Joint Transceiver Design for **Linear** MMSE Data Fusion in Coherent MAC Wireless Sensor Networks

Yang Liu^{a,*}, Jing Li^a, Xuanxuan Lu^a

^a*Electrical and Computer Engineering Department, Lehigh University, Bethlehem, P.A. United States*

Abstract

This paper considers the design of minimal mean square error (MMSE) transceivers in a wireless sensor network. The problem is nonconvex and challenging, and previous results (with partial solutions and/or with convergence unproved) left much to be desired. Here we propose several approaches — 2 block coordinate descent (BCD), essentially cyclic multi-block method and its variants and distributive method to solve this problem. The proposed 2-BCD approach formulates the subproblem of joint beamformer optimization as a general second-order cone programming problem, which lends itself to standard numerical solvers and which requires no extra assumptions like previous works do. The proposed essentially cyclic multi-block approach further decomposes the joint beamformer design subproblem into multiple blocks, and rigorously solves each with semi-closed-form solution. The distributive algorithm optimizes transmitters in a decentralized manner and has never been considered in existing literature. The distributive algorithm has time complexity independent of number of sensors and is especially suitable for large-scale networks. All the previous BCD-based approaches left some singularity issue unattended as well as the convergence property unaddressed, our proposals are the first to provide a complete and provably-converging analytical solution. Extensive analysis and simulations demonstrate the merits of the novel approaches relative to existing alternatives.

Keywords: Wireless Sensor Network, Data Fusion, Signal Estimation, Linear

^{*}Supported by National Science Foundation under Grants No.1343372.

^{*}Corresponding author

Email addresses: liuocean613@gmail.com (Yang Liu), jingli@ece.lehigh.edu (Jing Li), xul311@lehigh.edu (Xuanxuan Lu)

1. Introduction

Consider a typical wireless sensor network comprised of a fusion center (FC) and multiple sensors that are spatially distributed and wirelessly connected to provide surveillance of some physical event. After harvesting information from the environment, these sensors transmit distorted observations to the FC for data fusion. A central problem underlying this type of systems is how to design the sensors and the FC to collaboratively accomplish sensing, communication and fusion tasks in an efficient and trust-worthy manner.

In this paper we focus on the systems where all sensors and the fusion center utilize linear filters to perform signal processing. In fact it is shown in [1] and [2] that for Gaussian signaling model, where the statistics of source signal and the noise at sensors and fusion center are all Gaussian, the linear filtering scheme can achieve optima. Although the optimal filtering schemes for generic non-Gaussian systems are unknown, the linear filtering are still relatively appealing due to its easiness of analysis and implementation. When the sensors and the FC are all equipped with multiple antennas and linear filters, this problem may be regarded as a cooperative multi-input multi-output (MIMO) beamformer design problem, which has been tackled from a variety of perspectives [3, 4, 5, 6, 8, 7, 9, 10]. Early papers focused on orthogonal multiple access channels (MAC) with simple models involving either no power constraint (i.e. ideal channels) [3, 4] or a lump sum power constraint on scalar sources [5, 6]. Recent studies have looked into more sophisticated models, and have considered vector sources, coherent MAC and/or individual power constraints [7, 9, 12, 13, 14], but the results are restricted by one or more of the following assumptions: i) parallel sensor-FC channels that do not interfere with one another, ii) exactly the same number of antennas at the transmitter(s) and at the receiver(s) and square and invertible channel matrix, and iii) certain (parametric) matrix in the derivation steps being invertible.

Overcoming these limitations, our goal in this paper is to develop efficient methods to design beamformers of broader applicability. The system model we consider (see.

Fig. 1) is a rather general coherent MAC model, involving many parameters such as arbitrary signal dimension (i.e. vector sources), arbitrary number of sensor nodes, arbitrary number of transmit antennas and/ receiver antennas, and separate power constraint on each sensor. We must acknowledge that we are not the first to study this wireless sensor network model. This general model (which subsumes many previous studies as its special cases) was first introduced in [7]. However, due to the difficulty in solving this general model, [7] provided solution only to special (but nonetheless interesting) cases, such as scalar sources, noiseless sensor-FC channels, and parallel and independent sensor-FC channels. It was several years later that another noteworthy paper [9] succeeded in developing a block coordinated decent (BCD) based method to optimize the minimal mean square error (MMSE) beamformers for this general coherent MAC model. The BCD method developed in [9] is fairly general and capable of obtaining beamformers for most of the scenarios – except when a second-order coefficient matrix in the quadratic optimization problem becomes singular (in which case the method will fail). To avoid the break of the method, [9] had to introduce additional constraints that the number of transmitter antennas at the sensors must not exceed the number of receiver antennas at the FC nor the signal dimension (observation dimension).

It is also worth noting that several closely related problems for the same settings of wireless sensor network have been studied recently. [10] studies the beamformer design towards maximizing the channel capacity between the source and the fusion center. The decentralized detection with Rao test is developed in [16]. The joint optimization of linear filters in sensor networks with huge number of antennas/dimensions are researched in [17] and [18].

In this paper, we propose new approaches to design the linear beamformers for the general system model presented in [7]: 2-BCD, and essentially cyclic multi-block method and its variants and a distributive algorithm. The 2-BCD method optimizes all the beamformers (at the sensors) in a single batch. The essentially cyclic multi-block method handles the beamformers one at a time and its variants can further provide a faster and guaranteed convergence. The distributive algorithm optimizes the transceivers in a decentralized way. The results we obtained, especially the semi-closed-form results (Theorem 3) and the rigorous convergence analysis (Theorem 4),

60 are not only significant to our own problem, but also have easy extensions to a broader set of problems. Specifically, our contributions include:

- 1) In the proposed 2-BCD approach, we successfully transformed the subproblem into a rather general second-order cone programming (SOCP) formulation, which then lends itself to relatively efficient numerical solvers. To the best of
65 our knowledge, the only other batch mode approach that is applicable to the general model considered here is [15]. The specific SOCP formulation developed in [15] must rely on the stringent assumptions that all quadratic coefficient matrices in the objective and all constraints are strictly positive definite (which are too restrictive to apply to our problem). In comparison, our SOCP loosens these re-
70 strictions to positive semi-definiteness, which apply to broader scenarios than [15] and successfully solve our problem.
- 2) In the proposed multi-block BCD approach, we decompose the original problem into smaller blocks each optimizing an individual transmitter with per-transmitter power constraint. The essentially cyclic BCD method developed here provides a
75 complete semi-closed-form solution to the problem optimizing one transmitter, which previous researches dealing with the same/similar type of problems [8, 9, 12, 13, 14, 15] have not been able to. As noted in [28], problems of this nature would in general involve addressing two cases – the “usual case” and the “hard case”. The previous studies succeeded in solving the usual case(s), but left the
80 “hard case” either partially addressed [9] or totally unattended [8, 12, 13, 14, 15]. Here we report a complete solution that imposes no extra assumptions and leaves no case unaddressed. Our proposed semi-closed-form results are not only very efficient, but also shed great insight into the solution itself: they clearly delineate when the solution is unique and when not, and in the presence of multiple solu-
85 tions (all of which achieve MMSE), they spell out the one with the least energy consumption.
- 3) Additionally, all the solutions provided by the existing literature on the similar model have been centralized methods [3, 4, 5, 6, 8, 7, 9, 15]. Although BCD methods were widely adopted in these works, their blocks are updated sequen-

90 tially. Comparatively, we propose a semi-distributive algorithm which updates transmitters of all sensors in parallel. To the best of the authors' knowledge, our proposal is the first distributive method in the same line of research. Our proposed semi-distributive algorithm has time complexity independent of the number of sensors and is suitable for the networks with large number of sensors.

95 4) Last but not least, we also provide rigorous convergence proof for all of our approaches. We know that BCA/BCD does not guarantee convergence of the solution sequence¹, but to prove or disprove has been challenging. While previous papers in this line of research have also noted the importance of convergence analysis, none presented any — with the exception of [9] which proved the convergence of the target sequence but not the convergence of the solution sequence.

100 In this paper, we rigorously prove that not only the target sequence, but the solution points obtained by the proposed algorithms, will also guarantee to converge to the stationary points. Further, since Theorem 4 presents the convergence property of essentially cyclic BCD, the result readily extend to several previous cases (e.g. a coherent MAC case discussed in [9] which is a subcase of ours, and an

105 orthogonal MAC case discussed in [8]).

The rest of the paper is organized as follows: Section 2 introduces the system model. Section 3 discusses the 2-BCD beamformer design approach and analyzes its convergence. Section 4 presents the semi-closed-form solutions to update individual

110 transmitter, proposes an essentially cyclic BCD method and its variants and discusses their convergence. Section 5 proposes a distributive method to optimize all transmitters in a parallel manner. Section 6 provides simulation verification and Section 7 concludes this article.

Notations: We use bold lowercase letters to denote complex vectors and bold capital letters to denote complex matrices. $\mathbf{0}$, $\mathbf{O}_{m \times n}$, and $\mathbf{I}_m$ are used to denote zero vectors, zero matrices of dimension $m \times n$, and identity matrices of order m respectively. $\mathbf{A}^T$, $\mathbf{A}^*$ and $\mathbf{A}^H$ are used to denote transpose, conjugate and conjugate transpose

¹As exemplified by [23], (inappropriate) BCA/BCD approaches could generate diverging or oscillating solution sequence far from the stationary points.

(Hermitian) respectively of an arbitrary complex matrix $\mathbf{A}$. $\text{Tr}\{\cdot\}$ denotes the trace operation of a square matrix. $|\cdot|$ denotes the modulus of a complex scalar, and $\|\cdot\|_2$ denotes the l_2 -norm of a complex vector. $\text{vec}(\cdot)$ means vectorization operation of a matrix, which is performed by **stacking** the columns of a matrix into a long one column. $\otimes$ denotes the Kronecker product. $\text{diag}\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ denotes the block diagonal matrix with its i -th diagonal block being the square complex matrix $\mathbf{A}_i$, $i \in \{1, \dots, n\}$. $\text{Re}\{x\}$ denotes the **real part of a complex number** x .

2. System Model

The model of interest is a centralized wireless sensor network with L sensors and one fusion center where all the nodes are equipped with multiple antennas, as shown in Figure 1. Let M and N_i ($i = 1, 2, \dots, L$) be the number of antennas provisioned to the fusion center and the i -th sensor respectively. Denote $\mathbf{s}$ as the common source vector observed by all sensors. The source $\mathbf{s}$ is a complex vector of dimension K , i.e. $\mathbf{s} \in \mathbb{C}^{K \times 1}$, and is observed by all the sensors. At the i -th sensor, the source signal is linearly transformed by an observation matrix $\mathbf{K}_i \in \mathbb{C}^{J_i \times K}$ and corrupted by additive observation noise $\mathbf{n}_i$, which has zero mean and covariance matrix $\mathbf{\Sigma}_i$.

Figure 1: Multi-Sensor System Model

Here we focus on wireless sensor systems with linear filters. Each sensor applies some linear precoder, $\mathbf{F}_i \in \mathbb{C}^{N_i \times J_i}$, to its observation $(\mathbf{K}_i \mathbf{s} + \mathbf{n}_i)$ before sending it to

the common fusion center. Denote $\mathbf{H}_i \in \mathbb{C}^{M \times N_i}$ as the fading channel between the i -th sensor and the fusion center. Here we consider the coherent MAC model, where the transmitted data is superimposed and corrupted by additive noise at the fusion center. Without loss of generality, the channel noise is modeled as a vector $\mathbf{n}_0 \in \mathbb{C}^{M \times 1}$ with zero mean and white covariance $\sigma_0^2 \mathbf{I}_M$. The fusion center, after collecting all the results, applies a linear postcoder, $\mathbf{G}^H \in \mathbb{C}^{K \times M}$, to retrieve the original source $\mathbf{s}$.

The system model depicted in Figure 1 is the same as the general model presented in [7, 9]. Following the same norm, we assume that the system is perfectly time-synchronous (which may be realized via the GPS system) and that all the channel state information $\mathbf{H}_i$ is known (which may be achieved via channel estimation techniques). Since the sensors and the fusion center are usually distributed over a wide range of space, it is reasonable to assume that the noise $\mathbf{n}_i$ at different sensors and $\mathbf{n}_0$ at the fusion center are mutually uncorrelated.

