On Optimal Sensor Collaboration for Distributed Estimation with Individual Power Constraints

Sijia Liu*, Swarnendu Kar†, Makan Fardad*, 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 study the problem of sensor collaboration with individual power constraints, where sensor collaboration refers to the act of sharing measurements with neighboring sensors prior to transmission to a fusion center. In order to find the optimal collaboration strategy consisting of collaboration topology and power allocation scheme, we propose a non-convex formulation in which the estimation distortion is minimized subject to individual power constraints. By exploiting the problem structure, locally optimal collaboration strategies are found via bilinear relaxations and a convex-concave procedure. Numerical examples are provided to show the effectiveness of our approach.**

***Index Terms*—Distributed estimation, sensor collaboration, ℓ_1 norm, bilinear relaxation, convex-concave procedure, networks.**

I. Introduction

In a collaborative estimation system, each sensor accquires a local observation of the phenomenon of interest and transmits a processed message, possibly after inter-sensor collaboration, to a fusion center (FC) that determines the global estimate. The act of inter-sensor collaboration allows sensors to share their observations with a set of neighboring nodes prior to transmission to the FC. In the absence of collaboration, the estimation architecture reduces to a classical distributed estimation network, where sensor measurements are transmitted using an amplify-and-forward strategy [1]. However, it has been shown in [2] that sensor collaboration results in significant improvement of estimation performance.

Design of optimal sensor collaboration schemes for distributed estimation has attracted significant attention [2]–[7]. In [2], the optimal power allocation strategy was found for a fully-connected network, where all the sensors are allowed to collaborate. In [3] and [4], the authors investigated a family of sparsely connected networks which involve sensor collaboration. However, the work [2]–[4] 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 [5], the nonzero collaboration cost was taken into account, and a greedy method was developed for seeking sub-optimal collaboration strategies. Further in [6] and [7], the authors introduced a formulation that simultaneously optimizes both the collaboration topology and the power allocation scheme under a constraint on the *cumulative* power usage. However, the collaboration problem with *individual* power constraints is more practical since each sensor has its own power budget as dictated by the capacity of its battery [5]. In this paper, we focus on solving the sensor collaboration problem under individual power constraints.

Our proposed optimization framework can be regarded as an extension of that in [7], where a locally optimal solution for the cumulative-power constraint problem was found with the help of the solution of its *converse* problem, known as the information constrained problem. This refers to the problem of minimizing the total power consumption subject to a certain estimation quality. However, the approach in [7] is not applicable to addressing the problem with individual power constraints, since the existence of multiple constraints leads to no particular definition of a converse problem. Instead, despite the non-convex property of the individual-constraint problem, we use convex relaxations and a convex-concave procedure to obtain locally optimal collaboration strategies. We numerically compare the performance of our approach with that of the greedy method algorithm [5]. Numerical experiments show that our approach leads to more efficient power allocations as evidenced by improved estimation performance.

The rest of the paper is organized as follows. In Section II, we introduce the collaborative estimation system. In Section III, we formulate the main problem of optimal collaboration with individual power constraints and explore the associated problem structures. In Section IV, we develop an optimization approach to solve the proposed collaboration problem. In Section V, we demonstrate the effectiveness of our approach through numerical examples. Finally, in Section VI we summarize our work and discuss future research directions.

II. System Model

In this paper, the task of the sensor network is to estimate a scalar random parameter with known statistics. The estimation system is depicted in Fig. 1, which includes linear sensing, spatial collaboration, coherent multiple access channel (MAC) transmission, and linear estimation. Based on the system model, we will formalize the concepts of collaboration cost, transmission cost and Fisher information.

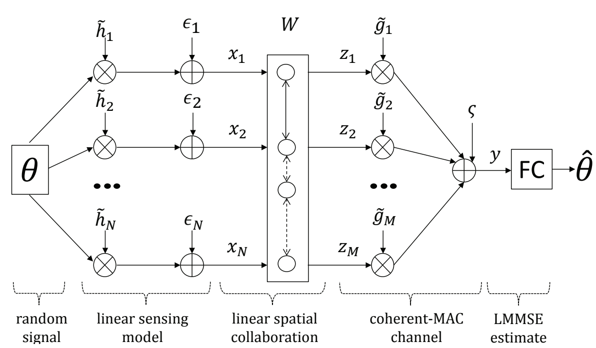

Fig. 1: The collaborative estimation architecture.

The linear sensing model is given by

$$\mathbf{x} = \tilde{\mathbf{h}}\theta + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{x} = [x_1, \ldots, x_N]^T$ denotes the vector of measurements from N sensors, $\tilde{\mathbf{h}}$ is the vector of observation gains with known second order statistics $\mathbb{E}[\tilde{\mathbf{h}}] = \mathbf{h}$ and $\mathrm{cov}(\tilde{\mathbf{h}}) = \mathbf{\Sigma}_h$, θ is the parameter of interest which has zero mean and variance η^2, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$ is the noise vector.

