Joint Precoder Design for Distributed Transmission of Correlated Sources in Sensor Networks

Jun Fang, Hongbin Li, *Senior Member, IEEE*, Zhi Chen, and Yu Gong

***Abstract*—We consider the problem of transmitting multiple spatially distributed correlated sources to a common destination (e.g. a fusion center or an access point) in wireless sensor networks (WSNs). The correlated data from multiple sensors are jointly transmitted to the destination via orthogonal channels. We assume that the channel between each sensor and the receiver is multiple-input multiple-output (MIMO), with each sensor and the receiver equipped with multiple transmit/receive antennas. In this framework, we study the problem of joint linear precoder design for all sensors by assuming the knowledge of the instantaneous channel state information (CSI), aiming at maximizing the mutual information between the sources and the received signals at the destination. We propose a Gauss-Seidel iterative approach which successively optimizes the precoding matrix associated with each sensor, while fixing the other precoding matrices. Numerical results show that the proposed algorithm that takes into account the spatial correlation across sensors can achieve higher capacity than conventional methods that neglect the spatial correlation.**

***Index Terms*—Precoder design, distributed correlated sources, Gauss-Seidel approach, sensor networks.**

I. Introduction

WIRELESS sensor networks (WSNs) have been of significant interest over the past few years due to their potential applications in environment monitoring, battlefield surveillance, target localization and tracking [1], [2]. In many monitoring/surveillance tasks, sensors are deployed to monitor spatio-temporally correlated physical processes. For example, sensors can be dispersed in a wild forest, collecting meteorological measurements (such as pressure, humidity, or temperature) for the purpose of obtaining a long term historical record and building models on the observed ecosystem. The sensor data, therefore, often exhibit both spatial and temporal correlations. These correlations can be exploited for efficient algorithm design to enhance the spectral/energy-efficiency of the network. A multitude of studies along this line have appeared recently in the context of data aggregation (e.g. [3], [4]), distributed detection (e.g. [5]–[7]), distributed compression-estimation (e.g. [8]–[12]), and distributed source coding (e.g. [13]–[16]).

This paper considers the problem of jointly transmitting multiple spatially distributed correlated sources to a common destination over noisy channels. Such a problem arises in sensor networks when multiple sensors observe a spatio-temporal physical process and send their data to a common destination (such as a fusion center) for processing and data fusion. As another example, in wireless surveillance systems, a set of wireless video camera sensors are deployed to provide different views of the same area or target of interest. There is significant inter-sensor correlation due to overlapping views between different cameras. In these scenarios, we are interested in studying the joint design of linear precoders for all nodes, intending to achieve a maximum overall channel capacity. As already pointed out in previous studies, precoding can improve the network spectral/energy efficiency through altering the effective channels in a beneficial manner.

Linear precoder design has been extensively investigated in a number of studies [17]–[29]. Among them, most [17]–[22] considered precoding for single-user point-to-point MIMO communications. Precoder and transceiver optimization for multiuser MIMO systems was examined in [23]–[25]. Nevertheless, in these works [23]–[25], sources are assumed mutually independent. The independence assumption is reasonable for cellular communications where each user is a unit working independently of other users. However, this assumption is not valid for sensor network scenarios where multiple sensors usually collaborate to monitor a common phenomena. Also, in these works [23]–[25], sources are transmitted to a common receiver via a coherent multiple access channel. Although spectral efficiency can be preserved by letting users transmit simultaneously via coherent multiple access channels, coherent transmission requires strict transmitter synchronization and causes substantial interference between data streams. Precoding was also commonly used in multiple antenna amplify-and-forward relay networks, e.g. [26]–[29]. In these works, the objective of precoding design for relay networks is to maximize the mutual information between the original single source and the reconstruction at the destination.

In this paper, linear precoder design for multiple spatially distributed but correlated sources is studied. Sources are assumed being transmitted to a common destination via orthogonal channels. In particular, orthogonal channels can be realized by multiaccess techniques such as frequency division multiple access (FDMA) which does not require strict synchronization among sensors. Under certain power constraints, we study

Manuscript received August 15, 2012; revised January 14, 2013; accepted March 25, 2013. The associate editor coordinating the review of this paper and approving it for publication was M. Rossi.

J. Fang and Z. Chen are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: {junfang, chenzhi}@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, 07030, USA (e-mail: hongbin.li@stevens.edu).

Y. Gong is with the School of Electronic, Electrical and System Engineering, Loughborough University, Leicestershire LE11 3TU UK (e-mail: y.gong@lboro.ac.uk).

This work was supported in part by the National Science Foundation of China under Grant 61172114, the National Science Foundation under Grant ECCS-0901066, the Program for New Century Excellent Talents in University (China) under Grant NCET-09-0261, and the National High-Tech R&D Program of China under Grant 2012AA011402.

Digital Object Identifier 10.1109/TCOMM.2013.050613.121221

the joint design of linear precoders for all sensors, aiming at maximizing the mutual information between sources and the received signals at the destination. This optimization problem is generally non-concave due to the block-diagonal structure imposed on the overall precoding matrix, and a closed-form globally optimal solution is difficult to obtain. To circumvent this difficulty, we propose a Gauss-Seidel iterative algorithm which successively optimizes the precoding matrix associated with each sensor, with other precoding matrices fixed. Our numerical results show that, by exploiting the spatial correlation among sources, the proposed algorithm achieves higher capacity than conventional methods that neglect the spatial correlation. Another important contribution of this paper is that, using majorization theory, we discovered and proved a new determinant inequality. This new inequality is crucial in solving our optimization problem that has a generalized form of the one studied in previous works, e.g. [19].

We note that there is a rich literature on joint design of non-linear precoders for multiple correlated distributed sources (e.g. [13]–[16]). Most of these works were studied in the framework of distributed source coding, in which the sources are firstly quantized, and then followed by a non-linear encoding that maps the quantization indices to a smaller number of transmission indices. The ultimate goal is to increase the rate-distortion efficiency through exploiting the correlation across sensors. Nevertheless, design of such non-linear precoders for MIMO systems is a daunting task. In this case, linear precoders considered in this work could be a practically feasible choice.

The following notations are used throughout this paper. The notations $[\cdot]^T$ and $[\cdot]^H$ stand for transpose and Hermitian transpose, respectively. $\mathrm{tr}(\mathbf{A})$ and $\det(\mathbf{A})$ denote the trace and the determinant of the matrix $\mathbf{A}$, respectively. The symbols $\mathbb{R}^{n\times m}$ and $\mathbb{R}^n$ stand for the set of $n\times m$ matrices and the set of n-dimensional column vectors with real entries, respectively. $\mathbb{C}^{n\times m}$ and $\mathbb{C}^n$ denote the set of $n \times m$ matrices and the set of n-dimensional column vectors with complex entries, respectively.

The rest of the paper is organized as follows. In Section II, we introduce the data model, basic assumptions, and the precoding design problem. The generally non-concave precoding design problem is then studied in Section III and a Gauss-Seidel iterative algorithm is developed for joint precoder design. In Section IV, numerical results are presented to examine the convergence properties and the effectiveness of the proposed algorithm, followed by concluding remarks in Section V.

II. Problem Formulation

Consider a distributed wireless sensor network in which N correlated sources are transmitted to a common destination through orthogonal multiple access channels (see Fig. 1). Examples of such multiaccess schemes include time/frequency division multiple access (TDMA/FDMA). Also, we assume that sensors and the receiver are equipped with multiple antennas. The received signal at the destination can therefore be written as:

$$\mathbf{y}_n = \mathbf{H}_n\mathbf{C}_n\mathbf{x}_n + \mathbf{v}_n \quad n = 1, \ldots, N \tag{1}$$

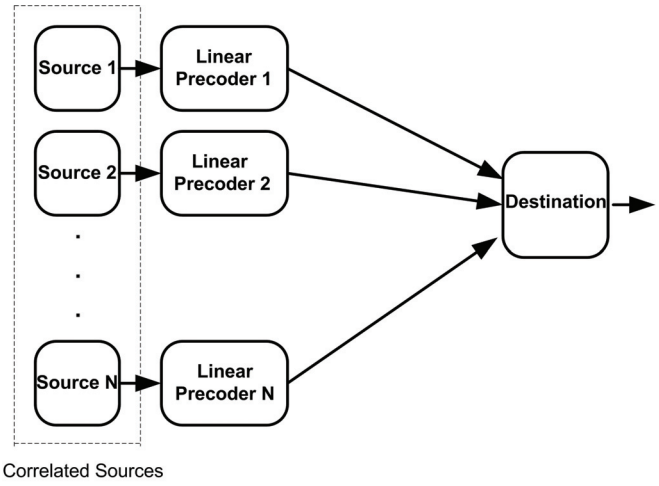


Fig. 1. Precoding design for multiple correlated sources which are jointly transmitted to a common destination.

where $\mathbf{x}_n \in \mathbb{C}^{p_n}$ and $\mathbf{y}_n \in \mathbb{C}^q$ denote the nth node's symbol vector and the received signal from node n, respectively, $\mathbf{C}_n \in \mathbb{C}^{p_n\times p_n}$ is a square source precoding matrix used to improve the spectral/energy efficiency (the extension to a non-square precoding matrix will be discussed in Section III-C), $\mathbf{H}_n \in \mathbb{C}^{q\times p_n}$ is the instantaneous MIMO channel matrix between node n and the destination, and $\mathbf{v}_n \in \mathbb{C}^q$ is zero-mean circularly symmetric complex Gaussian channel noise with arbitrary covariance matrix $\mathbf{R}_{v_n}$. Note that the assumption of multiple antennas per sensor node can be justified as follows. In practice, a sensor network could consist of two types of nodes [30]: low-cost sensor nodes with basic sensing and communication abilities, and high-end nodes equipped with more powerful transceivers and multiple antennas. Those high-end nodes can serve as cluster heads, collecting sensing data from their respective neighboring sensors and forwarding data to a common destination. In addition, even if each sensor has a single antenna, we can still convert a single-input multiple-output (SIMO) channel into a virtual MIMO model (1) using temporal multiplexing. The virtual channel matrix $\mathbf{H}_n$ in this case is a diagonal matrix with identical or non-identical diagonal elements, depending on the channel time-variation characteristics, which, however, makes no difference to the precoding design.

Let $\mathbf{y} \triangleq [\mathbf{y}_1^T\ \mathbf{y}_2^T\ \ldots\ \mathbf{y}_N^T]^T$ denote a column vector formed by stacking the data received from all source nodes. We have

$$\mathbf{y} = \mathbf{H}\mathbf{C}\mathbf{x} + \mathbf{v} \tag{2}$$

where $\mathbf{x} \triangleq [\mathbf{x}_1^T\ \mathbf{x}_2^T\ \ldots\ \mathbf{x}_N^T]^T$, $\mathbf{v} \triangleq [\mathbf{v}_1^T\ \mathbf{v}_2^T\ \ldots\ \mathbf{v}_N^T]^T$, $\mathbf{H} \triangleq \mathrm{diag}\{\mathbf{H}_1, \ldots, \mathbf{H}_N\}$, and $\mathbf{C} \triangleq \mathrm{diag}\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ is the overall precoding matrix with block-diagonal structure. The covariance matrices of the transmitted vector $\mathbf{x}$ and the noise vector $\mathbf{v}$ are given by $\mathbf{R}_x$ and $\mathbf{R}_v$, respectively, and their knowledge is assumed available at the destination. Since the sources are sent to the destination via orthogonal channels, it is reasonable to assume that the additive noise $\mathbf{v}$ is independent across channels. Therefore the covariance matrix $\mathbf{R}_v$ has a block-diagonal structure, with each diagonal block given as $\mathbf{R}_{v_n} = E[\mathbf{v}_n\mathbf{v}_n^H]$.

