

Optimized Sensor Collaboration for Estimation of Temporally Correlated Parameters

Sijia Liu, *Member, IEEE*, Swarnendu Kar, *Member, IEEE*, Makan Fardad, *Member, IEEE*,
and Pramod K. Varshney, *Fellow, IEEE*

Abstract—In this paper, we aim to design the optimal sensor collaboration strategy for the estimation of time-varying parameters, where collaboration refers to the act of sharing measurements with neighboring sensors prior to transmission to a fusion center. We begin by addressing the sensor collaboration problem for the estimation of uncorrelated parameters. We show that the resulting collaboration problem can be transformed into a special nonconvex optimization problem, where a difference of convex functions carries all the nonconvexity. This specific problem structure enables the use of a convex–concave procedure to obtain a near-optimal solution. When the parameters of interest are temporally correlated, a penalized version of the convex–concave procedure becomes well suited for designing the optimal collaboration scheme. In order to improve computational efficiency, we further propose a fast algorithm that scales gracefully with problem size via the alternating direction method of multipliers. Numerical results are provided to demonstrate the effectiveness of our approach and the impact of parameter correlation and temporal dynamics of sensor networks on estimation performance.

Index Terms—ADMM, convex–concave procedure, distributed estimation, semidefinite programming, sensor collaboration, wireless sensor networks.

I. INTRODUCTION

WIRELESS sensor networks (WSNs) consist of a large number of spatially distributed sensors that often cooperate to perform parameter estimation; example applications include environment monitoring, source localization and target tracking [1]–[3]. Under limited resources, such as limited communication bandwidth and sensor battery power, it is important to design an energy-efficient architecture for distributed estimation. In this paper, we employ a WSN to estimate time-varying parameters in the presence of inter-sensor communication that is referred to as sensor collaboration. Here sensors

Manuscript received March 10, 2016; revised July 13, 2016; accepted August 20, 2016. Date of publication September 21, 2016; date of current version October 19, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ashish Pandharipande. The work of S. Liu and P. K. Varshney was supported by the U.S. Air Force Office of Scientific Research under Grant FA9550-10-1-0458. The work of M. Fardad was supported by the National Science Foundation under awards EAGER ECCS-1545270 and CNS-1329885.

S. Liu was with the Department of Electrical Engineering & Computer Science, Syracuse University, Syracuse, NY 13244 USA. He is now with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: lsjxjtu@umich.edu).

S. Kar is with the New Devices Group, Intel Corporation, Hillsboro, OR 97124 USA (e-mail: swarnendu.kar@intel.com).

M. Fardad and P. K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: makan@syr.edu; varshney@syr.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2016.2612173

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 (FC). The presence of sensor collaboration smooths out the observation noise, thereby improving the quality of the signal and the eventual estimation performance.

Early research efforts [4]–[8] focused on the problem of distributed inference (estimation or detection) in the absence of sensor collaboration, where an amplify-and-forward transmission strategy is commonly used. In [4], the problem of designing optimal power amplifying factors (also known as power allocation problem) was studied for distributed estimation over an orthogonal multiple access channel (MAC). In [5], the power allocation problem was addressed when the MAC is coherent, where sensors coherently form a beam into a common channel received at the FC. In [6], a likelihood-based multiple access communication strategy was proposed for estimation, and was proved to be asymptotically efficient as the number of sensors increases. In [7], distributed detection problem was studied in the setting of identical Gaussian multiple access channels. It was shown that the centralized error exponent can be achieved via the transmission of the log-likelihood ratio as the number of sensors approaches infinity. In [8], the problem of power allocation for distributed detection was studied using a MAC. In the aforementioned literature [4]–[8], the act of inter-sensor communication was not considered. In contrast, here we seek the optimal sensor collaboration scheme for the estimation of temporally correlated parameters.

Recently, the problem of distributed estimation with sensor collaboration has attracted attention [9]–[17]. In [9], the optimal power allocation strategy was found for a fully connected network, where all the sensors are allowed to collaborate, namely, share their measurements with the other sensors. It was shown that sensor collaboration results in significant improvement of estimation performance compared with the conventional amplify-and-forward transmission scheme. In [10] and [11], optimal power allocation schemes were found for star, branch and linear network topologies. In [12], the sensor collaboration problem was studied for parameter estimation via the best linear unbiased estimator. In [13]–[15], the problem of sensor collaboration was studied given an arbitrary collaboration topology. It was observed that even a partially connected network can yield performance close to that of a fully connected network. In [16] and [17], nonzero collaboration costs were taken into account, and a sparsity inducing optimization framework was proposed to jointly design both sensor selection and sensor collaboration schemes.

In the existing literature [9]–[17], 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. In contrast, here we study the problem of sensor collaboration for the estimation of *temporally-correlated* parameters in *dynamic* networks that involve, for example, time-varying observation and channel gains. Solving such a problem is also motivated by real-life applications, in which the physical phenomenon to be monitored such as daily temperature, precipitation, soil moisture and seismic activities [18]–[20] is temporally correlated. For example, when monitoring daily temperature variations, temperatures at different times of the day are strongly correlated, e.g., a cold morning is likely to be followed by a cold afternoon.

Due to the presence of temporal dynamics and parameter correlation, optimal sensor collaboration schemes at multiple time steps are coupled with each other, and thus pose many challenges in problem formulation and optimization compared to the existing work [9]–[17]. For example, when parameters of interest are temporally correlated, expressing the estimation distortion in a succinct closed form (with respect to the collaboration variables) is not straightforward. 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. In this paper, we seek the optimal sensor collaboration scheme by minimizing the estimation distortion subject to individual energy constraints of sensors in the presence of (a) temporal dynamics in system, (b) temporal correlation of parameter, and (c) energy constraints in time.

Besides [9]–[17], our work is also related to but quite different from the problem of consensus-based decentralized estimation [21]–[25]. The common idea in [21]–[25] is that the task of centralized estimation can be performed using local estimators at sensors together with inter-sensor communications. It was shown in [24] and [25] that the success of decentralized estimation is based on the fact that the global estimation cost with respect to the parameter of interest can be converted into a sum of local cost functions subject to consensus constraints. Different from [21]–[25], the focus of this paper is to design the optimal energy allocation strategy (namely, the collaboration weights), rather than to find the optimal estimate. Here tasks of estimation and optimization are completed at an FC. Moreover, the studied sensor network is not necessarily connected. An extreme case is that in the absence of inter-sensor communication, the proposed sensor collaboration problem would reduce to the conventional power allocation problem (based on the amplify-and-forward transmission strategy) [4], [5]. Therefore, our problem is different from the consensus-based decentralized estimation problem, in which the network is assumed to be connected so that the consensus of estimate at local sensors can be achieved.

In our work, design of the optimal collaboration scheme is studied under two scenarios: a) parameters are temporally uncorrelated or prior knowledge about temporal correlation is not available, and b) parameters are temporally correlated. When parameters are uncorrelated, we derive the closed form of the estimation distortion with respect to sensor collaboration variables, which is in the form of a sum of quadratic ratios. We show that the resulting sensor collaboration problem is equivalent to a nonconvex quadratically constrained problem, in which the difference of convex functions carries all the nonconvexity. This specific problem structure enables the use of convex-concave

procedure (CCP) [26] to solve the sensor collaboration problem in a numerically efficient manner.

When parameters of interest are temporally correlated, expressing the estimation error as an explicit function of the collaboration variables becomes difficult. In this case, we show that the sensor collaboration problem can be converted into a semidefinite program together with a (nonconvex) rank-one constraint. After convexification, the method of penalty CCP [27] becomes well-suited for seeking the optimal sensor collaboration scheme. However, the proposed algorithm is computationally intensive for large-scale problems. To improve computational efficiency, we develop a fast algorithm that scales gracefully with problem size by using the alternating direction method of multipliers (ADMM) [28].

We summarize our contributions as follows.

- We propose a tractable optimization framework for the design of the optimal collaboration scheme that accounts for parameter correlation and temporal dynamics of sensor networks.
- We show that the problem of sensor collaboration for the estimation of temporally uncorrelated parameters 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.
- We provide valuable insights into the problem structure of sensor collaboration with correlated parameters, and propose an ADMM-based algorithm for improving the computational efficiency.

The rest of the paper is organized as follows. In Section II, we introduce the collaborative estimation system, and present the general formulation of the optimal sensor collaboration problem. In Section III, we discuss two types of sensor collaboration problems for the estimation of temporally uncorrelated and correlated parameters. In Section IV, we study the sensor collaboration problem with uncorrelated parameters. In Section V, we propose efficient optimization methods to solve the sensor collaboration problem with correlated parameters. In Section VI, we demonstrate the effectiveness of our approach through numerical examples. Finally, in Section VII we summarize our work and discuss future research directions.

II. SYSTEM MODEL

In this section, we introduce the collaborative estimation system and formulate the sensor collaboration problem considered in this work. The task here is to estimate a time-varying parameter θ_k over a time horizon of length K . In the estimation system, sensors first acquire 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 overall architecture of the collaborative estimation system is shown in Fig. 1.

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

$$\mathbf{x}_k = \mathbf{h}_k \theta_k + \boldsymbol{\epsilon}_k, \quad k \in [K], \quad (1)$$

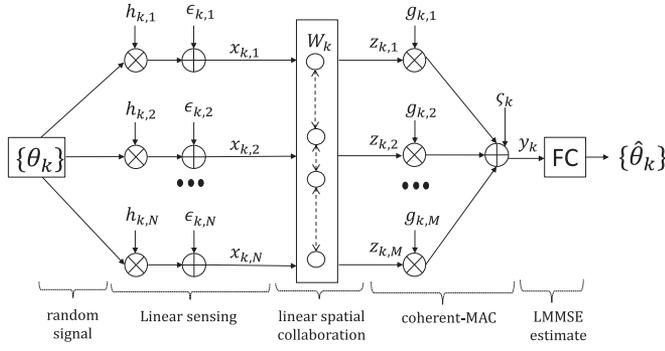


Fig. 1. Collaborative estimation architecture.

where for notational simplicity, let $[K]$ denote the integer set $\{1, 2, \dots, K\}$, $\mathbf{x}_k = [x_{k,1}, \dots, x_{k,N}]^T$ is the vector of measurements, $\mathbf{h}_k = [h_{k,1}, \dots, h_{k,N}]^T$ is the vector of observation gains, without loss of generality θ_k is assumed to be a random process with zero mean and variance σ_θ^2 , $\boldsymbol{\epsilon}_k = [\epsilon_{k,1}, \dots, \epsilon_{k,N}]^T$ 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 other sensors 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 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 we call \mathbf{A} a topology matrix, where $A_{mn} = 1$ signifies that the n th sensor shares its observation with the m th sensor, and $A_{mn} = 0$ indicates the absence of a collaboration link from the n th sensor to the m th sensor. Note that \mathbf{A} is essentially a truncated adjacency matrix. The bidirectional communication link between two sensors indicates that the underlying graph of the network is directed but not necessarily connected. In particular, the network given by $A_{mn} = 0$ for $n \neq m$ corresponds to the amplify-and-forward transmission strategy considered in [4].