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 with a relabelling of the sensors, we assume that the first M sensors communicate with the FC, $\mathbf{z} \in \mathbb{R}^M$ is the message after collaboration, and $\mathbf{W} \in \mathbb{R}^{M \times N}$ is the collaboration matrix that contains weights used to combine sensor measurements.

In (2), the nonzero entries of $\mathbf{W}$ correspond to the active collaboration links among sensors. For instance, $W_{mn} = 0$ indicates the absence of a collaboration link from the nth sensor to the mth sensor, where W_{mn} is the (m,n)th entry of $\mathbf{W}$. Thus, the sparsity structure of $\mathbf{W}$ characterizes the *collaboration topology*. To account for an active collaboration link, we use the cardinality function

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

According to (3), we introduce the *collaboration cost* of each sensor

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

for $n = 1, 2, \ldots, N$, where we assume that sharing of an observation is realized through a reliable (noiseless) communication link that consumes a known power C_{mn}. Note that $C_{mm} = 0$ since each node can collaborate with itself at no cost.

We then introduce the *transmission cost* of each sensor that communicates with the FC,

$$T_m(\mathbf{W}) = \mathbb{E}_{\theta,\tilde{\mathbf{h}},\boldsymbol{\epsilon}}[z_m^2] = \mathbf{e}_m^T \mathbf{W}\mathbf{\Sigma}_x \mathbf{W}^T \mathbf{e}_m \tag{5}$$

for $m = 1, 2, \ldots, M$, where $\mathbf{e}_m \in \mathbb{R}^M$ is a basis vector with a 1 at the mth coordinate and 0s elsewhere, and $\mathbf{\Sigma}_x = \mathbf{\Sigma}_\epsilon + \eta^2(\mathbf{h}\mathbf{h}^T + \mathbf{\Sigma}_h)$. We note that $T_n(\mathbf{W}) = 0$ if $n > M$, since the nth sensor is not used to communicate with the FC.

After transmitting $\mathbf{z}$ over a coherent MAC, the received signal at the FC is a coherent sum [8]

$$y = \tilde{\mathbf{g}}^T \mathbf{z} + \varsigma, \tag{6}$$

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

To estimate θ, we consider the linear minimum mean square error (LMMSE) estimator [9]

$$\hat{\theta} = a_{\text{LMMSE}}\, y, \text{ with MSE } D_{\mathbf{W}} = \mathbb{E}[(\theta - \hat{\theta})^2], \tag{7}$$

where $a_{\text{LMMSE}} = \frac{\mathbb{E}[y\theta]}{\mathbb{E}[y^2]}$, $\mathbb{E}[y^2] = \mathrm{tr}[\mathbf{\Sigma}_{\tilde{g}} \mathbf{W}\mathbf{\Sigma}_x \mathbf{W}^T] + \xi^2$, $\mathbb{E}[y\theta] = \eta^2 \mathbf{g}^T \mathbf{W}\mathbf{h}$, $\mathbf{\Sigma}_{\tilde{g}} = \mathbf{g}\mathbf{g}^T + \mathbf{\Sigma}_g$, and $\mathbf{\Sigma}_x$ has been defined in (5). We assume that the FC knows the second-order statistics of the observation gain, channel gain, and additive noises, and that the corresponding variance and covariance matrices are invertible.

We now introduce *Fisher information* $J_{\mathbf{W}}$ which is monotonically related to $D_{\mathbf{W}}$ [5],

$$J_{\mathbf{W}} = \frac{1}{D_{\mathbf{W}}} - \frac{1}{\eta^2} = \frac{(\mathbf{g}^T\mathbf{W}\mathbf{h})^2}{\mathrm{tr}[\mathbf{\Sigma}_{\tilde{g}}\mathbf{W}\mathbf{\Sigma}_x\mathbf{W}^T] - \eta^2(\mathbf{g}^T\mathbf{W}\mathbf{h})^2 + \xi^2}. \tag{8}$$

III. Problem Formulation

In this section, we formulate the main sensor collaboration problem that will be considered in the rest of the paper. Note that both the expressions of transmission cost (5) and Fisher information (8) involve quadratic matrix functions[1]. For simplicity of representation, we transform a quadratic matrix function to a quadratic vector function by concatenating the entries of matrix into its vector form. Specifically, a *row-wise* vector $\mathbf{w}$ of the collaboration matrix $\mathbf{W}$ can be written as

$$\mathbf{w} = [w_1, w_2, \ldots, w_L]^T, \quad w_l = W_{m_l n_l}, \tag{9}$$

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

According to (9), the expressions of collaboration cost (4), transmission cost (5), and Fisher information (8) can be converted into functions of the collaboration vector $\mathbf{w}$,

$$\begin{aligned} Q_n(\mathbf{w}) &= \sum_{m=1}^{M} c_{n,m}\,\mathrm{card}(w_{n,m}), \\ T_m(\mathbf{w}) &= \mathbf{w}^T \mathbf{\Omega}_{\text{T},m}\mathbf{w}, \quad J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{\Omega}_{\text{JN}}\mathbf{w}}{\mathbf{w}^T\mathbf{\Omega}_{\text{JD}}\mathbf{w} + \xi^2}, \end{aligned} \tag{10}$$

