkdiag$\{\mathbf{Q}_{k,m}\}_{k=1}^K$. We note that the inequality (7.20b) implicitly adds the additional constraint $u_k \geq 0$ since both $\mathbf{R}_k$ and $\mathbf{S}_k$ are positive semidefinite for $k \in [K]$.

We further introduce new variables r_k and s_k for $k \in [K]$ to rewrite (7.20b) as

$$\frac{r_k}{s_k} \leq u_k, \ s_k > 0, \ \mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\varsigma^2 \geq s_k, \ \sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\varsigma^2 \leq r_k, \tag{7.21}$$

where the equivalence between (7.20b) and (7.21) holds since the minimization of $\mathbf{1}^T \mathbf{u}$ with the above inequalities forces the variable s_k and r_k to achieve its upper and lower bound, respectively.

In (7.21), the ratio $r_k/s_k \leq u_k$ together with $s_k > 0$ can be reformulated as a quadratic

inequality of DC type

$$s_k^2 + u_k^2 + 2r_k - (s_k + u_k)^2 \leq 0, \tag{7.22}$$

where both $s_k^2 + u_k^2 + 2r_k$ and $(s_k + u_k)^2$ are convex quadratic functions.

From (7.21) and (7.22), problem (7.20) becomes

$$\begin{align}
\text{minimize} \quad & \mathbf{1}^T \mathbf{u} \tag{7.23a} \\
\text{subject to} \quad & s_k^2 + u_k^2 + 2r_k \leq (s_k + u_k)^2, \ k \in [K] \tag{7.23b} \\
& s_k - \mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k - \sigma_\varsigma^2 \leq 0, \ k \in [K] \tag{7.23c} \\
& \sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\varsigma^2 \leq r_k, \ k \in [K] \tag{7.23d} \\
& \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \ m \in [M] \tag{7.23e} \\
& \mathbf{s} > \mathbf{0}, \tag{7.23f}
\end{align}$$

where the optimization variables are $\mathbf{w}$, $\mathbf{u}$, $\mathbf{r}$ and $\mathbf{s}$, $\mathbf{r} = [r_1, r_2, \ldots, r_K]^T$, $\mathbf{s} = [s_1, s_2, \ldots, s_K]^T$, and $>$ denotes elementwise inequality. Note that the quadratic functions of DC type in (7.23b) and (7.23c) bring in the nonconvexity of problem (7.23). In what follows, we will show that CCP is a suitable convex restriction approach for solving this problem.

Convex restriction

Problem (7.23) is convex except for the nonconvex quadratic constraints (7.23b) and (7.23c), which have the DC form

$$f(\mathbf{v}) - g(\mathbf{v}) \leq 0, \tag{7.24}$$

where both f and g are convex functions. In (7.23b), we have $f(s_k, u_k, r_k) = s_k^2 + u_k^2 + 2r_k$, and $g(s_k, u_k) = (s_k + u_k)^2$. In (7.23c), $f(s_k) = s_k$, and $g(\mathbf{w}_k) = \mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\varsigma^2$.

We can convexify (7.24) by linearizing g around a feasible point $\hat{\mathbf{v}}$,

$$f(\mathbf{v}) - \hat{g}(\mathbf{v}) \leq 0, \tag{7.25}$$

where $\hat{g}(\mathbf{v}) := g(\hat{\mathbf{v}}) + (\mathbf{v} - \hat{\mathbf{v}})^T \frac{\partial g(\hat{\mathbf{v}})}{\partial \mathbf{v}}$, $\frac{\partial g(\hat{\mathbf{v}})}{\partial \mathbf{v}}$ is the first-order derivative of g at the point $\hat{\mathbf{v}}$. In (7.25), $\hat{g}$ is an affine *lower bound* on the convex function g, and therefore, the set of $\mathbf{v}$ that satisfy (7.25) is a strict *subset* of the set of $\mathbf{v}$ that satisfy (7.24). This implies that a solution of the optimization problem with the linearized constraint (7.25) is locally optimal for the problem with the original nonconvex constraint (7.24).

We can obtain a 'restricted' convex version of problem (7.23) by linearizing (7.23b) and (7.23c) as (7.25). We then solve a sequence of convex programs with iteratively updated linearization points. The use of linearization to convexify nonconvex problems with DC type functions is known as CCP [102, 103]. At each iteration of CCP, we solve

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{1}^T \mathbf{u} \\
\text{subject to} \quad & s_k^2 + u_k^2 + 2r_k - \hat{g}_1(s_k, u_k) \leq 0, \ k \in [K] \\
& s_k - \hat{g}_2(\mathbf{w}_k) \leq 0, \ k \in [K] \\
& \sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\varsigma^2 \leq r_k, \ k \in [K] \\
& \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \ m \in [M] \\
& \mathbf{s} > \mathbf{0},
\end{aligned}
\tag{7.26}
$$

where the optimization variables are $\mathbf{w}$, $\mathbf{u}$, $\mathbf{r}$, and $\mathbf{s}$, $\hat{g}_1$ and $\hat{g}_2$ are affine approximations of $(s_k + u_k)^2$ and $\mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\varsigma^2$, namely, $\hat{g}_1(s_k, u_k) := 2(s_k + u_k)(\hat{s}_k + \hat{u}_k) - (\hat{s}_k + \hat{u}_k)^2$, and $\hat{g}_2(\mathbf{w}_k) := 2\hat{\mathbf{w}}_k^T \mathbf{S}_k \mathbf{w}_k - \hat{\mathbf{w}}_k^T \mathbf{S}_k \hat{\mathbf{w}}_k + \sigma_\varsigma^2$.

The convergence of CCP is guaranteed, since at each iteration, we solve a restricted convex problem with a smaller feasible set which contains the linearization point (i.e., the solution after the previous iteration) [103]. In other words, for a given linearization point, we always obtain a new feasible point with a lower or equal objective value at each iteration.

The computation cost of CCP is dominated by the solution of the convex program with quadratic constraints. This has the computational complexity $O(a^3 + a^2 b)$ in the use of interior-point algorithm [104, Chapter. 10], where a and b denote the number of optimization variables and constraints, respectively. In problem (7.26), we have $a = 3K + KL$ and $b = 4K + M$. Therefore, the complexity of our algorithm is roughly given by $O(L^3)$ per iteration. Here we focus on the scenario in which the number of collaboration links L is much larger than K or M.

7.5 Sensor collaboration for estimation of correlated parameters

Different from (7.14), the presence of temporal correlation makes finding the solution of (7.19) more challenging. However, we show that (7.19) can be cast as an SDP with a (nonconvex) rank-one constraint. Spurred by the problem structure, we employ a *penalty CCP* to solve (7.19), and propose a fast optimization algorithm by using ADMM.

Equivalent optimization problem

From (7.18), problem (7.19) can be equivalently transformed to

$$
\begin{aligned}
& \text{minimize} && \mathrm{tr}(\mathbf{V}) \\
& \text{subject to} && \mathbf{P}_w^{-1} \succeq \mathbf{V}^{-1} \\
& && \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \; m \in [M] \\
& && \mathbf{U}_k = \mathbf{w}_k \mathbf{w}_k^T, \; k \in [K],
\end{aligned}
\tag{7.27}
$$

where $\mathbf{V} \in \mathbb{S}^K$ and $\mathbf{U}_k \in \mathbb{S}^L$ are newly introduced optimization variables for $k \in [K]$, $\mathbb{S}^n$ represents the set of $n \times n$ symmetric matrices, and the notation $\mathbf{X} \succeq \mathbf{Y}$ (or $\mathbf{X} \preceq \mathbf{Y}$) indicates that $\mathbf{X} - \mathbf{Y}$ (or $\mathbf{Y} - \mathbf{X}$) is positive semidefinite. The first inequality constraint of problem (7.27) is obtained from $\mathbf{P}_w \preceq \mathbf{V}$, where $\mathbf{P}_w$ is given by (7.18), and $\mathbf{P}_w^{-1}$ represents Bayesian Fisher

information corresponding to LMMSE.

We further introduce a new vector of optimization variables $\mathbf{p} = [p_1, \ldots, p_K]^T$ such that the first matrix inequality of problem (7.27) is expressed as

$$\mathbf{C} - \mathrm{diag}(\mathbf{p}) \succeq \mathbf{V}^{-1}, \tag{7.28}$$

$$p_k \geq \sigma_\epsilon^{-2} \mathbf{h}_k^T \left(\mathbf{I} + \sigma_\epsilon^2 \sigma_\varsigma^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \right)^{-1} \mathbf{h}_k, \ k \in [K], \tag{7.29}$$

where we use the expression of $\mathbf{P}_w$ given by (7.18), and the fact that $\mathbf{U}_k = \mathbf{w}_k \mathbf{w}_k^T$. Note that the minimization of $\mathrm{tr}(\mathbf{V})$ with inequalities (7.28) and (7.29) would force the variable p_k to achieve its lower bound. In other words, problem (7.27) is equivalent to the problem in which the first inequality constraint of (7.27) is replaced by the above two inequalities.

By employing the Schur complement, we can express (7.28) and (7.29) as LMIs

$$\begin{bmatrix} \mathbf{C} - \mathrm{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} p_k & \sigma_\epsilon^{-1} \mathbf{h}_k^T \\ \sigma_\epsilon^{-1} \mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\varsigma^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} \succeq 0, \ k \in [K]. \tag{7.30}$$

Replacing the first inequality of problem (7.27) with LMIs (7.30), we obtain an SDP together with a (nonconvex) rank-one constraint $\mathbf{U}_k = \mathbf{w}_k \mathbf{w}_k^T$. This nonconvex constraint can be recast as two inequalities

$$\mathbf{U}_k - \mathbf{w}_k \mathbf{w}_k^T \succeq 0, \ \mathbf{U}_k - \mathbf{w}_k \mathbf{w}_k^T \preceq 0, \ k \in [K]. \tag{7.31}$$

According to the Schur complement, the first matrix inequality is equivalent to an LMI

$$\begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} \succeq 0, \ k \in [K]. \tag{7.32}$$

And the second inequality in (7.31) involves a function of DC type, where $\mathbf{U}_k$ and $\mathbf{w}_k \mathbf{w}_k^T$ are matrix convex functions [59].

From (7.30) – (7.32), problem (7.27) or (7.19) is equivalent to

$$\text{minimize} \quad \text{tr}(\mathbf{V}) \tag{7.33a}$$

$$\text{subject to} \quad \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \ m \in [M] \tag{7.33b}$$

$$\text{LMIs in (7.30)} \tag{7.33c}$$

$$\text{LMIs in (7.32)} \tag{7.33d}$$

$$\mathbf{U}_k - \mathbf{w}_k \mathbf{w}_k^T \preceq 0, \ k \in [K], \tag{7.33e}$$

where the optimization variables are $\mathbf{w}$, $\mathbf{p}$, $\mathbf{V}$ and $\mathbf{U}_k$ for $k \in [K]$, and (7.33e) is a nonconvex constraint of DC type.

Convexification

Proceeding with the same logic as in the previous section to convexify the constraint (7.24), we linearize (7.33e) around a point $\hat{\mathbf{w}}_k$,

$$\mathbf{U}_k - \hat{\mathbf{w}}_k \mathbf{w}_k^T - \mathbf{w}_k \hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T \preceq 0, \ k \in [K]. \tag{7.34}$$

It is straightforward to apply CCP to solve problem (7.33) by replacing (7.33e) with (7.34). However, such an approach fails in practice. This is not surprising, since the feasible set determined by (7.33d) and (7.34) only contains the linearization point. Specifically, from (7.33d) and (7.34), we obtain

$$(\mathbf{w}_k - \hat{\mathbf{w}}_k)(\mathbf{w}_k - \hat{\mathbf{w}}_k)^T = \mathbf{w}_k \mathbf{w}_k^T - \hat{\mathbf{w}}_k \mathbf{w}_k^T - \mathbf{w}_k \hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T$$

$$\preceq \mathbf{U}_k - \hat{\mathbf{w}}_k \mathbf{w}_k^T - \mathbf{w}_k \hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T \preceq 0, \tag{7.35}$$

which indicates that $\mathbf{w}_k = \hat{\mathbf{w}}_k$. Therefore, CCP gets trapped in a linearization point.

Remark 4. *Dropping the nonconvex constraint* (7.33e) *is another method to convexify problem*

(7.33), *known as semidefinite relaxation [63]. However, such an approach makes the optimization variable* $\mathbf{U}_k$ *unbounded, since the minimization of* $\mathrm{tr}(\mathbf{V})$ *forces* $\mathbf{U}_k$ *to be as large as possible such that the variable* p_k *in (7.29) is as small as possible.*

In order to circumvent the drawback of the standard CCP, we consider its penalized version, known as penalty CCP [103], where we add new variables to allow for constraints (7.34) to be violated and penalize the sum of the violations in the objective function. As a result, the convexification (7.34) is modified by

$$\mathbf{U}_k - \hat{\mathbf{w}}_k\mathbf{w}_k^T - \mathbf{w}_k\hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k\hat{\mathbf{w}}_k^T \preceq \mathbf{Z}_k, \ k \in [K], \tag{7.36}$$

where $\mathbf{Z}_k \in \mathbb{S}^L$ is a newly introduced variable. The constraint (7.36) implicitly adds the additional constraint $\mathbf{Z}_k \succeq 0$ due to $\mathbf{U}_k \succeq \mathbf{w}_k\mathbf{w}_k$ from (7.33d).

After replacing (7.33e) with (7.36), we obtain the SDP,

$$\begin{aligned} &\text{minimize} \quad \mathrm{tr}(\mathbf{V}) + \tau \sum_{k=1}^{K} \mathrm{tr}(\mathbf{Z}_k) \\ &\text{subject to} \quad (7.33\text{b}) - (7.33\text{d}) \text{ and } (7.36) \end{aligned} \tag{7.37}$$

where the optimization variables are $\mathbf{w}$, $\mathbf{p}$, $\mathbf{V}$, $\mathbf{U}_k$ and $\mathbf{Z}_k$ for $k \in [K]$, and $\tau > 0$ is a penalty parameter. Compared to the standard CCP, problem (7.37) is optimized over a larger feasible set since we allow for constraints to be violated by adding variables $\mathbf{Z}_k$ for $k \in [K]$. We summarize the use of penalty CCP to solve (7.19) in Algorithm 7.1.

In Algorithm 7.1, the initial point $\hat{\mathbf{w}}$ is randomly picked from a standard uniform distribution. Note that $\hat{\mathbf{w}}$ is not necessarily feasible for (7.19) since violations of constraints are allowed. We also remark that when $\tau = \tau_{\max}$, penalty CCP reduces to CCP, and therefore, its convergence is guaranteed [103].

The computation cost of Algorithm 7.1 is dominated by the solution of the SDP (7.37) at Step 2. This leads to the complexity $O(a^2b^2 + ab^3)$ by using the interior-point alogrithm in off-the-shelf

solvers [104, Chapter. 11], where a and b are the number of optimization variables and the size of the semidefinite matrix, respectively. In (7.37), the number of optimization variables is proportional to L^2. Therefore, the complexity of Algorithm 7.1 is roughly given by $O(L^6)$. Clearly, computing solutions to SDPs becomes inefficient for problems of medium or large size. In what follows, we will develop an ADMM-based algorithm that is more amenable to large-scale optimization.

Algorithm 7.1 Penalty CCP for solving (7.19)

Require: an initial point $\hat{\mathbf{w}}$, $\epsilon_{\text{ccp}} > 0$, $\tau^0 > 0$, $\tau_{\max} > 0$ and $\mu > 1$.
 1: **for** iteration $t = 1, 2, \ldots$ **do**
 2:      solve problem (7.37) for its solution $\mathbf{w}^t$ via SDP solver or ADMM-based algorithm
 3:      update the linearization point, $\hat{\mathbf{w}} = \mathbf{w}^t$
 4:      update the penalty parameter $\tau^t = \min\{\mu\tau^{t-1}, \tau_{\max}\}$
 5:      let ψ^t be the objective value of (7.37)
 6:      **until** $|\psi^t - \psi^{t-1}| \leq \epsilon_{\text{ccp}}$ with $t \geq 2$.
 7: **end for**

Fast algorithm via ADMM

It has been shown in [60, 93, 107, 108] that ADMM is a powerful tool for solving large-scale optimization problems. The major advantage of ADMM is that it allows us to split the original problem into subproblems, each of which can be solved more efficiently or even analytically. In what follows, we will employ ADMM to solve problem (7.37).

We begin by reformulating problem (7.37) in a way that lends itself to the application of ADMM. We introduce slack variables $\boldsymbol{\lambda}_m \in \mathbb{R}^{KL+1}$ for $m \in [M]$ to rewrite (7.33b) as an equality constraint together with a second-order cone constraint,

$$\bar{\mathbf{Q}}_m \mathbf{w} - \boldsymbol{\lambda}_m + \mathbf{c}_m = \mathbf{0}, \; \|[\boldsymbol{\lambda}_m]_{1:KL}\|_2 \leq [\boldsymbol{\lambda}_m]_{KL+1}, \tag{7.38}$$

where $\bar{\mathbf{Q}}_m := [\mathbf{Q}_m^{\frac{1}{2}}, \mathbf{0}]^T$, $\mathbf{Q}_m^{\frac{1}{2}}$ is the square root of $\mathbf{Q}_m$ given by the matrix decomposition $\mathbf{Q}_m = (\mathbf{Q}_m^{\frac{1}{2}})^T \mathbf{Q}_m^{\frac{1}{2}}$, $\mathbf{c}_m = [\mathbf{0}^T, \sqrt{E_m}]^T$, and $[\mathbf{a}]_{1:n}$ denotes a subvector of $\mathbf{a}$ that consists of its first n entries.

We further introduce slack variables $\mathbf{\Lambda}_1 \in \mathbb{S}^{2K}$, $\mathbf{\Lambda}_{2,k} \in \mathbb{S}^{N+1}$, $\mathbf{\Lambda}_{3,k} \in \mathbb{S}^{L+1}$ and $\mathbf{\Lambda}_{4,k} \in \mathbb{S}^L$ for $k \in [K]$ to rewrite LMIs of problem (7.37) as a sequence of equality constraints together with positive semidefinite cone constraints

$$\begin{bmatrix} \mathbf{C} - \mathrm{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} - \mathbf{\Lambda}_1 = \mathbf{0} \tag{7.39}$$

$$\begin{bmatrix} p_k & \sigma_\epsilon^{-1}\mathbf{h}_k^T \\ \sigma_\epsilon^{-1}\mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\varsigma^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} - \mathbf{\Lambda}_{2,k} = \mathbf{0} \tag{7.40}$$

$$\begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} - \mathbf{\Lambda}_{3,k} = \mathbf{0} \tag{7.41}$$

$$\mathbf{Z}_k - \mathbf{U}_k + \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T - \mathbf{\Lambda}_{4,k} = \mathbf{0}, \tag{7.42}$$

where $\mathbf{\Lambda}_1 \succeq 0$, $\mathbf{\Lambda}_{2,k} \succeq 0$, $\mathbf{\Lambda}_{3,k} \succeq 0$, and $\mathbf{\Lambda}_{4,k} \succeq 0$ for $k \in [K]$. From $(7.38)-(7.42)$, problem (7.37) becomes

$$\begin{aligned} \text{minimize} \quad & \mathrm{tr}(\mathbf{V}) + \tau \sum_{k=1}^K \mathrm{tr}(\mathbf{Z}_k) + \sum_{m=1}^M \mathcal{I}_0(\boldsymbol{\lambda}_m) + \mathcal{I}_1(\mathbf{\Lambda}_1) + \sum_{i=2}^4 \sum_{k=1}^K \mathcal{I}_i(\mathbf{\Lambda}_{i,k}) \\ & \tag{7.43} \\ \text{subject to} \quad & \text{equality constraints in } (7.38)-(7.42), \end{aligned}$$

where the optimization variables are $\mathbf{w}$, $\mathbf{p}$, $\mathbf{V}$, $\mathbf{U}_k$, $\mathbf{Z}_k$, $\boldsymbol{\lambda}_m$, $\mathbf{\Lambda}_1$, and $\{\mathbf{\Lambda}_{i,k}\}_{i=2,3,4}$ for $m \in [M]$ and $k \in [K]$, and $\mathcal{I}_i$ is the indicator function specified by

$$\mathcal{I}_0(\boldsymbol{\lambda}_m) = \begin{cases} 0, & \text{if } \|[\boldsymbol{\lambda}_m]_{1:KL}\|_2 \leq [\boldsymbol{\lambda}_m]_{KL+1} \\ \infty & \text{otherwise,} \end{cases} \qquad \mathcal{I}_1(\mathbf{\Lambda}_1) = \begin{cases} 0, & \text{if } \mathbf{\Lambda}_1 \succeq 0 \\ \infty & \text{otherwise,} \end{cases}$$

$$\mathcal{I}_i(\mathbf{\Lambda}_{i,k}) = \begin{cases} 0, & \text{if } \mathbf{\Lambda}_{i,k} \succeq 0 \\ \infty & \text{otherwise,} \end{cases} \quad i = 2, 3, 4.$$

It is clear from problem (7.43) that the introduced indicator functions helps to isolate the second-

order cone and positive semidefinite cone constraints with respect to slack variables.

Problem (7.43) is now in a form suitable for the application of ADMM. The corresponding augmented Lagrangian [60] in ADMM is given by

$$
\begin{aligned}
\mathcal{L}_\rho(\mathscr{X}, \mathscr{Z}, \mathscr{Y}) = \ & \mathrm{tr}(\mathbf{V}) + \tau \sum_{k=1}^{K} \mathrm{tr}(\mathbf{Z}_k) + \sum_{m=1}^{M} \mathcal{I}_0(\boldsymbol{\lambda}_m) + \mathcal{I}_1(\boldsymbol{\Lambda}_1) + \sum_{i=2}^{4} \sum_{k=1}^{K} \mathcal{I}_i(\boldsymbol{\Lambda}_{i,k}) \\
& + \sum_{m=1}^{M} \boldsymbol{\pi}_m^T \mathbf{f}_m(\mathscr{X}, \mathscr{Z}) + \frac{\rho}{2} \sum_{m=1}^{M} \|\mathbf{f}_m(\mathscr{X}, \mathscr{Z})\|_2^2 + \mathrm{tr}\left(\boldsymbol{\Pi}_1^T \mathbf{F}_1(\mathscr{X}, \mathscr{Z})\right) \\
& + \frac{\rho}{2} \|\mathbf{F}_1(\mathscr{X}, \mathscr{Z})\|_F^2 + \sum_{i=2}^{4} \sum_{k=1}^{K} \mathrm{tr}\left(\boldsymbol{\Pi}_{i,k}^T \mathbf{F}_{i,k}(\mathscr{X}, \mathscr{Z})\right) \\
& + \frac{\rho}{2} \sum_{i=2}^{4} \sum_{k=1}^{K} \|\mathbf{F}_{i,k}(\mathscr{X}, \mathscr{Z})\|_F^2,
\end{aligned}
\tag{7.44}
$$

where $\mathscr{X}$ denotes the set of primal variables $\mathbf{w}$, $\mathbf{p}$, $\mathbf{V}$, $\mathbf{U}_k$ and $\mathbf{Z}_k$ for $k \in [K]$, $\mathscr{Z}$ denotes the set of primal slack variables $\boldsymbol{\lambda}_m$, $\boldsymbol{\Lambda}_1$ and $\{\boldsymbol{\Lambda}_{i,k}\}_{i=2,3,4}$ for $m \in [M]$ and $k \in [K]$, $\mathscr{Y}$ is the set of dual variables (also known as Lagrangian multipliers) $\boldsymbol{\pi}_m$, $\boldsymbol{\Pi}_1$ and $\{\boldsymbol{\Pi}_{i,k}\}_{i=2,3,4}$ for $m \in [M]$ and $k \in [K]$, $\mathbf{f}_m(\cdot)$, $\mathbf{F}_1(\cdot)$, and $\mathbf{F}_{i,k}(\cdot)$ for $i \in \{2, 3, 4\}$ represent linear functions at the left hand side of equality constraints in (7.38) − (7.42), $\rho > 0$ is a regularization parameter, and $\| \cdot \|_F$ denotes the Frobenius norm of a matrix.