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

$$\begin{aligned} \mathbf{z}_k &= \mathbf{W}_k \mathbf{x}_k, \quad k \in [K] \\ \mathbf{W}_k \circ (\mathbf{1}_M \mathbf{1}_N^T - \mathbf{A}) &= \mathbf{0}, \end{aligned} \quad (2)$$

where $\mathbf{z}_k = [z_{k,1}, z_{k,2}, \dots, z_{k,M}]^T$, $z_{k,m}$ 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 referring to vectors of all ones and all zeros, their dimensions will be omitted for simplicity but can be inferred from the context. In (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 [5]

$$y_k = \mathbf{g}_k^T \mathbf{z}_k + \varsigma_k, \quad k \in [K], \quad (3)$$

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

From (1)–(3), the vector of received signals at the FC can be compactly expressed as a linear function of parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_K]^T$,

$$\mathbf{y} = \mathbf{D}_W \mathbf{D}_h \boldsymbol{\theta} + \boldsymbol{\nu}, \quad \mathbf{D}_W := \text{blkdiag} \{ \mathbf{g}_k^T \mathbf{W}_k \}_{k=1}^K, \quad (4)$$

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

At the FC, we employ a linear minimum mean squared-error estimator (LMMSE) [29] 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 estimation error covariance is given by [29, Theorem 10.3]

$$\mathbf{P}_W = (\boldsymbol{\Sigma}_\theta^{-1} + \mathbf{D}_h^T \mathbf{D}_W^T \mathbf{D}_\nu^{-1} \mathbf{D}_W \mathbf{D}_h)^{-1}, \quad (5)$$

where $\boldsymbol{\Sigma}_\theta$ represents prior knowledge about the parameter correlation, particularly $\boldsymbol{\Sigma}_\theta = \sigma_\theta^2 \mathbf{I}_K$ for temporally uncorrelated parameters, \mathbf{I}_K is the $K \times K$ identity matrix, and $\mathbf{D}_\nu := \sigma_\epsilon^2 \mathbf{D}_W \mathbf{D}_W^T + \sigma_\varsigma^2 \mathbf{I}_K$. It is clear from (5) that the estimation error covariance matrix is a function 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 will be provided in Section III.

We next define the transmission cost of the m th sensor at time k , which refers to the energy consumption of transmitting the collaborative message $z_{k,m}$ to the FC. That is,

$$\begin{aligned} T_m(\mathbf{W}_k) &= \mathbb{E}_{\theta_k, \boldsymbol{\epsilon}_k} [z_{k,m}^2] \\ &= \mathbf{e}_m^T \mathbf{W}_k (\sigma_\theta^2 \mathbf{h}_k \mathbf{h}_k^T + \sigma_\epsilon^2 \mathbf{I}_N) \mathbf{W}_k^T \mathbf{e}_m, \end{aligned} \quad (6)$$

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

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

$$\begin{aligned} &\text{minimize} \quad \text{tr}(\mathbf{P}_W) \\ &\text{subject to} \quad \sum_{k=1}^K T_m(\mathbf{W}_k) \leq E_m, \quad m \in [M] \\ &\quad \mathbf{W}_k \circ (\mathbf{1}_M \mathbf{1}_N^T - \mathbf{A}) = \mathbf{0}, \quad k \in [K], \end{aligned} \quad (7)$$

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, $T_m(\mathbf{W}_k)$ is the transmission cost given by (6), E_m is a prescribed energy budget of the m th sensor, and \mathbf{A} characterizes the network

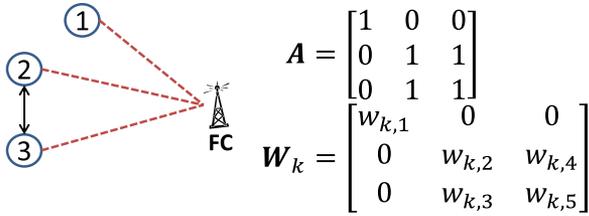


Fig. 2. Example of vectorization of \mathbf{W}_k .

topology. The problem structure and the solution of (7) will be elaborated on in the rest of the paper.

We end this section with the following remarks.

Remark 1: In the system model, the assumption of known observation and channel gains can be further relaxed to that of given knowledge about their second-order statistics. Our earlier work [17] has shown that under this weaker assumption, we can obtain similar expressions of the linear estimator. In this paper, we assume the observation and channel models are known for ease of presentation and analysis.

Remark 2: 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}_k is time varying in the presence of temporal dynamics of the sensor network. As will be evident later, the proposed sensor collaboration approach is also applicable to the problem with time-varying topologies.

III. REFORMULATION AND SIMPLIFICATION USING MATRIX VECTORIZATION

In this section, we simplify problem (7) by exploiting the sparsity structure of the topology matrix and concatenating the nonzero entries of a collaboration matrix into a collaboration vector. There exist two benefits to using matrix vectorization: a) the topology constraint in (7) can be eliminated without loss of performance, which renders a less complex problem; b) the structure of nonconvexities is more easily revealed via such a reformulation.

In problem (7), the only optimization variables are the nonzero entries of collaboration matrices. We concatenate these nonzero entries (columnwise) into a collaboration vector

$$\mathbf{w}_k = [w_{k,1}, w_{k,2}, \dots, w_{k,L}]^T, \quad (8)$$

where $w_{k,l}$ denotes the l th entry of \mathbf{w}_k , and L is the number of nonzero entries of the topology 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. 2, where we consider $N = 3$ sensor nodes, $M = 3$ communicating nodes, and 2 collaboration links.

A. Collaboration Problem for the Estimation of Uncorrelated Parameters

When the parameters of interest are uncorrelated, the estimation error covariance matrix (5) simplifies to

$$\begin{aligned} \mathbf{P}_W &= (\sigma_\theta^{-2} \mathbf{I} + \mathbf{D}_h^T \mathbf{D}_W^T (\sigma_\epsilon^2 \mathbf{D}_W \mathbf{D}_W^T + \sigma_\zeta^2 \mathbf{I})^{-1} \mathbf{D}_W \mathbf{D}_h)^{-1} \\ &= \left(\sigma_\theta^{-2} \mathbf{I} + \text{diag} \left\{ \frac{\mathbf{g}_k^T \mathbf{W}_k \mathbf{h}_k \mathbf{h}_k^T \mathbf{W}_k^T \mathbf{g}_k}{\sigma_\epsilon^2 \mathbf{g}_k^T \mathbf{W}_k \mathbf{W}_k^T \mathbf{g}_k + \sigma_\zeta^2} \right\}_{k=1}^K \right)^{-1} \\ &= \text{diag} \left\{ \frac{\sigma_\theta^2 \sigma_\epsilon^2 \mathbf{g}_k^T \mathbf{W}_k \mathbf{W}_k^T \mathbf{g}_k + \sigma_\theta^2 \sigma_\zeta^2}{\sigma_\theta^2 \mathbf{g}_k^T \mathbf{W}_k \mathbf{h}_k \mathbf{h}_k^T \mathbf{W}_k^T \mathbf{g}_k + \sigma_\epsilon^2 \mathbf{g}_k^T \mathbf{W}_k \mathbf{W}_k^T \mathbf{g}_k + \sigma_\zeta^2} \right\}_{k=1}^K, \end{aligned} \quad (9)$$

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

Let $\mathbf{w} \in \mathbb{R}^L$ be the vector obtained by stacking the nonzero entries of $\mathbf{W} \in \mathbb{R}^{M \times N}$ columnwise. Then

$$\mathbf{b}^T \mathbf{W} = \mathbf{w}^T \mathbf{B}, \quad (10)$$

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

$$B_{ln} = \begin{cases} b_{m_l} & n = n_l \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

and the indices m_l and n_l are such that $w_l = W_{m_l n_l}$ for $l \in [L]$. The proof of Eq. (10) is reported in [30, Appendix A] and omitted here for sake of brevity.

From (9) and (10), the objective function of problem (7) can be rewritten as

$$\phi(\mathbf{w}) := \text{tr}(\mathbf{P}_W) = \sum_{k=1}^K \frac{\sigma_\theta^2 \sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\theta^2 \sigma_\zeta^2}{\mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\zeta^2}, \quad (12)$$

where we used the fact that $\mathbf{g}_k^T \mathbf{W}_k = \mathbf{w}_k^T \mathbf{G}_k$, i.e., \mathbf{G}_k is derived from \mathbf{g}_k in the same way that \mathbf{B} is derived from \mathbf{b} in (10), and $\mathbf{S}_k := \mathbf{G}_k (\sigma_\theta^2 \mathbf{h}_k \mathbf{h}_k^T + \sigma_\epsilon^2 \mathbf{I}) \mathbf{G}_k^T$.

Moreover, the transmission cost (6) can be rewritten as

$$T_m(\mathbf{w}_k) := \mathbf{w}_k^T \mathbf{Q}_{k,m} \mathbf{w}_k, \quad (13)$$

$$\mathbf{Q}_{k,m} := \mathbf{E}_m (\sigma_\theta^2 \mathbf{h}_k \mathbf{h}_k^T + \sigma_\epsilon^2 \mathbf{I}) \mathbf{E}_m^T,$$

where \mathbf{E}_m is defined as in (10) 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 (12) and (13), the sensor collaboration problem for the estimation of temporally uncorrelated parameters becomes

$$\begin{aligned} &\text{minimize} && \phi(\mathbf{w}) \\ &\text{subject to} && \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \quad m \in [M], \end{aligned} \quad (\text{P1})$$

where $\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_K^T]^T$ is the optimization variable, $\phi(\mathbf{w})$ is the estimation distortion given by (12), and $\mathbf{Q}_m := \text{blkdiag}\{\mathbf{Q}_{k,m}\}_{k=1}^K$. Note that (P1) cannot be decomposed in time since sensor energy constraints are temporally inseparable.

Compared to problem (7), the topology constraint in terms of \mathbf{A} is eliminated without loss of performance in (P1) since the sparsity structure of the topology matrix has been taken into account while constructing the collaboration vector. In the special case of single-snapshot estimation (namely, $K = 1$), the objective function of (P1) simplifies to a single quadratic ratio. It has been shown in [13] and [17] that such a nonconvex problem can be readily solved via convex programming. In contrast, (P1) is a more complex nonconvex optimization problem, where the nonconvexity stems from the *sum* of quadratic ratios in the objective

function. As indicated in [31] and [32], the Karush-Kuhn-Tucker (KKT) conditions of such a complex fractional optimization problem are intractable to solve to obtain the globally optimal solution (or all locally optimal solutions). Therefore, an efficient local optimization method will be proposed to solve (P1) in Section IV. Also, the efficacy of the proposed solution will be shown in Section VI via extensive numerical experiments.