where for notational simplicity we have used, and henceforth will continue to use $c_{n,m}$ and $w_{n,m}$ instead of $c_{n+(m-1)N}$ and $w_{n+(m-1)N}$, $\mathbf{c}$ is the row-wise vector of the known cost matrix $\mathbf{C}$, and

$$\begin{aligned} &\mathbf{\Omega}_{\text{T},m} = (\mathbf{e}_m\mathbf{e}_m^T) \otimes \mathbf{\Sigma}_x, \ \mathbf{e}_m \text{ and } \mathbf{\Sigma}_x \text{ were defined in (5)}, \\ &\mathbf{\Omega}_{\text{JN}} = \mathbf{G}\mathbf{h}\mathbf{h}^T\mathbf{G}^T, \quad [\mathbf{G}]_{l,n} = \begin{cases} g_{m_l} & n = n_l, \\ 0 & \text{otherwise}, \end{cases} \\ &\mathbf{\Omega}_{\text{JD}} = \mathbf{G}(\mathbf{\Sigma}_\epsilon + \eta^2\mathbf{\Sigma}_h)\mathbf{G}^T + \eta^2\mathbf{H}\mathbf{\Sigma}_g\mathbf{H}^T + \eta^2\mathbf{\Sigma}_g \otimes \mathbf{\Sigma}_h \\ &\qquad + \mathbf{\Sigma}_g \otimes \mathbf{\Sigma}_\epsilon, \quad \mathbf{H} = \mathbf{I}_M \otimes \mathbf{h}, \end{aligned} \tag{11}$$

$\otimes$ denotes the Kronecker product, $\mathbf{h}$ is the mean of observation gain, $\mathbf{\Sigma}_g$, $\mathbf{\Sigma}_h$ and $\mathbf{\Sigma}_\epsilon$ are covariance matrices of channel gains, observation gains and measurement noises, and $\mathbf{I}_M$ is an identity matrix of size M. It is clear from (11) that the matrices $\mathbf{\Omega}_{\text{T},m}$, $\mathbf{\Omega}_{\text{JN}}$ and $\mathbf{\Omega}_{\text{JD}}$ are all positive semidefinite. We refer readers to [5, Sec. III-A] for more details on matrix derivations.

We now pose the problem of optimal sensor collaboration with individual power constraints

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & J(\mathbf{w}) \\ \text{subject to} & Q_n(\mathbf{w}) + T_n(\mathbf{w}) \leq \hat{P}_n, \ n = 1, 2, \ldots, N, \end{array} \tag{P0}$$

where recalling from (5) that $T_n(\mathbf{w}) = 0$ if $n > M$, and $\hat{P}_n$ is a known power budget at the nth sensor. As will be evident later, (P0) is a non-convex optimization problem. However, if the *collaboration topology* is *given*, the individual collaboration costs $\{Q_n(\mathbf{w})\}_{n=1,2,\ldots,N}$ are constants and the constraints of (P0) become homogeneous quadratic constraints (i.e., no linear term with respect to $\mathbf{w}$ is involved). As a result, (P0) can be cast as a quadratically constrained ratio with homogeneous quadratic functions. It has been shown in [5] and [10] that the globally optimal solution of such a problem can be obtained via semidefinite programming. In brief, (P0) is solvable if the collaboration topology is given. Motivated by that, our objective is to seek a good approximate collaboration topology by solving the individual-constraint problem under a convex optimization lens.

[1]A quadratic matrix function is a function $f : \mathbb{R}^{n \times r} \to \mathbb{R}$ of the form $f(\mathbf{X}) = \mathrm{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) + 2\,\mathrm{tr}(\mathbf{B}^T\mathbf{X}) + c$, where $\mathbf{A} \in \mathbb{R}^n$ is a symmetric matrix, $\mathbf{B} \in \mathbb{R}^{n \times r}$ and $c \in \mathbb{R}$.

We begin by elaborating on the problem structures of (P0) via its epigraph form

$$\begin{aligned} \underset{\mathbf{w},t}{\text{maximize}} \quad & t^2 \\ \text{subject to} \quad & Q_n(\mathbf{w}) + T_n(\mathbf{w}) \leq \hat{P}_n, \ n = 1, 2, \ldots, N \\ & J(\mathbf{w}) \geq t^2, \end{aligned} \tag{12}$$

where t is an auxiliary optimization variable, and we use the fact that $J(\mathbf{w}) \geq 0$. It is not diffcult to show that the *quadratic* objective function of (12) can be replaced with a *linear* function t without loss of performance. After specifying $J(\mathbf{w})$ and imposing a new variable $\mathbf{v}$ and constraint $\mathbf{v} = \mathbf{w}t$, problem (12) is then equivalent to

$$\underset{\mathbf{w},\mathbf{v},t}{\text{minimize}} \quad -t$$

$$\text{subject to} \quad Q_n(\mathbf{w}) + T_n(\mathbf{w}) \leq \hat{P}_n, \ n = 1, 2, \ldots, N \tag{13a}$$

$$\mathbf{v} = \mathbf{w}t \tag{13b}$$

$$\mathbf{v}^T \boldsymbol{\Omega}_{\text{JD}} \mathbf{v} + \xi^2 t^2 - \mathbf{w}^T \boldsymbol{\Omega}_{\text{JN}} \mathbf{w} \leq 0. \tag{13c}$$

