On Optimal Sensor Collaboration Topologies for Linear Coherent Estimation

Sijia Liu*, Makan Fardad*, Swarnendu Kar†, Pramod K. Varshney*

*Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, 13244 USA
†New Devices Group, Intel Corporation, Hillsboro, Oregon, 97124 USA
*{sliu17, makan, varshney}@syr.edu, †swarnendu.kar@intel.com

***Abstract*—In the context of distributed estimation we consider the problem of sensor collaboration, which refers to the act of sharing measurements with neighboring sensors prior to transmission to a fusion center. While incorporating the cost of sensor collaboration, we aim to find optimal sparse collaboration topologies subject to a certain information or energy constraint. To achieve this goal, we present a tractable optimization framework and propose efficient methods to solve the formulated sensor collaboration problems. The effectiveness of our approach is demonstrated by numerical examples.**

I. Introduction

In this paper, we study the problem of distributed estimation [1], where each sensor reports the local observation of the phenomenon of interest and transmits a *processed* message (after inter-sensor communication) to a fusion center (FC) that determines the global estimate. The act of inter-sensor communication is referred to as sensor collaboration, where sensors are allowed to share their observations with a set of neighboring nodes prior to transmission to the FC. Recently, there has been a growing interest in designing sensor collaboration schemes, spurred by a significant improvement of estimation performance resulting from collaboration [1]–[3].

In [1], the optimal collaboration strategy was studied in an orthogonal multiple access channel (MAC) setting with a fully connected network, where all the sensors are allowed to collaborate. For a coherent MAC, it was shown in [2] that even a partially connected network can yield performance close to that of a fully connected network; examples of partially connected networks include nearest-neighbor and random geometric graphs. However, the work of [1], [2] assumed that there is no cost associated with collaboration, the collaboration topologies are fixed and given in advance, and the only unknowns are the collaboration weights used to combine sensor observations. In [3], the nonzero collaboration cost was taken into account and a greedy algorithm was developed for seeking the optimal collaboration topology in energy constrained sensor networks. This work is an extension of [3]. We present a tractable framework to solve the collaboration problem with nonzero collaboration cost, and propose an efficient solution.

First, we describe collaboration through the collaboration matrix, in which the nonzero entries characterize the collaboration topology and the values of these entries characterize the collaboration weights. We introduce a formulation that simultaneously optimizes both the collaboration topology and the collaboration weights. This is in contrast to [3], where the optimization was performed in a sequential manner. The new formulation results in more efficient allocation of energy resources as evidenced by improved distortion performance in numerical simulations.

Second, in addition to the energy constrained problem considered in [3] where the Fisher information was maximized subject to a total energy budget, we also study an information constrained problem where the energy cost is minimized with a constraint on the Fisher information. For the information constrained problem, we apply the alternating directions method of multipliers (ADMM) [4] to find its locally optimal solutions. For the energy constrained problem, we exploit its relationship with the information constrained problem and propose a bisection algorithm to solve it. Numerical results are provided to illustrate the effectiveness of our proposed methods.

II. Sparsity-Aware Sensor Collaboration

We first introduce the architecture of distributed estimation with sensor collaboration in Fig. 1; see [3] for more details. Based on the system model, the collaboration problems considered in this work are then formulated.

A. System Model

To estimate a random parameter θ, which follows a Gaussian distribution with zero mean and variance η^2, we consider the estimation system shown in Fig. 1, which includes linear sensing, spatial collaboration, coherent-MAC transmission, and linear estimation.

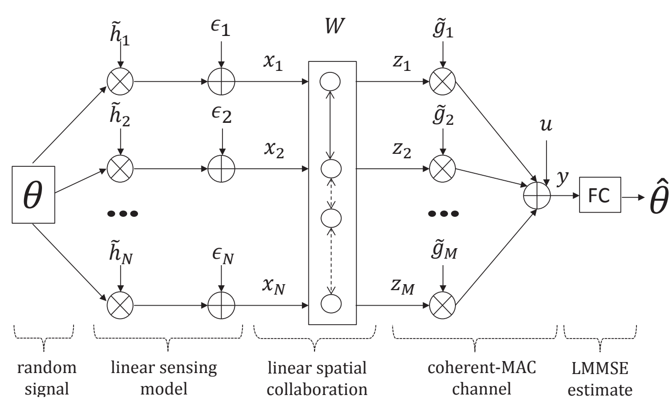

Fig. 1: The collaborative estimation architecture.