We iteratively execute the following three steps for ADMM iteration $t = 0, 1, \ldots$

$$
\mathscr{X}^{t+1} = \arg\min_{\mathscr{X}} \mathcal{L}(\mathscr{X}, \mathscr{Z}^t, \mathscr{Y}^t)
\tag{7.45}
$$

$$
\mathscr{Z}^{t+1} = \arg\min_{\mathscr{Z}} \mathcal{L}(\mathscr{X}^{t+1}, \mathscr{Z}, \mathscr{Y}^t)
\tag{7.46}
$$

$$
\begin{cases}
\boldsymbol{\pi}_m^{t+1} = \boldsymbol{\pi}_m^t + \rho \, \mathbf{f}_m(\mathscr{X}^{t+1}, \mathscr{Z}^{t+1}), \ \forall m \\
\boldsymbol{\Pi}_1^{t+1} = \boldsymbol{\Pi}_1^t + \rho \, \mathbf{F}_1(\mathscr{X}^{t+1}, \mathscr{Z}^{t+1}) \\
\boldsymbol{\Pi}_{i,k}^{t+1} = \boldsymbol{\Pi}_{i,k}^t + \rho \, \mathbf{F}_{i,k}(\mathscr{X}^{t+1}, \mathscr{Z}^{t+1}), \ \forall i, k,
\end{cases}
\tag{7.47}
$$

until both of the conditions $\|\mathscr{X}^{t+1} - \mathscr{Z}^t\|_F \leq \epsilon_{\mathrm{admm}}$ and $\|\mathscr{Z}^{t+1} - \mathscr{Z}^t\|_F \leq \epsilon_{\mathrm{admm}}$ are satisfied, where with an abuse of notation, $\|\mathscr{X}\|_F$ denotes the sum of Frobenius norms of variables in $\mathscr{X}$,

and ϵ_{admm} is a stopping tolerance.

The rationale behind using ADMM is that we can split the original problem into the $\mathscr{X}$-minimization step (7.45) and $\mathscr{Z}$-minimization step (7.46), of which the former can be treated as an unconstrained quadratic program and the latter renders an analytical solution.

$\mathscr{X}$-minimization step

Completing the squares with respect to primal variables in the augmented Lagrangian (7.44), problem (7.45) can be cast as

$$
\begin{aligned}
\text{minimize} \quad & \varphi(\mathbf{w}, \mathbf{p}, \mathbf{V}, \{\mathbf{U}_k\}, \{\mathbf{Z}_k\}) := \text{tr}(\mathbf{V}) + \tau \sum_{k=1}^{K} \text{tr}(\mathbf{Z}_k) + \frac{\rho}{2} \sum_{m=1}^{M} \left\| \bar{\mathbf{Q}}_m \mathbf{w} - \boldsymbol{\alpha}_m \right\|_2^2 \\
& + \frac{\rho}{2} \left\| \begin{bmatrix} \mathbf{C} - \text{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} - \boldsymbol{\Upsilon}_1 \right\|_F^2 + \frac{\rho}{2} \sum_{k=1}^{K} \left\| \begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} - \boldsymbol{\Upsilon}_{3,k} \right\|_F^2 \\
& + \frac{\rho}{2} \sum_{k=1}^{K} \left\| \begin{bmatrix} p_k & \sigma_\epsilon^{-1} \mathbf{h}_k^T \\ \sigma_\epsilon^{-1} \mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\varsigma^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} - \boldsymbol{\Upsilon}_{2,k} \right\|_F^2 \\
& + \frac{\rho}{2} \sum_{k=1}^{K} \left\| \mathbf{Z}_k - \mathbf{U}_k + \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T - \boldsymbol{\Upsilon}_{4,k} \right\|_F^2 \quad (7.48)
\end{aligned}
$$

where $\boldsymbol{\alpha}_m := \boldsymbol{\lambda}_m - \mathbf{c}_m - (1/\rho)\boldsymbol{\pi}_m$ for $m \in [M]$, $\boldsymbol{\Upsilon}_1 := \boldsymbol{\Lambda}_1 - (1/\rho)\boldsymbol{\Pi}_1$, and $\boldsymbol{\Upsilon}_{i,k} := \boldsymbol{\Lambda}_{i,k} - (1/\rho)\boldsymbol{\Pi}_{i,k}$ for $i \in \{2, 3, 4\}$ and $k \in [K]$, and for simplicity, we have omitted the ADMM iteration index t in the primal slack variables and dual variables.

We note that problem (7.48) is an unconstrained quadratic program (UQP). In what follows, we employ a gradient descent method [59] together with a backtracking line search [59, Chapter 9.2] to solve this UQP. The first order method yields low computational complexity and memory requirement at each iteration. In Proposition 7.2, we show the gradient of the objective function of problem (7.48).

Proposition 7.2. *The gradient of the objective function of problem* (7.48) *is given by*

$$\nabla_{\mathbf{w}}\varphi = \rho \sum_{m=1}^{M} \bar{\mathbf{Q}}_m^T(\bar{\mathbf{Q}}_m\mathbf{w} - \boldsymbol{\alpha}_m) + 2\rho(\mathbf{w} - \boldsymbol{\gamma}_3) + 2\rho\,\mathrm{blkdiag}\{\hat{\mathbf{w}}_k\mathbf{w}_k^T + \mathbf{w}_k\hat{\mathbf{w}}_k^T - \mathbf{H}_k\}_{k=1}^{K}\hat{\mathbf{w}}$$

$$\nabla_{\mathbf{p}}\varphi = \rho\,\mathrm{diag}\,(\mathbf{C} - \mathrm{diag}(\mathbf{p}) - \boldsymbol{\Upsilon}_1^{11}) + \rho(\mathbf{p} - \boldsymbol{\gamma}_2)$$

$$\nabla_{\mathbf{V}}\varphi = \mathbf{I} + \rho(\mathbf{V} - \boldsymbol{\Upsilon}_1^{22})$$

$$\nabla_{\mathbf{U}_k}\varphi = \rho\sigma_\epsilon^2\sigma_\varsigma^{-2}\mathbf{G}_k(\mathbf{I} + \sigma_\epsilon^2\sigma_\varsigma^{-2}\mathbf{G}_k^T\mathbf{U}_k\mathbf{G}_k - \boldsymbol{\Upsilon}_{2,k}^{22})\mathbf{G}_k^T + \rho(2\mathbf{U}_k - \boldsymbol{\Upsilon}_{3,k}^{11} - \mathbf{Z}_k - \mathbf{T}_k),\ k \in [K]$$

$$\nabla_{\mathbf{Z}_k}\varphi = \tau\mathbf{I} + \rho(\mathbf{Z}_k - \mathbf{U}_k + \mathbf{T}_k),\ k \in [K],$$

where $\boldsymbol{\gamma}_3 = [\boldsymbol{\gamma}_{3,1}^T, \ldots, \boldsymbol{\gamma}_{3,K}^T]^T$, $\boldsymbol{\gamma}_{3,k}$ *is the* $(L+1)$ *column of* $\boldsymbol{\Upsilon}_{3,k}$ *after the last entry is removed,* $\mathbf{H}_k := \mathbf{U}_k - \mathbf{Z}_k + \hat{\mathbf{w}}_k\hat{\mathbf{w}}_k^T + \boldsymbol{\Upsilon}_{4,k}$, $\hat{\mathbf{w}} = [\hat{\mathbf{w}}_1^T, \ldots, \hat{\mathbf{w}}_K^T]^T$, $\boldsymbol{\Upsilon}_1^{11}$ *is a submatrix of* $\boldsymbol{\Upsilon}_1$ *that contains its first* K *rows and columns,* $\boldsymbol{\gamma}_2 = [\gamma_{2,1}, \ldots, \gamma_{2,K}]^T$, $\gamma_{2,k}$ *is the first element of* $\boldsymbol{\Upsilon}_{2,k}$, $\mathrm{diag}(\cdot)$ *returns the diagonal entries of its matrix argument in vector form,* $\boldsymbol{\Upsilon}_1^{22}$ *is a submatrix of* $\boldsymbol{\Upsilon}_1$ *after the first* K *rows and columns are removed,* $\boldsymbol{\Upsilon}_{2,k}^{22}$ *is a submatrix of* $\boldsymbol{\Upsilon}_{2,k}$ *after the first row and column are removed,* $\boldsymbol{\Upsilon}_{3,k}^{11}$ *is a submatrix of* $\boldsymbol{\Upsilon}_{3,k}$ *after the last row and column are removed, and* $\mathbf{T}_k := \hat{\mathbf{w}}_k\mathbf{w}_k^T + \mathbf{w}_k\hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k\hat{\mathbf{w}}_k^T - \boldsymbol{\Upsilon}_{4,k}$.

Proof: See Appendix A.18. ∎

In Proposition 7.2, we note that the optimal value of $\mathbf{V}$ is achieved by letting $\nabla_{\mathbf{V}}\varphi = \mathbf{0}$. This leads to

$$\mathbf{V} = \boldsymbol{\Upsilon}_1^{22} - (1/\rho)\mathbf{I}. \tag{7.49}$$

To solve problem (7.48) for other variables, we employ the gradient descent method summarized in Algorithm 7.2. This algorithm calls on the backtracking line search (Algorithm 7.3) to properly determine the step size such that the convergence to a stationary point of problem (7.48) is accelerated.

Algorithm 7.2 Gradient descent method for solving UQP (7.48)

Require: values of $\mathbf{w}$, $\mathbf{p}$, $\{\mathbf{U}_k\}$ and $\{\mathbf{Z}_k\}$ at the previous ADMM iteration, $\epsilon_{\text{grad}} > 0$, and $\mathbf{V}$ given by (7.49)
1: **repeat**
2:    compute the gradient of ϕ following Proposition 7.2
3:    compute $c_{\text{grad}} := \sum_{k=1}^{K} \|\nabla_{\mathbf{U}_k}\varphi\|_F^2 + \|\nabla_{\mathbf{w}}\varphi\|_2^2 \quad + \|\nabla_{\mathbf{p}}\varphi\|_2^2 + \sum_{k=1}^{K} \|\nabla_{\mathbf{Z}_k}\varphi\|_F^2$
4:    call Algorithm 7.3 to determine a step size κ
5:    update variables

$$\mathbf{w} := \mathbf{w} + \kappa\nabla_{\mathbf{w}}\varphi, \ \mathbf{p} := \mathbf{p} + \kappa\nabla_{\mathbf{p}}\varphi$$
$$\mathbf{U}_k := \mathbf{U}_k + \kappa\nabla_{\mathbf{U}_k}\varphi, \ \mathbf{Z}_k := \mathbf{Z}_k + \kappa\nabla_{\mathbf{Z}_k}\varphi$$

6: **until** $c_{\text{grad}} \leq \epsilon_{\text{grad}}$.

Algorithm 7.3 Backtracking line search for choosing κ

1: Given $\kappa := 1$, $a_1 \in (0, 0.5)$, $a_2 \in (0, 1)$, and c_{grad}
2: **repeat**
3:    $\kappa := a_2\kappa$,
4:    let $\hat{\varphi}$ be the value of φ at the points $\mathbf{w} + \kappa\nabla_{\mathbf{w}}\varphi, \quad \mathbf{p} + \kappa\nabla_{\mathbf{p}}\varphi, \mathbf{V}, \mathbf{U}_k + \kappa\nabla_{\mathbf{U}_k}\varphi$, and $\mathbf{Z}_k + \kappa\nabla_{\mathbf{Z}_k}\varphi$
5: **until** $\hat{\varphi} < \varphi(\mathbf{w}, \mathbf{p}, \mathbf{V}, \{\mathbf{U}_k\}, \{\mathbf{Z}_k\}) - a_1\kappa\, c_{\text{grad}}$.

$\mathscr{Z}$-minimization step

From the expressions of the augmented Lagrangian and indicator functions, problem (7.46) can be decomposed into a sequence of subproblems with respect to slack variables in $\mathscr{Z}$, respectively.

Completing the squares with respect to $\boldsymbol{\lambda}_m$ in (7.44), problem (7.46) together with (7.38) yields the subproblem

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{\lambda}_m - \boldsymbol{\beta}_m\|_2^2 \\ \text{subject to} \quad & \|[\boldsymbol{\lambda}_m]_{1:KL}\|_2 \leq [\boldsymbol{\lambda}_m]_{KL+1}, \end{aligned} \tag{7.50}$$

where $\boldsymbol{\beta}_m := \bar{\mathbf{Q}}_m\mathbf{w} + \mathbf{c}_m + (1/\rho)\boldsymbol{\pi}_m$, and for notational simplicity, we have omitted the ADMM iteration index t. The solution of problem (7.50) is achieved by projecting $\boldsymbol{\beta}_m$ onto a second-order

cone [93, Sec. 6.3],

$$
\boldsymbol{\lambda}_m = \begin{cases} \mathbf{0} & \|[\boldsymbol{\beta}_m]_{1:KL}\|_2 \leq -[\boldsymbol{\beta}_m]_{KL+1} \\ \boldsymbol{\beta}_m & \|[\boldsymbol{\beta}_m]_{1:KL}\|_2 \leq [\boldsymbol{\beta}_m]_{KL+1} \\ \tilde{\boldsymbol{\beta}}_m & \|[\boldsymbol{\beta}_m]_{1:KL}\|_2 \geq |[\boldsymbol{\beta}_m]_{KL+1}|, \end{cases}
\tag{7.51}
$$

for $m \in [M]$, where

$$
\tilde{\boldsymbol{\beta}}_m := \frac{1}{2} \left(1 + \frac{[\boldsymbol{\beta}_m]_{KL+1}}{\|[\boldsymbol{\beta}_m]_{1:KL}\|_2} \right) \left[[\boldsymbol{\beta}_m]_{1:KL}^T, \|[\boldsymbol{\beta}_m]_{1:KL}\|_2 \right]^T .
$$

Completing the squares with respect to $\boldsymbol{\Lambda}_1$ in (7.44), problem (7.46) together with (7.39) yields the subproblem

$$
\begin{array}{ll}
\text{minimize} & \|\boldsymbol{\Lambda}_1 - \boldsymbol{\Phi}_1\|_F^2 \\
\text{subject to} & \boldsymbol{\Lambda}_1 \succeq 0,
\end{array}
\qquad
\boldsymbol{\Phi}_1 := \begin{bmatrix} \mathbf{C} - \mathrm{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} + (1/\rho)\boldsymbol{\Pi}_1.
\tag{7.52}
$$

The solution of problem (7.52) is given by [93, Sec. 6.3]

$$
\boldsymbol{\Lambda}_1 = \sum_{i=1}^{2K} (\sigma_i)_+ \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T ,
\tag{7.53}
$$

where $\sum_{i=1}^{2K} \sigma_i \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T$ is the eigenvalue decomposition of $\boldsymbol{\Phi}_1$, and $(\cdot)_+$ is the positive part operator.

Completing the squares with respect to $\boldsymbol{\Lambda}_{i,k}$ for $i \in \{2, 3, 4\}$ and $k \in [K]$ in (7.44), problem (7.46) yields a sequence of subproblems similar to (7.52). The resulting solution is the same as (7.53) except that $\boldsymbol{\Phi}_1$ is replaced with one of the following matrices

$$
\boldsymbol{\Phi}_{2,k} := \begin{bmatrix} p_k & \sigma_\epsilon^{-1} \mathbf{h}_k^T \\ \sigma_\epsilon^{-1} \mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\varsigma^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} + \tfrac{1}{\rho} \boldsymbol{\Pi}_{2,k}, \ \boldsymbol{\Phi}_{3,k} := \begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} + \tfrac{1}{\rho} \boldsymbol{\Pi}_{3,k}
$$

$$
\boldsymbol{\Phi}_{4,k} := \mathbf{Z}_k - \mathbf{U}_k + \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T + \tfrac{1}{\rho} \boldsymbol{\Pi}_{4,k}.
$$

Summary of the proposed ADMM algorithm

We initialize the ADMM algorithm by setting $\mathbf{w}^0 = \mathbf{1}$, $\mathbf{p}^0 = \mathbf{1}$, $\mathbf{V}^0 = \mathbf{I}$, $\mathbf{U}_k^0 = \mathbf{Z}_k^0 = \mathbf{I}$ for $k \in [K]$, $\boldsymbol{\lambda}_m^0 = \boldsymbol{\pi}_m^0 = \mathbf{0}$ for $m \in [M]$, $\boldsymbol{\Lambda}_1^0 = \boldsymbol{\Pi}_1^0 = \mathbf{0}$, and $\boldsymbol{\Lambda}_{i,k}^0 = \boldsymbol{\Pi}_{i,k}^0 = \mathbf{0}$ for $i \in \{2, 3, 4,\}$ and $k \in [K]$. The ADMM algorithm proceeds as $(7.45) - (7.47)$. The convergence of ADMM is guaranteed in solving convex problems [60], and it typically takes a few tens of iterations to converge with satisfactory accuracy.

At each iteration of ADMM, the computational complexity of the $\mathscr{X}$-minimization step is approximated by $O(L^4)$, where $O(L)$ roughly counts for the number of iterations of the gradient descent method, and $O(L^3)$ is the complexity of matrix multiplication while computing the gradient. Here we assume that L is much larger than K and N. In $\mathscr{Z}$-minimization step, the computational complexity is dominated by the eigenvalue decomposition used in (7.53). This leads to the complexity $O(L^{3.5})$. As a result, the total computation cost of the ADMM algorithm is given by $O(L^4)$. For additional perspective, we compare the computational complexity of the ADMM algorithm with the interior-point algorithm that takes complexity $O(L^6)$. Clearly, the complexity of ADMM decreases significantly in terms of the number of collaboration links by a factor L^2.

7.6   Numerical results

This section empirically shows the effectiveness of our approach for sensor collaboration in time-varying sensor networks. We assume that θ_k follows a Ornstein-Uhlenbeck process [45] with correlation $\text{cov}(\theta_{k_1}, \theta_{k_2}) = \sigma_\theta^2 e^{-|k_1 - k_2|/\rho_{\text{corr}}}$ for $k_1 \in [K]$ and $k_2 \in [K]$, where ρ_{corr} is a parameter that governs the correlation strength, namely, a larger (or smaller) ρ_{corr} corresponds to a weaker

(or stronger) correlation. The covariance matrix of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\Sigma_\theta} = \sigma_\theta^2 \begin{bmatrix} 1 & e^{-\rho_{\text{corr}}} & \cdots & e^{-(K-1)\rho_{\text{corr}}} \\ e^{-\rho_{\text{corr}}} & 1 & \cdots & e^{-(K-2)\rho_{\text{corr}}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-(K-1)\rho_{\text{corr}}} & e^{-(K-2)\rho_{\text{corr}}} & \cdots & 1 \end{bmatrix}.$$

where unless specified otherwise, we set $\sigma_\theta^2 = 1$ and $\rho_{\text{corr}} = 0.5$. The spatial placement and neighborhood structure of the sensor network is modeled by a random geometric graph [48], $\text{RGG}(N, d)$, where $N = 10$ sensors are randomly deployed over a unit square and bidirectional communication links are possible only for pairwise distances at most d. Clearly, the adjacency matrix $\mathbf{A}$ is determined by $\text{RGG}(N, d)$, and the number of collaboration links increases as d increases. In our numerical examples unless specified otherwise, we set $d = 0.3$ which leads to $\text{RGG}(10, 0.3)$ shown in Fig. 7.2.
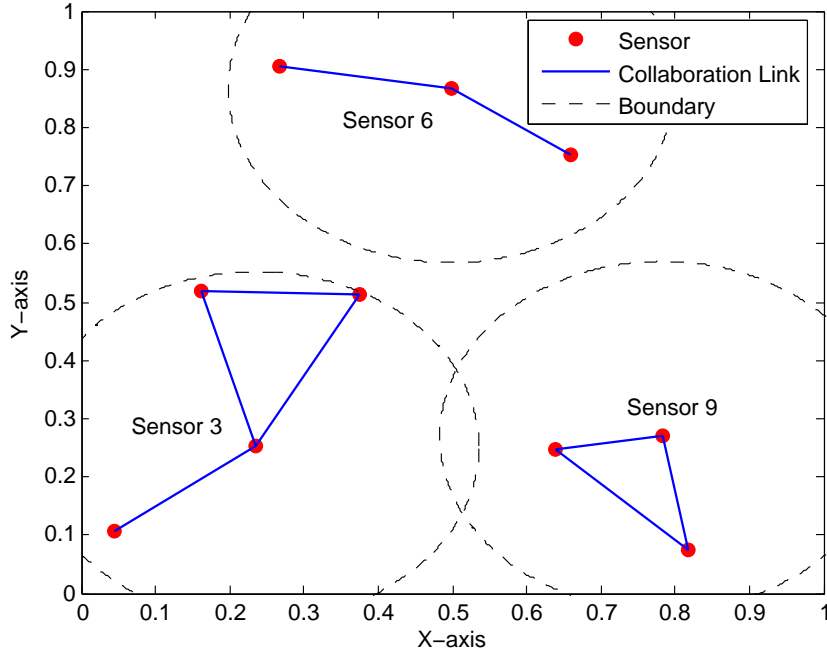
Fig. 7.2: $\text{RGG}(10, 0.3)$, collaboration is depicted for sensors $3, 6$ and 9.

In the collaborative estimation system, we assume that $M = 10$, $K = 3$, $\sigma_\epsilon^2 = \sigma_\varsigma^2 = 1$, and $E_m = E_{\text{total}}/M$ for $m \in [M]$, where $E_{\text{total}} = 1$ gives the total energy budget of M sensors. For

simplicity, the obverstion gain $\mathbf{h}_k$ and channel gain $\mathbf{g}_k$ are randomly chosen from the uniform distribution $\mathcal{U}(0.1, 1)$. Moreover, we select $\tau^0 = 0.1$, $\mu = 1.5$, $\tau_{\max} = 100$ in penalty CCP, $a_1 = 0.02$ and $a_2 = 0.5$ in backtracking line search and $\epsilon_{\text{ccp}} = \epsilon_{\text{admm}} = \epsilon_{\text{grad}} = 10^{-3}$ for the stopping tolerance of the proposed algorithms. Unless specified otherwise, the ADMM algorithm is adopted at Step 2 of penalty CCP, and we use CVX [94] for all other computations. The estimation performance is measured through the empirical mean squared error (MSE), which is computed over 1000 numerical trials.
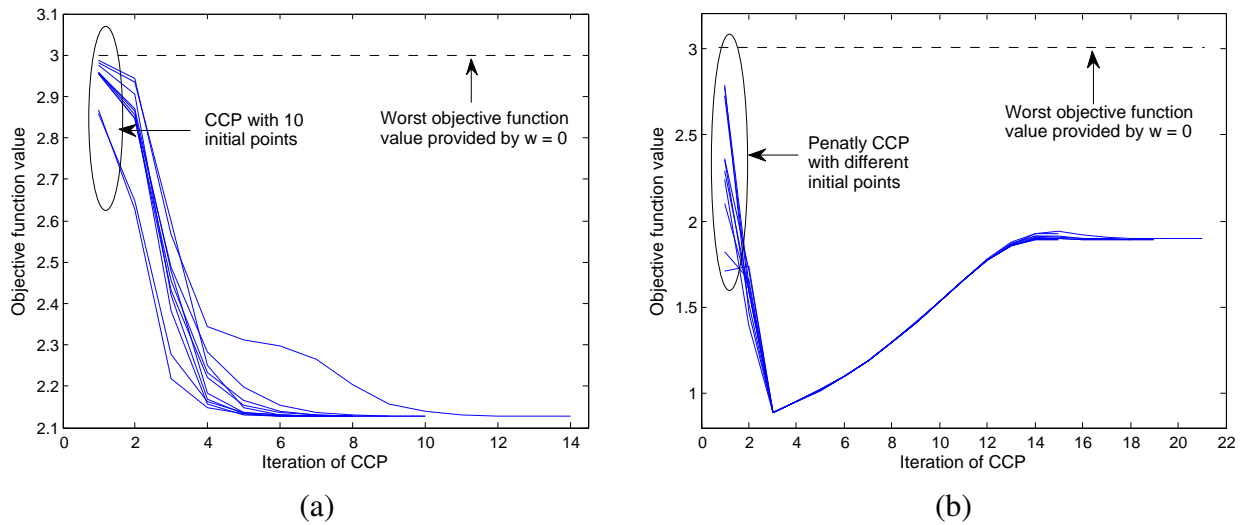
Fig. 7.3: Convergence of CCP and penalty CCP for different initial points.

In Fig. 7.3, we present convergence trajectories of CCP and penalty CCP as functions of interation index for 10 different initial points. For comparison, we plot the worst objective function value of collaboration problem (7.7) when $\mathbf{w} = \mathbf{0}$, namely, LMMSE is determined only by the prior information, which leads to the worst estimation error $\text{tr}(\mathbf{\Sigma}_\theta) = K = 3$. As we can see, much of the benefit of using CCP or penalty CCP is gained during the first few iterations. And each algorithm converges to almost the same objective function value for different initial points. Compared to CCP, the convergence trajectory of penalty CCP is not monotonically decreasing. Namely, penalty CCP is not a descent algorithm. The non-monotonicity of penalty CCP is caused by the penalization on the violation of constraints in the objective function. The objective function value of penalty CCP converges until the penalization ceases to change significantly (after 15

iterations in this example).