It is clear from (13) that there exist three nonconvex sources: a) the cardinality involved collaboration cost $Q_n(\mathbf{w})$, b) bilinear function $\mathbf{w}t$, c) the difference of convex quadratic functions in (13c). In what follows, we will apply relaxations and approximations to overcome the difficulties posed by nonconvexities in (13).

IV. Convex Relaxations and Convex-Concave Procedure

Due to the presence of the cardinality function (also known as ℓ_0 norm), the sensor collaboration problem is combinatorial in nature. A method for solving (13) is to replace the cardinality function with a weighted ℓ_1 norm [11]. This leads to the following optimization problem

$$\begin{aligned} \underset{\mathbf{w},\mathbf{v},t}{\text{minimize}} \quad & -t \\ \text{subject to} \quad & \tilde{Q}_n(\mathbf{w}) + T_n(\mathbf{w}) \leq \hat{P}_n, \ n = 1, 2, \ldots, N \\ & \text{bilinear constraint (13b)} \\ & \text{nonconvex inequality (13c)} \end{aligned} \tag{14}$$

where $\tilde{Q}_n(\mathbf{w}) = \sum_{m=1}^{M} c_{n,m} \alpha_{n,m}^k |w_{n,m}|$, $\{\alpha_{n,m}^k\}$ are the weights that are iteratively updated in order to make $\tilde{Q}_n(\mathbf{w})$ a good approximation for $Q(\mathbf{w})$, and $|\cdot|$ denotes the absolute value of a scalar. We summarize the reweighted ℓ_1 method for solving (13) in Algorithm 1.

Remark 1: *Let us define*

$$\begin{cases} \boldsymbol{\Omega}_{C,n} := \text{diag}(c_{n,1}\alpha_{n,1}^k, c_{n,2}\alpha_{n,2}^k, \ldots, c_{n,M}\alpha_{n,M}^k) \\ \boldsymbol{\Upsilon}_n := [\mathbf{e}_n, \mathbf{e}_{n+N}, \ldots, \mathbf{e}_{n+(M-1)N}]^T \in \mathbb{R}^{M \times L}, \end{cases} \tag{15}$$

where $\text{diag}(\cdot)$ *denotes a diagonal matrix, and* $\mathbf{e}_i \in \mathbb{R}^L$ *is the basis vector mentioned in* (5)*. According to* (15)*, the* ℓ_1 *relaxation* $\tilde{Q}_n(\mathbf{w})$ *in* (14) *can be compactly expressed as*

$$\tilde{Q}_n(\mathbf{w}) = \|\boldsymbol{\Omega}_{C,n} \boldsymbol{\Upsilon}_n \mathbf{w}\|_1. \tag{16}$$

Algorithm 1 Reweighted ℓ_1 method for solving (13)

Require: given $\varepsilon > 0$, $\epsilon_{\text{rw}} > 0$ and $K > 0$. Set $\alpha_{n,m}^0 = 1$ for all $(n,m) \in \{1, 2, \ldots, N\} \times \{1, 2, \ldots, M\}$.

1: **for** $k = 0, 1, \ldots, K$ **do**
2: solve (14) to obtain solution $\mathbf{w}^k$ from Algorithm 2.
3: update the weights $\alpha_{n,m}^{k+1} = \dfrac{1}{|w_{n,m}^k| + \varepsilon}$.
4: if $\|\mathbf{w}^{k+1} - \mathbf{w}^k\|_2 < \epsilon_{\text{rw}}$ or $k = K$, **quit**.
5: **end for**

Reference [11] shows that much of the benefit of using the reweighted ℓ_1 method is gained from its first few iterations. In Step 2 of Algorithm 1, the implementation of Algorithm 2 for solving (14) will be elaborated on later. In Step 3 of Algorithm 1, the positive scalar ε is a small number which ensures that the denominator is always nonzero and helps the convergence of Algorithm 1; for example, if $w_{n,m}^k \to 0$, the weight $\alpha_{n,m}^{k+1}$ converges to $\frac{1}{\varepsilon}$.

After ℓ_1 relaxation, it is still a nontrivial exercise to solve problem (14) due to the presence of nonconvex constraints (13b) and (13c).

A. Bilinear Relaxation

In nonconvex optimization, a commonly used method for handling the bilinear function is to replace it with its affine approximation over a trust region. However, this methodology could bring in two issues [12]: a) infeasibility of the approximate convex problem, and b) difficulty of choosing a proper size of trust region. To mitigate these issues, we relax the bilinear equality (13b) to a sequence of linear inequalities. To elaborate on the bilinear relaxation, we begin by reviewing a property of bilinear function.

