Robust MMSE Precoding in MIMO Channels With Pre-Fixed Receivers

Jiaheng Wang, *Student Member, IEEE*, and Daniel P. Palomar, *Senior Member, IEEE*

***Abstract*—In this paper, we design robust precoders, under the minimum mean square error (MMSE) criterion, for different types of channel state information (CSI) in multiple-input multiple-output (MIMO) channels. We consider low-complexity pre-fixed receivers that may adapt to the channel but are oblivious to the existence of a precoder at the transmitter. In particular, three types of CSI are taken into account: i) perfect CSI, ii) statistical CSI in the form of mean feedback, and iii) deterministic imperfect CSI assuming that the actual channel is within the neighborhood of a nominal channel, which leads to the worst-case robust design that is the focus of this paper. Interestingly, it is found that, under some mild conditions, the optimal transmit directions, i.e., the left singular vectors of the precoder, are equal to the right singular vectors of the channel, the channel mean, and the nominal channel for perfect CSI, statistical CSI, and the worst-case design, respectively. Consequently, the matrix-valued problems can be simplified to scalar power allocation problems that either admit closed-form solutions or can be efficiently solved by the proposed algorithm.**

***Index Terms*—Convex optimization, imperfect CSI, MIMO, MMSE, minimax, robust precoders, worst-case designs.**

I. Introduction

Using multiple transmit and receive antennas has been widely known as an effective way to increase the capacity of wireless communications [1], [2]. The performance of a multiple-input multiple-output (MIMO) system depends, to a substantial extent, on the quality of channel state information (CSI) available at the transmitter and receiver. Traditional MIMO transceiver optimization is based on accurate CSI at the transmitter (CSIT) and CSI at the receiver (CSIR) [3]–[7]. However, due to many factors such as inaccurate channel estimation, quantization, erroneous or outdated feedback, and time delays or frequency offsets between the reciprocal channels, CSI, especially CSIT, is usually imperfect and partially known in practice. To improve the robustness of communication systems, the imperfectness of CSI has to be taken into consideration.

Typically, there are two classes of models to characterize imperfect CSI: the stochastic and deterministic (or worst-case) models. The stochastic model usually assumes the channel to be a complex random matrix with normally distributed elements, and the mean and/or the covariance, i.e., the slowly-varying channel statistics that can be well estimated, are known. The system design is then based on optimizing the average or outage performance [8]–[17]. On the other hand, the deterministic model assumes that the instantaneous channel, although not exactly known, lies in a known set of possible values, often called the uncertainty region and defined by some norm. The size of this region represents the amount of uncertainty on the channel, i.e., the bigger the region is the more uncertainty there is. In this case, the goal of the robust design is to guarantee a performance level for any channel realization in the uncertainty region, which is achieved by optimizing the worst-case performance [18]–[28] and usually leads to a maximin or minimax problem formulation.

The goal of this paper is to design robust precoders, under the minimum mean square error (MMSE) criterion, for different types of CSI, including perfect CSI (as a warm-up), statistical CSI in the form of mean feedback, and deterministic imperfect CSI assuming that the actual channel is within the neighborhood of a nominal channel. Motivated by low-complexity systems, we consider simple pre-fixed receivers, as in [29]–[31], in the sense that they are independent of the transmitter but may depend on the channel, so that the efforts to combat imperfect CSI are undertaken by the transmitter. This separate structure not only reduces the computational complexity of the receiver, which is especially preferable in mobile wireless communication systems, but also increases the backward compatibility of the receiver since advanced and complicated techniques can be introduced at the transmitter without modifying the receiver. Specifically, the pre-fixed receiver uses an equalizer that depends only on the channel, or the channel mean, or the nominal channel, but not the precoder, i.e., the receiver is oblivious to the existence of the precoder. Therefore, common linear equalizers, such as the matched filter (MF), the zero-forcing (ZF) and MMSE equalizers, can be used. One choice that maximally reduces the computational workload at the receiver, as in [29] and [30], is to use no equalizer at all.

For perfect and statistical CSI, the robust precoder designs are easily-recognized convex problems [32], thus admitting globally optimal solutions that can be efficiently found in polynomial time using, e.g., an interior-point method [33]. Interestingly, it has been observed that, for the optimal precoders with both perfect CSI [3]–[7] and statistical CSI [8]–[16], the transmit directions, i.e., the left singular vectors of the precoder, in many cases are the right singular vectors of either the channel or the channel mean, hence leading to the eigenmode transmission that is carried out through a set of parallel subchannels or eigenmodes. In this paper, we first show that,

Manuscript received August 25, 2009; accepted June 25, 2010. Date of publication July 29, 2010; date of current version October 13, 2010. This work was supported by the Hong Kong RGC 618709 research grant. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ali Ghrayeb.