In Fig. 7.4, we present the trace of error covariance matrix $\mathbf{P}_W$ given by (7.6) as a function of the correlation parameter ρ_{corr}, where the sensor collaboration scheme is obtained from CCP and penalty CCP to solve (7.14) and (7.19), respectively. We observe that the estimation error resulting from the solution of (7.14) remains unchanged for different values of ρ_{corr} since the formulation of (7.14) is independent of the prior knowledge about parameter correlation. The estimation error resulting from the solution of (7.19) increases as ρ_{corr} increases, and it eventually converges to the error resulting from the solution of (7.14) at an extremely large ρ_{corr}, where parameters become uncorrelated. This is not surprising, since the prior information about parameter correlation was taken into account in (7.19), thereby significantly improving the estimation performance.
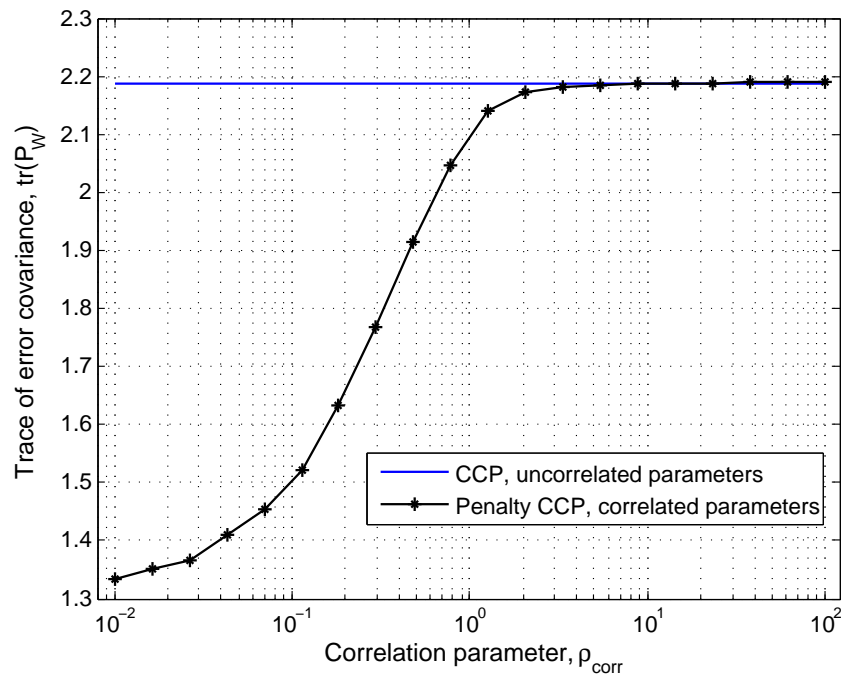
Fig. 7.4: Estimation error versus correlation parameter ρ_{corr}.

In Fig. 7.5, we present the MSE of collaborative estimation as a function of the total energy budget E_{total} for $\rho_{\text{corr}} = 0.5$. For comparison, we plot the estimation performance when using a time-invariant collaboration scheme to solve (7.14) and (7.19), respectively. The assumption of time-invariant collaboration implicitly adds the additional constraint $\mathbf{w}_1 = \ldots = \mathbf{w}_K$, which reduces the problem size. By fixing the type of algorithm, we observe that the MSE when using
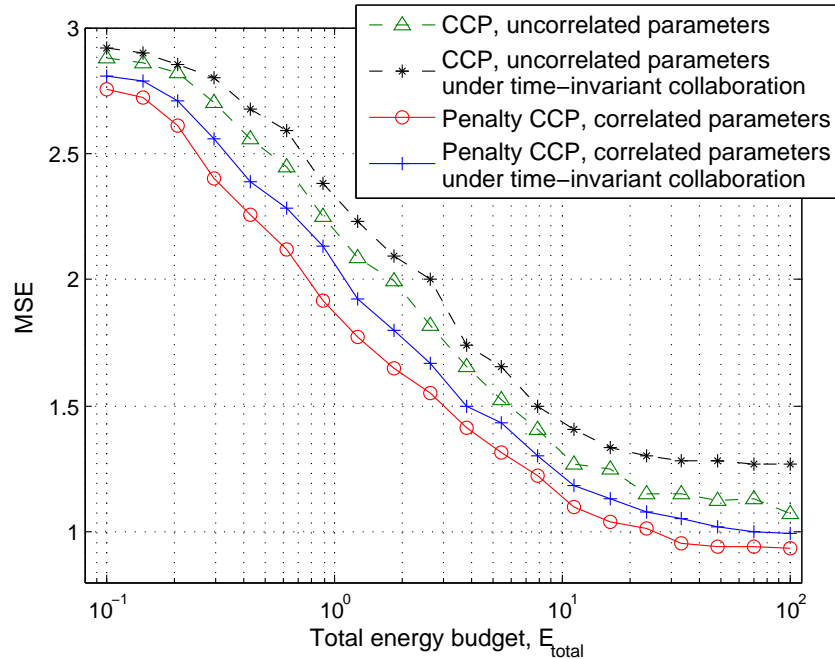
Fig. 7.5: MSE versus total energy budget.

time-invariant sensor collaboration is larger than that of the originally proposed algorithm. This is because the latter accounts for temporal dynamics of the network, where observation and channel gains vary in time. Moreover, the solution of (7.19) yields lower MSE than that of (7.14). This result is consistent with Fig. 3.3 for a fixed correlation parameter. Lastly, the estimation error is smaller as more energy is used in sensor collaboration.

In Fig. 7.6, we present the MSE and the number of collaboration links as functions of the collaboration radius d for $\rho_{\mathrm{corr}} = 0.5$ and $E_{\mathrm{total}} = 1$. We note that the estimation accuracy improves as d increases, since a larger value of d corresponds to more collaboration links in the network. For a fixed value of d, the MSE when solving (7.19) is lower than that when solving (7.14), since the latter ignores the information about parameter correlation. Moreover, we observe that the MSE tends to saturate beyond a collaboration radius $d \approx 0.7$. This indicates that a large part of the performance improvement is achieved only through partial collaboration.

In Fig. 7.7, we present the computation time of our algorithms as functions of problem size specified in terms of the number of collaboration links L. For comparison, we plot the computation time of penalty CCP when using an interior-point solver in CVX [94]. As we can see,
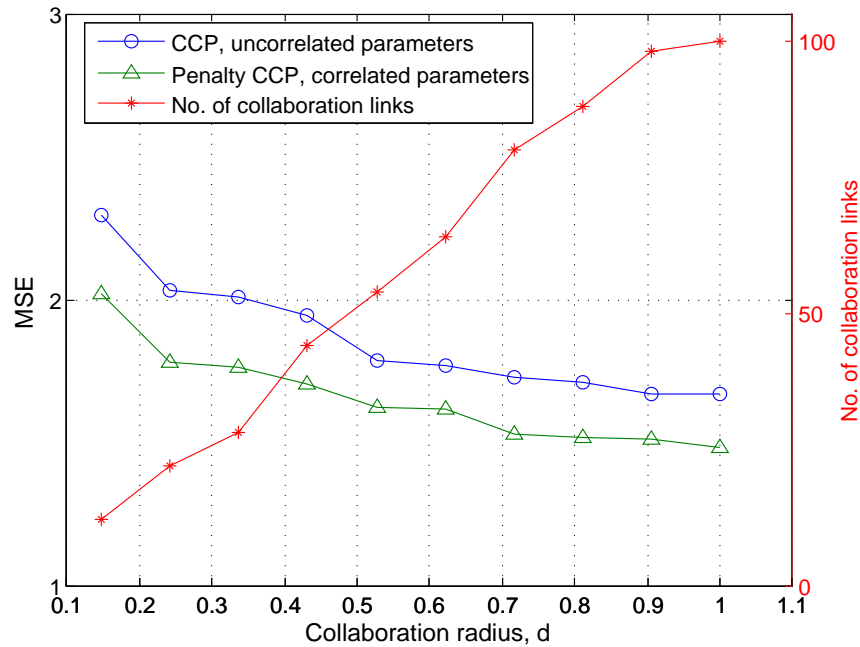
Fig. 7.6: MSE and collaboration links versus collaboration radius d.

penalty CCP requires much higher computation time than CCP, since the former requires solutions of SDPs. When L is small, we observe that the ADMM based penalty CCP has a higher computation time than when using the interior-point solver. This is because the gradient descent method in ADMM takes relatively more iterations (compared to small L) to converge with satisfactory accuracy. However, the ADMM based algorithm performs much faster for a relatively large problem with $L > 60$.

7.7   Summary

In this chapter, we studied the problem of sensor collaboration for estimation of time-varying parameters in dynamic sensor networks. Based on prior knowledge about parameter correlation, the resulting sensor collaboration problem was solved for estimation of temporally uncorrelated and correlated parameters. In the case of temporally uncorrelated parameters, we showed that the sensor collaboration problem can be cast as a special nonconvex optimization problem, where a difference of convex functions carries all the nonconvexity. By exploiting problem structure, we
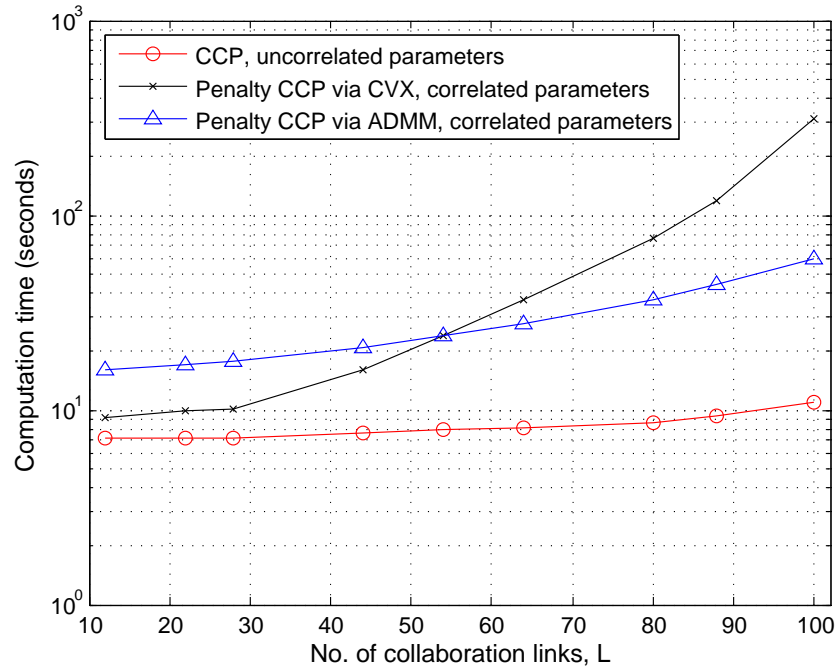
Fig. 7.7: Computation time versus number of collaboration links.

solved the problem by using a convex-concave procedure, which renders a near optimal solution evidenced by numerical results. In the case of correlated parameters, we showed that the sensor collaboration problem can be converted into a semidefinite program together with a nonconvex rank-one constraint. Spurred by the problem structure, we employed a semidefinite programming based penalty convex-concave procedure to solve the sensor collaboration problem. We further proposed an ADMM-based algorithm that is more scalable to large-scale optimization. We empirically showed the effectiveness of our approach for sensor collaboration in dynamic sensor networks. In the next chapter, we will study the problem of collaborative estimation with energy harvesting sensors.

CHAPTER 8

SENSOR COLLABORATION WITH ENERGY HARVESTING SENSORS

8.1 Introduction

In the previous chapters, sensors were assumed to be equipped with conventional power-limited batteries. There have been significant advances in the design of energy-harvesting sensors. Thus, it becomes quite attractive to study collaborative estimation problems in sensor networks with energy harvesting nodes, where at each time step sensors can replenish energy from the environment (such as solar and wind) without the need of battery replacement.

Different from the work [51–53] on distributed estimation with energy harvesting sensors, here we present a unified optimization framework for the joint design of optimal energy allocation and storage control polices while incorporating the cost of sensor collaboration. We show that the resulting optimization problem is nonconvex. However, by identifying the special types of nonconvexities, the methods of convex relaxation and convex-concave procedure can be effectively used to find locally optimal solutions. Extensive numerical results are provided to demonstrate the utility of our approach for energy allocation and storage control in collaborative estimation.

The rest of the chapter is organized as follows. In Section 8.2, we the collaborative estimation

system with energy harvesting sensors. In Section 8.3, we show that the problem of energy alloca-tion and storage control can be addressed as a special nonconvex problem, where the nonconvexity sources can be isolated to constraints that contain the difference of convex functions and the ℓ_0 norm. In Section 8.4, we present the optimization approach for determining the optimal energy allocation and storage control policies. In Section 8.5, we show the effectiveness of our approach via numerical examples. We summarize our work in Section 8.6.

8.2 Problem statement

System model

The task of the sensor network is to estimate a time-varying parameter $\theta_t \in \mathbb{R}$ over a time horizon of length T. We consider a collaborative estimation system where sensors are able to collaborate, namely, share observations with other neighboring sensors. The obtained collaborative messages are transmitted through a coherent MAC to the FC, which produces a global estimate of θ_t for $t = 1, 2, \ldots, T$. In the network, each sensor is equipped with an energy harvesting device and a finite-capacity energy storage device, where the former harvests the renewable energy from the environment, and the latter governs the storage charging/discharging actions across time for sensor collaboration and data transmission. An overview of the collaborative estimation system with energy harvesting and storage is shown in Fig. 8.1.

The vector of measurements from N sensors at time t is

$$\mathbf{x}_t = \mathbf{h}_t \theta_t + \boldsymbol{\epsilon}_t, \quad t \in [T], \tag{8.1}$$

where $[T]$ denotes the integer set $\{1, 2, \ldots, T\}$, $\mathbf{x}_t = [x_{t,1}, \ldots, x_{t,N}]^T$ is the vector of measure-ments, $\mathbf{h}_t = [h_{t,1}, \ldots, h_{t,N}]^T$ is the vector of observation gains, θ_t is the parameter of interest which has zero mean and variance σ_θ^2, $\boldsymbol{\epsilon}_t = [\epsilon_{t,1}, \ldots, \epsilon_{t,N}]^T$ is the noise vector with i.i.d. Gaussian variables $\epsilon_{t,i} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ for $t \in [T]$ and $i \in [N]$.
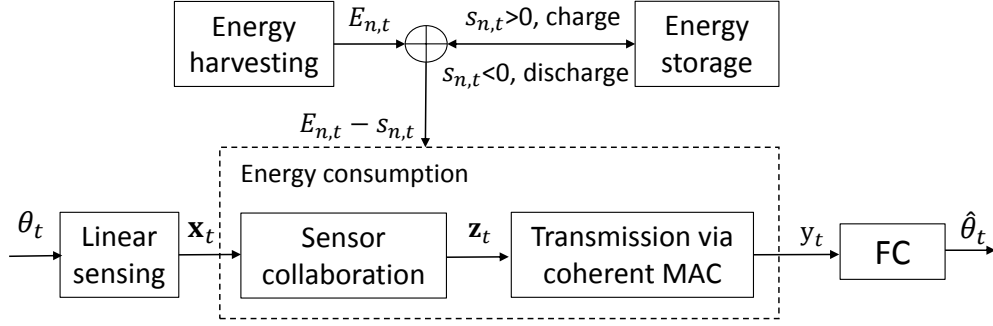
Fig. 8.1: Collaborative estimation with energy harvesting and storage. θ_t and $\hat{\theta}_t$ denote the parameter and its estimate at time t, respectively. $\mathbf{x}_t$, $\mathbf{z}_t$ and y_t denote the sensor measurements, collaborative signals and transmitted signal, respectively. $E_{n,t}$ and $s_{n,t}$ are the harvested energy (known in advance) and stored energy (to be designed) of the nth sensor at time t.

After linear sensing, each sensor may pass its observation to another sensor for collaboration prior to transmission to the FC. With a relabelling of the sensors, we assume that the first M sensors (out of a total of N sensor nodes) communicate with the FC. The process of sensor collaboration at time t is

$$\mathbf{z}_t = \mathbf{W}_t \mathbf{x}_t, \quad t \in [T] \tag{8.2}$$

where $\mathbf{z}_t = [z_{t,1}, z_{t,2}, \ldots, z_{t,M}]^T$ is the vector of collaborative signals, and $\mathbf{W}_t \in \mathbb{R}^{M \times N}$ is the collaboration matrix that contains collaboration weights (based on the energy allocated) used to combine sensor measurements.

After sensor collaboration, the message $\mathbf{z}_t$ is transmitted through a coherent MAC so that the received signal y_t at the FC is a coherent sum [30]

$$y_t = \mathbf{g}_t^T \mathbf{z}_t + \varsigma_t, \ t \in [T], \tag{8.3}$$

where $\mathbf{g}_t = [g_{t,1,}, g_{t,2}, \ldots, g_{t,M}]^T$ is the vector of channel gains, and ς_t is temporally white Gaussian noise with zero mean and variance σ_ς^2.

Best linear unbiased estimator (BLUE)

We assume that the FC knows the observation gains, channel gains, and variances of the parameters of interest and additive noises. We employ BLUE [58, Theorem 6.1] to estimate the parameters θ_t for $t \in [T]$. The use of BLUE is spurred by the lack of prior knowledge about the temporal correlation of time-varying parameters $\{\theta_t\}$ [46]. As shown in Appendix A.19, the estimation error resulting from BLUE is given by a sum of rational functions with respect to collaboration matrices

$$f(\mathbf{W}_1, \ldots, \mathbf{W}_T) = \sum_{t=1}^{T} \frac{\sigma_\varsigma^2 + \sigma_\epsilon^2 \operatorname{tr}\left(\mathbf{W}_t^T \mathbf{g}_t \mathbf{g}_t^T \mathbf{W}_t\right)}{\operatorname{tr}\left(\mathbf{W}_t^T \mathbf{g}_t \mathbf{g}_t^T \mathbf{W}_t \mathbf{h}_t \mathbf{h}_t^T\right)}. \tag{8.4}$$

Collaboration and transmission costs

It is clear from (8.2) that the sparsity structure of $\mathbf{W}_t$ characterizes the collaboration topology at time t. For instance, $[\mathbf{W}_t]_{mn} = 0$ indicates the absence of a collaboration link from the nth sensor to the mth sensor, where $[\mathbf{W}_t]_{mn}$ is the (m, n)th entry of $\mathbf{W}_t$. To account for an active collaboration link, we use the cardinality function

$$\operatorname{card}([\mathbf{W}_t]_{mn}) = \begin{cases} 0 & [\mathbf{W}_t]_{mn} = 0 \\ 1 & [\mathbf{W}_t]_{mn} \neq 0. \end{cases} \tag{8.5}$$

The *collaboration cost* of each sensor is then given by

$$Q_n(\mathbf{W}_t) = \sum_{m=1}^{M} C_{mn} \operatorname{card}([\mathbf{W}_t]_{mn}), \ n \in [N], \ t \in [T], \tag{8.6}$$

where we assume that sharing of an observation is realized through a reliable (noiseless) communication link that consumes a known power C_{mn}. Note that $C_{mm} = 0$ since each node can collaborate with itself at no cost.

The *transmission cost* of the mth sensor at time t refers to the energy consumption of transmit-

ting the collaborative message $z_{t,m}$ to the FC. Namely,

$$T_m(\mathbf{W}_t) = \mathbb{E}_{\theta_t, \epsilon_t}[z_{t,m}^2] = \text{tr}\left[\mathbf{W}_t^T \mathbf{e}_m \mathbf{e}_m^T \mathbf{W}_t (\sigma_\theta^2 \mathbf{h}_t \mathbf{h}_t^T + \sigma_\epsilon^2 \mathbf{I})\right], \tag{8.7}$$

for $m \in [M]$ and $t \in [T]$, where $\mathbf{e}_m \in \mathbb{R}^M$ is a basis vector with 1 at the mth coordinate and 0s elsewhere. Since only the first M sensors are used to communicate with the FC, we define $T_n(\mathbf{W}_t) := 0$ for $n > M$,

Energy harvesting and storage constraints

We assume that knowledge of the harvested energy is available in advance (also known as full side information [109]). Let $E_{n,t}$ denote the harvested energy of the nth sensor at time t, which is used for storage or sensor collaboration and data transmission. We introduce a variable $s_{n,t}$ to represent the charging/discharging operation of the storage device at the nth sensor at time t, with the following sign convention for $s_{n,t}$

$$\begin{cases} s_{n,t} \geq 0 & \text{charging with amount of energy } s_{n,t}, \\ s_{n,t} < 0 & \text{discharging with amount of energy } -s_{n,t}. \end{cases} \tag{8.8}$$

Given the harvested energy $E_{n,t}$ and the operating mode of the storage device, the energy consumption for sensor collaboration and data transmission satisfies the constraint

$$Q_n(\mathbf{W}_t) + T_n(\mathbf{W}_t) \leq E_{n,t} - s_{n,t}, \; n \in [N], \; t \in [T] \tag{8.9}$$

where $Q_n(\mathbf{W}_t)$ is the collaboration cost in (8.6), $T_n(\mathbf{W}_t)$ is the transmission cost in (8.7), and the quantity $E_{n,t} - s_{n,t}$ stands for the amount of energy available at the nth sensor at time t.

The stored energy at each sensor per time satisfies

$$-\sum_{k=0}^{t-1} s_{n,k} \leq s_{n,t} \leq E_{n,t}, \; n \in [N], \; t \in [T], \tag{8.10}$$

where $s_{n,0} \geq 0$ is the known amount of initially stored energy at the nth sensor, $\sum_{k=0}^{t-1} s_{n,k}$ is the amount of stored energy at time $t - 1$ which might be discharged from the storage device at the next time, and $E_{n,t}$ is the amount of harvested energy at time t which can be used to charge the storage device.

There is a capacity limit of the storage device [110], denoted by $\hat{S}$, so that

$$\sum_{k=0}^{t} s_{n,k} \leq \hat{S}, \ n \in [N], \ t \in [T]. \tag{8.11}$$

To increase the network lifetime, it is also desirable to have some remaining stored energy, denoted by $\check{S}$, at the end of time horizon. This gives the following constraint [111]

$$\sum_{k=0}^{T} s_{n,k} \geq \check{S}, \ n \in [N]. \tag{8.12}$$

Sensor collaboration with energy storage management

In order to seek the optimal collaboration schemes $\{\mathbf{W}_t\}$ and storage control policy $\{s_{n,t}\}$, we formulate the optimization problem, in which the estimation error (8.4) is minimized subject to energy constraints $(8.9) - (8.12)$,

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{W}_1, \ldots, \mathbf{W}_T) \\ \text{subject to} \quad & \text{constraints } (8.9) - (8.12), \end{aligned} \tag{8.13}$$

where the optimization variables are $\mathbf{W}_t \in \mathbb{R}^{M \times N}$ and $\mathbf{s}_t = [s_{1,t}, s_{2,t}, \ldots, s_{N,t}]^T$ for $t \in [T]$. As will be evident later, problem (8.13) is not convex due to its objective function and the constraint (8.9) that involves the cardinality function. We will elaborate on the problem structure in the next section.

8.3 Insights into nonconvexity

In this section, we first simplify problem (8.13) by converting quadratic matrix functions into quadratic vector functions. We then show that problem (8.13) is equivalent to an optimization problem with special types of nonconvexities.

It is clear from (8.4) and (8.7) that both the estimation distortion and the transmission cost contain a quadratic matrix function. For simplicity of representation, we transform a quadratic matrix function into a quadratic vector function by concatenating the entries of a matrix into a column vector.

Specifically, a columnwise vector $\mathbf{w}_t$ of the collaboration matrix $\mathbf{W}_t$ can be written as $\mathbf{w}_t = [w_{t,1}, w_{t,2}, \ldots, w_{t,L}]^T$, where $L = MN$, $w_{t,l}$ is the lth entry of $\mathbf{w}_t$, $w_{t,l} = [\mathbf{W}_t]_{m_l n_l}$, $m_l = l - (\lceil \frac{l}{M} \rceil - 1)M$, $n_l = \lceil \frac{l}{M} \rceil$, and $\lceil x \rceil$ is the ceiling function that yields the smallest integer not less than x.