The signal transmitted by the i -th sensor takes the form of $\mathbf{F}_i(\mathbf{K}_i \mathbf{s} + \mathbf{n}_i)$. The output $\hat{\mathbf{s}}$ of the postcoder at the fusion center is given as

$$\hat{\mathbf{s}} = \mathbf{G}^H \mathbf{r} = \mathbf{G}^H \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i (\mathbf{K}_i \mathbf{s} + \mathbf{n}_i) + \mathbf{n}_0 \right) \quad (1)$$

$$= \mathbf{G}^H \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right) \mathbf{s} + \underbrace{\mathbf{G}^H \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{n}_i + \mathbf{n}_0 \right)}_{\mathbf{n}}, \quad (2)$$

where the equivalent noise $\mathbf{n}$ obtained at the fusion center has covariance matrix $\Sigma_{\mathbf{n}}$ given by

$$\Sigma_{\mathbf{n}} = \sigma_0^2 \mathbf{I}_M + \sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \Sigma_i \mathbf{F}_i^H \mathbf{H}_i^H. \quad (3)$$

We take the mean square error as a figure of merit. The MSE matrix Φ is defined as

$$\Phi \triangleq \mathbb{E} \{ (\mathbf{s} - \hat{\mathbf{s}}) (\mathbf{s} - \hat{\mathbf{s}})^H \}. \quad (4)$$

Assume that the source signal $\mathbf{s}$ has zero mean and a covariance matrix $\Sigma_{\mathbf{s}} \triangleq \mathbb{E} \{ \mathbf{s} \mathbf{s}^H \}$. By plugging (2) into (4), we can express the MSE matrix Φ as a function of $\{\mathbf{F}_i\}$ and

$\mathbf{G}$ as:

$$\begin{aligned}\Phi(\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}) = & \mathbf{G}^H \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right) \Sigma_s \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right)^H \mathbf{G} \\ & - \mathbf{G}^H \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right) \Sigma_s - \Sigma_s \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right)^H \mathbf{G} \\ & + \sum_{i=1}^L \mathbf{G}^H \mathbf{H}_i \mathbf{F}_i \Sigma_i \mathbf{F}_i^H \mathbf{H}_i^H \mathbf{G} + \sigma_0^2 \mathbf{G}^H \mathbf{G} + \Sigma_s.\end{aligned}\quad (5)$$

The total MSE is then given by

$$\text{MSE}(\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}) \triangleq \text{Tr} \left\{ \Phi(\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}) \right\}. \quad (6)$$

Here we consider the case where each sensor has its own transmission power constrained in an average sense. The average transmission power is $\mathbb{E}\{\|\mathbf{F}_i(\mathbf{K}_i \mathbf{s} + \mathbf{n}_i)\|_2^2\} = \text{Tr}\{\mathbf{F}_i(\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i) \mathbf{F}_i^H\}$, which should be no larger than some predefined value P_i . The overall beamforming design problem can then be formulated as the following optimization problem:

$$(\text{P0}): \min_{\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}} \text{MSE}(\{\mathbf{F}_i\}_{i=1}^L, \mathbf{G}), \quad (7a)$$

$$s.t. \text{Tr}\{\mathbf{F}_i(\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i) \mathbf{F}_i^H\} \leq P_i, \quad i \in \{1, \dots, L\}. \quad (7b)$$

The problem we just formulated is nonconvex, which can be verified by checking the special case where $\{\mathbf{F}_i\}_{i=1}^L$ and $\mathbf{G}$ are all scalars.

3. Two-Block Coordinate Descent Method

In this section, we study a two block coordinate descent (2-BCD) method that decouples the design of the postcoder $\mathbf{G}$ (conditioned on the precoders), thereafter referred to as (P1), from the design of all the precoders $\{\mathbf{F}_i\}_{i=1}^L$ (conditioned on the postcoder), thereafter referred to as (P2).

3.1. (P1): Optimizing $\mathbf{G}$ given $\{\mathbf{F}_i\}$

For any given $\{\mathbf{F}_i\}_{i=1}^L$, minimizing MSE with respect to $\mathbf{G}$ becomes a strictly convex non-constrained quadratic problem (P1):

$$(\text{P1}) : \min_{\mathbf{G}} \text{Tr} \left\{ \Phi(\mathbf{G} | \{\mathbf{F}_i\}_{i=1}^L) \right\}. \quad (8)$$

By equating the derivative $\frac{\partial}{\partial \mathbf{G}^*} \text{MSE}(\mathbf{G})$ with zero, the optimal receiver is readily obtained as the well-known Wiener filter [25]

$$\mathbf{G}_{(P1)}^* = \left[\left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right) \Sigma_s \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right)^H + \Sigma_n \right]^{-1} \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i \right) \Sigma_s, \quad (9)$$

where Σ_n is given in (3).

3.2. (P2): Optimizing $\{\mathbf{F}_i\}$ given $\mathbf{G}$

With $\mathbf{G}$ being fixed, the subproblem (P2) minimizes MSE with respect to $\{\mathbf{F}_i\}_{i=1}^L$ is formulated as

$$(P2) : \min_{\{\mathbf{F}_i\}_{i=1}^L} \text{Tr} \left\{ \Phi \left(\{\mathbf{F}_i\}_{i=1}^L \middle| \mathbf{G} \right) \right\}, \quad (10a)$$

$$s.t. \text{Tr} \left\{ \mathbf{F}_i (\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i) \mathbf{F}_i^H \right\} \leq P_i, \quad i \in \{1, \dots, L\}. \quad (10b)$$

160 For the problem (P2) obtained above, we have the following conclusion.

Theorem 1. (P2) has an SOCP formulation.

Proof. First we show that the problem (P2) is convex with respect to $\{\mathbf{F}_i\}_{i=1}^L$. Consider the function $f(\mathbf{X}) : \mathbb{C}^{m \times n} \mapsto \mathbb{R}$, $f(\mathbf{X}) = \text{Tr} \{ \mathbf{A}^H \mathbf{X} \Sigma \mathbf{X}^H \mathbf{A} \}$, where the constant matrices $\mathbf{A}$ and Σ have appropriate dimensions and Σ is Hermitian and positive semidefinite. 165

By the identities $\text{Tr} \{ \mathbf{A} \mathbf{B} \} = \text{Tr} \{ \mathbf{B} \mathbf{A} \}$ and $\text{Tr} \{ \mathbf{A} \mathbf{B} \mathbf{C} \mathbf{D} \} = \text{vec}^T(\mathbf{D}^T) [\mathbf{C}^T \otimes \mathbf{A}] \text{vec}(\mathbf{B})$, $f(\mathbf{X})$ can be equivalently written as $f(\mathbf{X}) = \text{vec}^H(\mathbf{X}) [\Sigma^* \otimes (\mathbf{A} \mathbf{A}^H)] \text{vec}(\mathbf{X})$.

According to [24], i) $[\mathbf{A} \otimes \mathbf{B}]^H = \mathbf{A}^H \otimes \mathbf{B}^H$; ii) for any two Hermitian matrices $\mathbf{A}_{m \times m}$ and $\mathbf{B}_{n \times n}$ having eigenvalues $\{\lambda_i(\mathbf{A})\}_{i=1}^m$ and $\{\lambda_j(\mathbf{B})\}_{j=1}^n$ respectively, the 170 eigenvalues of their Kronecker product $\mathbf{A} \otimes \mathbf{B}$ are given by $\{\lambda_i(\mathbf{A}) \lambda_j(\mathbf{B})\}_{i=1, j=1}^{m, n}$. As a result, $\mathbf{A} \otimes \mathbf{B}$ is Hermitian and positive semidefinite when $\mathbf{A}$ and $\mathbf{B}$ are Hermitian and positive semidefinite.

Since $\mathbf{A} \mathbf{A}^H$ and Σ^* are both positive semidefinite, $[\Sigma^* \otimes (\mathbf{A} \mathbf{A}^H)]$ is positive semidefinite and therefore $f(\mathbf{X})$ is actually a convex homogeneous quadratic function 175 of $\text{vec}(\mathbf{X})$.

We now substitute $\mathbf{X}$ in $f(\mathbf{X})$ by $\sum_{i=1}^L (\mathbf{H}_i \mathbf{F}_i \mathbf{K}_i)$. Recall that affine operation preserves convexity [26]. The term $\text{Tr}\left\{\mathbf{G}^H \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i\right) \boldsymbol{\Sigma}_s \left(\sum_{i=1}^L \mathbf{H}_i \mathbf{F}_i \mathbf{K}_i\right)^H \mathbf{G}\right\}$ in the objective function (P2) is therefore convex with respect to $\{\mathbf{F}_i\}_{i=1}^L$. By the same reasoning, the remaining terms in the objective and the constraints of (P2) are either convex quadratic or affine functions of $\{\mathbf{F}_i\}_{i=1}^L$ and therefore the problem (P2) is convex with respect to $\{\mathbf{F}_i\}_{i=1}^L$. 180

Next we introduce the notations $\mathbf{f}_i \triangleq \text{vec}(\mathbf{F}_i)$, $\mathbf{g} \triangleq \text{vec}(\mathbf{G})$, $\mathbf{A}_{ij} \triangleq (\mathbf{K}_j \boldsymbol{\Sigma}_s \mathbf{K}_i^H)^T \otimes (\mathbf{H}_i^H \mathbf{G} \mathbf{G}^H \mathbf{H}_j)$, $\mathbf{B}_i \triangleq (\mathbf{K}_i \boldsymbol{\Sigma}_s)^T \otimes \mathbf{H}_i$, and $\mathbf{C}_i \triangleq \boldsymbol{\Sigma}_i^* \otimes (\mathbf{H}_i^H \mathbf{G} \mathbf{G}^H \mathbf{H}_i)$.

By the identity $\text{Tr}\{\mathbf{ABCD}\} = \text{vec}^T(\mathbf{D}^T) [\mathbf{C}^T \otimes \mathbf{A}] \text{vec}(\mathbf{B})$ and the above notations, we can rewrite the MSE in (P2) as

$$\begin{aligned} \text{MSE}\left(\{\mathbf{f}_i\}_{i=1}^L \middle| \mathbf{g}\right) &= \sum_{i=1}^L \sum_{j=1}^L \mathbf{f}_i^H \mathbf{A}_{ij} \mathbf{f}_j - 2\text{Re}\left(\sum_{i=1}^L \mathbf{g}^H \mathbf{B}_i \mathbf{f}_i\right) \\ &\quad + \sum_{i=1}^L \mathbf{f}_i^H \mathbf{C}_i \mathbf{f}_i + \sigma_0^2 \|\mathbf{g}\|^2 + \text{Tr}\{\boldsymbol{\Sigma}_s\}. \end{aligned} \quad (11)$$

By further denoting $\mathbf{f}^T \triangleq [\mathbf{f}_1^T, \dots, \mathbf{f}_L^T]$, $\mathbf{E}_i \triangleq (\mathbf{K}_i \boldsymbol{\Sigma}_s \mathbf{K}_i^H + \boldsymbol{\Sigma}_i)^T \otimes \mathbf{I}_{N_i}$, $c \triangleq \text{Tr}\{\boldsymbol{\Sigma}_s\} + \sigma_0^2 \|\mathbf{g}\|^2$, and the block matrices $[\mathbf{A}]_{i,j} \triangleq \mathbf{A}_{i,j}$, $\mathbf{B} \triangleq [\mathbf{B}_1, \dots, \mathbf{B}_L]$, $\mathbf{C} \triangleq \text{diag}\{\mathbf{C}_1, \dots, \mathbf{C}_L\}$, and $\mathbf{D}_i \triangleq \text{diag}\{\mathbf{O}_{\sum_{j=1}^{i-1} J_j N_j}, \mathbf{E}_i, \mathbf{O}_{\sum_{j=i+1}^L J_j N_j}\}$, the problem (P2) can be rewritten as (P2'):

$$(\text{P2}') : \min_{\mathbf{f}} \mathbf{f}^H (\mathbf{A} + \mathbf{C}) \mathbf{f} - 2\text{Re}\{\mathbf{g}^H \mathbf{B} \mathbf{f}\} + c, \quad (12a)$$

$$s.t. \quad \mathbf{f}^H \mathbf{D}_i \mathbf{f} \leq P_i, \quad i \in \{1, \dots, L\}. \quad (12b)$$

As we just proved, (P2') (or equivalently (P2)) is convex, which implies $(\mathbf{A} + \mathbf{C})$ is positive semidefinite. Thus the square root $(\mathbf{A} + \mathbf{C})^{\frac{1}{2}}$ exists. The problem (P2') can

therefore be reformulated in an SOCP form as follows

$$(\text{P2}_{SOCP}) : \min_{\mathbf{f}, t, s} t, \quad (13a)$$

$$\text{s.t. } s - 2\text{Re}\{\mathbf{g}^H \mathbf{B}\mathbf{f}\} + c \leq t; \quad (13b)$$

$$\left\| \begin{pmatrix} (\mathbf{A} + \mathbf{C})^{\frac{1}{2}} \mathbf{f} \\ \frac{s-1}{2} \end{pmatrix} \right\|_2 \leq \frac{s+1}{2}; \quad (13c)$$

$$\left\| \begin{pmatrix} \mathbf{D}_i^{\frac{1}{2}} \mathbf{f} \\ \frac{P_i-1}{2} \end{pmatrix} \right\|_2 \leq \frac{P_i+1}{2}, \quad i \in \{1, \dots, L\}; \quad (13d)$$

Thus the proof is complete. $\square$

Remark 1. We have just solved the beamformer design problem via SOCP. Note that the idea to jointly optimize all the transmitters also appeared in [7] and [15]. Specifically, [7] transformed the original problem into a semidefinite programming (SDP) problem by substituting the Wiener filter into the objective and introducing relaxation, but the approach thereof only applies to the special scenario where all the transmitters and the receiver have identical antenna numbers and all the channel matrices are diagonal. The work of [15] also presented an SOCP formulation but required the quadratic matrices in the objective and in all the constraints be strictly positive definite, an assumption that is actually hard to meet in our problem (P2') (see the definition of $\mathbf{D}_i$). Comparatively, we only require the matrix be positive semidefinite, a condition that is always satisfied. Detailed numerical results will be shown in Sec.6 comparing our algorithm with the one proposed in [9].