B. Collaboration Problem for the Estimation of Correlated Parameters

When parameters are temporally correlated, the covariance matrix Σ_θ is no longer diagonal and it is not straight forward to express the estimation error in a succinct form, as was done in (12). We recall from (5) that the dependence of the estimation error covariance on collaboration matrices is through \mathbf{D}_W . According to the matrix inversion lemma [29, A 1.1.3],

$$\begin{aligned} & \mathbf{D}_W^T (\sigma_\epsilon^2 \mathbf{D}_W \mathbf{D}_W^T + \sigma_\zeta^2 \mathbf{I})^{-1} \mathbf{D}_w \\ &= \sigma_\epsilon^{-2} \mathbf{I} - (\sigma_\epsilon^2 \mathbf{I} + \sigma_\epsilon^4 \sigma_\zeta^{-2} \mathbf{D}_W^T \mathbf{D}_W)^{-1}. \end{aligned} \quad (14)$$

Substituting (14) into (5), we obtain

$$\mathbf{P}_W = (\mathbf{C} - \sigma_\epsilon^{-2} \mathbf{D}_h^T (\mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{D}_W^T \mathbf{D}_W)^{-1} \mathbf{D}_h)^{-1} \quad (15)$$

with $\mathbf{C} := \Sigma_\theta^{-1} + \sigma_\epsilon^{-2} \mathbf{D}_h^T \mathbf{D}_h$. According to the definition of \mathbf{D}_W in (4), we obtain

$$\begin{aligned} \mathbf{D}_W^T \mathbf{D}_W &= \text{blkdiag} \{ \mathbf{W}_k^T \mathbf{g}_k \mathbf{g}_k^T \mathbf{W}_k \}_{k=1}^K \\ &= \text{blkdiag} \{ \mathbf{G}_k^T \mathbf{w}_k \mathbf{w}_k^T \mathbf{G}_k \}_{k=1}^K, \end{aligned} \quad (16)$$

where \mathbf{G}_k has been introduced in the paragraph that proceeds (12).

Combining (15) and (16), we can rewrite the estimation error covariance as a function of the collaboration vector

$$\begin{aligned} \mathbf{P}_w := & \left(\mathbf{C} - \sigma_\epsilon^{-2} \text{diag} \{ \mathbf{h}_k^T (\mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k^T \mathbf{w}_k \mathbf{w}_k^T \mathbf{G}_k)^{-1} \right. \\ & \left. \cdot \mathbf{h}_k \}_{k=1}^K \right)^{-1}. \end{aligned} \quad (17)$$

From (17), the sensor collaboration problem for the estimation of temporally correlated parameters becomes

$$\begin{aligned} & \text{minimize} \quad \text{tr}(\mathbf{P}_w) \\ & \text{subject to} \quad \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \quad m \in [M], \end{aligned} \quad (P2)$$

where $\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_K^T]^T$ is the optimization variable.

We note that (P2) is a nonconvex optimization problem. We will show in Section V that the rank-one matrix $\mathbf{w}_k \mathbf{w}_k^T$ that appears in (17) is the source of nonconvexity. Compared to (P1), (P2) is more involved due to the presence of the parameter correlation. We will also show that (P2) can be cast as a particular nonconvex optimization problem, where the objective function is linear, and the constraint set is formed by convex quadratic constraints, linear matrix inequalities and nonconvex rank constraints. The presence of generalized inequalities (with respect to positive semidefinite cones) and rank constraints make KKT conditions complex and intractable to find the globally optimal solution. Instead, we will employ an efficient convexification method to find a locally optimal solution of (P2). The efficacy

of the proposed optimization method will be empirically shown in Section VI.

We finally remark that both (P1) and (P2) are feasible optimization problems, namely, in the sense that an optimal solution exists for each of them. This can be examined as follows. First, there exists a non-empty constraint set. For example, $\mathbf{w}_k = \mathbf{0}$ is a feasible solution to (P1) and (P2). When $\mathbf{w}_k = \mathbf{0}$, the estimate of the unknown parameter is only determined by the prior knowledge about the parameter. Second, the optimal value is bounded due to the presence of the energy constraint.

IV. SPECIAL CASE: OPTIMAL SENSOR COLLABORATION FOR THE ESTIMATION OF UNCORRELATED PARAMETERS

In this section, we show that (P1) 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 (P1).

A. Equivalent Optimization Problem

We express (P1) in its epigraph form [33, Sections 3.1 & 7.5]

$$\text{minimize} \quad \mathbf{1}^T \mathbf{u} \quad (18a)$$

$$\text{subject to} \quad \frac{\sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\zeta^2}{\mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\zeta^2} \leq u_k, \quad k \in [K] \quad (18b)$$

$$\mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \quad m \in [M], \quad (18c)$$

where $\mathbf{u} = [u_1, u_2, \dots, u_K]^T$ is the vector of newly introduced optimization variables.

We further introduce new variables r_k and s_k for $k \in [K]$ to rewrite (18b) as

$$\begin{cases} \frac{r_k}{s_k} \leq u_k, \quad s_k > 0 \\ \mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\zeta^2 \geq s_k \\ \sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\zeta^2 \leq r_k, \end{cases} \quad (19)$$

where the equivalence between (18b) and (19) holds since the minimization of $\mathbf{1}^T \mathbf{u}$ with the above inequalities forces the variable s_k and r_k to achieve their upper and lower bounds, respectively.

In (19), the ratio $r_k/s_k \leq u_k$ together with $s_k > 0$ can be reformulated as a quadratic inequality of DC type

$$s_k^2 + u_k^2 + 2r_k - (s_k + u_k)^2 \leq 0, \quad (20)$$

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

From (19) and (20), problem (18) becomes

$$\text{minimize} \quad \mathbf{1}^T \mathbf{u} \quad (21a)$$

$$\text{subject to} \quad s_k^2 + u_k^2 + 2r_k \leq (s_k + u_k)^2, \quad k \in [K] \quad (21b)$$

$$s_k - \mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k - \sigma_\zeta^2 \leq 0, \quad k \in [K] \quad (21c)$$

$$\sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\zeta^2 \leq r_k, \quad k \in [K] \quad (21d)$$

$$\mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \quad m \in [M] \quad (21e)$$

$$s > \mathbf{0}, \quad (21f)$$

Algorithm 1: CCP for Solving (P1).

Require: initial points $\hat{\mathbf{w}}$, $\hat{\mathbf{s}}$ and $\hat{\mathbf{u}}$, and $\epsilon_{\text{ccp}} > 0$

- 1: **for** iteration $t = 1, 2, \dots$ **do**
- 2: solve problem (24) for the solution $(\mathbf{w}^t, \mathbf{s}^t, \mathbf{u}^t)$
- 3: update the linearization point, $\hat{\mathbf{w}} = \mathbf{w}^t$, $\hat{\mathbf{s}} = \mathbf{s}^t$,
and $\hat{\mathbf{u}} = \mathbf{u}^t$
- 4: **until** $|\mathbf{1}^T \mathbf{u}^t - \mathbf{1}^T \mathbf{u}^{t-1}| \leq \epsilon_{\text{ccp}}$ with $t \geq 2$.
- 5: **end for**

where the optimization variables are \mathbf{w} , \mathbf{u} , \mathbf{r} and \mathbf{s} , $\mathbf{r} = [r_1, r_2, \dots, r_K]^T$, $\mathbf{s} = [s_1, s_2, \dots, s_K]^T$, and $>$ denotes elementwise inequality. Note that the quadratic functions of DC type in (21b) and (21c) contain the nonconvexity of problem (21). In what follows, we will show that CCP is a suitable convex restriction approach for solving this problem.

B. Convex Restriction

Problem (21) is convex except for the nonconvex quadratic constraints (21b) and (21c), which have the DC form

$$f(\mathbf{v}) - g(\mathbf{v}) \leq 0, \quad (22)$$

where both f and g are convex functions. In (21b), 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 (21c), $f(\mathbf{s}_k) = s_k$, and $g(\mathbf{w}_k) = \mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\zeta^2$.

We can convexify (22) by linearizing g around a feasible point $\hat{\mathbf{v}}$,

$$f(\mathbf{v}) - \hat{g}(\mathbf{v}) \leq 0, \quad (23)$$

where $\hat{g}(\mathbf{v}) := g(\hat{\mathbf{v}}) + (\frac{\partial g(\hat{\mathbf{v}})}{\partial \mathbf{v}})^T (\mathbf{v} - \hat{\mathbf{v}})$, $\frac{\partial g(\hat{\mathbf{v}})}{\partial \mathbf{v}}$ is the first-order derivative of g at the point $\hat{\mathbf{v}}$. In (23), \hat{g} is an affine *lower bound* on the convex function g , and therefore, the set of \mathbf{v} that satisfy (23) is a strict *subset* of the set of \mathbf{v} that satisfy (22). This implies that a solution of the optimization problem with the linearized constraint (23) is locally optimal for the problem with the original nonconvex constraint (22).

We can obtain a restricted convex version of problem (21) by linearizing (21b) and (21c) as was done in (22) and (23). 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 [27]. At each iteration of CCP, we solve

$$\begin{aligned} & \text{minimize} && \mathbf{1}^T \mathbf{u} \\ & \text{subject to} && s_k^2 + u_k^2 + 2r_k - \hat{g}_1(s_k, u_k) \leq 0, && k \in [K] \\ & && s_k - \hat{g}_2(\mathbf{w}_k) \leq 0, && k \in [K] \\ & && \sigma_\epsilon^2 \mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k + \sigma_\zeta^2 \leq r_k, && k \in [K] \\ & && \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, && m \in [M] \\ & && \mathbf{s} > \mathbf{0}, \end{aligned} \quad (24)$$

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 $\mathbf{w}_k^T \mathbf{S}_k \mathbf{w}_k + \sigma_\zeta^2$, namely, $\hat{g}_1(s_k, u_k) := 2(\hat{s}_k + \hat{u}_k)(s_k + u_k) - (\hat{s}_k + \hat{u}_k)^2$, and $\hat{g}_2(\mathbf{w}_k) := 2\hat{\mathbf{w}}_k^T \mathbf{S}_k \mathbf{w}_k - \hat{\mathbf{w}}_k^T \mathbf{S}_k \hat{\mathbf{w}}_k + \sigma_\zeta^2$. We summarize CCP for solving problem (21) or (P1) in Algorithm 1.

To initialize Algorithm 1, we can choose random points, for example drawn from a standard uniform distribution, that are

then scaled to satisfy the constraints (21b)–(21e). Our extensive numerical examples show that Algorithm 1 is fairly robust with respect to the choice of the initial point; see Fig. 3(a) for an example.

It is known from [34, Theorem 10] that CCP is a descent algorithm that converges to a stationary point of the original nonconvex problem. To be specific, 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). Therefore, we always obtain a new feasible point with a lower or equal objective value. Moreover, reference [35] showed that CCP has *at least* linear convergence rate $O(1/t)$, where t is the number of iterations.¹ However, our numerical results and those in [26], [27] have shown that the empirical convergence rate is typically faster, and much of the benefit of using CCP is gained during its first few iterations.