The authors are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: jhwang@ust.hk; palomar@ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2062178

for both perfect and statistical CSI, the optimal transmit directions of the MMSE precoder with separate receivers are still the right singular vectors of the channel or the channel mean under some mild conditions. Hence, the matrix precoder designs can be simplified to scalar power allocation problems that have closed-form solutions.

For the worst-case design, as the focus of this paper, we consider an elliptical channel uncertainty region centered at a nominal channel and defined by the weighted Frobenius norm, which covers the model used in [18]–[28] as special cases. Given that the formulated minimax problem is not convex, we first transform it into a convex problem, or even further a semidefinite program (SDP) [34], so that it can be globally and efficiently solved in practice. In light of the optimality of the eigenmode transmission for perfect and statistical CSI, one may wonder whether the channel-diagonalizing structure is still optimal in the worst-case design. As the most important result, we prove that the optimal transmit directions for the worst-case design are the right singular vectors of the nominal channel under some mild conditions. Therefore, the eigenmode transmission is still optimal, which is consistent not only with the cases of perfect and statistical CSI, but also with the recent finding in [35], which considered maximizing the worst-case received signal-to-noise ratio (SNR) by choosing the optimal transmit covariance matrix. Consequently, the complicated matrix-valued problem can also be simplified to a scalar power allocation problem without losing any optimality. Then, an efficient algorithm based on primal/dual decomposition methods for optimization [36], [37] is proposed to solve the simplified problem. Finally, we derive the least favorable channels for both robust and non-robust MMSE precoders.

The paper is organized as follows. The signal model and the problem formulation are introduced in Section II. Sections III and IV address the precoding designs with perfect and statistical CSI, respectively. In Section V, we provide the convex reformulation for the worst-case design, and prove the optimality of the eigenmode transmission. The power allocation problem of the worst-case design is solved in Section VI, and the least favorable channels are derived in Section VII. Section VIII provides numerical examples. Finally, we conclude with Section IX.

Notation: Uppercase and lowercase boldface denote matrices and vectors, respectively. $\mathbb{R}^{m\times n}$ and $\mathbb{C}^{m\times n}$ denote the set of $m\times n$ matrices with real- and complex-valued entries, respectively, and $\mathbb{S}_+^n$ denotes the ensemble of all $n\times n$ positive semidefinite matrices. $[\mathbf{X}]_{ij}$ represents the (ith, jth) element of matrix $\mathbf{X}$. By $\mathbf{X}\succeq 0$ or $\mathbf{X}\succ 0$, we mean that $\mathbf{X}$ is a Hermitian positive semidefinite or definite matrix, respectively. The vectors $\mathbf{d}(\mathbf{A})$ and $\boldsymbol{\lambda}(\mathbf{A})$ contain the diagonal elements and the eigenvalues of a square matrix $\mathbf{A}$, respectively. The operators $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^{\dagger}$, $\mathrm{vec}(\cdot)$, and $\mathrm{Tr}(\cdot)$ denote the Hermitian, inverse, pseudo-inverse, stacking vectorization, and trace operations, respectively. The maximum eigenvalue of a Hermitian matrix is represented by $\lambda_{\max}(\cdot)$. $\|\cdot\|$ denotes a general norm of a matrix as well as the Euclidean norm of a vector, while $\|\cdot\|_F$ and $\|\cdot\|_2$ represent the Frobenius and spectral norms of a matrix. $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex value, and $\otimes$ represents the Kronecker product operator.

II. Problem Statement

A. Signal Model

We consider a MIMO channel with N transmit and M receive antennas. The transmit signal vector $\mathbf{s}\in\mathbb{C}^L$, whose elements are zero-mean, unit-variance and uncorrelated, i.e., $E[\mathbf{s}\mathbf{s}^H]=\mathbf{I}$, is linearly precoded by $\mathbf{F}\in\mathbb{C}^{N\times L}$ so that the received signal vector $\mathbf{y}\in\mathbb{C}^M$ can be expressed as

$$\mathbf{y}=\mathbf{HFs}+\mathbf{n} \tag{1}$$

where $\mathbf{H}\in\mathbb{C}^{M\times N}$ is the channel matrix, and $\mathbf{n}\in\mathbb{C}^M$ is a white circularly symmetric complex Gaussian noise vector, i.e., $\mathbf{n}\sim\mathcal{CN}(0,\sigma_n^2\mathbf{I})$. At the receiver, a linear equalizer $\mathbf{G}\in\mathbb{C}^{L\times M}$ is used to estimate $\mathbf{s}$ from $\mathbf{y}$, resulting in $\hat{\mathbf{s}}=\mathbf{Gy}$. Then, the system performance is measured by the MSE between $\hat{\mathbf{s}}$ and $\mathbf{s}$, which is given by

$$\mathrm{MSE}=E\{\|\hat{\mathbf{s}}-\mathbf{s}\|^2\}=\|\mathbf{GHF}-\mathbf{I}\|_F^2+\sigma_n^2\|\mathbf{G}\|_F^2. \tag{2}$$