The linear sensing model is given by

$$\mathbf{x} = \tilde{\mathbf{h}}\theta + \boldsymbol{\epsilon}, \qquad (1)$$

where $\mathbf{x} = [x_1, \dots, x_N]^T$ denotes the vector of measurements from N sensors, $\tilde{\mathbf{h}}$ is the vector of observation gains with known second order statistics $\mathbb{E}[\tilde{\mathbf{h}}] = \mathbf{h}$ and $\mathrm{cov}(\tilde{\mathbf{h}}) = \mathbf{\Sigma}_h$, and $\boldsymbol{\epsilon}$ represents the vector of zero-mean Gaussian noise with $\mathrm{cov}(\boldsymbol{\epsilon}) = \mathbf{\Sigma}$.

We assume that M predetermined nodes, out of a total of $N \geq M$ sensor nodes, communicate with the FC over a coherent MAC, and each sensor is able to pass its observation to one or more nodes among the M communicating nodes. The *sensor collaboration* process is described by

$$\mathbf{z} = \mathbf{W}\mathbf{x}, \tag{2}$$

where $\mathbf{z} \in \mathbb{R}^M$ denotes the message after collaboration, and $\mathbf{W} \in \mathbb{R}^{M \times N}$ is the collaboration matrix contains weights used to combine sensor measurements. In (2), with a relabelling of the sensors, we assume that the first M sensors communicate with the FC. Note that the *collaboration topology* can be inferred from the nonzero entries of $\mathbf{W}$, and is characterized by the cardinality function

$$\mathrm{card}(W_{mn}) = \begin{cases} 0 & W_{mn} = 0 \\ 1 & W_{mn} \neq 0, \end{cases} \tag{3}$$

where W_{mn} is the entry of the matrix $\mathbf{W}$ at the mth row and nth column. In (3), $\mathrm{card}(W_{mn}) = 0$ indicates no collaboration link from the nth sensor to the mth sensor, and $\mathrm{card}(W_{mn}) = 1$ signifies that the nth sensor shares its observation with the mth sensor. For instance, a matrix $\mathbf{W}$ with $\mathrm{card}(W_{mn}) = 1$ for all $m = 1, \dots, M$ and $n = 1, \dots, N$ corresponds to a fully-connected network. And a matrix $\mathbf{W}$ with $\mathrm{card}(W_{mn}) = 1$ only for $m = n = 1, \dots, M$ implies a distributed network.

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

$$Q_{\mathbf{W}} = \sum_{m=1}^{M} \sum_{n=1}^{N} C_{mn} \mathrm{card}(W_{mn}), \tag{4}$$

where $\mathbf{C}$ is a known cost matrix, and C_{mn} corresponds to the cost of sharing an observation from the nth sensor to the mth sensor. We assume that $C_{mm} = 0$, since each node can collaborate with itself without any cost.

The message $\mathbf{z}$ is transmitted to the FC through a coherent MAC, so that the received signal is given by a coherent sum

$$y = \tilde{\mathbf{g}}^T \mathbf{z} + u, \tag{5}$$

where $\tilde{\mathbf{g}} = [\tilde{g}_1, \dots, \tilde{g}_M]$ denotes the channel gains with known second order statistics $\mathbb{E}[\tilde{\mathbf{g}}] = \mathbf{g} = [g_1, \dots, g_M]^T$ and $\mathrm{cov}(\tilde{\mathbf{g}}) = \mathbf{\Sigma}_g$, and u is a zero-mean Gaussian noise with variance ζ^2. The *transmission cost* is defined by the energy required for transmitting the message $\mathbf{z}$

$$P_{\mathbf{W}} = \sum_{m=1}^{M} \mathbb{E}_{\theta, \tilde{\mathbf{h}}, \boldsymbol{\epsilon}}[z_m^2] = \mathrm{tr}[\mathbf{W}\mathbf{E}_x\mathbf{W}^T], \tag{6}$$

where $\mathbf{E}_x = \mathbf{\Sigma} + \eta^2(\mathbf{h}\mathbf{h}^T + \mathbf{\Sigma}_h)$.

We assume that the FC has full knowledge of the parameters η^2, $\mathbf{h}$, $\mathbf{\Sigma}_h$, $\mathbf{\Sigma}$, $\mathbf{g}$, $\mathbf{\Sigma}_g$ and ζ^2, where the variance and covariance matrices are invertible. Then for estimating the random parameter θ, we consider the linear minimum mean square error (LMMSE) estimator [3]

$$\hat{\theta} = a_{\text{LMMSE}} y, \tag{7}$$

where $a_{\text{LMMSE}} = \frac{\mathbb{E}[y\theta]}{\mathbb{E}[y^2]}$, $\mathbb{E}[y^2] = \mathrm{tr}[\mathbf{E}_g\mathbf{W}\mathbf{E}_x\mathbf{W}^T] + \zeta^2$, $\mathbb{E}[y\theta] = \eta^2\mathbf{g}^T\mathbf{W}\mathbf{h}$, $\mathbf{E}_g = \mathbf{g}\mathbf{g}^T + \mathbf{\Sigma}_g$, and $\mathbf{E}_x$ has been defined in (6). Given the estimator, we adopt an *equivalent* Fisher information [3] as the performance metric