Using the relationship between the Kronecker product and the vectorization of a matrix[1], we can rewrite the matrix quadratic functions in (8.4) and (8.7) as

$$f(\mathbf{w}_1, \ldots, \mathbf{w}_T) = \sum_{t=1}^{T} \frac{\sigma_\varsigma^2 + \sigma_\epsilon^2 \mathbf{w}_t^T \mathbf{\Omega}_{\mathrm{PN},t} \mathbf{w}_t}{\mathbf{w}_t^T \mathbf{\Omega}_{\mathrm{PD},t}}, \text{ and} \tag{8.14}$$

$$T_n(\mathbf{w}_t) = \begin{cases} \mathbf{w}_t^T \mathbf{\Gamma}_{m,t} \mathbf{w}_t & n \in [M], \\ 0 & n = m+1, \ldots, N, \end{cases} \tag{8.15}$$

where with an abuse of notation, we used f and T_n introduced in (8.4) and (8.7) to represent the estimation distortion and tranmission cost with respect to $\{\mathbf{w}_t\}$, $\mathbf{\Omega}_{\mathrm{PN},t} = \mathbf{I}_N \otimes (\mathbf{g}_t \mathbf{g}_t^T)$, $\mathbf{\Omega}_{\mathrm{PD},t} = (\mathbf{h}_t \mathbf{h}_t^T) \otimes (\mathbf{g}_t \mathbf{g}_t^T)$, and $\mathbf{\Gamma}_{m,t} = (\sigma_\theta^2 \mathbf{h}_t \mathbf{h}_t^T + \sigma_\epsilon^2 \mathbf{I}) \otimes (\mathbf{e}_m \mathbf{e}_m^T)$. Clearly, the matrices $\mathbf{\Omega}_{\mathrm{PN},t}$, $\mathbf{\Omega}_{\mathrm{PD},t}$ and $\mathbf{\Gamma}_{m,t}$ are positive semidefinite, and in particular $\mathbf{\Omega}_{\mathrm{PD},t}$ is of rank one.

---

[1]For appropriate matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$, we obtain $\mathrm{tr}(\mathbf{A}^T \mathbf{B} \mathbf{C} \mathbf{D}^T) = \mathrm{vec}(\mathbf{A})^T (\mathbf{D} \otimes \mathbf{B}) \mathrm{vec}(\mathbf{C})$, where $\otimes$ denotes the Kronecker product, and $\mathrm{vec}(\mathbf{Z})$ is the columnwise vector of $\mathbf{Z}$.

Moreover, we express the collaboration cost (8.6) as

$$Q_n(\mathbf{w}_t) = \sum_{m=1}^{M} c_{m+(n-1)M} \, \mathrm{card}(w_{t,m+(n-1)M}) \tag{8.16}$$

for $n \in [N]$ and $t \in [T]$, where c_i denotes the ith entry of the columnwise vector of the collaboration cost matrix $\mathbf{C} \in \mathbb{R}^{M \times N}$, whose (m,n)th entry is given by C_{mn} in (8.6).

Based on (8.15) and (8.16), the energy consumption constraint (8.9) can be cast as a quadratic inequality that involves the cardinality function with respect to $\mathbf{w}_t$,

$$Q_n(\mathbf{w}_t) + T_n(\mathbf{w}_t) \leq E_{n,t} - s_{n,t}, \ n \in [N], \ t \in [T]. \tag{8.17}$$

Problem (8.13) is further written as

$$\begin{aligned}
\text{minimize} \quad & f(\mathbf{w}_1, \ldots, \mathbf{w}_T) \\
\text{subject to} \quad & \text{storage constraints } (8.10)-(8.12) \\
& \text{energy consumption constraint } (8.17),
\end{aligned} \tag{8.18}$$

where $\mathbf{w}_t \in \mathbb{R}^L$ and $\mathbf{s}_t \in \mathbb{R}^N$ are optimization variables for $t \in [T]$, and we recall that $L = MN$.

Proposition 8.1. *Problem* (8.18) *is equivalent to*

$$\textit{minimize} \quad \mathbf{1}^T \mathbf{u} \tag{8.19a}$$

$$\textit{subject to} \quad \textit{storage constraints } (8.10)-(8.12) \tag{8.19b}$$

$$\textit{energy consumption constraint } (8.17) \tag{8.19c}$$

$$\sigma_\varsigma^2 + \sigma_\epsilon^2 \mathbf{w}_t^T \mathbf{\Omega}_{\mathrm{PN},t} \mathbf{w}_t \leq p_t, \ q_t > 0, t \in [T] \tag{8.19d}$$

$$2p_t + (q_t^2 + u_t^2) - (q_t + u_t)^2 \leq 0, \ t \in [T] \tag{8.19e}$$

$$q_t - \mathbf{w}_t^T \mathbf{\Omega}_{\mathrm{PD},t} \mathbf{w}_t \leq 0, \ t \in [T], \tag{8.19f}$$

where $\mathbf{u} \in \mathbb{R}^T$, $\mathbf{p} \in \mathbb{R}^T$, $\mathbf{q} \in \mathbb{R}^T$, $\mathbf{w}_t \in \mathbb{R}^L$ *and* $\mathbf{s}_t \in \mathbb{R}^N$ *are optimization variables for* $t \in [T]$,

and $\mathbf{1}$ *is the vector of all ones.*

Proof: See Appendix A.20. ∎

It is clear from (8.19) that there exist two sources of nonconvexity: a) the cardinality function in the collaboration cost (8.16), b) the difference of convex (DC) quadratic functions (8.19e) and (8.19f). In what follows, we will show that the difficulties caused by the cardinality function and the nonconvex function of DC type can be efficiently handled via certain convexifications.

8.4   Optimization approach

Problem (8.19) is combinatorial in nature due to the presence of the cardinality function (also known as ℓ_0 norm). In the sparsity-promoting optimization context [112], an ℓ_0 norm is commonly relaxed to an ℓ_1 norm. Based on that, the collaboration cost (8.16) is approximated as

$$\tilde{Q}_n(\mathbf{w}_t) := \|\mathbf{C}_n\mathbf{w}_t\|_1, \ n \in [N], \ t \in [T], \tag{8.20}$$

where $\mathbf{C}_n = [c_{1+(n-1)M}\mathbf{e}_{1+(n-1)M}, \ldots, c_{nM}\mathbf{e}_{nM}]^T \in \mathbb{R}^{M \times L}$, and $\mathbf{e}_i \in \mathbb{R}^L$ is a basis vector. Accordingly, an ℓ_1 proxy for the energy consumption constraint (8.17) is given by

$$\tilde{Q}_n(\mathbf{w}_t) + T_n(\mathbf{w}_t) \le E_{n,t} - s_{n,t}, \ n \in [N], \ t \in [T]. \tag{8.21}$$

We remark that the ℓ_1 norm given by (8.20) can be eliminated by introducing additional optimization variables so that the ℓ_1 constraint (8.21) is cast as a smooth convex constraint together with a linear inequality constraint [59]

$$\mathbf{1}^T\mathbf{r}_{n,t} + T_n(\mathbf{w}_t) \le E_{n,t} - s_{n,t}, \quad -\mathbf{r}_{n,t} \le \mathbf{C}_n\mathbf{w}_t \le \mathbf{r}_{n,t}, \ n \in [N], \ t \in [T], \tag{8.22}$$

where $\mathbf{r}_{n,t} \in \mathbb{R}^M$ is the newly introduced optimization variable for $n \in [N]$ and $t \in [T]$, and the linear inequality is defined in an elementwise fashion.

After ℓ_1 approximation, problem (8.19) becomes convex except for the nonconvex quadratic constraints (8.19e) and (8.19f). However, we will show that such a nonconvex optimization problem can be efficiently solved via a convex-concave procedure.

The nonconvex quadratic constraints (8.19e)–(8.19f) have the DC type. Proceeding with the same logic as used in Section 6.6, we employ CCP to solve the convexified problem

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{1}^T\mathbf{u} \\
\text{subject to} \quad & \text{storage constraints (8.10)–(8.12)} \\
& \text{energy consumption constraints (8.22)} \\
& \text{quadratic constraints (8.19d)} \\
& \text{linearized quadratic constraints (8.19e)–(8.19f):} \\
& 2p_t + (q_t^2 + u_t^2) \le \hat{\psi}_1(q_t, u_t), \text{ and} \\
& q_t \le \hat{\psi}_2(\mathbf{w}_t), \ t \in [T],
\end{aligned}
$$

(8.23)

where $\mathbf{u} \in \mathbb{R}^T$, $\mathbf{p} \in \mathbb{R}^T$, $\mathbf{q} \in \mathbb{R}^T$, $\mathbf{w}_t \in \mathbb{R}^L$, $\mathbf{s}_t \in \mathbb{R}^N$, and $\mathbf{r}_{n,t} \in \mathbb{R}^M$ are optimization variables for $n \in [N]$ and $t \in [T]$, $\psi_1(q_t, u_t) := 2(u_t + q_t)(\hat{u}_t + \hat{q}_t) - (\hat{u}_t + \hat{q}_t)^2$, $(\hat{u}_t, \hat{q}_t)$ is the linearization point used in $\hat{\psi}_1$, $\hat{\psi}_2(\mathbf{w}_t) = 2\mathbf{w}_t^T \mathbf{\Omega}_{\text{PD},t} \hat{\mathbf{w}}_t - \hat{\mathbf{w}}_t^T \mathbf{\Omega}_{\text{PD},t} \hat{\mathbf{w}}_t$, and $\hat{\mathbf{w}}_t$ is the linearization point used in $\hat{\psi}_2$. We summarize CCP for solving problem (8.19) in Algorithm 8.1.

For simplicity, we randomly choose the initial points (drawn from the standard uniform distribution) that are scaled to satisfy the constraints of problem (8.19). It is worth mentioning that reference [103] has proposed a penalized version of CCP which allows the use of infeasible initial points. Our numerical examples show that Algorithm 8.1 works effectively and converges to almost the same optimal value for different initial points; see Fig. 8.2. The convergence of Algorithm 8.1 is also guaranteed, since at each iteration, we solve a linearized convex problem with a smaller feasible set which contains the linearization point (i.e., the solution at the previous iteration) [103].

The computation cost of Algorithm 8.1 is dominated by the solution of a convex program with quadratic constraints at Step 2, which has the computational complexity $O(T^{3.5}L^{3.5})$ via the

---

Algorithm 8.1 CCP for solving problem (8.19)

Require: $\epsilon_{\mathrm{ccp}} > 0$ and feasible initial points $\{\hat{\mathbf{w}}_t\}$, $\hat{\mathbf{q}}$ and $\hat{\mathbf{u}}$

 1: **for** iteration $i = 1, 2, \ldots$ **do**

 2:     set $(\{\mathbf{w}_t^i\}, \mathbf{q}^i, \mathbf{u}^i)$ as the solution of problem (8.23)

 3:     update the linearizing point: $\hat{\mathbf{w}}_t = \mathbf{w}_t^i$, $\hat{\mathbf{q}} = \mathbf{q}^i$, $\hat{\mathbf{u}} = \mathbf{u}^i$

 4:     **until** $|\mathbf{1}^T \mathbf{u}^i - \mathbf{1}^T \mathbf{u}^{i-1}| \le \epsilon_{\mathrm{ccp}}$ for $i \ge 2$

 5: **end for**

---

interior-point algorithm [104, Sec. 10], where recalling that T is the length of time horizon, and $L = NM$ is the maximun number of collaboration links.

8.5   Numerical results
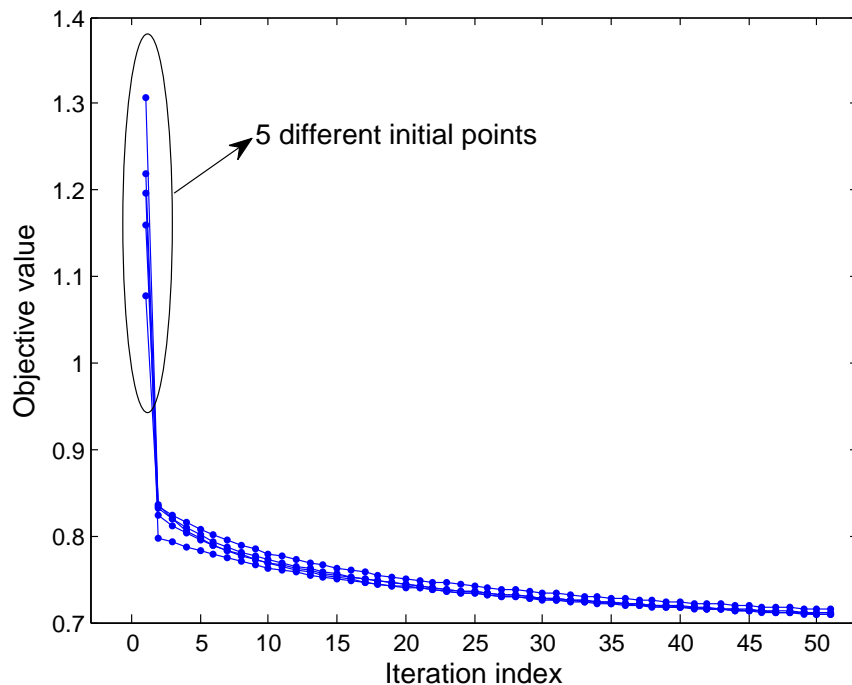
Fig. 8.2: Convergence of CCP for 5 different initial points.

In this section, we empirically show the effectiveness of the proposed energy allocation and storage algorithm. To specify the system model in Fig. 8.1, we assume that $N = 5$ sensors are deployed in a unit square region, and let $T = 10$, $M = 5$, and $\sigma_\theta^2 = \sigma_\epsilon^2 = \sigma_\varsigma^2 = 1$. The values of observation gains $\mathbf{h}_t$ and channel gains $\mathbf{g}_t$ are drawn from the uniform distribution $U(0.1, 1)$. The

amount of energy harvesting $E_{n,t}$ satisfies a Poisson process [52], namely, $\text{prob}\{E_{n,t} = m\mu_0\} = \frac{\lambda^m e^{-\lambda}}{m!}$ for $n \in [N]$ and $t \in [T]$, where $\mu_0 = 1$ is the average energy harvesting rate, and $\lambda = 5$. The initial energy in the storage device is set by $s_{n,0} = 0$ for $n \in [N]$. The capacity limit and the desired residual energy in the storage device are given by $\hat{S} = 100$ and $\check{S} = 0.2 \sum_{t=1}^{T} \sum_{n=1}^{N} \frac{E_{n,t}}{TN}$, respectively. The cost of establishing a collaboration link is modeled by $C_{mn} = \alpha_c \|\mathbf{a}_m - \mathbf{a}_n\|_2$ for $m, n \in [N]$, where $\mathbf{a}_i$ is the location of the ith sensor, and $\alpha_c = 1$ is the collaboration cost parameter. In CCP, we select $\epsilon_{\text{ccp}} = 10^{-3}$.
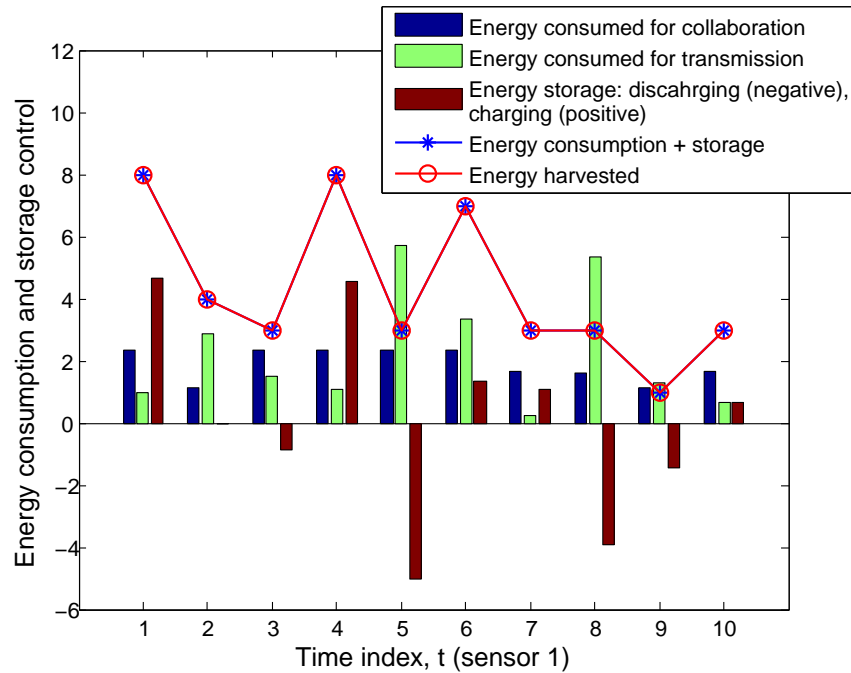
Fig. 8.3: Energy consumption and storage at the first sensor.

In Fig. 8.2, we present the convergence trajectory of CCP as a function of iteration index for 5 different initial points. As we can see, much of the benefit of using CCP is gained from its first few iterations. Moreover, CCP converges to almost the same objective value for different initial points. That is because the linearization of nonconvex quadratic functions, which have rank-one coefficient matrices, e.g, $\Omega_{\text{PD},t}$ in (8.23), leads to good proxies of DC type nonconvex functions.

In Fig. 8.3, we show the obtained energy allocation and storage schemes at the first sensor by solving problem (8.19). We observe that the harvested energy is not wasted, namely, the sum of consumed energy and stored energy is equal to the amount of the harvested energy at each time
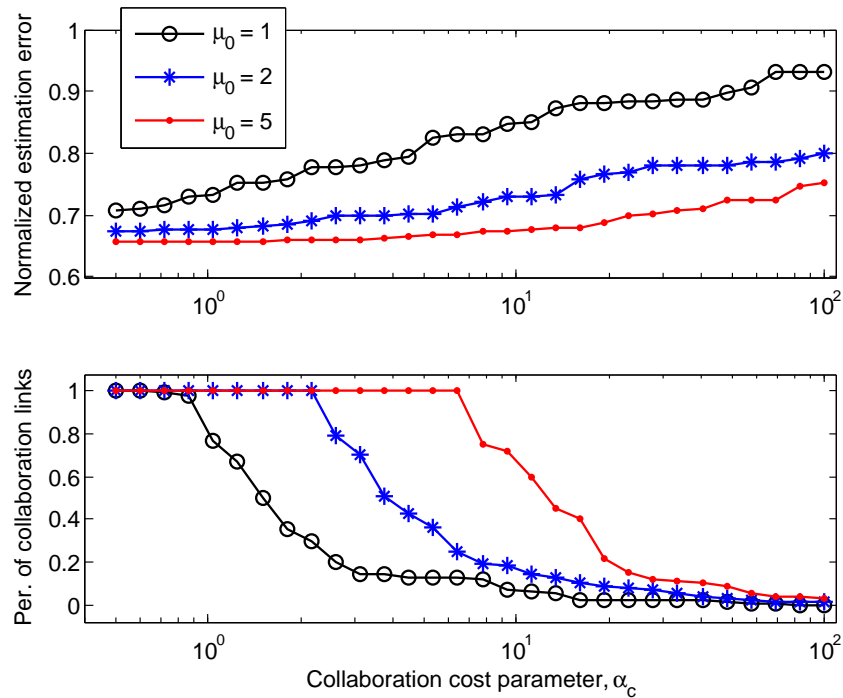
Fig. 8.4: Estimation error and collaboration links while varying α_c.

step. We also note that the storage device operates in a charging mode when there is a large amount of harvested energy (e.g., at time steps 1, 4 and 6). By contrast, the storage device operates in a discharging mode when the harvested energy is low and there is a sufficient amount of stored energy (e.g., at time steps 3, 5 and 8). We finally remark that the energy consumption for sensor collaboration and data transmission varies in time due to the time-varying sensor network.

In Fig. 8.4, we present the estimation error and the percentage of active collaboration links as a function of the collaboration cost parameter α_c for $\mu_0 \in \{1, 3, 5\}$. As we can see, by fixing the energy harvesting rate μ_0, both the estimation accuracy and the number of collaboration links decrease as α_c increases. This is expected, since a larger value of α_c corresponds to a larger cost of sensor collaboration. Moreover, if we fix the value of α_c, both the estimation accuracy and the number of collaboration links increase when the energy harvesting rate increases, since more energy would be harvested and can be used for sensor collaboration.

8.6 Summary

In this chapter, we proposed a tractable optimization framework to jointly design optimal energy allocation and storage control policies for distributed estimation with sensor collaboration. Although the resulting optimization problem is nonconvex in nature, we provided valuable insights into the sources of nonconvexity: a) cardinality function, and b) difference of convex functions. Based on the problem structure, the problem was convexified via an ℓ_1 relaxation and a convex-concave procedure, which renders a near-optimal solution as evidenced by our extensive numerical results.

CHAPTER 9

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

9.1   Concluding remarks

Although pervasive sensors are able to collect massive amounts of data, commonly used sensor networks only have limited capacity to store, transmit and process the accquired data. Therefore, the problem of resource management arises in order to achieve optimal trade-offs between inference accuracy and resource usage. In this thesis, we have addressed two types of resource management problems: a) sensor selection/scheduling in Chapters 3–5, and b) energy allocation for collaborative estimation in Chapters 6–8, where the former refers to finding the best subset of sensors to activate in order to minimize the estimation error subject to a constraint on the number of activations, and the latter refers to seeking the optimal inter-sensor communication strategy prior to the transmission of sensor measurements to a fusion center. We summarize the main contributions of the work reported in this thesis as follows.

Chapter 3 was dedicated to optimal sensor selection with correlated measurement noise. In this chapter, we circumvented drawbacks of the commonly used formulations for sensor selection, and demonstrated their validity only when sensor selection is restricted to the weak noise correlation

regime. We proposed a more general and tractable framework for sensor selection with correlated measurement noise that is valid for any arbitrary noise correlation regime. It was shown that the proposed sensor selection framework is able to yield a more accurate sensor selection scheme than those presented in the previous literature.

Chapter 4 addressed the problem of optimal time-periodic sensor schedules for estimating the state of discrete-time dynamical systems. In this chapter, the sensor scheduling problem was interpreted as a design problem in which both the sensor activation schedules, and the estimator gains used to combine the sensor measurements, are jointly optimized. To allow for additional design flexibility, we introduced sparsity-promoting penalty functions into the optimization formulation to encourage fewer measurements at every time instant of the periodic horizon. This helps to generate arbitrarily sparse sensor schedules that employ a minimal number of active sensors.

Chapter 5 took the individual power constraint of each sensor into account while promoting the sparsity of sensor schedules. In order to achieve a balance between activating the most informative sensors and uniformly allocating sensor energy, we proposed a novel sparsity-promoting penalty function that discourages successive selections of the same group of sensors. We showed the the resulting optimization is convex, and its optimal solution can be efficiently obtained via ADMM or a proximal gradient method.

Chapter 6 was dedicated to the design of optimal sensor collaboration schemes for linear coherent estimation. In this chapter, we described collaboration through a collaboration matrix, in which the nonzero entries characterize the collaboration topology and the values of these entries characterize the collaboration weights. Also, we explored the situation in which there is a cost associated with the involvement of each sensor in the estimation scheme. In such situations, the participating sensors must be chosen judiciously. We proposed a sparsity-inducing optimization framework that jointly designs the optimal sensor selection and collaboration schemes. It was shown that there exists a trade-off between sensor collaboration and sensor selection for a given estimation performance.

Chapter 7 addressed the problem of sensor collaboration in the the presence of (a) temporal

dynamics in system, (b) temporal correlation of parameters, and (c) energy constraints in time. In this chapter, we proposed a tractable optimization framework for the design of the optimal collaboration scheme that accounts for parameter correlation and temporal dynamics of sensor networks. We provided valuable insights into the problem of sensor collaboration for the estimation of correlated parameters. In order to improve computational efficiency, we proposed a fast algorithm that scales gracefully with problem size via ADMM.

Chapter 8 addressed the sensor collaboration problem with energy harvesting sensors. In this chapter, the presence of energy harvesting and storage constraints introduces challenges in problem formulation and optimization compared to the collaboration problem for estimation with battery-fixed sensors. Optimal energy allocation and storage control polices were designed under a non-convex optimization lens.

Despite our contributions in this thesis that advanced the state-of-the-art, the understanding of resource management problems remains incomplete. In the next section, we will discuss some interesting research topics that follow directly from this thesis.

9.2   Directions for future research

We conclude this thesis by listing some directions for future research.

- *Decentralized sensor selection/scheduling*: In order to reduce the computational burden of the fusion center, developing a decentralized architecture where the optimization procedure can be carried out in a distributed way and by the sensors themselves becomes quite interesting. For the problem of sensor selection with correlated noise, the presence of the positive semidefinite constraint restricts the implementation of sensor selection algorithms in a distributed fashion. Some potential ideas may include a) the use of block coordinate descent algorithm for solving SDP, and b) replacing the positive semidefinite constraint with other approximate sub-constraints. Moreover, we feel that a consensus-based formulation might be useful in the design of decentralized sensor scheduling framework.

- *Sensor collaboration with imperfect communications*: It will be worthwhile to extend the collaborative estimation model to include collaboration noise, collaboration link failures, and quantized communication. These conditions make the problem more realistic and practical since sensor collaboration may suffer from environmental noises, packet loss in wireless digital communications my cause links to fail intermittently, and sensors may have a limited bit budget for communication. The question of interest is whether or not sensor collaboration can improve the estimation performance in the presence of imperfect communications compared to the conventional amplify-and-forward transmission strategy.