Assume that the overall source $\mathbf{x}$ is memoryless and zero mean circularly symmetric complex Gaussian, and the instantaneous channel state information (CSI) $\{\mathbf{H}_n\}$ is known at the destination. The mutual information between the source and the received signal at the destination is given by [31]

$$\mathcal{I}(\mathbf{x},\mathbf{y}) = \log\det(\mathbf{I}+\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C}\mathbf{R}_x\mathbf{C}^H\mathbf{H}^H) \tag{3}$$

The mutual information can also be expressed as

$$\begin{aligned}\mathcal{I}(\mathbf{x},\mathbf{y}) &= \log\det(\mathbf{I}+\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C}\mathbf{R}_x\mathbf{C}^H\mathbf{H}^H)\\ &\overset{(a)}{=} \log\det(\mathbf{I}+\mathbf{R}_x^{\frac{1}{2}}\mathbf{C}^H\mathbf{H}^H\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C}\mathbf{R}_x^{\frac{1}{2}})\\ &\overset{(b)}{=} \log c + \log\det(\mathbf{R}_x^{-1}+\mathbf{C}^H\mathbf{H}^H\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C})\\ &= \log c + \log\det(\mathbf{R}_\epsilon^{-1})\end{aligned} \tag{4}$$

where (a) follows from the determinant identity $\det(\mathbf{I}+\mathbf{AB}) = \det(\mathbf{I}+\mathbf{BA})$, in (b), $c \triangleq \det(\mathbf{R}_x)$ is a constant independent of the choice of the precoding matrix, and $\mathbf{R}_\epsilon \triangleq E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^H]$ denotes the error covariance matrix, in which $\boldsymbol{\epsilon} \triangleq \mathbf{x}-\hat{\mathbf{x}}$, $\hat{\mathbf{x}}$ is the linear minimum-mean-square-error estimate (LMMSE) of $\mathbf{x}$.

Our objective is to determine a set of precoding matrices $\{\mathbf{C}_n\}$, or equivalently an overall block-diagonal precoding matrix $\mathbf{C}$, such that the mutual information (MI) between the sources $\mathbf{x}$ and the received signal $\mathbf{y}$ is maximized under certain transmit power constraints. The problem can be formulated as

$$\begin{aligned}\max_{\{\mathbf{C}_n\}} \quad & \det(\mathbf{R}_\epsilon^{-1}) = \det(\mathbf{R}_x^{-1}+\mathbf{C}^H\mathbf{H}^H\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C})\\ \text{s.t.} \quad & \text{tr}(\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H) \le P_n \qquad \forall n\end{aligned} \tag{5}$$

where we impose individual power constraints on the sources. MI measures the amount of reduced uncertainty concerning $\mathbf{x}$ given the knowledge of $\mathbf{y}$. It determines the maximum rate at which the information can be reliably transmitted over the noisy channel. In our paper, the transmitted signals $\{\mathbf{x}_n\}$ are communication sources which can be coded already. Since sensors' measurements are spatially correlated, the coded symbols remain correlated in space[1]. The MI is meaningful design criterion in this case. As shown further below, the use of the MI metric allows for an analytical solution to a key problem in solving (5). It is noted that, while MI is often used for the design of coded communication systems, it has also been a useful tool in solving many signal estimation problems which involves analog/uncoded signals (e.g. [32], [33]). This is not surprising since MI represents the amount of reduction in uncertainty about $\mathbf{x}$ given $\mathbf{y}$, by maximizing $\mathcal{I}(\mathbf{x},\mathbf{y})$, we may better infer/estimate $\mathbf{x}$ from $\mathbf{y}$.

For analog/uncoded signals, one may also consider optimizing the trade-off between rate/power and accuracy in the rate-distortion (e.g. [14]–[16]) or power-distortion (e.g. [21], [24]) sense, using the mean square error (MSE) for the reconstruction error at the receiver. There is an intimate relation between the MI and the mean square error (MSE) criteria and, more broadly, between information theory and estimation theory, as discussed in [34], [35]. An interesting connection also exists for the problem studied here. Specifically, as suggested by the objective function in (5), maximizing the mutual information is equivalent to minimizing the determinant of the error covariance matrix of the LMMSE estimator, i.e.

$$\max \ \text{MI} \Leftrightarrow \min \ \det(\mathbf{R}_\epsilon) = \prod_{k=1}^{l}\lambda_k(\mathbf{R}_\epsilon)$$

where $\lambda_k(\mathbf{R}_\epsilon)$ denotes the kth eigenvalue of the error covariance matrix. On the other hand, minimizing the MSE between the transmitted and reconstructed signals is equivalent to minimizing the trace of the error covariance matrix

$$\min \ \text{MSE} \Leftrightarrow \min \ \text{trace}(\mathbf{R}_\epsilon) = \sum_{k=1}^{l}\lambda_k(\mathbf{R}_\epsilon)$$

We see that both criteria are functions of the eigenvalues of the error covariance matrix, with the function being the arithmetic mean for the MSE-based criterion and the geometric mean for the MI-based criterion.

Let us return to the problem (5). Note that in (5), $\mathbf{C}$, $\mathbf{H}$ and $\mathbf{R}_v$ are block-diagonal matrices. If the source signals are mutually independent, the overall signal covariance matrix $\mathbf{R}_x$ is block-diagonal as well. Hence the optimization (5) can be decoupled into a set of simple and identical optimization problems

$$\begin{aligned}\max_{\mathbf{C}_n} \quad & \det(\mathbf{R}_{x_n}^{-1}+\mathbf{C}_n^H\mathbf{H}_n^H\mathbf{R}_{v_n}^{-1}\mathbf{H}_n\mathbf{C}_n)\\ \text{s.t.} \quad & \text{tr}(\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H) \le P_n\end{aligned} \tag{6}$$

The above optimization (6) has been thoroughly studied in previous works, e.g. [19], with the closed-form solution given by a channel-diagonalizing structure and the classical capacity-achieving water-filling for the power allocation. Nevertheless, when the sources are correlated, the signal covariance matrix $\mathbf{R}_x$ does not have a block-diagonal structure, which hinders us from decomposing (5) into a number of tractable optimization problems. In sensor network scenarios where multiple sensors are deployed to monitor a same area or target, the data across sensors are usually correlated. It is therefore very important to examine precoding design for the general case where $\mathbf{R}_x$ is non-diagonal.

III. Precoding Design: A Gauss-Seidel Approach

For the correlated case, the optimization (5) is generally non-concave due to the block-diagonal structure imposed on the overall precoding matrix. An analytical solution of the optimization (5) is therefore difficult to obtain. Moreover, a joint search over these N precoding matrices is practically infeasible since it has a complexity that grows exponentially with N. To simplify the problem, we consider a Gauss-Seidel approach which successively optimizes the precoding matrix associated with each sensor, given the other precoding matrices fixed. Without loss of generality, we study the precoding design for the first sensor node, supposing that the remaining precoding matrices are pre-determined and satisfy their respective power constraints.

For convenience, define $\boldsymbol{\Omega}_n \triangleq \mathbf{C}_n^H\mathbf{H}_n^H\mathbf{R}_{v_n}^{-1}\mathbf{H}_n\mathbf{C}_n$. We have

$$\mathbf{C}^H\mathbf{H}^H\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C} = \text{diag}(\boldsymbol{\Omega}_1,\ldots,\boldsymbol{\Omega}_N) = \begin{bmatrix}\boldsymbol{\Omega}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}_{\bar{1}}\end{bmatrix} \tag{7}$$

[1]Coding and interleaving may have a temporal whitening effect but this kind of temporal processing does not remove the spatial correlation across sensors.

where $\mathbf{\Omega}_{\bar{1}} \triangleq \text{diag}(\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_N)$ is a matrix independent of the optimization variable $\mathbf{C}_1$. Also, let

$$\mathbf{Q} \triangleq \mathbf{R}_x^{-1} = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}^H & \mathbf{Q}_{22} \end{bmatrix} \tag{8}$$

be partitioned into four blocks, with $\mathbf{Q}_{11}$ and $\mathbf{Q}_{22}$ having the same dimensions as $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_{\bar{1}}$, respectively. The objective function (5) can be re-expressed as

$$\begin{aligned} &\det(\mathbf{R}_x^{-1} + \mathbf{C}^H\mathbf{H}^H\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C}) \\ &= \det\left\{\begin{bmatrix} \mathbf{Q}_{11}+\mathbf{\Omega}_1 & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}^H & \mathbf{Q}_{22}+\mathbf{\Omega}_{\bar{1}} \end{bmatrix}\right\} \\ &\overset{(a)}{=} \det(\mathbf{Q}_{22}+\mathbf{\Omega}_{\bar{1}}) \\ &\quad \times \det(\mathbf{Q}_{11}+\mathbf{\Omega}_1 - \mathbf{Q}_{12}(\mathbf{Q}_{22}+\mathbf{\Omega}_{\bar{1}})^{-1}\mathbf{Q}_{12}^H) \end{aligned} \tag{9}$$

where (a) comes from the block matrix determinant identity:

$$\det\left\{\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\right\} = \det(\mathbf{D})\det(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}) \tag{10}$$

where $\mathbf{D}$ is invertible. Resorting to (9), and noting that both $\mathbf{\Omega}_{\bar{1}}$ and $\mathbf{Q}$ are independent of $\mathbf{C}_1$, the precoding design for the first sensor, given the others fixed, can therefore be formulated as

$$\begin{aligned} \max_{\mathbf{C}_1} \quad & \det(\mathbf{G}_1+\mathbf{\Omega}_1) = \det(\mathbf{G}_1 + \mathbf{C}_1^H\mathbf{H}_1^H\mathbf{R}_{v_1}^{-1}\mathbf{H}_1\mathbf{C}_1) \\ \text{s.t.} \quad & \text{tr}(\mathbf{C}_1\mathbf{R}_{x_1}\mathbf{C}_1^H) \le P_1 \end{aligned} \tag{11}$$

in which $\mathbf{G}_1 \triangleq \mathbf{Q}_{11} - \mathbf{Q}_{12}(\mathbf{Q}_{22}+\mathbf{\Omega}_{\bar{1}})^{-1}\mathbf{Q}_{12}^H$. Note that $\mathbf{G}_1$ is the Schur complement of the block $(\mathbf{Q}_{22}+\mathbf{\Omega}_{\bar{1}})$ in $\mathbf{M} \triangleq [\mathbf{Q}_{11}\ \mathbf{Q}_{12}; \mathbf{Q}_{12}^H\ \mathbf{Q}_{22}+\mathbf{\Omega}_{\bar{1}}]$, where $\mathbf{M}$ is equivalent to the matrix $\mathbf{R}_x^{-1} + \mathbf{C}^H\mathbf{H}^H\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C}$ when $\mathbf{C}_1 = \mathbf{0}$. Since $\mathbf{R}_x^{-1} + \mathbf{C}^H\mathbf{H}^H\mathbf{R}_v^{-1}\mathbf{H}\mathbf{C}$ is always positive definite (whether $\mathbf{C}_1 = \mathbf{0}$ or not), the Schur complement $\mathbf{G}_1$ is also positive definite.

