IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 8, 2019, accepted July 12, 2019, date of publication July 18, 2019, date of current version August 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929584

Optimal Pilot Design for MIMO Broadcasting Systems Based on the Positive Definite Matrix Manifold

WEN ZHOU[1], XUTAO LI[2], JUNJUAN XIA[3], WANGUO JIAO[1], AND MIN HUA[1]
[1]College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China
[2]Electronic Engineering Department, Shantou University, Shantou 515063, China
[3]School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510006, China

Corresponding author: Junjuan Xia (xiajunjuan@gzhu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61601275, Grant 61471229, Grant 61701241, Grant 61801225, and Grant 61871139, in part by the Innovation Team Project of Guangdong Province University under Grant 2016KCXTD017, in part by the Science and Technology Program of Guangzhou under Grant 201807010103, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170935, in part by the Jiangsu Province Science and Technology Research under Grant 17KJB510030, and in part by the Highly Educated Research Start-Up Fund of Nanjing Forestry University under Grant GXL2016030.

ABSTRACT In the MIMO broadcasting system, channel state information (CSI) is often used for data detection at the receiver or preprocessing techniques such as the power control and user scheduling at the transmitter and hence, the study of its acquisition is important. Usually, the CSI can be obtained by the training-based channel estimation. Assuming that the users have the channels' statistical information and adopt the minimum mean-square-error (MMSE) receivers, this paper studies how to design an optimal training sequence to minimize the total weighted mean square error (MSE) of channel estimation. Following this objective, we formulate this problem as a positive semidefinite problem (SDP) with two constraints, and further transform it into a dual problem. In the solving process of the dual problem, we model it as an optimization problem on the positive definite matrix manifold and solve it by use of the iterative geodesic equation on the matrix manifold. The convergence and correctness of the proposed method are demonstrated by computer simulation; the effects of key system parameters on the weighted MSE are also investigated. Under various system configurations, the superiority of the proposed method is shown by comparing with a few other existing schemes.

INDEX TERMS Pilot, MIMO broadcasting, positive definite matrix manifold.

I. INTRODUCTION

Multi-input multi-output (MIMO) broadcasting systems model a network in which there is one transmitter and multiple receivers/users, both of which are equipped with multiple antennas, and the transmitter sends coded data that usually consist of all users' desired information to the users for further decoding. Since such a system has wide applications in communications, it has been always attracting attention [1]–[15].

References [1]–[3] are pioneering research articles on MIMO broadcasting systems. In [1], [2], the authors established the duality between the MIMO broadcast channel (BC) and MIMO multiple access channel, and pointed out that the achievable region of the ''dirty paper'' achieves the sum-rate capacity of the MIMO BC. Jindal *et al.* proposed an iterative water-filling algorithm to maximize the sum-rate of the MIMO BC, subject to the users' power constraints [3]. Later, researchers studied MIMO broadcasting systems from various aspects and also exploited a few novel applications or systems of this type. In [7], assuming the channel state information (CSI) is not ideal, the authors studied the user scheduling problem. In [8], the authors considered the problem of the energy efficiency optimization in the MIMO BC and proposed a novel optimization framework. There are also researches that introduced novel techniques or applications into the MIMO BC, such as the cognitive ratio [9], terrestrial broadcast TV [11], energy harvesting [12], decode-forward relay [13], etc.

Most of the above researches are conducted with the aid of the CSI, including the user scheduling algorithm, robust energy harvesting maximization, and so on. Usually, the

The associate editor coordinating the review of this manuscript and approving it for publication was Prabhat Kumar Upadhyay.

acquisition of the CSI is completed by channel estimation at the receiver, which plays an important role in communication systems. According to the design criterion, there exist a few classical channel estimation methods, including the least square (LS) estimator [16], minimum mean square error (MMSE) estimator [16], [17], linear MMSE (LMMSE) estimator [18], maximum likelihood (ML) estimator, etc. In [16], the authors investigated channel estimation in MIMO systems including the conventional LS and MMSE algorithms and they also proposed the scaled LS (SLS) and relaxed MMSE methods with less channel statistics information. According to whether there exist training sequences or not, channel estimation is roughly divided into three categories: blind estimation [22], semi-blind estimation [21], and training-based estimation [17]–[20]. In [22], the authors proposed a novel blind channel estimation method by use of the 4th cumulants for MIMO space–time block coded systems. Notice that as the training-based estimation method is more popular compared with the other two kinds of methods, this article mainly focuses on the training-based one.

Although most researches in MIMO broadcasting systems require the CSI, few articles study the acquisition of it. In fact, there exists some researches addressing the CSI; however, their focuses lie in how the incomplete CSI affects the system performance [23] or how to design robust precoders to combat the CSI error. On the other front, we also note that a lot of researches about channel estimation have been done in point-to-point MIMO or other systems [17]–[20]; however, they cannot be applied to the MIMO BC directly. Based on the above observations, this article considers the training-based channel estimation for MIMO broadcasting systems.

Specifically, the aim of this article is to design an optimal pilot/training sequence to minimize the total weighted MSE of channel estimation, assuming that all users adopt the MMSE channel estimators. Consequently, we formulate this problem as a positive semidefinite problem (SDP) with two constraints including the positive semidefinite constraint and the power constraint of the pilot covariance matrix. The contributions of this article are briefly summarized as follows.

1) We propose a novel method to design the optimal pilot sequence which minimizes the weighted channel estimation MSE. The convergence of the proposed method is demonstrated and the correctness is also verified via another method based on convex optimization softwares. A few existing works are evaluated and analyzed, including the isotropic transmission, the minimum variance unbiased (MVU) estimator, etc. Under various system settings, we compare the system performance of the proposed methods with that of these existing methods.
2) In the pilot sequence design, we incorporate the theory of information geometry. More concretely, we transform the original SDP into a dual problem at first and solve the dual problem by use of the iterative geodesic equation on the positive definite matrix manifold. It is worth noting that in the search of minimum of the SDP's *Lagrangian* function, we prove that, the positive semidefinite matrix set, on which the matrix variables of this function are defined, can be replaced by the positive definite matrix set. In this way, the geodesic equation on the positive definite manifold can be applied to solving the SDP.
3) We provide another proof on the convexity of the SDP's objective function.

The rest of the article is organized as follows. Section 2 illustrates the MIMO broadcasting system model. Section 3 formulates the problem and proposes the pilot design algorithm. Simulation results are presented in Section 4, with concluding remarks in Section 5.

Notations. Vectors are denoted by boldface lowercase letters and matrices are denoted by boldface uppercase letters. The notation $E(\cdot)$ represents the statistical expectation; $\mathbb{C}^{m\times n}$ stands for the sets of $m \times n$ complex matrices; $\otimes$ is the Kronecker product. We write $\mathbb{CN}(\boldsymbol{\mu}, \mathbf{R})$ to denote a complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$; for a matrix $\mathbf{X}$, the notations $\mathbf{X}^{1/2}$, $\mathrm{Tr}(\mathbf{X})$, $\mathbf{X}^H$, and $\mathbf{X}*$ and denote its square root, trace, Hermitian transpose, and conjugate, respectively; $\mathrm{vec}(\mathbf{X})$ is a column vector created by stacking the columns of $\mathbf{X}$;$\mathbf{X}\succ\mathbf{0}$ and $\mathbf{X}\succ\mathbf{0}$ mean that $\mathbf{X}$ is positive semidefinite and positive definite, respectively; besides, $\mathbf{I}_m$ is an $m \times m$ identity matrix.

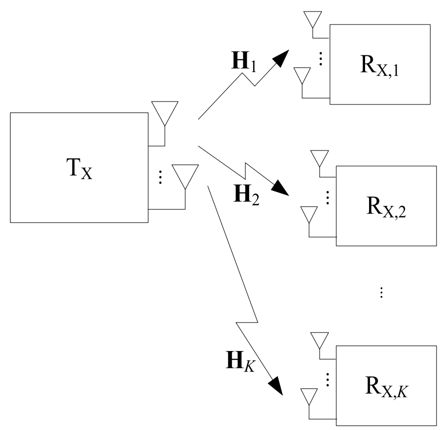


FIGURE 1. An MIMO broadcasting system model.

II. SYSTEM MODEL

As depicted in Fig. 1, consider an MIMO broadcasting system with one transmitter T_X and K receivers/users $\{\mathrm{R}_{\mathrm{X},i}\}$, in which Tx and $\{\mathrm{R}_{\mathrm{X},i}\}$ have N_T and $\{N_{R,i}\}$ antennas, respectively. The system input-output relationship can be modeled as

$$\mathbf{y}_i(n) = \mathbf{H}_i\mathbf{x}(n) + \mathbf{n}_i(n), \quad i = 1, \cdots, K \tag{1}$$

where $\mathbf{x}(n) \in \mathbb{C}^{N_T\times 1} = \sum_{i=1}^{K}\mathbf{x}_i(n)$ is the composite data sent from the transmitter at the n-th time slot, $\mathrm{tr}\left[\mathbf{x}(n)(\mathbf{x}(n))^H\right] \leq P_T$, $\mathbf{x}_i(n) \in \mathbb{C}^{N_T\times 1}$ is the data for user i, P_T is the maximum transmission power, and $\mathbf{y}_i(n) \in \mathbb{C}^{N_R\times 1}$ is the received signal for user i; $\mathbf{H}_i \in \mathbb{C}^{N_{R,i}\times N_T}$ is the flat fading MIMO channel from the transmitter to user i; $\mathbf{n}_i \in \mathbb{C}^{N_{R,i}\times 1}$ is the $i.i.d$ complex Gaussian noise vectors that follows $\mathbb{CN}\left(0, \sigma_i^2 \mathbf{I}_{N_{R,i}}\right)$.

The channel matrix $\mathbf{H}_i$ can be further modeled as

$$\mathbf{H}_i = \boldsymbol{\theta}_{R,i}^{1/2} \mathbf{H}_{\omega,i} \boldsymbol{\theta}_{T,i}^{1/2} \tag{2}$$

in which the positive definite matrices $\boldsymbol{\theta}_{R,i}$ and $\boldsymbol{\theta}_{T,i}$ are channel covariance matrices at the receiver and the transmitter end, respectively; $\mathbf{H}_{\omega,i}$ is a random matrix having *i.i.d.* complex Gaussian random variables with zero mean and unit variance, standing for the uncorrelated scattering. Note that the channel is assumed to be block fading and hence is time-invariant within the time period of interest.