Summarizing the above discussion, the problem (P0) can be solved by a 2-BCD algorithm: updating $\mathbf{G}$ by solving (P1) and updating $\{\mathbf{F}_i\}_{i=1}^L$ by solving (P2') alternatively, which is summarized in Algorithm 1.

3.3. Convergence of 2-BCD Algorithm

In this subsection we study the convergence of the above 2-BCD algorithm. For the proposed 2-BCD algorithm, we have the following convergence conclusion.

Algorithm 1: 2-BCD Algorithm to Solve (P0)

1 Initialization: Randomly generate feasible $\{\mathbf{F}_i^{(0)}\}_{i=1}^L, i \in \{1, \dots, L\}$;
 Compute $\mathbf{G}^{(0)}$ using (9);
2 repeat
 3 With $\mathbf{G}^{(j-1)}$ fixed, solve (P2') and obtain $\{\mathbf{F}_i^{(j)}\}_{i=1}^L$;
 4 With $\{\mathbf{F}_i^{(j)}\}_{i=1}^L$ fixed, compute $\mathbf{G}^{(j)}$ using (9);
5 until decrease of MSE is small enough or predefined number of iterations is reached;

Theorem 2. The objective sequence $\{\text{MSE}^{(j)}\}_{j=0}^\infty$ generated by the 2-BCD algorithm in Algorithm 1 is monotonically decreasing. If $\mathbf{K}_i \Sigma_s \mathbf{K}_i^H \succ 0$ or $\Sigma_i \succ 0$ for all $i \in \{1, \dots, L\}$, the solution sequence $\{\{\mathbf{F}_i^{(j)}\}_{i=1}^L, \mathbf{G}^{(j)}\}_{j=1}^\infty$ generated by the 2-BCD algorithm has limit points and each limit point of $\{\{\mathbf{F}_i^{(j)}\}_{i=1}^L, \mathbf{G}^{(j)}\}_{j=1}^\infty$ is a stationary point of (P0).

Proof. Since each block update solves a minimization problem, MSE keeps decreasing. Let $\mathcal{X}_i = \{\mathbf{X} \in \mathbb{C}^{N_i \times J_i} \mid \text{Tr}\{\mathbf{X}(\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i) \mathbf{X}^H\} \leq P_i\}$, for $i = 1, \dots, L$ and $\mathcal{X}_{L+1} = \mathbb{C}^{M \times K}$. Under the strictly positive definiteness assumption of $\mathbf{K}_i \Sigma_s \mathbf{K}_i^H$ or Σ_i , we have $(\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i) \succ 0$ and thus $(\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i)^T \otimes \mathbf{I}_{N_i} \succ 0$ for all $i \in \{1, \dots, L\}$. This implies that the null space of $(\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i)^T \otimes \mathbf{I}_{N_i}$ is $\{\mathbf{0}\}$ and consequently $\mathbf{f}_i$ has to be bounded to satisfy power constraint. Therefore $\mathcal{X}_i$ is bounded for all $i \in \{1, \dots, L\}$. Since the feasible set for each $\mathbf{F}_i$ is bounded, by Bolzano-Weierstrass theorem, there exists a convergent subsequence $\{\{\mathbf{F}_i^{(j_k)}\}_{i=1}^L\}_{k=1}^\infty$. Since $\mathbf{G}$ is updated by equation (9) as a continuous function of $\{\mathbf{F}_i\}_{i=1}^L$, the subsequence $\{\mathbf{G}^{(j_k+1)}\}_{k=1}^\infty$ also converges and thus bounded. By further restricting to a subsequence of $\{\{\mathbf{F}_i^{(j_k+1)}\}, \mathbf{G}^{(j_k+1)}\}_{k=1}^\infty$, we can obtain a convergent subsequence of $\{\{\mathbf{F}_i^{(j)}\}_{i=1}^L, \mathbf{G}^{(j)}\}_{j=1}^\infty$.

Since Algorithm 1 is a two-block BCD procedure, and since the problem (P0) has continuously differentiable objective and closed and convex feasible domain, Corollary 2 in [21] becomes applicable. We thus conclude that any limit point of $\{\{\mathbf{F}_i^{(j)}\}_{i=1}^L, \mathbf{G}^{(j)}\}_{j=1}^\infty$ is a stationary point of (P0). $\square$

4. Multi-Block Coordinate Descent Method

225 The previous 2-BCD approach has solved the subproblem (P2) via SOCP with a complexity $\mathcal{O}\left(\sqrt{L}\left(\sum_{i=1}^L N_i J_i\right)^3\right)$. The complexity can be daunting when the sensor network has a large number of sensors and/or antennas. Below we present more efficient ways to optimize the transmitters.

4.1. Further Decoupling of (P2) and Closed-Form Solution

We now consider solving (P2) analytically. Looking back to (P2), we realize that the difficulty comes from the quadratic terms in the objective, which involves all the different precoders, thus making the Karush-Kuhn-Tucker(KKT) conditions of (P2) analytically unsolvable. Once again we apply the BCD idea and further decompose the subproblem (P2) into L blocks. Instead of optimizing all the $\mathbf{F}_i$'s in a single batch, we optimize one $\mathbf{f}_i$ at a time with the others being fixed. By introducing the notation $\mathbf{q}_i \triangleq \sum_{j=1, j \neq i}^L \mathbf{A}_{ij} \mathbf{f}_j$, each subproblem (P2'_i) of (P2') is given as

$$(P2'_i) : \min_{\mathbf{f}_i} \mathbf{f}_i^H (\mathbf{A}_{ii} + \mathbf{C}_i) \mathbf{f}_i + 2\text{Re}\{\mathbf{q}_i^H \mathbf{f}_i\} - 2\text{Re}\{\mathbf{g}^H \mathbf{B}_i \mathbf{f}_i\} \quad (14a)$$

$$s.t. \mathbf{f}_i^H \mathbf{E}_i \mathbf{f}_i \leq P_i. \quad (14b)$$

230 Now our problem boils down to solving the simpler problem (P2'_i), for $i = 1, \dots, L$. The following theorem provides a semi-closed-form solution to (P2'_i). The only reason that this is not a *fully* closed-form solution is because it may involve a Newton's root finding algorithm or a bisection search to determine the value of a positive real number μ_i^* .

Theorem 3. Assume $\mathbf{K}_i \Sigma_s \mathbf{K}_i^H \succ 0$ or $\Sigma_i \succ 0$. Define parameters $\mathbf{M}_i$, $\mathbf{U}_i$ and $\mathbf{p}_i$ as in Equation (A.1) in the appendix. Let r_i be the rank of $\mathbf{M}_i$ and $p_{i,k}$ be the i -th entry of $\mathbf{p}_i$. The solution to (P2'_i) is given as:

CASE (I) — If either of the following two conditions holds:

- i) $\exists k \in \{r_i + 1, \dots, J_i N_i\}$ such that $|p_{i,k}| \neq 0$;
- or ii) $\sum_{k=r_i+1}^{J_i N_i} |p_{i,k}| = 0$ and $\sum_{k=1}^{r_i} \frac{|p_{i,k}|^2}{\lambda_{i,k}^2} > P_i$.

The optimal solution to (P2'_i) is given by

$$\mathbf{f}_i^* = (\mathbf{A}_{ii} + \mathbf{C}_i + \mu_i^* \mathbf{E}_i)^{-1} (\mathbf{B}_i^H \mathbf{g} - \mathbf{q}_i), \quad (15)$$

with the positive value μ_i^* being the unique solution to the equation:

$g_i(\mu_i) = \sum_{k=1}^{J_i N_i} \frac{|p_{i,k}|^2}{(\lambda_{i,k} + \mu_i)^2} = P_i$. An interval $[lbd_i, ubd_i]$ containing μ_i^* is determined by Lemma 1 which comes later.

CASE (II) — When $\sum_{k=r_i+1}^{J_i N_i} |p_{i,k}| = 0$ and $\sum_{k=1}^{r_i} \frac{|p_{i,k}|^2}{\lambda_{i,k}^2} \leq P_i$,

The optimal solution to $(P2'_i)$ is given by

$$\mathbf{f}_i^* = \mathbf{E}_i^{-\frac{1}{2}} \left(\mathbf{E}_i^{-\frac{1}{2}} (\mathbf{A}_{ii} + \mathbf{C}_i) \mathbf{E}_i^{-\frac{1}{2}} \right)^\dagger \mathbf{E}_i^{-\frac{1}{2}} (\mathbf{B}_i^H \mathbf{g} - \mathbf{q}_i). \quad (16)$$

235 *Proof.* Please see Appendix Appendix A for proof. $\square$

Remark 2. Recall that in CASE (I) of Theorem 3, μ_i^* is obtained as the solution to $g_i(\mu_i) = P_i$. The equation $g_i(\mu_i) = P_i$ generally has no analytic solution. In the special case of $K = J_i = 1$ (scalar source and scalar observation), μ_i^* and $\mathbf{f}_i^*$ can be obtained fully analytically without the need for numerical searches, and the eigenvalue decomposition is also unnecessary; Solving $(P2_i)$ therefore becomes extremely efficient. The details of the scalar case can be found in [11]². In the general vector-source case, μ_i^* can be numerically solved with high efficiency by a bisection search or by Newton's root finding method (e.g. [28]). A starting point for the numerical search is presented in Lemma 1 below.

245 **Remark 3.** The result obtained in Theorem 3 can be regarded as an enhancement of the standard solution to the trust region subproblem (e.g. [28]) for the case of *convex* objective. Unlike the existing approaches to the trust region subproblem, our Theorem 3 not only specifies the solution structures, but also delineates the conditions for different cases: the usual case and the hard case [28]. In CASE(I) or the usual case, the optimal beamformer $\mathbf{F}_i^*$ herein is unique and must utilize all the power available. In CASE(II) or the hard case (which is not rare, see Remark 4), we see that $\mu_i^* = 0$ always holds. This case is actually rather intriguing as it involves several scenarios. i) As illustrated in the proof, when $(\mathbf{A}_{ii} + \mathbf{C}_i)$ is singular and $\sum_{k=1}^{r_i} |p_{i,k}|^2 / \lambda_{i,k}^2 < P_i$,

²Reference [11] solved an “approximated *scalar-source* version of problem $(P2'_i)$, where a specific affine term of $\mathbf{f}_i$ in the objective of $(P2'_i)$ was approximated by its latest value. The technique of approximation is also discussed in Subsection 4.3 of this paper. Additionally, a similar line of reasoning developed in [11] can be applied to obtain the exact scalar-form of $(P2'_i)$.

(P2') has infinitely many optimal solutions, all of which achieve the same MMSE but
 255 may consume different amount of power. The existing methods (e.g. [28]) would gen-
 erate a solution that (wastefully) consumes all the power provisioned. In comparison,
 our result in (16) provides the power-preserving optimal solution, which consumes the
 minimal transmission power among all the optimal solutions to (P2'). ii) In the other
 scenarios (($\mathbf{A}_{ii} + \mathbf{C}_i$) is nonsingular and/or $\sum_{k=1}^{r_i} |p_{i,k}|^2 / \lambda_{i,k}^2 = P_i$), only a single op-
 260 timal solution exists, and this unique solution $\mathbf{F}_i^*$ may either take the full power (when
 $\sum_{k=1}^{r_i} |p_{i,k}|^2 / \lambda_{i,k}^2 = P_i$) or consumes only part of it (when $\sum_{k=1}^{r_i} |p_{i,k}|^2 / \lambda_{i,k}^2 < P_i$).