The computation cost of Algorithm 1 is dominated by the solution of the convex program with quadratic constraints at Step 2. This has the computational complexity $O(a^3 + a^2b)$ in the use of interior-point algorithm [36, Chapter 10], where a and b denote the number of optimization variables and constraints, respectively. In problem (24), 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 .

V. GENERAL CASE: OPTIMAL SENSOR COLLABORATION FOR THE ESTIMATION OF CORRELATED PARAMETERS

Different from (P1), the presence of temporal correlation makes finding the solution of (P2) more challenging. However, we demonstrate that (P2) can be recast as an optimization problem with the important property that the problem becomes a semidefinite program (SDP) if its rank-one constraint is replaced by a linear relaxation/approximation. Spurred by the problem structure, we employ a *penalty CCP* to solve (P2), and propose a fast optimization algorithm by using the alternating direction method of multipliers (ADMM).

A. Equivalent Optimization Problem

We transform (P2) into the following equivalent form

$$\begin{aligned} & \text{minimize} && \text{tr}(\mathbf{V}) \\ & \text{subject to} && \mathbf{P}_w^{-1} \succeq \mathbf{V}^{-1} \\ & && \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \quad m \in [M], \end{aligned} \quad (25)$$

where $\mathbf{V} \in \mathbb{S}^K$ is the newly introduced optimization variable, \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 (25) is obtained from $\mathbf{P}_w \preceq \mathbf{V}$, where \mathbf{P}_w is given by (17), and \mathbf{P}_w^{-1} represents the Bayesian Fisher information matrix.

We further introduce a new vector of optimization variables $\mathbf{p} = [p_1, \dots, p_K]^T$ such that the first matrix inequality of

¹Given the stopping tolerance ϵ_{ccp} , the linear convergence rate implies $O(1/\epsilon_{\text{ccp}})$ iterations to convergence.

problem (25) is expressed as

$$\mathbf{C} - \text{diag}(\mathbf{p}) \succeq \mathbf{V}^{-1}, \quad (26)$$

$$p_k \geq \sigma_\epsilon^{-2} \mathbf{h}_k^T (\mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k)^{-1} \mathbf{h}_k, \quad k \in [K], \quad (27)$$

$$\mathbf{U}_k = \mathbf{w}_k \mathbf{w}_k^T, \quad (28)$$

where we use the expression of \mathbf{P}_w given by (17), and $\mathbf{U}_k \in \mathbb{S}^L$ is the newly introduced optimization variable for $k \in [K]$. Note that the minimization of $\text{tr}(\mathbf{V})$ with inequalities (26) and (27) would force the variable p_k to achieve its lower bound. In other words, problem (25) is equivalent to the problem in which the first inequality constraint of (25) is replaced by the above two inequalities.

By employing the Schur complement, we can express (26) and (27) as the linear matrix inequalities (LMIs)

$$\begin{bmatrix} \mathbf{C} - \text{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} \succeq 0, \quad (29)$$

$$\begin{bmatrix} p_k & \sigma_\epsilon^{-1} \mathbf{h}_k^T \\ \sigma_\epsilon^{-1} \mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} \succeq 0, \quad k \in [K]. \quad (30)$$

Replacing the first inequality of problem (25) with LMIs (29)–(30), we obtain an optimization problem that is convex except for the rank-one constraint (28), which can be recast as two inequalities

$$\mathbf{U}_k - \mathbf{w}_k \mathbf{w}_k^T \succeq 0, \quad \mathbf{U}_k - \mathbf{w}_k \mathbf{w}_k^T \preceq 0, \quad k \in [K]. \quad (31)$$

According to the Shur complement, the first matrix inequality is equivalent to the LMI

$$\begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} \succeq 0, \quad k \in [K]. \quad (32)$$

And the second inequality in (31) involves a function of DC type, where \mathbf{U}_k and $\mathbf{w}_k \mathbf{w}_k^T$ are matrix convex functions [33].

From (29)–(32), problem (25) or (P2) is equivalent to

$$\text{minimize} \quad \text{tr}(\mathbf{V}) \quad (33a)$$

$$\text{subject to} \quad \mathbf{w}^T \mathbf{Q}_m \mathbf{w} \leq E_m, \quad m \in [M] \quad (33b)$$

$$\text{LMIs in (29)–(30)} \quad (33c)$$

$$\text{LMIs in (32)} \quad (33d)$$

$$\mathbf{U}_k - \mathbf{w}_k \mathbf{w}_k^T \preceq 0, \quad k \in [K], \quad (33e)$$

where the optimization variables are \mathbf{w} , \mathbf{p} , \mathbf{V} and \mathbf{U}_k for $k \in [K]$, and (33e) is a nonconvex constraint of DC type.

B. Convexification

Proceeding with the same logic as in Section IV to convexify the constraint (22), we linearize (33e) around a point $\hat{\mathbf{w}}_k$,

$$\mathbf{U}_k - \hat{\mathbf{w}}_k \mathbf{w}_k^T - \mathbf{w}_k \hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T \preceq 0, \quad k \in [K]. \quad (34)$$

It is straightforward to apply CCP to solve problem (33) by replacing (33e) with (34). However, such an approach fails

Algorithm 2: Penalty CCP for Solving (P2).

Require: an initial point $\hat{\mathbf{w}}$, $\epsilon_{\text{ccp}} > 0$, $\tau^0 > 0$, $\tau_{\text{max}} > 0$ and $\mu > 1$.

- 1: **for** iteration $t = 1, 2, \dots$ **do**
 - 2: solve problem (37) for its solution \mathbf{w}^t via SDP solver or ADMM-based algorithm in Section V-C
 - 3: update the linearization point, $\hat{\mathbf{w}} = \mathbf{w}^t$
 - 4: update the penalty parameter $\tau^t = \min\{\mu\tau^{t-1}, \tau_{\text{max}}\}$
 - 5: let ψ^t be the objective value of (37)
 - 6: **until** $|\psi^t - \psi^{t-1}| \leq \epsilon_{\text{ccp}}$ with $t \geq 2$.
 - 7: **end for**
-

in practice. This is not surprising, since the feasible set determined by (33d) and (34) only contains the linearization point. Specifically, from (33d) and (34), we obtain

$$\begin{aligned} & (\mathbf{w}_k - \hat{\mathbf{w}}_k)(\mathbf{w}_k - \hat{\mathbf{w}}_k)^T \\ &= \mathbf{w}_k \mathbf{w}_k^T - \hat{\mathbf{w}}_k \mathbf{w}_k^T - \mathbf{w}_k \hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T \\ &\preceq \mathbf{U}_k - \hat{\mathbf{w}}_k \mathbf{w}_k^T - \mathbf{w}_k \hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T \preceq 0, \end{aligned} \quad (35)$$

which indicates that $\mathbf{w}_k = \hat{\mathbf{w}}_k$. Therefore, CCP gets trapped in the linearization point.

Remark 3: Dropping the nonconvex constraint (33e) is another method to convexify problem (33), known as semidefinite relaxation [37]. However, such an approach makes the optimization variable \mathbf{U}_k unbounded, since the minimization of $\text{tr}(\mathbf{V})$ forces \mathbf{U}_k to be as large as possible such that the variable p_k in (27) is as small as possible.

In order to circumvent the drawback of the standard CCP, we consider its penalized version, known as penalty CCP [27], [38], where we add new variables to allow for constraints (34) to be violated and penalize the sum of the violations in the objective function. As a result, the convexification (34) is modified by

$$\mathbf{U}_k - \hat{\mathbf{w}}_k \mathbf{w}_k^T - \mathbf{w}_k \hat{\mathbf{w}}_k^T + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T \preceq \mathbf{Z}_k, \quad k \in [K], \quad (36)$$

where $\mathbf{Z}_k \in \mathbb{S}^L$ is a newly introduced variable. The constraint (36) implicitly adds the additional constraint $\mathbf{Z}_k \succeq 0$ due to $\mathbf{U}_k \succeq \mathbf{w}_k \mathbf{w}_k^T$ from (33d).

After replacing (33e) with (36), we obtain the SDP,

$$\begin{aligned} & \text{minimize} \quad \text{tr}(\mathbf{V}) + \tau \sum_{k=1}^K \text{tr}(\mathbf{Z}_k) \\ & \text{subject to} \quad (33b)–(33d) \text{ and } (36) \end{aligned} \quad (37)$$

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 (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 (P2) in Algorithm 2.

In Algorithm 2, the initial point $\hat{\mathbf{w}}$ is randomly picked from a standard uniform distribution. Note that $\hat{\mathbf{w}}$ is not necessarily feasible for (P2) since violations of constraints are allowed. We also remark that once $\tau = \tau_{\text{max}}$ (after at most $\log_\mu(\tau_{\text{max}}/\tau_0)$

iterations), the penalty CCP reduces to CCP. Therefore, the penalty CCP enjoys the same convergence properties of CCP.

The computation cost of Algorithm 2 is dominated by the solution of the SDP (37) at Step 2. This leads to the complexity $O(a^2b^2 + ab^3)$ by using the interior-point algorithm in off-the-shelf solvers [36, Chapter 11], where a and b are the number of optimization variables and the size of the semidefinite matrix, respectively. In (37), the number of optimization variables is proportional to L^2 . Therefore, the complexity of Algorithm 2 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.

C. Fast Algorithm via ADMM

It has been shown in [28], [39]–[41] 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 (37).

It is shown in Appendix A that problem (37) can be reformulated in a way that lends itself to the application of ADMM. This is achieved by introducing slack variables and indicator functions to express the inequality constraints of problem (37) as linear equality constraints together with cone constraints with respect to slack variables, including second-order cone and positive semidefinite cone constraints.

ADMM is performed based on the augmented Lagrangian [28] of the reformulated problem (37), and leads to two problems, the first of which can be treated as an unconstrained quadratic program and the latter renders an analytical solution. These two problems are solved iteratively and communicate to each other through special quadratic terms in their objectives; the quadratic term in each problem contains information about the solution of the other problem and also about dual variables (also known as Lagrange multipliers). In what follows, we refer to these problems as the ‘ \mathcal{X} -minimization’ and ‘ \mathcal{Z} -minimization’ problems. Here \mathcal{X} denotes the set of primal variables \mathbf{w} , \mathbf{p} , \mathbf{V} , \mathbf{U}_k and \mathbf{Z}_k for $k \in [K]$, and \mathcal{Z} denotes the set of slack variables λ_m , $\mathbf{\Lambda}_1$ and $\{\mathbf{\Lambda}_{i,k}\}_{i=2,3,4}$ for $m \in [M]$ and $k \in [K]$. We also use \mathcal{Y} to denote the set of dual variables π_m , $\mathbf{\Pi}_1$ and $\{\mathbf{\Pi}_{i,k}\}_{i=2,3,4}$ for $m \in [M]$ and $k \in [K]$. The ADMM algorithm is precisely described by (56)–(58) in Appendix A.

We emphasize that the crucial property of the ADMM approach is that, as we demonstrate in the rest of this section, the solution of each of the \mathcal{X} - and \mathcal{Z} -minimization problems can be found exactly and efficiently.

