Robust Transmission for Massive MIMO Downlink with Imperfect CSI

An-An Lu, Member, IEEE, Xiqi Gao, Fellow, IEEE, Wen Zhong and Chengshan Xiao, Fellow, IEEE

Abstract

In this paper, the design of robust linear precoders for the massive multi-input multi-output (MIMO) downlink with imperfect channel state information (CSI) is investigated, where each user equipment (UE) is equipped with multiple antennas. The imperfect CSI for each UE obtained at the BS is modeled as statistical CSI under a jointly correlated channel model with both channel mean and channel variance information, which includes the effects of channel estimation error, channel aging and spatial correlation. The design objective is to maximize the expected weighted sum-rate. By combining the minorize-maximize (MM) algorithm with the deterministic equivalent method, an algorithm for robust linear precoder design is derived. The proposed algorithm achieves a local optimum of the expected weighted sum-rate maximization problem. To reduce the computational complexity of the proposed algorithm, two low-complexity algorithms are then derived. One for the general case, and the other for the case when all the channel means are zeros. The optimality of the beam domain transmissions when all the channel means are zeros is also proved. Simulation results show that the proposed robust linear precoder designs apply to various mobile scenarios and achieve high spectral efficiency.

Index Terms

Massive multi-input multi-output (MIMO), minorize-maximize (MM) algorithm, deterministic equivalents, robust linear precoders, imperfect CSI.

A.-A. Lu, X. Q. Gao and W. Zhong are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, 210096 China, E-mail: aalu@seu.edu.cn, xqgao@seu.edu.cn, wzhang@seu.edu.cn.
C. Xiao is with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015. Email: xiaoc@lehigh.edu.
I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) [1]–[8] is one of the key technologies of next generation wireless networks. It provides huge potential capacity gains by employing hundreds of antennas at a base station (BS) and supports multi-user MIMO (MU-MIMO) transmissions on the same time and frequency resource. With massive antenna arrays at the BS, it is also possible to achieve high energy efficiency. Most of the literature on massive MIMO assume that each user equipment (UE) has a single antenna. Since UEs with multiple antennas have already been used in current wireless systems, it is also important to investigate massive MIMO systems with multiple antenna users.

In this paper, we focus on linear precoder designs for massive MIMO downlinks. The sum-rate is a commonly used metric for evaluating the performance of the massive MIMO downlink. To alleviate the multi-user interference and improve the sum-rate performance, the linear precoders for all the UEs at the BS should be properly designed. The precoder designs are based on the available channel state information (CSI) at the BS. When the BS has perfect CSI of all UEs, the classic iterative weighted minimum mean square error (WMMSE) method [9]–[11] from conventional MU-MIMO can be used. It converges to a local optimum of the weighted sum-rate maximization problem. When the BS has statistical CSI with only channel covariance information, there exists the beam division multiplex access (BDMA) transmission [5], [12] and the joint spatial division and multiplexing (JSDM) approach [13]. In the BDMA transmission, the BS serves multiple users via different beams simultaneously. In the JSDM approach, the users are partitioned into groups with approximately the same channel covariance eigenspace. In conclusion, when the channel of all users are quasi-static, the iterative WMMSE method can be used to design linear precoders for massive MIMO downlinks. On the contrary, when all users are in medium or high mobility scenarios, the BDMA transmission or the JSDM approach can be used.

In practical massive MIMO systems, perfect CSI at the BS are usually not available due to channel estimation error, channel aging, etc. Furthermore, different user usually has different moving speed. Thus, the two kinds of methods mentioned above are insufficient for use in practical systems. Then, it is naturally to ask whether there exist any unified linear precoding method which is robust against imperfect CSI for massive MIMO downlinks. The goal of this paper is to answer this important question. We consider a massive MIMO downlink where the a priori CSI for each UE available at the BS before channel estimation is expressed as a jointly correlated channel model [14] with only channel covariance information. After channel estimation, we model the a posteriori CSI for each UE at the BS as statistical CSI under a jointly correlated channel model with both channel mean and channel covariance information. We assume that the UEs obtain the perfect CSI of their corresponding channel matrices in the precoding domain. At each UE, we treat the aggregate interference-plus-noise as Gaussian noise and assume only their covariance matrices are known. The design objective is to maximize the expected weighted sum-rate of the considered downlink.

Before proceeding to our method, we point out here a relevant research, the stochastic weighted MMSE approach [15], [16]. It was extended from the iterative WMMSE method and proposed in [15] to maximize the ergodic sum rate for a MIMO interference channel. It is an sample average approximation (SAA) method [17], [18], which
use a sample average problem to approximate the original optimization problem. Furthermore, a block successive upper-bound minimization (BSUM) algorithm for the linear precoder designs was proposed. The precoders depend on all of the past channel realizations, which can be the actual CSI or generated virtually using known channel statistics \[15\].

In this paper, we propose to use the MM (majorize-minimize or minorize-maximize) algorithm \[19\], \[20\] to solve the problem of maximizing the expected weighted sum-rate. The MM algorithm is a widely used method to find the local optimums of complicated optimization problems. It substitutes a simple optimization problem for a difficult optimization problem. Inspired by the stochastic weighted WMMSE method, we find a convex quadratic minorizing function of the objective function which can be used to apply the MM algorithm. The optimal solution of the surrogate problem needs calculating the expected values of several random matrices with respect to the channel matrices based on the established \textit{a posteriori} channel model. To compute the optimal solution, we use the deterministic equivalent method \[21\], \[22\], which can be used to compute the approximations of the matrix expectations needed. Based on the obtained approximations, we propose an algorithm for robust linear precoder design. Furthermore, we derive two low-complexity algorithms by reducing the number of large dimensional matrix inversions and avoiding large dimensional matrix inversions, respectively.

The rest of this article is organized as follows. The system model and problem formulation are presented in Section II. The robust linear precoder designs based on the deterministic equivalents are shown in Section III. Simulation results are provided in Section IV. The conclusion is drawn in Section V. Proofs of Lemmas and Theorems are provided in Appendices.

\textbf{Notations:} Throughout this paper, uppercase and lowercase boldface letters are used for matrices and vectors, respectively. The superscripts $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote the conjugate, transpose and conjugate transpose operations, respectively. $E\{\cdot\}$ denotes the mathematical expectation operator. In some cases, where it is not clear, subscripts will be employed to emphasize the definition. The operators $\text{tr}(\cdot)$ and $\det(\cdot)$ represent the matrix trace and determinant, respectively. The operator $\otimes$ denotes the kronecker product. The Hadamard product of two matrices $A$ and $B$ of the same dimensions is represented by $A \odot B$. The $N \times N$ identity matrix is denoted by $I_N$. The $(i,j)$-th entry of the matrix $A$ is denoted by $[A]_{ij}$.

\section{System Model and Problem Formulation}

\subsection{System Model}

We consider a massive MIMO system with block flat fading channels, where the channel coefficients remain constant for a coherence interval of $T$ symbol periods. The system consists of one BS and $K$ UEs. The number of antennas at the BS is $M_t$. The $k$-th UE is equipped with $M_k$ antennas, and $\sum_{k=1}^{K} M_k = M_r$. We divide the time resources into slots and each time slot contains $N_b$ blocks. In this paper, we focus on the case where the considered massive MIMO system operates in time division duplexing (TDD) mode. However, the results of this paper can be extended to the system operating in frequency division duplexing (FDD) mode easily. For simplicity, we assume that there only exists the uplink training phase and the downlink transmission phase. At each slot, the uplink training sequences are sent once at the first block. The second block to the $N_b$-th block are used for downlink
transmission. The length of the uplink pilot sequences is $T$ symbols, i.e., the length of each block. Furthermore, the uplink training sequences assigned to different antennas are orthogonal to each other ($M_r \leq T$). For illustration purpose, we plot the time slot structure in Fig. 1.

We restrict our considerations to stationary channels and use the jointly correlated channel model to describe the spatial correlations of each channel. Specifically, the channel matrix $H_{k,m,n}$ from the BS to the $k$-th UE at the $n$th block of slot $m$ has the following structure [14]

$$H_{k,m,n} = U_k (M_k \odot W_{k,m,n}) V_k^H$$

(1)

where $U_k$ and $V_k$ are deterministic unitary matrices, $M_k$ is an $M_k \times M_t$ deterministic matrix with nonnegative elements, and $W_{k,m,n}$ is a complex Gaussian random matrix with independent and identically distributed (i.i.d.), zero mean and unit variance entries. In massive MIMO systems, $M_t$ can go to a very large value, and $V_k$ can be assumed to be independent of $k$. In this paper, we assume that uniform linear arrays (ULAs) are employed in the BS and $M_t$ grows very large. Then, each $V_k$ is closely approximated by a discrete Fourier transform (DFT) matrix [23], [24]. Further, the channel model in (1) can be rewritten as

$$H_{k,m,n} = U_k (M_k \odot W_{k,m,n}) V_{M_t}^H$$

(2)

where $V_{M_t}$ denotes the $M_t \times M_t$ DFT matrix. The channel model in (2) can be seen as an a priori model of the channels before channel estimation. As in [25]–[27], we model the time variation of the channel from block to block by a first order Gauss-Markov process as

$$H_{k,m,n+1} = \alpha_k H_{k,m,n} + \sqrt{1 - \alpha_k^2} U_k (M_k \odot W_{k,m,n+1}) V_{M_t}^H$$

(3)

where $\alpha_k$ is the temporal correlation coefficient which is related to the moving speed. An often used metric for $\alpha_k$ in the literature [26] is related to Jakes’ autocorrelation model, i.e., $\alpha_k = J_0(2\pi v_k f_c T/c)$, where $J_0(\cdot)$ is the zero order Bessel function of the first kind, $v_k$ is the moving speed of the $k$-th user, $f_c$ is the carrier frequency and $c$ is the light speed.

We define the channel coupling matrices (CCMs) $\Omega_k$ as $\Omega_k = M_k \odot M_k$ and assume the BS knows $U_k$ and $\Omega_k$ through a channel sounding process. Exploiting channel reciprocity, the channel state information of the downlink channels can be obtained from uplink training signals [28].

Fig. 1. Time slot structure.
Let $Y_{m,1}^{BS} \in \mathbb{C}^{M_t \times T}$ denote the received matrix at the BS on the first block of slot $m$. It can be written as

$$Y_{m,1}^{BS} = \sum_{k=1}^{K} H_{k,m,1}^{T} X_{k,m,1}^{UE} + Z_{m,1}^{BS}$$  \hspace{1cm} (4)$$

where $X_{k,m,1}^{UE} \in \mathbb{C}^{M_t \times T}$ denotes the uplink training matrix sent by the $k$-th user on the first block of slot $m$, and $Z_{m,1}^{BS} \in \mathbb{C}^{M_t \times T}$ is a noise random matrix whose elements are i.i.d. complex Gaussian entries with zero mean and variance $\sigma_B^2$.

In the following, we model the a posteriori CSI for each UE at the BS given $Y_{m,1}^{BS}$ as statistical CSI under a jointly correlated channel model with both channel mean and channel covariance information. Vectorizing the received matrix $Y_{m,1}^{BS}$, we obtain

$$\text{vec}(Y_{m,1}^{BS}) = \sum_{k=1}^{K} ((X_{k,m,1}^{UE})^T \otimes I_{M_t}) \text{vec}(H_{k,m,1}^{T}) + \text{vec}(Z_{m,1}^{BS}).$$  \hspace{1cm} (5)$$

Let $K_{k,m,1}$ denote the covariance matrix of $\text{vec}(H_{k,m,1}^{T})$. From (2), we then obtain

$$K_{k,m,1} = \mathbb{E} \{ \text{vec}(H_{k,m,1}^{T})\text{vec}(H_{k,m,1}^{T})^T \} = (U_k \otimes V_{M_t}^*) \text{diag}(\text{vec}(\Omega_k^T))(U_k^H \otimes V_{M_t}^T).$$  \hspace{1cm} (6)$$

Let $\hat{H}_{k,m,n}^{T}$ denote the MMSE estimator of $H_{k,m,n}^{T}$ given $Y_{m,1}^{BS}$, we have that $\hat{H}_{k,m,n}$ is the conditional mean of $H_{k,m,n}^{T}$ given $Y_{m,1}^{BS}$, i.e.,

$$\hat{H}_{k,m,n}^{T} = \mathbb{E} \{ H_{k,m,n}^{T} | Y_{m,1}^{BS} \}.$$  \hspace{1cm} (7)$$

From (2), (5) and (6), we then obtain

$$\text{vec}(\hat{H}_{k,m,n}^{T}) = \alpha_k^{-1} K_{k,m,1} \left( (X_{k,m,1}^{UE})^* \otimes I_{M_t} \right) \left( \sum_{k=1}^{K} ((X_{k,m,1}^{UE})^T \otimes I_{M_t}) K_{k,m,1} \left( (X_{k,m,1}^{UE})^* \otimes I_{M_t} \right) + \sigma_B^2 I \right)^{-1} \text{vec}(Y_{m,1}^{BS}).$$  \hspace{1cm} (8)$$

Since the pilot sequences assigned to transmitted antennas are orthogonal, we obtain $X_{k,m,1}^{UE}(X_{k,m,1}^{UE})^H = I_{M_t}$ and $X_{l,m,1}^{UE}(X_{k,m,1}^{UE})^H = 0$ for $l \neq k$. Using these conditions, we can rewrite (8) as

$$\text{vec}(\hat{H}_{k,m,n}^{T}) = \alpha_k^{-1} \left( K_{k,m,1} + \sigma_B^2 I \right)^{-1} K_{k,m,1} \left( (X_{k,m,1}^{UE})^* \otimes I_{M_t} \right) \text{vec}(Y_{m,1}^{BS}).$$  \hspace{1cm} (9)$$