Since the number of degrees of freedom is upper bounded by $\mathrm{rank}(\mathbf{H})\le\min\{M,N\}$, we assume that $L\le\mathrm{rank}(\mathbf{H})$.

Due to many practical issues, the channel state information, i.e., $\mathbf{H}$, is usually imperfectly known and partially available at the transmitter and/or receiver. There are typically two kinds of models to characterize the imperfectness of CSI: the stochastic and deterministic (or worst-case) models. In the stochastic model, the channel is usually assumed to be a complex random matrix with nonzero-mean and normally distributed elements, and the transmitter knows the channel mean, so it is often called mean feedback [9]–[16].[1] To be more specific, the channel is regarded as consisting of two components

$$\mathbf{H}=\bar{\mathbf{H}}+\mathbf{E} \tag{3}$$

where $\bar{\mathbf{H}}$ is the channel mean known to the transmitter, and the elements of $\mathbf{E}$ are independent and identically-distributed (i.i.d.), zero-mean, unit-variance, circularly symmetric complex Gaussian random variables, i.e.,

$$E\{\mathbf{E}\}=\mathbf{0},\ E\{\mathrm{vec}(\mathbf{E})\mathrm{vec}(\mathbf{E})^H\}=\mathbf{I}. \tag{4}$$

On the other hand, the deterministic (or worst-case) model assumes that the actual channel $\mathbf{H}$ lies in the neighborhood of a nominal channel $\hat{\mathbf{H}}$ known to the transmitter. Specifically, $\mathbf{H}$ is assumed to belong to an uncertainty region $\mathcal{H}\triangleq\{\mathbf{H}:\|\mathbf{H}-\hat{\mathbf{H}}\|\le\varepsilon\}$ defined by some norm $\|\cdot\|$, which is an ellipsoid centered at $\hat{\mathbf{H}}$ with the radius ε (also known to the transmitter). By introducing the channel error

$$\boldsymbol{\Delta}\triangleq\mathbf{H}-\hat{\mathbf{H}} \tag{5}$$

$\mathbf{H}\in\mathcal{H}$ can be equally described by $\boldsymbol{\Delta}\in\mathcal{E}=\{\boldsymbol{\Delta}:\|\boldsymbol{\Delta}\|\le\varepsilon\}$. In this paper, we consider $\mathcal{E}$ defined by the weighted Frobenius norm $\|\cdot\|_F^{\mathbf{T}}$, i.e.,

$$\mathcal{E}\triangleq\{\boldsymbol{\Delta}:\|\boldsymbol{\Delta}\|_F^{\mathbf{T}}\le\varepsilon\}=\{\boldsymbol{\Delta}:\mathrm{Tr}(\boldsymbol{\Delta}\mathbf{T}\boldsymbol{\Delta}^H)\le\varepsilon^2\} \tag{6}$$

[1]There are also other stochastic models assuming that the covariance or both the mean and the covariance is available [8], [11], [13], [17].

where $\mathbf{T}$ is a given positive definite matrix. The ellipsoid reduces to a sphere when $\mathbf{T} = \mathbf{I}$, which is the most frequently used model [18]–[28].

B. Problem Formulation

The focus of this paper is on designing robust precoders for different types of CSI, so we consider simple pre-fixed receivers that are independent of the precoder design as in [29]–[31]. This separate structure lets the transmitter undertake the most computational workload to cope with imperfect CSI, thus decreasing the receiver's complexity and meanwhile increasing its backward compatibility. To be more exact, the receiver uses an equalizer that is a function of the channel $\mathbf{H}$, or the channel mean $\bar{\mathbf{H}}$, or the nominal channel $\hat{\mathbf{H}}$, but not the precoder $\mathbf{F}$. We formulate the precoder designs as the following problems.

1) *Perfect CSI:* With perfect CSI at both ends of the link, the optimal precoder is the solution to the problem

$$\underset{\mathbf{F}\in\mathcal{F}}{\text{minimize}}\ \|\mathbf{G}(\mathbf{H})\mathbf{H}\mathbf{F}-\mathbf{I}\|_F^2 \tag{7}$$

which will be solved as a warm-up for the later robust designs.