1) *\mathcal{X} -Minimization Step:* The \mathcal{X} -minimization problem can be cast as

$$\text{minimize } \varphi(\mathbf{w}, \mathbf{p}, \mathbf{V}, \{\mathbf{U}_k\}, \{\mathbf{Z}_k\}). \quad (38)$$

The objective function of problem (38) is given by (39), where $\alpha_m := \lambda_m^t - \mathbf{c}_m - (1/\rho)\pi_m^t$ for $m \in [M]$, $\mathbf{\Upsilon}_1 := \mathbf{\Lambda}_1^t - (1/\rho)\mathbf{\Pi}_1^t$, and $\mathbf{\Upsilon}_{i,k} := \mathbf{\Lambda}_{i,k}^t - (1/\rho)\mathbf{\Pi}_{i,k}^t$ for $i \in \{2, 3, 4\}$ and $k \in [K]$, and t denotes the ADMM iteration. For ease of notation, we will omit the ADMM iteration index t in

what follows.

$$\begin{aligned} \varphi(\mathbf{w}, \mathbf{p}, \mathbf{V}, \{\mathbf{U}_k\}, \{\mathbf{Z}_k\}) &:= \text{tr}(\mathbf{V}) + \tau \sum_{k=1}^K \text{tr}(\mathbf{Z}_k) + \frac{\rho}{2} \sum_{m=1}^M \\ &\times \|\bar{\mathbf{Q}}_m \mathbf{w} - \alpha_m\|_2^2 + \frac{\rho}{2} \left\| \begin{bmatrix} \mathbf{C} - \text{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} - \mathbf{\Upsilon}_1 \right\|_F^2 \\ &+ \frac{\rho}{2} \sum_{k=1}^K \left\| \begin{bmatrix} p_k & \sigma_\epsilon^{-1} \mathbf{h}_k^T \\ \sigma_\epsilon^{-1} \mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} - \mathbf{\Upsilon}_{2,k} \right\|_F^2 \\ &+ \frac{\rho}{2} \sum_{k=1}^K \left\| \begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} - \mathbf{\Upsilon}_{3,k} \right\|_F^2 + \frac{\rho}{2} \sum_{k=1}^K \\ &\times \|\mathbf{Z}_k - \mathbf{U}_k + \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T - \mathbf{\Upsilon}_{4,k}\|_F^2 \end{aligned} \quad (39)$$

We note that problem (38) is an unconstrained quadratic program (UQP) with large amounts of variables. In order to reduce the computational complexity and memory requirement in optimization, we will employ a gradient descent method [33] together with a backtracking line search [33, Chapter 9.2] to solve this UQP. In Proposition 1, we show the gradient of the objective function of problem (38).

Proposition 1: The gradient of the objective function of problem (38) is given by

$$\begin{aligned} \nabla_{\mathbf{w}} \varphi &= \rho \sum_{m=1}^M \bar{\mathbf{Q}}_m^T (\bar{\mathbf{Q}}_m \mathbf{w} - \alpha_m) + 2\rho(\mathbf{w} - \gamma_3) \\ &\quad + 2\rho \text{blkdiag} \{ \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \mathbf{H}_k \}_{k=1}^K \hat{\mathbf{w}} \\ \nabla_{\mathbf{p}} \varphi &= 2\rho \mathbf{p} + \rho(\text{diag}(\mathbf{\Upsilon}_1^{11}) - \text{diag}(\mathbf{C}) - \gamma_2) \\ \nabla_{\mathbf{V}} \varphi &= \mathbf{I} + \rho(\mathbf{V} - \mathbf{\Upsilon}_1^{22}) \\ \nabla_{\mathbf{U}_k} \varphi &= \rho \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k (\mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k - \mathbf{\Upsilon}_{2,k}^{22}) \mathbf{G}_k^T \\ &\quad + \rho(2\mathbf{U}_k - \mathbf{\Upsilon}_{3,k}^{11} - \mathbf{Z}_k - \mathbf{T}_k), \quad k \in [K] \\ \nabla_{\mathbf{Z}_k} \varphi &= \tau \mathbf{I} + \rho(\mathbf{Z}_k - \mathbf{U}_k + \mathbf{T}_k), \quad k \in [K], \end{aligned}$$

where $\gamma_3 = [\gamma_{3,1}^T, \dots, \gamma_{3,K}^T]^T$, $\gamma_{3,k}$ is the $(L+1)$ column of $\mathbf{\Upsilon}_{3,k}$ after the last entry is removed, $\mathbf{H}_k := \mathbf{U}_k - \mathbf{Z}_k + \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T + \mathbf{\Upsilon}_{4,k}$, $\hat{\mathbf{w}} = [\hat{\mathbf{w}}_1^T, \dots, \hat{\mathbf{w}}_K^T]^T$, $\mathbf{\Upsilon}_1^{11}$ is a submatrix of $\mathbf{\Upsilon}_1$ that contains its first K rows and columns, $\gamma_2 = [\gamma_{2,1}, \dots, \gamma_{2,K}]^T$, $\gamma_{2,k}$ is the first element of $\mathbf{\Upsilon}_{2,k}$, $\text{diag}(\cdot)$ returns the diagonal entries of its matrix argument in vector form, $\mathbf{\Upsilon}_1^{22}$ is a submatrix of $\mathbf{\Upsilon}_1$ after the first K rows and columns are removed, $\mathbf{\Upsilon}_{2,k}^{22}$ 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 $\mathbf{T}_k := \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T - \mathbf{\Upsilon}_{4,k}$.

Proof: See [30, Appendix C]. \blacksquare

In Proposition 1, the optimal values of \mathbf{p} and \mathbf{V} are achieved by letting $\nabla_{\mathbf{p}} \varphi = \mathbf{0}$ and $\nabla_{\mathbf{V}} \varphi = \mathbf{0}$, which yield

$$\mathbf{p} = \frac{1}{2} (\text{diag}(\mathbf{C}) + \gamma_2 - \text{diag}(\mathbf{\Upsilon}_1^{11})), \quad \mathbf{V} = \mathbf{\Upsilon}_1^{22} - \frac{1}{\rho} \mathbf{I}. \quad (40)$$

To solve problem (38) for other variables, we employ the gradient descent method summarized in Algorithm 3. This

Algorithm 3: Gradient Descent Method for Solving UQP (38).

Require: values of \mathbf{w} , $\{\mathbf{U}_k\}$ and $\{\mathbf{Z}_k\}$ at the previous ADMM iteration, \mathbf{p} and \mathbf{V} given by (40), and $\epsilon_{\text{grad}} > 0$

- 1: **repeat**
- 2: compute the gradient of ϕ following Proposition 1
- 3: compute $c_{\text{grad}} := \sum_{k=1}^K \|\nabla_{\mathbf{U}_k} \phi\|_F^2 + \|\nabla_{\mathbf{w}} \phi\|_2^2 + \sum_{k=1}^K \|\nabla_{\mathbf{Z}_k} \phi\|_F^2$
- 4: call Algorithm 4 to determine a step size κ
- 5: update variables $\mathbf{w} := \mathbf{w} + \kappa \nabla_{\mathbf{w}} \phi$, $\mathbf{U}_k := \mathbf{U}_k + \kappa \nabla_{\mathbf{U}_k} \phi$, $\mathbf{Z}_k := \mathbf{Z}_k + \kappa \nabla_{\mathbf{Z}_k} \phi$
- 6: **until** $c_{\text{grad}} \leq \epsilon_{\text{grad}}$.

Algorithm 4: 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 $\hat{\phi}$ be the value of ϕ at the points $\mathbf{w} + \kappa \nabla_{\mathbf{w}} \phi$, $\mathbf{U}_k + \kappa \nabla_{\mathbf{U}_k} \phi$, and $\mathbf{Z}_k + \kappa \nabla_{\mathbf{Z}_k} \phi$
- 5: **until** $\hat{\phi} < \phi(\mathbf{w}, \{\mathbf{U}_k\}, \{\mathbf{Z}_k\}) - a_1 \kappa c_{\text{grad}}$.

algorithm calls on the backtracking line search (Algorithm 4) to properly determine the step size such that the convergence to a stationary point of problem (38) is accelerated.

2) \mathcal{L} -Minimization Step: The \mathcal{L} -minimization problem is decomposed with respect to each of slack variables.

- Subproblem with respect to λ_m :

$$\begin{aligned} & \text{minimize} && \|\lambda_m - \beta_m\|_2^2 \\ & \text{subject to} && \|[\lambda_m]_{1:KL}\|_2 \leq [\lambda_m]_{KL+1}, \end{aligned} \quad (41)$$

where $\beta_m := \bar{\mathbf{Q}}_m \mathbf{w}^{t+1} + \mathbf{c}_m + (1/\rho) \boldsymbol{\pi}_m^t$, and t is the ADMM iteration index. For notational simplicity, the ADMM iteration will be omitted in what follows. The solution of problem (41) is achieved by projecting β_m onto a second-order cone [39, Section 6.3],

$$\lambda_m = \begin{cases} \mathbf{0} & \|[\beta_m]_{1:KL}\|_2 \leq -[\beta_m]_{KL+1} \\ \beta_m & \|[\beta_m]_{1:KL}\|_2 \leq [\beta_m]_{KL+1} \\ \tilde{\beta}_m & \|[\beta_m]_{1:KL}\|_2 \geq |[\beta_m]_{KL+1}|, \end{cases} \quad (42)$$

for $m \in [M]$, where

$$\tilde{\beta}_m = \frac{1}{2} \left(1 + \frac{[\beta_m]_{KL+1}}{\|[\beta_m]_{1:KL}\|_2} \right) [[\beta_m]_{1:KL}^T, \|[\beta_m]_{1:KL}\|_2]^T.$$

- Subproblem with respect to $\boldsymbol{\Lambda}_1$:

$$\begin{aligned} & \text{minimize} && \|\boldsymbol{\Lambda}_1 - \Phi_1\|_F^2 \\ & \text{subject to} && \boldsymbol{\Lambda}_1 \succeq \mathbf{0}, \end{aligned} \quad (43)$$

where $\Phi_1 := \begin{bmatrix} \mathbf{C} - \text{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} + (1/\rho) \boldsymbol{\Pi}_1$. The solution of problem (43) is given by [39, Section 6.3]

$$\boldsymbol{\Lambda}_1 = \sum_{i=1}^{2K} (\sigma_i)_+ \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T, \quad (44)$$

Algorithm 5: ADMM for Solving Problem (37).

- 1: Initialize variables and set ρ and ϵ_{admm}
- 2: **for** iteration $t = 1, 2, \dots$ **do**
- 3: obtain optimal values of primal variables \mathcal{X}^t using Algorithm 3 and (40)
- 4: obtain optimal values of slack variables \mathcal{Z}^t using (42), (44) and (45)
- 5: update dual variables based on (58)
- 6: **until** both $\|\mathcal{X}^{t+1} - \mathcal{Z}^t\|_F$ and $\|\mathcal{Z}^{t+1} - \mathcal{Z}^t\|_F$ are less than ϵ_{admm} .
- 7: **end for**

where $\sum_{i=1}^{2K} \sigma_i \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T$ is the eigenvalue decomposition of Φ_1 , and $(\cdot)_+$ is the positive part operator.