Substituting (9) into (5), we obtain

$$\text{vec}(\hat{H}_{k,m,n}^{T}) = \alpha_k^{-1} (U_k \otimes V_{M_t}^*) \text{diag}(\text{vec}(\Omega_k^T)) + \sigma_B^2 I \text{diag}(\text{vec}(\Omega_k^T))(U_k^H (X_{k,m,1}^{UE})^* \otimes V_{M_t}^T) \text{vec}(Y_{m,1}^{BS}).$$  \hspace{1cm} (10)$$

Let $\Delta_k$ denote the matrix whose entries are defined by

$$|\Delta_k|_{ij} = \frac{|M_k|_{ij}^2}{|M_k|_{ij} + \sigma_B^2}.$$  \hspace{1cm} (11)$$

Then, (10) can be re-expressed as

$$\text{vec}(\hat{H}_{k,m,n}^{T}) = \alpha_k^{-1} (U_k \otimes V_{M_t}^*) \text{diag}(\text{vec}(\Delta_k^T)) (U_k^H (X_{k,m,1}^{UE})^* \otimes V_{M_t}^T) \text{vec}(Y_{m,1}^{BS}).$$  \hspace{1cm} (12)$$
From \((2\), \(3\), \(5\) and \(12\), we obtain the posterior distribution of the random vector \(\text{vec}(H_{k,m,n}^T)\) given \(Y_{m,1}\) is a multivariate Gaussian distribution, and its conditional covariance matrix is given by

\[
\mathbb{E}\left\{ (\text{vec}(H_{k,m,n}^T) - \text{vec}(\hat{H}_{k,m,n}^T))(\text{vec}(H_{k,m,n}^T) - \text{vec}(\hat{H}_{k,m,n}^T))^H | Y_{m,1} \right\} = (U_k \otimes V^*_M) \text{diag}(\text{vec}(\Xi_{k,m,n}^T \otimes \Xi_{k,m,n}^T))(U_k^H \otimes V^T_M)
\]

(13)

where the square of the elements in \(\Xi_{k,m,n} \in \mathbb{C}^{M_k \times M_t}\) are computed by

\[
[\Xi_{k,m,n}]^2_{ij} = [M_k]_{ij}^2 - \alpha_k^{2(n-1)} \frac{[M_k]_{ij}^4}{[M_k]_{ij}^2 + \sigma^2_{BS}}.
\]

(14)

Finally, we obtain the \textit{a posteriori} model of \(H_{k,m,n}\) given \(Y_{m,1}\) as

\[
H_{k,m,n} = \tilde{H}_{k,m,n} + U_k(\Xi_{k,m,n} \otimes W_{k,m,n})V^H_M
\]

(15)

where \(\tilde{H}_{k,m,n}\) is obtained from \((12)\) as

\[
\tilde{H}_{k,m,n} = \alpha_k^{n-1} U_k(\Delta_k \otimes U_k^H(X_{k,m,n}^{UE})^*(Y_{m,1}^{BS})^TV^*_M)V^H_M
\]

(16)

and \(W_{k,m,n}\) is a complex Gaussian random matrix with i.i.d., zero mean and unit variance entries. With \((15)\), the available imperfect CSI for each UE obtained at the BS is modeled as statistical CSI under a jointly correlated channel model with both channel mean and channel variance information, which includes the effects of channel estimation error, channel aging and spatial correlation. The \textit{a posteriori} model described in \((15)\) is a generic model for the available imperfect CSI obtained by the BS in the massive MIMO system under various mobile scenarios. When \(\alpha_k\) is very close to 1, it is suitable for the quasi-static scenario. When \(\alpha_k\) becomes very small, it is used to describe high speed scenario. Furthermore, in such case, \(\tilde{H}_{k,m,n}\) becomes close to zero, and the difference between the \textit{a posteriori} model in \((15)\) and \textit{a priori} model in \((2)\) also become very small. By setting the values of the \(\alpha_k\)s according to their moving speeds, we are able to describe the channel models of massive MIMO for various typical wireless communications.

\section*{B. Problem Formulation}

We now consider the downlink transmission for slot \(m\). Let \(x_{k,m,n}\) denote the \(M_k \times 1\) transmitted vector to the \(k\)-th UE at the \(n\)-th block of slot \(m\). The covariance matrix of \(x_{k,m,n}\) is the identity matrix \(I_{d_k}\). The received signal \(y_{k,m,n}\) at the \(k\)-th UE for a single symbol interval at the \(n\)-th block of slot \(m\) can be written as

\[
y_{k,m,n} = H_{k,m,n}P_{k,m,n}x_{k,m,n} + H_{k,m,n} \sum_{l \neq k} P_{l,m,n}x_{l,m,n} + z_{k,m,n}
\]

(17)

where \(P_{k,m,n}\) is the \(M_t \times d_k\) precoding matrix of the \(k\)-th UE, and \(z_{k,m,n}\) is a complex Gaussian noise vector distributed as \(CN(0, \sigma^2_{\text{BS}}I_{M_k})\).

We assume that the UEs obtain the perfect CSI of their corresponding channel matrices \(H_{k,m,n}P_{k,m,n}\) from the precoding domain training signals as in the BDMA transmission \([5]\). The DL training phase is included in the DL
data transmission and omitted in the slot structure for simplicity. At each UE, we treat the aggregate interference-plus-noise \( z'_{k,m,n} = H_{k,m,n} \sum_{l \neq k} P_{l,m,n} x_{l,m,n} + z_{k,m,n} \) as Gaussian noise. Let \( R_{k,m,n} \) denote the covariance matrix of \( z'_{k,m,n} \), we have that

\[
R_{k,m,n} = \sigma^2 I_k + \sum_{l \neq k} \mathbb{E}_{H_{k,m,n}} \left\{ H_{k,m,n} P_{l,m,n} P^H_{k,m,n} H^H_{k,m,n} \right\}
\]  

(18)

where the notation \( \mathbb{E}_{H_{k,m,n}} \{ \cdot \} \) denotes the expectation with respect to \( H_{k,m,n} \) according to the long-term statistics of the channel matrices at the user end. Owing to the channel reciprocity, the long-term channel statistics at the user end are the same as that of the BS, which have been provided in (15). Thus, the expectation \( \mathbb{E}_{H_{k,m,n}} \{ \cdot \} \) can be computed according to (2). We assume the covariance matrix \( R_{k,m,n} \) is known at the \( k \)-th UE. In such case, the expected rate of the \( k \)-th user at the \( n \)-th block of slot \( m \) is given by

\[
R_{k,m,n}(\sigma^2) = \mathbb{E}_{H_{k,m,n} \mid Y_{m,n}} \left\{ \log \det \left( R_{k,m,n} + H_{k,m,n} P_{k,m,n} P^H_{k,m,n} H^H_{k,m,n} \right) \right\} - \log \det \left( R_{k,m,n} \right)
\]

(19)

where the notation \( \mathbb{E}_{H_{k,m,n} \mid Y_{m,n}} \{ \cdot \} \) denotes the expectation with respect to \( H_{k,m,n} \) according to the a posteriori model obtained in (15). In this paper, we are interested in finding the precoding matrices \( P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n} \) that maximize the expected weighted sum-rate. The optimization problem can be formulated as

\[
\left( P^0_{1,m,n}, P^0_{2,m,n}, \cdots, P^0_{K,m,n} \right) = \arg \max_{P_{1,m,n}, \cdots, P_{K,m,n}} \sum_{k=1}^K \mathbb{E}_{H_{k,m,n}} \left\{ \log \det \left( R_{k,m,n}(\sigma^2) \right) \right\}
\]

(20)

s.t. \( \sum_{k=1}^K \text{tr} \left( P_{k,m,n} P^H_{k,m,n} \right) \leq P \)

where \( w_k \) is the weight of the rate of the \( k \)-th user and \( P \) denotes the total power budget.

### III. ROBUST LINEAR PRECODER DESIGN BASED ON DETERMINISTIC EQUIVALENTS

#### A. MM Algorithm for Precoder Design

The expected weighted sum-rate \( \sum_{k=1}^K w_k R_{k,m,n}(\sigma^2) \) is a very complicated function of the precoding matrices \( P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n} \), and thus also very difficult to be optimized directly. In the following, we use the MM algorithm to find a local optimum of the optimization problem (20).

Let \( f(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}) \) denote the objective function \( \sum_{k=1}^K w_k R_{k,m,n}(\sigma^2) \) in the optimization problem (20). Let \( P_{1,m,n}^d, P_{2,m,n}^d, \cdots, P_{K,m,n}^d \) be a fixed family of the precoding matrices and let

\[
g(P_{1,m,n}^d, P_{2,m,n}^d, \cdots, P_{K,m,n}^d) = \text{arg min}_{P_{1,m,n}, \cdots, P_{K,m,n}} f(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n})
\]

represent a real-valued continuous function of \( P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n} \) whose form depends on the fixed precoding matrices \( P_{1,m,n}^d, P_{2,m,n}^d, \cdots, P_{K,m,n}^d \). The function \( g \) is said to minorize \( f \) at \( P_{1,m,n}^d, P_{2,m,n}^d, \cdots, P_{K,m,n}^d \) provided that

\[
g(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}) \leq f(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n})
\]

(21)

\[
g(P_{1,m,n}^d, P_{2,m,n}^d, \cdots, P_{K,m,n}^d) = f(P_{1,m,n}^d, P_{2,m,n}^d, \cdots, P_{K,m,n}^d)
\]

(22)
When both the functions \( g \) and \( f \) are continuously differentiable with respect to \( \mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n} \), the conditions in (21) and (22) ensures
\[
\frac{\partial g}{\partial \mathbf{P}_{k,m,n}^{(d)}} \bigg|_{\mathbf{P}_{k,m,n} = \mathbf{P}_{k,m,n}^{(d)}} = \frac{\partial f}{\partial \mathbf{P}_{k,m,n}^{(d)}} \bigg|_{\mathbf{P}_{k,m,n} = \mathbf{P}_{k,m,n}^{(d)}} , \quad k = 1, \ldots, K. \tag{23}
\]

The key of the MM algorithms for the considered problem is to obtain a surrogate function which minorize the objective function at any point. When we find a good minorizing function, we will maximize it rather than the original function. Let \( \mathbf{P}_{1,m,n}^{(d+1)}, \mathbf{P}_{2,m,n}^{(d+1)}, \ldots, \mathbf{P}_{K,m,n}^{(d+1)} \) denote the maximizer of
\[
g(\mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n} | \mathbf{P}_{1,m,n}^{(d)}, \mathbf{P}_{2,m,n}^{(d)}, \ldots, \mathbf{P}_{K,m,n}^{(d)})
\]
under the constraint. From the conditions (21) and (22), we obtain
\[
f(\mathbf{P}_{1,m,n}^{(d+1)}, \mathbf{P}_{2,m,n}^{(d+1)}, \ldots, \mathbf{P}_{K,m,n}^{(d+1)}) \geq f(\mathbf{P}_{1,m,n}^{(d)}, \mathbf{P}_{2,m,n}^{(d)}, \ldots, \mathbf{P}_{K,m,n}^{(d)}). \tag{24}
\]
Let \( \mathbf{P}_{1,m,n}^{(d)}, \mathbf{P}_{2,m,n}^{(d)}, \ldots, \mathbf{P}_{K,m,n}^{(d)} \) be the current iterate and \( \mathbf{P}_{1,m,n}^{(d+1)}, \mathbf{P}_{2,m,n}^{(d+1)}, \ldots, \mathbf{P}_{K,m,n}^{(d+1)} \) be the next iterate in a search, we obtain a sequence of the precoding matrices. From (23) and (24), we observe that the sequence will converge to a local maximum of the original function \( f(\mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n}) \). The proof of the convergence depends on the conditions (21) and (22) and has been provided in the literature [29], [30]. Thus, we omit it here.

For simplicity, we define
\[
\eta_{k,m,n}^{pri} (\tilde{\mathbf{C}}) = \mathbb{E}_{\mathbf{H}_{k,m,n}} \{ \mathbf{H}_{k,m,n} \tilde{\mathbf{C}} \mathbf{H}_{k,m,n}^{H} \} \tag{25}
\]
\[
\eta_{k,m,n}^{pri} (\mathbf{C}) = \mathbb{E}_{\mathbf{H}_{k,m,n}} \{ \mathbf{H}_{k,m,n}^{H} \mathbf{C} \mathbf{H}_{k,m,n} \} \tag{26}
\]
where \( \tilde{\mathbf{C}} \in \mathbb{C}^{M_{k} \times M_{l}} \) and \( \mathbf{C} \in \mathbb{C}^{M_{k} \times M_{k}} \). Let \( \mathbf{R}_{k,m,n}^{(d)} \) denote
\[
\sigma_{z}^{2} \mathbf{I}_{M_{k}} + \sum_{l \neq k}^{K} \eta_{l,m,n}^{pri} (\mathbf{P}_{l,m,n}^{(d)} (\mathbf{P}_{l,m,n}^{(d)})^{H}) .
\]
Let \( \mathbf{E}_{k,m,n} \) and \( \mathbf{E}_{k,m,n}^{(d)} \) be defined as
\[
\mathbf{E}_{k,m,n} = \left( \mathbf{I}_{d_{k}} + \mathbf{P}_{k,m,n}^{H} \mathbf{H}_{k,m,n}^{H} \mathbf{R}_{k,m,n}^{-1} \mathbf{H}_{k,m,n} \mathbf{P}_{k,m,n} \right)^{-1} \tag{27}
\]
and
\[
\mathbf{E}_{k,m,n}^{(d)} = \left( \mathbf{I}_{d_{k}} + (\mathbf{P}_{k,m,n}^{(d)})^{H} \mathbf{H}_{k,m,n}^{H} (\mathbf{R}_{k,m,n}^{(d)})^{-1} \mathbf{H}_{k,m,n} \mathbf{P}_{k,m,n}^{(d)} \right)^{-1} . \tag{28}
\]