$$J_{\mathbf{W}} = \frac{(\mathbf{g}^T\mathbf{W}\mathbf{h})^2}{\mathrm{tr}[\mathbf{E}_g\mathbf{W}\mathbf{E}_x\mathbf{W}^T] - \eta^2(\mathbf{g}^T\mathbf{W}\mathbf{h})^2 + \zeta^2}, \tag{8}$$

which is monotonically related to the actual estimation distortion (mean square error)

$$D_{\mathbf{W}} = \mathbb{E}[(\theta - a_{\text{LMMSE}}\, y)^2] = \frac{\eta^2}{1 + \eta^2 J_{\mathbf{W}}}. \tag{9}$$

B. Problem Formulation

We concatenate the elements of $\mathbf{W}$ into a vector $\mathbf{w} = [w_1, w_2, \dots, w_L]^T$, where $L = MN$. For $l = 1, 2, \dots, L$, we have $w_l = W_{m_l n_l}$, where $m_l = \lceil \frac{l}{N} \rceil$, $n_l = l - (\lceil \frac{l}{N} \rceil - 1)N$ and $\lceil x \rceil$ is the ceiling function that yields the smallest integer not less than x. Then the collaboration cost (4) can be rewritten as $Q_{\mathbf{w}} = \sum_{l=1}^{L} C_{m_l n_l} \mathrm{card}(w_l)$.

According to [3], the expressions of transmission cost (6) and Fisher information (8) can be equivalently written as,

$$P_{\mathbf{w}} = \mathbf{w}^T\mathbf{\Omega}_{\text{P}}\mathbf{w}, \quad J_{\mathbf{w}} = \frac{\mathbf{w}^T\mathbf{\Omega}_{\text{JN}}\mathbf{w}}{\mathbf{w}^T\mathbf{\Omega}_{\text{JD}}\mathbf{w} + \zeta^2}, \tag{10}$$

where

$$\mathbf{\Omega}_{\text{P}} = \mathbf{I}_M \otimes \mathbf{E}_x, \ \mathbf{I}_M \text{ is the } M \times M \text{ identity matrix}, \tag{11}$$

$$\mathbf{\Omega}_{\text{JN}} = \mathbf{G}\mathbf{h}\mathbf{h}^T\mathbf{G}^T, \quad [\mathbf{G}]_{l,n} = \begin{cases} g_{m_l} & n = n_l, \\ 0 & \text{otherwise}, \end{cases} \tag{12}$$

$$\begin{aligned} \mathbf{\Omega}_{\text{JD}} = \mathbf{G}(\mathbf{\Sigma} + \eta^2\mathbf{\Sigma}_h)\mathbf{G}^T + \eta^2\mathbf{H}\mathbf{\Sigma}_g\mathbf{H}^T + \eta^2\mathbf{\Sigma}_g \otimes \mathbf{\Sigma}_h \\ + \mathbf{\Sigma}_g \otimes \mathbf{\Sigma}, \quad \mathbf{H} = \mathbf{I}_M \otimes \mathbf{h}, \end{aligned} \tag{13}$$

$\mathbf{E}_x$ has been defined in (6), $\mathbf{h}$ is the mean of observation gain, and $\otimes$ denotes the Kronecker product.

Based on the collaboration cost $Q_{\mathbf{w}}$, transmission cost $P_{\mathbf{w}}$, and performance measure $J_{\mathbf{w}}$, we pose the sensor collaboration problems considered in this work as follows.

- *Information constrained* sensor collaboration

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\mathbf{\Omega}_{\text{P}}\mathbf{w} + \sum_{l=1}^{L} C_{m_l n_l}\mathrm{card}(w_l) \\ \text{subject to} & \dfrac{\mathbf{w}^T\mathbf{\Omega}_{\text{JN}}\mathbf{w}}{\mathbf{w}^T\mathbf{\Omega}_{\text{JD}}\mathbf{w} + \zeta^2} \geq J, \end{array} \tag{P$_1$}$$

where $J > 0$ is a given information threshold.

- *Energy constrained* sensor collaboration

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \dfrac{\mathbf{w}^T\mathbf{\Omega}_{\text{JN}}\mathbf{w}}{\mathbf{w}^T\mathbf{\Omega}_{\text{JD}}\mathbf{w} + \zeta^2} \\ \text{subject to} & \mathbf{w}^T\mathbf{\Omega}_{\text{P}}\mathbf{w} + \sum_{l=1}^{L} C_{m_l n_l}\mathrm{card}(w_l) \leq P, \end{array} \tag{P$_2$}$$

where $P \in (0, +\infty)$ is a given energy budget.