- Subproblem with respect to $\boldsymbol{\Lambda}_{i,k}$ for $i \in \{2, 3, 4\}$ and $k \in [K]$:

$$\begin{aligned} & \text{minimize} && \|\boldsymbol{\Lambda}_{i,k} - \Phi_{i,k}\|_F^2 \\ & \text{subject to} && \boldsymbol{\Lambda}_{i,k} \succeq \mathbf{0}, \end{aligned} \quad (45)$$

where

$$\Phi_{2,k} := \begin{bmatrix} p_k & \sigma_\epsilon^{-1} \mathbf{h}_k^T \\ \sigma_\epsilon^{-1} \mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} + \frac{1}{\rho} \boldsymbol{\Pi}_{2,k}$$

$$\Phi_{3,k} := \begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} + \frac{1}{\rho} \boldsymbol{\Pi}_{3,k}$$

$$\Phi_{4,k} := \mathbf{Z}_k - \mathbf{U}_k + \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T + \frac{1}{\rho} \boldsymbol{\Pi}_{4,k}.$$

The solution of problem (45) is the same as (44) except that Φ_1 is replaced with $\Phi_{i,k}$ for $i \in \{2, 3, 4\}$ and $k \in [K]$.

3) *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 $k \in [K]$, $\lambda_m^0 = \boldsymbol{\pi}_m^0 = \mathbf{0}$ 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 approach is summarized in Algorithm 5.

The global convergence of ADMM has been widely studied in [42], [43]. It is known from [42], [43] that ADMM has a linear convergence rate $O(1/t)$ for general convex optimization problems such as problem (37), where t is the number of iterations. In practice, our numerical results and those in [28], [39]–[41] have shown that ADMM can converge to modest accuracy—sufficient for many applications—within a few tens of iterations.

At each iteration of ADMM, the computational complexity of the \mathcal{L} -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 \mathcal{L} -minimization step, the computational complexity is dominated by the eigenvalue decomposition used in (44). 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 $O(L^6)$. The complexity of ADMM decreases significantly in terms of the number of collaboration links by a factor L^2 . We refer the

reader to Section VI for numerical results on the running time improvement.

VI. 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 [15] 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 θ is given by

$$\Sigma_\theta = \sigma_\theta^2 \begin{bmatrix} 1 & e^{-\rho_{\text{corr}}} & \dots & e^{-(K-1)\rho_{\text{corr}}} \\ e^{-\rho_{\text{corr}}} & 1 & \dots & e^{-(K-2)\rho_{\text{corr}}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-(K-1)\rho_{\text{corr}}} & e^{-(K-2)\rho_{\text{corr}}} & \dots & 1 \end{bmatrix}.$$

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 [13], $\text{RGG}(N, d)$, 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 topology 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 $d = 0.3$. We refer interested readers to [30, Fig. 3] for an example of the random geometric graph used.

In the collaborative estimation system shown in Fig. 1, we assume that $M = N$, $K = 3$, $\sigma_c^2 = \sigma_e^2 = 1$, 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 observation 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 $\tau^0 = 0.1$, $\mu = 1.5$, $\tau_{\text{max}} = 100$ in penalty CCP (namely, Algorithm 2), $a_1 = 0.02$ and $a_2 = 0.5$ in backtracking line search (namely, Algorithm 4) 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 [44] for all other computations. The estimation performance is measured through the empirical mean squared error (MSE), which is computed over 1000 numerical trials.

In Fig. 3, we present convergence trajectories of CCP (namely, Algorithm 1) and penalty CCP (namely, Algorithm 2) as functions of iteration index for 10 different initial points. For comparison, we plot the worst objective function value of collaboration problem (7) when $\mathbf{w} = \mathbf{0}$, namely, LMMSE is determined only by the prior information, which leads to the worst estimation error $\text{tr}(\Sigma_\theta) = K = 3$. 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).

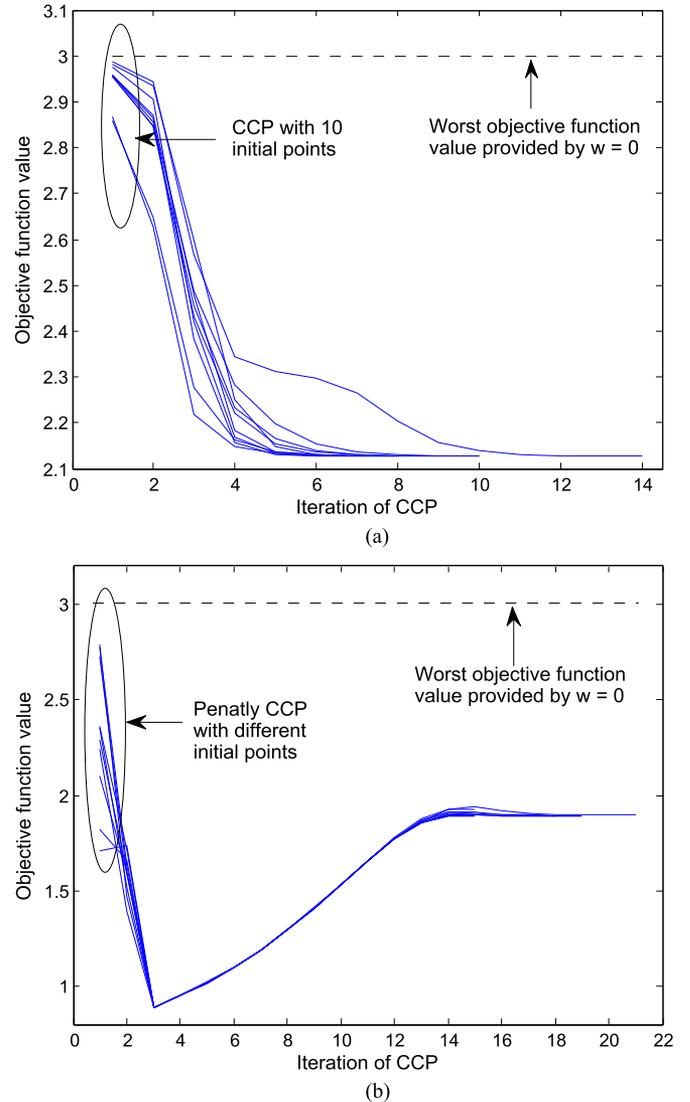


Fig. 3. Convergence of Algorithm 1 and 2 for different initial points.

In Fig. 4, we present the trace of error covariance matrix \mathbf{P}_W given by (5) as a function of the correlation parameter ρ_{corr} , where the sensor collaboration scheme is obtained from Algorithm 1 and Algorithm 2 to solve (P1) and (P2), respectively. We observe that the estimation error resulting from the solution of (P1) remains unchanged for different values of ρ_{corr} since the formulation of (P1) is independent of the prior knowledge about parameter correlation. The estimation error resulting from the solution of (P2) increases as ρ_{corr} increases, and it eventually converges to the error resulting from the solution of (P1) 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 (P2), thereby significantly improving the estimation performance.

In Fig. 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 (P1) and (P2), respectively. The assumption of time-invariant collaboration implicitly adds the additional constraint $\mathbf{w}_1 = \dots = \mathbf{w}_K$, which reduces the problem size. By fixing the type of algorithm, we

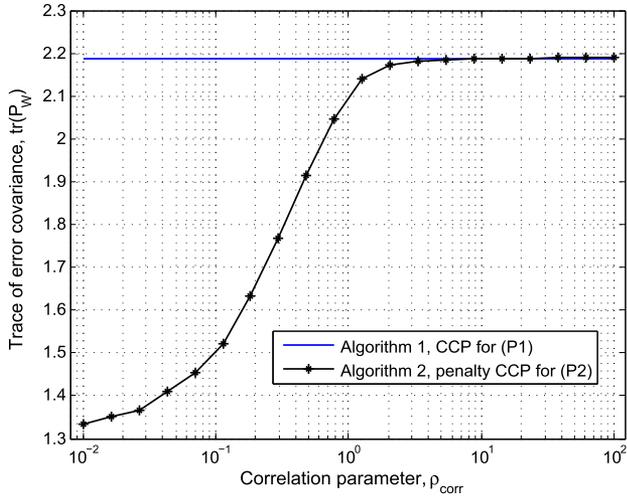
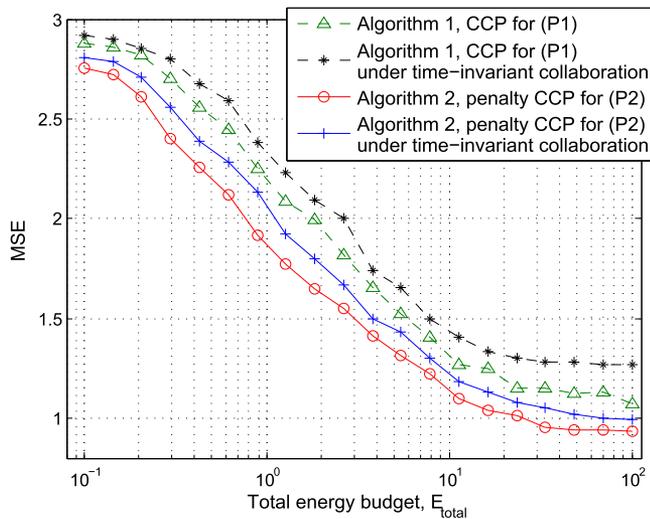
Fig. 4. Estimation error versus correlation parameter ρ_{corr} .

Fig. 5. MSE versus total energy budget.

observe that the MSE when using 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 (P2) yields lower MSE than that of (P1). This result is consistent with Fig. 4 for a fixed correlation parameter. Lastly, the estimation error is smaller as more energy is used in sensor collaboration.