Remark 4. The subproblem (P2') — minimizing MSE for a single beamformer with
 one power constraint — is a fairly general problem that also appeared (possibly in
 different forms) in a number of previous beamformer papers [12, 14, 9, 13, 15, 8],
 265 but a complete solution was missing. The previous solutions would, at some point,
 involve a matrix inversion, but the relevant matrix may not have a full rank. This
 singularity difficulty was sometimes neglected [12, 13, 14] or other times circumvented
 by the inclusion of extra restrictive assumptions [9, 8, 15]. For example, [9] introduced
 additional constraints on signal dimensions (i.e. antenna numbers at the transmitters
 270 and the receiver, and dimensions of the source and the observations) to ensure that the
 quadratic matrix in their objective is always invertible. Similarly, [8] also introduced
 (somewhat stringent) constraints (see $a1$)' and the hypothesis of Theorem 4 therein) to
 ensure that only CASE(I) would happen. The extra constraints sadly undermined the
 usability of these previous results. We are happy to provide a full solution in Theorem
 275 3, which completed the previous results. Please also note that the case of singular
 matrix ($\mathbf{A}_{ii} + \mathbf{C}_i$) is actually not rare. In fact, whenever $K < N_i$ or $M < N_i$ holds,
 the matrices $\mathbf{A}_{ii}$ and $\mathbf{C}_i$ are both born rank deficient. If they share common nonzero
 components of null space, ($\mathbf{A}_{ii} + \mathbf{C}_i$) will be rank deficient. For example, the simple
 case when $\mathbf{K}_i = \mathbf{I}_K$, $\Sigma_s = \sigma_s^2 \mathbf{I}$, $\Sigma_i = \sigma_i^2 \mathbf{I}$ and $\min(K, M) < N_i$ would result in
 280 singular ($\mathbf{A}_{ii} + \mathbf{C}_i$). If, additionally, P_i is large, then μ_i^* will be zero.

We now provide an interval $[lbd_i, ubd_i]$ containing the positive μ_i^* in CASE(I).
 These bounds can serve as the starting point for the bisection search or the Newton's
 root finding method, in the quest for the solution of $g(\mu_i) = P_i$.

Lemma 1. The positive μ_i^* in $(P2'_i)$ (i.e. CASE (I) in Theorem 3) has the following lower bound lbd_i and upper bound ubd_i :

i) For subcase i)

$$lbd_i = \left[\frac{\|\mathbf{p}_i\|_2}{\sqrt{P_i}} - \lambda_{i,1} \right]^+, \quad ubd_i = \frac{\|\mathbf{p}_i\|_2}{\sqrt{P_i}}; \quad (17)$$

ii) For subcase ii)

$$lbd_i = \left[\frac{\|\mathbf{p}_i\|_2}{\sqrt{P_i}} - \lambda_{i,1} \right]^+, \quad ubd_i = \frac{\|\mathbf{p}_i\|_2}{\sqrt{P_i}} - \lambda_{i,r_i}, \quad (18)$$

where $[x]^+ = \max\{0, x\}$.

Proof. For subcase i), by definition of $g_i(\mu_i)$ in (A.5), we have

$$\begin{aligned} \frac{\|\mathbf{p}_i\|_2^2}{(\mu_i + \lambda_{i,1})^2} &= \frac{\sum_{k=1}^{J_i N_i} |p_{i,k}|^2}{(\mu_i + \lambda_{i,1})^2} \leq g_i(\mu_i) = P_i \\ &\leq \frac{\sum_{k=1}^{J_i N_i} |p_{i,k}|^2}{\mu_i^2} = \frac{\|\mathbf{p}_i\|_2^2}{\mu_i^2}, \end{aligned} \quad (19)$$

which can be equivalently written as

$$\frac{\|\mathbf{p}_i\|_2}{\sqrt{P_i}} - \lambda_{i,1} \leq \mu_i \leq \frac{\|\mathbf{p}_i\|_2}{\sqrt{P_i}}. \quad (20)$$

Also notice that μ_i^* should be positive; the bounds in (17) thus follow.

For subcase ii), by assumption, $\sum_{k=r_i+1}^{J_i N_i} |p_{i,k}|^2 = 0$. This leads to

$$\begin{aligned} \frac{\|\mathbf{p}_i\|_2^2}{(\mu_i + \lambda_{i,1})^2} &= \frac{\sum_{k=1}^{r_i} |p_{i,k}|^2}{(\mu_i + \lambda_{i,1})^2} \leq g_i(\mu_i) = P_i \\ &\leq \frac{\sum_{k=1}^{r_i} |p_{i,k}|^2}{(\mu_i + \lambda_{i,r_i})^2} = \frac{\|\mathbf{p}_i\|_2^2}{(\mu_i + \lambda_{i,r_i})^2}. \end{aligned} \quad (21)$$

Following the same line of derivation as in subcase i), we obtain the bounds in (18). $\square$

4.2. Essentially Cyclic $(L+1)$ -BCD Algorithm

Based on the semi-closed-form solution established in previous subsection, we decompose the whole set of variables into $(L+1)$ blocks and update each variable by (9) or Theorem 3 respectively. Compared to the 2-BCD algorithm, the block updating rule for multiple-block coordinate descent method can have various patterns. Here we adopt a very general and useful updating rule called *essentially cyclic rule* [22]. For a

specific update rule, if there exists a positive integer T (termed period), such that each block of variables is updated at least once within any consecutive T updates, the the
 295 update rule is said to be essentially cyclic with period T .

The convergence of BCD algorithm usually requires the solutions in each block update be unique [21, 20, 22]. Without the uniqueness condition, convergence may not be guaranteed [23]. In retrospect to Remark 3, (P2'_{*i*}) can have non-unique solution. To overcome this difficulty, we adopt the proximal method (Exercise 2.7.1 in [20]).

Specifically, to update the i -th transmitter, we resort to the proximal version objective $\text{MSE}(\mathbf{f}_i | \{\mathbf{f}_j\}_{j \neq i}, \mathbf{g}) + \kappa \|\mathbf{f}_i - \hat{\mathbf{f}}_i\|_2^2$ of (P2'_{*i*}) with κ being any positive real constant and $\hat{\mathbf{f}}_i$ being the latest value of $\mathbf{f}_i$ until the current update. The modified problem of updating the i -th precoder is as (P4_{*i*})

$$\begin{aligned} (\text{P4}_i): \min_{\mathbf{f}_i} & \mathbf{f}_i^H (\mathbf{A}_{ii} + \mathbf{C}_i + \kappa \mathbf{I}_{N_i J_i}) \mathbf{f}_i + 2 \text{Re} \{ (\mathbf{q}_i^H - \mathbf{g}^H \mathbf{B}_i - \kappa \hat{\mathbf{f}}_i^H) \mathbf{f}_i \} \\ \text{s.t.} & \mathbf{f}_i^H \mathbf{E}_i \mathbf{f}_i \leq P_i. \end{aligned} \quad (22)$$

300 For the proximal $(L+1)$ -BCD algorithm, we have the following statement.

Theorem 4. Assume that $\mathbf{K}_i \Sigma_s \mathbf{K}_i^H \succ 0$ or $\Sigma_i \succ 0$, $\forall i \in \{1, \dots, L\}$. By updating $\mathbf{G}$ and $\mathbf{F}_i$ via (P1) and (P4_{*i*}) respectively, any essentially cyclic $(L+1)$ -BCD algorithm generates monotonically decreasing MSE sequence and the resulting solution sequence has limit points with each limit point being a stationary point of the original problem
 305 (P0).

Proof. We relegate this proof to Appendix Appendix B. □

Note that the solution to (P4_{*i*}) can be easily obtained by Theorem 3 with the terms $(\mathbf{A}_{ii} + \mathbf{C}_i)$ and $(\mathbf{B}_i^H \mathbf{g} - \mathbf{q}_i)$ replaced by $(\mathbf{A}_{ii} + \mathbf{C}_i + \kappa \mathbf{I}_{N_i J_i})$ and $(\mathbf{B}_i^H \mathbf{g} - \mathbf{q}_i + \kappa \hat{\mathbf{f}}_i)$, respectively; no additional complexity is introduced.

310 **Remark 5.** The essentially cyclic update framework we just discussed also applies to other scenarios. For example, it subsumes the algorithm in [9] (Table I therein) as a special realization. Further, although [8] considered a system model of orthogonal MAC (rather than coherent MAC), the convergence of their approach can actually be

proved following a very similar line of reasoning as in our proof (by repeating the proof
 315 associated with each receiver multiple times).

Despite the advantage of provable convergence, one drawback of the above proximal update is its relatively slow convergence rate, as will be shown in Section 6. However this shortcoming can be well compensated by the following acceleration scheme.

4.3. Accelerating the Convergence

The aforementioned essentially cyclic $(L+1)$ -BCD algorithm can be accelerated by introducing an approximation during the update of individual beamformers $\mathbf{F}_i$ in $(P2'_i)$. In addition to setting $\{\mathbf{F}_j\}_{j \neq i}$ as known and fixed, we assume that the term $\mathbf{A}_{ii}\mathbf{f}_i$ is also known by leveraging the value of $\mathbf{f}_i$ in the previous updates. In other words, we define $\hat{\mathbf{q}}_i = \sum_{j=1, j \neq i}^L \mathbf{A}_{ij}\mathbf{f}_j + \mathbf{A}_{ii}\hat{\mathbf{f}}_i = \mathbf{q}_i + \mathbf{A}_{ii}\hat{\mathbf{f}}_i$ with $\hat{\mathbf{f}}_i$ being the latest value of $\mathbf{f}_i$. Thus, to update $\mathbf{f}_i$ we solve the approximate version $(P5_i)$ of $(P2'_i)$:

$$(P5_i) : \min_{\mathbf{f}_i} \mathbf{f}_i^H \mathbf{C}_i \mathbf{f}_i + 2\text{Re}\{\hat{\mathbf{q}}_i^H \mathbf{f}_i\} - 2\text{Re}\{\mathbf{g}^H \mathbf{B}_i \mathbf{f}_i\} \quad (23a)$$

$$s.t. \mathbf{f}_i^H \mathbf{E}_i \mathbf{f}_i \leq P_i. \quad (23b)$$

320 The problem $(P5_i)$ can still be efficiently solved by Theorem 3.

Remark 6. Similar approximation idea also appeared in [12], where the precoders of multiusers was updated in a cyclic manner. In Implementation 2 (Table II) of [12], with others being fixed, one separate precoder is updated by minimizing the total MSE function with some terms of the to-be-updated precoder approximated by previous values. Results in [12] show that the approximation can lead to a considerable improvement in
 325 convergence speed, which is consistent with what we observe in our case. We would also like to point out one interesting fact regarding the approximation treatment we just did: the approximation in $(P5'_i)$ has identical directional derivatives with the original problem in $(P2'_i)$ with respect to $\mathbf{f}_i$ at the point $\hat{\mathbf{f}}$ along any directions. Since the
 330 optimality conditions, e.g. KKT conditions or stationary point conditions, are usually primarily related to the first-order property, this fact therefore provides some intuition of the excellent convergence of the approximation in (23).

The very fast convergence of the approximate update inspires us to treat it as a useful complement to the proximal update. In the actual implementation, $(L + 1)$ -BCD algorithm can be performed in an *approximate-proximal* manner — in the first few outer-loop iterations, we run the approximate update to give it a jump start, and then convert to the proximal update in the subsequent updates. This approximate-proximal combination exhibits a fast speed (see Sec.6) and at the same time guarantees a stationary-point convergence.

5. A Decentralized Approach

In this section we introduce a decentralized algorithm, which optimizes all the transmitters in a distributive manner.

5.1. Distributively Optimizing $\{\mathbf{F}_i\}$

The major drawback of the centralized algorithm is that when the number of sensors/antennas grows large, the entire optimization can be time-requiring. Thus a distributed and closed form solution is highly desirable. The major difficulty to distributively optimize separate transmitters in (P2) lies in the coefficient matrix $\mathbf{A}$, which is non-diagonal and thus couples different transmitters. To overcome this difficulty, we can adopt the majorization-minimization (MM, [29]) techniques to modify the objective by its upper bound. Assume that $\hat{\mathbf{f}}$ is the value of the entire transmitters in the latest update. The second-order Taylor expansion of the quadratic term $\mathbf{f}^H \mathbf{A} \mathbf{f}$ in (P1) at the point of $\hat{\mathbf{f}}$ is given as

$$\begin{aligned} \mathbf{f}^H \mathbf{A} \mathbf{f} &= (\mathbf{f} - \hat{\mathbf{f}})^H \mathbf{A} (\mathbf{f} - \hat{\mathbf{f}}) + 2\text{Re}\{\hat{\mathbf{f}}^H \mathbf{A} (\mathbf{f} - \hat{\mathbf{f}})\} + \hat{\mathbf{f}}^H \mathbf{A} \hat{\mathbf{f}}, \\ &\leq \lambda_{\max}(\mathbf{A}) \|\mathbf{f} - \hat{\mathbf{f}}\|_2^2 + 2\text{Re}\{\hat{\mathbf{f}}^H \mathbf{A} (\mathbf{f} - \hat{\mathbf{f}})\} + \hat{\mathbf{f}}^H \mathbf{A} \hat{\mathbf{f}}, \end{aligned} \quad (24)$$

where $\lambda_{\max}(\mathbf{A})$ denotes the maximal eigenvalue of matrix $\mathbf{A}$. In fact $\mathbf{A}$ is positive semidefinite and thus $\lambda_{\max}(\mathbf{A}) > 0$. Substitute the term $\mathbf{f}^H \mathbf{A} \mathbf{f}$ in (P1) by the upper bound given in (24) and noticing that matrix $\mathbf{C}$ is block diagonal, we obtain the