Before the user data transmission, a pilot sequence will be sent from the transmitter in order to perform channel estimation at the receivers. Denote the pilot sequence as $\mathbf{P} \in \mathbb{C}^{N_T \times D}$ and it also satisfies the transmission power constraint $\mathrm{tr}\left(\mathbf{P}^H \mathbf{P}\right) = P_T$, where D is the number of time slots and P_T is the total power for pilot transmission. Stacking the user received signal set $\{\mathbf{y}_i(n), i = 1, \ldots, K, n = 1, \ldots, D\}$ during the pilot transmission, we have

$$\mathbf{Y}_i \triangleq \left[\, \mathbf{y}_i(1)\ \mathbf{y}_i(2)\ \ldots\ \mathbf{y}_i(D) \,\right] = \mathbf{H}_i \mathbf{P} + \mathbf{N}_i \tag{3}$$

where $\mathbf{N}_i = \left[\, \mathbf{n}_i(1)\ \mathbf{n}_i(2)\ \ldots\ \mathbf{n}_i(D) \,\right] \in \mathbb{C}^{N_{R,i}\times D}$ is the stacked noise matrix for the i-th user.

In this article, the aim of the pilot design is to minimize the weighted total mean square errors of all channel estimators at their respective receivers, meanwhile satisfying the total power constraint for pilot transmission. In addition, for the sake of performance evaluation, assume that the noise power $\sigma_i^2 = \sigma_n^2, \forall i$, and define the signal-to-noise ratio (SNR) as

$$\mathrm{SNR} = P_T / \sigma_n^2.$$

Note that as the definition of the pilot-to-noise ratio (PNR) is the same with SNR, we use the latter one in subsequent sections.

III. THE PROPOSED OPTIMUM PRECODER DESIGN

A. PROBLEM FORMULATION

The MMSE estimator is widely adopted for channel estimation in communication systems, since it has moderate computational complexity and good performance. Hence, we consider MMSE estimation during the pilot transmission, which means that all receivers are equipped with MMSE estimators and have their respective channel statistical information.

First, from the definition of $\mathbf{H}_i$ in (2), the covariance of $\mathrm{vec}(\mathbf{H}_i)$, θ_i, is given by

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_{T,i}^T \otimes \boldsymbol{\theta}_{R,i} \tag{4}$$

It is well known that, the MMSE estimator $\hat{\mathbf{H}}_{\mathrm{MMSE},i}$ for the channel matrix $\mathbf{H}_i$ is given by [17]

$$\mathrm{vec}\left(\hat{\mathbf{H}}_{\mathrm{MMSE},i}\right) = \boldsymbol{\theta}_i \tilde{\mathbf{P}}^H \left(\tilde{\mathbf{P}} \boldsymbol{\theta}_i \tilde{\mathbf{P}}^H + \mathbf{I}_{DN_{R,i}}\right) \mathrm{vec}\left(\mathbf{Y}_i\right), \tag{5}$$

in which $\tilde{\mathbf{P}} = \mathbf{P}^T \otimes \mathbf{I}_{N_{R,i}}$. The error covariance matrix $\mathbf{C}_{\mathrm{MMSE,i}}$ becomes

$$\begin{aligned} \mathbf{C}_{\mathrm{MMSE},i} &\triangleq E\left\{\left[\mathrm{vec}\left(\mathbf{H}_i\right) - \mathrm{vec}\left(\mathbf{H}_{\mathrm{MMSE},i}\right)\right] \right. \\ &\qquad \left. \times \left[\mathrm{vec}\left(\mathbf{H}_i\right) - \mathrm{vec}\left(\mathbf{H}_{\mathrm{MMSE},i}\right)\right]\right\}^H \\ &= \left(\boldsymbol{\theta}_i^{-1} + \tilde{\mathbf{P}}^H \tilde{\mathbf{P}} / \sigma_i^2\right)^{-1} \end{aligned} \tag{6}$$

Subsequently, the channel estimation MSE and normalized MSE (NMSE) of user i can be expressed as

$$\begin{aligned} \mathrm{MSE}_i &= E\left\{\left\|\mathrm{vec}\left(\mathbf{H}_i\right) - \mathrm{vec}\left(\mathbf{H}_{\mathrm{MMSE},i}\right)\right\|^2\right\} \\ &= \mathrm{tr}\left[\left(\boldsymbol{\theta}_i^{-1} + \tilde{\mathbf{P}}^H \tilde{\mathbf{P}} / \sigma_i^2\right)^{-1}\right], \end{aligned} \tag{7a}$$

$$\mathrm{NMSE}_i = \mathrm{MSE}_i / \mathrm{tr}\left(\boldsymbol{\theta}_i\right). \tag{7b}$$

Therefore, with (7), we define the total weighted NMSE for all users as

$$f(\mathbf{P}) = \sum_{i=1}^{K} w_i \mathrm{NMSE}_i, \tag{8}$$

where w_i is the weight for user i, $0 \leq w_i \leq 1$, and $\sum_i w_i = 1$.

Then, we formulate the problem on the pilot design. From (8), it is seen that the pilot matrix $\mathbf{P}$ plays an important role in the total NMSE equation. The aim of this article is to find the optimum pilot matrix to minimize the total NMSE, meanwhile meeting the pilot power constraint. Based on the above, the optimization problem, denoted as P1, can be formulated as

$$\begin{aligned} \mathrm{P1}: \min_{\mathrm{P}} \quad & f(\mathrm{P}) = \sum_{i=1}^{K} w_i \mathrm{NMSE}_i = \sum_{i=1}^{K} w_i \\ & \mathrm{tr}\left[\boldsymbol{\theta}_i^{-1} + \left(\mathbf{P}^T \otimes \mathbf{I}_{N_{R,i}}\right)^H \right. \\ & \left. \times \left(\mathbf{P}^T \otimes \mathbf{I}_{N_{B,i}}\right) / \sigma_i^2\right]^{-1} / \mathrm{tr}\left(\boldsymbol{\theta}_i\right) \quad (9) \\ \mathrm{s.t.} \quad & \mathrm{tr}\left(\mathbf{P}^H \mathbf{P}\right) = P_T \quad (10) \end{aligned}$$

Let $\mathbf{Q} = \mathbf{P}^* \mathbf{P}^T$ and we refer to it as the pilot covariance matrix. It is clear that $\mathbf{Q} \in \mathbb{C}^{N_T \times N_T}$ is Hermitian and positive semidefinite. Considering

$$\left(\mathbf{P}^T \otimes \mathbf{I}_{N_{R,i}}\right)^H \left(\mathbf{P}^T \otimes \mathbf{I}_{N_{R,i}}\right) = \left(\mathbf{P}^* \mathbf{P}^T\right) \otimes \mathbf{I}_{N_{R,i}},$$

the problem P1 can be written as another form

$$\begin{aligned} \mathrm{P2}: \min_{\mathrm{Q}} \quad & f(\mathrm{Q}) = \sum_{i=1}^{K} w_i \mathrm{tr} \\ & \left[\left(\boldsymbol{\theta}_i^{-1} + \mathbf{Q} \otimes \mathbf{I}_{N_{Ri}} / \sigma_i^2\right)^{-1}\right] / \mathrm{tr}\left(\boldsymbol{\theta}_i\right) \\ \mathrm{s.t.} \quad & \mathrm{tr}(\mathbf{Q}) = P_T, \quad \mathbf{Q} \succeq 0 \end{aligned}$$

Remark 1: The weights in Eq. (8) are preconfigured according to the importance of users and the users of high importance are given large weights. Often, the users can be regarded as equally important and their weights are identical.

Remark 2: In the pilot design problem, the statistical CSI including $\{\boldsymbol{\theta}_{R,i}\}$, and $\{\boldsymbol{\theta}_{T,i}\}$, and $\{\sigma_i^2\}$ is required and this can be obtained by a period of measurement. Measuring $\{\sigma_i^2\}$ is easy to complete by the receivers and its process is omitted for brevity. The measurement of covariance matrices includes the following steps. First, the transmitter sends pilots and the receivers perform channel estimation to obtain the channel estimates, denoted by $\left\{\hat{\mathbf{H}}_k^{(i)}, i = 1 \cdots N_{sa}, k = 1 \cdots K\right\}$, where $\hat{\mathbf{H}}_k^{(i)}$ is the $i-$ th channel estimate for user k, N_{sa} is the number of samples. Second, the transmitter and receiver covariance matrices can be estimated as

$$\hat{\boldsymbol{\theta}}_{T,k} = \frac{1}{N_{sa}} \sum_{i=0}^{N_{sa}-1} \left(\hat{\mathbf{H}}_k^{(i)}\right)^H \hat{\mathbf{H}}_k^{(i)}$$

and

$$\hat{\boldsymbol{\theta}}_{R,k} = \frac{1}{N_{sa}} \sum_{i=0}^{N_{sa}-1} \hat{\mathbf{H}}_k^{(i)} \left(\hat{\mathbf{H}}_k^{(i)}\right)^H . \tag{11}$$

If N_{sa} is large, $\hat{\boldsymbol{\theta}}_{T,k} \approx \boldsymbol{\theta}_{T,k}$ and $\hat{\boldsymbol{\theta}}_{R,k} \approx \boldsymbol{\theta}_{R,k}$. Finally, when the measurement is completed, the receivers have the statistical CSI and then feed it back to the transmitter.

B. PRELIMINARIES OF THE POSITIVE DEFINITE MATRIX MANIFOLD

In recent years, matrix information geometry has been investigated and applied in neural network and information engineering [24]–[28]. By use of it, some problems can be converted into optimization problems on matrix manifolds. In [25], the authors formulated a network control problem as a cost function that involves matrix variables subjected to a few constraints. They further modeled this problem as an optimization problem on matrix manifolds to search its minimum value. In [26], for a system with multiple sensors for data collection, the power spectral density (PSD) matrices of the received signals were modeled as a so-called PSD matrix manifold in the signal space; by introducing new Riemannian metrics, novel algorithms were exploited to find the means and medians of the PSD matrices. In [27], the author proposed a novel metric-learning method that operates on logarithms of symmetric positive definite (SPD) matrices, in order to overcome the low efficiency of conventional algorithms in image set classifications.