Lemma 1: Given $\mathcal{C} = \{\check{\omega} \leq \omega \leq \hat{\omega}, \check{\nu} \leq \nu \leq \hat{\nu}\} \subset \mathbb{R}^2$, then the bilinear function $\omega\nu$ satisfies

$$\text{vex}(\omega\nu) \leq \omega\nu \leq \text{cav}(\omega\nu), \tag{17}$$

where $\text{vex}(\omega\nu)$ and $\text{cav}(\omega\nu)$ are convex and concave envelops

$$\begin{cases} \text{vex}(\omega\nu) = \max\{\check{\nu}\omega + \check{\omega}\nu - \check{\omega}\check{\nu}, \hat{\nu}\omega + \hat{\omega}\nu - \hat{\omega}\hat{\nu}\} \\ \text{cav}(\omega\nu) = \min\{\hat{\nu}\omega + \check{\omega}\nu - \check{\omega}\hat{\nu}, \check{\nu}\omega + \hat{\omega}\nu - \hat{\omega}\check{\nu}\}, \end{cases} \tag{18}$$

and $\text{vex}(\omega\nu) = \omega\nu = \text{cav}(\omega\nu)$ for all (ω, ν) at the boundary of $\mathcal{C}$.

Proof: See proofs of Theorems 6, 7, and 10 in [13]. ■

We next show that the optimization variables $\mathbf{w}$ and t of (12) are bounded such that the bilinear functions $\mathbf{w}t$ satisfies the property of Lemma 1.

Proposition 1: The lower and upper bounds of variables t and $\mathbf{w}$ of (12) are given by

$$\begin{aligned} & 0 \leq t \leq \sqrt{\lambda_{\max}(\boldsymbol{\Omega}_{\text{JN}}, \boldsymbol{\Omega}_{\text{JD}} + \xi^2 \boldsymbol{\Omega}_{\text{T}}/\hat{P})} \\ & -\sqrt{\hat{P}\mathbf{e}_l^T \boldsymbol{\Omega}_{\text{T}}^{-1} \mathbf{e}_l} \leq w_l \leq \sqrt{\hat{P}\mathbf{e}_l^T \boldsymbol{\Omega}_{\text{T}}^{-1} \mathbf{e}_l}, \end{aligned} \tag{19}$$

for $l = 1, 2, \ldots, L$, where $\boldsymbol{\Omega}_{\text{T}} = \sum_{m=1}^{M} \boldsymbol{\Omega}_{\text{T},m}$, $\hat{P} = \sum_{n=1}^{N} \hat{P}_n$, $\lambda_{\max}(\mathbf{A}, \mathbf{B})$ denotes the maximum eigenvalue of the generalized eigenvalue problem $\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v}$, and $\mathbf{e}_l$ is the basis vector with a 1 at the lth coordinate and 0s elsewhere.

Proof: See Appendix. ■

According to Lemma 1 and Proposition 1, we can relax the bilinear constraint (13b) to inequalities

$$\text{vex}(w_l t) \leq v_l \leq \text{cav}(w_l t), \ l = 1, 2, \ldots, L, \tag{20}$$

where the functions $\text{vex}(w_l t)$ and $\text{cav}(w_l t)$ are determined by (18) and (19). Due to the features of pointwise minimum and maximum in (18), inequalities (20) can be rewritten as $4L$ linear inequalities.

Substituting (13b) with (20), we relax problem (14) as

$$\begin{aligned} \underset{\mathbf{w},\mathbf{v},t}{\text{minimize}} \quad & -t \\ \text{subject to} \quad & \tilde{Q}_n(\mathbf{w}) + T_n(\mathbf{w}) \leq \hat{P}_n, \ n = 1, 2, \ldots, N \\ & \text{vex}(w_l t) \leq v_l \leq \text{cav}(w_l t), \ l = 1, 2, \ldots, L \\ & \text{nonconvex inequality (13c).} \end{aligned} \tag{21}$$

Problem (21) is now convex except for the last nonconvex inequality. However, the difference of two convex quadratic functions in (13c) can be efficiently handled via a convex-concave procedure [14], [15].

B. Convex-concave procedure

We recall that in (21), the nonconvex inequality (13c) is

$$\mathbf{v}^T \boldsymbol{\Omega}_{\text{JD}} \mathbf{v} + \xi^2 t^2 \leq \mathbf{w}^T \boldsymbol{\Omega}_{\text{JN}} \mathbf{w}, \tag{22}$$

where $\boldsymbol{\Omega}_{\text{JD}}$ and $\boldsymbol{\Omega}_{\text{JN}}$ are positive semidefinite; see (11). We linearize the right hand side of (22) around a feasible point $\bar{\mathbf{w}}$

$$\mathbf{v}^T \boldsymbol{\Omega}_{\text{JD}} \mathbf{v} + \xi^2 t^2 \leq g(\mathbf{w}; \bar{\mathbf{w}}), \tag{23}$$