Then, the ergodic rate of the \( k \)-th user \( \mathcal{R}_{k,m,n}(\sigma_{z}^{2}) \) can be rewritten as
\[
\mathcal{R}_{k,m,n}(\sigma_{z}^{2}) = \mathbb{E} \left\{ \log \det \left( \mathbf{E}_{k,m,n}^{-1} \right) \right\} . \tag{29}
\]
We obtain a minorizing function of the objective function of the considered optimization problem in the following theorem.
Theorem 1. Let \( g_1(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)}) \) be a function defined as

\[
g_1(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)}) = \sum_{k=1}^{K} w_k c_{k,m,n}^{(d)} + \sum_{k=1}^{K} w_k \text{tr} \left( (A_{k,m,n}^{(d)})^H P_{k,m,n} \right) + \sum_{k=1}^{K} w_k \text{tr} \left( A_{k,m,n}^{(d)} P_{k,m,n}^H \right)
\]

where \( c_{k,m,n}^{(d)} \) is a constant defined as

\[
c_{k,m,n}^{(d)} = \mathbb{E}_{H_{k,m,n} | Y_{m,i}^{BS}} \left\{ \text{log det} \left( \left( E_{k,m,n}^{(d)} \right)^{-1} \right) \right\} + \text{tr} (I_{d_k}) - \mathbb{E}_{H_{k,m,n} | Y_{m,i}^{BS}} \left\{ \text{tr} \left( (E_{k,m,n}^{(d)})^{-1} \right) \right\}
- \sigma^2_{E} \mathbb{E}_{H_{k,m,n} | Y_{m,i}^{BS}} \left\{ \text{tr} \left( (P_{d,k,m,n})^H H_{k,m,n} (R_{k,m,n}^{(d)})^{-1} (R_{k,m,n}^{(d)})^{-1} H_{k,m,n} P_{k,m,n}^H E_{k,m,n} \right) \right\}
\]

(30)

and

\[
A_{k,m,n}^{(d)} = \mathbb{E}_{H_{k,m,n} | Y_{m,i}^{BS}} \left\{ H_{k,m,n} (R_{k,m,n}^{(d)})^{-1} H_{k,m,n} \right\} P_{k,m,n}^{(d)}
\]

(32)

\[
B_{k,m,n}^{(d)} = \mathbb{E}_{H_{k,m,n} | Y_{m,i}^{BS}} \left\{ H_{k,m,n} (R_{k,m,n}^{(d)})^{-1} H_{k,m,n} \right\}
\]

(33)

\[
C_{k,m,n}^{(d)} = N_{k,m,n}^{pri} \left( R_{k,m,n}^{(d)} \right)^{-1}
\]

(34)

\[
D_{k,m,n}^{(d)} = w_k B_{k,m,n}^{(d)} + \sum_{l \neq k} w_l C_{l,m,n}^{(d)}.
\]

(35)

Then, it is a minorizing function of \( f(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}) \) at \( P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)} \).

Proof: The proof is provided in Appendix A.

Using the minorizing function provided in Theorem 1, we update the precoding matrices sequence by

\[
P_{1,m,n}^{(d+1)}, P_{2,m,n}^{(d+1)}, \ldots, P_{K,m,n}^{(d+1)} = \arg \max_{P_{1,m,n}, \ldots, P_{K,m,n}} g_1(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)})
\]

\[
\text{s.t. } \sum_{k=1}^{K} \text{tr} \left( (P_{k,m,n}^H P_{k,m,n}) \right) \leq P.
\]

(36)

The limit point of the sequence provided by (36) is a local optimum of (20). The optimization problem in (36) is a concave quadratical optimization problem of \( P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n} \). Its optimal solution can be found by using the Lagrange multiplier methods. We define the Lagrangian as

\[
\mathcal{L}(\mu, P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n})
\]

\[
= -g_1(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)}) + \mu \left( \sum_{k=1}^{K} \text{tr} \left( (P_{k,m,n}^H P_{k,m,n}) \right) - P \right)
\]

\[
= -\sum_{k=1}^{K} w_k c_{k,m,n}^{(d)} - \sum_{k=1}^{K} w_k \text{tr} \left( (A_{k,m,n}^{(d)})^H P_{k,m,n} \right) - \sum_{k=1}^{K} w_k \text{tr} \left( A_{k,m,n}^{(d)} P_{k,m,n}^H \right)
\]

September 14, 2017

DRAFT
Equation (33) shows the first part of $B$. Robust Linear Precoder Design based on Deterministic Equivalents

From (32) and (41), we obtain

$$
\sum_{k=1}^{K} \text{tr} \left( D_{k,m,n}^{(d)} P_{k,m,n} P_{k,m,n}^{H} \right) + \mu \left( \sum_{k=1}^{K} \text{tr} \left( P_{k,m,n} P_{k,m,n}^{H} - P \right) \right)
$$

(37)

where $\mu \geq 0$ is the Lagrange multiplier. From the first order optimal conditions of (37), we obtain

$$
P_{k,m,n}^{(d+1)} = \left( D_{k,m,n}^{(d)} + \mu \ast I_{M_t} \right)^{-1} \left( w_{k} A_{k,m,n}^{(d)} \right).
$$

(38)

Similar to that in [10], the function $\sum_{k=1}^{K} P_{k,m,n} P_{k,m,n}^{H}$ is a monotonically decreasing function of $\mu$. Thus, if $\mu^* = 0$ and $\sum_{k=1}^{K} P_{k,m,n}^{(d+1)} (P_{k,m,n}^{(d+1)})^{H} \leq P$, we have obtained the optimal solution $P_{k,m,n}^{(d+1)} = \left( D_{k,m,n}^{(d)} \right)^{-1} w_{k} A_{k,m,n}^{(d)}$.

Otherwise, we can obtain the value of $\mu^*$ by using a bisection method.

To calculate the optimal solution in (38), we need to calculate $A_{k,m,n}^{(d)}$, $B_{k,m,n}^{(d)}$ and $C_{k,m,n}^{(d)}$ using (32), (33) and (34). Recall that the expectation $E_{H_{k,m,n} | Y_{BS_{m,1}}}$ used in (32), (33) and (34) is defined according to the a posteriori model of $H_{k,m,n}$, i.e., $H_{k,m,n} = H_{k,m,n}^{*} + U_{k} (\Xi_{k,m,n} \otimes W_{k,m,n}) V_{Y_{M_t}}^{H}$. Let $\tilde{H}_{k,m,n}$ denote $U_{k} (\Xi_{k,m,n} \otimes W_{k,m,n}) V_{Y_{M_t}}^{H}$. We define

$$
\eta_{k,m,n}^{\ast, post} (\tilde{C}) = E_{\tilde{H}_{k,m,n}} \left\{ \tilde{H}_{k,m,n} \tilde{C} H_{k,m,n}^{H} \right\}
$$

(39)

$$
\tilde{\eta}_{k,m,n}^{\ast, post} (\tilde{C}) = E_{\tilde{H}_{k,m,n}} \left\{ \tilde{H}_{k,m,n}^{H} \tilde{C} H_{k,m,n} \right\}.
$$

(40)

Then, we obtain

$$
E_{H_{k,m,n} | Y_{BS_{m,1}}} \left\{ H_{k,m,n} \tilde{C} H_{k,m,n}^{H} \right\} = \tilde{H}_{k,m,n} \tilde{C} H_{k,m,n}^{H} + \eta_{k,m,n}^{\ast, post} (\tilde{C})
$$

(41)

$$
E_{H_{k,m,n} | Y_{BS_{m,1}}} \left\{ H_{k,m,n}^{H} \tilde{C} H_{k,m,n} \right\} = \tilde{H}_{k,m,n}^{H} \tilde{C} H_{k,m,n} + \tilde{\eta}_{k,m,n}^{\ast, post} (\tilde{C}).
$$

(42)

From (32) and (41), we obtain $A_{k,m,n}^{(d)}$ as

$$
A_{k,m,n}^{(d)} = \tilde{H}_{k,m,n}^{H} \left( R_{k,m,n}^{(d)} \right)^{-1} \tilde{H}_{k,m,n} P_{k,m,n}^{(d)} + \eta_{k,m,n}^{\ast, post} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) P_{k,m,n}^{(d)}.
$$

(43)

Equation (33) shows the first part of $B_{k,m,n}^{(d)}$ can be obtained similarly as $A_{k,m,n}^{(d)}$. However, the second part of $B_{k,m,n}^{(d)}$ is very complicated and it is hardly possible to obtain a closed-form expression. Similarly, the computation of $C_{k,m,n}^{(d)}$ has also no closed-form expression. In the next subsection, we will provide the approximations of $B_{k,m,n}^{(d)}$ and $C_{k,m,n}^{(d)}$ by using the deterministic equivalent method.

B. Robust Linear Precoder Design based on Deterministic Equivalents

Observing (33) and (34), we find that $B_{k,m,n}^{(d)}$ and $C_{k,m,n}^{(d)}$ are closely related to the derivatives of $R_{k,m,n}^{2}$ with respect to $P_{k,m,n} P_{k,m,n}^{H}$ and $P_{l,m,n} P_{l,m,n}^{H}$, $l \neq k$. Thus, to derive the deterministic equivalents of $B_{k,m,n}^{(d)}$ and $C_{k,m,n}^{(d)}$, we begin from the deterministic equivalent of $R_{k,m,n}^{2}$. The channel model provided in (15) is a jointly correlated channel model with a nonzero mean. For such model, the deterministic equivalent of $R_{k,m,n}^{2}$ has been provided in [21] and [31]. Using the results from [21], we obtain the following Lemma.

**Lemma 1.** The deterministic equivalent of $R_{k,m,n}^{2}$ is

$$
\overline{R}_{k,m,n}^{2} = \log \det \left( \Phi_{k,m,n} \right) + \log \det \left( \Phi_{k,m,n} \right) - \text{tr} \left( \eta_{k,m,n}^{\ast, post} \left( P_{k,m,n} G_{k,m,n} P_{k,m,n}^{H} \right) R_{k,m,n}^{-1/2} \left( \Phi_{k,m,n} \right) R_{k,m,n}^{-1/2} \right)
$$

(44)
where $\Phi_{k,m,n}$, $\tilde{\Phi}_{k,m,n}$, $\Gamma_{k,m,n}$, $\tilde{\Gamma}_{k,m,n}$, $G_{k,m,n}$ and $\tilde{G}_{k,m,n}$ are obtained by the iterative equations:

\[
\Phi_{k,m,n} = I_{d_k} + P_{k,m,n}^{H} \eta_{k,m,n}^{post} \left( R_{k,m,n}^{-1/2} \tilde{G}_{k,m,n} R_{k,m,n}^{-1/2} \right) P_{k,m,n} \tag{46}
\]

\[
\tilde{\Phi}_{k,m,n} = I_{M_k} + R_{k,m,n}^{-1/2} \eta_{k,m,n}^{post} \left( R_{k,m,n}^{-1/2} \tilde{G}_{k,m,n} R_{k,m,n}^{-1/2} \right) \tilde{P}_{k,m,n} \tag{47}
\]

\[
\Gamma_{k,m,n} = \tilde{\eta}_{k,m,n}^{post} \left( R_{k,m,n}^{-1/2} \tilde{G}_{k,m,n} R_{k,m,n}^{-1/2} \right) + \tilde{H}_{k,m,n}^{H} R_{k,m,n}^{-1/2} \tilde{\Phi}_{k,m,n} R_{k,m,n}^{-1/2} \tilde{H}_{k,m,n} \tag{48}
\]

\[
\tilde{\Gamma}_{k,m,n} = \eta_{k,m,n}^{post} \left( P_{k,m,n} G_{k,m,n} P_{k,m,n}^{H} \right) + \tilde{H}_{k,m,n} P_{k,m,n} \Phi_{k,m,n}^{-1} P_{k,m,n}^{H} \tilde{H}_{k,m,n}^{H} \tag{49}
\]

\[
G_{k,m,n} = \left( I_{d_k} + P_{k,m,n}^{H} \Gamma_{k,m,n} P_{k,m,n} \right)^{-1} \tag{50}
\]

\[
\tilde{G}_{k,m,n} = \left( I_{M_k} + R_{k,m,n}^{-1/2} \tilde{\Gamma}_{k,m,n} R_{k,m,n}^{-1/2} \right)^{-1} \tag{51}
\]

**Proof:** The Lemma is a corollary of Theorem 3 in [21]. Thus, the proof is omitted here for brevity. ■

Lemma 1 provides two equivalent deterministic equivalents of $R_{k,m,n}(\sigma^2)$ in (44) and (45). From the two deterministic equivalents, we can obtain the derivatives of $\mathcal{R}_{k,m,n}(\sigma^2)$ with respect to $P_{k,m,n} P_{k,m,n}^{H}$ and $P_{l,m,n} P_{l,m,n}^{H}$, $l \neq k$, respectively. With the obtained derivatives, we then obtain the deterministic equivalents of $B^{(d)}_{k,m,n}$ and $C^{(d)}_{k,m,n}$ in the following theorem.

**Theorem 2.** The deterministic equivalents of $B^{(d)}_{k,m,n}$ and $C^{(d)}_{k,m,n}$ are

\[
B^{(d)}_{k,m,n} = \tilde{H}_{k,m,n}^{H} \left( R_{k,m,n}^{(d)} \right)^{-1} \tilde{H}_{k,m,n} + \eta_{k,m,n}^{post} \left( R_{k,m,n}^{(d)} \right)^{-1} \Gamma_{k,m,n} \tag{52}
\]

and

\[
C^{(d)}_{k,m,n} = \tilde{\eta}_{k,m,n}^{pri} \left( R_{k,m,n}^{(d)} \right)^{-1} - \eta_{k,m,n}^{pri} \left( R_{k,m,n}^{(d)} + \tilde{\Gamma}_{k,m,n} \right)^{-1} \tag{53}
\]

**Proof:** The proof is provided in Appendix B. ■

With the deterministic equivalents of $B^{(d)}_{k,m,n}$ and $C^{(d)}_{k,m,n}$ provided in Theorem 2, the update step in (38) using the minorizing function $g_1$ becomes

\[
P_{k,m,n}^{(d+1)} = \left( D^{(d)}_{k,m,n} + \mu I_{M_k} \right)^{-1} (w_k A^{(d)}_{k,m,n}) \tag{54}
\]

where

\[
D^{(d)}_{k,m,n} = w_k B^{(d)}_{k,m,n} + \sum_{l \neq k} w_l C^{(d)}_{l,m,n}. \tag{55}
\]

We now present an algorithm for the design of the robust linear precoder using the minorizing function $g_1$.