In this article, the theory or results on positive definite matrix manifold will be utilized to solve the problem of pilot design. Therefore, a few preliminaries are provided in the followings.

Generally, an $n \times n$ complex Hermitian matrix $\mathbf{M}$ is said to be positive definite if the scalar $\mathbf{z}^H \mathbf{M}\mathbf{z}$ positive for an arbitrary nonzero $n \times 1$ complex column vector $\mathbf{z}$. All $n \times n$ positive definite matrices construct a positive definite Hermitian matrix manifold, denoted as $SPD\,(n)$. Define the Riemann metric on $SPD\,(n)$ as follows [24],

$$\langle \mathbf{A}_1, \mathbf{A}_2 \rangle_{\mathbf{M}} = \operatorname{tr}\left(\mathbf{M}^{-1}\mathbf{A}_1\mathbf{M}^{-1}\mathbf{A}_2\right) \tag{12}$$

where $\mathbf{M} \in SPD\,(n)$, $\mathbf{A}_1, \mathbf{A}_2 \in T_{\mathbf{M}}SPD\,(n)$, and $T_{\mathbf{M}}SPD\,(n)$ is the tangent space at the point $\mathbf{M}$. It can be proved that the positive definite matrix manifold $SPD\,(n)$ with such a metric is a complete Riemann manifold. Then, we present the *natural gradient* of a function.

Proposition 1: If the function $f : SPD\,(n) \to \mathbb{R}$ is smooth, with a Riemann metric defined by (12), its *natural gradient* $\nabla_{\mathbf{M}} f\,(\mathbf{M})$ is given by [24]

$$\nabla_{\mathbf{M}} f\,(\mathbf{M}) = \mathbf{M}\frac{\partial f}{\partial \mathbf{M}}\mathbf{M}, \tag{13}$$

where $\frac{\partial f}{\partial \mathbf{M}}$ denotes the partial derivative at the point $\mathbf{M}$ in Euclidean space.

Besides, the concept of geodesic is very important in the optimization of information geometry. The term geodesic comes from geodesy, the science of measuring the Earth. Originally, a geodesic was the shortest route between two points on the Earth's surface. Now, it is generalized and defined as the shortest route between two points on a manifold. The following proposition presents the geodesic equation.

Proposition 2: For the positive definite manifold $SPD\,(n)$, the geodesic that starts from $\mathbf{M}_0 \in SPD\,(n)$ with its direction $\mathbf{A} \in T_{\mathbf{M}_0}SPD\,(n)$, is parameterized as

$$\gamma(t) = \mathbf{M}_0^{1/2} \exp\left(t\mathbf{M}_0^{-1/2}\mathbf{A}\mathbf{M}_0^{-1/2}\right) \mathbf{M}_0^{1/2} \tag{14}$$

C. THE PROPOSED PILOT DESIGN METHOD

In this subsection, a numerical algorithm will be proposed to solve P2. We can prove that the objective function $f(\mathbf{Q})$ of P2 is convex. Though the conclusion can be inferred from *Proposition 1* in Ref. [20], herein, we provide another proof which is shown in Appendix A. Observe that there are two constraints in P2. Clearly, both the trace constraint and the positive semidefinite constraint are linear and hence convex. When the optimal $\mathbf{Q}$ is acquired, by taking the eigen-decomposition of $\mathbf{Q}$, i.e.,$\mathbf{Q} = \mathbf{U_Q}\boldsymbol{\Lambda}_\mathbf{Q}\mathbf{U}_\mathbf{Q}^H$, the optimal precoder $\mathbf{P}$ can be readily chosen as $\mathbf{P} = \mathbf{U_Q}\Lambda_\mathbf{Q}^{1/2}$.

First, the *Lagrangian* of P2 can be expressed as

$$\begin{aligned} L(\mathbf{Q}, u) &= (\mathbf{Q}) + u\,[\operatorname{tr}(\mathbf{Q}) - P_T] \\ &= \sum_{i=1}^{K} w_i \operatorname{tr}\left[\left(\boldsymbol{\theta}_i^{-1} + \mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}/\sigma_i^2\right)^{-1}\right] / \operatorname{tr}(\boldsymbol{\theta}_i) \\ &\quad + u\,[\operatorname{tr}(\mathbf{Q}) - P_T] \end{aligned} \tag{15}$$

where $u \geq 0$ is an introduced auxiliary variable.

Then, with (15), we define a *dual function* as

$$g(u) = \min_{Q \geq 0} L(\mathbf{Q}, u). \tag{16}$$

With (16), the dual problem of P2 can be formulated as (P3)

$$\text{P3} : J_{du} = \max_{u \geq 0} g(u). \tag{17}$$

It is obvious that there exists a strict feasible point for P2 so that the *Slater condition* is satisfied. Therefore, there is no dual gap between P2 and P3. In other words, the two problems are equivalent.

Next, we consider how to solve P3. The problem-solving trait is divided into two steps: 1) given u, obtain the optimal $\mathbf{Q}$ and $g(u)$; 2) search over u and find the minimum value of $g(u)$. Notice that $\mathbf{Q}$ in the function $L(\mathbf{Q}, u)$ must be positive semidefinite. If utilizing the conventional gradient decent algorithm to solve (16), we cannot guarantee the next iterative $\mathbf{Q}$ still be positive semidefinite. Hence, we resort to the theory of matrix information geometry. Specifically, we change the positive semidefinite constraint into positive definite and adopt the geodesic equation (14) to solve (16). It can be proved that solving $\min\limits_{\mathbf{Q}\succ\mathbf{0}} L(\mathbf{Q}, u)$ is equivalent to solving $\min\limits_{Q\succeq 0} L(\mathbf{Q}, u)$ (see Appendix C). In fact, this also illustrates that the positive semidefinite constraint of $\mathbf{Q}$ in P2 can be replaced by the positive definite constraint.

Differentiating $L(\mathbf{Q}, u)$ with respect to $\mathbf{Q}$, one has (see Appendix B)

$$\frac{\partial L(\mathbf{Q}, u)}{\partial \mathbf{Q}} = -\sum_{i=1}^{K} \left[w_i / \mathrm{tr}(\boldsymbol{\theta}_i)\right] \mathbf{G}_i + u\mathbf{I}_{N_T}, \tag{18}$$

where $\mathbf{G}_i \in \mathbb{C}^{N_T \times N_T} = \left(G_{i,mn}\right)$, $G_{i,mn} = \mathrm{tr}\left(\mathbf{B}_{i,nm}\right)$, $\mathbf{B}_i \in \mathbb{C}^{(N_T N_{R,i}) \times (N_T N_{R,i})} = \left(\mathbf{B}_{i,mn}\right)$, $\mathbf{B}_{i,mn} \in \mathbb{C}^{N_{R,i} \times N_{R,i}}$ is the (m, n)-th block of $\mathbf{B}_i$, $\mathbf{B}_i = \mathbf{A}_i^{-2}/\sigma_i^2$, and $\mathbf{A}_i = \boldsymbol{\theta}_i^{-1} + \left(\mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}\right)/\sigma_i^2$.

According to **Proposition 1**, the *natural gradient* of. $L(\mathbf{Q}, u)$ is given by

$$\nabla_{\mathbf{Q}} L(\mathbf{Q}, u) = \mathbf{Q} \frac{\partial L(\mathbf{Q}, u)}{\partial \mathbf{Q}} \mathbf{Q}. \tag{19}$$

With the geodesic equation in **Proposition 2**, it is easy to present its iterative form:

$$\mathbf{Q}_{i+1} = \mathbf{Q}_i^{1/2} \exp\left[-t\mathbf{Q}_i^{-1/2} \nabla_{\mathbf{Q}_i} L(\mathbf{Q}_i, u)\, \mathbf{Q}_i^{-1/2}\right] \mathbf{Q}_i^{1/2}, \quad t > 0. \tag{20}$$

Consequently, given u, we propose the following sub-algorithm to solve $g(u)$ and obtain $\mathbf{Q}$ afterwards.

Sub-algorithm I: Solve $g(u)$

1: Initialize the iteration index $i := 0$ and the searching step $t, 0 \leq t \leq 1$, and choose an arbitrary $N_T \times N_T$ complex matrix as $\mathbf{Q}^{(0)}$.
2: At the i-th iteration, compute the search direction $\nabla_{\mathbf{Q}^{(0)}} L\left(\mathbf{Q}^{(0)}, u\right)$ according to (19).
3: Update the matrix $\mathbf{Q}$ according to (20).
4: Increase the iteration index $i := i + 1$ and repeat 2-4 until $L(\mathbf{Q}, u)$ converges.

The above proposed sub-algorithm will compute the optimal $\mathbf{Q}$ and corresponding $g(u)$, given u. Based on Sub-Algorithm 1, we further propose a decent algorithm, termed as the Main algorithm 1, to solve the dual problem P3, summarized as follows.

Main algorithm I: Solve the Dual Problem P3 and Output the Optimal $\mathbf{Q}$

1: Initialize the iteration index $i = 0, {}_{u}^{(0)} \geq 0$, the searching step $s > 0$, and the constriction factor $0 < \beta < 1$; compute $g\left(u^{(0)}\right)$ and $\mathbf{Q}^{(0)}$ according to **Sub-algorithm 1**.
2: Update $u^{(i+1)}$ according to $u^{(i+1)} = \max\left(0, u^{(i)} + s\Delta u^{(i)}\right)$, where $\Delta u^{(i)}$ is the derivative of $g(u)$ with respect to u at $u^{(i)}$ and it is derived as

$$\Delta u^{(i)} \left.\frac{dg(u)}{du}\right|_{u=u^{(i)}} = \mathrm{Tr}\left[\mathbf{Q}^{(i)}\right] - P_T. \tag{21}$$

3: Compute $g\left(u^{(i+1)}\right)$ and $\mathbf{Q}^{(i+1)}$ according to *Sub-algorithm 1*. If $g\left(u^{(i+1)}\right) < g\left(u^{(i)}\right)$, $s := s \times \beta$, and go to step 2.
4: Increase the iteration index $i := i + 1$ and repeat 2-4 until $g(u)$ converges.