where $g(\mathbf{w}; \bar{\mathbf{w}}) := \bar{\mathbf{w}}^T \boldsymbol{\Omega}_{\text{JN}} \bar{\mathbf{w}} + 2\bar{\mathbf{w}}^T \boldsymbol{\Omega}_{\text{JN}} (\mathbf{w} - \bar{\mathbf{w}})$. Note that $g(\mathbf{w}; \bar{\mathbf{w}})$ is an affine *lower bound* on the convex function $\mathbf{w}^T \boldsymbol{\Omega}_{\text{JN}} \mathbf{w}$. This implies that the set of $\mathbf{w}$ that satisfy (23) is a strict subset of the set of $\mathbf{w}$ that satisfy (22).

By replacing (22) with (23), we obtain a 'restricted' convex version of problem (21). By iteratively updating the linearizing point $\bar{\mathbf{w}}$, we solve a sequence of convex problems. This methodology is first proposed by [14] and named concave-convex procedure (or convex-concave procedure). Recently, a more robust method called *penalty* convex-concave procedure was presented in [15], which allows us to initialize the algorithm with an infeasible linearizing point. To be specific, we convexify (21) as

$$\begin{array}{ll} \underset{\mathbf{w},\mathbf{v},t,u}{\text{minimize}} & -t + \gamma u \\ \text{subject to} & \tilde{Q}_n(\mathbf{w}) + T_n(\mathbf{w}) \leq \hat{P}_n, \ n = 1, 2, \ldots, N \\ & \text{vex}(w_l t) \leq v_l \leq \text{cav}(w_l t), \ l = 1, 2, \ldots, L \\ & \mathbf{v}^T \boldsymbol{\Omega}_{\text{JD}} \mathbf{v} + \xi^2 t^2 - g(\mathbf{w}; \bar{\mathbf{w}}) \leq u, \\ & u \geq 0, \end{array} \tag{24}$$

where $\gamma > 0$ is a known regularization parameter, and the new variable u provides the tolerance of violating the last inequality constraint. Clearly, problem (24) is convex, and as γ grows, the variable u approaches to zero and the inequality (23) is recovered.

We note that the convex problem (24) involves an ℓ_1 norm due to the presence of $\tilde{Q}_n(\mathbf{w})$ given in (16). However, we can eliminate the ℓ_1 norm by introducing auxiliary variables and constraints such that problem (24) is cast as a quadratically constrained convex program [16]

$$\begin{array}{ll} \underset{\mathbf{w},\mathbf{v},\mathbf{r},t,u}{\text{minimize}} & -t + \gamma u \\ \text{subject to} & \mathbf{1}^T \mathbf{r}_n + T_n(\mathbf{w}) \leq \hat{P}_n, \ n = 1, 2, \ldots, N \\ & -\mathbf{r}_n \leq \boldsymbol{\Omega}_{\text{C},n} \boldsymbol{\Upsilon}_n \mathbf{w} \leq \mathbf{r}_n, \ n = 1, 2, \ldots, N \\ & \text{vex}(w_l t) \leq v_l \leq \text{cav}(w_l t), \ l = 1, 2, \ldots, L \\ & \mathbf{v}^T \boldsymbol{\Omega}_{\text{JD}} \mathbf{v} + \xi^2 t^2 - g(\mathbf{w}; \bar{\mathbf{w}}) \leq u, \\ & u \geq 0, \end{array} \tag{25}$$

where $\mathbf{r}_n \in \mathbb{R}^M$ and $\mathbf{r} = [\mathbf{r}_1^T, \ldots, \mathbf{r}_N^T]$ is a vector of auxiliary variables. We now summarize the penalty convex-concave procedure for solving problem (21) in Algorithm 2.

Algorithm 2 Penalty convex-concave procedure

Require: given an initial point $\mathbf{w}^0$, $t^0 = -\infty$, $u^0 = \infty$, $\gamma > 0$, $\gamma_{\max} > 0$, $\mu > 1$, and $\epsilon_{\text{ccp}} > 0$.
1: **for** iteration $s = 1, 2, \ldots$ **do**
2: set $\bar{\mathbf{w}} = \mathbf{w}^{s-1}$.
3: set $(\mathbf{w}^s, t^s, u^s)$ as the solution of problem (25).
4: update γ by setting $\gamma = \min\{\mu\gamma, \gamma_{\max}\}$.
5: **until** $|t^{s-1} - t^s| \leq \epsilon_{\text{ccp}}$.
6: **end for**

We remark that when $\gamma = \gamma_{\max}$, Algorithm 2 reduces to a standard convex-concave procedure, and therefore its convergence is guaranteed [15]. In our extensive numerical examples, Algorithm 2 converges fast and typically takes a few iterations. In Step 2 of Algorithm 2, problem (25) can be solved via the interior-point algorithm, which takes the computational complexity $O(L^{3.5})$ [17, Sec. 10].