Similarly, the precoding design for the nth sensor with other precoding matrices fixed can be written as

$$\begin{aligned} \max_{\mathbf{C}_n} \quad & \det(\mathbf{G}_n+\mathbf{\Omega}_n) = \det(\mathbf{G}_n + \mathbf{C}_n^H\mathbf{H}_n^H\mathbf{R}_{v_n}^{-1}\mathbf{H}_n\mathbf{C}_n) \\ \text{s.t.} \quad & \text{tr}(\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H) \le P_n \end{aligned} \tag{12}$$

for some properly defined $\mathbf{G}_n$, and $\mathbf{G}_n$ is positive definite. At the first sight, there is a close resemblance between (6) and (12). Nevertheless, the optimization (12) is much more complicated than (6) because $\mathbf{G}_n$ is generally not equivalent to $\mathbf{R}_{x_n}^{-1}$. This inequivalence spoils the inherent symmetry which is the key to solving (6). To see this, we note that the optimization (6) can be readily reformulated as

$$\begin{aligned} \max_{\mathbf{C}_n} \quad & \det(\mathbf{I} + \mathbf{R}_{v_n}^{-\frac{1}{2}}\mathbf{H}_n\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H\mathbf{H}_n^H\mathbf{R}_{v_n}^{-\frac{1}{2}}) \\ \text{s.t.} \quad & \text{tr}(\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H) \le P_n \end{aligned} \tag{13}$$

in which the quadratic term, $\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H$, appears in both the objective function and the constraint, and therefore can be treated as a single optimization variable. It can be easily verified that the optimization (13) is concave with respect to the new variable. Let $\mathbf{R}_{v_n}^{-\frac{1}{2}}\mathbf{H}_n = \mathbf{U}_h\mathbf{D}_h\mathbf{V}_h^H$ denote the singular value decomposition (SVD), and let $\mathbf{\Psi} \triangleq \mathbf{V}_h^H\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H\mathbf{V}_h$. Using Hadamard's inequality [36, Theorem 7.8.1], it is easy to show that $\mathbf{\Psi} \triangleq \mathbf{V}_h^H\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H\mathbf{V}_h$ is a diagonal matrix and the optimal solution is obtained by the well-known water-filling algorithm.

From the above derivation, we see that the key step to solve (6) is to formulate the optimization into the amiable form (13) in which the quadratic term $\mathbf{C}_n\mathbf{R}_{x_n}\mathbf{C}_n^H$, instead of the precoding matrix itself, becomes the variable to be optimized. This significantly simplifies the optimization problem. However, for the optimization problem (12), since $\mathbf{G}_n$ is not equivalent to $\mathbf{R}_{x_n}^{-1}$, there is no way to convert (12) into an amiable form like (13). The optimization (12), therefore, is totally different from (6).

A. An Analytical Solution to the Optimization (12)

To gain an insight into (12), we first convert the problem to an amiable form easier for analysis. Let $\bar{\mathbf{C}}_n \triangleq \mathbf{C}_n\mathbf{R}_{x_n}^{\frac{1}{2}}$, and $\bar{\mathbf{G}}_n \triangleq \mathbf{R}_{x_n}^{\frac{1}{2}}\mathbf{G}_n\mathbf{R}_{x_n}^{\frac{1}{2}}$, the optimization (12) can be re-expressed as

$$\begin{aligned} \max_{\bar{\mathbf{C}}_n} \quad & \det(\bar{\mathbf{G}}_n + \bar{\mathbf{C}}_n^H\mathbf{H}_n^H\mathbf{R}_{v_n}^{-1}\mathbf{H}_n\bar{\mathbf{C}}_n) \\ \text{s.t.} \quad & \text{tr}(\bar{\mathbf{C}}_n\bar{\mathbf{C}}_n^H) \le P_n \end{aligned} \tag{14}$$

To further simplify the problem, we carry out the SVD: $\bar{\mathbf{C}}_n = \mathbf{U}_c\mathbf{D}_c\mathbf{V}_c^H$ and the eigenvalue decomposition (EVD): $\mathbf{T}_n \triangleq \mathbf{H}_n^H\mathbf{R}_{v_n}^{-1}\mathbf{H}_n = \mathbf{U}_t\mathbf{D}_t\mathbf{U}_t^H$, and $\bar{\mathbf{G}}_n = \mathbf{U}_g\mathbf{D}_g\mathbf{U}_g^H$, where $\mathbf{U}_c$, $\mathbf{V}_c$, $\mathbf{U}_t$, and $\mathbf{U}_g$ are $p_n \times p_n$ unitary matrices, $\mathbf{D}_c$, $\mathbf{D}_t$ and $\mathbf{D}_g$ are diagonal matrices given respectively as

$$\begin{aligned} \mathbf{D}_c &\triangleq \text{diag}(d_{c,1}, d_{c,2}, \ldots, d_{c,p_n}) \\ \mathbf{D}_t &\triangleq \text{diag}(\lambda_1(\mathbf{T}_n), \lambda_2(\mathbf{T}_n), \ldots, \lambda_{p_n}(\mathbf{T}_n)) \\ \mathbf{D}_g &\triangleq \text{diag}(\lambda_1(\bar{\mathbf{G}}_n), \lambda_2(\bar{\mathbf{G}}_n), \ldots, \lambda_{p_n}(\bar{\mathbf{G}}_n)) \end{aligned} \tag{15}$$

in which $\lambda_k(\mathbf{T}_n)$ and $\lambda_k(\bar{\mathbf{G}}_n)$ denote the k-th eigenvalue associated with $\mathbf{T}_n$ and $\bar{\mathbf{G}}_n$, respectively. Without loss of generality, we assume that the diagonal elements of $\mathbf{D}_c$, $\mathbf{D}_t$ and $\mathbf{D}_g$ are arranged in descending order. We have

$$\begin{aligned} &\det(\bar{\mathbf{G}}_n + \bar{\mathbf{C}}_n^H\mathbf{H}_n^H\mathbf{R}_{v_n}^{-1}\mathbf{H}_n\bar{\mathbf{C}}_n) \\ &= \det(\mathbf{V}_c^H\bar{\mathbf{G}}_n\mathbf{V}_c + \mathbf{D}_c^H\mathbf{U}_c^H\mathbf{T}_n\mathbf{U}_c\mathbf{D}_c) \\ &\triangleq \det(\bar{\mathbf{V}}_c^H\mathbf{D}_g\bar{\mathbf{V}}_c + \mathbf{D}_c^H\bar{\mathbf{U}}_c^H\mathbf{D}_t\bar{\mathbf{U}}_c\mathbf{D}_c) \end{aligned} \tag{16}$$

where $\bar{\mathbf{V}}_c \triangleq \mathbf{U}_g^H\mathbf{V}_c$, and $\bar{\mathbf{U}}_c \triangleq \mathbf{U}_t^H\mathbf{U}_c$ are unitary matrices. Resorting to (16), the optimization (12) (or (14)) can be transformed into a new optimization that searches for an optimal set $\{\bar{\mathbf{U}}_c, \mathbf{D}_c, \bar{\mathbf{V}}_c\}$, in which $\bar{\mathbf{U}}_c$ and $\bar{\mathbf{V}}_c$ are also unitary matrices

$$\begin{aligned} \max_{\{\bar{\mathbf{U}}_c, \mathbf{D}_c, \bar{\mathbf{V}}_c\}} \quad & \det(\bar{\mathbf{V}}_c^H\mathbf{D}_g\bar{\mathbf{V}}_c + \mathbf{D}_c^H\bar{\mathbf{U}}_c^H\mathbf{D}_t\bar{\mathbf{U}}_c\mathbf{D}_c) \\ \text{s.t.} \quad & \text{tr}(\mathbf{D}_c\mathbf{D}_c^H) \le P_n \\ & \bar{\mathbf{U}}_c\bar{\mathbf{U}}_c^H = \mathbf{I}, \qquad \bar{\mathbf{V}}_c\bar{\mathbf{V}}_c^H = \mathbf{I} \end{aligned} \tag{17}$$

The optimization involves searching for multiple optimization variables. To solve (17), we firstly find the optimal $\{\bar{\mathbf{U}}_c, \bar{\mathbf{V}}_c\}$ given that $\mathbf{D}_c$ is fixed. Then substituting the derived optimal unitary matrices into (17), we determine the optimal diagonal matrix $\mathbf{D}_c$. Optimizing $\{\bar{\mathbf{U}}_c, \bar{\mathbf{V}}_c\}$ conditional on a given $\mathbf{D}_c$ can be formulated as

$$\begin{aligned} \max_{\{\bar{\mathbf{U}}_c, \bar{\mathbf{V}}_c\}} \quad & \det(\bar{\mathbf{V}}_c^H\mathbf{D}_g\bar{\mathbf{V}}_c + \mathbf{D}_c^H\bar{\mathbf{U}}_c^H\mathbf{D}_t\bar{\mathbf{U}}_c\mathbf{D}_c) \\ \text{s.t.} \quad & \bar{\mathbf{U}}_c\bar{\mathbf{U}}_c^H = \mathbf{I}, \qquad \bar{\mathbf{V}}_c\bar{\mathbf{V}}_c^H = \mathbf{I} \end{aligned} \tag{18}$$

To gain an insight into (18), we first introduce the following new determinant inequality.

Lemma 1: Suppose $\mathbf{A} \in \mathbb{C}^{N\times N}$ and $\mathbf{B} \in \mathbb{C}^{N\times N}$ are positive semi-definite matrices with eigenvalues $\{\lambda_k(\mathbf{A})\}$ and $\{\lambda_k(\mathbf{B})\}$ arranged in descending order, $\mathbf{D} \in \mathbb{R}^{N\times N}$ is a diagonal matrix with non-negative diagonal elements $\{d_k\}$ arranged in descending order. Then the following determinant inequality holds

$$\det\left(\mathbf{D}^H\mathbf{A}\mathbf{D}+\mathbf{B}\right) \le \prod_{k=1}^{N}(d_k^2\lambda_k(\mathbf{A})+\lambda_{N+1-k}(\mathbf{B})) \quad (19)$$

The above inequality becomes an equality if $\mathbf{A}$ and $\mathbf{B}$ are diagonal, and the diagonal elements of $\mathbf{A}$ and $\mathbf{B}$ are sorted in descending order and ascending order, respectively, i.e. $\mathbf{A} = \text{diag}(\lambda_1(\mathbf{A}),\ldots,\lambda_N(\mathbf{A}))$, and $\mathbf{B} = \text{diag}(\lambda_N(\mathbf{B}),\ldots,\lambda_1(\mathbf{B}))$.