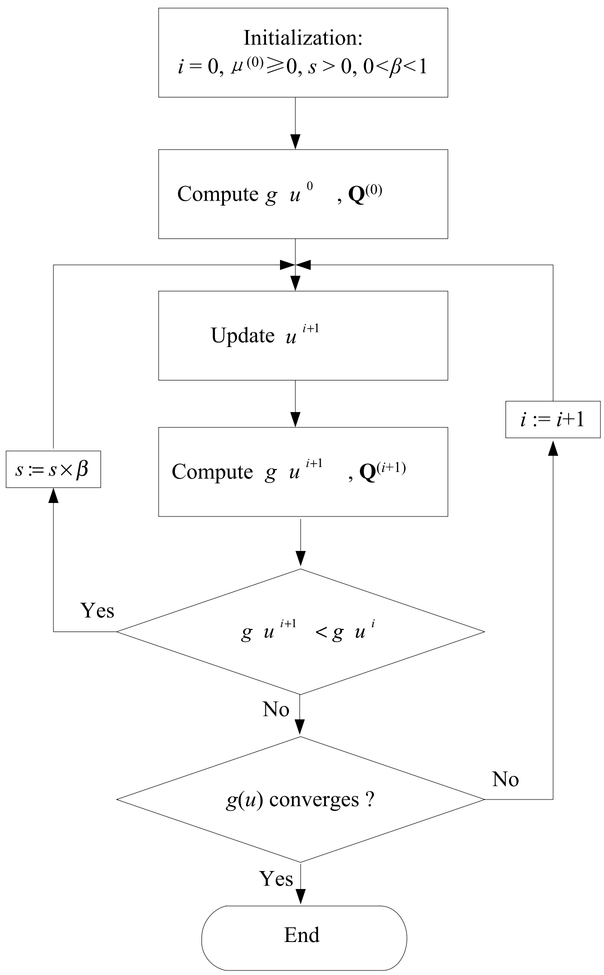


FIGURE 2. The flowchart of the proposed algorithm.

The flow chart of Main-algorithm 1 is depicted in Fig. 2. With the proposed Main-algorithm 1 above, we will obtain the optimal $\mathbf{Q}$, denoted as $\mathbf{Q}^{OPT}$, and the minimum $g(u)$.

According to the duality principle, the minimal MSE for the original problem P2 is $g(u)$'s minimum and the optimal $\mathbf{Q}$ for P2 is also $\mathbf{Q}^{OPT}$.

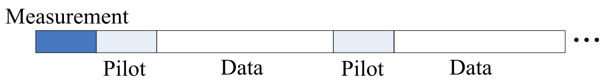


FIGURE 3. The system working sequence diagram.

Besides, the whole system process is shown in Fig. 3 and summarized as follows.

1) Obtain the statistical CSI including $\{\boldsymbol{\theta}_{R,i}\}$, $\{\boldsymbol{\theta}_{T,i}\}$, and $\{\sigma_i^2\}$ by a period of measurement.

2) Solve the optimal pilot matrix according to Main algorithm 1.

3) Enter the pilot transmission phase: the transmitter send the optimal pilot matrix and the receivers conduct channel estimation.

4) Enter the data transmission phase.

Note that, as aforementioned in Section II, the channel is block-fading and a block consists of a pilot transmission phase and a data transmission phase.

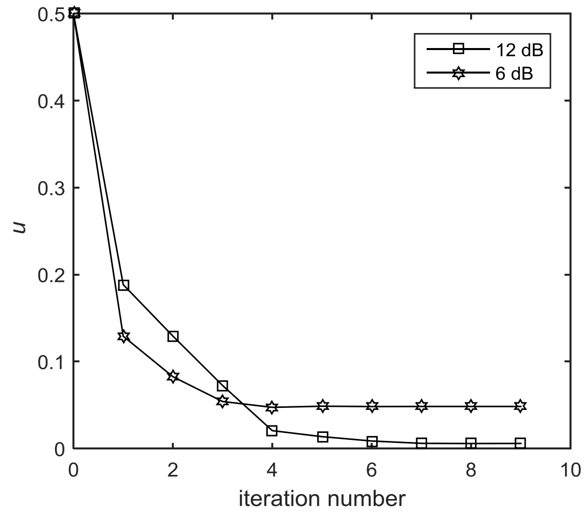
D. COMPUTATIONAL COMPLEXITY ANALYSIS

Usually, the number of floating point flops is used to measure of the complexity of an algorithm. Herein, a flop is defined as one floating point operation and has computational complexity $O(1)$. Note that the matrix product $\mathbf{X}_1\mathbf{X}_2$, requires $O(mnk)$ flops, where $X_1 \in \mathbb{C}^{m\times n}$, $X_2 \in \mathbb{C}^{n\times k}$; for an $n \times n$ matrix, both its inverse and eigen-decomposition operations requires $O(n^3)$ flops [31].

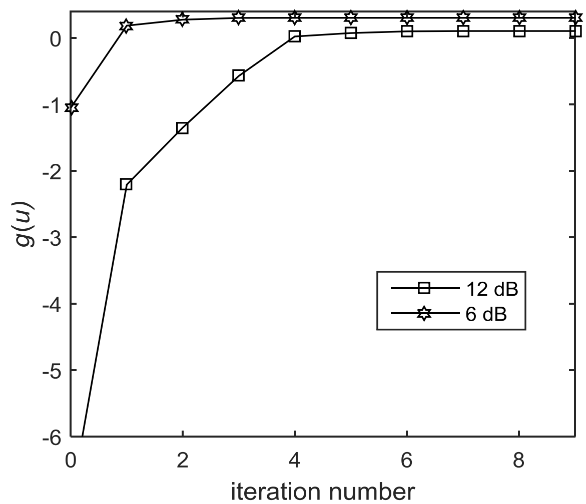
For the Main algorithm 1 and Sub-algorithm 1, suppose that numbers of iterations required to reach the given accuracy are T_1 and T_2, respectively. Observe that, the main computation of the proposed method lies in computing (19) and (20). The computation of (19) requires $O\left[N_T^3\left(N_R^{\max}\right)^3\right]$ flops, where $N_R^{\max} = \max_{i=1...K} N_{R,i};$; if the *eigenvector approach* is adopted to calculate the matrix exponential function [31], the computation of (20) requires $O\left(N_T^3\right)$ flops. Therefore, the complexity of the proposed method is $T_1T_2 \times O\left[N_T^3\left(N_R^{\max}\right)^3 + N_T^3\right] = T_1T_2O\left[N_T^3\left(N_R^{\max}\right)^3\right]$, where $N_R^{\max} = \max_{i=1\cdots K} N_{R,i}$.

IV. SIMULATION RESULTS

Computer simulation has been deployed to evaluate the performance of the proposed pilot-design algorithm, and also compare a few other existing methods. The MIMO broadcasting system model depicted in Fig. 1 is considered, in which the Rayleigh flat fading channel is adopted [32], [33]. The exponential correlation model is adopted as the channel correlation matrix, with its (i, j)-th entry being $\rho^{|i-j|}$, in which the constant ρ is the correlation coefficient. Further, define ρ_i^T as the transmitter correlation coefficient for $\boldsymbol{\theta}_{T,i}$ and ρ_i^R as the receiver correlation coefficient for $\boldsymbol{\theta}_{R,i}$, respectively. Unless otherwise specified, the weights in (8) are identical, i.e., $w_i = 1/K, \forall i$. We use the notation (K, N_T, N_R) to represent a K-user $N_T \times N_R$ system, where each user has N_R antennas. However, when users have different number of antennas, such a notation is no longer suitable and this will be specially stated.



a) u



b) $g(u)$

FIGURE 4. The iterates of $g(u)$ and u.

A. THE CONVERGENCE AND CORRECTNESS OF THE PROPOSED METHOD

Fig. 4 shows the convergence of the u and $g(u)$, in which a (2, 2, 2) broadcasting system is considered. The correlation coefficients ρ_1^T, ρ_2^T, ρ_1^R, and ρ_2^R are set to 0.5, 0.2, 0.1, and 0.5, respectively. Two cases of SNR = 6 dB and 10 dB are included. For both cases, we set $u = 0$ as its initial. Observe from Fig. 4a that, for SNR = 6 dB and 10 dB, after a few iterations, u converges to about 0.0058 and 0.0484, respectively. Meanwhile, for SNR = 6 dB in Fig. 4b, the function

TABLE 1. The transmitter and receiver coefficients with different system configurations.

System configuration	(2, 2, 2)		(2, 3, 1)		(3, 4, 2)		
Transmitter coefficients ρ_i^T	i=1	i=2	i=1	i=2	i=1	i=2	i=3
	0.5	0.2	0.5	0.2	0.82	0.45	0.53
Receiver coefficients ρ_j^R	j=1	j=2	j=1	j=2	j=1	j=2	j=3
	0.1	0.5	0.1	0.5	0.65	0.13	0.28

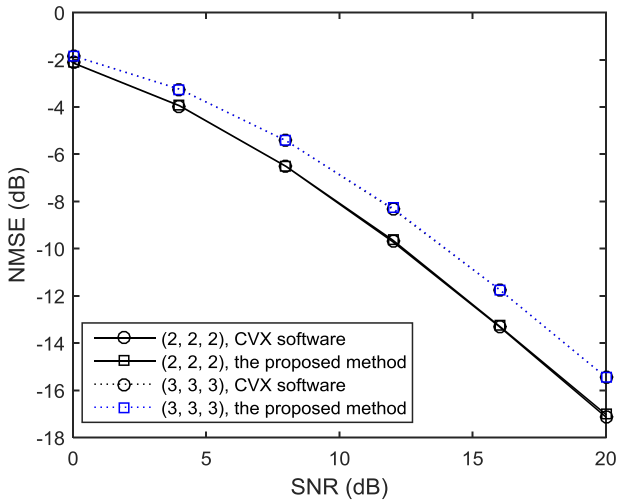


FIGURE 5. The NMSE comparison between the proposed method and the one based on the CVX software.