Remark 1: To guarantee the feasibility of (P$_1$), the information threshold is *upper bounded*, $J < J_0$, where J_0 is the resulting Fisher information for a *fully* connected collaboration topology with an *infinite* energy budget. The expression of J_0 is given by [3] $J_0 = \mathbf{h}^T(\mathbf{\Sigma} + \eta^2\mathbf{\Sigma}_h)^{-1}\mathbf{h}\left[1 + \frac{1 + \eta^2\mathbf{h}^T(\mathbf{\Sigma} + \eta^2\mathbf{\Sigma}_h)^{-1}\mathbf{h}}{\mathbf{g}^T\mathbf{\Sigma}_g^{-1}\mathbf{g}}\right]^{-1}$. Therefore, the estimation distortion $D_{\mathbf{w}}$ in (9) belongs to (D_0, η^2), where

$D_0 = \frac{\eta^2}{1+\eta^2 J_0}$ denotes the minimum distortion, and η^2 signifies the maximum distortion, which is achieved when $J_{\mathbf{w}} = 0$.

We note that both (P_1) and (P_2) are nonconvex optimization problems not only due to the presence of the cardinality function but also the nonconvexity of the expression for Fisher information. In what follows, we employ the reweighted ℓ_1 norm, the alternating direction method of multipliers, and a bisection algorithm, to find locally optimal solutions of (P_1) and (P_2). The proposed bisection algorithm is in contrast to the heuristic method in [3], where the optimization is performed in a greedy manner. Numerical experiments show that our proposed approach yields more efficient solutions as evidenced by improved estimation performance.

III. Information Constrained Collaboration

Due to the presence of the cardinality function, the proposed sensor collaboration problems are combinatorial in nature. A standard method for solving (P_1) is to replace the cardinality function with the ℓ_1 norm. However, the use of the ℓ_1 norm leads to the undesired dependence on the magnitude of elements in a vector [5]. Therefore, the authors in [5] proposed an iterative *reweighted* ℓ_1 method for optimization. Specifically, we replace $\mathrm{card}(w_l)$ with $\alpha_l^t |w_l|$, where α_l^t is a positive weight corresponding to the optimization variable w_l at the tth iteration. During each iteration, we solve the weighted ℓ_1 minimization problem

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T \boldsymbol{\Omega}_{\text{P}} \mathbf{w} + \|\boldsymbol{\Psi}^t \mathbf{w}\|_1 \\ \text{subject to} & \mathbf{w}^T (J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}})\mathbf{w} + J\zeta^2 \leq 0, \end{array} \qquad (\mathrm{P}_1')$$

where $\boldsymbol{\Psi}^t = \mathrm{diag}(\alpha_1^t C_{m_1 n_1}, \ldots, \alpha_L^t C_{m_L n_L})$ and its solution is denoted by $\mathbf{w}^{t+1}$. Then, we update the weights $\alpha_l^{t+1} = \frac{1}{|w_l^{t+1}|+\nu}$, where the parameter $\nu > 0$ is used to ensure the validity of inversion of zero-value component in $\mathbf{w}^{t+1}$, and $\alpha_l^0{=}1$ for $l{=}1,\ldots,L$. The reweighted ℓ_1-based method typically takes few reweighting iterations to converge [5]. In our numerical experiments, only 4 or 5 iterations are required for satisfactory accuracy. We refer readers to [5] for the detailed algorithm and more applications. Given $\{\alpha_l^t\}_{l=1,\ldots,L}$, the resulting problem (P_1') is a nonconvex optimization problem since the matrix $J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}$ is not positive semidefinite, where both $\boldsymbol{\Omega}_{\text{JD}}$ and $\boldsymbol{\Omega}_{\text{JN}}$ are positive semidefinite. We also remark that to solve the original information constrained problem (P_1), the iterative reweighted ℓ_1 method corresponding to (P_1') is used as the outer loop, and the following ADMM algorithm constitutes the inner loop.

Alternating direction method of multipliers

It has been recently observed in [4], [6] that the alternating direction method of multipliers (ADMM) is a powerful tool in solving optimization problems that involve sparsity-promoting functions (e.g., cardinality functions and ℓ_1 norms). For a nonconvex problem, such as the one considered here, ADMM yields a locally optimal solution when it converges, although its convergence is not guaranteed [4]. However, our numerical experiments and those in other works such as [4], [6] demonstrate that ADMM indeed works well in practice.