In summary, for solving the original problem (P0) we first replace the cardinality function with the weighted ℓ_1 norm, which yields the nonconvex problem (14). We then use the bilinear relaxation and convex-concave procedure to tackle problem (14). Since the ℓ_1 approximation and bilinear relaxation were used to solve (P0), the resulting collaboration vector may not satisfy the original individual power constraints. However, we can infer the *collaboration topology* from the sparsity structure of the obtained collaboration vector. Once the collaboration topology is given, the individual-constraint problem becomes a quadratically constrained ratio with homogeneous quadratic functions, which can be solved by [10, Theorem 1].

V. Numerical Results

Consider the system model shown in Fig.1, where we assume that the network is homogeneous and equicorrelated with $N = M = 10$ sensors deployed on a 10×10 grid. As in [7, Sec. VII], the system parameters are modeled by $\mathbf{h} = h_0\sqrt{\alpha_h}\mathbf{1}$, $\boldsymbol{\Sigma}_h = h_0^2(1-\alpha_h)\mathbf{I}$, $\boldsymbol{\Sigma}_\epsilon = \zeta^2[(1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T]$, $\mathbf{g} = g_0\sqrt{\alpha_g}\mathbf{1}$, and $\boldsymbol{\Sigma}_g = g_0^2(1-\alpha_g)\mathbf{I}$, where we set $h_0 = g_0 = 1$, $\alpha_h = \alpha_g = 0.7$, $\rho = 0.5$, $\zeta^2 = \xi^2 = 1$, $\eta^2 = 0.1$. The (m,n)th entry of the collaboration cost matrix is given by $C_{mn} = 0.1\|\mathbf{s}_m - \mathbf{s}_n\|_2$, where $\mathbf{s}_i$ is the location of the ith sensor. The power budgets at the N sensors are

$$\hat{P}_1 = P_0, \hat{P}_2 = \varrho P_0, \ldots, \hat{P}_N = \varrho^{N-1} P_0,$$

where P_0 is a constant, $\hat{P} = P_0 \sum_{n=1}^N \varrho^n$ gives the total power budget, and $\varrho \in [0,1]$ is a skewness parameter which measures the dissimilarity of individual powers [5]. For example, the value $\varrho = 0$ implies that only one of the sensors has all the power while other sensors have no power at all. And the value $\varrho = 1$ implies that all sensors have equal power budgets. While employing Algorithm 1 and 2, we select $\varepsilon = 0.1$, $K = 50$, $\epsilon_{\text{rw}} = 0.01$, $\gamma = 0.1$, $\gamma_{\max} = 100$, $\mu = 1.1$, and $\epsilon_{\text{ccp}} = 10^{-3}$.

For a better depiction of estimation performance, we define the *normalized distortion* $\frac{D_{\mathbf{w}} - D_0}{\eta^2 - D_0} \in (0,1]$, where $D_{\mathbf{w}}$ is monotonically related to the value of Fisher information $J_{\mathbf{w}}$ as shown in (8), D_0 is the lowest estimation distortion which corresponds to the upper bound of Fisher information for an infinite power budget, namely, $\lambda_{\max}(\boldsymbol{\Omega}_{\text{JN}}, \boldsymbol{\Omega}_{\text{JD}})$ in (19), and η^2 signifies the maximum distortion which is only determined by the prior information of θ while $J_{\mathbf{w}} = 0$. Further to characterize the number of established collaboration links, we consider the percentage of collaboration links $\frac{\sum_{l=1}^{MN} \text{card}(w_l) - M}{MN - M} \times 100$ (%).

In Fig. 2, we present the estimation performance and number of collaboration links as functions of the total power budget $\hat{P}$ for different skewness conditions $\varrho = 1$, 0.6, and 0.2. As we can see, a higher power budget leads to a lower estimation distortion and more collaboration links. By fixing the value of $\hat{P}$ and varying the skewness parameter, we note that the estimation performance improves by increasing ϱ. This is because a larger value of ϱ corresonds to less skewed power constraints, which result in more effective collaboration links and lower estimation error in the homogeneous network.

In Fig. 3, we compare the performance of our approach with that of the greedy method in [5] when $\varrho = 0.8$. As we can see, our approach outperforms the greedy method in terms of estimation error. This is because the latter easily gets trapped in unexpected local optima. Indeed, the bottom plots of Fig. 3 show that even though a large

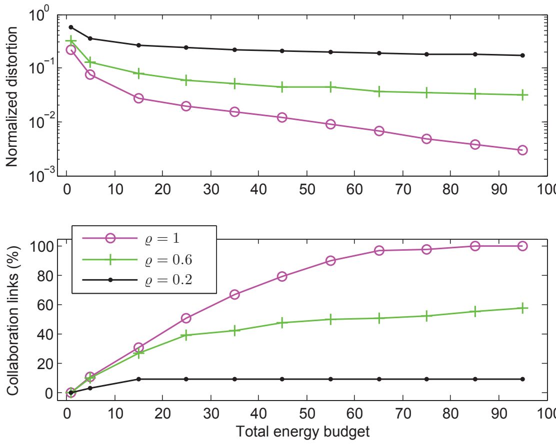

Fig. 2: Estimation error and number of collaboration links versus total power budgets for different skewed power constraints.