- *Causality of energy harvesting in sensor collaboration*: In the thesis, we have assumed that full knowledge about the harvested energy is known in advance. Instead of using non-causal information, it will be interesting to incorporate causality of energy harvesting in sensor collaboration. Causality introduces inherent temporal dynamics of sensor collaboration and energy storage. Meanwhile, the stability of sensor collaboration over an infinite time horizon could be investigated. The steady state of sensor collaboration will enable the use of an offline protocol in energy management.

- *Hardware helping resource management implementation*: At a higher level, previous research efforts in resource management protocols focus on theoretical foundations and software-based solvers of optimal sensor scheduling and energy allocation. However, performing the software-based algorithms on voluminous data sets suffer from limited scalability due to the relatively high memory requirement and computational complexity, which thereby limits the applicability of resource management solutions on big data problems. It has recently been found that a crossbar array of memristor devices (known as memristor crossbar) exhibits a unique type of parallelism and can be utilized to solve systems of linear equations with an astonishing $O(1)$ time complexity. Therefore, it will be interesting to study resource management problems from the aspects of novel computing hardware architecture and paradigm such that the mathematical programming solvers are accelarated and scalability is enhanced.

APPENDIX A

APPENDIX

A.1 Proof of Proposition 3.1

Problem (3.11) can be equivalently transformed to

$$
\begin{aligned}
\underset{\mathbf{w},\mathbf{Z}}{\text{minimize}} \quad & \text{tr}\left(\mathbf{Z}\right) \\
\text{subject to} \quad & \mathbf{C} - \mathbf{B}^T\left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1}\mathbf{B} \succeq \mathbf{Z}^{-1}, \\
& \mathbf{1}^T\mathbf{w} \leq s, \ \mathbf{w} \in \{0,1\}^m,
\end{aligned}
\tag{A.1}
$$

where $\mathbf{Z} \in \mathbb{S}^n$ is an auxiliary variable, and the first inequality constraint is obtained from $(\mathbf{C} - \mathbf{B}^T\left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1}\mathbf{B})^{-1} \preceq \mathbf{Z}$, which implicitly adds the additional constraint $\mathbf{Z} \succeq 0$, since the left hand side of the above inequality is the inverse of the Fisher information matrix.

We further introduce another auxiliary variable $\mathbf{V} \in \mathbb{S}^n$ such that the first matrix inequality of (A.1) is expressed as

$$
\mathbf{C} - \mathbf{V} \succeq \mathbf{Z}^{-1}, \ \text{and} \ \mathbf{V} \succeq \mathbf{B}^T\left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1}\mathbf{B}.
\tag{A.2}
$$

Note that the minimization of $\text{tr}(\mathbf{Z})$ with inequalities in (A.2) would force the variable $\mathbf{V}$ to achieve its lower bound. In other words, problem (A.1) is equivalent to the problem in which the inequality

constraint in (A.1) is replaced by the two inequalities in (A.2). According to the Schur complement, the inequalities in (A.2) can be then cast as the following linear matrix inequalities (LMIs)

$$\begin{bmatrix} \mathbf{C} - \mathbf{V} & \mathbf{I} \\ \mathbf{I} & \mathbf{Z} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \mathbf{V} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{S}^{-1} + a^{-1} \operatorname{diag}(\mathbf{w}) \end{bmatrix} \succeq 0. \tag{A.3}$$

Substituting (A.3) into (A.1), the sensor selection problem becomes

$$\begin{aligned} \underset{\mathbf{w}, \mathbf{Z}, \mathbf{V}}{\text{minimize}} \quad & \operatorname{tr}\left(\mathbf{Z}\right) \\ \text{subject to} \quad & \text{LMIs in (A.3), } \mathbf{1}^T \mathbf{w} \le s, \ \mathbf{w} \in \{0, 1\}^m. \end{aligned} \tag{A.4}$$

The Boolean constraint on the entries of $\mathbf{w}$ can be enforced by $\operatorname{diag}(\mathbf{w}\mathbf{w}^T) = \mathbf{w}$, where $\operatorname{diag}(\cdot)$ returns in vector form the diagonal entries of its matrix argument. By introducing an auxiliary variable $\mathbf{W}$ together with the rank-one constraint $\mathbf{W} = \mathbf{w}\mathbf{w}^T$, the energy and Boolean constraints in (A.4) can be expressed as

$$\operatorname{tr}(\mathbf{W}) \le s, \ \operatorname{diag}(\mathbf{W}) = \mathbf{w}, \ \mathbf{W} = \mathbf{w}\mathbf{w}^T. \tag{A.5}$$

Combining (A.4) and (A.5), we obtain problem 3.12. The proof is now complete. ∎

A.2 Proof of Proposition 3.2

Given the sensor selection scheme $\tilde{\mathbf{w}}$, it is clear from (3.5) that Fisher information can be written as

$$\mathbf{J}_{\tilde{w}} = \mathbf{\Sigma}^{-1} + [\mathbf{H}_w^T, \mathbf{h}_j]\mathbf{R}_v^{-1}\begin{bmatrix} \mathbf{H}_w \\ \mathbf{h}_j^T \end{bmatrix}, \quad \mathbf{R}_{\tilde{w}} := \begin{bmatrix} \mathbf{R}_w & \mathbf{r}_j \\ \mathbf{r}_j^T & r_{jj} \end{bmatrix} \tag{A.6}$$

where $\mathbf{H}_w := \mathbf{\Phi}_w \mathbf{H}$.

If $\mathbf{w} \neq \mathbf{0}$, the inverse of $\mathbf{R}_{\tilde{w}}$ in (A.6) is given by

$$\mathbf{R}_{\tilde{w}}^{-1} = c_j \begin{bmatrix} c_j^{-1}\mathbf{R}_w^{-1} + \mathbf{R}_w^{-1}\mathbf{r}_j\mathbf{r}_j^T\mathbf{R}_w^{-1} & -\mathbf{R}_w^{-1}\mathbf{r}_j \\ -\mathbf{r}_j^T\mathbf{R}_w^{-1} & 1 \end{bmatrix} \tag{A.7}$$

where $c_j := 1/(r_{jj} - \mathbf{r}_j^T\mathbf{R}_w^{-1}\mathbf{r}_j)$, and $c_j > 0$ following from the Schur complement of $\mathbf{R}_{\tilde{w}}$. Substituting (A.7) into (A.6), we obtain

$$\mathbf{J}_{\tilde{w}} = \mathbf{J}_w + c_j\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^T, \tag{A.8}$$

where $\mathbf{J}_w = \mathbf{\Sigma}^{-1} + \mathbf{H}_w^T\mathbf{R}_w^{-1}\mathbf{H}_w$ as indicated by (3.5), and $\boldsymbol{\alpha}_j := \mathbf{H}_w^T\mathbf{R}_w^{-1}\mathbf{r}_j - \mathbf{h}_j$.

If $\mathbf{w} = \mathbf{0}$, namely, $\mathbf{J}_w = \mathbf{\Sigma}^{-1}$, we can immediately obtain from (A.6) that

$$\mathbf{J}_{\tilde{w}} = \mathbf{J}_w + \frac{1}{r_{jj}}\mathbf{h}_j^T\mathbf{h}_j. \tag{A.9}$$

Equations (A.8) and (A.9) imply that $\mathbf{J}_{\tilde{w}} - \mathbf{J}_w \succeq \mathbf{0}$ since $c_j > 0$.

We apply the matrix inversion lemma to (A.8). This yields $\mathbf{J}_{\tilde{w}}^{-1} = [\mathbf{J}_w + c_j\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^T]^{-1} = \mathbf{J}_w^{-1} - \frac{c_j\mathbf{J}_w^{-1}\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^T\mathbf{J}_w^{-1}}{1+c_j\boldsymbol{\alpha}_j\mathbf{J}_w^{-1}\boldsymbol{\alpha}_j}$. The improvement in estimation error is then given by $\text{tr}(\mathbf{J}_w^{-1}) - \text{tr}(\mathbf{J}_{\tilde{w}}^{-1}) = \frac{c_j\boldsymbol{\alpha}_j^T\mathbf{J}_w^{-2}\boldsymbol{\alpha}_j}{1+c_j\boldsymbol{\alpha}_j\mathbf{J}_w^{-1}\boldsymbol{\alpha}_j}$.. The proof is now complete. ∎

A.3 Proof of Proposition 3.3

According to (3.15), we obtain

$$
\begin{aligned}
\mathbf{R}_w^{-1} &= (\boldsymbol{\Phi}_w \mathbf{R} \boldsymbol{\Phi}_w^T)^{-1} = (\boldsymbol{\Phi}_w \boldsymbol{\Lambda} \boldsymbol{\Phi}_w^T + \epsilon \boldsymbol{\Phi}_w \boldsymbol{\Upsilon} \boldsymbol{\Phi}_w^T)^{-1} \\
&\stackrel{(1)}{=} (\mathbf{I} + \epsilon \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T \boldsymbol{\Phi}_w \boldsymbol{\Upsilon} \boldsymbol{\Phi}_w^T)^{-1} \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T \\
&\stackrel{(2)}{=} (\mathbf{I} - \epsilon \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T \boldsymbol{\Phi}_w \boldsymbol{\Upsilon} \boldsymbol{\Phi}_w^T) \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T + O(\epsilon^2) \quad (\text{as } \epsilon \to 0) \\
&\stackrel{(3)}{=} \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T - \epsilon \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \mathbf{D}_w \boldsymbol{\Upsilon} \mathbf{D}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T + O(\epsilon^2) \quad (\text{as } \epsilon \to 0),
\end{aligned}
\tag{A.10}
$$

where $\mathbf{D}_w := \operatorname{diag}(\mathbf{w})$. In (A.10), step (1) holds since we use the facts that $\boldsymbol{\Lambda}$ is a diagonal matrix and $(\boldsymbol{\Phi}_w \boldsymbol{\Lambda} \boldsymbol{\Phi}_w^T)^{-1} = \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T$; step (2) is obtained from the Taylor series expansion $(\mathbf{I} + \epsilon \mathbf{X})^{-1} = \sum_{i=0}^{\infty} (-\epsilon \mathbf{X})^i$ as $\epsilon \to 0$ (namely, the spectrum of $\epsilon \mathbf{X}$ is contained inside the open unit disk); step (3) is true since $\boldsymbol{\Phi}_w^T \boldsymbol{\Phi}_w = \mathbf{D}_w$ as in (3.4).

Substituting (A.10) into (3.5), we obtain

$$
\begin{aligned}
\mathbf{J}_w &= \boldsymbol{\Sigma}^{-1} + \mathbf{H}^T \boldsymbol{\Phi}_w^T \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T \boldsymbol{\Phi}_w \mathbf{H} - \epsilon \mathbf{H}^T \boldsymbol{\Phi}_w^T \boldsymbol{\Phi}_w \boldsymbol{\Lambda}^{-1} \mathbf{D}_w \boldsymbol{\Upsilon} \mathbf{D}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_w^T \boldsymbol{\Phi}_w \mathbf{H} + O(\epsilon^2) \quad (\text{as } \epsilon \to 0) \\
&\stackrel{(1)}{=} \boldsymbol{\Sigma}^{-1} + \mathbf{H}^T (\mathbf{D}_w \boldsymbol{\Lambda}^{-1} \mathbf{D}_w - \epsilon \mathbf{D}_w \boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon} \boldsymbol{\Lambda}^{-1} \mathbf{D}_w) \mathbf{H} + O(\epsilon^2) \quad (\text{as } \epsilon \to 0) \\
&= \boldsymbol{\Sigma}^{-1} + \mathbf{H}^T \mathbf{D}_w (\boldsymbol{\Lambda}^{-1} - \epsilon \boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon} \boldsymbol{\Lambda}^{-1}) \mathbf{D}_w \mathbf{H} + O(\epsilon^2) \quad (\text{as } \epsilon \to 0) \\
&\stackrel{(2)}{=} \boldsymbol{\Sigma}^{-1} + \mathbf{H}^T \mathbf{D}_w \mathbf{R}^{-1} \mathbf{D}_w \mathbf{H} + O(\epsilon^2) \quad (\text{as } \epsilon \to 0) \\
&\stackrel{(3)}{=} \boldsymbol{\Sigma}^{-1} + \mathbf{H}^T (\mathbf{w} \mathbf{w}^T \circ \mathbf{R}^{-1}) \mathbf{H} + O(\epsilon^2) \quad (\text{as } \epsilon \to 0),
\end{aligned}
$$

where step (1) is achieved by using the fact that $\mathbf{D}_w \boldsymbol{\Lambda}^{-1} = \boldsymbol{\Lambda}^{-1} \mathbf{D}_w = \mathbf{D}_w \boldsymbol{\Lambda}^{-1} \mathbf{D}_w$, step (2) holds due to $\mathbf{R}^{-1} = \boldsymbol{\Lambda}^{-1} - \epsilon \boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon} \boldsymbol{\Lambda}^{-1} + O(\epsilon^2)$, and step (3) is true since $\mathbf{D}_w$ is diagonal and has only binary elements. $\blacksquare$

A.4   Proof of Proposition 3.4

We begin by simplifying the objective function in (P2),

$$\phi(\mathbf{w}) := \text{tr}(\boldsymbol{\Sigma}^{-1}) + \text{tr}\left((\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})(\mathbf{H}^T\mathbf{H})\right)$$

$$= \text{tr}(\boldsymbol{\Sigma}^{-1}) + \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j \bar{R}_{ij}\mathbf{h}_i^T\mathbf{h}_j = \text{tr}(\boldsymbol{\Sigma}^{-1}) + \mathbf{w}^T\boldsymbol{\Omega}\mathbf{w}, \tag{A.11}$$

where $\bar{R}_{ij}$ is the (i,j)th entry of $\mathbf{R}^{-1}$, and $\bar{R}_{ij}\mathbf{h}_i^T\mathbf{h}_j$ corresponds to the (i,j)th entry of $\boldsymbol{\Omega}$ which yields the succinct form $\boldsymbol{\Omega} = \mathbf{A}(\mathbf{R}^{-1} \otimes \mathbf{I}_n)\mathbf{A}^T$. Here let $\otimes$ denote the Kronecker product, $\mathbf{A} \in \mathbb{R}^{m \times mn}$ is a block-diagonal matrix whose diagonal blocks are given by $\{\mathbf{h}_i^T\}_{i=1}^{m}$, and $\boldsymbol{\Omega} \succeq 0$ due to $\mathbf{R}^{-1} \otimes \mathbf{I}_n \succeq 0$. According to (A.11), (P2) can be rewritten as

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mathbf{w}^T\boldsymbol{\Omega}\mathbf{w} \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{w} \le s, \ \mathbf{w} \in \{0,1\}^m. \end{aligned} \tag{A.12}$$

Next, we prove that problem (3.17) is equivalent to problem (A.12). We recall that the former is a relaxation of the latter, where the former entails the maximization of a convex quadratic function over a bounded polyhedron $\mathcal{P} := \{\mathbf{w}|\mathbf{1}^T\mathbf{w} \le s, \mathbf{w} \in [0,1]^m\}$. It has been shown in [113] that optimal solutions of such a problem occur at vertices of the polyhedron $\mathcal{P}$, which are zero-one vectors. This indicates that solutions of problem (3.17) are feasible for problem (A.12). Therefore, solutions of (3.17) are solutions of (A.12), and vice versa. ∎

A.5   Proof of Proposition 4.1

The optimization problem (4.17) is equivalent to

$$
\begin{aligned}
\text{minimize} \quad & \phi(\{L_k\}) = \sum_{k=0}^{K-1} \text{tr}(P_k) + \frac{\rho}{2} \sum_{k=0}^{K-1} \text{tr}[(L_k - U_k)^T (L_k - U_k)] \\
\text{subject to} \quad & P_{k+1} = (A - L_k C) P_k (A - L_k C)^T + BQB^T + L_k R L_k^T.
\end{aligned}
\tag{L-Φ}
$$

To find the necessary conditions for optimality, we find the gradient of ϕ and set $\nabla_{L_k} \phi = 0$ for $k = 0, 1, \cdots, K - 1$.

We begin by assuming an incremental change in the unknown variables $\{L_k\}$ and finding the resulting incremental change to the value of the objective. Replacing L_k with $L_k + \delta L_k$ and ϕ with $\phi + \delta\phi$ in the objective function of (L-Φ) and collecting first order variation terms on both sides, we obtain

$$
\delta\phi = \sum_{k=0}^{K-1} \text{tr}(\delta P_k) + \frac{\rho}{2} \sum_{k=0}^{K-1} \text{tr}[(L_k - U_k)^T \delta L_k + \delta L_k^T (L_k - U_k)].
$$

We note that for δL_k to constitute a legitimate variation of L_k, it has to satisfy the periodicity property $\delta L_{k+K} = \delta L_k$. Similarly, replacing L_k with $L_k + \delta L_k$ and P_k with $P_k + \delta P_k$ in the constraint equation of (L-Φ) and collecting first-order variation terms on both sides, we obtain

$$
\delta P_{k+1} = (A - L_k C) P_k (A - L_k C)^T + \delta M_k,
$$

where

$$
\delta M_k = -\delta L_k C P_k (A - L_k C)^T - (A - L_k C) P_k C^T \delta L_k^T + \delta L_k R L_k^T + L_k R \, \delta L_k^T.
$$

The difficulty with finding the gradient of ϕ from the above equation is the dependence of $\delta\phi$ on δP_k, with the dependence of δP_k on δL_k being through a Lyapunov recursion. In what follows, we aim to express $\sum_{k=0}^{K-1} \text{tr}(\delta P_k)$ in terms of $\{\delta L_k\}$.

It is easy to see that

$$\delta P_k = \delta M_{k-1} + \sum_{n=k-1}^{-\infty} (A - L_{k-1}C) \cdots (A - L_nC)\delta M_{n-1}(A - L_nC)^T \cdots (A - L_{k-1}C)^T.$$

Taking the trace of both sides of the equation and summing over k, we have

$$\sum_{k=0}^{K-1} \text{tr}(\delta P_k) = \sum_{k=0}^{K-1} \text{tr}(\delta M_{k-1}) + \sum_{k=0}^{K-1} \sum_{n=k-1}^{-\infty} \text{tr}[(A - L_nC)^T \cdots$$

$$(A - L_{k-1}C)^T(A - L_{k-1}C) \cdots (A - L_nC)\delta M_{n-1}],$$

where we have used the property of the trace to change the order of the terms inside the square brackets. Now exploiting the periodicity properties $L_{k+K} = L_k$, $\delta L_{k+K} = \delta L_k$, $P_{k+K} = P_k$, which also imply the periodicity $\delta M_{k+K} = \delta M_k$ of $\{\delta M_k\}$, the double sum in the last equation above can be rewritten to give

$$\sum_{k=0}^{K-1} \text{tr}(\delta P_k)$$

$$= \sum_{n=0}^{K-1} \text{tr}(\delta M_{n-1}) + \sum_{n=0}^{K-1} \sum_{k=n+1}^{+\infty} \text{tr}[(A - L_nC)^T \cdots (A - L_{k-1}C)^T(A - L_{k-1}C) \cdots (A - L_nC)\delta M_{n-1}]$$

$$= \sum_{n=0}^{K-1} \text{tr}\{[I + \sum_{k=n+1}^{+\infty} (A - L_nC)^T \cdots (A - L_{k-1}C)^T I(A - L_{k-1}C) \cdots (A - L_nC)]\delta M_{n-1}\}.$$

To help with the simplification of the above sums, we define the new matrix variable V_n as

$$V_n = I + \sum_{k=n+1}^{+\infty} (A - L_nC)^T \cdots (A - L_{k-1}C)^T I(A - L_{k-1}C) \cdots (A - L_nC).$$

It can be seen that $\{V_n\}$ is periodic, $V_{n+K} = V_n$, and satisfies

$$V_n = (A - L_nC)^T V_{n+1}(A - L_nC) + I.$$

Returning to $\sum_{k=0}^{K-1} \text{tr}(\delta P_k)$ and using the definition of V_n, we obtain

$$\sum_{k=0}^{K-1} \text{tr}(\delta P_k) = \sum_{n=0}^{K-1} \text{tr}(V_n \delta M_{n-1}) = \sum_{n=0}^{K-1} \text{tr}(V_{n+1} \delta M_n),$$

where the last equality results from the periodicity of $\{V_n\}$ and $\{\delta M_n\}$. Recalling that δM_n can be written explicity in terms of $\{\delta L_k\}$, we have thus achieved our goal of expressing $\sum_{k=0}^{K-1} \text{tr}(\delta P_k)$ in terms of $\{\delta L_k\}$. We next carry out the last step required to find the gradient of ϕ.

Replacing $\sum_{k=0}^{K-1} \text{tr}(\delta P_k)$ with $\sum_{k=0}^{K-1} \text{tr}(V_{k+1} \delta M_k)$ in the expression for $\delta\phi$, and using the definition of δM_k, we obtain

$$\delta\phi = \sum_{k=0}^{K-1} \text{tr}(V_{k+1} \delta M_k) + \frac{\rho}{2} \sum_{k=0}^{K-1} \text{tr}[(L_k - U_k)^T \delta L_k + \delta L_k^T (L_k - U_k)]$$

$$= 2 \sum_{k=0}^{K-1} \text{tr}[-CP_k(A - L_kC)^T V_{k+1} \delta L_k + RL_k^T V_{k+1} \delta L_k] + \rho \sum_{k=0}^{K-1} \text{tr}[(L_k - U_k)^T \delta L_k],$$

where we have used the properties of the trace to arrive at the last equality. Thus

$$\nabla_{L_k}\phi = [-2CP_k(A - L_kC)^T V_{k+1} + RL_k^T V_{k+1} + \rho(L_k - U_k)^T]^T.$$

Setting $\nabla_{L_k}\phi = 0$ gives the necessary condition for optimality

$$0 = -2V_{k+1}(A - L_kC)P_kC^T + 2V_{k+1}L_kR + \rho(L_k - U_k),$$

where P_k and V_k satisfy the recursion euqations

$$P_{k+1} = (A - L_kC)P_k(A - L_kC)^T + BQB^T + L_kRL_k^T$$

$$V_k = (A - L_kC)^T V_{k+1}(A - L_kC) + I.$$

The proof is now complete. ∎

A.6 Proof of Proposition 4.3

Problem (4.24) is equivalent to

$$\text{minimize} \quad \psi_m(\boldsymbol{\mathcal{G}}_m) := \sum_{k=0}^{K-1} \gamma \operatorname{card}\big(\|\mathbf{G}_{k,m}\|_2\big) + \sum_{k=0}^{K-1} \frac{\rho}{2}\|\mathbf{G}_{k,m} - \mathbf{S}_{k,m}\|_2^2$$

$$\text{subject to} \quad \sum_{k=0}^{K-1} \operatorname{card}\big(\|\mathbf{G}_{k,m}\|_2\big) = q, \quad q \in \{0, \ldots, \eta\}$$

where $\boldsymbol{\mathcal{G}}_m := [\mathbf{G}_{0,m}, \cdots, \mathbf{G}_{K-1,m}]$. Similar to $\boldsymbol{\mathcal{G}}_m$, we form the matrix $\boldsymbol{\mathcal{S}}_m$ by picking out the mth column from each of the matrices in the set $\{\mathbf{S}_k\}$ and stacking them to obtain $\boldsymbol{\mathcal{S}}_m :=$ $[\mathbf{S}_{0,m}, \cdots, \mathbf{S}_{K-1,m}]$. We define $\kappa := \sum_{k=0}^{K-1} \operatorname{card}\big(\|\mathbf{S}_{k,m}\|_2\big)$, which gives the column-cardinality of $\boldsymbol{\mathcal{S}}_m$.

It can be shown that if $q = \kappa$ then the minimizer $\boldsymbol{\mathcal{G}}_m^\kappa$ of problem (4.24) is $\boldsymbol{\mathcal{S}}_m$, and $\psi_m(\boldsymbol{\mathcal{G}}_m^\kappa) = \gamma\kappa$. If $q > \kappa$, we have $\psi_m(\boldsymbol{\mathcal{G}}_m) > \psi_m(\boldsymbol{\mathcal{G}}_m^\kappa)$ for arbitrary values of $\boldsymbol{\mathcal{G}}_m \in F_q$ since $\psi_m(\boldsymbol{\mathcal{G}}_m) = \gamma q + \frac{\rho}{2}\|\boldsymbol{\mathcal{G}}_m - \boldsymbol{\mathcal{S}}_m\|_F^2$ which is greater than $\psi_m(\boldsymbol{\mathcal{G}}_m^\kappa) = \gamma\kappa$. Therefore, the solution of (4.23) is only determined by solving the sequence of minimization problems (4.24) for $q = 0, 1, \ldots, \min\{\eta, \kappa\}$ rather than $q = 0, 1, \cdots, \eta$.