In Fig. 6, we present the MSE and the number of collaboration links as functions of the collaboration radius d for $\rho_{\text{corr}} = 0.5$ and $E_{\text{total}} = 1$. 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 (P2) is lower than that when solving (P1), 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, we present the MSE as a function of the signal-to-noise ratio (SNR), $10 \log_{10}(\sigma_\theta^2/\sigma_v^2)$, where $\sigma_\theta^2 = 1$ is the variance of the parameter to be estimated, and $\sigma_v^2 \in [10^{-3}, 10^3]$

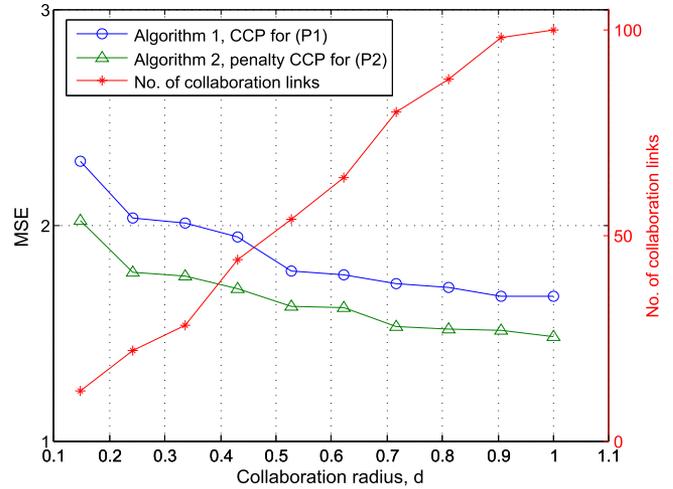
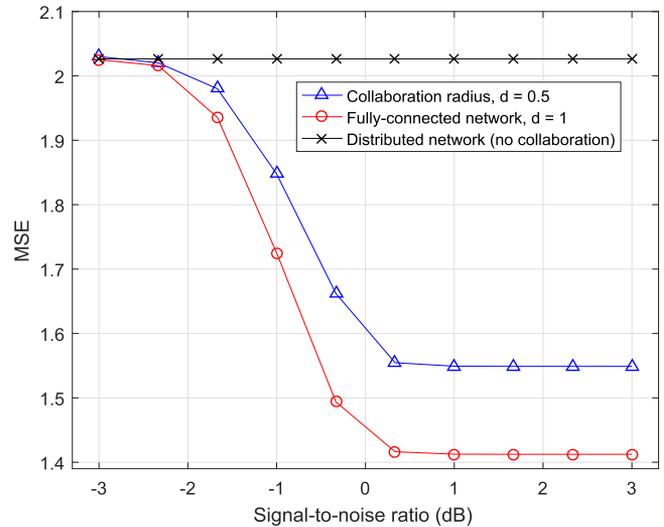
Fig. 6. MSE and collaboration links versus collaboration radius d .

Fig. 7. Noisy collaboration: MSE versus SNR.

is the variance of the additive communication noise when inter-sensor collaboration occurs. In this numerical example, we study the impact of *noisy* collaboration links on estimation performance, where the collaboration scheme is obtained by the solution of (P2) for $d \in \{0.5, 1\}$. As we can see, estimation distortion increases when SNR decreases. Moreover, the MSE in the presence of noisy collaboration under the lowest SNR is consistent with that of using the classical amplify-and-forward transmission strategy in the absence of sensor collaboration. This is because each sensor has access to its own measurement in a noiseless manner (collaboration noise only occurs if two different sensors are communicating). At a fixed value of SNR, we observe that the MSE decreases as d increases, and it converges to the MSE in the absence of collaboration noise. This implies that the act of sensor collaboration is able to improve estimation performance even if the collaboration link is noisy.

In Fig. 8, 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 [44]. As we can see, penalty CCP requires much higher computation time than CCP, since the former requires solutions

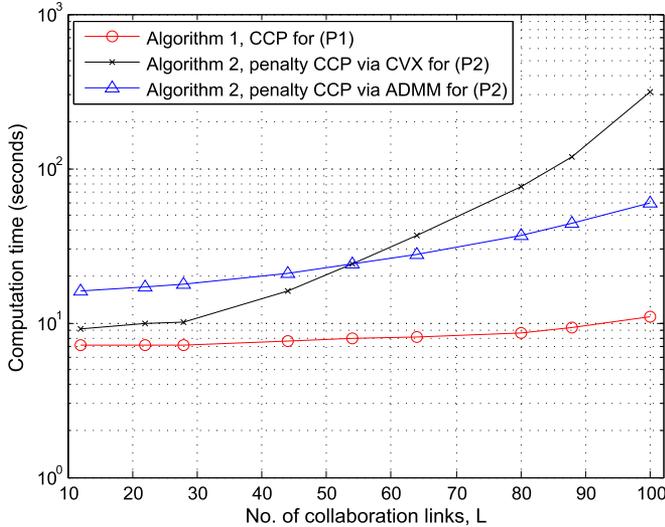


Fig. 8. Computation time versus number of collaboration links.

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 $L > 80$.

VII. CONCLUSION

We study the problem of sensor collaboration for estimation of time-varying parameters in sensor networks. Based on prior knowledge about parameter correlations, the resulting sensor collaboration problem is solved for 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 good locally optimal solution as 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 problem structure, we employ a semidefinite programming based penalty convex-concave procedure to solve the sensor collaboration problem. Moreover, we propose an ADMM-based algorithm that scales more gracefully for large problems. Numerical results are provided to demonstrate the effectiveness of our approach and the impact of parameter correlation and temporal dynamics of sensor networks on the performance of distributed estimation with sensor collaboration.

There are multiple directions for future research. We would like to consider noise-corrupted or quantization-based imperfect communication links in sensor collaboration. It will also be of interest to seek the duality gap between the nonconvex sensor collaboration problems in order to gain theoretical insights on the performance of the proposed optimization methods. Another direction of future work is to seek an approach that jointly designs the optimal power allocation scheme and the

collaboration topology. Last but not the least, it will be worthwhile to study the sensor collaboration problem in the framework of consensus-based decentralized estimation.

APPENDIX A APPLICATION OF ADMM

We introduce slack variables $\lambda_m \in \mathbb{R}^{KL+1}$ for $m \in [M]$ to rewrite (33b) as an equality constraint together with a second-order cone constraint,

$$\bar{\mathbf{Q}}_m \mathbf{w} - \lambda_m + \mathbf{c}_m = \mathbf{0}, \quad \|\lambda_m\|_2 \leq [\lambda_m]_{KL+1}, \quad (46)$$

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 $\Lambda_1 \in \mathbb{S}^{2K}$, $\Lambda_{2,k} \in \mathbb{S}^{N+1}$, $\Lambda_{3,k} \in \mathbb{S}^{L+1}$ and $\Lambda_{4,k} \in \mathbb{S}^L$ for $k \in [K]$ to rewrite LMIs of problem (37) as a sequence of equality constraints together with positive semidefinite cone constraints

$$\begin{bmatrix} \mathbf{C} - \text{diag}(\mathbf{p}) & \mathbf{I} \\ \mathbf{I} & \mathbf{V} \end{bmatrix} - \Lambda_1 = \mathbf{0} \quad (47)$$

$$\begin{bmatrix} p_k & \sigma_\epsilon^{-1} \mathbf{h}_k^T \\ \sigma_\epsilon^{-1} \mathbf{h}_k & \mathbf{I} + \sigma_\epsilon^2 \sigma_\zeta^{-2} \mathbf{G}_k^T \mathbf{U}_k \mathbf{G}_k \end{bmatrix} - \Lambda_{2,k} = \mathbf{0} \quad (48)$$

$$\begin{bmatrix} \mathbf{U}_k & \mathbf{w}_k \\ \mathbf{w}_k^T & 1 \end{bmatrix} - \Lambda_{3,k} = \mathbf{0} \quad (49)$$

$$\mathbf{Z}_k - \mathbf{U}_k + \hat{\mathbf{w}}_k \mathbf{w}_k^T + \mathbf{w}_k \hat{\mathbf{w}}_k^T - \hat{\mathbf{w}}_k \hat{\mathbf{w}}_k^T - \Lambda_{4,k} = \mathbf{0}, \quad (50)$$

where $\Lambda_1 \succeq \mathbf{0}$, $\Lambda_{2,k} \succeq \mathbf{0}$, $\Lambda_{3,k} \succeq \mathbf{0}$, and $\Lambda_{4,k} \succeq \mathbf{0}$ for $k \in [K]$. From (46)–(50), problem (37) becomes

$$\begin{aligned} \text{minimize} \quad & \text{tr}(\mathbf{V}) + \tau \sum_{k=1}^K \text{tr}(\mathbf{Z}_k) + \sum_{m=1}^M \mathcal{I}_0(\lambda_m) \\ & + \mathcal{I}_1(\Lambda_1) + \sum_{i=2}^4 \sum_{k=1}^K \mathcal{I}_i(\Lambda_{i,k}) \\ \text{subject to} \quad & \text{equality constraints in (46)–(50)}, \end{aligned} \quad (51)$$

where the optimization variables are \mathbf{w} , \mathbf{p} , \mathbf{V} , \mathbf{U}_k , \mathbf{Z}_k , λ_m , Λ_1 , and $\{\Lambda_{i,k}\}_{i=2,3,4}$ for $m \in [M]$ and $k \in [K]$, and \mathcal{I}_i is the indicator function specified by

$$\mathcal{I}_0(\lambda_m) = \begin{cases} 0, & \text{if } \|\lambda_m\|_2 \leq [\lambda_m]_{KL+1} \\ \infty & \text{otherwise,} \end{cases} \quad (52)$$

$$\mathcal{I}_1(\Lambda_1) = \begin{cases} 0, & \text{if } \Lambda_1 \succeq \mathbf{0} \\ \infty & \text{otherwise,} \end{cases} \quad (53)$$

$$\mathcal{I}_i(\Lambda_{i,k}) = \begin{cases} 0, & \text{if } \Lambda_{i,k} \succeq \mathbf{0} \\ \infty & \text{otherwise,} \end{cases} \quad i = 2, 3, 4. \quad (54)$$

It is clear from problem (51) that the introduced indicator functions helps to isolate the second-order cone and positive semidefinite cone constraints with respect to slack variables.

Problem (51) is now in a form suitable for the application of ADMM. The corresponding augmented Lagrangian [28] in ADMM is given by

$$\begin{aligned}
\mathcal{L}_\rho(\mathcal{X}, \mathcal{Z}, \mathcal{Y}) &= \text{tr}(\mathbf{V}) + \tau \sum_{k=1}^K \text{tr}(\mathbf{Z}_k) + \sum_{m=1}^M \mathcal{I}_0(\lambda_m) \\
&+ \mathcal{I}_1(\Lambda_1) + \sum_{i=2}^4 \sum_{k=1}^K \mathcal{I}_i(\Lambda_{i,k}) + \sum_{m=1}^M \pi_m^T \mathbf{f}_m(\mathcal{X}, \mathcal{Z}) \\
&+ \frac{\rho}{2} \sum_{m=1}^M \|\mathbf{f}_m(\mathcal{X}, \mathcal{Z})\|_2^2 + \text{tr}(\Pi_1^T \mathbf{F}_1(\mathcal{X}, \mathcal{Z})) \\
&+ \frac{\rho}{2} \|\mathbf{F}_1(\mathcal{X}, \mathcal{Z})\|_F^2 + \sum_{i=2}^4 \sum_{k=1}^K \text{tr}(\Pi_{i,k}^T \mathbf{F}_{i,k}(\mathcal{X}, \mathcal{Z})) \\
&+ \frac{\rho}{2} \sum_{i=2}^4 \sum_{k=1}^K \|\mathbf{F}_{i,k}(\mathcal{X}, \mathcal{Z})\|_F^2, \tag{55}
\end{aligned}$$

where \mathcal{X} denotes the set of primal variables \mathbf{w} , \mathbf{p} , \mathbf{V} , \mathbf{U}_k and \mathbf{Z}_k for $k \in [K]$, \mathcal{Z} denotes the set of primal slack variables λ_m , Λ_1 and $\{\Lambda_{i,k}\}_{i=2,3,4}$ for $m \in [M]$ and $k \in [K]$, \mathcal{Y} is the set of dual variables π_m , Π_1 and $\{\Pi_{i,k}\}_{i=2,3,4}$ 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 (46)–(50), $\rho > 0$ is a regularization parameter, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