2) *Stochastic CSI:* Given the channel statistics, i.e., the channel mean, the precoding design is based on optimizing the average performance, hence leading to

$$\underset{\mathbf{F}\in\mathcal{F}}{\text{minimize}}\ E\left\{\|\mathbf{G}(\mathbf{H})\mathbf{H}\mathbf{F}-\mathbf{I}\|_F^2\right\} \tag{8}$$

and

$$\underset{\mathbf{F}\in\mathcal{F}}{\text{minimize}}\ E\left\{\|\bar{\mathbf{G}}(\bar{\mathbf{H}})\mathbf{H}\mathbf{F}-\mathbf{I}\|_F^2\right\} \tag{9}$$

where the expectation is taken on the channel $\mathbf{H}$ modeled by (3) and (4). Note that $\mathbf{G}(\mathbf{H})$ accounts for the perfect CSIR case, while $\bar{\mathbf{G}}(\bar{\mathbf{H}})$ corresponds to the case where both transmitter and receiver can only access the channel mean.

3) *Worst-case Design:* For deterministic imperfect CSI, the robustness is embodied by a guaranteed performance level for any channel realization in the uncertainty region. Therefore, one needs to find the optimal precoder in the worst channel, leading to

$$\min_{\mathbf{F}\in\mathcal{F}}\max_{\mathbf{H}\in\mathcal{H}}\ \|\hat{\mathbf{G}}(\hat{\mathbf{H}})\mathbf{H}\mathbf{F}-\mathbf{I}\|_F^2 \tag{10}$$

where $\mathbf{H}\in\mathcal{H}$ can be replaced by $\mathbf{\Delta}\in\mathcal{E}$ defined in (6), and $\hat{\mathbf{G}}(\hat{\mathbf{H}})$ corresponds to the situation where both transmitter and receiver are subject to deterministic imperfect CSI [26]–[28]. We note that for $\mathbf{G}(\mathbf{H})$, i.e., perfect CSIR, (10) is still an open problem. The focus of this paper is the worst-case robust design.

The transmit power constraint, in (7)–(10), is represented by $\mathbf{F}\in\mathcal{F}$. We first consider a general set $\mathcal{F}$ that is a nonempty compact convex set, covering all common power constraints as special cases. Then, some particular constraints are considered:

1) *Sum Power Constraint:* $\mathcal{F}_1 \triangleq \{\mathbf{F} : \mathrm{Tr}(\mathbf{F}\mathbf{F}^H) \le P_s\}$.
2) *Maximum Power Constraint:* $\mathcal{F}_2 \triangleq \{\mathbf{F} : \lambda_{\max}(\mathbf{F}\mathbf{F}^H) \le P_m\}$.
3) *Per-antenna Power Constraint:* $\mathcal{F}_3 \triangleq \{\mathbf{F} : \max_i [\mathbf{F}\mathbf{F}^H]_{ii} \le P_a\}$ or $\mathcal{F}_4 \triangleq \{\mathbf{F} : [\mathbf{F}\mathbf{F}^H]_{ii} \le P_{a,i}, i = 1,\ldots,N\}$.

III. Perfect CSI

Given a convex set $\mathcal{F}$, the formulation (7) is evidently a convex problem whose solution can be found in polynomial time. Nevertheless, in this section, we show that with perfect CSI, just like [3]–[7], the optimal transmit directions for the MMSE precoder with a pre-fixed receiver are equal to the right singular vectors of the equivalent channel under some mild conditions, which paves the path to finding the optimal transmit directions for robust precoders later. Then, we provide the closed-form solutions to the resulting power allocation problems. Although [29] and [30] have considered similar problems to (7), only suboptimal results were given.

A. Optimal Transmit Directions

For perfect CSI, the precoding design depends totally on the equivalent channel $\mathbf{W} \triangleq \mathbf{G}(\mathbf{H})\mathbf{H}$. Denote the singular value decomposition (SVD) of $\mathbf{W}$ by $\mathbf{W} = \mathbf{U}_w\mathbf{\Sigma}_w\mathbf{V}_w^H$ with $\mathbf{\Sigma}_w = [\mathbf{\Lambda}_w\ \mathbf{0}]$ and the diagonal matrix $\mathbf{\Lambda}_w \in \mathbb{R}^{L\times L}$ containing the singular values $w_1 \ge \cdots \ge w_L$ in decreasing order. Denote the SVD of $\mathbf{F}$ by $\mathbf{F} = \mathbf{U}_f\mathbf{\Sigma}_f\mathbf{V}_f^H$ with $\mathbf{\Sigma}_f = [\mathbf{\Lambda}_f\ \mathbf{0}]^T$ and the diagonal matrix $\mathbf{\Lambda}_f \in \mathbb{R}^{L\times L}$ containing the singular values $\{f_i\}_{i=1}^L$.