**Algorithm 1:** Robust linear precoder design using the minorizing function $g_1$
Step 1: Set \( d = 0 \). Randomly generate \( \mathbf{P}^{(d)}_{1,m,n}, \mathbf{P}^{(d)}_{2,m,n}, \ldots, \mathbf{P}^{(d)}_{K,m,n} \) and normalize them to satisfy the constraint 
\[
\sum_{k=1}^{K} \text{tr} \left( \mathbf{P}^{(d)}_{k,m,n} (\mathbf{P}^{(d)}_{k,m,n})^H \right) = P.
\]

Step 2: Calculate \( \mathbf{R}^{(d)}_{k,m,n} \) according to
\[
\mathbf{R}^{(d)}_{k,m,n} = \sigma^2 \mathbf{I}_{M_k} + \sum_{l \neq k}^{K} \eta^{\text{pri}}_{l,m,n} \left( (\mathbf{P}^{(d)}_{l,m,n})^H \mathbf{P}^{(d)}_{l,m,n} \right).
\]

Step 3: Calculate \( \Gamma_{k,m,n} \) and \( \tilde{\Gamma}_{k,m,n} \) according to Lemma 1.

Step 4: Compute \( \mathbf{A}^{(d)}_{k,m,n}, \mathbf{B}^{(d)}_{k,m,n} \) and \( \mathbf{C}^{(d)}_{k,m,n} \) according to
\[
\begin{align*}
\mathbf{A}^{(d)}_{k,m,n} &= \mathbf{H}^{H}_{k,m,n} \left( \mathbf{R}^{(d)}_{k,m,n} \right)^{-1} \tilde{\mathbf{H}}_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} + \tilde{\eta}^{\text{post}}_{k,m,n} \left( \left( \mathbf{R}^{(d)}_{k,m,n} \right)^{-1} \right) \mathbf{P}^{(d)}_{k,m,n} \\
\mathbf{B}^{(d)}_{k,m,n} &= \mathbf{H}^{H}_{k,m,n} \left( \mathbf{R}^{(d)}_{k,m,n} \right)^{-1} \tilde{\mathbf{H}}_{k,m,n} + \tilde{\eta}^{\text{post}}_{k,m,n} \left( \left( \mathbf{R}^{(d)}_{k,m,n} \right)^{-1} \right) \\
&\quad - \left( \mathbf{I}_{M_t} + \Gamma_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} (\mathbf{P}^{(d)}_{k,m,n})^H \right)^{-1} \Gamma_{k,m,n} \\
\mathbf{C}^{(d)}_{k,m,n} &= \tilde{\eta}^{\text{pri}}_{k,m,n} \left( \left( \mathbf{R}^{(d)}_{k,m,n} \right)^{-1} \right) - \tilde{\eta}^{\text{pri}}_{k,m,n} \left( \left( \mathbf{R}^{(d)}_{k,m,n} + \tilde{\Gamma}_{k,m,n} \right)^{-1} \right).
\end{align*}
\]

Step 5: Update \( \mathbf{P}^{(d+1)}_{k,m,n} \) by
\[
\mathbf{P}^{(d+1)}_{k,m,n} = \left( \mathbf{D}^{(d)}_{k,m,n} + \mu^* \mathbf{I}_{M_t} \right)^{-1} (w_k \mathbf{A}^{(d)}_{k,m,n})
\]

where
\[
\mathbf{D}^{(d)}_{k,m,n} = w_k \mathbf{B}^{(d)}_{k,m,n} + \sum_{l \neq k}^{K} w_l \mathbf{C}^{(d)}_{l,m,n}.
\]

Set \( d = d + 1 \).

Repeat Step 2 through Step 5 until convergence or until a pre-set target is reached.

For very large \( M_t \), the computational complexity of Algorithm 1 are dominated by the number of \( M_t \times M_t \) matrix inversions. Observing Algorithm 1, we find there are an \( M_t \times M_t \) inversion \( \left( \mathbf{I}_{M_t} + \Gamma_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} (\mathbf{P}^{(d)}_{k,m,n})^H \right)^{-1} \) in each computation of \( \mathbf{B}^{(d)}_{k,m,n} \), and an \( M_t \times M_t \) inversion \( \left( \mathbf{D}^{(d)}_{k,m,n} + \mu^* \mathbf{I}_{M_t} \right)^{-1} \) in each computation of \( \mathbf{P}^{(d+1)}_{k,m,n} \). Thus, there are total \( 2K M_t \times M_t \) matrix inversions per iteration.

C. Low-Complexity Robust Linear Precoder Designs

In this subsection, we introduce two low-complexity algorithms for robust linear precoder designs. The first algorithm is based on an alternative minorizing function modified from \( g_1 \). The second algorithm is designed for the case when \( \mathbf{H}_{k,m,n} = 0 \).

We begin with the first low-complexity algorithm. As shown in the previous subsection, the computational complexity of Algorithm 1 per iteration are dominated by \( 2K \) large dimensional matrix inversions. The first \( K \) large dimensional matrix inversions in Algorithm 1 can be avoid by rewriting them as
\[
\begin{align*}
\left( \mathbf{I}_{M_t} + \Gamma_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} (\mathbf{P}^{(d)}_{k,m,n})^H \right)^{-1} & \Gamma_{k,m,n} \\
= \Gamma_{k,m,n} & - \Gamma_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} (\mathbf{P}^{(d)}_{k,m,n})^H \left( \mathbf{I}_{M_t} + \Gamma_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} (\mathbf{P}^{(d)}_{k,m,n})^H \right)^{-1} \Gamma_{k,m,n} \\
= \Gamma_{k,m,n} & - \Gamma_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} \left( \mathbf{I}_{d_k} + (\mathbf{P}^{(d)}_{k,m,n})^H \mathbf{\Gamma}_{k,m,n} \mathbf{P}^{(d)}_{k,m,n} \right)^{-1} (\mathbf{P}^{(d)}_{k,m,n})^H \mathbf{\Gamma}_{k,m,n}
\end{align*}
\] (56)
where the second equality is due to the matrix inversion lemma. The second \( K \times M_t \) matrix inversions \((D_{k,m,n}^{(d)} + \mu^* M_t)^{-1}\) can be reduced to one matrix inversion. For this purpose, we provide the following theorem which presents an alternative minorizing function modified from the minorizing function \( g_1 \).

**Theorem 3.** Let \( g_2(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}, P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)}) \) be a function defined as

\[
g_2(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}, P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)}) = c_{m,n}^{(d)} + \sum_{k=1}^{K} \text{tr} \left( w_k A_{k,m,n}^{(d)} + F_{k,m,n}^{(d)} \right) + \sum_{k=1}^{K} \text{tr} \left( w_k A_{k,m,n}^{(d)} + F_{k,m,n}^{(d)} P_{k,m,n}^{(d)} \right)
\]

\[
- \sum_{k=1}^{K} \text{tr} \left( (D_{k,m,n}^{(d)} + F_{k,m,n}^{(d)}) P_{k,m,n}^{(d)} B_{k,m,n}^{(d)} \right).
\]

where \( F_{k,m,n}^{(d)} \) is a positive semidefinite matrix and

\[
c_{m,n}^{(d)} = \sum_{k=1}^{K} w_k c_{k,m,n}^{(d)} - \sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} P_{k,m,n}^{(d)} (P_{k,m,n}^{(d)})^H \right).
\]

Then, it is also a minorizing function of \( f(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}) \) at \( P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \ldots, P_{K,m,n}^{(d)} \).

**Proof:** The proof is provided in Appendix [C].

The process of using the minorizing function \( g_2 \) obtained in Theorem 3 to obtain a local optimum is similar to that of using \( g_1 \). For brevity, we omit the details and give the solution directly as

\[
P_{k,m,n}^{(d+1)} = ((D_{k,m,n}^{(d)} + F_{k,m,n}^{(d)}) + \mu^* M_t)^{-1}(w_k A_{k,m,n}^{(d)} + F_{k,m,n}^{(d)} P_{k,m,n}^{(d)}).
\]

The minorizing function \( g_2 \) can be used to reduce the complexity of the solutions of the surrogate optimization problem. Let \( F_{k,m,n}^{(d)} \) be defined as

\[
F_{k,m,n}^{(d)} = w_k C_{k,m,n}^{(d)} + \sum_{l \neq k} w_l B_{l,m,n}^{(d)}.
\]

Recall that \( D_{k,m,n}^{(d)} = w_k B_{k,m,n}^{(d)} + \sum_{l \neq k} w_l C_{l,m,n}^{(d)} \), we obtain

\[
D_{k,m,n}^{(d)} + F_{k,m,n}^{(d)} = \sum_{k=1}^{K} w_k (B_{k,m,n}^{(d)} + C_{k,m,n}^{(d)})
\]

which is the same for all \( k \). Thus, the \( M_t \times M_t \) matrix inversion \((D_{k,m,n}^{(d)} + F_{k,m,n}^{(d)}) + \mu^* M_t)^{-1}\) only need to be done once per iteration. For brevity, we define

\[
D_{m,n}^{(d)} = \sum_{k=1}^{K} w_k (B_{k,m,n}^{(d)} + C_{k,m,n}^{(d)}).
\]

From (52) and (53), we obtain the deterministic equivalents of \( D_{m,n}^{(d)} \) and \( F_{k,m,n}^{(d)} \) as

\[
\overline{D}_{m,n}^{(d)} = \sum_{k=1}^{K} w_k (\overline{B}_{k,m,n}^{(d)} + \overline{C}_{k,m,n}^{(d)})
\]

\[
\overline{F}_{k,m,n}^{(d)} = w_k C_{k,m,n}^{(d)} + \sum_{l \neq k} w_l B_{l,m,n}^{(d)}.
\]
We now present an algorithm for the design of the robust linear precoder using the minorizing function $g_2$.

**Algorithm 2: Robust linear precoder design using the minorizing function $g_2$**

1. **Step 1:** Set $d = 0$. Randomly generate $P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)}$ and normalize them to satisfy the constraint $\sum_{k=1}^{K} \text{tr} \left( P_{k,m,n}^{(d)} (P_{k,m,n}^{(d) \dagger}) \right) = P$.

2. **Step 2:** Calculate $R_{k,m,n}^{(d)}$ according to

   $$ R_{k,m,n}^{(d)} = \sigma_s^2 I_{M_k} + \sum_{l \neq k}^{K} \eta_{l,m,n} \left( P_{l,m,n}^{(d)} (P_{l,m,n}^{(d) \dagger}) \right). $$

3. **Step 3:** Calculate $\Gamma_{k,m,n}$ and $\tilde{\Gamma}_{k,m,n}$ according to Lemma 1.

4. **Step 4:** Compute $A_{k,m,n}^{(d)}, B_{k,m,n}^{(d)}, C_{k,m,n}^{(d)}$ and $F_{k,m,n}^{(d)}$ according to

   $$ A_{k,m,n}^{(d)} = \tilde{H}_{k,m,n}^{H} \left( R_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} P_{k,m,n}^{(d)} + \eta_{k,m,n}^{post} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) P_{k,m,n}^{(d)} $$

   $$ B_{k,m,n}^{(d)} = \tilde{H}_{k,m,n}^{H} \left( R_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} + \eta_{k,m,n}^{post} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) - \Gamma_{k,m,n} + \tilde{\Gamma}_{k,m,n} \left( I_{d_k} + (P_{k,m,n}^{(d)})^{H} \Gamma_{k,m,n} P_{k,m,n}^{(d)} \right)^{-1} (P_{k,m,n}^{(d)})^{H} \Gamma_{k,m,n} $$

   $$ C_{k,m,n}^{(d)} = \eta_{k,m,n}^{pri} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) - \eta_{k,m,n}^{pri} \left( \left( R_{k,m,n}^{(d)} + \tilde{\Gamma}_{k,m,n} \right)^{-1} \right) $$

   $$ F_{k,m,n}^{(d)} = w_k C_{k,m,n}^{(d)} + \sum_{l \neq k}^{K} w_l B_{l,m,n}^{(d)}. $$

5. **Step 5:** Update $P_{k,m,n}^{(d+1)}$ by

   $$ P_{k,m,n}^{(d+1)} = \left( D_{m,n}^{(d)} + \mu^{*} I_{M_t} \right)^{-1} (w_k A_{k,m,n}^{(d)} + F_{k,m,n}^{(d)} P_{k,m,n}^{(d)}) $$

   where

   $$ D_{m,n}^{(d)} = \sum_{k=1}^{K} w_k (B_{k,m,n}^{(d)} + C_{k,m,n}^{(d)}). $$

Set $d = d + 1$.

Repeat Step 2 through Step 5 until convergence or until a pre-set target is reached.

For Algorithm 2, there only need one $M_t \times M_t$ matrix inversion per iteration. Thus, the complexity of Algorithm 2 is reduced compared with that of Algorithm 1 for very large $M_t$.