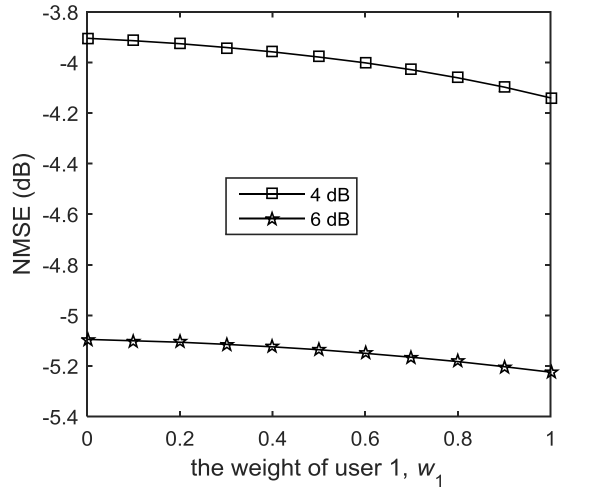


FIGURE 6. The effect of the weight of user 1 on the NMSE.

$g(u)$ rapidly increases from -1.05 to 0.19 at the first iteration, and then approaches 0.31 gradually. Similar phenomena can be found for SNR = 12 dB. **These results have well demonstrated the convergence of the function $g(u)$.**

Fig. 5 compares the NMSE performance between the proposed method and convex (CVX) software based method. As P2 is convex, it can be solved by convex optimization softwares directly [29]. In this figure, two scenarios are considered, including a two-user 2×2 system and a three-user 3×3 system. For the two-user system, the correlation coefficients ρ_1^T, ρ_2^T, ρ_1^R, and ρ_2^R are set to 0.1, 0.2, 0.5, and 0.6, respectively; for the three-user system, the correlation coefficients ρ_1^T, ρ_2^T, ρ_3^T, ρ_1^R, ρ_2^R, and ρ_3^R are set to 0.1, 0.5, 0.26, 0.2, 0.6, and 0.5, respectively. Observe that, the system NMSE is decreased with increasing SNR. At fixed SNR, the NMSE of the (2, 2, 2) system is less than that of the (3, 3, 3) system. At SNR = 12 dB, the NMSE gap between the two cases is about 1.4 dB. For both cases, there is nearly no NMSE gap between the proposed method and the CVX-software based method. **The results have well verified the correctness of the proposed method.**

Besides, TABLE 2 listed the optimal **Q**'s and the optimal pilot matrices under several system configurations, where the corresponding transmitter and receiver coefficients are listed in TABEL 1.

B. THE EFFECT OF KEY SYSTEM PARAMETERS

Fig. 6 shows the effect of the weight of user 1, w_1, on the NMSE. A two-user 2×2 system is studied, in which the transmitter and receiver correlation coefficients can be found in TABEL I. Two cases of SNR = 4 dB and SNR = 6 dB are include. Observe that for both cases, the total NMSE is decreased with increasing w_1. For $w_1 = 0/1$, the pilot is designed only with respect to user 2/user 1; for $0< w_1 <1$, both two users are considered. Given $0< w_1 <1$, the design will make a compromise between two users and satisfice their partial performance. The optimal pilot of P1 is not optimal for each user and both the two users cannot achieve their respective minimum values. For instance, at SNR = 4 dB, when considered separately, the minimum NMSE_1 and NMSE_2 are -4.14 dB and -3.90 dB, respectively. Given $w_1 = 0.5$, substituting the optimal pilot of P1, one has $\text{NMSE}_1 = -4.09\text{ dB} > -4.14\text{ dB}$ and $\text{NMSE}_2 = -3.87\text{ dB} > -3.90\text{ dB}$.

Fig. 7 shows the effect of the SNR on the eigenvalues of the pilot covariance matrix **Q**. Two scenarios are considered, including a two-user 2×2 system and a two-user 3×1 system, where the covariance coefficients can be found in TABEL I. The notation p_i denotes the i-th eigenvalue of **Q**. For the two-user 2×2 system, observe that at SNR = 0 dB, the gap between two eigenvalues is about 7 dB. As SNR increases, the gap is decreased and nearly disappears for SNR $\geq$ 15 dB. Similar phenomena can be found in the other scenario. This indicates that, **at high SNR, the optimal Q approaches $P_T/N_T\mathbf{I}_{N_T}$ and the optimal transmission scheme approaches the isotropic transmission.**

Fig. 8 presents the effect of the transmitter/receiver antenna number on the system NMSE, where the system has two users and the correlation coefficients ρ_1^T, ρ_2^T, ρ_1^R, and ρ_2^R are set to

TABLE 2. The optimal Q'S and pilot matrices.

System configuration	(2, 2, 2)	(2, 3, 1)	(3, 4, 2)
Q	$\begin{bmatrix} 3.15 & 0.43 \\ 0.43 & 3.15 \end{bmatrix}$	$\begin{bmatrix} 2.15 & 0.39 & 0.023 \\ 0.39 & 2.01 & 0.39 \\ 0.023 & 0.39 & 2.15 \end{bmatrix}$	$\begin{bmatrix} 1.767 & 0.786 & 0.084 & 0.041 \\ 0.786 & 1.387 & 0.75 & 0.084 \\ 0.084 & 0.75 & 1.39 & 0.79 \\ 0.041 & 0.084 & 0.79 & 1.77 \end{bmatrix}$
Pilot matrix **P**	$\begin{bmatrix} 1.17 & 1.34 \\ -1.17 & 1.34 \end{bmatrix}$	$\begin{bmatrix} 0.57 & -1.03 & 0.87 \\ -0.93 & 0 & 1.07 \\ 0.87 & 1.07 & 0.87 \end{bmatrix}$	$\begin{bmatrix} -0.17 & 0.57 & -0.91 & 0.76 \\ 0.34 & -0.47 & -0.45 & 0.92 \\ -0.34 & -0.47 & 0.45 & 0.92 \\ 0.17 & 0.57 & 0.91 & 0.76 \end{bmatrix}$

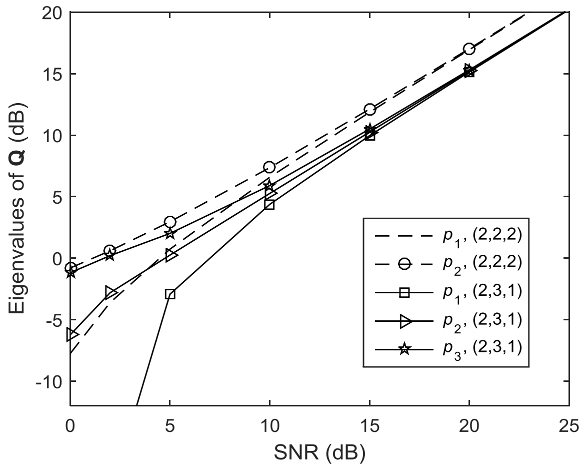


FIGURE 7. The effect of SNR on the eigenvalues of the pilot covariance matrix Q.

0.8, 0.7, 0.1, and 0.2, respectively. In Fig. 8a, we study the effect of transmitter antenna number N_T on the NMSE in a (2, N_T, 2) system. Clearly, as N_T increases, the NMSE increases for both SNR = 5 dB and 10 dB. In Fig. 8b, the number of transmitter antennas $N_T = 3$ and the number of antennas for user 1 is $N_{R,1} = 2$. This subfigure shows the effect of receiver antenna number of user 2, $N_{R,2}$, on the NMSE. Observe that there is nearly no variation as $N_{R,2}$ increases from 2 to 6. Hence, **these indicate that only the transmitter antenna number affects the NMSE and there is a positive correlation between them.**

Fig. 9 presents the effect of the correlation coefficients of user 1 on the system NMSE, where a (2, 4, 3) system and a (3, 4, 3) system are included. In Fig. 9a, for the (2, 4, 3) system, the coefficients ρ_2^T, ρ_1^R, and ρ_2^R are 0.65, 0.72, and 0.15, respectively; for the (3, 4, 3) system, the coefficients ρ_2^T, ρ_3^T, ρ_1^R, ρ_2^R, and ρ_3^R are 0.65, 0.53, 0.72, 0.15, and 0.45, respectively; for both systems, only ρ_1^T varies. Observe that the NMSE of the 3-user system is larger than or equal to that of the 2-user system on the whole. Increasing the transmitter coefficient decreases the NMSE, no matter what the user number or the SNR is. Similar phenomena can be found in Fig. 9b. **Therefore, the above illustrates that both the transmitter and receiver coefficients are negative correlated with the NMSE.**