Proof: See Appendix A. ∎

Letting $\mathbf{A} \triangleq \bar{\mathbf{U}}_c^H\mathbf{D}_t\bar{\mathbf{U}}_c$, $\mathbf{B} \triangleq \bar{\mathbf{V}}_c^H\mathbf{D}_g\bar{\mathbf{V}}_c$, and utilizing Lemma 1, we know that for any feasible solution of (18), its objective function value is upper bounded by

$$\begin{aligned}&\det(\bar{\mathbf{V}}_c^H\mathbf{D}_g\bar{\mathbf{V}}_c+\mathbf{D}_c^H\bar{\mathbf{U}}_c^H\mathbf{D}_t\bar{\mathbf{U}}_c\mathbf{D}_c)\\ &\le\prod_{k=1}^{p_n}\left(d_{c,k}^2\lambda_k(\mathbf{T}_n)+\lambda_{p_n+1-k}(\bar{\mathbf{G}}_n)\right)\end{aligned} \quad (20)$$

where the upper bound is attained when $\bar{\mathbf{U}}_c=\mathbf{I}$ and $\bar{\mathbf{V}}_c=\mathbf{J}$, $\mathbf{J}$ is an anti-identity matrix which has ones along the anti-diagonal and zeros elsewhere. Hence $\{\bar{\mathbf{U}}_c=\mathbf{I},\bar{\mathbf{V}}_c=\mathbf{J}\}$ is an optimal solution to (18). Note that the optimal solution to (18) may not be unique, i.e. other choices of unitary matrices $\bar{\mathbf{U}}_c$ and $\bar{\mathbf{V}}_c$ may also achieve the upper bound (20). Nevertheless, this does not affect the optimality of our solution. Notice that the upper bound (20) is independent of the specific choice of $\{\bar{\mathbf{U}}_c,\bar{\mathbf{V}}_c\}$. Hence whatever the optimal solution to (18) is, by substituting the optimal $\{\bar{\mathbf{U}}_c,\bar{\mathbf{V}}_c\}$ back into (17), we reach the following optimization which searches for the optimal diagonal matrix $\mathbf{D}_c$:

$$\begin{aligned}\max_{\{d_{c,k}\}}\quad&\prod_{k=1}^{p_n}\left(d_{c,k}^2\lambda_k(\mathbf{T}_n)+\lambda_{p_n+1-k}(\bar{\mathbf{G}}_n)\right)\\ \text{s.t.}\quad&\sum_{k=1}^{p_n}d_{c,k}^2\le P_n,\quad d_{c,k}\ge 0\quad\forall k\end{aligned} \quad (21)$$

The above optimization (21) can be solved analytically by resorting to the Lagrangian function and KKT conditions, whose details are elaborated in Appendix D. Briefly speaking, for a threshold ϕ that is uniquely determined by a procedure described in Appendix D, we have

$$d_{c,k}=\begin{cases}\sqrt{\frac{1}{\phi}-\frac{\lambda_{p_n+1-k}(\bar{\mathbf{G}}_n)}{\lambda_k(\mathbf{T}_n)}} & \text{if } \phi<\frac{\lambda_k(\mathbf{T}_n)}{\lambda_{p_n+1-k}(\mathbf{G}_n)}\\ 0 & \text{otherwise}\end{cases} \quad (22)$$

Clearly, the global maximum of (21) is also the maximum achievable objective function value of (17). Hence any set of $\{\bar{\mathbf{U}}_c,\mathbf{D}_c,\bar{\mathbf{V}}_c\}$ that can attain this global maximum is an optimal solution to (17). Based on the above discussions, the optimal solution to (17) can therefore be summarized as follows.

Lemma 2: With $\mathbf{D}_c$, $\mathbf{D}_g$, and $\mathbf{D}_t$ defined in (15), the optimal solution to (17) is given by

$$\bar{\mathbf{U}}_c^*=\mathbf{I},\qquad \bar{\mathbf{V}}_c^*=\mathbf{J},\qquad \mathbf{D}_c^*=\text{diag}(\{d_{c,k}^*\}) \quad (23)$$

where $d_{c,k}^*$ is given in (22).

Given Lemma 2, we can readily derive the optimal solution to (12), whose results are summarized as follows.

Theorem 1: The optimal solution to (12) is given by

$$\mathbf{C}_n^*=\mathbf{U}_t\mathbf{D}_c^*\mathbf{J}\mathbf{U}_g^H\mathbf{R}_{x_n}^{-\frac{1}{2}} \quad (24)$$

where $\mathbf{U}_t$ and $\mathbf{U}_g$ are unitary matrices obtained from the EVD of $\mathbf{T}_n \triangleq \mathbf{H}_n^H\mathbf{R}_{v_n}^{-1}\mathbf{H}_n$ and $\bar{\mathbf{G}}_n \triangleq \mathbf{R}_{x_n}^{\frac{1}{2}}\mathbf{G}_n\mathbf{R}_{x_n}^{\frac{1}{2}}$, respectively; $\mathbf{D}_c$ is a diagonal matrix with its diagonal elements $\{d_{c,k}\}$ given by (22).

Proof: Clearly, we have $\mathbf{C}_n^*=\bar{\mathbf{C}}_n^*\mathbf{R}_{x_n}^{-\frac{1}{2}}$, in which $\bar{\mathbf{C}}_n^*=\mathbf{U}_c^*\mathbf{D}_c^*(\mathbf{V}_c^*)^H$, $\mathbf{U}_c^*=\mathbf{U}_t\bar{\mathbf{U}}_c^*$, and $\mathbf{V}_c^*=\mathbf{U}_g\bar{\mathbf{V}}_c^*$. Recalling results in Lemma 2, we arrive at (24). ∎

B. Summary of the Gauss-Seidel Approach

By using the above results, we can establish an iterative algorithm to solve (5) by successively optimizing and replacing each precoding matrix $\mathbf{C}_n$. The algorithm is described as follows.

1) Randomly generate a set of precoding matrices $\{\mathbf{C}_n^{(0)}\}$ as an initialization.
2) At iteration $i+1$ ($i=0,1,\ldots$): determine $\mathbf{C}_1^{(i+1)}$ from (24) given: $\{\mathbf{C}_2^{(i)},\ldots,\mathbf{C}_N^{(i)}\}$; determine $\mathbf{C}_k^{(i+1)}$ from (24) given: $\{\mathbf{C}_1^{(i+1)},\ldots,\mathbf{C}_{k-1}^{(i+1)},\mathbf{C}_{k+1}^{(i)},\ldots,\mathbf{C}_N^{(i)}\}$ for $k=2,\ldots,N$.
3) Go to Step 2 if $|f(\{\mathbf{C}_n^{(i+1)}\})-f(\{\mathbf{C}_n^{(i)}\})|>\epsilon$, where $f(\cdot)$ denotes the objective function defined in (5), ϵ is a prescribed tolerance value; otherwise stop.

Remark 1: Clearly, in this algorithm, every iteration results in a non-decreasing objective function value. In this manner, the iterative algorithm converges to a stationary point and finds an effective set of precoding matrices. Due to the complexity of the objective function (5) for block-diagonal matrix $\mathbf{C}$, we are unable to provide a theoretical analysis concerning the convergence of $\{\mathbf{C}_n^{(t)}\}$ to the global optimal solution. Nevertheless, we found from our simulations that, starting from different (randomly generated) initial points, the iteration ends with the same precoding matrix almost every time, which suggests that the proposed algorithm is able to achieve the globally optimal solution.

Remark 2: We discuss the implementation issue of the proposed algorithm. In cellular systems, precoder is usually computed at the transmitter side using the CSI provided by the receiver (e.g. base station) via a feedback channel. For our proposed algorithm, however, a global CSI and the correlation parameters of all sources are required to compute the precoding matrix associated with each sensor. In addition, due to the nature of the Gauss-Seidel approach, each precoding matrix needs to be computed along with other sensors' precoding matrices, which increases the computational complexity at the transmitter. To overcome these difficulties, we can let a central processing unit (located at the receiver) undertake the computationally intensive precoder design task, supposing the computational complexity for the central processing unit

is not a major issue. In doing this way, the computational burden on sensors can be relieved. Also, by letting each sensor periodically send training data to the destination, the receiver can obtain all required channels for joint precoder design. The computed precoding matrices are then assigned to sensors. Note that the knowledge of the autocorrelation matrix of the correlated sources may come either from specific data models or from sample estimation after a training phase.

Remark 3: In our paper, perfect global CSI is assumed available (at the receiver) for procoding design. In practice, CSI is typically unknown and needs to be estimated, and the estimate inevitably suffers from estimation errors. Imperfect CSI will certainly lead to deteriorated performance. It is, however, difficult to theoretically analyze the impact of the channel estimation error on the performance of the proposed algorithm due to the iterative nature of the Gauss-Seidel approach. Nevertheless, our simulation results demonstrate that the proposed algorithm is effective in the presence of channel estimation errors and achieves higher ergodic capacity than other schemes even when the estimate is coarse.

C. Extension to Non-Square Precoding Matrices

For simplicity of mathematical exposition, linear precoders considered in previous subsections are assumed to be square. In this subsection, we discuss the extension of our proposed algorithm to the non-square case. In the following, we assume $\mathbf{C}_n \in \mathbb{C}^{p_n \times r}$ with $p_n < r$. The case of $p_n > r$ follows a similar argument and thus omitted here.

It can be readily verified that the problem of precoder design for the nth sensor with other precoding matrices fixed can still be formulated as (12), except that $\mathbf{G}_n$ and $\mathbf{R}_{x_n}$ are $r \times r$ matrices, instead of $p_n \times p_n$. Define $\bar{\mathbf{C}}_n \triangleq \mathbf{C}_n \mathbf{R}_{x_n}^{\frac{1}{2}}$, $\bar{\mathbf{G}}_n \triangleq \mathbf{R}_{x_n}^{\frac{1}{2}} \mathbf{G}_n \mathbf{R}_{x_n}^{\frac{1}{2}}$, we can rewrite (12) as (14). Similarly, we conduct the EVD: $\mathbf{T}_n \triangleq \mathbf{H}_n^H \mathbf{R}_{v_n}^{-1} \mathbf{H}_n = \mathbf{U}_t \mathbf{D}_t \mathbf{U}_t^H$, $\bar{\mathbf{G}}_n = \mathbf{U}_g \mathbf{D}_g \mathbf{U}_g^H$, and the reduced SVD of $\bar{\mathbf{C}}_n$: $\bar{\mathbf{C}}_n = \mathbf{U}_c \mathbf{D}_c \mathbf{V}_c^H$, where $\mathbf{U}_c \in \mathbb{C}^{p_n \times p_n}$, $\mathbf{V}_c \in \mathbb{C}^{r \times p_n}$, and

$$\begin{aligned} \mathbf{D}_c &\triangleq \text{diag}(d_{c,1}, d_{c,2}, \ldots, d_{c,p_n}) \\ \mathbf{D}_t &\triangleq \text{diag}(\lambda_1(\mathbf{T}_n), \lambda_2(\mathbf{T}_n), \ldots, \lambda_{p_n}(\mathbf{T}_n)) \\ \mathbf{D}_g &\triangleq \text{diag}(\lambda_1(\bar{\mathbf{G}}_n), \lambda_2(\bar{\mathbf{G}}_n), \ldots, \lambda_r(\bar{\mathbf{G}}_n)) \end{aligned} \tag{25}$$