In the following, we introduce another low-complexity algorithm for a special case. When the *a posteriori* channel models $H_{k,m,n} = \hat{H}_{k,m,n} + U_k (\xi_{k,m,n} \otimes W_{k,m,n})V_{M_t}^{H}$ have zero means, *i.e.*, $H_{k,m,n} = 0$, they will reduce to the *a priori* channel models $H_{k,m,n} = U_k (M_k \otimes W_{k,m,n})V_{M_t}^{H}$. In such case, we obtain

$$ A_{k,m,n}^{(d)} = \eta_{k,m,n}^{pri} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) P_{k,m,n}^{(d)} $$

$$ B_{k,m,n}^{(d)} = \eta_{k,m,n}^{pri} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) - \Gamma_{k,m,n} $$

$$ + \Gamma_{k,m,n} P_{k,m,n}^{(d)} \left( I_{d_k} + (P_{k,m,n}^{(d)})^{H} \Gamma_{k,m,n} P_{k,m,n}^{(d)} \right)^{-1} (P_{k,m,n}^{(d)})^{H} \Gamma_{k,m,n}. $$

$$ \text{September 14, 2017} \quad \text{DRAFT} $$
Furthermore, the computations of $\Gamma_{k,m,n}$ and $\tilde{\Gamma}_{k,m,n}$ become

$$\Gamma_{k,m,n} = \eta_{k,m,n}^{pri} \left( R_{k,m,n}^{-1/2} \hat{\Gamma}_{k,m,n} R_{k,m,n}^{-1/2} \right)$$  \hspace{1cm} (67)$$

and

$$\tilde{\Gamma}_{k,m,n} = \eta_{k,m,n}^{pri} \left( P_{k,m,n} \tilde{\Gamma}_{k,m,n} H_{k,m,n} \right)$$  \hspace{1cm} (68)$$

and $\tilde{\Phi}_{k,m,n}, \Phi_{k,m,n}, \tilde{G}_{k,m,n}$ and $\tilde{G}_{k,m,n}$ are now obtained by the iterative equations

$$\tilde{\Phi}_{k,m,n} = \eta_{k,m,n}^{pri} \left( \left( \Phi_{k,m,n}^d + \tilde{\Gamma}_{k,m,n} \right)^{-1} \right)$$  \hspace{1cm} (69)$$

$$\Phi_{k,m,n} = \left( \tilde{\Phi}_{k,m,n} \right)^{-1}$$  \hspace{1cm} (70)$$

Let $\Lambda_{k,m,n}(\tilde{C})$ and $\tilde{\Lambda}_{k,m,n}(C)$ be two diagonal matrix valued functions defined as

$$\left[ \Lambda_{k,m,n}(\tilde{C}) \right]_{ii} = \sum_{j=1}^{M_k} \left[ V_{M_k} H_{k,m,n} \tilde{C} V_{M_k} \right]_{jj}$$  \hspace{1cm} (75)$$

$$\left[ \tilde{\Lambda}_{k,m,n}(C) \right]_{ii} = \sum_{j=1}^{M_k} \left[ U_{k}^H C U_{k} \right]_{jj}$$  \hspace{1cm} (76)$$

Then, we obtain

$$\eta_{k,m,n}^{pri}(\tilde{C}) = U_k \Lambda_{k,m,n}(\tilde{C}) U_k^H$$  \hspace{1cm} (77)$$

$$\eta_{k,m,n}^{pri}(C) = V_{M_k} \tilde{\Lambda}_{k,m,n}(C) V_{M_k}^H$$  \hspace{1cm} (78)$$

In such case, we observe from Algorithms 1 and 2 that once the left singular vector matrix of $P_{k,m,n}$ is $V_{M_k}$ right multiplying a permutation matrix, it will remain the same forever. Thus, we obtain that $V_{M_k}$ right multiplying a permutation matrix must be the left singular vector matrix for $P_{k,m,n}$ at certain stationary points. In the following, we will show that this conclusion actually holds for all stationary points.

Let $f(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n})$ denote $K \sum_{k=1}^{K} w_k \overline{r}_{k,m,n}(\sigma_z^2)$. From Lemma 1, we obtain $\overline{f}$ is the deterministic equivalent of $f$. According to (73) and (78), $\Gamma_{k,m,n}$ can be written as

$$\Gamma_{k,m,n} = V_{M_k} \Sigma_{k,m,n}^2 V_{M_k}^H$$  \hspace{1cm} (79)$$

where $\Sigma_{k,m,n}^2$ is a diagonal matrix whose value depends on $P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}$. Then, we obtain the following theorem.
Theorem 4. Assume $\hat{H}_{k,m,n} = 0$. Then, the left singular vector matrix of the linear precoders at the stationary points of the optimization problem

$$
\max_{P_1, m,n, \ldots, P_{K,m,n}} \mathcal{J}(P_1, m,n, P_2, m,n, \ldots, P_{K,m,n})
$$

\[\text{s.t. } \sum_{k=1}^{K} \text{tr}(P_{k,m,n}P_{k,m,n}^H) \leq P \] (80)

can be written as

$$
U_{P_{k,m,n}} = V_{M_t, \Pi_{k,m,n}}
$$

where $\Pi_{k,m,n}$ is a permutation matrix.

Proof: The proof is provided in Appendix D.

Theorem 4 proves the optimality of beam domain transmission when $\hat{H}_{k,m,n} = 0$ and the objective function of the optimization problem (20) is replaced by its deterministic equivalent. Using Theorem 4 we obtain that the optimal precoders can be written as

$$
P_{k,m,n} = V_{M_t, \Pi_{k,m,n}} J_{k,m,n} V_{k,m,n}^H
$$

where $J_{k,m,n}$ are $M_t \times d_k$ matrices with nonzero elements on the main diagonal and zeros elsewhere, and $V_{k,m,n}^H$ are any $d_k \times d_k$ unitary matrices. Since $V_{k,m,n}^H$ has no impact on the expected weighted sum-rate, it can be set to a fixed unitary matrix. For brevity, we set $V_{k,m,n}^H = I_{d_k}$. Then, the optimal precoders can be rewritten as

$$
P_{k,m,n} = V_{M_t, \Pi_{k,m,n}} J_{k,m,n}.
$$

(83)

To achieve an algorithm with a complexity lower than Algorithm 2, we also set each $\Pi_{k,m,n}$ to a fixed permutation matrix. Let $a_k$ be an $M_t \times 1$ row vector defined as

$$
[a_k]_j = \sum_{i=1}^{M_t} [\Omega_{k}]_{ij}.
$$

(84)

The permutation matrix $\Pi_{k,m,n}$ is set to make the elements in $a_k \Pi_{k,m,n}$ are of descend order. Then, we only need to optimize $J_{k,m,n}$. Substituting the optimal structures of the precoders with fixed permutation matrices and the condition $\hat{H}_{k,m,n} = 0$ into Algorithm 2, we obtain the following new algorithm with lower complexity.

---

Algorithm 3: Robust linear precoder design using the minorizing function $g_2$ when $\hat{H}_{k,m,n} = 0$

Step 1: Set $d = 0$. Initialize all the $J_{k,m,n}^{(d)}$ with ones along the main diagonal and zeros elsewhere, and normalize them to satisfy the constraint $\sum_{k=1}^{K} \text{tr}\left( J_{k,m,n}^{(d)}(J_{k,m,n}^{(d)})^H \right) = P$.

Step 2: Calculate $R_{k,m,n}^{(d)}$ according to

$$
R_{k,m,n}^{(d)} = a_{2}^{2} I_{M_t} + \sum_{l \neq k}^{K} \text{tr}(V_{M_t, \Pi_{l,m,n}} J_{l,m,n}^{(d)}(J_{l,m,n}^{(d)})^H \Pi_{l,m,n} V_{M_t}^H).
$$

Step 3: Calculate $\Gamma_{k,m,n}$ and $\tilde{\Gamma}_{k,m,n}$ according to (67) and (68).
Step 4: Compute $A_{k,m,n}^{(d)}$, $A_{B,k,m,n}^{(d)}$, $A_{C,k,m,n}^{(d)}$ and $A_{F,k,m,n}^{(d)}$ according to

$$A_{k,m,n}^{(d)} = \hat{A}_{k,m,n} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) J_{k,m,n}^{(d)}$$

$$A_{B,k,m,n}^{(d)} = \hat{A}_{k,m,n} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) - \Sigma_{k,m,n}^2 + \Sigma_{k,m,n}^2 \Pi_{k,m,n} J_{k,m,n}^{(d)}$$

$$A_{C,k,m,n}^{(d)} = \hat{A}_{k,m,n} \left( \left( R_{k,m,n}^{(d)} \right)^{-1} \right) - \Sigma_{k,m,n}^2$$

$$A_{F,k,m,n}^{(d)} = w_k A_{C,k,m,n}^{(d)} + \sum_{l \neq k} w_l A_{F,l,m,n}^{(d)}.$$

Step 5: Update $J_{k,m,n}^{(d+1)}$ by

$$J_{k,m,n}^{(d+1)} = (A_{D,m,n}^{(d)} + \mu^* M_t)^{-1} \left( w_k A_{C,k,m,n}^{(d)} + F_{k,m,n} J_{k,m,n}^{(d)} \right)$$

where

$$A_{D,m,n}^{(d)} = \sum_{k=1}^{K} w_k \left( A_{B,k,m,n}^{(d)} + A_{C,k,m,n}^{(d)} \right).$$

Set $d = d + 1$.

Repeat Step 2 through Step 5 until convergence or until a pre-set target is reached. Then the optimal precoders are obtained as $P^*_k,m,n = V_{M_t} \Pi_{k,m,n} J^*_k,m,n$.

In Algorithm 3, we have used the commutativity of the permutation matrices with the diagonal matrices to simplify the formulas. Since $A_{D,m,n}^{(d)}$ is a diagonal matrix, the $M_t \times M_t$ matrix inversion $(A_{D,m,n}^{(d)} + \mu^* M_t)^{-1}$ in Algorithm 3 can be implemented element-wisely. Thus, the computational complexity are further reduced when $\hat{H}_{k,m,n} = 0$.

IV. SIMULATION RESULTS

In this section, we provide simulation results to show the performance of the proposed algorithms. We use the 3GPP stochastic channel model (SCM) \cite{cc} to generate $U_k$ and the CCM $\Omega_k$. The antenna arrays used at the BS and the UEs are both ULAs with 0.5λ spacing. The shadow fading and path loss are not considered. The scenario used to generate $U_k$ and $\Omega_k$ is “urban_metro”. The users in the cell are random uniformly distributed. In all simulations, we set $P = 1$, $w_k = 1$, $d_k = M_k$ and $N_t = 7$. The $M_k$ for all the users are set to be the same. For simplicity, we set $\sigma_{BS}^2 = \sigma^2$. The signal-to-noise ratio (SNR) is given by $\text{SNR} = \frac{1}{\sigma^2}$.

We first investigate the performance of the three proposed algorithms. We consider a massive MIMO downlink system with $M_t = 128$, $M_k = 4$ and $K = 10$. The values of $\alpha_k$s used in the simulations are presented in Table \[1\] It indicates that the channels of the first 5 users are quasi-static, and that the other 5 users move slowly. Let $N_s$ denote the number of time slots used in the simulations. Fig.2\[2\] shows the simulation results of the average sum-rate performance of the three algorithms for this massive MIMO downlink over $N_s = 100$ time slots. From Fig.2\[2\] we see that the average sum-rates of the three algorithms increase almost linearly as the SNR increases. Furthermore,
TABLE I
THE VALUES OF $\alpha_k$s IN SCENARIO 1.

| $\alpha_1 - \alpha_2$ | $\alpha_6 - \alpha_{10}$ |
|-----------------------|--------------------------|
| 0.999                 | 0.9                      |

Fig. 2. Average sum-rate performance of the three proposed algorithms for a massive MIMO downlink with $M_t = 128$, $M_k = 4$, $K = 10$ and the $\alpha$s presented in Table I.

we see that the differences between the performance of Algorithms 1 and 2 are negligible. The performance gaps between Algorithm 3 and that of Algorithms 1 and 2 increase as the SNR increases. At SNR=20 dB, the performance loss of Algorithm 3 is about 25 percent. This is because Algorithm 3 is designed for the case when the channel means of all users equal zeros, and thus does not exploit the full benefits of the available statistical CSI at the BS. We also present in Fig. 2 the deterministic equivalents of the average sum-rate to show the accuracy of the deterministic equivalents. The deterministic equivalent results of Algorithms 1 and 2 are very accurate. For Algorithm 3, small deviations of the deterministic equivalent results from the simulation results are observed. This is also because the mismatch of the statistical CSI that Algorithm 3 used with the actual available statistical CSI at the BS.

To investigate the performance of the three proposed algorithms when high speed users exist. We keep $M_t = 128$, $M_k = 4$, $K = 10$ and $N_s = 100$, but change the values of $\alpha_k$s to those presented in Table II. Fig. 3 show the simulation results of the average sum-rate performance of the three algorithms for this scenario. From Fig. 3 we see that the average sum-rates of the three algorithms are still increase almost linearly as the SNR increases, and that the differences between the performance of Algorithms 1 and 2 are also negligible. Furthermore, the performance gaps between Algorithm 3 and that of Algorithms 1 and 2 are become smaller. At SNR=20 dB, the performance loss of Algorithm 3 is less than 15 percent. This is because the difference between the statistical CSI that Algorithm 3 used and the available statistical CSI at the BS become smaller at this scenario. The deterministic equivalents of the
TABLE II
THE VALUES OF $\alpha_k$s IN SCENARIO 2.

| $\alpha_1$, $\alpha_2$ | $\alpha_3$, $\alpha_4$ | $\alpha_5$, $\alpha_6$ | $\alpha_7$, $\alpha_8$ | $\alpha_9$, $\alpha_{10}$ |
|------------------------|------------------------|------------------------|------------------------|------------------------|
| 0.999                  | 0.9                    | 0.5                    | 0.1                    | 0                      |

Fig. 3. Average sum-rate performance of the three proposed algorithms for a massive MIMO downlink with $M_t = 128$, $M_k = 4$, $K = 10$ and the $\alpha$s presented in Table II.

average sum-rates are also presented in Fig. 3 to show their accuracy. In this scenario, the deterministic equivalent results of all three Algorithms are very accurate.