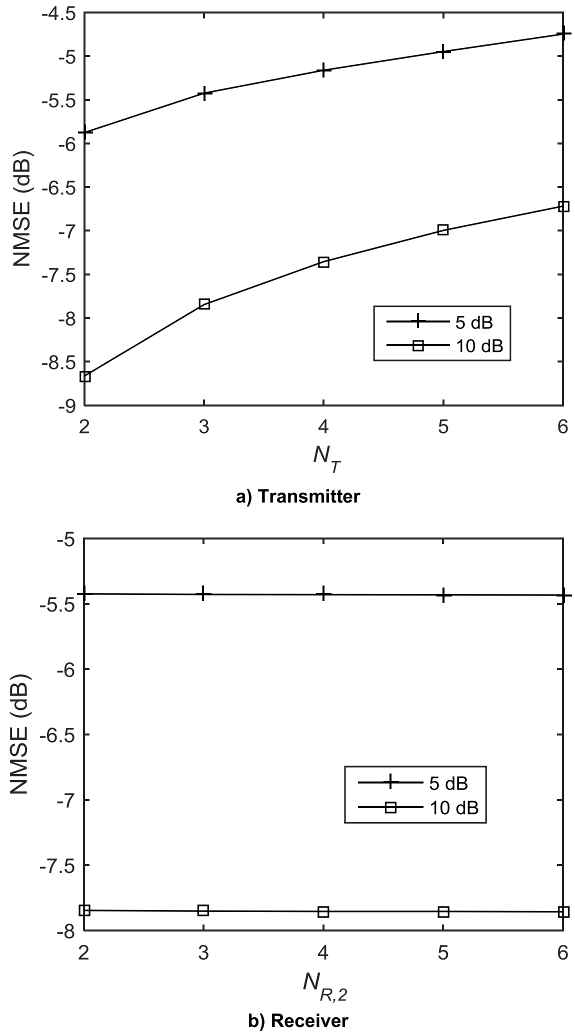
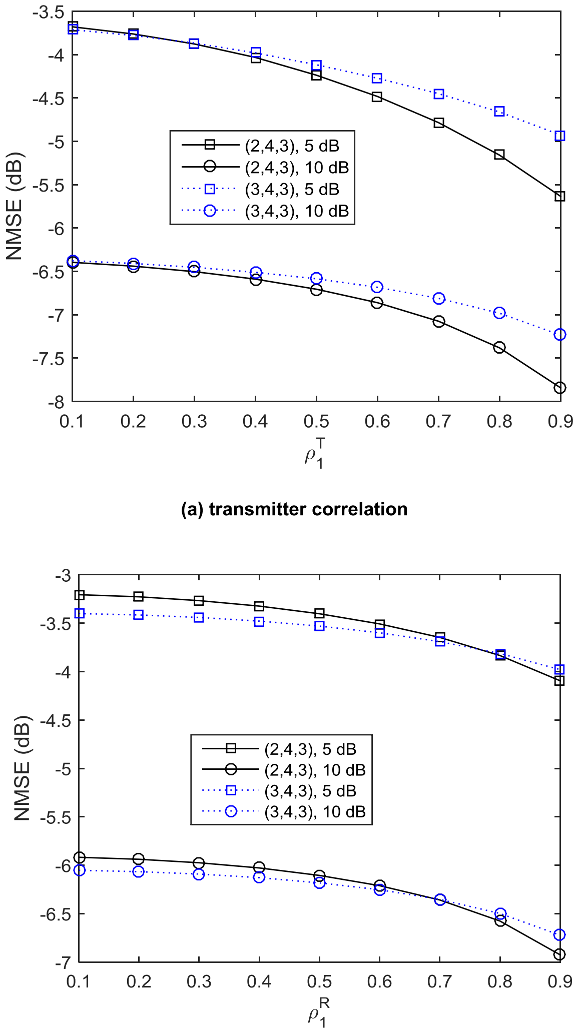


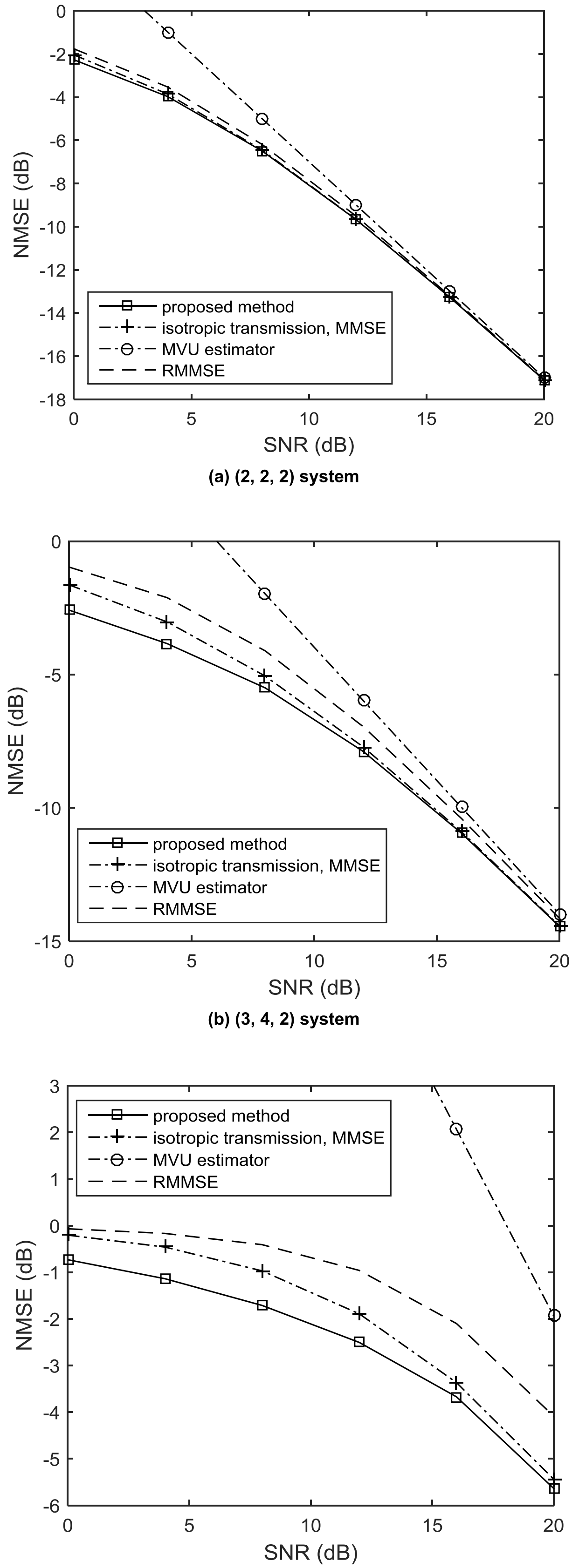
FIGURE 8. The effect of the transmitter/receiver antenna number on NMSE.



(a) transmitter correlation

(b) receiver correlation

FIGURE 9. The effect of the correlation coefficients of user 1 on NMSE.



(a) (2, 2, 2) system

(b) (3, 4, 2) system

(c) (16, 64, 2) system

FIGURE 10. The NMSE comparison of four methods.

C. THE COMPARISON BETWEEN THE PROPOSED METHOD AND A FEW EXISTING METHODS

In Fig. 10, the NMSE performance of four different methods is compared. The isotropic transmission refers to $\mathbf{P} = \sqrt{P_T/N_T}\mathbf{I}_{N_T}$ and the receivers adopt the MMSE estimators. The minimum variance unbiased (MVU) estimator and the relaxed MMSE (RMMSE) estimator were proposed in Ref. [8] and the details are omitted for brevity. The above three existing methods are used as baseline comparisons.

We consider a 2-user 2×2 system and a 3-user 4×2 system, where their respective correlation coefficients are set according to TABLE 1. Besides, considering the future scenario of massive MIMO, the system with a higher level of quantity is also included, in which $N_T = 64$, $K = 16$, $\rho_k^T = (k-1)/K$, and $\rho_k^R = 1-k/K$, $k = 1\cdots K$. In Fig. 10a, observe that the MVU estimator performs poorly since it is unaware of the channel statistics, especially at low SNR.

The proposed method has best performance among the four methods. However, the NMSE gaps are not noticeable among the proposed method, RMMSE, and isotropic transmission; this is because N_T is small. In Fig. 10b, with increasing N_T, the NMSE gaps among the four methods are a little larger, compared with Fig. 10a. In Fig. 10c, the superiority of the proposed method is more obvious. In addition, we also find that the isotropic transmission with MMSE estimator approaches the proposed method when SNR is large; **therefore, it is an optional choice instead of the proposed method, at high SNR.**

V. CONCLUSION

For the MIMO broadcasting system with correlated Rayleigh fading channels, the optimal pilot sequence has been studied and designed aiming at minimizing the system users' MSE of channel estimation. The design problem can be formulated as a SDP; by using the convex optimization technique, and after a few transformations, we transform it into a dual one. Both the two problems have no duality gap according to the *Slater's condition*. In the process of solving the *dual function*, we model it as an optimization problem on the positive definite matrix manifold. By introducing the concept of the *nature gradient*, we have solved this problem based on the iterative geodesic equation on this kind of manifold. Under various system configurations, the proposed method has been thoroughly investigated by computer simulation.

The results reveal the followings. **First**, the proposed method is convergent and also correct by mutual validation with the CVX-software based one. **Second**, only the transmitter antenna number affects the total NMSE and there is a positive correlation between them. **Third**, both the transmitter and receiver coefficients are negative correlated with the NMSE. **Fourth**, compared with a few existing methods, including the isotropic transmission with MMSE estimation, the RMMSE, and the MVU estimation, the proposed method has a better performance in terms of the total NMSE. At high SNR, the performance gap between the isotropic transmission with MMSE estimation and the proposed method is small, and hence it can be chosen as an optional scheme.

The proposed pilot-design method and above results can be a reference for practical system implementation. Our further work will consider how to enhance the robustness of the design in the case of incomplete channel statistical information.

In the future works, we will incorporate the wireless caching technique [36]–[39] to enhance the network transmission performance, especially for the transmission of video files. Moreover, we will apply the intelligent algorithms [40]–[43] to optimize the system design, especially when the system cannot gather all of the required channel parameters.