Note that $\mathbf{V}_c$ obtained from the reduced SVD of $\bar{\mathbf{C}}_n$ is a non-square matrix consisting of p_n orthomornal columns. Again, we assume that the diagonal elements of $\mathbf{D}_c$, $\mathbf{D}_t$, and $\mathbf{D}_g$ are arranged in descending order. The optimization problem (14) can be re-expressed as

$$\begin{aligned} \max_{\{\bar{\mathbf{U}}_c, \mathbf{D}_c, \bar{\mathbf{V}}_c\}} \quad & \det(\bar{\mathbf{V}}_c^H \mathbf{D}_g \bar{\mathbf{V}}_c + \mathbf{D}_c^H \bar{\mathbf{U}}_c^H \mathbf{D}_t \bar{\mathbf{U}}_c \mathbf{D}_c) \\ \text{s.t.} \quad & \text{tr}(\mathbf{D}_c \mathbf{D}_c^H) \le P_n \\ & \bar{\mathbf{U}}_c \bar{\mathbf{U}}_c^H = \mathbf{I}, \qquad \bar{\mathbf{V}}_c^H \bar{\mathbf{V}}_c = \mathbf{I} \end{aligned} \tag{26}$$

in which $\bar{\mathbf{V}}_c \triangleq \mathbf{U}_g^H \mathbf{V}_c$, and $\bar{\mathbf{U}}_c \triangleq \mathbf{U}_t^H \mathbf{U}_c$. This optimization problem is similar to (17) except that $\bar{\mathbf{V}}_c \in \mathbb{C}^{r \times p_n}$ is no longer a unitary matrix, but a sub-matrix of a unitary matrix consisting of p_n orthonormal columns. Given $\mathbf{D}_c$ fixed, searching for $\{\bar{\mathbf{U}}_c, \bar{\mathbf{V}}_c\}$ yields the following optimization

$$\begin{aligned} \max_{\{\bar{\mathbf{U}}_c, \bar{\mathbf{V}}_c\}} \quad & \det(\bar{\mathbf{V}}_c^H \mathbf{D}_g \bar{\mathbf{V}}_c + \mathbf{D}_c^H \bar{\mathbf{U}}_c^H \mathbf{D}_t \bar{\mathbf{U}}_c \mathbf{D}_c) \\ \text{s.t.} \quad & \bar{\mathbf{U}}_c \bar{\mathbf{U}}_c^H = \mathbf{I}, \qquad \bar{\mathbf{V}}_c^H \bar{\mathbf{V}}_c = \mathbf{I} \end{aligned} \tag{27}$$

It can be proved that (the proof can be found in Appendix C) for any feasible solution of (27), its objective function value is upper bounded by

$$\begin{aligned} &\det(\bar{\mathbf{V}}_c^H \mathbf{D}_g \bar{\mathbf{V}}_c + \mathbf{D}_c^H \bar{\mathbf{U}}_c^H \mathbf{D}_t \bar{\mathbf{U}}_c \mathbf{D}_c) \\ &\le \prod_{k=1}^{p_n} \left(d_{c,k}^2 \lambda_k(\mathbf{T}_n) + \lambda_{p_n+1-k}(\bar{\mathbf{G}}_n) \right) \end{aligned} \tag{28}$$

which has exactly the same form as its counterpart for the square case (20). The upper bound is attained when $\bar{\mathbf{U}}_c = \mathbf{I}$ and $\bar{\mathbf{V}}_c = [\mathbf{J}\ \mathbf{0}]^T$, where $\mathbf{J}$ is the anti-diagonal matrix of dimension $p_n \times p_n$. Therefore the optimal solution to the optimization problem (27) is given by

$$\bar{\mathbf{U}}_c^* = \mathbf{I}, \qquad \bar{\mathbf{V}}_c^* = [\mathbf{J}\ \mathbf{0}]^T \tag{29}$$

Substituting the optimal $\bar{\mathbf{U}}_c$ and $\bar{\mathbf{V}}_c$ back into (26), we can solve $\mathbf{D}_c$ as we did previously. Clearly, the above result presents a generalization of the results for the square precoder case. When $p_n = r$, the optimal solution (29) reduces to the conventional form (23).

IV. Simulation Results

We present numerical results to illustrate the performance of the proposed Gauss-Seidel iterative algorithm. Two different scenarios are respectively considered.

A. Scenarios of Monitoring a Stochastic Process

The first example involves transmitting noisy observations of a common random process from multiple distributed sensors to the destination. Suppose we have $N = 5$ sensors monitoring a physical process. The process is modeled as an autoregressive (AR)(1) stochastic process which is given as

$$\theta(k) = \beta\theta(k-1) + u(k) \qquad k = 2, \ldots, K \tag{30}$$

in which β is the AR coefficient; $\{u(k)\}$ is a white Gaussian noise process with zero mean and variance $\sigma_u^2 = 1$ for $k = 1$ and $\sigma_u^2 = 1 - \beta^2$ for $k > 1$, which ensures that the process $\{\theta(k)\}$ is stationary with zero mean and variance $E[\theta^2(k)] = 1$. The observed measurements at each sensor, say sensor n, are given by

$$x_n(k) = \theta(k) + w_n(k) \tag{31}$$

where $\{w_n(k)\}$ denotes the white Gaussian observation noise with zero mean and variance σ_w^2. We assume that each sensor and the receiver are equipped with 3 antennas, i.e. $q = 3$, and $p_n = 3$ for all n. At each sensor, the observation sequence is organized into a number of consecutive data blocks with the length of each block equal to p_n. In other words, the lth data block is defined as $\mathbf{x}_n(l) \triangleq [x_n(lp_n)\ x_n(lp_n - 1)\ \ldots\ x_n(lp_n - p_n + 1)]^T$. The data in each block are respectively emitted from the p_n transmit antennas. Clearly, for this scenario, $\mathbf{R}_x$ is a non block-diagonal matrix with its diagonal blocks given by $\mathbf{R}_{x_n} = E[\mathbf{x}_n(l)\mathbf{x}_n(l)^H] = \mathbf{R}_\theta + \sigma_w^2 \mathbf{I}$, and its off-diagonal blocks equal to $\mathbf{R}_\theta$, where $\mathbf{R}_\theta = E[\boldsymbol{\theta}\boldsymbol{\theta}^H]$ can be easily determined from the AR model (30), and $\boldsymbol{\theta} \triangleq [\theta(lp_n)\ \theta(lp_n - 1)\ \ldots\ \theta(lp_n - p + 1)]^T$.

In our simulations, the entries of the channel matrix $\mathbf{H}_n$ are randomly generated from a complex Gaussian distribution with zero-mean, independent real and imaginary parts, each

with variance $1/2$. Such a choice models a Rayleigh fading scenario in which the receiving/transmitting antennas are sufficiently spatially separated [31]. The covariance matrix of the channel noise is set to be $\mathbf{R}_{v_n} = 0.1\mathbf{I}$ for all n. Also, the transmit power is assumed the same for all sensors, i.e. $P_n = P, \forall n$. We first provide numerical results to show the convergence behavior of the proposed iterative algorithm. The ergodic capacity vs. the number of iterations is plotted in Fig. 2, where we set $\sigma_w^2 = 0.1$, $P = 1$ and $\beta = 0.8$. Results are averaged over 10^4 independent runs, with the channel matrices $\{\mathbf{H}_n\}$ randomly generated for each run. From Fig. 2, we see that the algorithm has a fast convergence rate, and usually provides a reasonable convergence performance within only a few iterations. Also, our simulations show that for a fixed channel realization, the proposed algorithm, when starting from different initial points, converges to a same solution almost every time. This suggests that the proposed algorithm achieves the globally optimal solution.

To illustrate the effectiveness of the proposed algorithm, we compare the proposed algorithm with two other precoding schemes, namely, a "correlation-ignored precoding" which designs the precoder without considering the spatial correlation across different sensors, i.e. the precoder associated with each sensor is given by the optimal solution to (6); and a "no precoding" scheme in which no precoding is used, i.e. $\mathbf{C}_n = \alpha_n\mathbf{I}, \forall n$, α_n is a scaling factor chosen to satisfy the prescribed power constraint. Fig. 3 depicts the ergodic capacities of the three schemes vs. the transmit power P, where we set $\beta = 0.8$ and $\sigma_w^2 = 0.1$. It can be seen from Fig. 3 that our proposed algorithm presents a clear performance advantage over the other two schemes. The performance gain is primarily due to the fact that our algorithm exploits spatial correlation in joint precoder design, whereas the correlation-ignored scheme, ignoring the correlation across sensors, designs the precoder for each sensor individually. Next, we examine the behavior of three schemes when the spatial correlation across sensors varies. Note that the spatial correlation is dependent on the observation noise variance σ_w^2: a smaller σ_w^2 leads to higher correlation across sensors and vice versa. Fig. 4 plots the ergodic capacities of three schemes as a function of σ_w^2 for $\beta = 0.5$ and $\beta = 0.8$, respectively. It can be observed that the performance of all three schemes improves with an incremental σ_w^2. This is because when σ_w^2 becomes larger, the average signal power $E[x_n^2(k)]$ increases correspondingly. Meanwhile, we notice that the performance of the other two schemes has the tendency to approach that of the proposed algorithm as σ_w^2 increases. The reason is that sensors are less correlated for a larger σ_w^2. In particular, the correlation coefficient between $x_n(k)$ and $x_m(k)$ reduces to $1/3$ when $\sigma_w^2 = 2$. In this case the correlation-ignored scheme is closer to be an optimal strategy. Thus the performance benefit gained by the proposed algorithm decreases accordingly.

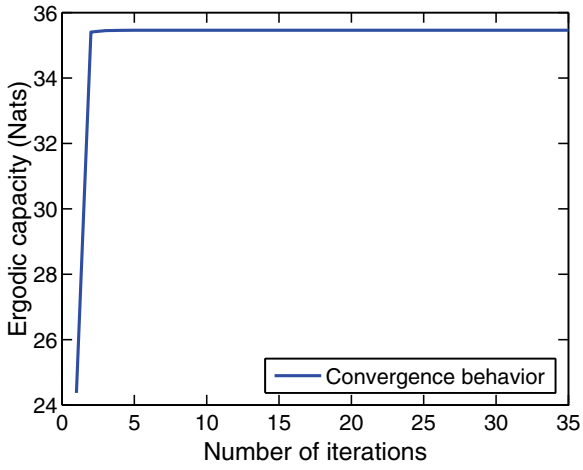


Fig. 2. Convergence behavior of the proposed iterative algorithm.

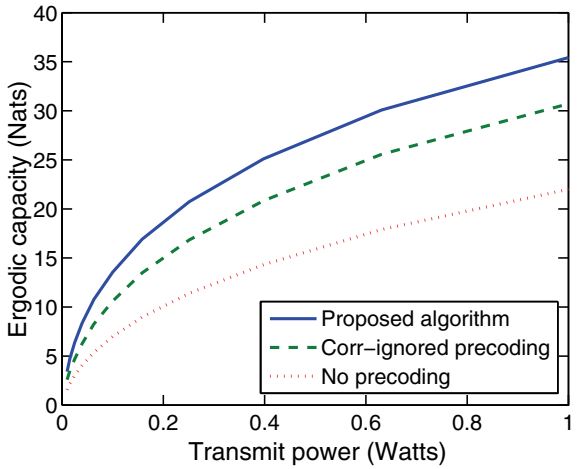


Fig. 3. Ergodic capacity of the three respective algorithms.