For a given $q \in \{0, 1, \min\{\eta, \kappa\}\}$, problem (4.24) can be written as

$$\underset{\mathcal{G}_m}{\text{minimize}} \quad \sum_{k=0}^{K-1} \frac{\rho}{2}\|\mathbf{G}_{k,m} - \mathbf{S}_{k,m}\|_2^2$$

$$\text{subject to} \quad \sum_{k=0}^{K-1} \operatorname{card}\big(\|\mathbf{G}_{k,m}\|_2\big) = q.$$

For $q = 0$, the minimizer $\boldsymbol{\mathcal{G}}_m^q$ of the optimization problem is $\mathbf{0}$. For $q \neq 0$, it was demonstrated in [87, Appendix B] that the solution is obtained by projecting the minimizer ($\boldsymbol{\mathcal{G}}_m = \boldsymbol{\mathcal{S}}_m$) of the objective function onto the constraint set $\sum_{k=0}^{K-1} \operatorname{card}\big(\|\mathbf{G}_{k,m}\|_2\big) = q$. This gives

$$\mathbf{G}_{k,m} = \begin{cases} \mathbf{S}_{k,m} & \|\mathbf{S}_{k,m}\|_2 \geq \|[\boldsymbol{\mathcal{S}}_m]_q\|_2 \\ \mathbf{0} & \|\mathbf{S}_{k,m}\|_2 < \|[\boldsymbol{\mathcal{S}}_m]_q\|_2 \end{cases}, \tag{A.13}$$

for $k = 0, 1, \cdots, K - 1$, where $[\boldsymbol{S}_m]_q$ is the qth largest column of $\boldsymbol{S}_m$ in the 2-norm sense. The proof is now complete. ∎

A.7 Proof of Proposition 4.4

According to Prop. 4.3, substituting the minimizer $\boldsymbol{\mathcal{G}}_m^q$ of problem (4.24) into its objective function yields

$$\psi_m(\boldsymbol{\mathcal{G}}_m^q) = \sum_{k=0}^{K-1} \frac{\rho}{2}\|\mathbf{G}_{k,m} - \mathbf{S}_{k,m}\|_2^2 + \sum_{k=0}^{K-1} \gamma\mathrm{card}\left(\|\mathbf{G}_{k,m}\|_2\right) = \sum_{\substack{k=0 \\ k\notin\chi_q}}^{K-1} \frac{\rho}{2}\|\mathbf{S}_{k,m}\|_2^2 + \gamma q,$$

where $\mathbf{S}_{k,m}$ denotes the mth column of $\mathbf{S}_k$, χ_q is a set that is composed by indices of the first q largest columns of $\boldsymbol{\mathcal{S}}_m$ (refer to Appendix A.6) in the 2-norm sense, and $\chi_0 = \emptyset$. We have

$$\psi_m(\boldsymbol{\mathcal{G}}_m^q) - \psi_m(\boldsymbol{\mathcal{G}}_m^{q+1}) = \frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_{q+1}\|_2^2 - \gamma \tag{A.14}$$

for $q = 0, 1, \ldots, \min\{\eta, \kappa\} - 1$, where $[\boldsymbol{\mathcal{S}}_m]_{q+1}$ denotes the $(q+1)$th largest column of $\boldsymbol{\mathcal{S}}_m$ in the 2-norm sense.

Since $[\boldsymbol{\mathcal{S}}_m]_1 \geq [\boldsymbol{\mathcal{S}}_m]_2 \geq \cdots \geq [\boldsymbol{\mathcal{S}}_m]_{\min\{\eta,\kappa\}}$, for $\gamma \in (\frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_1\|_2^2, \infty)$ equation (A.14) yields $\psi_m(\boldsymbol{\mathcal{G}}_m^0) - \psi_m(\boldsymbol{\mathcal{G}}_m^1) < 0$ and $\psi_m(\boldsymbol{\mathcal{G}}_m^q) - \psi_m(\boldsymbol{\mathcal{G}}_m^{q+1}) < 0$ for other $q \in \{1, \ldots, \min\{\eta, \kappa\} - 1\}$. Therefore, the minimizer of (4.23) is given by $\boldsymbol{\mathcal{G}}_m^0$. Similarly, for $\gamma \in (\frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_{q+1}\|_2^2, \frac{\rho}{2}\|[\boldsymbol{\mathcal{S}}_m]_q\|_2^2]$ equation (A.14) yields $\psi_m(\boldsymbol{\mathcal{G}}_m^{l-1}) - \psi_m(\boldsymbol{\mathcal{G}}_m^l) \geq 0$ for $l = 1, \ldots, q$, and $\psi_m(\boldsymbol{\mathcal{G}}_m^l) - \psi_m(\boldsymbol{\mathcal{G}}_m^{l+1}) < 0$ for $l = q, \ldots, \min\{\eta, \kappa\} - 1$. Therefore, the minimizer of (4.23) is given by $\boldsymbol{\mathcal{G}}_m^q$. Finally, we can write the solution of (4.23) in the form given in the statement of Prop. 4.4. The proof is now complete. $\blacksquare$

A.8   Proof of Proposition 5.1

To prove the convexity of problem (5.10), it is sufficient to study the convexity of the function $g_m(\mathbf{w}) = p^2(\mathbf{w})$, where $p(\mathbf{w}) := (\sum_{k=1}^{K} \|\mathbf{w}_{k,m}\|_2)^2$, which is a convex function.

Since $p(\mathbf{w})$ is convex, for any two points $\mathbf{v}$, $\mathbf{u}$ and $\theta \in [0, 1]$, we obtain

$$p^2(\theta\mathbf{v} + (1 - \theta)\mathbf{u}) \leq (\theta p(\mathbf{v}) + (1 - \theta)p(\mathbf{u}))^2. \tag{A.15}$$

By comparing the right hand side of (A.15) and the function $\theta g_m(\mathbf{v}) + (1 - \theta)g_m(\mathbf{u})$, we have

$$(\theta p(\mathbf{v}) + (1 - \theta)p(\mathbf{u}))^2 - \theta g_m(\mathbf{u}) - (1 - \theta)g_m(\mathbf{v})$$
$$= \theta^2 p^2(\mathbf{v}) + (1 - \theta)^2 p^2(\mathbf{u}) + 2\theta(1 - \theta)p(\mathbf{v})p(\mathbf{u}) - \theta p^2(\mathbf{v}) - (1 - \theta)p^2(\mathbf{u})$$
$$= -\theta(1 - \theta)(p(\mathbf{v}) - p(\mathbf{u}))^2 \leq 0. \tag{A.16}$$

From (A.15) and (A.16), we obtain

$$g_m(\theta\mathbf{v} + (1 - \theta)\mathbf{u}) \leq \theta g_m(\mathbf{v}) + (1 - \theta)g_m(\mathbf{u}),$$

which proves that the function $g_m(\mathbf{w})$ is convex. This proves that the optimization problem (5.10) is convex. ∎

A.9 Proof of Proposition 5.2

Problem (5.20) can be written as

$$
\begin{aligned}
&\underset{\{\mathbf{u}_{k,m}\},\mathbf{r}}{\text{minimize}} \quad \mathbf{r}^T \mathbf{1}\mathbf{1}^T \mathbf{r} + \frac{1}{2\hat{\eta}} \sum_{k=1}^{K} \|\mathbf{u}_{k,m} - \mathbf{c}_{k,m}\|_2^2 \\
&\text{subject to} \quad r_k = \|\mathbf{u}_{k,m}\|_2, \ k = 1, 2, \ldots, K,
\end{aligned}
\tag{A.17}
$$

where we use $\{\mathbf{u}_{k,m}\}$ instead of $\{\mathbf{u}_{k,m}\}_{k=1,2,\ldots,K}$ for simplicity.

Problem (A.17) can be further transformed to

$$
\underset{\mathbf{r} \geq \mathbf{0}}{\text{minimize}} \left\{ \underset{\{\|\mathbf{u}_{k,m}\|_2 = r_k\}}{\text{minimize}} \ \mathbf{r}^T \mathbf{1}\mathbf{1}^T \mathbf{r} + \frac{1}{2\hat{\eta}} \sum_{k=1}^{K} \|\mathbf{u}_{k,m} - \mathbf{c}_{k,m}\|_2^2 \right\},
\tag{A.18}
$$

where the inner minimization problem is with respect to $\{\mathbf{u}_{k,m}\}$, while the outer is with respect to $\mathbf{r}$. Note that the equivalence between (A.17) and (A.18) can be verified, proceeding by contradiction and using the fact that the solution of problem (A.18) is unique.

We first consider the inner minimization problem of (A.18)

$$
\begin{aligned}
&\underset{\{\mathbf{u}_{k,m}\}}{\text{minimize}} \quad \frac{1}{2\hat{\eta}} \sum_{k=1}^{K} \|\mathbf{u}_{k,m} - \mathbf{c}_{k,m}\|_2^2 \\
&\text{subject to} \quad \|\mathbf{u}_{k,m}\|_2 = r_k, \ k = 1, 2, \ldots, K,
\end{aligned}
\tag{A.19}
$$

which can be decomposed into K subproblems

$$
\begin{aligned}
&\underset{\mathbf{u}_{k,m}}{\text{minimize}} \quad \frac{1}{2\hat{\eta}} \|\mathbf{u}_{k,m} - \mathbf{c}_{k,m}\|_2^2 \\
&\text{subject to} \quad \|\mathbf{u}_{k,m}\|_2 = r_k
\end{aligned}
\tag{A.20}
$$

for $k = 1, 2, \ldots, K$.

From a geometrical point of view, the minimizer of (A.20) can be interpreted as a point lying at the surface of Euclidean ball $\|\mathbf{u}_{k,m}\|_2 \leq r_k$ such that its distance from a give point $\mathbf{c}_{k,m}$ is minimized. Therefore, the solution of (A.20) is a vector with *length* r_k and *direction* $\frac{\mathbf{c}_{k,m}}{\|\mathbf{c}_{k,m}\|_2}$.

Furthermore, the solution of (A.19) can be given by $\mathbf{u}_{k,m}^{*} = r_k \frac{\mathbf{c}_{k,m}^{T}}{\|\mathbf{c}_{k,m}^{T}\|_2}$ for $k = 1, 2, \ldots, K$. Substituting $\{\mathbf{u}_{k,m}^{*}\}$ into (A.18), the outer minimization problem becomes QP (5.22). The proof is now complete. ∎

A.10 Proof of Proposition 5.4

Without loss of generality, we assume that $f_1 \leq f_2 \leq \ldots \leq f_K$. Then according to Lemma 5.3, we have $r_1^* \leq r_2^* \leq \ldots \leq r_K^*$ for the solution of problem (5.22). Our goal is to find this value of $\mathbf{r}^*$ in a closed-form.

Note that the solution of problem (5.22) is unique since it is strongly convex [59]. Also, the optimal primal-dual feasible pair $(\mathbf{r}, \boldsymbol{\nu})$ is given by KKT conditions of (5.22):

- primal and dual feasibility: $\mathbf{r} \geq \mathbf{0}$ and $\boldsymbol{\nu} \geq \mathbf{0}$, $\qquad$ (A.21a)

- complementary slackness: $r_i \nu_i = 0$ for $i = 1, 2, \ldots, K$, $\qquad$ (A.21b)

- stationary condition: $\mathbf{r} = (\mathbf{I} + 2\hat{\eta}\mathbf{1}\mathbf{1}^T)^{-1}(\mathbf{f} + \hat{\eta}\boldsymbol{\nu})$. $\qquad$ (A.21c)

According to (A.21c), we obtain

$$\mathbf{r} = \left(\mathbf{I} - \frac{2\hat{\eta}}{1 + 2\hat{\eta}K}\mathbf{1}\mathbf{1}^T\right)(\mathbf{f} + \hat{\eta}\boldsymbol{\nu}), \qquad (\text{A.22})$$

where we use the fact that $(\mathbf{I} + 2\hat{\eta}\mathbf{1}\mathbf{1}^T)(\mathbf{I} - \frac{2\hat{\eta}}{1+2K\hat{\eta}}\mathbf{1}\mathbf{1}^T) = \mathbf{I}$.

Now, from (A.21b), consider different cases for the values of optimal dual variable $\boldsymbol{\nu}$. If $\boldsymbol{\nu} = \mathbf{0}$, we have

$$r_i = f_i - \frac{2\hat{\eta}\sum_{k=1}^{K} f_k}{1 + 2\hat{\eta}K}, \ i = 1, 2, \ldots, K. \qquad (\text{A.23})$$

According to (A.21a) and (A.21b), we can conclude that $\mathbf{r}^* = \mathbf{r}$ only if $f_1 > \frac{2\hat{\eta}\sum_{k=1}^{K} f_k}{1+2\hat{\eta}K}$. If $f_1 \leq \frac{2\hat{\eta}\sum_{k=1}^{K} f_k}{1+2\hat{\eta}K}$, then the solution of (5.22) is not given by (A.23), which yields that the vector of optimal dual variables $\boldsymbol{\nu}$ is not zero. Suppose that $\nu_j \neq 0$ for some $j \in \{1, 2, \ldots, K\}$, then from (A.21b), we have $r_j^* = 0$. Recall that $0 \leq r_1^* \leq r_2^* \leq \ldots \leq r_K^*$, implying that $r_1^* = 0$. Therefore, we obtain if $f_1 \leq \frac{2\hat{\eta}\sum_{k=1}^{K} f_k}{1+2\hat{\eta}K}$, $r_1^* = 0$. In summary, we have the following result.

If $f_1 > \frac{2\hat{\eta}\sum_{k=1}^{K} f_k}{1+2\hat{\eta}K}$, then $r_i^* = f_i - \frac{2\hat{\eta}\sum_{k=1}^{K} f_k}{1+2\hat{\eta}K}$, for $i = 1, 2, \ldots, K$. If $f_1 \leq \frac{2\hat{\eta}\sum_{k=1}^{K} f_k}{1+2\hat{\eta}K}$, then

$r_1^* = 0$.

In the aforementioned conclusion, note that the values of r_i^* for $i = 2, 3 \ldots, K$ have not been determined when $f_1 \leq \frac{2\hat{\eta} \sum_{k=1}^{K} f_k}{1+2\hat{\eta} K}$. However, they can be found by letting $r_1 = 0$ and solving problem (5.22) with $K-1$ variables. Similar to the previous analysis, by exploring KKT conditions we obtain the following results.

If $f_1 \leq \frac{2\hat{\eta} \sum_{k=1}^{K} f_k}{1+2\hat{\eta} K}$ and $f_2 > \frac{2\hat{\eta} \sum_{k=2}^{K} f_k}{1+2\hat{\eta}(K-1)}$, then $r_i^* = f_i - \frac{2\hat{\eta} \sum_{k=2}^{K} f_k}{1+2\hat{\eta}(K-1)}$, for $i = 2, 3, \ldots, K$. If $f_1 \leq \frac{2\hat{\eta} \sum_{k=1}^{K} f_k}{1+2\hat{\eta} K}$ and $f_2 \leq \frac{2\hat{\eta} \sum_{k=2}^{K} f_k}{1+2\hat{\eta}(K-1)}$, then $r_2^* = 0$.

Continuing further, the solution of (5.22) can be compactly written as

$$
r_i^* = \begin{cases} 0 & 1 \leq i \leq \iota - 1, i \in \mathbb{N} \\ f_i - \dfrac{2\hat{\eta} \sum_{k=\iota}^{K} f_k}{1 + 2\hat{\eta}(K - \iota + 1)} & \iota \leq i \leq K, i \in \mathbb{N}, \end{cases} \tag{A.24}
$$

for $i = 1, 2, \ldots, K$, where ι is the index of the first positive element in the numbers $\left\{ f_i - \frac{2\hat{\eta} \sum_{k=i}^{K} f_k}{1+2\hat{\eta}(K-i+1)} \right\}$ and $\iota = K + 1$ if no positive element exists. The proof is now complete. $\blacksquare$

A.11   Proof of Proposition 5.5

We denote the optimal solution to (5.25) by $\{\mathbf{w}'_{k,m}, \mathbf{u}'_{k,m}\}$ for $k = 1, 2, \ldots, K$ and $m = 1, 2, \ldots, M$.

Our goal is to prove that $u'^{,n}_{k,m} = |w'^{,n}_{k,m}|$ for all $k = 1, 2, \ldots, K$, $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$, where $u'^{,n}_{k,m}$ and $w'^{,n}_{k,m}$ denote the nth entry of $\mathbf{u}'_{k,m}$ and $\mathbf{w}'_{k,m}$, respectively. We prove it by contradiction. Suppose this is not true, then there exists an index set

$$\mathbb{S} := \left\{ (k, m, n) | - u'^{,n}_{k,m} < w'^{,n}_{k,m} < u'^{,n}_{k,m}, \ (k, m, n) \in \mathbb{I} \right\},$$

where $\mathbb{I} := [K] \times [M] \times [N]$, $[K] := \{1, 2, \ldots, K\}$, $[M] := \{1, 2, \ldots, M\}$, and $[N] := \{1, 2, \ldots, N\}$.

Now, consider the pair $\{\mathbf{w}^*_{k,m}, \mathbf{u}^*_{k,m}\}$ such that

$$
\begin{aligned}
\mathbf{w}^*_{k,m} &= \mathbf{w}'_{k,m}, \\
u^{*,n}_{k,m} &= u'^{,n}_{k,m} \text{ for } (k, m, n) \notin \mathbb{S}, \\
u^{*,n}_{k,m} &= |w'^{,n}_{k,m}| < u'^{,n}_{k,m} \text{ for } (k, m, n) \in \mathbb{S}.
\end{aligned}
\tag{A.25}
$$

It is clear from (5.25) and (A.25) that $\{\mathbf{w}^*_{k,m}, \mathbf{u}^*_{k,m}\}$ are feasible points. Also, since $\mathbb{S}$ is not empty, there exists $k \in [K]$ and $m \in [M]$ such that $\mathbf{1}^T \mathbf{u}^*_{k,m} < \mathbf{1}^T \mathbf{u}'_{k,m}$. Further, we have $(\mathbf{1}^T \mathbf{u}^*_{k,m})^2 < (\mathbf{1}^T \mathbf{u}'_{k,m})^2$ since $\mathbf{u}^n_{k,m} > 0$ for any $(k, m, n) \in \mathbb{I}$. Therefore, the pair $\{\mathbf{w}^*_{k,m}, \mathbf{u}^*_{k,m}\}$ can yield a smaller objective value of (5.25) than that of $\{\mathbf{w}'_{k,m}, \mathbf{u}'_{k,m}\}$. This contradicts our assumption, in which $\{\mathbf{w}'_{k,m}, \mathbf{u}'_{k,m}\}$ is the solution to (5.25). Hence, we can conclude that $\mathbf{u}'_{k,m} = |\mathbf{w}'_{k,m}|$ for all k and m. ∎

A.12 Quadratic functions transformation

According to [48, Sec. III], it is straightforward to derive the quadratic vector functions (6.16), and the corresponding coefficient matrices are given by

$$\boldsymbol{\Omega}_{\text{T}} = \mathbf{I}_{\text{N}} \otimes \boldsymbol{\Sigma}_x, \ \mathbf{I}_{\text{N}} \text{ is the } N \times N \text{ identity matrix,} \tag{A.26}$$

$$\boldsymbol{\Omega}_{\text{JN}} = \mathbf{G}\mathbf{h}\mathbf{h}^T\mathbf{G}^T, \quad [\mathbf{G}]_{l,n} = \begin{cases} g_{m_l} & n = n_l, \\ 0 & \text{otherwise,} \end{cases} \tag{A.27}$$

$$\boldsymbol{\Omega}_{\text{JD}} = \mathbf{G}(\boldsymbol{\Sigma}_\epsilon + \eta^2\boldsymbol{\Sigma}_h)\mathbf{G}^T + \eta^2\mathbf{H}\boldsymbol{\Sigma}_g\mathbf{H}^T + \eta^2\boldsymbol{\Sigma}_g \otimes \boldsymbol{\Sigma}_h + \boldsymbol{\Sigma}_g \otimes \boldsymbol{\Sigma}_\epsilon, \quad \mathbf{H} = \mathbf{I}_N \otimes \mathbf{h}, \tag{A.28}$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{\Sigma}_x$ is defined in (6.4), and (m_l, n_l) is given by (6.15). It is clear from (A.26)-(A.28) that $\boldsymbol{\Omega}_{\text{T}}$, $\boldsymbol{\Omega}_{\text{JN}}$, and $\boldsymbol{\Omega}_{\text{JD}}$ are all symmetric positive semidefinite matrices, and $\boldsymbol{\Omega}_{\text{JN}}$ is of rank one. ∎

A.13 Sensor collaboration with fixed topologies

It has been shown in [48] that the sensor collaboration problems with *given* collaboration topologies can be solved analytically, since the collaboration cost $Q(\mathbf{w}) = \sum_{l=1}^{L} c_l \text{card}(w_l)$ is a constant, and (6.17) and (6.18) become problems with homogeneous quadratic functions, in which no linear terms with respect to $\mathbf{w}$ are involved. The solutions of (6.17) and (6.18) for a fully-connected network are shown in Theorem 1.

Theorem 1 [48, Theorem 1]: For a fully-connected network, the optimal values $\tilde{P}$ and J^* and solutions $\tilde{\mathbf{w}}$ and $\mathbf{w}^*$ of (6.17) and (6.18) are given by

$$
\begin{cases}
\tilde{P} = \lambda_{\min}^{\text{pos}} \left(\boldsymbol{\Omega}_{\text{T}}, -\boldsymbol{\Omega}_{\text{JD}} + \dfrac{\boldsymbol{\Omega}_{\text{JN}}}{\check{J}} \right) \xi^2 + \mathbf{1}^T \mathbf{c} \\[2ex]
\tilde{\mathbf{w}} = \sqrt{\dfrac{\tilde{P} - \mathbf{1}^T \mathbf{c}}{\tilde{\mathbf{v}}^T \boldsymbol{\Omega}_{\text{T}} \tilde{\mathbf{v}}}} \, \tilde{\mathbf{v}},
\end{cases}
\tag{A.29a}
$$

and

$$
\begin{cases}
J^* = \lambda_{\max} \left(\boldsymbol{\Omega}_{\text{JN}}, \boldsymbol{\Omega}_{\text{JD}} + \dfrac{\xi^2 \boldsymbol{\Omega}_{\text{T}}}{\hat{P} - \mathbf{1}^T \mathbf{c}} \right) \\[2ex]
\mathbf{w}^* = \sqrt{\dfrac{\hat{P} - \mathbf{1}^T \mathbf{c}}{(\mathbf{v}^*)^T \boldsymbol{\Omega}_{\text{T}} \mathbf{v}^*}} \, \mathbf{v}^*,
\end{cases}
\tag{A.29b}
$$

where $\lambda_{\min}^{\text{pos}}(\mathbf{A}, \mathbf{B})$ and $\lambda_{\max}(\mathbf{A}, \mathbf{B})$ denote the minimum positive eigenvalue and the maximum eigenvalue of the generalized eigenvalue problem $\mathbf{A}\mathbf{v} = \lambda\mathbf{B}\mathbf{v}$, respectively, and $\tilde{\mathbf{v}}$ and $\mathbf{v}^*$ are the corresponding eigenvectors. $\blacksquare$

It is clear from (A.29b) that the optimal Fisher information is upper bounded by $J_0 := \lambda_{\max}(\boldsymbol{\Omega}_{\text{JN}}, \boldsymbol{\Omega}_{\text{JD}})$ as $P \to +\infty$. This implies that to guarantee the *feasibility* of (6.17), the information threshold $\check{J}$ must lie in the interval $[0, J_0)$. Accordingly, the estimation distortion in (6.10) belongs to $(D_0, \eta^2]$, where $D_0 = \eta^2/(1 + \eta^2 J_0)$ denotes the minimum distortion, and η^2 signifies the maximum distortion which is determined by the prior information of θ. We summarize the boundedness of Fisher information and estimation distortion in Lemma 1.

Lemma 1. *For problems (6.17) and (6.18), the values of Fisher information and estimation distortion are bounded as $J(\mathbf{w}) \in [0, J_0)$ and $D(\mathbf{w}) \in (D_0, \eta^2]$, respectively, where $J_0 = \lambda_{\max}(\boldsymbol{\Omega}_{\text{JN}}, \boldsymbol{\Omega}_{\text{JD}})$,*

$D_0 = \eta^2/(1 + \eta^2 J_0)$, *and η^2 is the variance of the random parameter to be estimated.*

For a given $\check{J} \in [0, J_0)$, we demonstrate in Lemma 2 that the matrix $\check{J}\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}$ is not positive semidefinite.