We then study the convergence behavior of the three proposed algorithms. The considered massive MIMO downlink is still that with $M_t = 128$, $M_k = 4$ and $K = 10$, and the values of $\alpha_k$s are those presented in Table II. As shown in Step 1 in each algorithm, we use random initializations for Algorithms 1 and 2, whereas the initializations for Algorithm 3 are fixed. Fig. 4 shows the convergence behaviors of the three proposed algorithms at the second block of the first time slot for the massive MIMO downlink at two different SNRs. The expected sum-rate results presented in Fig. 4 are the deterministic equivalent results. From Fig. 4 we see that all three algorithms quickly converge to the local maximums of the ergodic sum-rate at SNR = 0 dB. We also observe that all three algorithms take more iterations to converge as the SNR increases. At SNR = 10 dB, Algorithms 1 and 3 take 20 iterations to converge, whereas Algorithms 2 needs 30 iterations to converge. Algorithm 3 only need to be performed once when the $a$ priori statistical CSI changes. Thus, the number of the iterations needed to make Algorithm 3 converge is not a problem. On the contrary, Algorithms 1 and 2 are performed once for each block. Thus, the number of the iterations needed to make Algorithms 1 and 2 converge is an issue. To reduce the number of the iterations used in Algorithms 2 and 3, we use the resulting precoders from the previous block as initial instead of random initials at each block (not the first data block). Fig. 5 plots the convergence behaviors of Algorithms 1.
Fig. 4. Convergence trajectories of the three proposed algorithms at the second block of the first time slot for a massive MIMO downlink with $M_t = 128$, $M_k = 4$, $K = 10$ and the $\alpha$s presented in Table II.

Fig. 5. Convergence trajectories of Algorithms 1 and 2 using different initials at the third block of the first time slot for a massive MIMO downlink at SNR = 10 dB with $M_t = 128$, $M_k = 4$, $K = 10$ and the $\alpha$s presented in Table II.
TABLE III
THE VALUES OF $\alpha_k$s IN SCENARIO 3.

|       | $\alpha_1 - \alpha_4$ | $\alpha_5 - \alpha_8$ | $\alpha_9, \alpha_{10}$ | $\alpha_{11}, \alpha_{12}$ | $\alpha_{13} - \alpha_{16}$ | $\alpha_{17} - \alpha_{20}$ |
|-------|------------------------|------------------------|--------------------------|-----------------------------|-----------------------------|-----------------------------|
| Case 1| 0.999                  | 0.999                  | 0.999                    | 0.999                       | 0.999                       | 0.999                       |
| Case 2| 0.999                  | 0.999                  | 0.999                    | 0.9                         | 0.9                         | 0.9                         |
| Case 3| 0.999                  | 0.9                    | 0.5                      | 0.5                         | 0.1                         | 0                           |

Fig. 6. Average sum-rate performance of Algorithm 2 and the robust RZF precoder for a massive MIMO downlink with $M_t = 128$, $M_k = 1$ and $K = 20$.

and 2 at the third block of the first time slot using two different initials for the same massive MIMO downlink as that of Fig. 4 at SNR = 10dB. From Fig. 5 we see that the resulting precoders from the previous block are very good initials, and thus only a few iterations are needed to achieve good performance.

Finally, we investigate the performance of Algorithm 2 for massive MIMO downlinks with single antenna users. The regularized zero forcing (RZF) precoder [33], [34], also called the transmit Wiener filter (TxWF) or the MMSE precoder, is widely used in massive downlink when each user has a single antenna. When the BS has imperfect CSI, it can be extended to the robust RZF [35]. To show the sum-rate performance of Algorithm 2, we compare it with the sum-rate performance of the robust RZF precoder. We consider a massive MIMO downlink with $M_t = 128$ transmit antennas at the BS and $K = 20$ single antenna users. The $\alpha_k$s used in the simulations are presented in Table III. Fig. 6 plots the average sum-rate performance of Algorithm 2 and the robust RZF for the considered scenario over $N_s = 100$ time slots. As shown in Fig. 6, Algorithm 2 can achieve much better performance than that of the robust RZF precoder at all three cases. Furthermore, we observe that the performance gains are small at low SNR, but become significant as SNR increases. Specifically, while the average sum-rates of Algorithm 2 increase significantly with SNR at all SNRs, the average sum-rates of the robust RZF precoder change very slowly
after SNR = 5 dB. For Case 2, the average sum-rate of the robust RZF precoder even becomes slightly smaller at SNR = 20 dB compared with that of SNR = 15 dB. The reason is as follows. The robust RZF precoder is designed by minimizing the MSE, which is not monotonically with sum-rate. Meanwhile, Algorithm 2 can achieve good average sum-rate performance at all SNRs due to using the expected sum-rate maximization. In conclusion, Algorithm 2 are more effective in improving the average sum-rate performance in comparison with the robust RZF precoder.

V. CONCLUSION

In this paper, we investigated the design of robust linear precoders for the massive MIMO downlink with imperfect CSI. The available imperfect CSI for each UE obtained at the BS is modeled as statistical CSI under a jointly correlated channel model with both channel mean and channel variance information, which includes the effects of channel estimation error, channel aging and spatial correlation. We derived an algorithm for robust linear precoder design by using the MM algorithm. The derived algorithm can achieve a local optimum of the expected weighted sum-rate maximization problem. We then used the deterministic equivalent method to compute the approximations of several key matrices used in the robust linear precoder design. Then, we proposed an algorithm for robust linear precoder design based on the deterministic equivalent method. The proposed algorithm needs $2K$ large dimensional matrix inversions per iteration. To reduce the computational complexity, we then derived two low-complexity algorithms, one for the general case, and the other for the case when all the channel means are zeros. We also proved the optimality of the beam domain transmissions when all the channel means are zeros. Simulation results showed that the proposed robust linear precoder designs apply to various mobile scenarios and achieve high spectral efficiency.

APPENDIX A

PROOF OF THEOREM 1

For brevity, we omit the subscript of $E_{H_{k,m,n}|Y_{m1}}$ in the Appendix A. The function $E \left\{ \log \det \left( E_{k,m,n}^{-1} \right) \right\}$ is a convex function of $E_{k,m,n}$ on $S_{++}^n$. Using the first order condition of convex functions [36], we obtain

$$
E \left\{ \log \det \left( E_{k,m,n}^{-1} \right) \right\} \geq E \left\{ \log \det \left( \left( E_{d_{k,m,n}} \right)^{-1} \right) \right\} - E \left\{ \text{tr} \left( \left( E_{k,m,n}^{-1} \right) \left( E_{k,m,n} - E_{d_{k,m,n}} \right) \right) \right\}
$$

$$
= E \left\{ \log \det \left( \left( E_{d_{k,m,n}} \right)^{-1} \right) \right\} + \text{tr} \left( I_{d_k} \right) - E \left\{ \text{tr} \left( \left( E_{d_{k,m,n}} \right)^{-1} E_{k,m,n} \right) \right\}.
$$

The step in (85) is inspired by [37]. The first and second items at the right hand side (RHS) of the inequality in (85) are constants, and the third item $-E \left\{ \text{tr} \left( \left( E_{d_{k,m,n}} \right)^{-1} E_{k,m,n} \right) \right\}$ is still not a simple function of the precoding matrices. Inspired by [10], let $G_{k,m,n}^H$ denote the linear receiver of the $k$-th user. The mean-square error (MSE) matrix of the $k$-th estimate $\hat{x}_{k,m,n} = G_{k,m,n}^H y_{k,m,n}$ is given by

$$
\Theta_{k,m,n} = E \left\{ (\hat{x}_{k,m,n} - x_{k,m,n}) (\hat{x}_{k,m,n} - x_{k,m,n})^H \right\}
$$

$$
= (I_{d_k} - G_{k,m,n}^H H_{k,m,n} P_{k,m,n}) (I_{d_k} - G_{k,m,n}^H H_{k,m,n} P_{k,m,n})^H
$$

$$
+ G_{k,m,n}^H \sum_{l \neq k} r_{k,m,n}^{pri} (P_{l,m,n} P_{l,m,n}^H) G_{k,m,n} + \sigma_z^2 G_{k,m,n} G_{k,m,n}.
$$

(86)

September 14, 2017 DRAFT
From (86), we observe that the function \( \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \Theta_{k,m,n} \right) \) is a convex function of \( G_{k,m,n} \). Furthermore, the global minimum of \( \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \Theta_{k,m,n} \right) \) is achieved when \( (G^{*}_{k,m,n})^H \) is the linear minimum mean-square error (MMSE) receiver, i.e.,

\[
(G^{*}_{k,m,n})^H = P^H_{k,m,n} H^H_{k,m,n} \left( R_{k,m,n} + H_{k,m,n} P_{k,m,n} P^H_{k,m,n} H^H_{k,m,n} \right)^{-1}. \tag{87}
\]

Substituting \( G^{*}_{k,m,n} \) into (86), we obtain \( \Theta_{k,m,n}(G^{*}_{k,m,n}) = E_{k,m,n} \). Thus, we have

\[
E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} E_{k,m,n} \right) \right\} \leq E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \Theta_{k,m,n} \right) \right\} \tag{88}
\]

for any \( G_{k,m,n} \). From (85) and (88), we obtain

\[
E \left\{ \log \text{det} \left( E^{-1}_{k,m,n} \right) \right\} \geq E \left\{ \log \text{det} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \right) \right\} + \text{tr} (I_{dk}) - E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \Theta_{k,m,n} \right) \right\}. \tag{89}
\]

Furthermore, to make the equality in (89) hold at \( P^{(d)}_{1,k,m,n}, \ldots, P^{(d)}_{K,k,m,n} \), we set \( G_{k,m,n} = G^{(d)}_{k,m,n} \), which is defined by

\[
(G^{(d)}_{k,m,n})^H = (P^{(d)}_{k,m,n})^H H^H_{k,m,n} \left( R^{(d)}_{k,m,n} + H_{k,m,n} P^{(d)}_{k,m,n} P^H_{k,m,n} H^H_{k,m,n} \right)^{-1}. \tag{90}
\]

When \( G_{k,m,n} = G^{(d)}_{k,m,n} \), we obtain from (86) that

\[
E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \Theta_{k,m,n} \right) \right\} = E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} (I_{dk}) - (G^{(d)}_{k,m,n})^H H_{k,m,n} P_{k,m,n} (I_{dk} - (G^{(d)}_{k,m,n})^H H_{k,m,n} P_{k,m,n})^H \right) \right\} \]

\[
- \sigma^2 E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} (G^{(d)}_{k,m,n})^H G^{(d)}_{k,m,n} \right) \right\}. \tag{91}
\]

It follows that

\[
E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \Theta_{k,m,n} \right) \right\} = E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} \right) \right\} - \sigma^2 E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} G^{(d)}_{k,m,n} H^H G^{(d)}_{k,m,n} \right) \right\} + E \left\{ \text{tr} \left( \left( E^{(d)}_{k,m,n} \right)^{-1} (G^{(d)}_{k,m,n})^H H_{k,m,n} P_{k,m,n} (G^{(d)}_{k,m,n})^H \right) \right\} - E \left\{ \text{tr} \left( H^H_{k,m,n} G^{(d)}_{k,m,n} E^{(d)}_{k,m,n} \right) \right\} - E \left\{ \text{tr} \left( G^{(d)}_{k,m,n} P_{k,m,n} \right) \right\} - E \left\{ \text{tr} \left( G^{(d)}_{k,m,n} \sum_{l \neq k} \eta_{l,m,n} P_{l,m,n} H^H_{l,m,n} \right) \right\}. \tag{92}
\]

The last term on the RHS of the equality in (92) can be rewritten as

\[
E \left\{ \text{tr} \left( G^{(d)}_{k,m,n} \left( E^{(d)}_{k,m,n} \right)^{-1} (G^{(d)}_{k,m,n})^H \sum_{l \neq k} \eta_{l,m,n} P_{l,m,n} \right) \right\} = \text{tr} \left( \sum_{l \neq k} \eta_{l,k,m,n} \left( E^{(d)}_{k,m,n} \right)^{-1} G^{(d)}_{k,m,n} \right) \sum_{l \neq k} P_{l,m,n} \right). \tag{93}
\]
Let \( c_{k,m,n}^{(d)} \), \( A_{k,m,n}^{(d)} \), \( B_{k,m,n}^{(d)} \) and \( C_{k,m,n}^{(d)} \) be defined as
\[
c_{k,m,n}^{(d)} = \mathbb{E} \left\{ \log \det \left( E_{k,m,n}^{(d)} \right)^{-1} \right\} + \text{tr} \left( I_{dk} \right) - \mathbb{E} \left\{ \text{tr} \left( E_{k,m,n}^{(d)} \right)^{-1} \right\} - \sigma^2 \mathbb{E} \left\{ \text{tr} \left( E_{k,m,n}^{(d)} \right)^{-1} \left( G_{k,m,n}^{(d)} \right)^H G_{k,m,n}^{(d)} \right\}
\]
(94)
\[
(A_{k,m,n}^{(d)})^H = \mathbb{E} \left\{ \left( E_{k,m,n}^{(d)} \right)^{-1} \left( G_{k,m,n}^{(d)} \right)^H H_{k,m,n} \right\}
\]
(95)
\[
B_{k,m,n}^{(d)} = \mathbb{E} \left\{ H_{k,m,n} G_{k,m,n}^{(d)} \left( E_{k,m,n}^{(d)} \right)^{-1} \left( G_{k,m,n}^{(d)} \right)^H H_{k,m,n} \right\}
\]
(96)
\[
C_{k,m,n}^{(d)} = 2 \text{pri}_{H_{k,m,n}} \left( \mathbb{E} \left\{ G_{k,m,n}^{(d)} \left( E_{k,m,n}^{(d)} \right)^{-1} \left( G_{k,m,n}^{(d)} \right)^H \right\} \right).
\]
(97)

From (89), (92) and (94) to (97), we obtain
\[
\mathbb{E} \left\{ \log \det \left( E_{k,m,n}^{(d)} \right) \right\} 
\geq c_{k,m,n}^{(d)} + \text{tr} \left( (A_{k,m,n}^{(d)})^H P_{k,m,n} \right) + \text{tr} \left( A_{k,m,n}^{(d)} P_{k,m,n}^H \right) - \text{tr} \left( B_{k,m,n}^{(d)} P_{k,m,n}^H + C_{k,m,n}^{(d)} \sum_{l \neq k} P_{l,m,n} P_{l,m,n}^H \right).
\]
(98)

Furthermore, we have
\[
\mathbb{E} \left\{ \log \det \left( E_{k,m,n}^{(d)} \right)^{-1} \right\} = c_{k,m,n}^{(d)} + \text{tr} \left( (A_{k,m,n}^{(d)})^H P_{k,m,n} \right) + \text{tr} \left( A_{k,m,n}^{(d)} (P_{k,m,n}^H) \right) - \text{tr} \left( B_{k,m,n}^{(d)} P_{k,m,n}^H + C_{k,m,n}^{(d)} \sum_{l \neq k} P_{l,m,n} (P_{l,m,n}^H) \right).
\]
(99)

Substituting the expression of \( G_{k,m,n} \) into (94), and the expressions of \( E_{k,m,n} \) and \( G_{k,m,n}^{(d)} \) into (95), (96) and (97), we obtain \( c_{k,m,n}^{(d)}, (A_{k,m,n}^{(d)})^H, B_{k,m,n}^{(d)} \) and \( C_{k,m,n}^{(d)} \) as in (31), (32), (33) and (34), the details are provided in the following.