We begin by reformulating the optimization problem (P_1') in a way that lends itself to the application of ADMM. After introducing the indicator function corresponding to the constraint set and a new vector variable $\mathbf{v}$, the problem (P_1') can be written as

$$\begin{array}{ll} \underset{\mathbf{w},\mathbf{v}}{\text{minimize}} & \mathbf{w}^T \boldsymbol{\Omega}_{\text{P}} \mathbf{w} + \|\boldsymbol{\Psi} \mathbf{v}\|_1 + \mathcal{I}(\mathbf{w}) \\ \text{subject to} & \mathbf{w} = \mathbf{v}, \end{array} \qquad (14)$$

where we replaced $\boldsymbol{\Psi}^t$ with $\boldsymbol{\Psi}$ for notational simplicity, and the indicator function $\mathcal{I}(\mathbf{w})$ is defined as

$$\mathcal{I}(\mathbf{w}) = \begin{cases} 0 & \text{if } \mathbf{w}^T (J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}})\mathbf{w} + J\zeta^2 \leq 0 \\ +\infty & \text{otherwise.} \end{cases} \qquad (15)$$

Then, the augmented Lagrangian of (14) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \mathbf{v}, \boldsymbol{\lambda}) = {} & \mathbf{w}^T \boldsymbol{\Omega}_{\text{P}} \mathbf{w} + \|\boldsymbol{\Psi} \mathbf{v}\|_1 + \mathcal{I}(\mathbf{w}) \\ & + \boldsymbol{\lambda}^T(\mathbf{w} - \mathbf{v}) + \frac{\rho}{2}\|\mathbf{w} - \mathbf{v}\|_2^2, \end{aligned} \qquad (16)$$

where the vector $\boldsymbol{\lambda}$ is the Lagrangian multiplier, the scalar $\rho > 0$ is a penalty weight, and $\| \cdot \|_2$ denotes the Euclidean norm.

The ADMM algorithm can be described as follows [4]. For $k = 0, 1, \ldots$, we iteratively execute the following three steps

$$\mathbf{w}^{k+1} = \arg\min_{\mathbf{w}} \ \mathcal{L}(\mathbf{w}, \mathbf{v}^k, \boldsymbol{\lambda}^k), \qquad (17)$$

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \ \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}, \boldsymbol{\lambda}^k), \qquad (18)$$

and $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho(\mathbf{w}^{k+1} - \mathbf{v}^{k+1})$, until $\|\mathbf{w}^{k+1} - \mathbf{v}^{k+1}\|_2 \leq \delta$ and $\|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2 \leq \delta$, where δ is a stopping tolerance. To initialize ADMM, we choose $\mathbf{w}^0$ as the value of the optimal collaboration vector within a fully-connected network, see the derivation in [3, Theorem 1], and let $\mathbf{v}^0 = \mathbf{w}^0$.

The rationale behind using ADMM is that we can effectively separate the original non-differentiable problem into a '$\mathbf{w}$-minimization' step (17) and a '$\mathbf{v}$-minimization' step (18), of which the former can be treated as a quadratic optimization problem with only one constraint and the latter can be solved analytically.

1) $\mathbf{w}$*-minimization step:* Completing the squares with respect to $\mathbf{w}$ in (16), the $\mathbf{w}$-minimization step (17) is given by

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T \boldsymbol{\Omega}_{\text{P}} \mathbf{w} + \frac{\rho}{2}\|\mathbf{w} - \mathbf{a}^k\|_2^2 \\ \text{subject to} & \mathbf{w}^T (J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}})\mathbf{w} + J\zeta^2 \leq 0, \end{array} \qquad (19)$$

where we have applied the definition of $\mathcal{I}(\mathbf{w})$ in (15), and $\mathbf{a}^k = \mathbf{v}^k - 1/\rho \boldsymbol{\lambda}^k$. The problem (19) is a nonconvex quadratic program over one inequality quadratic constraint (QP1QC).

To seek the global minimizer of a nonconvex QP1QC, we solve the following semidefinite program (SDP)

$$\begin{array}{ll} \underset{\mathbf{Z} \in \mathbb{R}^{L+1}}{\text{minimize}} & \mathrm{tr}(\mathbf{B}_0 \mathbf{Z}) \\ \text{subject to} & \mathrm{tr}(\mathbf{B}_1 \mathbf{Z}) = 1, \ \ \mathrm{tr}(\mathbf{B}_2 \mathbf{Z}) \leq 0, \ \ \mathbf{Z} \succeq 0, \end{array} \qquad (20)$$

where $\mathbf{B}_0 = \begin{bmatrix} 0 & -\frac{\rho}{2}(\mathbf{a}^k)^T \\ -\frac{\rho}{2}\mathbf{a}^k & \boldsymbol{\Omega}_{\text{P}} + \frac{\rho}{2}\mathbf{I} \end{bmatrix}$, $\mathbf{B}_1 = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and $\mathbf{B}_2 = \begin{bmatrix} J\zeta^2 & \mathbf{0}^T \\ \mathbf{0} & J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}} \end{bmatrix}$.