approximate version of (P2) as follows:

$$(P6) : \min_{\{\mathbf{f}_i\}} h(\mathbf{f}; \hat{\mathbf{f}} | \mathbf{G}) \triangleq \quad (25)$$

$$\sum_{i=1}^L \left(\mathbf{f}_i^H (\lambda_{\max}(\mathbf{A}) \mathbf{I}_{J_i N_i} + \mathbf{C}_i) \mathbf{f}_i - 2 \operatorname{Re} \{ \mathbf{t}_i^H \mathbf{f}_i \} \right) + c_2, \quad (26)$$

$$\text{s.t. } \mathbf{f}_i^H \mathbf{E}_i \mathbf{f}_i \leq P_i, \quad i = \{1, \dots, L\}.$$

where $\mathbf{t}_i \triangleq [\mathbf{B}_i^H \mathbf{g} + \lambda_{\max}(\mathbf{A}) \hat{\mathbf{f}}_i - \mathbf{s}_i]$ with $\mathbf{s}_i \triangleq [\mathbf{A} \hat{\mathbf{f}}]_{\sum_{j=1}^{i-1} (J_j N_j) + 1 : \sum_{j=1}^i (J_j N_j)}$ and c_2 denotes all the constant terms independent of $\mathbf{f}$. Here the objective function $h(\mathbf{f}; \hat{\mathbf{f}} | \mathbf{G}, \mathbf{W})$ is an approximation of the function $\text{MSE}(\{\mathbf{F}_i\}_{i=1}^L | \mathbf{G})$ at the point $\hat{\mathbf{f}}$. Denote $\hat{\mathbf{f}}$ as the vector obtained by reshaping $\{\hat{\mathbf{F}}_i\}_{i=1}^L$. According to (24) we have,

$$h(\mathbf{f}; \hat{\mathbf{f}} | \mathbf{G}) \geq \text{MSE}(\{\mathbf{F}_i\}_{i=1}^L | \mathbf{G}), \quad (27a)$$

$$h(\hat{\mathbf{f}}; \hat{\mathbf{f}} | \mathbf{G}) = \text{MSE}(\{\hat{\mathbf{F}}_i\}_{i=1}^L | \mathbf{G}), \quad \forall \mathbf{f}, \hat{\mathbf{f}}. \quad (27b)$$

Obviously the problem (P6) has decoupled constraints and separable objective. Thus each optimal $\mathbf{f}_i$ can be obtained by solving the following problem

$$(P6_i) : \min_{\mathbf{f}_i} \mathbf{f}_i^H (\lambda_{\max}(\mathbf{A}) \mathbf{I}_{J_i N_i} + \mathbf{C}_i) \mathbf{f}_i - 2 \operatorname{Re} \{ \mathbf{t}_i^H \mathbf{f}_i \}, \quad (28a)$$

$$\text{s.t. } \mathbf{f}_i^H \mathbf{E}_i \mathbf{f}_i \leq P_i. \quad (28b)$$

Assume that the eigenvalue decomposition $\mathbf{E}_i^{-1/2} (\lambda_{\max}(\mathbf{A}) \mathbf{I}_{J_i N_i} + \mathbf{C}_i) \mathbf{E}_i^{-1/2} = \mathbf{V}_i \mathbf{\Psi}_i \mathbf{V}_i^H$ and the diagonal matrix $\mathbf{\Psi}_i$ has its diagonal elements $\{\psi_{i,k}\}_{k=1}^{J_i N_i}$ arranged in decreasing order. Denote $\mathbf{w}_i \triangleq \mathbf{V}_i \mathbf{E}_i^{-1/2} \mathbf{t}_i$ and let $w_{i,k}$ be the k -th component of the vector $\mathbf{w}_i$. By utilizing Theorem 3 again, the optimal $\mathbf{f}_i^*$ is given as

$$\mathbf{f}_i^* = (\lambda_{\max}(\mathbf{A}) \mathbf{I}_{J_i N_i} + \mathbf{C}_i + \mu_i \mathbf{E}_i)^{-1} \mathbf{t}_i, \quad (29)$$

with the parameter μ_i being determined in exactly one of the following two cases: i) if $\sum_{k=1}^{J_i N_i} |w_{i,k}|^2 / (\psi_{i,k} + \lambda_{\max}(\mathbf{A}))^2 > P_i$, μ_i is the unique solution of the equation $\sum_{k=1}^{J_i N_i} \frac{|w_{i,k}|^2}{(\psi_{i,k} + \lambda_{\max}(\mathbf{A}) + \mu_i)^2} = P_i$; ii) otherwise $\mu_i = 0$.

In practice, to update the transmitters, we solve (P6) with its objective approximated at the point of transmitters of last update. The overall distributive algorithm is summarized in Alg.2.

Algorithm 2: Distributive Algorithm

```

1 Initialization: Randomly generate feasible  $\{\mathbf{F}_i^{(0)}\}_{i=1}^L$ ; Compute  $\mathbf{G}^{(0)}$  using (9);
2 repeat
3   Update  $h(\mathbf{f}; \mathbf{f}^{(t-1)} | \mathbf{G}^{(t-1)})$  in (P6);
4   for  $i = 1 : L$  do // In Parallel
5     Update  $\mathbf{f}_i^{(t)}$  by (29);
6   end
7   Update  $\mathbf{G}^{(t)}$  by (9);
8    $t++$ ;
9 until decrease of MSE is small enough or predefined number of iterations is
    reached;

```

5.2. Convergence

355 For the distributive algorithm proposed in the above subsection, we has the following statement

Theorem 5. *The distributive algorithm 2 gives a monotonic decreasing MSE. The solution sequence has limit points. Each limit point of the solution sequence is a stationary point of (P0).*

360 *Proof.* We leave this proof in Appendix Appendix C. □

6. Numerical Results

In this section, numerical results are presented to verify our proposals. First we test our proposed algorithms in two different settings (CASE-1 and CASE-2) in the sequel.

In CASE-1, a wireless sensor network with $L = 3$ sensors is considered. The
365 antenna numbers of the sensors and the fusion center are set as $N_1 = 3, N_2 = 4, N_3 = 5$ and $M = 4$ respectively. All observation matrices $\mathbf{K}_i$ are set as identity matrices. The source signal $\mathbf{s}$ has dimension $K = 3$ with zero mean, unit-power and uncorrelated components. The observation noise at each sensor is colored and has covariance matrix $\Sigma_i = \sigma_i^2 \Sigma_{0,i}, i \in \{1, \dots, L\}$, where the $J_i \times J_i$ matrix $\Sigma_{0,i}$ has the Toeplitz structure

Figure 2: MSE Performance of 2-BCD v.s. $(L+1)$ B-FG Algorithms

with its (j, k) -th element $[\Sigma_{0,i}]_{j,k} = \rho^{|k-j|}$. The parameter ρ is set as 0.5 for all the L sensors. The transmission power and observation noise at each sensor are set as $P_1 = 2$, $P_2 = 2$, $P_3 = 3$, $\sigma_1^{-2} = 6\text{dB}$, $\sigma_2^{-2} = 7\text{dB}$ and $\sigma_3^{-2} = 8\text{dB}$, respectively. In the test of each algorithm, channel noise level increases from $\text{SNR}_0 = 0\text{dB}$ to 18dB. For one specific channel noise level, 500 channel realizations $\{\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3\}$ are randomly generated with each matrix entry following standard complex circular Gaussian distribution $\mathcal{CN}(0, 1)$. The mean square error averaged over all 500 random channel realizations are evaluated as a function of the number of (outer-loop) iterations and the channel SNR.

The 2-BCD algorithm is implemented by utilizing CVX (with SDPT3 solver) to solve its subproblem (P2). For the essentially cyclic $(L+1)$ -BCD algorithm, here we test a special realization: $(L+1)$ -BCD algorithm, where blocks are cyclically updated with each $\mathbf{F}_i$ updated immediately followed by the calibration of $\mathbf{G}$, i.e. $\mathbf{F}_1, \mathbf{G}, \mathbf{F}_2, \mathbf{G}, \dots$, which we denote as “ $(L+1)$ B-FG”. The performance of 2-BCD and the “ $(L+1)$ B-FG” are presented in Fig.2. On average, the $(L+1)$ B-FG algorithm converges in 40-50 outer-loop iterations to the identical MSE as that of the 2-BCD algorithm.

The proximal (with $\kappa = 1$) version of the $(L+1)$ B-FG algorithm are presented in

Figure 3: MSE Performance of 2-BCD v.s. Proximal $(L+1)$ B-FG (CASE-1)

Fig.3. The proximal method, whose convergence to stationary points can be proved, exhibits a rather slow convergence.

Comparatively the performance the approximate-proximal version shown in Fig.4 is surprisingly fast and exhibits excellent performance within only 3 to 5 outer-loop iterations. Here in the first 10 outer-loop iterations, approximate version of $(L+1)$ B-FG is performed and after that the proximal method is used. The approximate-proximal method has the advantage of fast and stationary-point-reaching convergence.

In CASE-2, we apply our proposed algorithms to a second network setting with a random combination of dimensions and power constraints as: $L = 4$, $K = 2$, $M = 3$, $N_1 = 2$, $N_2 = 3$, $N_3 = 4$, $N_4 = 3$, $P_1 = 1$, $P_2 = 2$, $P_3 = 2$, $P_4 = 1$ and the noise level at sensors are set as 9dB, 7dB, 8dB and 10dB respectively. The test results are plotted in Fig.5-Fig.7., with 2-BCD's curves in each figure as bench-mark. Similar numerical results can be observed as those obtained in CASE-1.

Next we compare our algorithms with the one proposed in [9]. We apply our 2-BCD algorithm and the method in [9] to the test scenarios CASE-1 and CASE-2 (specified previously) and present the performance with different iteration numbers and channel noise levels in Fig.8 and Fig.9. It can be seen that our proposed algorithms obtain

Figure 4: MSE Performance of 2-BCD v.s. Approximate-Proximal $(L+1)$ B-FG (CASE-1)

Figure 5: MSE Performance of 2-BCD v.s. $(L+1)$ B-FG Algorithms (CASE-2)

Figure 6: MSE Performance of 2-BCD v.s. Proximal $(L+1)$ B-FG (CASE-2)

Figure 7: MSE Performance of 2-BCD v.s. Approximate-Proximal $(L+1)$ B-FG (CASE-2)

Figure 8: MSE Performance of 2-BCD v.s. [9] (CASE-1)

almost identical convergent MSE with [9].

405 In Fig.10 and Fig. 11 we randomly generate channel realizations at three different channel noise levels (0dB, 4dB and 16dB) and perform our approximate-proximal algorithm and the method in [9] starting from identical initial points for each test. Fig.10 and Fig. 11 clearly suggest that our proposal exhibits much faster convergence than the one in [9].

410 In Figure 12, we examine the performance of the proposed distributive algorithm in Alg.2. We also present the performance of the centralized 2-BCD and the Jacobi distributive algorithm in [19]. Note that the Jacobi algorithm is a widely adopted naive distributive method, where each separate block of variables is updated with the other variables being given as the values of last update. In our experiment, the Jacobi algorithm distributively updates each $\mathbf{F}_i$ with other $\mathbf{F}_j$'s being given as the value of last iteration. Here in Fig.12 two cases are tested: case-(a), $L = 5$, $M = N_i = 4$ and the channel SNR is 6dB; case-(b), $L = 10$, $M = N_i = 3$ and the channel SNR is 2dB. For both test cases, $K = 3$, all the observation matrices are identities and each observation noise is set at the level SNR = 7dB. From Fig.12 we see that the proposed distributive algorithm converges to a very close MSE with that of 2-BCD within 250

420

Figure 9: MSE Performance of 2-BCD v.s. [9] (CASE-2)

Figure 10: MSE Performance of Approximate-Proximal Algorithm v.s. [9] (CASE-1)

Figure 11: MSE Performance of Approximate-Proximal Algorithm v.s. [9] (CASE-2)

iterations. Please note that each iteration of the distributive algorithm is every efficient. The Jacobi algorithm usually does not give monotonically decreasing objective and cannot guarantee its solution converging to stationary points. Fig.12 shows that the Jacobi algorithm usually gives out poor solutions.

425 We present in Table 1 the average MATLAB running time for different algorithms (running in a regular laptop). For simplicity, we focus on the homogeneous sensor network, where each sensor has the same number of antennas and $J_i = N_i$. Different values of K (size of the source vector) and L (number of sensors) are tested to take into account different problem sizes. The algorithms are run multiple times and the average
430 MATLAB running time per outer-loop iteration is recorded. For the 2-BCD algorithm, CVX is utilized to solve its subproblem and the solver SDPT3 is chosen. In Table 1, the average running time of 2-BCD, $(L+1)$ B-FG and the distributive algorithm is presented. Note that the approximate, proximal and approximate-proximal $(L+1)$ B-FG algorithms have the same complexity as that of $(L+1)$ B-FG algorithm. The analytic
435 solutions obtained in Theorem 3 makes it possible for the essentially cyclic $(L+1)$ -BCD algorithm and the distributive algorithm to be implemented highly efficiently.