Let \( \tilde{R}_{k,m,n}^{(d)} \) denote \( H_{k,m,n} P_{k,m,n}^{(d)} (P_{k,m,n}^{(d)})^H H_{k,m,n}^H \). Substituting (28) and (90) into (95), we obtain
\[
(A_{k,m,n}^{(d)})^H = \mathbb{E} \left\{ \left( E_{k,m,n}^{(d)} \right)^{-1} \left( G_{k,m,n}^{(d)} \right)^H H_{k,m,n} \right\}
\]
(100)
\[
= \mathbb{E} \left\{ \left( I_{dk} + P_{k,m,n}^{(d)} \right)^H H_{k,m,n} \left( R_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} P_{k,m,n}^{(d)} \right\}
\]
\[
= \mathbb{E} \left\{ (P_{k,m,n}^{(d)})^H H_{k,m,n} \left( R_{k,m,n}^{(d)} + \tilde{R}_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} \right\}
\]
\[
+ \mathbb{E} \left\{ (P_{k,m,n}^{(d)})^H H_{k,m,n} \left( P_{k,m,n}^{(d)} \right)^{-1} \tilde{R}_{k,m,n}^{(d)} \left( R_{k,m,n}^{(d)} + \tilde{R}_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} \right\}.
\]
(100)

The second item on the RHS of the equality of the above equation can be re-expressed as
\[
\mathbb{E} \left\{ (P_{k,m,n}^{(d)})^H H_{k,m,n} \left( R_{k,m,n}^{(d)} + \tilde{R}_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} \right\}
\]
\[
= \mathbb{E} \left\{ (P_{k,m,n}^{(d)})^H H_{k,m,n} \left( R_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} \right\}
\]
\[
- \mathbb{E} \left\{ (P_{k,m,n}^{(d)})^H H_{k,m,n} \left( R_{k,m,n}^{(d)} + \tilde{R}_{k,m,n}^{(d)} \right)^{-1} H_{k,m,n} \right\}.
\]
(101)
From (100) and (101), we then obtain the expression of $A_{k,m,n}^{(d)}$ in (32) and

$$
(E_{k,m,n}^{(d)})^{-1}(G_{k,m,n}^{(d)})^H = (P_{k,m,n}^{(d)}G_{k,m,n}^{(d)}H_{k,m,n}^{(d)}R_{k,m,n}^{(d)})^{-1}.
$$

(102)

According to (96) and (102), we obtain

$$
B_{k,m,n}^{(d)} = E\left\{H_{k,m,n}^{(d)}G_{k,m,n}^{(d)}(E_{k,m,n}^{(d)})^{-1}(G_{k,m,n}^{(d)})^H H_{k,m,n}^{(d)}\right\}
= E\left\{H_{k,m,n}^{(d)}G_{k,m,n}^{(d)}(P_{k,m,n}^{(d)}G_{k,m,n}^{(d)}H_{k,m,n}^{(d)}R_{k,m,n}^{(d)})^{-1} H_{k,m,n}^{(d)}\right\}.
$$

(103)

Substituting (90) into the above equation, we then obtain

$$
B_{k,m,n}^{(d)} = E\left\{H_{k,m,n}^{(d)}(R_{k,m,n}^{(d)} + H_{k,m,n}^{(d)}P_{k,m,n}^{(d)}(P_{k,m,n}^{(d)}H_{k,m,n}^{(d)}R_{k,m,n}^{(d)})^{-1} H_{k,m,n}^{(d)})ight\}.
$$

(104)

The RHS of the equality in (104) is similar to the LHS of the equality in (101). Thus, we obtain the expression of $B_{k,m,n}^{(d)}$ in (33). Similarly, we can obtain the expression of $C_{k,m,n}^{(d)}$ in (31) and the expression of $C_{k,m,n}^{(d)}$ in (34).

Let $D_{k,m,n}^{(d)}$ be defined as $D_{k,m,n}^{(d)} = w_kB_{k,m,n}^{(d)} + \sum_{l\neq k} w_lC_{l,m,n}^{(d)}$. Recall that the objective function $f$ can be written as

$$
f(P_{1,m,n}, P_{2,m,n}, \ldots, P_{K,m,n}) = \sum_{k=1}^{K} w_k E\left\{ \log \det \left( (E_{k,m,n})^{-1} \right) \right\}.
$$

(105)

From the above equation, (98) and (99), we obtain the function $g_1$ defined in (30) is a minorizing function of the objective function.

**Appendix B**

**Proof of Theorem 2**

From (44), we can obtain the gradient of $\mathcal{R}_{k,m,n}(\sigma^2)$ with respect to $P_{k,m,n}P_{k,m,n}^H$ as

$$
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2)}{\partial (P_{k,m,n}P_{k,m,n}^H)} = (I_M + \Gamma_{k,m,n}P_{k,m,n}P_{k,m,n}^H)^{-1}\Gamma_{k,m,n}
$$

$$
+ \sum_{i,j} \frac{\partial \mathcal{R}_{k,m,n}(\sigma^2)}{\partial [\eta_{k,m,n}^{post} (P_{k,m,n}G_{k,m,n}P_{k,m,n}^H)]_{ij}} \frac{\partial [\eta_{k,m,n}^{post} (P_{k,m,n}G_{k,m,n}P_{k,m,n}^H)]_{ij}}{\partial (P_{k,m,n}P_{k,m,n}^H)}
$$

$$
+ \sum_{i,j} \frac{\partial \mathcal{R}_{k,m,n}(\sigma^2)}{\partial [\gamma_{k,m,n}]_{ij}} \frac{\partial [\gamma_{k,m,n}]_{ij}}{\partial (P_{k,m,n}P_{k,m,n}^H)}.
$$

(106)

Using methods similar to that in the proof of Theorem 4 in [21], we obtain

$$
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2)}{\partial [\eta_{k,m,n}^{post} (P_{k,m,n}G_{k,m,n}P_{k,m,n}^H)]_{ij}} = 0
$$

(107)

$$
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2)}{\partial [\gamma_{k,m,n}]_{ij}} = 0.
$$

(108)

Thus, we have

$$
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2)}{\partial (P_{k,m,n}P_{k,m,n}^H)} \bigg|_{P_{k,m,n} = P_{k,m,n}^{(d)}} = (I_M + \Gamma_{k,m,n}P_{k,m,n}^{(d)}(P_{k,m,n}^{(d)})^H)^{-1}\Gamma_{k,m,n}.
$$

(109)
From (19), we obtain the gradient of $\mathcal{R}_{k,m,n}(\sigma^2_z)$ with respect to $\mathbf{P}_{k,m,n}^H\mathbf{P}_{k,m,n}$ as
\[
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2_z)}{\partial (\mathbf{P}_{k,m,n}^H\mathbf{P}_{k,m,n})} \bigg|_{\mathbf{P}_{k,m,n}=\mathbf{P}_{k,m,n}^{(d)}} = \mathbb{E}_{\mathbf{Y}^{\text{BS}}_{m,1}} \left\{ \mathbf{H}_{k,m,n}^H \left( \mathbf{R}_{k,m,n}^{(d)} + \mathbf{H}_{k,m,n} \mathbf{P}_{k,m,n}^{(d)} (\mathbf{P}_{k,m,n}^{(d)})^H \mathbf{H}_{k,m,n}^H \right)^{-1} \mathbf{H}_{k,m,n} \right\}. \tag{110}
\]
According to Lemma 1 we have that $\mathcal{R}_{k,m,n}(\sigma^2_z)$ is the deterministic equivalent of $\mathcal{R}_{k,m,n}(\sigma^2_z)$. From (109) and (110), we obtain $\mathbb{E}_{\mathbf{H}_{k,m,n}} \left\{ \mathbf{H}_{k,m,n}^H \left( \mathbf{R}_{k,m,n}^{(d)} + \mathbf{H}_{k,m,n} \mathbf{P}_{k,m,n}^{(d)} (\mathbf{P}_{k,m,n}^{(d)})^H \mathbf{H}_{k,m,n}^H \right)^{-1} \mathbf{H}_{k,m,n} \right\}$.

Recall that
\[
\mathbf{B}_{k,m,n}^{(d)} = \mathbb{E}_{\mathbf{H}_{k,m,n}} \mathbf{Y}^{\text{BS}}_{m,1} \left\{ \mathbf{H}_{k,m,n}^H \left( \mathbf{R}_{k,m,n}^{(d)} \right)^{-1} \mathbf{H}_{k,m,n} \right\} - \mathbb{E}_{\mathbf{H}_{k,m,n}} \mathbf{Y}^{\text{BS}}_{m,1} \left\{ \mathbf{H}_{k,m,n}^H \mathbf{H}_{k,m,n} \left( \mathbf{R}_{k,m,n}^{(d)} + \mathbf{H}_{k,m,n} \mathbf{P}_{k,m,n}^{(d)} (\mathbf{P}_{k,m,n}^{(d)})^H \mathbf{H}_{k,m,n}^H \right)^{-1} \mathbf{H}_{k,m,n} \right\}. \tag{111}
\]
Then, we obtain that the matrix $\mathbf{B}_{k,m,n}^{(d)}$ provided in (52) is the deterministic equivalent of $\mathbf{B}_{k,m,n}^{(d)}$.

Similarly to process of obtaining the gradient of $\mathcal{R}_{k,m,n}(\sigma^2_z)$ with respect to $\mathbf{P}_{k,m,n}^H\mathbf{P}_{k,m,n}$, we can obtain the gradient of $\mathcal{R}_{k,m,n}(\sigma^2_z)$ with respect to $\mathbf{R}_{k,m,n}$ from (45) as
\[
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2_z)}{\partial \mathbf{R}_{k,m,n}} = -\mathbf{R}_{k,m,n}^{-1} (\mathbf{I}_{M_k} + \tilde{\Gamma}_{k,m,n} \mathbf{R}_{k,m,n}^{-1})^{-1} \tilde{\Gamma}_{k,m,n} \mathbf{R}_{k,m,n}^{-1}. \tag{112}
\]
The gradients of $\mathcal{R}_{k,m,n}(\sigma^2_z)$ with respect to $\mathbf{P}_{l,m,n}^H\mathbf{P}_{l,m,n}$, $l \neq k$, are then obtained from the above equation by applying the chain rule. Using a method similar to that in Lemma 4 of [38] and Theorem 2 of [39], we then obtain
\[
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2_z)}{\partial (\mathbf{P}_{l,m,n}^H\mathbf{P}_{l,m,n})} \bigg|_{\mathbf{P}_{l,m,n}=\mathbf{P}_{l,m,n}^{(d)}} = -\eta^{\text{pri}}_{k,m,n} \left( \left( \mathbf{R}_{k,m,n}^{(d)} \right)^{-1} (\mathbf{I}_{M_k} + \tilde{\Gamma}_{k,m,n} \left( \mathbf{R}_{k,m,n}^{(d)} \right)^{-1}) \right) - \eta^{\text{pri}}_{k,m,n} \left( \left( \mathbf{R}_{k,m,n}^{(d)} \right)^{-1} \right)
= -\mathcal{C}_{k,m,n}^{(d)}. \tag{113}
\]
From (19) and the chain rule, we then obtain
\[
\frac{\partial \mathcal{R}_{k,m,n}(\sigma^2_z)}{\partial (\mathbf{P}_{l,m,n}^H\mathbf{P}_{l,m,n})} \bigg|_{\mathbf{P}_{l,m,n}=\mathbf{P}_{l,m,n}^{(d)}} = \eta^{\text{pri}}_{k,m,n} \left( \mathbb{E}_{\mathbf{H}_{k,m,n}} \mathbf{Y}^{\text{BS}}_{m,1} \left\{ \left( \mathbf{R}_{k,m,n}^{(d)} + \mathbf{H}_{k,m,n} \mathbf{P}_{k,m,n}^{(d)} (\mathbf{P}_{k,m,n}^{(d)})^H \mathbf{H}_{k,m,n}^H \right)^{-1} \right\} \right) - \eta^{\text{pri}}_{k,m,n} \left( \left( \mathbf{R}_{k,m,n}^{(d)} \right)^{-1} \right)
= -\mathcal{C}_{k,m,n}^{(d)}. \tag{114}
\]
From Lemma 1 (113) and (114), we obtain $\mathcal{C}_{k,m,n}^{(d)}$ is the deterministic equivalent of $\mathcal{C}_{k,m,n}^{(d)}$. Thus, (53) holds.
APPENDIX C

PROOF OF THEOREM 3

We first rewrite the minorizing function $g_1$ provided by Theorem 1 as

$$g_1(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)})$$

$$= \sum_{k=1}^{K} w_k c_{k,m,n}^{(d)} + \sum_{k=1}^{K} w_k \text{tr} \left( (A_{k,m,n}^{(d)})^H P_{k,m,n} \right) + \sum_{k=1}^{K} \text{tr} \left( A_{k,m,n}^{(d)} P_{k,m,n}^H \right)$$

$$- \sum_{k=1}^{K} \text{tr} \left( (D_{k,m,n}^{(d)} + F_{k,m,n}^{(d)}) P_{k,m,n} P_{k,m,n}^H \right) + \sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} P_{k,m,n} P_{k,m,n}^H \right).$$

The fourth item on the RHS of the equality of (115) is a convex quadratic function of $P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}$.