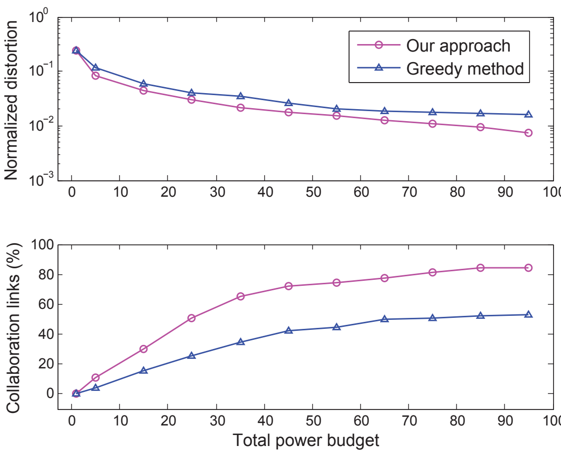

Fig. 3: Performance comparison for our approach and the greedy method in [5].

power budget is assigned, the number of collaboration links increases slightly for the greedy method.

VI. Conclusion

We studied the problem of sensor collaboration with individual power constraints for linear coherent estimation. We showed that the problem is nonconvex and combinatorial in nature. By exploiting problem structure, we convexified the individual-constraint problem by using reweighted ℓ_1 approximation, bilinear relaxation and the convex-concave procedure. The convexified problem renders a locally optimal collaboration topology as well as collaboration weights. Numerical results were provided to show the effectiveness of our approach. In future work, we will take collaboration noise into account. We will also explore the problem of sensor collaboration for estimating a time-varying parameter.

Acknowledgment

This work was supported by U.S. Air Force Office of Scientific Research (AFOSR) under grant FA9550-10-1-0458, and National Science Foundation under awards CNS-1329885 and CMMI-0927509.

Appendix: Proof of Proposition 1

It is clear from (12) that the maximum value of t^2 is equal to the maximum value of $J(\mathbf{w})$ in (P0). By dropping collaboration costs and combining transmission costs, we can relax (P0) to

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & J(\mathbf{w}) \\ \text{subject to} & \mathbf{w}^T \boldsymbol{\Omega}_{\text{T}} \mathbf{w} \leq \hat{P}, \end{array} \qquad (26)$$

where $\boldsymbol{\Omega}_{\text{T}}$ and $\hat{P}$ were defined in (19). Due to the relaxation, the optimal value of (26) is an upper bound on the optimal value of (P0), and is given by [5, Theorem 1], $\lambda_{\max}(\boldsymbol{\Omega}_{\text{JN}}, \boldsymbol{\Omega}_{\text{JD}} + \frac{\xi^2}{\hat{P}} \boldsymbol{\Omega}_{\text{T}})$, which provides the upper bound of t. On the other hand, we can restrict $t \geq 0$ since $J(\mathbf{w}) \geq 0$. Similarly, the lower and upper bounds of w_l can be computed through the optimal values of relaxed problems

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{e}_l^T \mathbf{w} \\ \text{subject to} & \mathbf{w}^T \boldsymbol{\Omega}_{\text{T}} \mathbf{w} \leq \hat{P}, \end{array} \quad \text{and} \quad \begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{e}_l^T \mathbf{w} \\ \text{subject to} & \mathbf{w}^T \boldsymbol{\Omega}_{\text{T}} \mathbf{w} \leq \hat{P}, \end{array}$$

where by studying KKT conditions we obtain the result of Prop. 1.

References

[1] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, 2007.

[2] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3822–3832, 2009.

[3] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2649–2661, July 2008.

[4] S. Kar and P.K. Varshney, "Controlled collaboration for linear coherent estimation in wireless sensor networks," in *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 334–341.

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

[6] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "On optimal sensor collaboration topologies for linear coherent estimation," in accepted *IEEE International Symposium on Information Theory (ISIT)*, 2014.

[7] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "Sparsity-aware sensor collaboration for linear coherent estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2582–2596, May 2015.

[8] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 757–770, 2008.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[10] F. Jiang, J. Chen, and A. L. Swindlehurst, "Optimal power allocation for parameter tracking in a distributed amplify-and-forward sensor network," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2200–2211, May 2014.

[11] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[12] S. Boyd, "Sequential convex programming, lecture notes for ee364b: Convex optimization ii, stanford university," http://stanford.edu/class/ee364b/lectures/seq_slides.pdf, 2011.

[13] F. A. Al-Khayyal, "Jointly constrained bilinear programs and related problems: An overview," *Computers & Mathematics with Applications*, vol. 19, no. 11, pp. 53–62, 1990.

[14] A. L. Yuille and Anand Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003.

[15] T. Lipp and S. Boyd, "Variations and extensions of the convex-concave procedure," http://web.stanford.edu/~boyd/papers/pdf/cvx_ccv.pdf, 2014.

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

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