Figure 12: MSE Performance of 2-BCD v.s. the Jacobi Distributed algorithm in [19] v.s. the Proposed Distributive Algorithm

Table 1: MATLAB Running Time Per Iteration (in Sec.)

Dim. \ L	Algorithms	$L = 5$	$L = 10$	$L = 20$
$K = 3$ $M = 3$ $N_i = 3$	2-BCD	0.5179	0.7471	2.7150
	$(L+1)$ B-FG	0.0681	0.2134	0.8347
	Dstr	0.0149	0.0221	0.0610
$K = 6$ $M = 6$ $N_i = 6$	2-BCD	1.2978	1.9347	2.8435
	$(L+1)$ B-FG	0.0743	0.2533	0.9009
	Dstr	0.0201	0.0410	0.0839
$K = 9$ $M = 9$ $N_i = 9$	2-BCD	8.4613	8.5483	18.9952
	$(L+1)$ B-FG	0.0921	0.2871	1.0069
	Dstr	0.0388	0.0731	0.1574

Notes: (i) SDPT3 solver of CVX is chosen to implement 2-BCD.

7. Conclusion

We have studied the joint transceiver design problem for the wireless sensor network under the MMSE criterion. Since the original problem is nonconvex, the block coordinate descent framework is adopted and three specific mechanisms are developed: 2-BCD, essentially cyclic $(L+1)$ -BCD and a distributive algorithm. Rigorous convergence proof is provided for all the three algorithms.

In particular, we have completely solved the subproblem of single beamformer optimization with one power constraint, by carefully addressing all possible cases. The resultant analytical results not only give birth to highly efficient essentially cyclic $(L+1)$ -BCD and distributive algorithm, but also complete the solutions of many previous beamformer studies in the literature. Our proof of the convergence of the essentially cyclic BCD algorithm also extends naturally to several previous BCD algorithms. Moreover our proposed distributive algorithm serves as a paradigm to optimize the entire transceivers in a decentralized way.

Appendix A. Proof of Theorem 3

Proof. The assumption implies $(\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i) \succ 0$. Therefore $\mathbf{E}_i = (\mathbf{K}_i \Sigma_s \mathbf{K}_i^H + \Sigma_i) \otimes \mathbf{I}_{N_i} \succ 0$. We introduce the following notations

$$\tilde{\mathbf{f}} \triangleq \mathbf{E}_i^{-\frac{1}{2}} \mathbf{f}; \quad (\text{A.1a})$$

$$\mathbf{M}_i \triangleq \mathbf{E}_i^{-\frac{1}{2}} (\mathbf{A}_{ii} + \mathbf{C}_i) \mathbf{E}_i^{-\frac{1}{2}} = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{U}_i^H; \quad (\text{A.1b})$$

$$\mathbf{b}_i \triangleq \mathbf{E}_i^{-\frac{1}{2}} (\mathbf{B}_i^H \mathbf{g} - \mathbf{q}_i); \quad (\text{A.1c})$$

$$\mathbf{p}_i \triangleq \mathbf{U}_i^H \mathbf{b}_i; \quad (\text{A.1d})$$

where the j -th column $\mathbf{u}_{i,j}$ of $\mathbf{U}_i$ is the eigenvector associated with the eigenvalue $\lambda_{i,j} \triangleq [\mathbf{\Lambda}_i]_{j,j}$. Without loss of generality, we assume that the eigenvalues of $\mathbf{M}_i$ are arranged in a decreasing order and that $\mathbf{M}_i$ has rank r_i , $r_i \leq J_i N_i$. In other words

$\lambda_{i,1} \geq \dots \geq \lambda_{i,r_i} > \lambda_{i,r_i+1} = \dots = \lambda_{i,J_i N_i} = 0$.

Then the problem (P2'_i) is rewritten as

$$(P3_i) : \min_{\tilde{\mathbf{f}}_i} \tilde{\mathbf{f}}_i^H \mathbf{M}_i \tilde{\mathbf{f}}_i - 2\text{Re}\{\mathbf{b}_i^H \tilde{\mathbf{f}}_i\}, \quad (\text{A.2a})$$

$$s.t. \|\tilde{\mathbf{f}}_i\|_2^2 \leq P_i. \quad (\text{A.2b})$$

Since $\mathbf{M}_i$ is positive semidefinite, (P3_i) is convex and is obviously strictly feasible. Thus solving (P3_i) is equivalent to solving its KKT conditions:

$$(\mathbf{M}_i + \mu_i \mathbf{I}) \tilde{\mathbf{f}}_i = \mathbf{b}_i; \quad (\text{A.3a})$$

$$\|\tilde{\mathbf{f}}_i\|_2^2 \leq P_i; \quad (\text{A.3b})$$

$$\mu_i (\|\tilde{\mathbf{f}}_i\|_2^2 - P_i) = 0; \quad (\text{A.3c})$$

$$\mu_i \geq 0. \quad (\text{A.3d})$$

The Lagrangian multiplier μ_i should be either positive or zero, our next discussion focuses on identifying the positivity of μ_i .

Assume that $\mu_i > 0$, then $(\mathbf{M}_i + \mu_i \mathbf{I})$ is strictly positive definite and thus invertible. Consequently $\tilde{\mathbf{f}}_i = (\mathbf{M}_i + \mu_i \mathbf{I})^{-1} \mathbf{b}_i$. By the slackness condition (A.3c), the power constraint (A.3b) should be active. Plugging $\tilde{\mathbf{f}}_i$ into (A.3b) and noting the eigenvalue decomposition in (A.1b), we get

$$\|\tilde{\mathbf{f}}_i\|^2 = \mathbf{b}_i^H \mathbf{U}_i (\mathbf{\Lambda}_i + \mu_i \mathbf{I})^{-2} \mathbf{U}_i^H \mathbf{b}_i = P_i. \quad (\text{A.4})$$

By the definition of $\mathbf{p}_i$ in (A.1d), we rewrite (A.4) as

$$\|\tilde{\mathbf{f}}_i\|^2 = g_i(\mu_i) = \sum_{k=1}^{r_i} \frac{|p_{i,k}|^2}{(\lambda_{i,k} + \mu_i)^2} + \sum_{k=r_i+1}^{J_i N_i} \frac{|p_{i,k}|^2}{\mu_i^2} = P_i. \quad (\text{A.5})$$

Note that here $g_i(\mu_i)$ is a positive, continuous and strictly decreasing function in μ_i .

To identify the positivity of μ_i , the following different cases are considered:

460 CASE (I)— $\mu_i^* > 0$ This case further involves two subcases:

case i)— $\exists k \in \{r_i + 1, \dots, J_i N_i\}$ s.t. $|p_{i,k}| \neq 0$:

In this case, it is easily seen that $g_i(\mu_i) \rightarrow +\infty$ when $\mu_i \rightarrow 0^+$, so $g_i(\mu_i)$ has the range of $(0, \infty)$ for positive μ_i . So in case i) there always exists a unique positive μ_i satisfying (A.5). Suppose that the unique solution of (A.5) is μ_i^* . Plugging μ_i^* back

into the KKT condition (A.3a), we obtain the optimal solution $\tilde{\mathbf{f}}_i^*$ as

$$\tilde{\mathbf{f}}_i^* = (\mathbf{M}_i + \mu_i^* \mathbf{I})^{-1} \mathbf{b}_i. \quad (\text{A.6})$$

Plugging (A.1) into the above, (15) is obtained. It is easily verified that the μ_i^* and $\mathbf{f}_i^*$ in (15) satisfy all the KKT conditions in (A.3) and therefore is the optimal solution to (P2').

465 case ii) $\sum_{k=r_i+1}^{J_i N_i} |p_{i,k}|^2 = 0$ and $\sum_{k=1}^{r_i} \frac{|p_{i,k}|^2}{\lambda_{i,k}^2} > P_i$:

In this case, the second part in the summation of $g_i(\mu_i)$ in (A.5) vanishes and $g_i(\mu_i)$ has the bounded range $\left(0, \sum_{k=1}^{r_i} \frac{|p_{i,k}|^2}{\lambda_{i,k}^2}\right]$, with its maximum value achieved at $\mu_i = 0$. When $\sum_{k=1}^{r_i} |p_{i,k}|^2 \lambda_{i,k}^2 > P_i$, a positive μ_i^* satisfying (A.5) still exists and is unique. Consequently, the optimal solution $\mathbf{f}_i^*$ can be determined by (A.6) as in the subcase i).

470 CASE (II)— $\sum_{k=r_i+1}^{J_i N_i} |p_{i,k}|^2 = 0$ and $\sum_{k=1}^{r_i} \frac{|p_{i,k}|^2}{\lambda_{i,k}^2} \leq P_i$

In this case, a positive μ_i satisfying KKT conditions does not exist any more, and $\mu_i^* = 0$. As such, the optimal solution $\mathbf{f}_i^*$ should satisfy (A.3a):

$$\mathbf{M}_i \tilde{\mathbf{f}}_i = \mathbf{b}_i. \quad (\text{A.7})$$

We now claim that the above equation (A.7) has a feasible solution. Indeed, this equation is solvable if and only if the right hand side $\mathbf{b}_i$ belongs to the column space $\mathcal{R}(\mathbf{M}_i)$. Recall that $\mathbf{M}_i$ is Hermitian and has rank r_i ; so $\mathcal{R}(\mathbf{M}_i) = \text{span}(\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,r_i})$ and the null space of $\mathbf{M}_i$ satisfies $\mathcal{N}(\mathbf{M}_i) = \mathcal{R}^\perp(\mathbf{M}_i) = \text{span}(\mathbf{u}_{i,r_i+1}, \dots, \mathbf{u}_{i,J_i N_i})$.
475 In fact, $\mathbb{C}^{J_i N_i} = \mathcal{R}(\mathbf{M}_i) \oplus \mathcal{N}(\mathbf{M}_i)$. Invoking the assumption of CASE (II) that $|p_{i,k}| = 0, \forall k \in \{r_i + 1, \dots, J_i N_i\}$ and the definition of $\mathbf{p}_i$, we obtain $p_{i,k} = \mathbf{u}_{i,k}^H \mathbf{b}_i, \forall k \in \{r_i + 1, \dots, J_i N_i\}$. Actually this implies $\mathbf{b}_i \in \mathcal{N}^\perp(\mathbf{M}_i) = \mathcal{R}(\mathbf{M}_i)$ and thus the consistency (i.e. the feasibility) of (A.7) is guaranteed.

Next we proceed to analytically identify one special feasible solution of (A.7). Eigenvalue decomposing $\mathbf{M}_i$, (A.7) can be equivalently written as

$$\mathbf{\Lambda}_i \mathbf{U}_i^H \tilde{\mathbf{f}}_i = \mathbf{p}_i. \quad (\text{A.8})$$

Denote $\bar{\mathbf{\Lambda}}_i$ as the top-left $r_i \times r_i$ sub-matrix of $\mathbf{\Lambda}_i$, i.e. $\mathbf{\Lambda}_i = \text{diag}\{\bar{\mathbf{\Lambda}}_i, \mathbf{O}_{J_i N_i - r_i}\}$. Let $\bar{\mathbf{U}}_i$ and $\tilde{\mathbf{U}}_i$ represent the left-most r_i columns and the remaining columns of $\mathbf{U}_i$ respectively, i.e. $\mathbf{U}_i = [\bar{\mathbf{U}}_i, \tilde{\mathbf{U}}_i]$. Denote $\bar{\mathbf{p}}_i$ as the first r_i elements of $\mathbf{p}_i$. We can then

simplify (A.8) to

$$\bar{\mathbf{\Lambda}}_i \bar{\mathbf{U}}_i^H \tilde{\mathbf{f}}_i = \bar{\mathbf{p}}_i. \quad (\text{A.9})$$

Since the columns of $\mathbf{U}_i$ form a set of orthonormal basis for $\mathbb{C}^{J_i N_i}$, $\tilde{\mathbf{f}}_i$ can be expressed via columns of $\mathbf{U}_i$ as $\tilde{\mathbf{f}}_i = \sum_{k=1}^{J_i N_i} \alpha_{i,k} \mathbf{u}_{i,k}$. Noticing the key fact that $\bar{\mathbf{U}}_i^H \mathbf{u}_{i,k} = \mathbf{0}, \forall k \in \{r_i+1, \dots, J_i N_i\}$, we know that the values of $\{\alpha_{i,r_i+1}, \dots, \alpha_{i,J_i N_i}\}$ have no impact on (A.9) and can therefore be safely set to zeros to save energy. As for $\alpha_{i,k}, \forall k \in \{1, \dots, r_i\}$, we substitute $\tilde{\mathbf{f}}_i = \sum_{k=1}^{r_i} \alpha_{i,k} \mathbf{u}_{i,k}$ into (A.9) and obtain

$$\alpha_{i,k} = \lambda_{i,k}^{-1} p_{i,k}, \quad \forall k \in \{1, \dots, r_i\}. \quad (\text{A.10})$$