Using the first order condition of convex functions, we obtain

$$\sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} P_{k,m,n} P_{k,m,n}^H \right)$$

$$\geq \sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} P_{k,m,n}^H (P_{k,m,n}^{(d)})^H \right) + \sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} (P_{k,m,n} - P_{k,m,n}^{(d)})(P_{k,m,n}^{(d)})^H \right)$$

$$+ \sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} P_{k,m,n}^H (P_{k,m,n} - P_{k,m,n}^{(d)})^H \right).$$

(116)

From (115) and (116), we then obtain

$$g_1(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)})$$

$$\geq \sum_{k=1}^{K} w_k c_{k,m,n}^{(d)} - \sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} P_{k,m,n}^H (P_{k,m,n}^{(d)})^H \right)$$

$$+ \sum_{k=1}^{K} \text{tr} \left( (w_k A_{k,m,n}^{(d)} + F_{k,m,n}^{(d)} P_{k,m,n}^H) P_{k,m,n} \right) + \sum_{k=1}^{K} \text{tr} \left( (w_k A_{k,m,n}^{(d)} + F_{k,m,n}^{(d)} P_{k,m,n}) P_{k,m,n}^H \right)$$

$$- \sum_{k=1}^{K} \text{tr} \left( (D_{k,m,n}^{(d)} + F_{k,m,n}^{(d)}) P_{k,m,n} P_{k,m,n}^H \right).$$

(117)

Let $c_{m,n}$ be defined as

$$c_{m,n} = \sum_{k=1}^{K} w_k c_{k,m,n}^{(d)} - \sum_{k=1}^{K} \text{tr} \left( F_{k,m,n}^{(d)} P_{k,m,n}^H (P_{k,m,n}^{(d)})^H \right)$$

(118)

and $g_2(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)})$ be defined as in (115). From (117), we have

$$g_2(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)})$$

$$\leq g_1(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)}).$$

(119)

Furthermore, it is easy to verify that

$$g_2(P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)})$$

$$= g_1(P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)}).$$

(120)

Thus, $g_2(P_{1,m,n}, P_{2,m,n}, \cdots, P_{K,m,n}|P_{1,m,n}^{(d)}, P_{2,m,n}^{(d)}, \cdots, P_{K,m,n}^{(d)})$ is also a minorizing function of the objective function.
APPENDIX D

PROOF OF THEOREM 4

From (19), we obtain
\[
\frac{\partial \tilde{\mathbf{R}}_{k,m,n}(\sigma^2)}{\partial (\mathbf{P}_{k,m,n}^*)} = (\mathbf{I}_M + \Gamma_{k,m,n} \mathbf{P}_{k,m,n} \mathbf{P}_{k,m,n}^H)^{-1} \Gamma_{k,m,n} \mathbf{P}_{k,m,n}.
\] (121)

and
\[
\frac{\partial \tilde{\mathbf{R}}_{l,m,n}(\sigma^2)}{\partial (\mathbf{P}_{k,m,n}^*)} = -\eta_{l,m,n}^{pri} \left( \mathbf{R}_{l,m,n}^{-1} - (\mathbf{R}_{l,m,n} + \tilde{\Gamma}_{l,m,n})^{-1} \right) \mathbf{P}_{k,m,n}.
\] (122)

Since \( \mathcal{J}(\mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n}) \) denotes \( \sum_{k=1}^{K} w_k \tilde{\mathbf{R}}_{k,m,n}(\sigma^2) \), we obtain
\[
\frac{\partial \mathcal{J}(\mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n})}{\partial (\mathbf{P}_{k,m,n}^*)} = (\mathbf{I}_M + \Gamma_{k,m,n} \mathbf{P}_{k,m,n} \mathbf{P}_{k,m,n}^H)^{-1} \Gamma_{k,m,n} \mathbf{P}_{k,m,n} \nonumber
\]
\[
- \sum_{l \neq k} \eta_{l,m,n}^{pri} \left( \mathbf{R}_{l,m,n}^{-1} - (\mathbf{R}_{l,m,n} + \tilde{\Gamma}_{l,m,n})^{-1} \right) \mathbf{P}_{k,m,n}.
\] (123)

We define the Lagrangian as
\[
\mathcal{L}(\mu, \mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n}) = -\mathcal{J}(\mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n}) + \mu \left( \sum_{k=1}^{K} \text{tr} \left( \mathbf{P}_{k,m,n} \mathbf{P}_{k,m,n}^H - P \right) \right).
\] (124)

From the first order optimal conditions of (124), we obtain
\[
- w_k (\mathbf{I}_M + \Gamma_{k,m,n} \mathbf{P}_{k,m,n} \mathbf{P}_{k,m,n}^H)^{-1} \Gamma_{k,m,n} \mathbf{P}_{k,m,n} \nonumber
\]
\[
+ \sum_{l \neq k} w_l \eta_{l,m,n}^{pri} \left( \mathbf{R}_{l,m,n}^{-1} - (\mathbf{R}_{l,m,n} + \tilde{\Gamma}_{l,m,n})^{-1} \right) \mathbf{P}_{k,m,n} + \mu \mathbf{P}_{k,m,n} = \mathbf{0}.
\] (125)

From (78), we obtain
\[
\sum_{l \neq k} w_l \eta_{l,m,n}^{pri} \left( \mathbf{R}_{l,m,n}^{-1} - (\mathbf{R}_{l,m,n} + \tilde{\Gamma}_{l,m,n})^{-1} \right) = \mathbf{V}_M \tilde{\Sigma}^2_{k,m,n} \mathbf{V}_M^H.
\] (126)

where \( \tilde{\Sigma}^2_{k,m,n} \) is a diagonal matrix whose value depends on \( \mathbf{P}_{1,m,n}, \mathbf{P}_{2,m,n}, \ldots, \mathbf{P}_{K,m,n} \). Then, the first order conditions in (125) become
\[
w_k (\mathbf{I}_M + \Gamma_{k,m,n} \mathbf{P}_{k,m,n} \mathbf{P}_{k,m,n}^H)^{-1} \Gamma_{k,m,n} \mathbf{P}_{k,m,n} = \mathbf{V}_M \tilde{\Sigma}^2_{k,m,n} \mathbf{V}_M^H \mathbf{P}_{k,m,n} + \mu \mathbf{P}_{k,m,n}.
\] (127)

When \( \mathbf{H}_{k,m,n} = \mathbf{0} \), we have
\[
\Gamma_{k,m,n} = \mathbf{V}_M \tilde{\Sigma}^2_{k,m,n} \mathbf{V}_M^H.
\] (128)

We define \( \mathbf{T}_{k,m,n} = \mu \mathbf{I}_M + \mathbf{V}_M \tilde{\Sigma}^2_{k,m,n} \mathbf{V}_M^H \), \( \Gamma'_{k,m,n} = \mathbf{T}_{k,m,n}^{-1/2} \Gamma_{k,m,n} \mathbf{T}_{k,m,n}^{-1/2} \) and \( \mathbf{P}'_{k,m,n} = \mathbf{T}_{k,m,n}^{1/2} \mathbf{P}_{k,m,n} \). Then, the conditions in (127) become
\[
w_k \Gamma'_{k,m,n} \mathbf{P}'_{k,m,n} (\mathbf{I}_M + (\mathbf{P}'_{k,m,n})^H \Gamma'_{k,m,n} \mathbf{P}'_{k,m,n})^{-1} = \mathbf{P}'_{k,m,n}.
\] (129)
It follows that
\[ w_k \Gamma'_{k,m,n} \mathbf{P}'_{k,m,n} = \mathbf{P}'_{k,m,n} (\mathbf{I}_M + (\mathbf{P}'_{k,m,n})^H \Gamma'_{k,m,n} \mathbf{P}'_{k,m,n}). \] (130)

Right multiplying both sides of the equality in (130) by \((\mathbf{P}'_{k,m,n})^H\), we obtain
\[ w_k \Gamma'_{k,m,n} \mathbf{P}'_{k,m,n} (\mathbf{P}'_{k,m,n})^H = \mathbf{P}'_{k,m,n} (\mathbf{I}_M + (\mathbf{P}'_{k,m,n})^H \Gamma'_{k,m,n} \mathbf{P}'_{k,m,n}) (\mathbf{P}'_{k,m,n})^H. \] (131)

Thus, we obtain \( \Gamma'_{k,m,n} \mathbf{P}'_{k,m,n} (\mathbf{P}'_{k,m,n})^H = \mathbf{P}'_{k,m,n} (\mathbf{P}'_{k,m,n})^H \Gamma'_{k,m,n} \), which indicates that \( \mathbf{P}'_{k,m,n} (\mathbf{P}'_{k,m,n})^H \) commutes with \( \Gamma'_{k,m,n} \). From Theorem 9-33 of [40] we then obtain \( \mathbf{P}'_{k,m,n} (\mathbf{P}'_{k,m,n})^H \) and \( \Gamma'_{k,m,n} \) have the same eigenvectors. From \( \Gamma'_{k,m,n} = \mathbf{T}_{k,m,n}^{-1/2} \Gamma_{k,m,n} \mathbf{T}_{k,m,n}^{1/2} \), we have that the eigenvectors of \( \Gamma'_{k,m,n} \) and \( \Gamma_{k,m,n} \) are the same. Thus, we have that the left singular vector matrix of \( \mathbf{P}'_{k,m,n} \) can be written as
\[ \mathbf{U}_{\mathbf{P}'_{k,m,n}} = \mathbf{V}_M \Pi'_{k,m,n} \] (132)
where \( \Pi'_{k,m,n} \) is a permutation matrix. From \( \mathbf{P}_{k,m,n} = \mathbf{T}_{k,m,n}^{-1/2} \mathbf{P}'_{k,m,n} \) and (132), we obtain (81) holds finally.

REFERENCES

[1] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
[2] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, “Secure massive MIMO transmission with an active eavesdropper,” *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3880–3900, 2016.
[3] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, “Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits,” *IEEE Inf. Theory, vol. 60*, no. 11, pp. 7112–7139, 2014.
[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.
[5] C. Sun, X. Q. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, “Beam division multiple access transmission for massive MIMO communications,” *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170 – 2184, 2015.
[6] X. Meng, X. Q. Gao, and X.-G. Xia, “Omnidirectional precoding based transmission in massive MIMO systems,” *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 174–186, 2016.
[7] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.
[8] A.-A. Lu, X. Q. Gao, Y. R. Zheng, and C. Xiao, “Low complexity polynomial expansion detector with deterministic equivalents of the moments of channel Gram matrix for massive MIMO uplink,” *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 586–600, 2016.
[9] S. S. Christensen, R. Agarwal, E. Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.
[10] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted mmse approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.
[11] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, “A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing,” *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, 2016.
[12] C. Sun, X. Q. Gao, and Z. Ding, “BDMA in multi-cell massive MIMO communications: Power allocation algorithms,” *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2962–2974, 2017.
[13] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing–the large-scale array regime,” *IEEE Trans. Inf. Theory*, vol. 10, no. 59, pp. 6441–6463, 2013.
[14] X. Q. Gao, B. Jiang, X. Li, A. B. Gershman, and M. R. McKay, “Statistical eigenmode transmission over jointly correlated MIMO channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3735–3750, 2009.
[15] M. Razaviyayn, M. S. Boroujeni, and Z.-Q. Luo, “A stochastic weighted mmse approach to sum rate maximization for a mimo interference channel,” in *Signal Processing Advances in Wireless Communications (SPAWC), 2013 IEEE 14th Workshop on*. IEEE, 2013, pp. 325–329.
[16] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, “A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks,” *Mathematical Programming*, vol. 157, no. 2, pp. 515–545, 2016.

[17] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, “Robust stochastic approximation approach to stochastic programming,” *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.

[18] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on stochastic programming: modeling and theory*. Siam, 2014, vol. 16.

[19] D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[20] Y. Sun, P. Babu, and D. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2016.

[21] A.-A. Lu, X. Q. Gao, and C. Xiao, “Free deterministic equivalents for the analysis of MIMO multiple access channel,” *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4604–4629, 2016.

[22] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[23] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, “Pilot beam pattern design for channel estimation in massive MIMO systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, 2014.

[24] L. You, X. Q. Gao, A. L. Swindlehurst, and W. Zhong, “Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots,” *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, 2016.

[25] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Multiuser MIMO achievable rates with downlink training and channel state feedback,” *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

[26] M. K. Member and G. Caire, “Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge,” *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1468–1477, 2007.

[27] R. H. Etkin and D. N. Tse, “Degrees of freedom in some underspread MIMO fading channels,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, 2006.

[28] J. Hoydis, S. Ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[29] F. Vaida, “Parameter convergence for EM and MM algorithms,” *Statistica Sinica*, pp. 831–840, 2005.

[30] M. W. Jacobson and J. A. Fessler, “An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms,” *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2411–2422, 2007.

[31] C.-K. Wen, S. Jin, and K.-K. Wong, “On the sum-rate of multiuser MIMO uplink channels with jointly-correlated Rician fading,” *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2883–2895, 2011.

[32] J. Salo, G. Del Galdo, J. Salmi, P. Kyöstjäntti, M. Milojevic, D. Laselva, and C. Schneider. (2005, Jan.) MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996). [Online]. Available: http://www.tkk.fi/Units/Radio/scm/

[33] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization,” *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, 2005.

[34] M. Joham, W. Utschick, and J. A. Nossek, “Linear transmit processing in MIMO communications systems,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, 2005.

[35] L. You, X. Q. Gao, X.-G. Xia, N. Ma, and Y. Peng, “Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, 2015.

[36] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.