In (20), if $\mathbf{Z} = \begin{bmatrix} 1 & \mathbf{w}^T \\ \mathbf{w} & \mathbf{w}\mathbf{w}^T \end{bmatrix}$, which implies $\mathrm{rank}(\mathbf{Z}) = 1$, then the SDP is equivalent to the original QP1QC (19). Therefore, problem (20) is a relaxed version of (19), obtained by removing the rank-one constraint. However, it has been shown in [7] that strong duality holds between a nonconvex QP1QC and its SDP relaxation. Further, it has been proved in [8] that a rank-one optimal solution $\mathbf{Z}$ to such an SDP (20) exists. For instances where $\mathbf{Z}$ is not of rank one, a standard rank-one decomposition procedure [9] can be applied to generate the global minimizer of QP1QC from $\mathbf{Z}$. In our extensive experiments, we have observed that the solution to (20) always satisfies $\operatorname{rank}(\mathbf{Z}) = 1$.

2) $\mathbf{v}$-minimization step: Completing the squares with respect to $\mathbf{v}$ in (16), the $\mathbf{v}$-minimization step (18) becomes

$$\underset{\mathbf{v}}{\text{minimize}} \quad \|\mathbf{\Psi}\mathbf{v}\|_1 + \frac{\rho}{2}\|\mathbf{v} - \mathbf{b}^k\|_2^2, \tag{21}$$

where $\mathbf{b}^k = \frac{1}{\rho}\lambda^k + \mathbf{w}^{k+1}$.

The solution of (21) is given by a soft thresholding [6]

$$v_l = \begin{cases} (1 - \frac{\alpha_l^t C_{m_l n_l}}{\rho |b_l^k|}) b_l^k & |b_l^k| > \frac{\alpha_l^t C_{m_l n_l}}{\rho} \\ 0 & |b_l^k| \leq \frac{\alpha_l^t C_{m_l n_l}}{\rho} \end{cases} \tag{22}$$

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

The ADMM algorithm has a fast convergence and typically takes few tens of iterations to converge [4]. In our experiments, the number of iterations is around 40. At each iteration, the computational complexity of ADMM is dominated by the cost of solving the SDP problem (20), which is approximated by $O(M^{4.5}N^{4.5})$ [10], where MN is the dimension of $\mathbf{w}$.

IV. Energy Constrained Collaboration

The energy constrained problem (P_2) is much more difficult due to the nonconvex objective function and the cardinality function in the inequality constraint. Particularly, even if we replace the cardinality function with ℓ_1 norm, the resulting ℓ_1-based energy constrained problem is still hard to solve and does not guarantee a feasible solution to (P_2) because of the change of inequality constraints.

However, if the collaboration topology is given, the collaboration cost $\sum_{l=1}^{L} C_{m_l n_l} \operatorname{card}(w_l)$ is a constant and the constraint in (P_2) becomes a homogeneous quadratic constraint (i.e., no linear term with respect to $\mathbf{w}$ is involved). Therefore, (P_2) becomes a problem with quadratic constraints and an objective that is a ratio of homogeneous quadratic functions. Reference [3] demonstrates that this problem can be solved analytically by solving the KKT optimality conditions.

Setting $\mathbf{w} = \beta\hat{\mathbf{w}}$ for some fixed vector $\hat{\mathbf{w}}$, it can be shown that the objective and constraint functions in (P_2) are strictly increasing functions of β when $\beta > 1$, and strictly decreasing functions of β when $\beta < 1$. Then, from Proposition 1, we can conclude that (P_1) is a "converse formulation" of (P_2).

Proposition 1: Consider the two problems (P_1) and (P_2), and denote their optimal values by $P_{\mathrm{opt}}(J)$ and $J_{\mathrm{opt}}(P)$, respectively. Then, the two problems are converse with each other in the sense that if $J = J_{\mathrm{opt}}(P)$, the optimal solution of (P_1) is equal to the solution of (P_2); if $P = P_{\mathrm{opt}}(J)$, the optimal solution of (P_2) is equal to the solution of (P_1).

Proof: The proof is omitted for brevity and will be reported elsewhere. ■

According to Proposition 1, the solution of the energy constrained problem (P_2) can be obtained by seeking the global minimizer of the information constrained problem (P_1), only if the information threshold in (P_1) is set by using the optimal value of (P_2). However, this methodology is practically intractable since the optimal value of (P_2) is unknown in advance, and problem (P_1) cannot be solved exactly using the methods proposed in Sec. III.