Lemma 2. *Given $\check{J} \in [0, J_0)$, the matrix $\check{J}\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}$ is not positive semidefinite.*

Proof: Since $J_0 = \lambda_{\max}(\boldsymbol{\Omega}_{\text{JN}}, \boldsymbol{\Omega}_{\text{JD}})$, there exists an eigenvector $\mathbf{v}_0$ such that $\boldsymbol{\Omega}_{\text{JN}}\mathbf{v}_0 = J_0\boldsymbol{\Omega}_{\text{JD}}\mathbf{v}_0$, which yields $\mathbf{v}_0^T\boldsymbol{\Omega}_{\text{JN}}\mathbf{v}_0 = \mathbf{v}_0^T J_0\boldsymbol{\Omega}_{\text{JD}}\mathbf{v}_0$. Since $\check{J} \in [0, J_0)$, we obtain $\mathbf{v}_0^T\boldsymbol{\Omega}_{\text{JN}}\mathbf{v}_0 > \mathbf{v}_0^T \check{J}\boldsymbol{\Omega}_{\text{JD}}\mathbf{v}_0$, namely, $\mathbf{v}_0^T(\check{J}\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}})\mathbf{v}_0 < 0$. Therefore, we find a vector $\mathbf{v}_0$ such that $\mathbf{v}_0^T(\check{J}\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}})\mathbf{v}_0 < 0$. This implies that $\check{J}\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}$ is not positive semidefinite. ∎

A.14   The KKT-based solution for a QP1QC

Consider a more general case of QP1QC

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \mathbf{A}_0 \mathbf{w} + 2\mathbf{b}_0^T \mathbf{w} \\
\text{subject to} \quad & \mathbf{w}^T \mathbf{A}_1 \mathbf{w} + 2\mathbf{b}_1^T \mathbf{w} + r_1 \leq 0,
\end{aligned}
\tag{A.30}
$$

where $\mathbf{A}_0$ is a symmetric positive definite matrix, and $\mathbf{A}_1$ is a symmetric matrix.

Upon defining $\mathbf{P} := \frac{1}{r_1} \mathbf{A}_0^{-\frac{1}{2}} \mathbf{A}_1 \mathbf{A}_0^{-\frac{1}{2}}$, we obtain the eigenvalue decomposition of $\mathbf{P}$

$$
\mathbf{P} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T,
$$

where $\mathbf{U}$ is an orthogonal matrix that includes the eigenvectors of $\mathbf{P}$, and $\mathbf{\Lambda}$ is a diagonal matrix that includes the eigenvalues of $\mathbf{P}$. Let $\mathbf{u} := \mathbf{U}^T \mathbf{A}_0^{\frac{1}{2}} \mathbf{w}$, $\mathbf{g} := \mathbf{U}^T \mathbf{A}_0^{-\frac{1}{2}} \mathbf{b}_0$ and $\mathbf{e} := \mathbf{U}^T \mathbf{A}_0^{-\frac{1}{2}} \frac{\mathbf{b}_1}{r_1}$, then problem (A.30) can be written as

$$
\begin{aligned}
\underset{\mathbf{u}}{\text{minimize}} \quad & \mathbf{u}^T \mathbf{u} + 2\mathbf{u}^T \mathbf{g} \\
\text{subject to} \quad & \mathbf{u}^T \mathbf{\Lambda} \mathbf{u} + 2\mathbf{u}^T \mathbf{e} + 1 \leq 0.
\end{aligned}
\tag{A.31}
$$

The rationale behind using the eigenvalue decomposition technique to reformulate (A.30) is that the KKT conditions of (A.31) are more compact and easily solved since $\mathbf{\Lambda}$ is a diagonal matrix.

We demonstrate the KKT conditions for problem (A.31).

Primal feasibility: $\mathbf{u}^T \mathbf{\Lambda} \mathbf{u} + 2\mathbf{u}^T \mathbf{e} + 1 \leq 0$.

Dual feasibility: $\mu \geq 0$, where μ is the dual variable.

Complementary slackness: $\mu(\mathbf{u}^T \mathbf{\Lambda} \mathbf{u} + 2\mathbf{u}^T \mathbf{e} + 1) = 0$.

Stationary of the Lagrangian: $\mathbf{u} = -(\mathbf{I} + \mu \mathbf{\Lambda})^{-1}(\mathbf{g} + \mu \mathbf{e})$.

If $\mu = 0$, we have

$$
\mathbf{u} = -\mathbf{g},
\tag{A.32}
$$

where $\mathbf{u}^T \mathbf{\Lambda} \mathbf{u} + 2\mathbf{u}^T \mathbf{e} + 1 \leq 0$.

If $\mu > 0$, eliminating $\mathbf{u}$ by substituting stationary condition into complementary slackness, we have

$$(\mathbf{g} + \mu\mathbf{e})^T (\mathbf{I} + \mu\mathbf{\Lambda})^{-1} \mathbf{\Lambda} (\mathbf{I} + \mu\mathbf{\Lambda})^{-1} (\mathbf{g} + \mu\mathbf{e}) - 2(\mathbf{g} + \mu\mathbf{e})^T (\mathbf{I} + \mu\mathbf{\Lambda})^{-1} \mathbf{e} + 1 = 0.$$

Since $\mathbf{\Lambda}$ is a diagonal matrix, we finally obtain that

$$\sum_{l=1}^{L} \left(\frac{\lambda_l (\mu e_l + g_l)^2}{(\mu\lambda_l + 1)^2} - \frac{2e_l(\mu e_l + g_l)}{\mu\lambda_l + 1} \right) + 1 = 0, \tag{A.33}$$

where λ_l is the lth diagonal element of $\mathbf{\Lambda}$.

If $\lambda_l > 0$ for $l = 1, 2, \ldots, L$ and thus $\mathbf{A}_1$ is positive definite, Eq. (A.33) can be simplified as

$$\sum_{l=1}^{L} \left\{ \left[\frac{\sqrt{\lambda_l}(\mu e_l + g_l)}{1 + \mu\lambda_l} - \frac{e_l}{\sqrt{\lambda_l}} \right]^2 - \frac{e_l^2}{\lambda_l} \right\} + 1 = \sum_{l=1}^{L} \frac{(\lambda_l g_l - e_l)^2}{\lambda_l (1 + \mu\lambda_l)^2} - \sum_{l=1}^{L} \frac{e_l^2}{\lambda_l} + 1 = 0. \tag{A.34}$$

In (A.34), the function $\frac{(\lambda_l g_l - e_l)^2}{\lambda_l(1 + \mu\lambda_l)^2}$ is monotonically decreasing with respect to μ when $\mu > 0$. This implies that there exists only one positive root for $f(\mu) = 0$ if $\mu > 0$ satisfies the KKT conditions.

Let $\mathbf{A}_0 = \tilde{\mathbf{\Omega}}_{\text{T}}$, $\mathbf{b}_0 = -\frac{\rho}{2}\mathbf{a}$, $\mathbf{A}_1 = \check{J}\mathbf{\Omega}_{\text{JD}} - \mathbf{\Omega}_{\text{JN}}$, $\mathbf{b}_1 = \mathbf{0}$ and $r_1 = \check{J}\xi^2$, we can obtain the result in Proposition 6.1, while let $\mathbf{A}_0 = \tilde{\mathbf{\Omega}}_{\text{T}}$, $\mathbf{b}_0 = -\frac{\rho}{2}\mathbf{a}$, $\mathbf{A}_1 = \tilde{\mathbf{\Omega}}_{\text{JD}}$, $\mathbf{b}_1 = -\tilde{\boldsymbol{\beta}}$ and $r_1 = \tilde{\gamma}$, we obtain the result in Proposition 6.4.

■

A.15 Proof of Proposition 6.2

We recall that the eigenvalues $\{\lambda_l\}_{l=1,2,\ldots,L}$ in (6.28) are obtained from the eigenvalue decomposition of the positive definite matrix $\boldsymbol{\Omega}_{\text{JD}}$ modified by the rank one matrix $\boldsymbol{\Omega}_{\text{JN}}$. Then it can be concluded that there exists only one negative eigenvalue in $\{\lambda_l\}_{l=1,2,\ldots,L}$ [114, Sec. 5].

Without loss of generality, we assume that $\lambda_1 < 0 < \lambda_2 \leq \lambda_3 < \ldots < \lambda_N$, where the case of $\lambda_l = 0$ ($l > 1$) is excluded since it is trivial to obtain that $\frac{\lambda_l g_l^2}{(\mu\lambda_l+1)^2} = 0$ in (6.30). Given $-\frac{1}{\lambda_1} > 0 > -\frac{1}{\lambda_L} \geq \ldots \geq -\frac{1}{\lambda_3} \geq -\frac{1}{\lambda_2}$, we have $f(-\frac{1}{\lambda_1}) \to -\infty$ and $f(-\frac{1}{\lambda_l}) \to \infty$ for $l = 2, 3, \ldots, L$.

Next, we take the first-order derivative of $f(\mu)$,

$$\frac{df(\mu)}{d\mu} = \sum_{l=1}^{L}[-2\lambda_l^2 g_l^2 (1 + \mu\lambda_l)^{-3}]. \tag{A.35}$$

When $\mu \in (0, -\frac{1}{\lambda_1})$, we have

$$1 + \mu\lambda_1 > 0, \text{ and } 1 + \mu\lambda_l > 0 \text{ for } i = 2, 3, \ldots, L.$$

From (A.35) we obtain $\frac{df(\mu)}{d\mu} \leq 0$. Therefore, $f(\mu)$ is monotonically decreasing as $\mu \in (0, -\frac{1}{\lambda_1})$. Together with $f(-\frac{1}{\lambda_1}) \to -\infty$ and $f(-\frac{1}{\lambda_L}) \to \infty$, we conclude that there exists only one positive root of $f(\mu) = 0$ if $f(0) > 0$. The proof of Lemma 3 is now complete. ∎

When $\mu \in (-\frac{1}{\lambda_1}, \infty)$, we have

$$1 + \mu\lambda_1 < 0, \text{ and } 1 + \mu\lambda_l > 0 \text{ for } i = 2, 3, \ldots, L. \tag{A.36}$$

It is clear from (A.35) that the sign of $\frac{df(\mu)}{d\mu}$ is difficult to determine since $-2\lambda_1^2 g_1^2 (1 + \mu\lambda_1)^{-3} > 0$ and $-2\lambda_l^2 g_l^2 (1 + \mu\lambda_l)^{-3} < 0$ for $l = 2, 3, \ldots, L$. Therefore, the function $f(\mu)$ may not be monotonic, and the number of positive roots of $f(\mu) = 0$ is uncertain.

A.16 Proof of Proposition 6.3

We recall that

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{\Omega}_{\text{JN}} \mathbf{w}}{\mathbf{w}^T \mathbf{\Omega}_{\text{JD}} \mathbf{w} + \xi^2},$$

and

$$P(\mathbf{w}) = \mathbf{w}^T \mathbf{\Omega}_{\text{T}} \mathbf{w} + \sum_{l=1}^{L} c_l \operatorname{card}(w_l).$$

Setting $\mathbf{w} = c\hat{\mathbf{w}}$ for some fixed vector $\hat{\mathbf{w}}$, $J(\mathbf{w})$ and $P(\mathbf{w})$ are strictly increasing functions of c when $c > 1$, and strictly decreasing functions of c when $c < 1$. Thus, the optimality is achieved for (6.17) or (6.18) when the inequality constraints are satisfied with equality.

Given the energy budget $\hat{P}$, we have $P(\mathbf{w}_2) = \hat{P}$, where $\mathbf{w}_2$ is the optimal solution of the energy constrained problem (6.18). Our goal is to show $\mathbf{w}_2$ is also a solution of the information constrained problem (6.17) when $\check{J} = J(\mathbf{w}_2)$.

If $\mathbf{w}_2$ is not the solution of (6.17), we assume a better solution $\mathbf{w}'_2$ such that $P(\mathbf{w}'_2) < P(\mathbf{w}_2)$. Since $P(\cdot)$ strictly increases as multiplying the optimization variables by a scalar $c > 1$, there exists a scalar $c > 1$ such that

$$P(\mathbf{w}'_2) < P(c\mathbf{w}'_2) \le P(\mathbf{w}_2). \tag{A.37}$$

On the other hand, since $J(\cdot)$ strictly increases as multiplying the optimization variables by a scalar $c > 1$, we have $J(c\mathbf{w}'_2) > J(\mathbf{w}'_2)$. Further, because $\mathbf{w}'_2$ is a feasible vector for (6.17), we have $J(\mathbf{w}'_2) \ge \check{J}$, where recalling that $\check{J} = J(\mathbf{w}_2)$. We can then conclude that

$$J(c\mathbf{w}'_2) > J(\mathbf{w}'_2) \ge J(\mathbf{w}_2). \tag{A.38}$$

From (A.37) and $P(\mathbf{w}_2) = \hat{P}$, we obtain that $P(c\mathbf{w}'_2) \le P$, which implies $c\mathbf{w}'_2$ is a feasible point for (6.18). From (A.38), we have $J(c\mathbf{w}'_2) > J(\mathbf{w}_2)$, which implies $c\mathbf{w}'_2$ yields a higher objective value of (6.18) than $\mathbf{w}_2$. This *contradicts* to the fact that $\mathbf{w}_2$ is the optimal solution of

(6.18). Therefore, we can conclude that $\mathbf{w}_2$ is the solution of (6.17).

On the other hand, if $\mathbf{w}_1$ is the solution of (6.17), it is similar to prove that $\mathbf{w}_1$ is the solution of (6.18) when $\hat{P} = P(\mathbf{w}_1)$. The proof is now complete. ∎

A.17 Proof of Proposition 6.5

For notational convenience, we define $\kappa_1 := \frac{1}{\rho}$, $\kappa_2 := \frac{\tilde{d}_n}{\rho}$ and $h_{\kappa_2}^{\kappa_1}(\mathbf{v}_{G_n}) := \kappa_1 \|\mathbf{F}_n \mathbf{v}_{G_n}\|_1 + \kappa_2 \|\mathbf{v}_{G_n}\|_2$. Then problem (6.41) can be written as

$$\underset{\mathbf{v}_{G_n}}{\text{minimize}} \quad \phi_{\kappa_2}^{\kappa_1}(\mathbf{v}_{G_n}) := h_{\kappa_2}^{\kappa_1}(\mathbf{v}_{G_n}) + \frac{1}{2}\|\mathbf{v}_{G_n} - \mathbf{b}_{G_n}\|_2^2. \tag{A.39}$$

Let $\mathbf{v}_{G_n}^*$ be the unique minimizer of the following problem

$$\underset{\mathbf{v}_{G_n}}{\text{minimize}} \quad h_{\kappa_2}^0(\mathbf{v}_{G_n}) + \frac{1}{2}\|\mathbf{v}_{G_n} - \boldsymbol{\nu}\|_2^2, \tag{A.40}$$

where $\boldsymbol{\nu} = \text{sgn}(\mathbf{b}_{G_n}) \odot \max(|\mathbf{b}_{G_n}| - \kappa_1 \mathbf{f}_{G_n}, 0)$.

We aim to show $\mathbf{v}_{G_n}^*$ is also the minimizer of problem (A.39). The optimality of $\mathbf{v}_{G_n}^*$ for problem (A.40) yields

$$\mathbf{0} \in \mathbf{v}_{G_n}^* - \boldsymbol{\nu} + \partial h_{\kappa_2}^0(\mathbf{v}_{G_n}^*), \tag{A.41}$$

where $\partial h_{\kappa_2}^0(\cdot)$ denotes the subgradient of $h_{\kappa_2}^0(\cdot)$. We then derive the subgradient of $\phi_{\kappa_2}^{\kappa_1}(\mathbf{v}_{G_n})$ at $\mathbf{v}_{G_n}^*$

$$\partial \phi_{\kappa_2}^{\kappa_1}(\mathbf{v}_{G_n}^*) = \mathbf{v}_{G_n}^* - \mathbf{b}_{G_n} + \kappa_1 \mathbf{F}_n \text{sgn}(\mathbf{v}_{G_n}^*) + \partial h_{\kappa_2}^0(\mathbf{v}_{G_n}^*), \tag{A.42}$$

where $\text{sgn}(\cdot)$ is defined in a component-wise fashion

$$\text{SGN}(x) = \begin{cases} \{1\} & x > 0 \\ [-1, 1] & x = 0 \\ \{-1\} & x < 0, \end{cases}$$

for $\forall x \in \mathbb{R}$.

The definition of $\boldsymbol{\nu} = \text{sgn}(\mathbf{b}_{\text{G}_n}) \odot \max(|\mathbf{b}_{\text{G}_n}| - \kappa_1 \mathbf{f}_{\text{G}_n}, 0)$ implies

$$
\nu_i = \begin{cases}
[\mathbf{b}_{\text{G}_n}]_i - \kappa_1 [\mathbf{f}_{\text{G}_n}]_i & [\mathbf{b}_{\text{G}_n}]_i > \kappa_1 [\mathbf{f}_{\text{G}_n}]_i \\
0 & |[\mathbf{b}_{\text{G}_n}]_i| \leq \kappa_1 [\mathbf{f}_{\text{G}_n}]_i \\
[\mathbf{b}_{\text{G}_n}]_i + \kappa_1 [\mathbf{f}_{\text{G}_n}]_i & [\mathbf{b}_{\text{G}_n}]_i < -\kappa_1 [\mathbf{f}_{\text{G}_n}]_i,
\end{cases}
\tag{A.43}
$$

where $[\mathbf{x}]_i$ denotes the ith entry of a vector $\mathbf{x}$.

From (A.43), we have $\boldsymbol{\nu} \in \mathbf{b}_{\text{G}_n} - \kappa_1 \mathbf{F}_n \text{SGN}(\boldsymbol{\nu})$. Since $\mathbf{v}_{\text{G}_n}^*$ is the minimizer of problem (A.40), according to [91, Lemma 1], we can obtain that $\text{SGN}(\boldsymbol{\nu}) \subseteq \text{SGN}(\mathbf{v}_{\text{G}_n}^*)$. Thus,

$$
\boldsymbol{\nu} \in \mathbf{b}_{\text{G}_n} - \kappa_1 \mathbf{F}_n \text{SGN}(\mathbf{v}_{\text{G}_n}^*).
\tag{A.44}
$$

Combining (A.41) and (A.44), we obtain that

$$
\mathbf{0} \in \mathbf{v}_{\text{G}_n}^* - \mathbf{b}_{\text{G}_n} + \kappa_1 \mathbf{F}_n \text{SGN}(\mathbf{v}_{\text{G}_n}^*) + \partial h_{\kappa_2}^0(\mathbf{v}_{\text{G}_n}^*),
$$

which implies that $\mathbf{0} \in \partial \phi_{\kappa_2}^{\kappa_1}(\mathbf{v}_{\text{G}_n}^*)$ from (A.42). Thus, $\mathbf{v}_{\text{G}_n}^*$ is the minimizer of problem (A.39).

Finally, the closed form of $\mathbf{v}_{\text{G}_n}^*$ in problem (A.40) is given by a block soft thresholding operator [93]

$$
\mathbf{v}_{\text{G}_n}^* = \begin{cases}
(1 - \frac{\kappa_2}{\|\boldsymbol{\nu}\|_2})\boldsymbol{\nu} & \|\boldsymbol{\nu}\|_2 \geq \kappa_2 \\
0 & \|\boldsymbol{\nu}\|_2 < \kappa_2.
\end{cases}
$$

The proof is now complete. ∎

A.18   Proof of Proposition 7.2

We begin by collecting terms in φ associated with $\mathbf{w}$,

$$\varphi_{\mathbf{w}} := \frac{\rho}{2} \sum_{m=1}^{M} \left\| \bar{\mathbf{Q}}_m \mathbf{w} - \boldsymbol{\alpha}_m \right\|_2^2 + \rho \sum_{k=1}^{K} \| \mathbf{w}_k - \boldsymbol{\gamma}_{3,k} \|_2^2 + \frac{\rho}{2} \sum_{k=1}^{K} \| \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \mathbf{H}_k \|_F^2, \quad \text{(A.45)}$$

where $\boldsymbol{\gamma}_{3,k}$ is the $(L+1)$ column of $\boldsymbol{\Upsilon}_{3,k}$ after the last entry is removed, and $\mathbf{H}_k := \mathbf{U}_k - \mathbf{Z}_k + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T + \boldsymbol{\Upsilon}_{4,k}$, which is a symmetric matrix.

In (A.45), we assume an incremental change $\delta \mathbf{w}$ in $\mathbf{w}$. Replacing $\mathbf{w}$ with $\mathbf{w} + \delta \mathbf{w}$ and $\varphi_{\mathbf{w}}$ with $\varphi_{\mathbf{w}} + \delta \varphi_{\mathbf{w}}$ and collecting first order variation terms on both sides of (A.45), we obtain

$$\delta \varphi_{\mathbf{w}} = \rho \sum_{m=1}^{M} (\bar{\mathbf{Q}}_m \mathbf{w} - \boldsymbol{\alpha}_m)^T \bar{\mathbf{Q}}_m \delta \mathbf{w} + 2\rho (\mathbf{w} - \boldsymbol{\gamma}_3)^T \delta \mathbf{w}$$
$$+ 2\rho \hat{\mathbf{w}}^T \text{blkdiag} \{ \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \mathbf{H}_k \} \delta \mathbf{w}, \quad \text{(A.46)}$$

where $\boldsymbol{\gamma}_3 = [\boldsymbol{\gamma}_{3,1}^T, \ldots, \boldsymbol{\gamma}_{3,K}^T]^T$, and $\hat{\mathbf{w}} = [\hat{\mathbf{w}}_1^T, \ldots, \hat{\mathbf{w}}_K^T]^T$. It is clear from (A.46) that the gradient of φ with respect to $\mathbf{w}$ is given by

$$\nabla_{\mathbf{w}} \varphi = \rho \sum_{m=1}^{M} \bar{\mathbf{Q}}_m^T (\bar{\mathbf{Q}}_m \mathbf{w} - \boldsymbol{\alpha}_m) + 2\rho (\mathbf{w} - \boldsymbol{\gamma}_3) + 2\rho \, \text{blkdiag} \{ \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \mathbf{H}_k \} \hat{\mathbf{w}}. \quad \text{(A.47)}$$

Second, we collect the terms associated with $\mathbf{p}$ in φ to construct the function

$$\varphi_{\mathbf{p}} := \frac{\rho}{2} \| \mathbf{C} - \text{diag}(\mathbf{p}) - \boldsymbol{\Upsilon}_1^{11} \|_F^2 + \frac{\rho}{2} \| \mathbf{p} - \boldsymbol{\gamma}_2 \|_2^2, \quad \text{(A.48)}$$

where $\boldsymbol{\Upsilon}_1^{11}$ is a matrix that consists of the first K rows and columns of $\boldsymbol{\Upsilon}_1$, and $\boldsymbol{\gamma}_2$ is a vector whose kth entry is given by the first entry of $\boldsymbol{\Upsilon}_{2,k}$ for $k \in [K]$.

In (A.48), replacing $\mathbf{p}$ with $\mathbf{p} + \delta \mathbf{p}$ and $\varphi_{\mathbf{p}}$ with $\varphi_{\mathbf{p}} + \delta \varphi_{\mathbf{p}}$ and collecting first order variation

terms on both sides, we obtain

$$\delta\varphi_{\mathbf{p}} = \rho[\operatorname{diag}(\mathbf{C} - \operatorname{diag}(\mathbf{p}) - \mathbf{\Upsilon}_1^{11}) + (\mathbf{p} - \boldsymbol{\gamma}_2)]^T \delta\mathbf{p}, \tag{A.49}$$

where $\operatorname{diag}(\cdot)$ returns in vector form the diagonal entries of its matrix argument. Therefore, the gradient of φ with respect to $\mathbf{p}$ is given by

$$\nabla_{\mathbf{p}}\varphi = \rho\operatorname{diag}(\mathbf{C} - \operatorname{diag}(\mathbf{p}) - \mathbf{\Upsilon}_1^{11}) + \rho(\mathbf{p} - \boldsymbol{\gamma}_2). \tag{A.50}$$

Third, given the terms associated with $\mathbf{V}$ in φ, the gradient of φ with respect to $\mathbf{V}$ is readily cast as

$$\nabla_{\mathbf{V}}\varphi = \mathbf{I} + \rho(\mathbf{V} - \mathbf{\Upsilon}_1^{22}), \tag{A.51}$$

where $\mathbf{\Upsilon}_1^{22}$ is a submatrix of $\mathbf{\Upsilon}_1$ after the first K rows and columns are removed.