B. Distance-Dependent Correlation Model

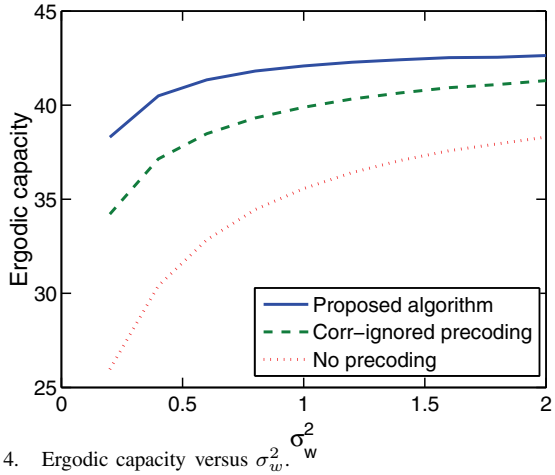
In this subsection, we use a distance-dependent function to model the spatial correlation among sensors. Suppose we have $N = 10$ sensors which are placed uniformly at random on a two-dimensional unit area. The symbols are assumed spatially correlated but temporally uncorrelated, i.e.

$$\text{cov}(x_m(k), x_n(t)) = \begin{cases} 0 & k \neq t, \forall\{m,n\} \\ e^{-d_{m,n}} & k = t, \forall\{m,n\} \end{cases} \tag{32}$$

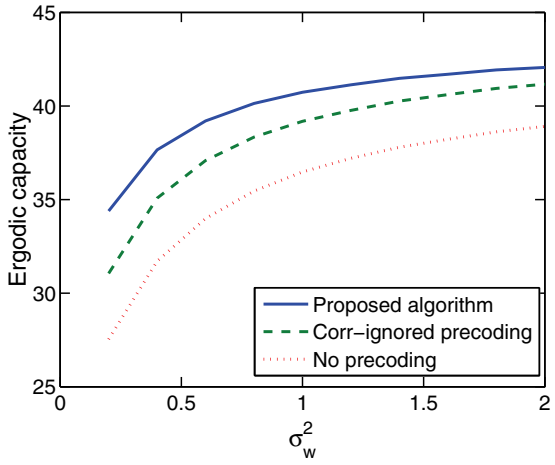
where $d_{m,n}$ denotes the Euclidean distance between nodes m and n. Such a distance-dependent correlation model has been widely used to characterize spatially distributed phenomena such as temperature, wind speed, and concentration of some chemical material. We assume that each sensor and the receiver are equipped with 3 antennas. The symbol vector $\mathbf{x}_n$ is defined in a similar way as in the previous example. It can be readily verified that the cross-covariance matrix between $\mathbf{x}_m$ and $\mathbf{x}_n$ is given by

$$\mathbf{R}_{x_{m,n}} = \begin{cases} \mathbf{I} & \text{if } m = n \\ e^{-d_{m,n}}\mathbf{I} & \text{otherwise} \end{cases} \tag{33}$$

In the simulations, the channel matrices $\{\mathbf{H}_n\}$ are generated in the same way as those in the previous example. The covariance matrix of the channel noise is set to be $\mathbf{R}_{v_n} = 0.1\mathbf{I}$. Fig. 5 plots the ergodic capacities of the three schemes vs. the transmit power. We can see that our proposed algorithm which exploits the spatial correlation across sensors outperforms the



(a) $\beta = 0.8$



(b) $\beta = 0.5$

Fig. 4. Ergodic capacity versus σ_w^2.

correlation-ignored precoding scheme by a significant margin. Also, we notice that for this example, correlation-ignored precoding scheme performs similarly (if not worse) to no precoding. This example indicates that the correlation-ignored precoding scheme does not necessarily perform better than the no-precoding scheme. In fact, when sources are correlated, both correlation-ignored and no-precoding schemes are sub-optimal solutions, and the correlation-ignored scheme could render a solution that is far away from optimality due to an unfavorable beamforming and power allocation strategy (The strategy is optimized without taking spatial correlation into account, and thus could even lead to an adverse effect on the spectral/energy efficiency of the network).

Previous simulations assume that CSI is perfectly known. In practice, the channel state has to be estimated and thus inevitably suffers from estimation errors. In the following, we examine the impact of channel estimation errors on the performance of respective schemes. We model imperfect CSI through a Gauss-Markov uncertainty of the form [37], [38]

$$\mathbf{H}_n = \sqrt{1-\rho^2}\hat{\mathbf{H}}_n + \rho\mathbf{E}_n \quad \forall n \tag{34}$$

where $\mathbf{H}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the true channel matrix, $\hat{\mathbf{H}}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ denotes the estimated channel, $\mathbf{E}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the estimation error. The parameter ρ characterizes the estimation accuracy since $\rho = 0$ corresponds to perfect channel knowledge and $\rho = 1$ corresponds to no CSI knowledge. Fig. 6 depicts the ergodic capacity versus the parameter ρ for three schemes. Results are averaged over 10^4 independent runs, with the channel matrices randomly generated for each run. We see that the performance of the proposed algorithm gradually degrades with an increasing estimation error. In contrast, the other two schemes incur little performance loss when the channel estimate deteriorates. The reason can be explained as follows. For the no precoding scheme, since the precoder, $\mathbf{C}_n = \alpha_n \mathbf{I}, \forall n$, is independent of the CSI, the channel estimation error does not affect its performance. For the correlation-ignored scheme, even with perfect CSI, the precoding matrices are far away from optimality due to its neglect of the spatial correlation across sensors. Therefore its performance is much less dependent on the channel estimation quality. We notice that although the proposed algorithm suffers from a certain performance degradation in the presence of estimation errors, the proposed algorithm still presents a clear performance advantage over the other two schemes even when there is a significant mismatch between the true and estimated channel, say, $\rho = 0.8$.

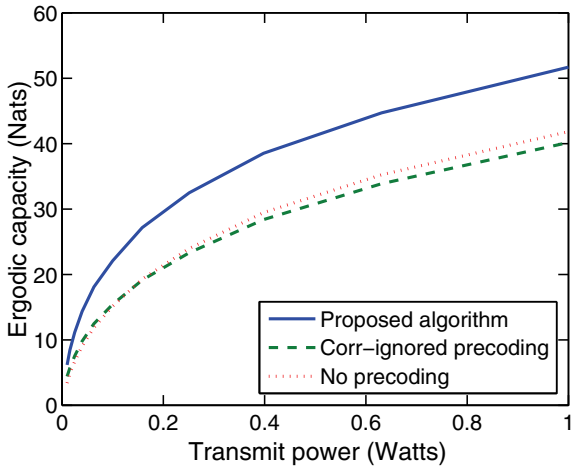


Fig. 5. Ergodic capacity of the three respective algorithms.

V. Conclusions

This paper studied precoding design for multiple spatially distributed but correlated sources. These multiple sources are transmitted to a common destination via orthogonal channels. Assuming a perfect instantaneous channel state information, our objective is to maximize the overall channel capacity between these correlated sources and the received signals. A Gauss-Seidel iterative algorithm was proposed, in which the precoding matrix associated with each sensor is successively optimized, given other precoding matrices fixed. Our simulations indicate that the proposed iterative algorithm has a fast convergence rate, and starting from different initial points, the iteration ends with the same precoding matrix, which suggests that the algorithm is able to achieve the globally optimal solution (although a theoretical analysis of global convergence is not available). In addition, numerical results show that the proposed algorithm which takes into account the spatial correlations can achieve considerably higher capacity

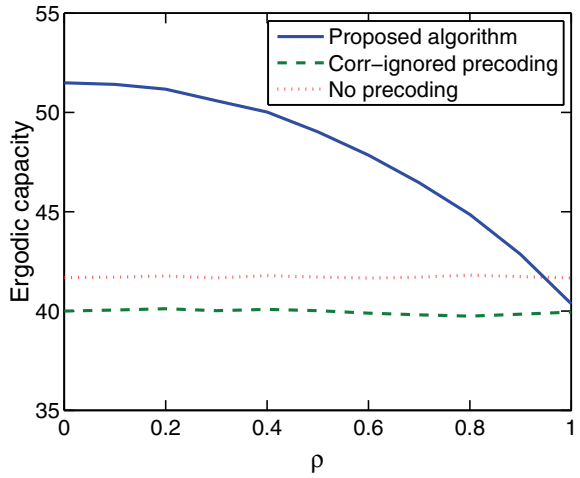


Fig. 6. Imperfect CSI: Ergodic capacity of the three schemes versus the parameter ρ.

than methods that neglect the spatial correlations and design linear precoders for each sensor individually.

Appendix A
Proof of Lemma 1

Define $\mathbf{\Gamma} \triangleq \mathbf{D}^H\mathbf{A}\mathbf{D}$, and its eigenvalues $\{\lambda_k(\mathbf{\Gamma})\}$ are arranged in descending order. Then we have

$$\det\left(\mathbf{D}^H\mathbf{A}\mathbf{D}+\mathbf{B}\right) \le \prod_{k=1}^{N}(\lambda_k(\mathbf{\Gamma})+\lambda_{N+1-k}(\mathbf{B})) \tag{35}$$

The above inequality comes from the following well-known matrix inequality [36]:

$$\begin{aligned}\prod_{k=1}^{N}(\lambda_k(\mathbf{X})+\lambda_k(\mathbf{Y})) \le& \det(\mathbf{X}+\mathbf{Y})\\ &\le \prod_{k=1}^{N}(\lambda_k(\mathbf{X})+\lambda_{N+1-k}(\mathbf{Y}))\end{aligned} \tag{36}$$

in which $\mathbf{X}$ and $\mathbf{Y}$ are positive semidefinite Hermitian matrices, with eigenvalues $\{\lambda_k(\mathbf{X})\}$ and $\{\lambda_k(\mathbf{Y})\}$ arranged in descending order respectively.