APPENDIX A

First, we prove that the function $\tilde{f}(\mathbf{X}) = \mathrm{tr}\left(\tilde{\boldsymbol{\theta}}^{-1} + \mathbf{X} \otimes \mathbf{I}\right)$ is convex with respect to $\mathbf{X}$, where $\tilde{\boldsymbol{\theta}}$ is positive definite and $\mathbf{X}$ is positive semidefinite. Let $\mathbf{X}_1 \succeq \mathbf{0}$, $\mathbf{X}_2 \succeq \mathbf{0}$, and construct a function $g(t)$ as follows.

$$g(t) = \mathrm{tr}\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]^{-1}\}, \ 0 \le t \le 1 \tag{A-1}$$

Then, we would like to find the second derivative of $g(t)$.

$$\begin{aligned} dg(t) &= -\mathrm{tr}\left\{\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]^{-1}\right.\\ &\quad \times d\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]\\ &\quad \left.\times\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]^{-1}\right\}\\ &= -\mathrm{tr}\left\{\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]^{-1}\right.\\ &\quad \times [(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\\ &\quad \left.\times\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]^{-1}\right\} dt \end{aligned}$$

Therefore, the first derivative of $g(t)$ is given by

$$\begin{aligned} \frac{dg(t)}{dt} &= -\mathrm{tr}\left\{\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]^{-1}\right.\\ &\quad \times [(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\\ &\quad \left.\left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]^{-1}\right\}. \end{aligned} \tag{A-2}$$

For the sake of easy manipulation in the followings, let $\mathbf{D} = \left[\tilde{\boldsymbol{\theta}}^{-1} + (t\mathbf{X}_1 + (1-t)\mathbf{X}_2) \otimes \mathbf{I}\right]$.

$$\begin{aligned} d\left(\frac{dg(t)}{dt}\right) &= \mathrm{tr}\left\{\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\, dt\right.\\ &\quad \left.\times \mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-1}\right\}\\ &\quad + \mathrm{tr}\left\{\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-1}\right.\\ &\quad \left.\times [(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\, dt\mathbf{D}^{-1}\right\}\\ &= 2\mathrm{tr}\left\{\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-1}\right.\\ &\quad \left.\times [(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-1}\right\} dt \end{aligned}$$

Therefore, the second derivative of $g(t)$ is given by

$$\frac{d^2 g(t)}{dt^2} = 2\mathrm{tr}\left\{\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-1}\right\}. \tag{A-3}$$

Further manipulating $\frac{d^2 g(t)}{dt^2}$ yields

$$\begin{aligned} \frac{d^2 g(t)}{dt^2} &= 2\mathrm{tr}\left\{\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-\frac{1}{2}}\right.\\ &\quad \left.\times \mathbf{D}^{-\frac{1}{2}}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-1}\right\}\\ &= 2\mathrm{tr}\left\{\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-\frac{1}{2}}\right.\\ &\quad \left.\times\left[\mathbf{D}^{-1}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]\,\mathbf{D}^{-\frac{1}{2}}\right]^H\right\}. \end{aligned} \tag{A-4}$$

In the derivation of the above formula, we have used a few matrix properties including $\mathbf{D}^H = \mathbf{D}$ and $[(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]^H = [(\mathbf{X}_1 - \mathbf{X}_2) \otimes \mathbf{I}]$. Clearly, from (A-4), we have $\frac{d^2 g(t)}{dt^2} \geq 0$. In other words, the function $g(t)$ is convex with respect to t. It follows that $tg(0) + (1-t)g(1) \geq g(t)$. Since $g(0) = \tilde{f}(\mathbf{X}_1)$, $g(1) = \tilde{f}(\mathbf{X}_2)$, and $g(t) = \tilde{f}[t\mathbf{X}_1 + (1-t)\mathbf{X}_2]$, one has

$$t\tilde{f}(\mathbf{X}_1) + (1-t)\tilde{f}(\mathbf{X}_2) \geq \tilde{f}[t\mathbf{X}_1 + (1-t)\mathbf{X}_2]. \quad \text{(A-5)}$$

Hence, the function $\tilde{f}(\mathbf{X}) = \mathrm{tr}\left(\tilde{\boldsymbol{\theta}}^{-1} + \mathbf{X} \otimes \mathbf{I}\right)$ is convex.

Second, it is straightforward that the function $\tilde{f}\left(\mathbf{X}/\sigma^2\right) = \mathrm{tr}\left(\tilde{\boldsymbol{\theta}}^{-1} + \mathbf{X} \otimes \mathbf{I}/\sigma^2\right)$ is also convex. Observe that the objective function $f(\mathbf{Q})$ of P2 is a nonnegative weighted sum of convex functions that are similar to $\tilde{f}\left(\mathbf{X}/\sigma^2\right)$. According the rule of preserving convexity [29], $f(\mathbf{Q})$ is convex.

APPENDIX B

To start with, consider the partial derivative of the function $\mathrm{tr}\left[\left(\boldsymbol{\theta}_i^{-1} + \mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}/\sigma_i^2\right)^{-1}\right]$, with respect to $\mathbf{Q}$.

$$\begin{aligned}
&\partial\left\{\mathrm{tr}\left[\left(\boldsymbol{\theta}_i^{-1} + \mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}/\sigma_i^2\right)^{-1}\right]\right\} \\
&\quad = -\mathrm{tr}\left\{\mathbf{A}_i^{-1}\partial\mathbf{A}_i\mathbf{A}_i^{-1}\right\} \\
&\quad = -\mathrm{tr}\left\{\mathbf{A}_i^{-1}\left(\partial\mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}\right)/\sigma_i^2\mathbf{A}_i^{-1}\right\} \\
&\quad = -\mathrm{tr}\left\{\left(\partial\mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}\right)/\sigma_i^2\mathbf{A}_i^{-1}\mathbf{A}_i^{-1}\right\} \\
&\quad = -\mathrm{tr}\left\{\left(\partial\mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}\right) \cdot \mathbf{B}_i\right\} \\
&\quad = -\mathrm{tr}\left(\mathbf{C}_i\right) \qquad \text{(B-1)}
\end{aligned}$$

where $\mathbf{A}_i = \boldsymbol{\theta}_i^{-1} + \left(\mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}\right)/\sigma_i^2$, $\mathbf{B}_i = \mathbf{A}_i^{-2}/\sigma_i^2$, and $\mathbf{C}_i \in \mathbb{C}^{(N_T N_{R,i}) \times (N_T N_{R,i})} = \left(\partial\mathbf{Q} \otimes \mathbf{I}_{N_{R,i}}\right) \cdot \mathbf{B}_i$.

Then, letting $\mathbf{Y} = \partial\mathbf{Q} = (\partial Q_{mk})$, it can be verified that the (m, n)-th block of the matrix $\mathbf{C}_i$ is given by

$$\mathbf{C}_{i,mn} = \sum_{k=1}^{N_T} Y_{mk}\mathbf{B}_{i,kn} \quad \text{(B-2)}$$

where $\mathbf{Y} = (Y_{mk})$ and $\mathbf{B}_{i,kn} \in \mathbb{C}^{N_{R,i} \times N_{R,i}}$ is the (k, n)-th block of $\mathbf{B}_i$. With (B-2), we have

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{C}_i\right) &= \mathrm{tr}\left(\sum_{m=1}^{N_T}\mathbf{C}_{i,mm}\right) = \mathrm{tr}\left[\sum_{m=1}^{N_T}\sum_{k=1}^{N_T} Y_{mk}\mathbf{B}_{i,km}\right] \\
&= \sum_{m=1}^{N_T}\sum_{k=1}^{N_T} Y_{mk}\mathrm{tr}\left(\mathbf{B}_{i,km}\right) = \sum_{m=1}^{N_T}\sum_{k=1}^{N_T} \partial Q_{mk}\mathrm{tr}\left(\mathbf{B}_{i,km}\right) \\
&= -\mathrm{tr}\left(\mathbf{A}\right). \qquad \text{(B-3)}
\end{aligned}$$

With (B-3), one obtains

$$-\frac{\partial\mathrm{tr}\left(\mathbf{A}\right)}{\partial Q_{mk}} = \mathrm{tr}\left(\mathbf{B}_{i,km}\right). \quad \text{(B-4)}$$

Consequently, we finally have

$$\frac{\partial\mathrm{tr}\left(\mathbf{A}\right)}{\partial\mathbf{Q}} = -\mathbf{G}_i \quad \text{(B-5)}$$

where $\mathbf{G}_i \in \mathbb{C}^{N_T \times N_T} = \left(G_{i,mn}\right)$ and $\mathbf{B}_{i,kn} \in \mathbb{C}^{N_{R,i} \times N_{R,i}}$ is the (k, n)-th block of $\mathbf{B}_i$. With (B-5) and Eq. (15), we obtain that

$$\frac{\partial L\left(\mathbf{Q}, u\right)}{\partial\mathbf{Q}} = -\sum_{i=1}^{K}\left[w_i/\mathrm{tr}\left(\theta_i\right)\right]\mathbf{G}_i + u\mathbf{I}_{N_T}. \quad \text{(B-6)}$$

APPENDIX C

This appendix shows the equivalence of $\min\limits_{\mathbf{Q}\succeq 0} L(\mathbf{Q}, u)$ and $\min\limits_{\mathbf{Q}\succeq 0} L(\mathbf{Q}, u)$.

First, when $u = 0$, it is easy to find that $\min\limits_{Q\succ 0} L(\mathbf{Q}, u) = \min\limits_{Q\succeq 0} L(\mathbf{Q}, u) = 0$, for $\mathbf{Q} = c\,\mathbf{I}$ and $c \to \infty$.

Second, when $u > 0$, this issue can be divided into two cases.

For one thing, if the optimal $\mathbf{Q}$ that satisfying $\min\limits_{\mathbf{Q}\succeq 0} L(\mathbf{Q}, u)$ is positive definite, obviously, it is also the solution of $\min\limits_{\mathbf{Q}\succ 0} L(\mathbf{Q}, u)$.

For another, to begin with, denote $\mathbf{Q}_A$ as the solution of $\min\limits_{\mathbf{Q}\succ 0} L(\mathbf{Q}, u)$ and $Q_B^{(i)} \succ 0$ as the solution of $\min\limits_{\mathbf{Q}\succ 0} L(\mathbf{Q}, u)$ at the i-th iteration, respectively. For $\mathbf{Q}_A$ is positive semidefinite but not positive definite, it is deduced that $\mathbf{Q}_B^{(i)}$ converges to $\mathbf{Q}_A$. This can be proved by using reduction to absurdity. Assume that $\mathbf{Q}_B^{(i)}$ converge to some matrix $\mathbf{Q}_B$, but $\mathbf{Q}_B \neq \mathbf{Q}_A$. Observe that the function $L\left(\mathbf{Q}, u\right)$ is convex and hence has only one extreme point which is also the global minimum. The matrix $\mathbf{Q}_B$ cannot be positive definite; if so, $\mathbf{Q}_B$ is also the solution of $\min\limits_{\mathbf{Q}\succ 0} L(\mathbf{Q}, u)$, which contradicts the prerequisite. Therefore, $\mathbf{Q}_B$ can only be positive semidefinite. Since $\mathbf{Q}_B \neq \mathbf{Q}_A$, we have. Introduce $\mathbf{Q}_A + \varepsilon\mathbf{I} \succ \mathbf{0}$ for $\varepsilon > 0$. If ε is small enough, the inequality $L\left(\mathbf{Q}_A, u\right) < L\left(\mathbf{Q}_A + \varepsilon\mathbf{I}, u\right) < L\left(\mathbf{Q}_B, u\right)$ holds. However, $L\left(\mathbf{Q}_A + \varepsilon\mathbf{I}, u\right) < L\left(\mathbf{Q}_B, u\right)$ cannot hold because $L\left(\mathbf{Q}_B, u\right)$ shall be larger than all $L\left(\mathbf{Q}, u\right)'s$ for $\mathbf{Q} \succ \mathbf{0}$. Therefore, the assumption that $\mathbf{Q}_B \neq \mathbf{Q}_A$ does not hold, and we have $\mathbf{Q}_B = \mathbf{Q}_A$, that is, $\mathbf{Q}_B^{(i)}$ converges to $\mathbf{Q}_A$.