Summarizing the above analysis, the optimal solution $\tilde{\mathbf{f}}_i^*$ to (P3_i) is given by

$$\tilde{\mathbf{f}}_i^* = \mathbf{U}_i \mathbf{\Lambda}_i^\dagger \mathbf{U}_i^H \mathbf{b}_i, \quad (\text{A.11})$$

with $\mathbf{\Lambda}_i^\dagger$ being the Moore-Penrose pseudoinverse of $\mathbf{\Lambda}_i$ given as $\text{diag}\{\bar{\mathbf{\Lambda}}_i^{-1}, \mathbf{0}_{J_i N_i - r_i}\}$. Matrix theory suggests that an arbitrary matrix $\mathbf{X}$ with its singular value decomposition (SVD) given by $\mathbf{X} = \mathbf{U}_\mathbf{X} \mathbf{\Lambda}_\mathbf{X} \mathbf{V}_\mathbf{X}^H$ has its unique Moore-Penrose pseudoinverse $\mathbf{X}^\dagger = \mathbf{V}_\mathbf{X} \mathbf{\Lambda}_\mathbf{X}^\dagger \mathbf{U}_\mathbf{X}^H$, where $\mathbf{U}_\mathbf{X}$ and $\mathbf{V}_\mathbf{X}$ are left and right singular square matrices, respectively, and $\mathbf{\Lambda}_\mathbf{X}$ is a diagonal matrix with appropriate dimensions. Hence, (A.11) can be equivalently written as

$$\tilde{\mathbf{f}}_i^* = \mathbf{M}_i^\dagger \mathbf{b}_i. \quad (\text{A.12})$$

Obviously $\mu_i^* = 0$, and μ_i^* and $\tilde{\mathbf{f}}_i^*$ satisfy the KKT conditions (A.3a), (A.3c) and (A.3d). What remains to be shown is that $\tilde{\mathbf{f}}_i^*$ satisfies the power constraint. We verify this using (A.10) and get

$$\|\tilde{\mathbf{f}}_i^*\|^2 = \sum_{k=1}^{r_i} |\alpha_{i,k}|^2 = \sum_{k=1}^{r_i} \frac{|p_{i,k}|^2}{\lambda_{i,k}^2} \leq P_i, \quad (\text{A.13})$$

where the inequality in the above follows the assumption of CASE (II). Plugging (A.1)

into (A.12), (16) is obtained. The proof is complete. $\square$

Appendix B. Proof of Theorem 4

Proof. This proof is inspired by Proposition 2.7.1 in [20]. To simplify the following exposition, we define $\mathbf{x} \triangleq [\mathbf{x}_1^T, \dots, \mathbf{x}_{L+1}^T] = [\mathbf{f}_1^T, \dots, \mathbf{f}_L^T, \mathbf{g}^T]$ and $\mathbf{x} \in \mathcal{X} \triangleq \mathcal{X}_1 \times$

$\cdots \times \mathcal{X}_{L+1}$ with $\mathcal{X}_i = \{\mathbf{f}_i \in \mathbb{C}^{J_i N_i} \mid \mathbf{f}_i^H \mathbf{E}_i \mathbf{f}_i \leq P_i\}$, for $i = 1, \dots, L$ and $\mathcal{X}_{L+1} =$
485 $\mathbb{C}^{KM}$. For any specific essentially cyclic update BCD algorithm, we assume that it starts from an initial feasible solution $\mathbf{x}^{(0)} \triangleq [\mathbf{x}_1^{T(0)}, \dots, \mathbf{x}_{L+1}^{T(0)}]$ and the iteration index (k) increases by one after any block's update. Denote $\mathbf{x}_i^{(k)}$ as the i -th block of $\mathbf{x}^{(k)}$ and $\mathbf{x}_{\bar{i}} = [\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_{L+1}]$, $i \in \{1, \dots, L+1\}$, $i \in \{1, \dots, L+1\}$. Assume that T is a period of the essentially cyclic update rule and $\{t_1, \dots, t_T\}$,
490 with $t_j \in \{1, \dots, L+1\} \forall j \in \{1, \dots, T\}$, as the indices of the updated blocks in a period in order. If $\mathbf{x}_{t_j}$ is updated in the (k) -th iteration, then $\mathbf{x}_{t_{j \oplus 1}}$ is updated in the $(k+1)$ -th iteration. Define $j \oplus 1 \triangleq j \pmod{T} + 1, \forall j \in \{1, \dots, T\}$ and $j \oplus m$ as $j \oplus 1$ by m times.

By repeatedly invoking Bolzano-Weierstrass theorem to $\mathbf{f}_i$ to $\mathbf{f}_L$ and noticing that
495 $\mathbf{g}$ is updated in closed form by equation (9), the existence of limit points of $\{\mathbf{x}^{(k)}\}_{k=0}^\infty$ can be proved.

Then we prove that $\text{MSE}^{(k)}$ is decreasing. If $\mathbf{x}_{L+1}$ (or $\mathbf{g}$) is updated in the $(k+1)$ -th iteration, then (P1) is solved and thus MSE is decreasing. Assume that in the $(k+1)$ -th iteration, the $(t_{j \oplus 1})$ -th block is updated, $t_{j \oplus 1} \in \{1, \dots, L\}$. Then

$$\mathbf{x}_{t_{j \oplus 1}}^{(k+1)} = \arg \min_{\mathbf{x}_{t_{j \oplus 1}} \in \mathcal{X}_{t_{j \oplus 1}}} \text{MSE}(\mathbf{x}_{t_{j \oplus 1}} \mid \mathbf{x}_{\bar{t}_{j \oplus 1}}^{(k)}) + \kappa \|\mathbf{x}_{t_{j \oplus 1}} - \mathbf{x}_{t_{j \oplus 1}}^{(k)}\|_2^2.$$

Since $\mathbf{x}_{t_j}^{(k)}$ is feasible, it should give no smaller objective than $\mathbf{x}_{t_j}^{(k+1)}$ for the above problem. This implies

$$\text{MSE}(\mathbf{x}^{(k+1)}) \leq \text{MSE}(\mathbf{x}^{(k)}) - \kappa \|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\|_2^2 \leq \text{MSE}(\mathbf{x}^{(k)}).$$

Thus $\text{MSE}^{(k)}$ is decreasing. At the same time notice that MSE should be nonnegative, thus $\text{MSE}^{(k)}$ converges.

Next we prove that any limit point is stationary. Assume that a subsequence of
500 solution $\mathbf{x}^{(k_j)}$ converges to a limit point $\bar{\mathbf{x}} \triangleq [\bar{\mathbf{x}}_1^T, \dots, \bar{\mathbf{x}}_{L+1}^T]$. Since there are finite blocks, we assume the block $i \in \{1, \dots, L+1\}$ is updated infinitely many times and assume that $i = t_l$ for some $l \in \{1, \dots, T\}$. It should be noted that such l may be non-unique and arbitrary one can be chosen to do the job.

We assert that $\mathbf{x}^{(k_j+1)} \rightarrow \bar{\mathbf{x}}$, i.e. $\mathbf{x}_{t_{l \oplus 1}}^{(k_j+1)} \rightarrow \bar{\mathbf{x}}_{t_{l \oplus 1}}$. This claim can be proved in two
505 cases—i) $t_{l \oplus 1} = L+1$ and ii) $t_{l \oplus 1} \in \{1, \dots, L\}$.

i) $t_{l\oplus 1} = L + 1$. Notice that $\mathbf{x}_{L+1} = \mathbf{g}$ is updated in a closed form (9), which is a continuous function of $[\mathbf{x}_1^T, \dots, \mathbf{x}_L^T]$. Since $\mathbf{x}_{\frac{L+1}{2}}^{(k_j)}$ converges, by taking $j \rightarrow \infty$, $\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)}$ should converge to some limit, i.e. $\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} \rightarrow \tilde{\mathbf{x}}_{L+1}$. Notice that $\text{MSE}^{(k)}$ converges, so $\text{MSE}(\bar{\mathbf{x}}_{\frac{L+1}{2}}, \bar{\mathbf{x}}_{L+1}) = \text{MSE}(\bar{\mathbf{x}}_{\frac{L+1}{2}}, \tilde{\mathbf{x}}_{L+1})$. This means both $\bar{\mathbf{x}}_{L+1}$ and $\tilde{\mathbf{x}}_{L+1}$ are solutions to the problem (P1) with sensors' beamformers $[\bar{\mathbf{x}}_1^T, \dots, \bar{\mathbf{x}}_L^T]$ given. Since (P1) is strictly convex and thus has unique solution, we conclude $\tilde{\mathbf{x}}_{L+1} = \bar{\mathbf{x}}_{L+1}$. So $\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} \rightarrow \bar{\mathbf{x}}_{t_{l\oplus 1}}$ holds for the case $t_{l\oplus 1} = L + 1$.

ii) $t_{l\oplus 1} \in \{1, \dots, L\}$. By contradiction, we assume that $\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)}$ does not converge to $\bar{\mathbf{x}}_{t_{l\oplus 1}}$. By denoting $\gamma^{(k_j)} \triangleq \|\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} - \bar{\mathbf{x}}_{t_{l\oplus 1}}\|_2$ and possibly restricting to a subsequence, we assume that there exists a $\bar{\gamma} > 0$ such that $\gamma^{(k_j)} \geq \bar{\gamma}$ for all j . Let $\mathbf{s}_l^{(k_j)} = (\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} - \mathbf{x}_{t_{l\oplus 1}}^{(k_j)})/\gamma^{(k_j)}$. Since $\mathbf{s}^{(k_j)}$ is bounded, by Bolzano-Weierstrass theorem and restricting to a subsequence, we assume that $\mathbf{s}^{(k_j)} \rightarrow \bar{\mathbf{s}}$. Then we obtain

$$\text{MSE}(\mathbf{x}^{(k_j+1)}) = \text{MSE}(\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}) \quad (\text{B.1})$$

$$\leq \text{MSE}(\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}) + \kappa \|\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} - \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}\|_2^2 \quad (\text{B.2})$$

$$= \text{MSE}(\mathbf{x}_{t_{l\oplus 1}}^{(k_j)} + \gamma^{(k_j)} \mathbf{s}^{(k_j)} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}) + \kappa \|\gamma^{(k_j)} \mathbf{s}^{(k_j)}\|_2^2 \quad (\text{B.3})$$

$$\leq \text{MSE}(\mathbf{x}_{t_{l\oplus 1}}^{(k_j)} + \epsilon \bar{\gamma} \mathbf{s}^{(k_j)} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}) + \kappa \|\epsilon \bar{\gamma} \mathbf{s}^{(k_j)}\|_2^2, \forall \epsilon \in [0, 1] \quad (\text{B.4})$$

$$\leq \text{MSE}(\mathbf{x}_{t_{l\oplus 1}}^{(k_j)} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}) = \text{MSE}(\mathbf{x}^{(k_j)}), \quad (\text{B.5})$$

where the last two inequalities follow the fact that $\text{MSE}(\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}) + \kappa \|\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} - \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}\|_2^2$ is strictly convex and attains the minimum at point $\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)}$. Noting $\text{MSE}^{(k_j)}$ converges and letting $j \rightarrow \infty$, we obtain

$$\begin{aligned} \text{MSE}(\bar{\mathbf{x}}) &\leq \text{MSE}(\bar{\mathbf{x}}_{t_{l\oplus 1}} + \epsilon \bar{\gamma} \bar{\mathbf{s}} | \bar{\mathbf{x}}_{t_{l\oplus 1}}) + \kappa \epsilon^2 \bar{\gamma}^2 \\ &\leq \text{MSE}(\bar{\mathbf{x}}), \quad \forall \epsilon \in [0, 1], \end{aligned} \quad (\text{B.6})$$

which immediately implies

$$\text{MSE}(\bar{\mathbf{x}}_{t_{l\oplus 1}} + \epsilon \bar{\gamma} \bar{\mathbf{s}} | \bar{\mathbf{x}}_{t_{l\oplus 1}}) + \kappa \epsilon^2 \bar{\gamma}^2 = \text{MSE}(\bar{\mathbf{x}}), \forall \epsilon \in [0, 1]. \quad (\text{B.7})$$

However the above is impossible. Notice that $\text{MSE}(\bar{\mathbf{x}}_{t_{l\oplus 1}} + \epsilon \bar{\gamma} \bar{\mathbf{s}} | \bar{\mathbf{x}}_{t_{l\oplus 1}})$ is a quadratic function of ϵ with nonnegative quadratic coefficient and $\bar{\gamma}, \kappa > 0$. Thus the left hand side(LHS) of equation (B.7) is a strictly convex quadratic function of ϵ , which has at

most two different ϵ giving the function value of $\text{MSE}(\bar{\mathbf{x}})$. Contradiction has been reached.