We iteratively execute the following three steps for ADMM iteration $t = 0, 1, \dots$

$$\mathcal{X}^{t+1} = \arg \min_{\mathcal{X}} \mathcal{L}(\mathcal{X}, \mathcal{Z}^t, \mathcal{Y}^t) \tag{56}$$

$$\mathcal{Z}^{t+1} = \arg \min_{\mathcal{Z}} \mathcal{L}(\mathcal{X}^{t+1}, \mathcal{Z}, \mathcal{Y}^t) \tag{57}$$

$$\begin{cases} \pi_m^{t+1} = \pi_m^t + \rho \mathbf{f}_m(\mathcal{X}^{t+1}, \mathcal{Z}^{t+1}), & \forall m \\ \Pi_1^{t+1} = \Pi_1^t + \rho \mathbf{F}_1(\mathcal{X}^{t+1}, \mathcal{Z}^{t+1}) \\ \Pi_{i,k}^{t+1} = \Pi_{i,k}^t + \rho \mathbf{F}_{i,k}(\mathcal{X}^{t+1}, \mathcal{Z}^{t+1}), & \forall i, k, \end{cases} \tag{58}$$

until both of the conditions $\|\mathcal{X}^{t+1} - \mathcal{X}^t\|_F \leq \epsilon_{\text{admm}}$ and $\|\mathcal{Z}^{t+1} - \mathcal{Z}^t\|_F \leq \epsilon_{\text{admm}}$ are satisfied, where with an abuse of notation, $\|\mathcal{X}\|_F$ denotes the sum of Frobenius norms of variables in \mathcal{X} , and ϵ_{admm} is a stopping tolerance.

Substituting (55) into (56) and completing the squares with respect to primal variables, the \mathcal{X} -minimization problem (56) becomes the unconstrained quadratic program given by (38).

Substituting (55) into (57), the \mathcal{Z} -minimization problem (57) is decomposed into a sequence of subproblems with respect to each of slack variables, given by (41), (43) and (45). ■

REFERENCES

- [1] L. Oliveira and J. Rodrigues, "Wireless sensor networks: A survey on environmental monitoring," *J. Commun.*, vol. 6, no. 2, pp. 143–151, 2011.
- [2] Y. Zou and K. Chakrabarty, "Sensor deployment and target localization in distributed sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 3, no. 1, pp. 61–91, Feb. 2004.
- [3] T. He *et al.*, "Achieving real-time target tracking using wireless sensor networks," in *Proc. IEEE Real Time Technol. Appl. Symp.*, 2006, pp. 37–48.
- [4] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.
- [5] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.
- [6] S. Marano, V. Matta, L. Tong, and P. Willett, "A likelihood-based multiple access for estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5155–5166, Nov. 2007.
- [7] K. Liu and A. M. Sayeed, "Type-based decentralized detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1899–1910, May 2007.
- [8] J. A. Maya, L. R. Vega, and C. G. Galarza, "Optimal resource allocation for detection of a gaussian process using a MAC in WSNs," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2057–2069, Apr. 2015.
- [9] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3822–3832, Jul. 2009.
- [10] G. Thatte and U. Mitra, "Power allocation in linear and tree WSN topologies," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Oct. 2006, pp. 1342–1346.
- [11] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2649–2661, Jul. 2008.
- [12] M. Fanaei, M. C. Valenti, A. Jamalipour, and N. A. Schmid, "Optimal power allocation for distributed blue estimation with linear spatial collaboration," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 5452–5456.
- [13] S. Kar and P. K. Varshney, "Linear coherent estimation with spatial collaboration," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3532–3553, Jun. 2013.
- [14] S. Kar and P. K. Varshney, "On linear coherent estimation with spatial collaboration," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 1448–1452.
- [15] S. Kar and P. K. Varshney, "Controlled collaboration for linear coherent estimation in wireless sensor networks," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, 2012, pp. 334–341.
- [16] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "On optimal sensor collaboration topologies for linear coherent estimation," in *Proc. IEEE Int. Symp. Inf. Theory*, 2014, pp. 2624–2628.
- [17] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "Sparsity-aware sensor collaboration for linear coherent estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2582–2596, May 2015.
- [18] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: Theory and applications for wireless sensor networks," *Comput. Netw.*, vol. 45, no. 3, pp. 245–259, Jun. 2004.
- [19] M. C. Vuran and O. B. Akan, "Spatio-temporal characteristics of point and field sources in wireless sensor networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2006, vol. 1, pp. 234–239.
- [20] Phaeton C. Kyriakidis, "A spatial time series framework for modeling daily precipitation at regional scales," *J. Hydrol.*, vol. 297, no. 4, pp. 236–255, Apr. 2004.
- [21] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
- [22] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *Proc. 46th IEEE Conf. Decision Control*, Dec. 2007, pp. 5492–5498.
- [23] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed kalman filtering based on consensus strategies," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 622–633, May 2008.
- [24] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.
- [25] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc WSNs with noisy links—Part II: Distributed estimation and smoothing of random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1650–1666, Apr. 2008.
- [26] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, pp. 915–936, 2003.
- [27] T. Lipp and S. Boyd, "Variations and extensions of the convex-concave procedure," 2014. [Online]. Available: http://web.stanford.edu/~boyd/papers/pdf/cvx_ccv.pdf

- [28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [29] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [30] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "Optimized sensor collaboration for estimation of temporally correlated parameters," 2016. [Online]. Available: <http://arxiv.org/abs/1603.03448>
- [31] P. Shen, Y. Chen, and Y. Ma, "Solving sum of quadratic ratios fractional programs via monotonic function," *Appl. Math. Comput.*, vol. 212, no. 1, pp. 234–244, 2009.
- [32] F. Bugarin, D. Henrion, and J.-B. Lasserre, "Minimizing the sum of many rational functions," *Math. Program. Comput.*, vol. 8, pp. 83–111, 2015.
- [33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [34] B. Sriperumbudur and G. Lanckriet, "On the convergence of the concave-convex procedure," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1–9.
- [35] I. E. H. Yen, N. Peng, P.-W. Wang, and S.-D. Lin, "On convergence rate of concave-convex procedure," in *Proc. NIPS Optim. Workshop*, 2012, pp. 1–5.
- [36] A. Nemirovski, "Interior point polynomial time methods in convex programming," 2012. [Online]. Available: http://www2.isye.gatech.edu/~nemirovs/Lect_IPM.pdf
- [37] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
- [38] M. Fardad and M. R. Jovanović, "On the design of optimal structured and sparse feedback gains via sequential convex programming," in *Proc. Amer. Control Conf.*, Jun. 2014, pp. 2426–2431.
- [39] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.
- [40] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Operator splitting for conic optimization via homogeneous self-dual embedding," 2013. [Online]. Available: <http://arxiv.org/abs/1312.3039>
- [41] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, Sep. 2015.
- [42] B. He and X. Yuan, "On the $O(1/n)$ convergence rate of the douglas-rachford alternating direction method," *SIAM J. Numer. Anal.*, vol. 50, no. 2, pp. 700–709, 2012.
- [43] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," *J. Sci. Comput.*, vol. 66, no. 3, pp. 889–916, 2016.
- [44] CVX Research, Inc., "CVX: Matlab software for disciplined convex programming, version 2.0," Aug. 2012. [Online]. Available: <http://cvxr.com/cvx>



Sijia Liu (S'13–M'16) received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2008 and 2011, respectively. He received the Ph.D. degree in electrical and computer engineering from Syracuse University, Syracuse, NY, USA, in 2016. He is currently working as a Postdoctoral Research Fellow at the University of Michigan, Ann Arbor, MI, USA. His research interests include statistical signal processing, convex optimization, online learning, object tracking, and resource management for wireless sensor networks. In

2016, he received the Syracuse University Graduate School All University Doctoral Prize and the IEEE SPS Travel Grant Award for ICASSP in 2015. He was among the seven finalists of the Best Student Paper Award at the Asilomar Conference on Signals, Systems, and Computers in 2013. He was the winner of the Nunan research poster competition at Syracuse University in 2012.



Swarnendu Kar (S'07–M'12) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 2004. He received the M.S. degree in mathematics in 2009 and the Ph.D. degree in electrical engineering in 2013, both from Syracuse University, Syracuse, NY, USA. In the past, he has worked for Ittiam Systems Pvt. Ltd., Bangalore, India. He was a Visiting Student at the University of Melbourne, Parkville, Australia, during October–December 2009. He joined the New Devices Group, Intel Corporation, Hillsboro, OR, USA, in September 2012. He is currently involved in the research and development of algorithms related to audio and gestural inputs in wearable devices, on which he has filed several patents. His research interests include detection and estimation theory, distributed and collaborative estimation in sensor networks, and acoustic signal enhancement algorithms. He received the Syracuse University Graduate School All University Doctoral Prize in 2013.



Makan Fardad (M'08) received the B.S. degree from Sharif University of Technology, Tehran, Iran, the M.S. degrees in electrical engineering from Iran University of Science and Technology, Tehran, Iran, and the Ph.D. degree in mechanical engineering from the University of California, Santa Barbara, CA, USA, in 2006. He was a Postdoctoral Associate with the University of Minnesota before joining the Department of Electrical Engineering and Computer Science at Syracuse University in 2008. His research interests include modeling, analysis, and optimization of large-scale dynamical networks.



Pramod K. Varshney (S'72–M'77–SM'82–F'97) was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1972, 1974, and 1976, respectively.

Since 1976, he has been with Syracuse University, Syracuse, NY, USA, where he is currently a Distinguished Professor of electrical engineering and computer science and the Director of CASE: Center for Advanced Systems and Engineering. He is also an Adjunct Professor of radiology at Upstate Medical University, Syracuse. His current research interests include distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing, and remote sensing. He is the author of *Distributed Detection and Data Fusion* (New York, NY, USA: Springer-Verlag, 1997).

Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow while at the University of Illinois. He is a Member of Tau Beta Pi. He received the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the Guest Editor of the Special Issue on Data Fusion of the IEEE Proceedings January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor's Citation for exceptional academic achievement at Syracuse University. He received the IEEE 2012 Judith A. Resnik Award, an honorary Doctor of Engineering degree from Drexel University in 2014, and the ECE Distinguished Alumni Award from UIUC in 2015. He is on the Editorial Boards of the *Journal on Advances in Information Fusion* and *IEEE SIGNAL PROCESSING MAGAZINE*. He was the President of International Society of Information Fusion during 2001.