Instead of deriving a solution of (P_2) from its converse problem (P_1), we can infer the collaboration topology of the energy constrained problem (P_2) from the sparsity structure of the solution to the information constrained problem (P_1) using a bisection algorithm. We recall that objective function of (P_2) (in terms of Fisher information) is bounded over an interval $[0, J_0]$; see Remark 1. And there is a one-to-one correspondence between Fisher information and energy budget P in (P_2). Therefore, a bisection procedure can be performed on the interval $[0, J_0]$, and then we solve the information constrained problem to obtain the resulting energy cost and collaboration topology. The procedure terminates if the resulting energy cost is reasonably close to the energy budget P. We summarize the bisection algorithm in Algorithm 1.

Algorithm 1 Bisection algorithm for seeking the optimal collaboration topology of (P_2)

Require: Given $J_l = 0$ and $J_u = J_0$ for the lower bound and upper bound on the Fisher information.

1: **repeat** $J = \frac{J_l + J_u}{2}$
2: For a given information threshold J, solve (P_1) to obtain the collaboration topology (in terms of the sparsity structure of $\mathbf{w}$) and the resulting energy cost $P(J)$.
3: **if** $P(J) < P$ **then** $J_l = J$
4: **else** $J_u = J$
5: **end if**
6: **until** $J_u - J_l < \delta$ or $|P - P(J)| < \delta$

In Step 2 of Algorithm 1, we can apply the proposed ADMM method to solve the information constrained problem. And the bisection procedure is convergence-guaranteed and at most requires $\lceil \log_2(J_0/\delta) \rceil$ iterations. In our experiments, the number of iterations is typically around 10 for $\delta = 10^{-3}$. Once the collaboration topology is obtained, the energy constrained problem can be formulated as a quadratically constrained ratio with homogeneous quadratic functions, which has been analytically solved in [3, Theorem 1]. For the sake of brevity, we refer readers to [3] for more details.

V. Numerical Results

Consider the system model shown in Fig.1, and set the values of system parameters as [3, Example 3], where $\eta^2 = \zeta^2 = 1$, $\mathbf{h} = h_0\sqrt{\alpha_h}\mathbf{1}$, $\mathbf{g} = g_0\sqrt{\alpha_g}\mathbf{1}$, $\mathbf{\Sigma}_h = h_0^2(1 - \alpha_h)\mathbf{I}$, $\mathbf{\Sigma}_g = g_0^2(1 - \alpha_g)\mathbf{I}$ and $\mathbf{\Sigma} = \sigma^2[(1 - \mu)\mathbf{I} + \mu\mathbf{1}\mathbf{1}^T]$. For simplicity, we select $h_0 = g_0 = \sigma^2 = 1$, $\alpha_h = 0.9$, $\alpha_g = 0.8$ and $\mu = 0.5$. The collaboration cost matrix $\mathbf{C}$ is given by $C_{mn} = \tau\|\mathbf{s}_m - \mathbf{s}_n\|_2$, where $\tau = 0.5$ and $\mathbf{s}_i$ is the location of sensor i. To perform the proposed optimization methods, we select $\rho = 20$ as the ADMM parameter and $\delta = 10^{-3}$ for the stopping tolerance.

In Fig. 2, we apply the reweighted ℓ_1-norm based ADMM to solve the information constrained problem (P_1), and compare its performance with that of ADMM with unweighted ℓ_1 norm and an exhaustive search that enumerates all possible sensor collaboration schemes. For the tractability of an exhaustive search, we assume $N = 5$ and $M = 3$. The top plot shows the minimum energy cost as a function of the normalized distortion, which is monotonically related to the information threshold and given by $\frac{D_{\mathbf{w}} - D_0}{\eta^2 - D_0}$; see Remark 1. In this example, $D_0 \approx 0.4$. The rationale behind showing normalized distortion is that it belongs to the interval $(0, 1)$ since $D_0 < D_{\mathbf{w}} < \eta^2$, and that we can compare the effectiveness of collaboration across different operating regimes. As expected, the unweighted ℓ_1-based ADMM yields much worse performance than the reweighted ℓ_1-norm based ADMM. The bottom plot shows the percentage of (active) collaboration links, obtained by $\frac{\mathrm{card}(\mathbf{w})}{MN}$, as a function of the normalized distortion. As we can see, a smaller estimation distortion enables more sensors to collaborate. By contrast, for large values of distortion, the network tends to perform in a distributed manner where no inter-sensor collaboration exists, and thus the energy is allocated *only* for the measurement transmission since each sensor can collaborate with itself at no cost.