To prove (19), we only need to show that the term on the right-hand side of (35) is upper bounded by

$$\prod_{k=1}^{N}(\lambda_k(\mathbf{\Gamma})+\lambda_{N+1-k}(\mathbf{B})) \le \prod_{k=1}^{N}(d_k^2\lambda_k(\mathbf{A})+\lambda_{N+1-k}(\mathbf{B})) \tag{37}$$

Before proceeding to prove (37), we introduce the following inequalities for the two sequences $\{\lambda_k(\mathbf{\Gamma})\}_{k=1}^N$ and $\{d_k^2\lambda_k(\mathbf{A})\}_{k=1}^N$.

$$\begin{aligned}\prod_{k=1}^{K}\lambda_k(\mathbf{\Gamma}) &\le \prod_{k=1}^{K} d_k^2\lambda_k(\mathbf{A}) \quad 1\le K<N\\ \prod_{k=1}^{N}\lambda_k(\mathbf{\Gamma}) &= \prod_{k=1}^{N} d_k^2\lambda_k(\mathbf{A})\end{aligned} \tag{38}$$

The proof of the inequalities (38) is provided in Appendix B. The inequality relations between these two sequences can be characterized by the notion of "multiplicative majorization" (also termed log-majorization). Multiplicative majorization is a notion parallel to the concept of additive majorization. For two vectors $\mathbf{a}\in\mathbb{R}_+^N$ and $\mathbf{b}\in\mathbb{R}_+^N$ with elements sorted in descending order ($\mathbb{R}_+$ stands for the set of non-negative real numbers), we say that $\mathbf{a}$ is multiplicatively majorized by $\mathbf{b}$, denoted by $\mathbf{a}\prec_\times\mathbf{b}$, if

$$\begin{aligned}\prod_{k=1}^{K}a_k &\le \prod_{k=1}^{K} b_k \quad 1\le K<N\\ \prod_{k=1}^{N}a_k &= \prod_{k=1}^{N} b_k\end{aligned} \tag{39}$$

Here we use the symbol $\prec_\times$ to differentiate the multiplicative majorization from the conventional additive majorization $\prec$. Another important concept that is closely related to majorization is schur-convex or schur-concave functions. A function $f:\mathbb{R}^N\to\mathbb{R}$ is said to be multiplicatively schur-convex if for $\mathbf{a}\prec_\times\mathbf{b}$, then $f(\mathbf{a})\le f(\mathbf{b})$. Clearly, establishing (37) is equivalent to showing the function

$$f(\mathbf{a}) \triangleq \prod_{k=1}^{N}(a_k+c_{N+1-k}) \tag{40}$$

is multiplicatively schur-convex for elements $\mathbf{c}=[c_k]\in\mathbb{R}_+^N$ arranged in descending order. This multiplicatively schur-convex property can also be summarized as follows.

Lemma 3: For vectors $\mathbf{a}\in\mathbb{R}_+^N$, $\mathbf{b}\in\mathbb{R}_+^N$, and $\mathbf{c}\in\mathbb{R}_+^N$, with their elements arranged in descending order, if $\mathbf{a}\prec_\times\mathbf{b}$, then we have $f(\mathbf{a})\le f(\mathbf{b})$, i.e.

$$\prod_{k=1}^{N}(a_k+c_{N+1-k}) \le \prod_{k=1}^{N}(b_k+c_{N+1-k}) \tag{41}$$

Proof: We prove Lemma 3 by induction. For $N=2$, we have

$$\begin{aligned}f(\mathbf{a})-f(\mathbf{b}) =& [a_1+c_2][a_2+c_1]-[b_1+c_2][b_2+c_1]\\ \overset{(a)}{=}& [a_1-b_1]c_1+[a_2-b_2]c_2\\ \overset{(b)}{\le}& c_2[a_1+a_2-b_1-b_2] \overset{(c)}{\le} 0\end{aligned} \tag{42}$$

where (a) can be easily derived by noting that $a_1a_2=b_1b_2$; (b) comes from the fact that $a_1-b_1\le 0$ and $c_1\ge c_2$; (c) is a result of the following inequality: $a_1+a_2\le b_1+b_2$, that is, for any two non-negative elements, if their product remains constant, then their sum increases as the two elements are further apart.

Now suppose that for M-dimensional vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, the inequality (41) holds true. We show that (41) is also valid for $(M+1)$-dimensional vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. From the inequalities (39), we know that $b_1\ge a_1$. For the special case where $b_1=a_1$, it is easy to verify that the truncated vector $\mathbf{a}_t\triangleq[a_2\ \ldots\ a_{M+1}]$ is multiplicatively majorized by the truncated vector $\mathbf{b}_t\triangleq[b_2\ \ldots\ b_{M+1}]$, i.e. $\mathbf{a}_t\prec_\times\mathbf{b}_t$. Therefore we have

$$f(\mathbf{a}_t)\le f(\mathbf{b}_t) \tag{43}$$

and consequently we arrive at $f(\mathbf{a})\le f(\mathbf{b})$ given $b_1=a_1$.

Now consider the general case where $b_1 > a_1$. There must be at least one index such that $b_l < a_l$ since the overall products of the two sequences $\{a_k\}_{k=1}^{M+1}$ and $\{b_k\}_{k=1}^{M+1}$ are identical[2]. Without loss of generality, let l_1 denote the smallest index for which $b_l < a_l$. We adopt a pairwise transformation to convert the sequence $\{b_l\}_{k=1}^{M+1}$ into a new sequence $\{\beta_k\}_{k=1}^{M+1}$. Specifically, the first and the l_1th entries of $\{b_l\}_{k=1}^{M+1}$ are updated as

$$\begin{cases} \beta_1 = a_1, \beta_{l_1} = \frac{b_1 b_{l_1}}{a_1} & \text{if } b_1 b_{l_1} \le a_1 a_{l_1} \\ \beta_1 = \frac{b_1 b_{l_1}}{a_{l_1}}, \beta_{l_1} = a_{l_1} & \text{if } b_1 b_{l_1} > a_1 a_{l_1} \end{cases} \tag{44}$$

whereas other entries remain unaltered, i.e. $\beta_k = b_k$, $\forall k \neq 1, l_1$. Clearly, the entries β_1 and β_{l_1} satisfy

$$\begin{aligned} \beta_1 \beta_{l_1} &= b_1 b_{l_1} \\ \beta_1 &\le b_1 \end{aligned} \tag{45}$$

That is, $[\beta_1 \ \beta_{l_1}]^T \prec_\times [b_1 \ b_{l_1}]^T$. By following the same argument of (42) and noting that $\beta_k = b_k$, $\forall k \neq 1, l_1$, we have

$$f(\boldsymbol{\beta}) \le f(\mathbf{b}) \tag{46}$$

where $\boldsymbol{\beta} \triangleq [\beta_1 \ \ldots \ \beta_{M+1}]$. Our objective now is to show

$$f(\mathbf{a}) \le f(\boldsymbol{\beta}) \tag{47}$$

It can be easily verified that $\mathbf{a}$ is multiplicatively majorized by $\boldsymbol{\beta}$, i.e. $\mathbf{a} \prec_\times \boldsymbol{\beta}$, by noting $\beta_l \ge a_l$ for any $l < l_1$ and $\beta_1 \beta_{l_1} = b_1 b_{l_1}$.

Now we proceed to prove (47). Consider two different cases in (44).

- If $b_1 b_{l_1} \le a_1 a_{l_1}$, then $\beta_1 = a_1$. In this case, it is easy to verify that the truncated vector $\mathbf{a}_t \triangleq [a_2 \ \ldots \ a_{M+1}]$ is multiplicatively majorized by the truncated vector $\boldsymbol{\beta}_t \triangleq [\beta_2 \ \ldots \ \beta_{M+1}]$, i.e. $\mathbf{a}_t \prec_\times \boldsymbol{\beta}_t$. Therefore we have
$$f(\mathbf{a}_t) \le f(\boldsymbol{\beta}_t) \tag{48}$$
and consequently $f(\mathbf{a}) \le f(\boldsymbol{\beta})$ as we have $\beta_1 = a_1$.
- For the second case where $b_1 b_{l_1} > a_1 a_{l_1}$, we have $\beta_{l_1} = a_{l_1}$. Define two new vectors $\mathbf{a}_p \triangleq [a_1 \ \ldots \ a_{l_1-1} \ a_{l_1+1} \ \ldots, a_{M+1}]$ and $\boldsymbol{\beta}_p \triangleq [\beta_1 \ \ldots \ \beta_{l_1-1} \ \beta_{l_1+1} \ \ldots \ \beta_{M+1}]$. From $\mathbf{a} \prec_\times \boldsymbol{\beta}$, we can readily verify that $\mathbf{a}_p$ is multiplicatively majorized by $\boldsymbol{\beta}_p$, i.e. $\mathbf{a}_p \prec_\times \boldsymbol{\beta}_p$. Therefore we have
$$f(\mathbf{a}_p) \le f(\boldsymbol{\beta}_p) \tag{49}$$
and consequently $f(\mathbf{a}) \le f(\boldsymbol{\beta})$ as we have $\beta_{l_1} = a_{l_1}$.

Combining (46)–(47), we arrive at (41). The proof is completed here. ∎

APPENDIX B
PROOF OF (38)

Recall the following theorem [39, Chapter 9: Theorem H.1]

Theorem: If $\mathbf{X}$ and $\mathbf{Y}$ are $N \times N$ complex matrices, then

$$\begin{aligned} &\prod_{k=1}^{K} \sigma_k(\mathbf{XY}) \le \prod_{k=1}^{K} \sigma_k(\mathbf{X})\sigma_k(\mathbf{Y}), \qquad K = 1, \ldots, N-1 \\ &\prod_{k=1}^{N} \sigma_k(\mathbf{XY}) = \prod_{k=1}^{N} \sigma_k(\mathbf{X})\sigma_k(\mathbf{Y}) \end{aligned} \tag{50}$$

where $\{\sigma_i(\cdot)\}$ are singular values arranged in a descending order.

By utilizing the above results, we have

$$\begin{aligned} \prod_{k=1}^{K} \lambda_k(\boldsymbol{\Gamma}) &= \prod_{k=1}^{K} \sigma_k(\boldsymbol{\Gamma}) \le \prod_{k=1}^{K} \sigma_k(\mathbf{D}^H\mathbf{A})\sigma_k(\mathbf{D}) \\ &= \Big(\prod_{k=1}^{K} \sigma_k(\mathbf{D}^H\mathbf{A})\Big)\Big(\prod_{k=1}^{K} d_k\Big) \\ &\le \Big(\prod_{k=1}^{K} \sigma_k(\mathbf{D}^H)\sigma_k(\mathbf{A})\Big)\Big(\prod_{k=1}^{K} d_k\Big) \\ &= \prod_{k=1}^{K} d_k^2 \lambda_k(\mathbf{A}), \qquad K = 1, \ldots, N-1 \end{aligned} \tag{51}$$

The proof is completed here.

APPENDIX C
PROOF OF (28)

To prove that the objective function value (27) is upper bounded by (28), we first introduce the following lemma.