To sum up, solving $\min\limits_{\mathbf{Q}\succ 0} L(\mathbf{Q}, u)$ is equivalent to solving $\min\limits_{\mathbf{Q}\succ 0} L(\mathbf{Q}, u)$.

REFERENCES

[1] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[2] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[3] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[4] M. Trivellato, F. Boccardi, and H. Huang, "On transceiver design and channel quantization for downlink multiuser MIMO systems with limited feedback," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1494–1504, Oct. 2008.

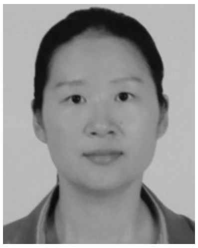
[5] K. Hamdi, W. Zhang, and K. B. Letaief, "Low-complexity antenna selection and user scheduling in cognitive MIMO broadcast systems," in *Proc. IEEE Int. Conf. Commun.*, May 2008, pp. 4038–4042.

[6] C. K. Au-Yeung, S. Y. Park, and D. J. Love, "A simple dual-mode limited feedback multiuser downlink system," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1514–1522, May 2009.

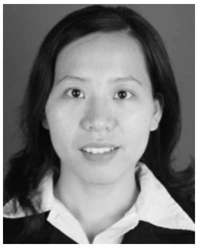
[7] D. Lee and K. Kim, "Sum-rate capacity analysis of the MIMO broadcast scheduling system with zero-forcing receivers under channel estimation error," *IEEE Commun. Lett.*, vol. 14, no. 3, pp. 223–225, Mar. 2010.
[8] J. Xu and L. Qiu, "Energy efficiency optimization for MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690–701, Feb. 2013.
[9] J. Mao, G. Xie, J. Gao, and Y. Liu, "Energy efficiency optimization for cognitive radio MIMO broadcast channels," *IEEE Commun. Lett.*, vol. 17, no. 2, pp. 337–340, Feb. 2013.
[10] X. Wang, X. Wang, and Y. Qi, "Low-complexity transceiver design for MIMO broadcast systems," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1661–1664, Sep. 2014.
[11] D. Vargas, Y. J. D. Kim, J. Bajcsy, D. Gómez-Barquero, and N. Cardona, "A MIMO-channel-precoding scheme for next generation terrestrial broadcast TV systems," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 445–456, Sep. 2015.
[12] J. Tang, D. K. C. So, N. Zhao, A. Shojaeifard, and K.-K. Wong, "Energy efficiency optimization with SWIPT in MIMO broadcast channels for Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2605–2619, Aug. 2018.
[13] F. Benkhelifa, A. S. Salem, and M.-S. Alouini, "Sum-rate enhancement in multiuser MIMO decode-and-forward relay broadcasting channel with energy harvesting relays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3675–3684, Dec. 2016.
[14] H. Niu, D. Guo, Y. Huang, B. Zhang, and B. Gao, "Outage constrained robust energy harvesting maximization for secure MIMO SWIPT systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 614–617, Oct. 2017.
[15] Z. Zhu, Z. Chu, N. Wang, Z. Wang, and I. Lee, "Energy harvesting fairness in AN-aided secure MU-MIMO SWIPT systems with cooperative jammer," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[16] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.
[17] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.
[18] K.-C. Hung and D. W. Lin, "Pilot-based LMMSE channel estimation for OFDM systems with power–delay profile approximation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 150–159, Jan. 2010.
[19] P. Jiyong, L. Jiandong, L. Zhuo, Z. Linjing, and C. Liang, "Optimal training sequences for MIMO systems under correlated fading," *J. Syst. Eng. Electron.*, vol. 19, no. 1, pp. 33–38, Feb. 2008.
[20] N. Shariati, J. Wang, and M. Bengtsson, "Robust training sequence design for correlated MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 107–120, Jan. 2014.
[21] K. Mawatwal, D. Sen, and R. Roy, "A semi-blind channel estimation algorithm for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 70–73, Feb. 2017.
[22] Z. Albataineh, "Robust blind channel estimation algorithm for linear STBC systems using fourth order cumulant matrices," *Telecommun. Syst.*, vol. 68, no. 3, pp. 573–582, 2018.
[23] D. Lee and H. Kang, "Symbol error rate of MIMO broadcast scheduling systems with partial channel state information," *IEEE Commun. Lett.*, vol. 15, no. 6, pp. 638–640, Jun. 2011.
[24] H. F. Sun, Z. N. Zhang, L. Y. Peng, *An Elementary Introduction to Information Geometry*. Beijing, China: Science Press, 2016.
[25] G. Li, T. Pei, Z. Meng, C. Wen, J. Pei, and L. Shi, "Optimization on matrix manifold based on gradient information and its applications in network control," *Phys. A, Stat. Mech. Appl.*, vol. 508, pp. 481–500, Oct. 2018.
[26] K. M. Wong, J.-K. Zhang, and H. Jiang, "Multi-sensor signal processing on a PSD matrix manifold," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.
[27] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen, "Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 720–729.
[28] Y. Wu, Y. Jia, P. Li, J. Zhang, and J. Yuan, "Manifold Kernel sparse representation of symmetric positive-definite matrices and its applications," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3729–3741, Nov. 2015.
[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Press, 2004.
[30] M. Grant and S. Boyd. *CVX: MATLAB Software for Disciplined Convex Programming*. Accessed: Jul. 23, 2019. [Online]. Available: http://cvxr.com/cvx
[31] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2013.
[32] X. Lai, L. Fan, X. Lei, J. Li, N. Yang, and G. K. Karagiannidis, "Distributed secure switch-and-stay combining over correlated fading channels," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 8, pp. 2088–2101, Aug. 2019.
[33] X. Lai, W. Zou, D. Xie, X. Li, and L. Fan, "DF relaying networks with randomly distributed interferers," *IEEE Access*, vol. 5, pp. 18909–18917, 2017.
[34] G. Liu, "Deep learning based channel prediction for edge computing networks towards intelligent connected vehicles," *IEEE Access*, to be published.
[35] Z. Zhao, "A novel framework of three-hierarchical offloading optimization for MEC in industrial IoT networks," *IEEE Trans. Ind. Informat.*, to be published.
[36] F. Shi, J. Xia, Z. Na, X. Liu, Y. Ding, and Z. Wang, "Secure probabilistic caching in random multi-user multi-UAV relay networks," *Phys. Commun.*, vol. 32, pp. 31–40, Feb. 2019.
[37] X. Lin, J. Xia, and Z. Wang, "Probabilistic caching placement in UAV-assisted heterogeneous wireless networks," *Phys. Commun.*, vol. 33, pp. 54–61, Apr. 2019.
[38] L. Fan, N. Zhao, X. Lei, Q. Chen, N. Yang, and G. K. Karagiannidis, "Outage probability and optimal cache placement for multiple amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12373–12378, Dec. 2018.
[39] J. Xia, "Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers," *IEEE Trans. Commun.*, to be published.
[40] C. Li, Y. Xu, J. Xia, and J. Zhao, "Protecting secure communication under UAV smart attack with imperfect channel estimation," *IEEE Access*, vol. 6, pp. 76395–76401, 2018.
[41] Y. Xu and J. Xia, "Q-learning based physical-layer secure game against multi-agent attacks," *IEEE Access*, vol. 7, pp. 49212–49222, 2019.
[42] C. Li, W. Zhou, K. Yu, L. Fan, and J. Xia, "Enhanced secure transmission against intelligent attacks," *IEEE Access*, vol. 7, pp. 53596–53602, 2019.
[43] X. Lin, J. Xia, X. Lei, J. Xia, Q. Zhou, H. Wu, and L. Fan, "MARL-based distributed cache placement for wireless networks," *IEEE Access*, vol. 7, pp. 62606–62615, 2019.

WEN ZHOU received the Ph.D. degree in engineering from Hong Kong University, in 2010. He is currently an Associate Professor with the College of Information Science and Technology, Nanjing Forestry University. His research interests include optimization designs in MIMO systems, information geometry in engineering, and forestry Internet of Things.

XUTAO LI (M'09) received the B.S., M.E., and Ph.D. degrees in electronics engineering from the Huazhong University of Science and Technology, in 1995, 2003, and 2006, respectively. From 2006 to 2008, he was a Postdoctoral Fellow with the South China University of Technology. Since 2013, he has been a Professor with the Department of Electronic Engineering, Shantou University. He has published more than 30 research papers in international referred journals. His current research interests include array signal processing, radar systems, computer vision, machine learning, and nonlinear presentation to signals.



JUNJUAN XIA received the bachelor's degree from the Department of Computer Science, Tianjin University, in 2003, and the master's degree from the Department of Electronic Engineering, Shantou University, in 2015. She is currently with the School of Computer Science, Guangzhou University, as a Laboratory Assistant. Her current research interests include wireless caching, physical-layer security, cooperative relaying, and interference modeling.



WANGUO JIAO received the B.S. degree in network engineering from Shijiazhuang Tiedao University, Shijiazhuang, China, in 2009, and the Ph.D. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2015. Since 2015, she has been with the College of Information Science and Technology, Nanjing Forestry University, where she is currently an Assistant Professor. Her research interests include performance analysis and protocol design of multi-hop wireless networks, cognitive radio networks, and wireless sensor networks.

MIN HUA received the Ph.D. degree from the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2018. Since 2018, she has been with the College of Information Science and Technology, Nanjing Forestry University, where she is currently a Lecturer. She has over ten IEEE journal publications. Her research interests include wireless communications, signal processing, and the Internet of Things.