In the above we have proved that $\mathbf{x}^{(k_j+1)} \rightarrow \bar{\mathbf{x}}$. Next we show that $\nabla_{\mathbf{x}_{t_{l\oplus 1}}} \text{MSE}(\bar{\mathbf{x}})^T (\mathbf{x}_{t_{l\oplus 1}} - \bar{\mathbf{x}}_{t_{l\oplus 1}}) \geq 0, \forall \mathbf{x}_{t_{l\oplus 1}} \in \mathcal{X}_{t_{l\oplus 1}}$, which is also proved in two cases:

When $t_{l\oplus 1} \in \{1, \dots, L\}$, we have

$$\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} = \arg \min_{\mathbf{x}_{t_{l\oplus 1}} \in \mathcal{X}_{t_{l\oplus 1}}} \text{MSE}(\mathbf{x}_{t_{l\oplus 1}} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}) + \kappa \|\mathbf{x}_{t_{l\oplus 1}} - \mathbf{x}_{t_{l\oplus 1}}^{(k_j)}\|_2^2.$$

By optimality condition, the above implies

$$\begin{aligned} & \nabla_{\mathbf{x}_{t_{l\oplus 1}}} \text{MSE}(\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} | \mathbf{x}_{t_{l\oplus 1}}^{(k_j)})^T (\mathbf{x}_{t_{l\oplus 1}} - \mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)}), \\ & + 2\kappa (\mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)} - \mathbf{x}_{t_{l\oplus 1}}^{(k_j)})^T (\mathbf{x}_{t_{l\oplus 1}} - \mathbf{x}_{t_{l\oplus 1}}^{(k_j+1)}) \geq 0, \forall \mathbf{x}_{t_{l\oplus 1}} \in \mathcal{X}_{t_{l\oplus 1}}. \end{aligned} \quad (\text{B.8})$$

Let $j \rightarrow \infty$ in the above equation and note that MSE is continuously differentiable, we obtain

$$\nabla_{\mathbf{x}_{t_{l\oplus 1}}} \text{MSE}(\bar{\mathbf{x}})^T (\mathbf{x}_{t_{l\oplus 1}} - \bar{\mathbf{x}}_{t_{l\oplus 1}}) \geq 0, \forall \mathbf{x}_{t_{l\oplus 1}} \in \mathcal{X}_{t_{l\oplus 1}}. \quad (\text{B.9})$$

520 When $t_{l\oplus 1} = L + 1$, the above reasoning still works except that the proximal term is absent (i.e. $\kappa = 0$). So we also obtain $\nabla_{\mathbf{x}_{t_{l\oplus 1}}} \text{MSE}(\bar{\mathbf{x}})^T (\mathbf{x}_{t_{l\oplus 1}} - \bar{\mathbf{x}}_{t_{l\oplus 1}}) \geq 0$.

Now replace the subsequence $\{k_j\}$ with $\{k_j + 1\}$, $t_{l\oplus 1}$ with $t_{l\oplus 2}$ and utilize the verbatim argument as above, we can prove

$$\nabla_{\mathbf{x}_{t_{l\oplus 2}}} \text{MSE}(\bar{\mathbf{x}})^T (\mathbf{x}_{t_{l\oplus 2}} - \bar{\mathbf{x}}_{t_{l\oplus 2}}) \geq 0, \forall \mathbf{x}_{t_{l\oplus 2}} \in \mathcal{X}_{t_{l\oplus 2}}. \quad (\text{B.10})$$

Repeating this argument for $(T - 1)$ times and recalling that for essentially cyclic update rule, $\{t_{l\oplus 1}, \dots, t_{l\oplus T}\} = \{1, \dots, L\}$, we have proved that

$$\nabla_{\mathbf{x}_i} \text{MSE}(\bar{\mathbf{x}})^T (\mathbf{x}_i - \bar{\mathbf{x}}_i) \geq 0, \forall \mathbf{x}_i \in \mathcal{X}_i, \forall i \in \{1, \dots, L+1\}. \quad (\text{B.11})$$

Summing up the above $(L + 1)$ inequalities, we obtain

$$\nabla_{\mathbf{x}} \text{MSE}(\bar{\mathbf{x}})^T (\mathbf{x} - \bar{\mathbf{x}}) \geq 0, \forall \mathbf{x} \in \mathcal{X}. \quad (\text{B.12})$$

So $\bar{\mathbf{x}}$ is actually a stationary point of (P0). $\square$

Appendix C. Proof of Theorem 5

Proof. First we prove that the obtained objective sequence for both problems is monotonic. Obviously the update of $\mathbf{G}$ will not increase MSE. We have

$$\begin{aligned} & \text{MSE}(\{\mathbf{F}_i^{(t+1)}\}_{i=1}^L, \mathbf{G}^{(t+1)}) \\ & \leq \text{MSE}(\{\mathbf{F}_i^{(t+1)}\}_{i=1}^L | \mathbf{G}^{(t)}) \stackrel{(a)}{\leq} h(\mathbf{f}^{(t+1)}; \mathbf{f}^{(t)} | \mathbf{G}^{(t)}) \\ & \stackrel{(b)}{\leq} h(\mathbf{f}^{(t)}; \mathbf{f}^{(t)} | \mathbf{G}^{(t)}) \stackrel{(c)}{=} \text{MSE}(\{\mathbf{F}_i^{(t)}\}_{i=1}^L, \mathbf{G}^{(t)}). \end{aligned} \quad (\text{C.1})$$

where (a) and (c) follows (27) and (b) follows the fact that $\mathbf{f}^{(t+1)}$ is the solution to (P6) approximated at $\mathbf{f}^{(t)}$. Thus the MSE sequence is non-increasing. So the objective MSE converges. Denote its limit as $\bar{s}$.

The existence of limit points of the solution sequence follows the arguments in the proof of Theorem 2. Assume that $(\mathbf{f}^{(t_k)}, \mathbf{G}^{(t_k)})$ is a subsequence converging to $(\bar{\mathbf{f}}, \bar{\mathbf{G}})$. Obviously $\bar{\mathbf{f}}$ is feasible. By (C.1), for arbitrary $\mathbf{f}$ we have

$$h(\bar{\mathbf{f}}; \bar{\mathbf{f}} | \bar{\mathbf{G}}) = -\bar{s}, \quad (\text{C.2a})$$

$$h(\mathbf{f}; \mathbf{f}^{(t_k)} | \mathbf{G}^{(t_k)}) \geq h(\mathbf{f}^{(t_k+1)}; \mathbf{f}^{(t_k)} | \mathbf{G}^{(t_k)}). \quad (\text{C.2b})$$

Taking $k \rightarrow \infty$ to both sides of the inequality in (C.2b) we obtain

$$h(\mathbf{f}; \bar{\mathbf{f}} | \bar{\mathbf{G}}) \geq -\bar{s}, \forall \mathbf{f}. \quad (\text{C.3})$$

Combining (C.2a) and (C.2b), it is known that $\bar{\mathbf{f}}$ is the optimal solution of (P6) with the objective $h(\mathbf{f}; \bar{\mathbf{f}} | \bar{\mathbf{G}})$. As optimality condition, $\bar{\mathbf{f}}$ is stationary, which implies

$$\nabla_{\mathbf{f}} h(\bar{\mathbf{f}})^T (\mathbf{f} - \bar{\mathbf{f}}) \geq 0, \quad \forall \mathbf{f} \in \Pi_{i=1}^L \mathcal{X}_i, \quad (\text{C.4})$$

where $\mathcal{X}_i$ is defined in the proof of Theorem 4. Recall that $h(\cdot)$ is an approximation of $\text{MSE}(\cdot)$ by introducing second order Taylor expansion at the point $\bar{\mathbf{f}}$ (see (24)). So $h(\cdot)$ and $\text{MSE}(\cdot)$ have identical first order derivative at the point $\bar{\mathbf{f}}$. Thus the (C.4) can be equivalently written as

$$\nabla_{\mathbf{f}} \text{MSE}(\bar{\mathbf{f}})^T (\mathbf{f} - \bar{\mathbf{f}}) \geq 0, \quad \forall \mathbf{f} \in \Pi_{i=1}^L \mathcal{X}_i, \quad (\text{C.5})$$

Following the same arguments in the proof for Theorem 4, we can prove that

$$\nabla_{\mathbf{g}} \text{MSE}(\bar{\mathbf{g}})^T (\mathbf{g} - \bar{\mathbf{g}}) \geq 0, \quad \forall \mathbf{g}. \quad (\text{C.6})$$

(C.5) and (C.6) imply that $(\{\bar{\mathbf{F}}_i\}_{i=1}^L, \bar{\mathbf{G}})$ is in fact a stationary point for (P0). $\square$

Acknowledgement

This work was supported by National Science Foundation under Grants No.1133027
 530 and No.1343372.

References

- [1] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science*. New York: Springer, Apr. 2003, vol. 2634, pp. 162-177.
- [2] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," presented at the *Inf. Theory Appl. Workshop*, San Diego, CA, Jan. 2007.
 535
- [3] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1631-1639, May 2005.
- [4] J. Fang and H. Li, "Optimal/near-optimal dimensionality reduction for distributed estimation in homogeneous and certain inhomogeneous scenarios," *IEEE Trans. Signal Process.*,
 540 vol. 58, no. 8, pp. 4339-4353, 2010.
- [5] 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, July 2009.
- [6] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data,"
 545 *IEEE Trans. Signal Processing*, vol. 57, no. 8, pp. 3292-3297, Aug. 2009.
- [7] J. Xiao, S. Cui, Z. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757-770, Feb. 2008.
- [8] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4284-4299,
 550 Aug. 2007.

- [9] A. S. Behbahani, A. M. Eltawil, H. Jafarkhani, "Linear Decentralized Estimation of Correlated Data for Power-Constrained Wireless Sensor Networks," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 6003-6016, Nov. 2012.
- [10] J. Fang, H. Li, Z. Chen and Y. Gong, "Joint Precoder Design for Distributed Transmission of Correlated Sources in Sensor Networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2918-2929, June 2013.
- [11] Y. Liu, J. Li, X. Lu and C. Yuen, "Optimal Linear Precoding and Postcoding for MIMO Multi-Sensor Noisy Observation Problem," *IEEE International Conference on Communications (ICC)*, Sydney, Jun. 2014.
- [12] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, pp. 214-226, Jan. 2004.
- [13] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331-4340, Dec. 2011.
- [14] M. R. A. Khandaker and Y. Rong, "Joint transceiver Optimization for multiuser MIMO relay communication systems," *IEEE Trans. Signal Processing*, vol. 60, pp. 5977-5986, Nov. 2012.
- [15] C. Xing, S. Li, Z. Fei and J. Kuang, "How to Understand Linear Minimum Mean-Square-Error Transceiver Design for Multiple-Input-Multiple-Output Systems from Quadratic Matrix Programming," *IET Communications*, vol. 7, no. 12, pp. 1231-1242, Aug. 2013.
- [16] D. Ciuonzo, G. Papa, G. Romano, P. Salvo Rossi and P. Willett, "One-bit decentralized detection with a Rao test for multisensor fusion." *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 861-864, Sep. 2013.
- [17] A. Shirazinia, S. Dey, D. Ciuonzo, and P. Salvo Rossi, "Massive MIMO for decentralized estimation over coherent multiple access channels," *International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015.
- [18] A. Shirazinia, S. Dey, D. Ciuonzo and P. Salvo Rossi, "Massive MIMO for Decentralized Estimation of a Correlated Source." *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2499-2512, Jan. 2016.

- 580 [19] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [20] D. Bertsekas, *Nonlinear Programming, 2nd ed*, Belmont, MA: Athena Scientific, 1999.
- [21] L. Grippo, and M. Sciandrone, “On the convergence of the block nonlinear Gauss-Seidel method under convex constraints,” *Operations Research Letters*, vol. 26, pp. 127-136, 585 2000.
- [22] P. Tseng, “Convergence of a block coordinate descent method for nondifferentiable minimization,” *Journ. of Optim. Theory and Appl.*, vol. 109, no. 2, pp. 475-494, June 2001.
- [23] M.J.D. Powell, “On search direction for minimization algorithms,” *Math. Programming*, vol. 4, no. 1, pp. 193-201 1973.
- 590 [24] T. P. Minka. *Old and New Matrix Algebra Useful for Statistics*, Notes, December 2000.
- [25] S. Haykin, *Adaptive Filter Theory (4th ed.)*, Upper Saddle River, NJ: Prentice-Hall, 2002.
- [26] S. Boyd, and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.
- [27] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming (web 595 page and software),” [Online]. Available: <http://cvxr.com/cvx> Apr. 2010
- [28] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, NY, 2nd ed., 2006.
- [29] D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” *Amer. Statistician*, vol. 58, no. 1, pp. 30-37, 2004.