Lemma 4: Suppose that $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{B} \in \mathbb{C}^{m\times m}$ are positive semi-definite matrices with eigenvalues $\{\lambda_k(\mathbf{A})\}$ and $\{\lambda_k(\mathbf{B})\}$ arranged in descending order, $\mathbf{V} \in \mathbb{C}^{n\times m}$ is a tall matrix ($n > m$) consisting of orthonormal columns, i.e. $\mathbf{V}^H\mathbf{V} = \mathbf{I}$. Then the following inequality holds

$$\det(\mathbf{V}^H\mathbf{AV} + \mathbf{B}) \le \prod_{k=1}^{m} (\lambda_{m+1-k}(\mathbf{A}) + \lambda_k(\mathbf{B})) \tag{52}$$

Proof: From [36, Corollary 4.3.16], we know that the kth largest eigenvalue of the matrix $\mathbf{V}^H\mathbf{AV}$ is no greater than its corresponding eigenvalue of the matrix $\mathbf{A}$, i.e.

$$\lambda_k(\mathbf{V}^H\mathbf{AV}) \le \lambda_k(\mathbf{A}) \qquad k = 1, \ldots, m \tag{53}$$

Utilizing this result, we can immediately reach that

$$\begin{aligned} \det(\mathbf{V}^H\mathbf{AV} + \mathbf{B}) &\overset{(a)}{\le} \prod_{k=1}^{m} (\lambda_{m+1-k}(\mathbf{V}^H\mathbf{AV}) + \lambda_k(\mathbf{B})) \\ &\le \prod_{k=1}^{m} (\lambda_{m+1-k}(\mathbf{A}) + \lambda_k(\mathbf{B})) \end{aligned} \tag{54}$$

where (a) comes from the determinant inequality (36). The proof is completed here. ∎

Letting $\mathbf{Z} \triangleq \mathbf{D}_c^H \bar{\mathbf{U}}_c^H \mathbf{D}_t \bar{\mathbf{U}}_c \mathbf{D}_c$, we then have

$$\begin{aligned} \det(\bar{\mathbf{V}}_c^H \mathbf{D}_g \bar{\mathbf{V}}_c + \mathbf{Z}) &\overset{(a)}{\le} \prod_{k=1}^{p_n} (\lambda_{p_n+1-k}(\bar{\mathbf{G}}_n) + \lambda_k(\mathbf{Z})) \\ &\overset{(b)}{\le} \prod_{k=1}^{p_n} \left(d_{c,k}^2 \lambda_k(\mathbf{T}_n) + \lambda_{p_n+1-k}(\bar{\mathbf{G}}_n)\right) \end{aligned} \tag{55}$$

[2]When $\mathbf{a}$ and $\mathbf{b}$ contain zero elements, the overall product is zero. In this case, we may not find an index such that $b_l < a_l$. Nevertheless, since we have $b_k \ge a_k$ for all k, proof of (41) is evident.

where (a) follows from Lemma 4, (b) comes from the inequality (37) already proved in Appendix A (note that $\mathbf{Z}$ corresponds to $\mathbf{\Gamma}$ in (37)). This completes our proof.

APPENDIX D
ANALYTICAL SOLUTION TO THE OPTIMIZATION (21)

For notational simplicity, we let $a_k \triangleq \lambda_k(\mathbf{T}_n)$ and $b_k \triangleq \lambda_{p_n+1-k}(\bar{\mathbf{G}}_n)$, (21) is equivalent to solving

$$\begin{aligned} \min_{\{x_k\}} \quad & -\sum_{k=1}^{p_n} \ln(x_k a_k + b_k) \\ \text{s.t.} \quad & \sum_{k=1}^{p_n} x_k = P_n \\ & x_k \geq 0 \end{aligned} \tag{56}$$

The Lagrangian function associated with (56) is given by

$$\begin{aligned} & L(x_k;\phi;\nu_k) \\ &= -\sum_{k=1}^{p_n} \ln(x_k a_k + b_k) - \phi\left(P_n - \sum_{k=1}^{p_n} x_k\right) - \sum_{k=1}^{p_n} \nu_k x_k \end{aligned} \tag{57}$$

which gives the following KKT conditions:

$$\begin{aligned} -\frac{a_k}{x_k a_k + b_k} + \phi - \nu_k &= 0 \quad \forall k \\ P_n - \sum_{k=1}^{p_n} x_k &= 0 \\ \nu_k x_k &= 0 \quad \forall k \\ \nu_k &\geq 0 \quad \forall k \\ x_k &\geq 0 \quad \forall k \end{aligned} \tag{58}$$

By solving the first equation of the above KKT conditions, we obtain

$$x_k = \frac{1}{\phi - \nu_k} - \frac{b_k}{a_k} \quad \forall k \tag{59}$$

Also, the last three KKT conditions imply that we have either $\{\nu_k = 0, x_k > 0\}$ or $\{\nu_k > 0, x_k = 0\}$. Therefore (59) becomes

$$x_k = \left[\frac{1}{\phi} - \frac{b_k}{a_k}\right]^+ \quad \forall k \tag{60}$$

where $[x]^+$ is equal to x if $x > 0$, otherwise it is zero. The Lagrangian multiplier ϕ and the number of nonzero elements ($x_k > 0$) can be uniquely determined from the second equation of the KKT conditions. The procedure is described as follows.

Suppose we have $K \in \{1, \ldots, p_n\}$ nonzero elements, i.e. $x_k > 0, \forall k = 1, \ldots, K$ (note that $\{x_k\}$ are in descending order since $\{b_k\}$ are arranged in ascending order and $\{a_k\}$ are arranged in descending order). Therefore ϕ can be solved by substituting $\{x_1, x_2, \ldots, x_K\}$ into the second KKT condition:

$$\phi = \frac{K}{P_n + \sum_{k=1}^{K}(b_k/a_k)} \tag{61}$$

Now substituting ϕ back to (60), we get a new solution $\{x'_1, x'_2, \ldots, x'_K, x'_{K+1}, \ldots, x'_{p_n}\}$. If for this new solution, we have $x_k = 0$ for $k > K$. Then it is the true solution we are looking for; otherwise we have to choose another K to repeat the above procedure.

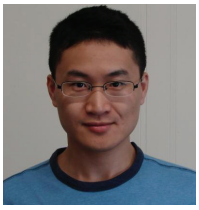
ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their valuable comments and suggestions that have helped improve the quality of our work.

REFERENCES

[1] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 17–29, Mar. 2002.
[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, pp. 102–114, Aug. 2002.
[3] B. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," in *Proc. 2002 Int. Conf. Distributed Computing Syst. Workshops*.
[4] A. Deligiannakis, Y. Kotidis, and N. Roussopoulos, "Compressing historical information in sensor networks," in *Proc. 2004 ACM SIGMOD Conf.*
[5] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2617–2624, Aug. 2005.
[6] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors—part I: fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
[7] R. S. Blum, S. A.Kassam, and H. V. Poor, "Distributed detection with multiple sensors—part II: advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.
[8] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—part I: Gaussian PDF," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.
[9] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: from delta modulation to maximum likelihood," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5246–5257, Oct. 2008.
[10] K. Zhang, X. R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion—part VI: sensor data compression," in *Proc. 2003 Int. Conf. Inf. Fusion*.
[11] 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, Aug. 2007.
[12] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.
[13] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 773–787, Nov. 1987.
[14] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *Proc. 2003 IEEE Data Compression Conf.*, pp. 13–22.
[15] G. Maierbacher and J. Barros, "Low-complexity coding and source-optimized clustering for large-scale sensor networks," *ACM Trans. Sensor Netw.*, vol. 5, no. 3, article 24, May 2009.
[16] K. Viswanatha, S. Ramaswamy, A. Saxena, and K. Rose, "Error-resilient and complexity-constrained distributed coding for large scale sensor networks," in *Proc. 2012 ACM/IEEE Conf. Inf. Process. Sensor Netw.*
[17] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.
[18] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers—part I: unification and optimal designs," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1988–2006, July 1999.
[19] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
[20] Y. Jiang, J. Li, and W. W. Hager, "Joint transceivers design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.
[21] D. P. Palomar and Y. Jiang, *MIMO Transceiver Design via Majorization Theory*. Now Publishers Inc., 2007.
[22] H. R. Bahrami and T. Le-Ngoc, "Precoder design based on correlation matrices for MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3579–3587, Dec. 2006.
[23] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.
[24] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 214–226, Jan. 2004.

[25] S. Järmyr, B. Ottersten, and E. Jorswieck, "Statistical precoding and detection ordering in MIMO multiple-access channels with decision feedback equalization," in *Proc. 2011 IEEE Int. Conf. Commun.*
[26] O. Munoz, J. Vidal, and A. Agustin, "Non-regenerative MIMO relaying with channel state information," in *Proc. 2005 IEEE Int. Conf. Acoust., Speech, Signal Process.*
[27] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.
[28] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.
[29] N. Fawaz, K. Zarifi, M. Debbah, and D. Gesbert, "Asymptotic capacity and optimal precoding in MIMO multi-hop relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2050–2069, Apr. 2011.
[30] R. Chandra, C. Fetzer, and K. Hogsedt, "A mesh based robust topology discovery algorithm for hybrid wireless networks," *Informatics*, Sep. 2002.
[31] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.
[32] Y. Yang and R. S. Blum, "MIMO radar waveform design based on mutual information and minimum mean-square error estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 330–343, Jan. 2007.
[33] M. Wax and T. Kailath, "Detection of signals by information theoretic criterion," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, pp. 387–392, Apr. 1985.
[34] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[35] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.
[36] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
[37] C. Wang, E. K. S. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 805–810, Mar. 2007.
[38] L. Musavian, M. R. Nakhai, M. Dohler, and A. H. Aghvami, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2798–2806, Sep. 2007.
[39] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. Academic Press, 1979.



Jun Fang (M'08) received the B.Sc. and M.Sc. degrees in electrical engineering from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree in electrical engineering from the National University of Singapore, Singapore, in 2006.

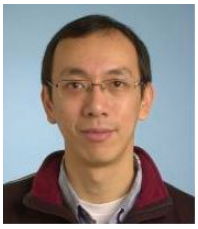
During 2006, he was with the Department of Electrical and Computer Engineering, Duke University, as a postdoctoral research associate. From January 2007 to December 2010, he was a research associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Since January 2011, he has been with the National Key Laboratory on Communications, University of Electronic Science and Technology of China. His research interests include statistical signal processing, distributed signal processing for wireless sensor networks, and wireless communications. Dr. Fang has authored/coauthored more than 20 papers in leading IEEE journals. He received the Outstanding Paper Award at the 2011 IEEE Africon Conference (co-authored with P. Wang, H. Li, and N. Han). He is currently an Associate Technical Editor for *IEEE Communications Magazine*.

Hongbin Li (M'99-SM'08) received B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant in the Department of Electrical and Computer Engineering at the University of Florida. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory during the summers of 2003, 2004, and 2009. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he is a Professor. His current research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for excellence in research in 2001 from Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He is presently a member of the Signal Processing Theory and Methods (SPTM) Technical Committee and served on the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society. He has been an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor for the *EURASIP Journal on Applied Signal Processing*, and a General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, Hoboken, NJ, held June 17–20, 2012.

Zhi Chen received B. Eng, M. Eng., and Ph.D. degrees in electrical engineering from the University of Electronic Science and Technology of China (UESTC) in 1997, 2000, 2006, respectively. In April 2006, he joined the National Key Laboratory on Communications (NCL), UESTC, where he is an Associate Professor. He was a visiting scholar at the University of California, Riverside, during 2010–2011. His current research interests include relay and cooperative communications, multi-user beamforming in cellular networks, interference coordination and cancellation, and THz communication. Dr. Chen has served as a reviewer for various international journals and conferences, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, et al.



Yu Gong has been with the School of Electronic, Electrical and Systems Engineering, Loughborough University, UK, since July 2012. Dr. Gong obtained his B.Eng. and M.Eng. in electronic engineering in 1992 and 1995, respectively, both at the University of Electronics and Science Technology of China. In 2002, he received his Ph.D. in communications from the National University of Singapore. After his Ph.D. graduation, he took several research positions in the Institute for Infocomm Research in Singapore and Queen's University of Belfast in the UK, respectively. From 2006 to 2012, Dr. Gong was an academic member in the School of Systems Engineering, University of Reading, UK. His research interests are in the areas of signal processing and communications, including wireless communications, cooperative networks, non-linear and non-stationary system identification, and adaptive filters.