Further, we collect the terms in φ with respect to the variable $\mathbf{U}_k$, and consider the function

$$\varphi_{\mathbf{U}_k} := \frac{\rho}{2}\left\|\mathbf{I} + \sigma_\epsilon^2\sigma_\varsigma^{-2}\mathbf{G}_k^T\mathbf{U}_k\mathbf{G}_k - \mathbf{\Upsilon}_{2,k}^{22}\right\|_F^2 + \frac{\rho}{2}\|\mathbf{U}_k - \mathbf{\Upsilon}_{3,k}^{11}\|_F^2 + \frac{\rho}{2}\|\mathbf{U}_k - \mathbf{Z}_k - \mathbf{T}_k\|_F^2, \tag{A.52}$$

where $\mathbf{\Upsilon}_{2,k}^{22}$ is a submatrix of $\mathbf{\Upsilon}_{2,k}$ after the first row and column are removed, $\mathbf{\Upsilon}_{3,k}^{11}$ is a submatrix of $\mathbf{\Upsilon}_{3,k}$ after the last row and column are removed, and $\mathbf{T}_k := \hat{\mathbf{w}}_k\mathbf{w}_k^T + \mathbf{w}_k\hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k\hat{\mathbf{w}}_k^T - \mathbf{\Upsilon}_{4,k}$.

In (A.52), replacing $\mathbf{U}_k$ with $\mathbf{U}_k + \delta\mathbf{U}_k$ and $\varphi_{\mathbf{U}_k}$ with $\varphi_{\mathbf{U}_k} + \delta\varphi_{\mathbf{U}_k}$ and collecting first order variation terms on both sides, we obtain

$$\delta\varphi_{\mathbf{U}_k} = \frac{\rho\sigma_\epsilon^2}{\sigma_\varsigma^2}\operatorname{tr}\left(\mathbf{G}_k(\mathbf{I} + \frac{\sigma_\epsilon^2}{\sigma_\varsigma^2}\mathbf{G}_k^T\mathbf{U}_k\mathbf{G}_k - \mathbf{\Upsilon}_{2,k}^{22})^T\mathbf{G}_k^T\delta\mathbf{U}_k\right)$$
$$+ \rho\operatorname{tr}\left((\mathbf{U}_k - \mathbf{\Upsilon}_{3,k}^{11})^T\delta\mathbf{U}_k + (\mathbf{U}_k - \mathbf{Z}_k - \mathbf{T}_k)^T\delta\mathbf{U}_k\right).$$

Therefore, the gradient of φ with respect to $\mathbf{U}_k$ is given by

$$\nabla_{\mathbf{U}_k}\varphi = \rho\sigma_\epsilon^2\sigma_\varsigma^{-2}\mathbf{G}_k(\mathbf{I} + \sigma_\epsilon^2\sigma_\varsigma^{-2}\mathbf{G}_k^T\mathbf{U}_k\mathbf{G}_k - \mathbf{\Upsilon}_{2,k}^{22})\mathbf{G}_k^T + \rho(\mathbf{U}_k - \mathbf{\Upsilon}_{3,k}^{11}) + \rho(\mathbf{U}_k - \mathbf{Z}_k - \mathbf{T}_k).$$

$$(\text{A.53})$$

Finally, the gradient of φ with respect to $\mathbf{Z}_k$ is given by

$$\nabla_{\mathbf{Z}_k}\varphi = \tau\mathbf{I} + \rho(\mathbf{Z}_k - \mathbf{U}_k + \mathbf{T}_k), \qquad (\text{A.54})$$

where $\mathbf{T}_k$ is defined in (A.52). We now complete the proof by combining (A.47), (A.50), (A.51), (A.53) and (A.54). ∎

A.19   Estimation distortion of BLUE

According to (8.1), (8.2) and (8.3), the received signals at the FC over T time steps can be written as the succinct form

$$\mathbf{y} = \mathbf{G}_w \mathbf{D}_h \boldsymbol{\theta} + \boldsymbol{v}, \tag{A.55}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_T]^T$, $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_T]^T$, $\boldsymbol{v} = [v_1, v_2, \ldots, v_T]^T$, $v_t = \mathbf{g}_t^T \mathbf{W}_t \boldsymbol{\epsilon}_t + \varsigma_t$, $\mathbf{D}_h = \mathrm{blkdiag}\{\mathbf{h}_t\}_{t=1}^T$, $\mathbf{G}_w = \mathrm{blkdiag}\{\mathbf{g}_t^T \mathbf{W}_t\}_{t=1}^T$, and $\mathrm{blkdiag}\{\mathbf{A}_i\}_{i=1}^n$ denotes the block-diagonal matrix with diagonal blocks $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$.

The estimation distortion resulting from BLUE is given by [58, Theorem 6.1]

$$f(\{\mathbf{W}_t\}) := \mathrm{tr}\left(\mathbf{D}_h^T \mathbf{G}_w^T (\sigma_\epsilon^2 \mathbf{G}_w \mathbf{G}_w^T + \sigma_\varsigma^2 \mathbf{I})^{-1} \mathbf{G}_w \mathbf{D}_h\right)^{-1}$$
$$= \sum_{t=1}^T \frac{\sigma_\varsigma^2 + \sigma_\epsilon^2\, \mathrm{tr}\left(\mathbf{W}_t^T \mathbf{g}_t \mathbf{g}_t^T \mathbf{W}_t\right)}{\mathrm{tr}\left(\mathbf{W}_t^T \mathbf{g}_t \mathbf{g}_t^T \mathbf{W}_t \mathbf{h}_t \mathbf{h}_t^T\right)},$$

where $\mathbf{D}_v = \mathbb{E}[\boldsymbol{v}\boldsymbol{v}^T] = \sigma_\epsilon^2 \mathbf{G}_w \mathbf{G}_w^T + \sigma_\varsigma^2 \mathbf{I}$. ∎

A.20   Proof of Proposition 8.1

We introduce a new vector of optimization variables $\mathbf{u} = [u_1, \ldots, u_T]^T$, and express problem (8.18) as

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{1}^T \mathbf{u} \\
\text{subject to} \quad & \frac{\sigma_\varsigma^2 + \sigma_\epsilon^2 \mathbf{w}_t^T \mathbf{\Omega}_{\text{PN},t} \mathbf{w}_t}{\mathbf{w}_t^T \mathbf{\Omega}_{\text{PD},t} \mathbf{w}_t} \leq u_t, \ t \in [T], \\
& \text{constraints of (8.18)},
\end{aligned}
\tag{A.56}
$$

where $\mathbf{u} \in \mathbb{R}^T$, $\mathbf{w}_t \in \mathbb{R}^L$ and $\mathbf{s}_t \in \mathbb{R}^N$ are optimization variables for $t \in [T]$.

By introducing additional variables p_t and q_t for $t \in [T]$, problem (A.56) can be then rewritten as

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{1}^T \mathbf{u} \\
\text{subject to} \quad & \frac{p_t}{q_t} \leq u_t, \ t \in [T], \\
& \sigma_\varsigma^2 + \sigma_\epsilon^2 \mathbf{w}_t^T \mathbf{\Omega}_{\text{PN},t} \mathbf{w}_t \leq p_t, \ t \in [T], \\
& \mathbf{w}_t^T \mathbf{\Omega}_{\text{PD},t} \mathbf{w}_t \geq q_t, \ q_t > 0, \ t \in [T], \\
& \text{constraints of (8.18)},
\end{aligned}
\tag{A.57}
$$

where $\mathbf{u} \in \mathbb{R}^T$, $\mathbf{p} = [p_1, \ldots, p_T]^T \in \mathbb{R}^T$, $\mathbf{q} = [q_1, \ldots, q_T]^T \in \mathbb{R}^T$, $\mathbf{w}_t \in \mathbb{R}^L$ and $\mathbf{s}_t \in \mathbb{R}^N$ are optimization variables for $t \in [T]$. Note that the ratio $(p_t/q_t) \leq u_t$ can be reformulated as a quadratic constraint given $q_t > 0$,

$$
2p_t \leq (q_t + u_t)^2 - (q_t^2 + u_t^2).
\tag{A.58}
$$

Substituting (A.58) into (A.57), we eventually reach the optimization problem (8.19). ∎

References

[1] L. Oliveira and J. Rodrigues, "Wireless sensor networks: a survey on environmental monitoring," *Journal of Communications*, vol. 6, no. 2, 2011. 1

[2] E. D. Zubiete, L. F. Luque, A. V. M. Rodríguez, and I. G. Gonzalez, "Review of wireless sensors networks in health applications," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 1789–1793. 1

[3] G. B. Giannakis, V. Kekatos, N. Gatsis, S. J. Kim, H. Zhu, and B. F. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 107–128, Sept. 2013. 1

[4] T. He, P. Vicaire, T. Yan, L. Luo, L. Gu, G. Zhou, S. Stoleru, Q. Cao, J. A. Stankovic, and T. Abdelzaher, "Achieving real-time target tracking using wireless sensor networks," in *Proceedings of IEEE Real Time Technology and Applications Symposium*, 2006, pp. 37–48. 1

[5] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, Mar. 2002. 2

[6] E. Masazade, R. Niu, and P. K. Varshney, "Dynamic bit allocation for object tracking in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5048–5063, Oct. 2012. 2

[7] H. Zhang, J. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2383 –2395, June 2009. 2, 66

[8] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, no. 10, pp. 2482–2493, 2012. 2, 3

[9] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2649–2661, July 2008. 2, 15, 16, 90

[10] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, vol. 10, no. 3, pp. 273–304, 1995. 2

[11] F. Lin, M. Fardad, and M. R. Jovanović, "Algorithms for leader selection in stochastically forced consensus networks," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1789–1802, July 2014. 2, 32

[12] E. Ertin, J. W. Fisher, and L. C. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *Proceedings of the 2nd international conference on Information processing in sensor networks*, 2003, pp. 405–416. 2

[13] H. Wang, G. Pottie, K. Yao, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proceedings of the 3rd international conference on Information processing in sensor networks*, 2004, pp. 36–45. 2

[14] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "Sensor scheduling algorithms requiring limited computation," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, vol. 3, pp. 825–828. 2

[15] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 451 –462, Feb. 2009. 2, 28, 41

[16] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 684–698, Feb. 2015. 2, 15, 28, 76

[17] A. S. Chhetri, D. Morrell, and A. Papandreou-Suppappola, "Nonmyopic sensor scheduling and its efficient implementation for target tracking applications," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–18, 2006. 3

[18] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, 2011. 3, 13, 40, 44, 66

[19] X. Shen and P. K. Varshney, "Sensor selection based on generalized information gain for target tracking in large sensor networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 363–375, Jan 2014. 3, 13, 24, 29, 33, 34, 35, 36

[20] E. Masazade, R. Niu, and P. K. Varshney, "An approximate dynamic programming based non-myopic sensor selection method for target tracking," in *Proceedings of the 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012, pp. 1–6. 3

[21] L. Shi, P. Cheng, and J. Chen, "Optimal periodic sensor scheduling with limited resources," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2190–2195, Sept. 2011. 3, 14, 51, 71

[22] T. H. McLoughlin and M. Campbell, "Solutions to periodic sensor scheduling problems for formation flying missions in deep space," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 1351 –1368, april 2011. 3, 14, 35, 51, 65

[23] P. Hovareshti, V. Gupta, and J.S. Baras, "Sensor scheduling using smart sensors," in *Proceedings of the 46th IEEE Conference on Decision and Control*, Dec. 2007, pp. 494 –499. 3

[24] W. Zhang, M.P. Vitus, J. Hu, A. Abate, and C.J. Tomlin, "On the optimal solutions of the infinite-horizon linear sensor scheduling problem," in *Proceedings of the 49th IEEE Conference on Decision and Control*, Dec. 2010, pp. 396 –401. 3, 49, 51, 53, 63

[25] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, 2007. 4

[26] J. Matamoros and C. Anton-Haro, "Scaling law of an opportunistic power allocation scheme for amplify-and-forward wireless sensor networks," *IEEE Communications Letters*, vol. 15, no. 2, pp. 169–171, February 2011. 4

[27] F. Jiang, J. Chen, and A. L. Swindlehurst, "Optimal power allocation for parameter tracking in a distributed amplify-and-forward sensor network," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2200–2211, May 2014. 4

[28] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003. 4

[29] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part i: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, 2006. 4

[30] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 757–770, 2008. 4, 92, 150

[31] Z.-Q. Luo, G. B. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," in *Proceedings of IEEE International Symposium on Information Theor (ISIT)*, Sept 2005, pp. 1441–1445. 4

[32] A. S. Leong, S. Dey, and J. S. Evans, "Asymptotics and power allocation for state estimation over fading channels," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 611–633, January 2011. 4

[33] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3822–3832, 2009. 4, 15, 16, 90

[34] H. Jamali-Rad, A. Simonetto, G. Leus, and X. Ma, "Sparsity-aware sensor selection for correlated noise," in *Proceedings of 17th International Conference on Information Fusion (FUSION)*, July 2014, pp. 1–7. 13, 24, 29, 33, 34

[35] E. Rigtorp, "Sensor selection with correlated noise," M.S. thesis, KTH Royal Institute of Technology, Aug. 2010. 13, 24, 29, 33, 34

[36] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering Tracking*, Wiley-IEEE press, 2007. 13, 38

[37] J. L. Ny, E. Feron, and M. A. Dahleh, "Scheduling continuous-time Kalman filters," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1381–1394, 2011. 14, 68, 69, 70

[38] R. Niu, P. K. Varshney, K. Mehrotra, and C. Mohan, "Temporally staggered sensors in multi-sensor target tracking systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 3, pp. 794 –808, July 2005. 14, 71

[39] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 845–848, Dec. 2012. 14, 15, 64

[40] S. Liu, M. Fardad, E. Masazade, and P. K. Varshney, "Sparsity-aware field estimation via ordinary Kriging," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3976–3980. 15, 76

[41] E. Ertin, J. W. Fisher, and L. C. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *Proceedings of the 2nd International Conference on Information Processing in Sensor Networks*, 2003, pp. 405–416. 15, 74

[42] S. Liu, E. Masazade, and P. K. Varshney, "Temporally staggered sensing for field estimation with quantized data in wireless sensor networks," in *IEEE Statistical Signal Processing Workshop (SSP)*, Aug. 2012, pp. 1–4. 15, 19, 74

[43] Y. Chen and Q. Zhao, "On the lifetime of wireless sensor networks," *IEEE Communications Letters*, vol. 9, no. 11, pp. 976–978, Nov 2005. 15, 74

[44] S. Kar and P. K. Varshney, "On linear coherent estimation with spatial collaboration," in *Proceedings of the 2012 IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012, pp. 1448–1452. 15, 16, 90, 95

[45] S. Kar and P. K. Varshney, "Controlled collaboration for linear coherent estimation in wireless sensor networks," in *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 334–341. 15, 16, 90, 95, 141

[46] M. Fanaei, M. C. Valenti, A. Jamalipour, and N. A. Schmid, "Optimal power allocation for distributed blue estimation with linear spatial collaboration," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5452–5456. 15, 16, 90, 151

[47] G. Thatte and U. Mitra, "Power allocation in linear and tree wsn topologies," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Oct 2006, pp. 1342–1346. 15, 16

[48] S. Kar and P. K. Varshney, "Linear coherent estimation with spatial collaboration," *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3532–3553, June 2013. 16, 95, 102, 103, 105, 111, 124, 125, 142, 183, 184

[49] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Computer Networks*, vol. 45, no. 3, pp. 245–259, June 2004. 16, 118

[50] Phaeton C. Kyriakidis, "A spatial time series framework for modeling daily precipitation at regional scales," *Journal of Hydrology*, vol. 297, no. 4, pp. 236 – 255, Apr. 2004. 16, 118

[51] Y. Zhao, B. Chen, and R. Zhang, "Optimal power allocation for an energy harvesting estimation system," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 4549–4553. 17, 148

[52] C. Huang, Y. Zhou, T. Jiang, P. Zhang, and S. Cui, "Power allocation for joint estimation with energy harvesting constraints," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 4804–4808. 17, 148, 159

[53] M. Nourian, S. Dey, and A. Ahlen, "Distortion minimization in multi-sensor estimation with energy harvesting," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 524–539, Mar. 2015. 17, 148

[54] M. Calvo-Fullana, J. Matamoros, and C. Antón-Haro, "Sensor selection in energy harvesting wireless sensor networks," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2015, pp. 43–47. 17

[55] X. Shen, S. Liu, and P. K. Varshney, "Sensor selection for nonlinear systems in large sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2664–2678, October 2014. 19, 35

[56] S. Liu, E. Masazade, M. Fardad, and P. K. Varshney, "Sensor selection with correlated measurements for target tracking in wireless sensor networks," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 4030–4034. 19, 36

[57] S. Liu, E. Masazade, X. Shen, and P. K. Varshney, "Adaptive non-myopic quantizer design for target tracking in wireless sensor networks," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 1085–1089. 19

[58] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, 1993. 20, 41, 93, 121, 151, 196

[59] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004. 21, 59, 85, 100, 101, 131, 137, 156, 180

[60] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011. 22, 23, 102, 134, 136, 141

[61] S. P. Chepuri and G. Leus, "Sparse sensing for distributed gaussian detection," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2394–2398. 27

[62] A. Nemirovski, "Interior point polynomial time methods in convex programming," 2012 [Online], Available: *http://www2.isye.gatech.edu/ nemirovs/Lect_IPM.pdf*. 31, 65, 82, 84

[63] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010. 31, 42, 133

[64] A. Bertrand, J. Szurley, P. Ruckebusch, I. Moerman, and M. Moonen, "Efficient calculation of sensor utility and sensor removal in wireless sensor networks for adaptive signal estimation and beamforming," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5857–5869, Nov. 2012. 32

[65] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. of 49th IEEE Conference on Decision and Control (CDC)*, Dec. 2010, pp. 2572–2577. 32

[66] V. V. Williams, "Multiplying matrices faster than Coppersmith-Winograd," in *Proc. the 44th symposium on Theory of Computing*, 2012, pp. 887–898. 32, 40

[67] F. E. Udwadia, "Methodology for optimum sensor locations for parameter identification in dynamic systems," *Journal of Engineering Mechanics*, vol. 120, no. 2, pp. 368–390, Feb. 1994. 35

[68] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3292–3297, Aug. 2009. 35

[69] P. M. Pardalos, "Global optimization algorithms for linearly constrained indefinite quadratic problems," *Computers & Mathematics with Applications*, vol. 21, no. 6âĂŞ7, pp. 87 – 97, 1991. 36

[70] E. Masazade, R. Niu, and P. K. Varshney, "Dynamic bit allocation for object tracking in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5048–5063, Oct. 2012. 37

[71] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 845–848, Dec 2012. 37, 38, 43

[72] P. Tichavsky, C.H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processin*, vol. 46, no. 5, pp. 1386–1396, may 1998. 38

[73] S. P. Chepuri and G. Leus, "Sparsity-promoting adaptive sensor selection for non-linear filtering," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5080–5084. 38

[74] S. Liu, E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-aware field estimation via ordinary Kriging," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3976–3980. 41

[75] A. d'Aspremont and S. Boyd, "Relaxations and randomized methods for non-convex qcqps," Stanford, CA: Stanford Univ., Autumn 2003. [Online]. Available: *http://web.stanford.edu/class/ee392o/relaxations.pdf*. 42

[76] F. Peng and B. Chen, "Decentralized estimation with correlated additive noise: Does dependency always imply redundancy?," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 677–681. 43

[77] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4519–4528, Dec. 2006. 44

[78] B. M. Yu, K. V. Shenoy, and M. Sahani, "Derivation of extended kalman filtering and smoothing equations," 2004. 44

[79] E. Feron and C. Olivier, "Targets, sensors and infinite-horizon tracking optimality," in *Proceedings of the 29th IEEE Conference on Decision and Control*, 1990, vol. 4, pp. 2291–2292. 51

[80] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME - Journal of Basic Engineering*, , no. 82, pp. 35–45, 1960. 52, 64

[81] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2426–2431, 2013. 54, 56, 57, 58, 59, 60, 63, 64

[82] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011. 56, 57, 60, 63, 64, 80

[83] F. Lin, M. Fardad, and M. R. Jovanović, "Augmented Lagrangian approach to design of structured optimal state feedback gains," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, pp. 2923–2929, Dec. 2011. 58, 59, 64

[84] T. Chen and B. Francis, *Optimal Sampled-Data Control Systems*, Springer-Verlag, 1995. 58

[85] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *http://arxiv.org/abs/1111.6188v1*, 2011. 59

[86] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999. 59

[87] Fu Lin, M. Fardad, and M.R. Jovanović, "Algorithms for leader selection in large dynamical networks: Noise-corrupted leaders," in *Proceedings of the 50th IEEE Conference on Decision and Control*, Dec. 2011, pp. 2932 –2937. 64, 174

[88] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX + XB = C$," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, 1972. 64

[89] S. Liu, M. Fardad, E. Masazade, and P. K. Varshney, "On optimal periodic sensor scheduling for field estimation in wireless sensor networks," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 137–140. 76

[90] S. Liu, F. Chen, A. Vempaty, M. Fardad, L. Shen, and P. K. Varshney, "Sparsity-promoting sensor management for estimation: An energy balance point of view," in *Proc. 18th International Conference on Information Fusion (Fusion)*, July 2015, pp. 231–238. 76

[91] L. Yuan, J. Liu, and J. Ye, "Efficient methods for overlapping group lasso," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2104–2116, Sept 2013. 78, 192

[92] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006. 79, 95

[93] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013. 81, 85, 134, 140, 192

[94] Inc. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," *http://cvxr.com/cvx*, Aug 2012. 84, 143, 145

[95] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery problems," in *Convex Optimization in Signal Processing and Communications,* (D. Palomar and Y. Eldar, eds.), pp. 1–49, Cambridge University Press, 2010. 85

[96] A. J. Stothers, *On the Complexity of Matrix Multiplication*, Ph.D. Thesis, U. Edinburgh, 2010. 85

[97] Emmanuel Candes, Michael Wakin, and Stephen Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008. 89

[98] A. Beck, "Quadratic matrix programming," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 1224–1238, 2007. 95, 123

[99] L. Liu, X. Zhang, and H. Ma, "Dynamic node collaboration for mobile target tracking in wireless camera sensor networks," in *Proceedings of IEEE INFOCOM 2009*, April 2009, pp. 1188–1196. 97

[100] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008. 98, 102, 106

[101] V. Y. Pan and Z. Q. Chen, "The complexity of the matrix eigenproblem," in *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, 1999, pp. 507–516. 103

[102] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003. 106, 107, 129

[103] T. Lipp and S. Boyd, "Variations and extensions of the convex-concave procedure," `http://web.stanford.edu/~boyd/papers/pdf/cvx_ccv.pdf`, 2014. 107, 129, 133, 157

[104] A. Nemirovski, "Interior point polynomial time methods in convex programming," 2012 [Online], Available: `http://www2.isye.gatech.edu/~nemirovs/Lect_IPM.pdf`. 108, 130, 134, 158

[105] A. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, NY, USA, 2005. 111

[106] F. Bugarin, D. Henrion, and J. B. Lasserre, "Minimizing the sum of many rational functions," *Mathematical Programming Computation*, pp. 1–29, 2015. 125

[107] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Operator splitting for conic optimization via homogeneous self-dual embedding," Arxiv preprint `http://arxiv.org/abs/1312.3039`, 2013. 134

[108] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4729–4743, Sept. 2015. 134

[109] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4808–4818, Sept. 2012. 152

[110] L. Huang, J. Walrand, and K. Ramchandran, "Optimal demand response with energy storage management," in *Proc. IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, Nov. 2012, pp. 61–66. 153

[111] Y. Wang, X. Lin, and M. Pedram, "A near-optimal model-based control algorithm for households equipped with residential photovoltaic power generation and energy storage systems," *IEEE Transactions on Sustainable Energy*, vol. PP, no. 99, pp. 1–10, 2015. 153

[112] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, Jan. 2012. 156

[113] H. Konno, "Maximization of a convex quadratic function under linear constraints," *Mathematical Programming*, vol. 11, no. 1, pp. 117–127, 1976. 170

[114] G. H. Golub, "Some modified matrix eigenvalue problems," *SIAM Review*, vol. 15, no. 2, pp. 318–334, 1973. 188

VITA

NAME OF AUTHOR: Sijia Liu

PLACE OF BIRTH: Zhengzhou, Henan, China

DATE OF BIRTH: Dec. 19, 1985

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Xi'an Jiaotong University, Xi'an, China

DEGREES AWARDED:

B. S., 2008, Xi'an Jiaotong University, Xi'an, China

M. S., 2011, Xi'an Jiaotong University, Xi'an, China