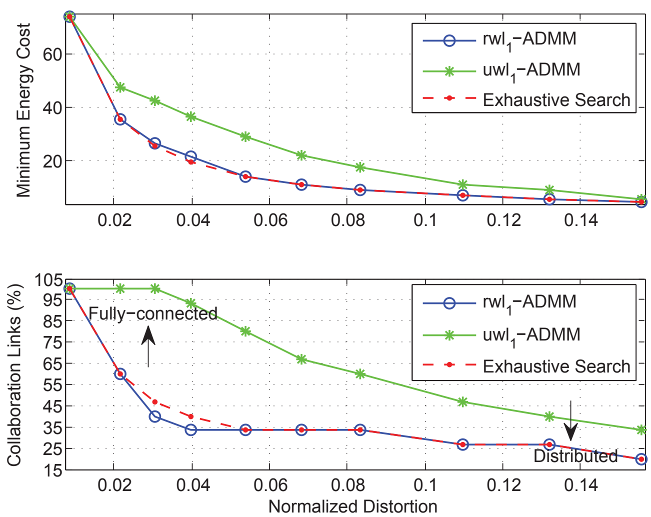

Fig. 2: Performance comparison for solving (P_1).

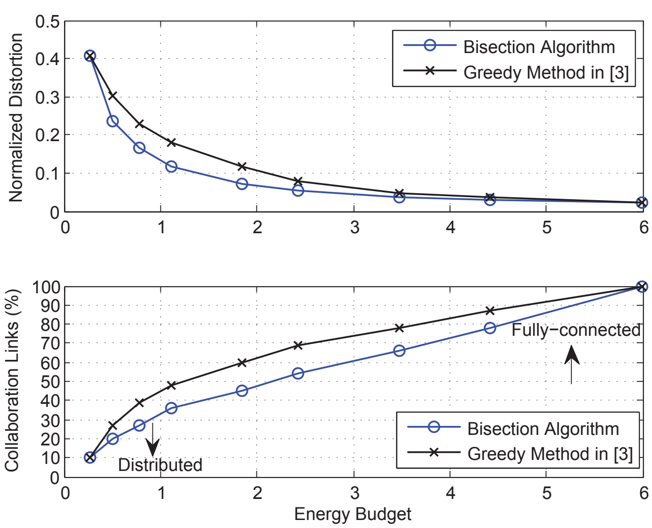

Fig. 3: Performance comparison for solving (P_2).

In Fig. 3, we employ the proposed bisection algorithm to solve the energy constrained problem (P_2) for a relatively large network with $N = M = 10$. We present the resulting estimation distortion and the percentage of (active) collaboration links as a function of the energy budget. Also, we show the results of using the greedy method in [3]. Note that the method in [3] yields worse estimation performance than our approach, even though its resulting number of collaboration links is larger except for two extreme cases: the distributed network and the fully-connected network. This indicates that compared to the number of collaboration links, the optimality of the sparsity structure of $\mathbf{w}$ has a more significant impact on the estimation performance. We also see that the normalized distortion ceases to decrease significantly beyond the activation of 60% collaboration links, which means that only a sparse collaboration network can yield near optimal estimation accuracy.

VI. Conclusion

We studied the problem of distributed estimation with sensor collaboration. By establishing a correspondence between the collaboration topology and the sparsity structure of the collaboration matrix, we explicitly formulated the sensor collaboration problems as sparsity-aware optimization problems. Further, we studied two types of sensor collaboration problems: information constrained problem and energy constrained problem, where we employed the reweighted ℓ_1-norm based ADMM and the bisection algorithm to find their locally optimal solutions. Finally, we presented numerical results to show the effectiveness of our approach by comparing it with state of the art methods. In future work, we will consider the case of estimating a vector parameter, and study sensor collaboration problems with individual energy constraints. Also, we will consider the effects of sensor selection on the performance of distributed estimation with sensor collaboration.

Acknowledgment

This work was supported by U.S. Air Force Office of Scientific Research (AFOSR) under Grants FA9550-10-1-0263 and FA9550-10-1-0458, and National Science Foundation under award CNS-1329885.

References

[1] 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, 2009.

[2] S. Kar and P. K. Varshney, "Controlled collaboration for linear coherent estimation in wireless sensor networks," in *Proc. the 50th Allerton Conf. Commun., Control, and Comput.*, 2012, pp. 334–341.

[3] S. Kar and P. K. Varshney, "Linear coherent estimation with spatial collaboration," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3532–3553, 2013.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[5] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.

[6] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2426–2431, 2013.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[8] Y. Ding, "On efficient semidefinite relaxations for quadratically constrained quadratic programming," *Master thesis, Univ. of Waterloo*, 2007.

[9] Y. Huang, D. A. Maio, and S. Zhang, "Semidefinite programming, matrix decomposition, and radar code design," in *Convex Optimization in Signal Processing and Communications*, pp. 192–228. Cambridge University Press, New York, 2010.

[10] A. Nemirovski, "Interior point polynomial time methods in convex programming," 2012 [Online]. Available: *http://www2.isye.gatech.edu/ nemirovs/Lect_IPM.pdf*.