Theorem 1: Let $\mathcal{F} = \{\mathbf{F} : \psi_n(\boldsymbol{\lambda}(\mathbf{F}^H\mathbf{F})) \le P_n,\ \forall n\}$, where each $\psi_n(\mathbf{x})$ is a Schur-convex function and componentwise nondecreasing, and $\mathbf{G}(\mathbf{H})$ be a function of $\mathbf{H}$. Then, $\mathbf{U}_f = \mathbf{V}_w$ and $\mathbf{V}_f = \mathbf{U}_w$ are optimal for the problem (7).

Proof: By defining $\mathbf{V}_w^H\mathbf{F}\mathbf{U}_w \triangleq \hat{\mathbf{F}} = [\hat{\mathbf{F}}_1^T\ \hat{\mathbf{F}}_2^T]^T$ with $\hat{\mathbf{F}}_1 \in \mathbb{C}^{L\times L}$ and $\hat{\mathbf{F}}_2 \in \mathbb{C}^{(N-L)\times L}$, the problem (7) is equivalent to

$$\underset{\hat{\mathbf{F}}\in\mathcal{F}}{\text{minimize}}\ \|\mathbf{\Lambda}_w\hat{\mathbf{F}}_1-\mathbf{I}\|_F^2. \tag{11}$$

Let $\tilde{\mathbf{F}} = [\hat{\mathbf{F}}_1^T\ \mathbf{0}]^T$. Since $\tilde{\mathbf{F}}^H\tilde{\mathbf{F}} = \hat{\mathbf{F}}_1^H\hat{\mathbf{F}}_1 \preceq \hat{\mathbf{F}}_1^H\hat{\mathbf{F}}_1 + \hat{\mathbf{F}}_2^H\hat{\mathbf{F}}_2 = \hat{\mathbf{F}}^H\hat{\mathbf{F}}$, implying that $\boldsymbol{\lambda}(\tilde{\mathbf{F}}^H\tilde{\mathbf{F}}) \le \boldsymbol{\lambda}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})$, then if $\hat{\mathbf{F}}$ is feasible, so is $\tilde{\mathbf{F}}$. Hence, we consider without losing any optimality the following problem:

$$\begin{aligned}\underset{\hat{\mathbf{F}}_1}{\text{minimize}}\quad & \|\mathbf{\Lambda}_w\hat{\mathbf{F}}_1-\mathbf{I}\|_F^2\\ \text{subject to}\quad & \psi_n(\boldsymbol{\lambda}(\hat{\mathbf{F}}_1^H\hat{\mathbf{F}}_1)) \le P_n,\quad \forall n.\end{aligned} \tag{12}$$

Now, we show that the optimal $\hat{\mathbf{F}}_1$ can be made diagonal. Divide $\hat{\mathbf{F}}_1$ into two parts as $\hat{\mathbf{F}}_1 = \hat{\mathbf{D}}_1 + \hat{\mathbf{B}}_1$, where $\hat{\mathbf{D}}_1$ is a diagonal matrix containing the diagonal elements of $\hat{\mathbf{F}}_1$, and $\hat{\mathbf{B}}_1$ contains the off-diagonal elements. It follows that

$$\|\mathbf{\Lambda}_w\hat{\mathbf{F}}_1-\mathbf{I}\|_F^2 = \|\mathbf{\Lambda}_w\hat{\mathbf{D}}_1-\mathbf{I}+\mathbf{\Lambda}_w\hat{\mathbf{B}}_1\|_F^2 \ge \|\mathbf{\Lambda}_w\hat{\mathbf{D}}_1-\mathbf{I}\|_F^2. \tag{13}$$

Therefore, given any feasible $\hat{\mathbf{F}}_1$, one can always achieve a smaller objective value by using $\tilde{\mathbf{F}}_1 = \hat{\mathbf{D}}_1$. The only question is whether $\hat{\mathbf{D}}_1$ is feasible or not. Considering that $\mathbf{d}(\hat{\mathbf{D}}_1^H\hat{\mathbf{D}}_1) \le$

$\mathbf{d}(\hat{\mathbf{F}}_1^H\hat{\mathbf{F}}_1)$ and $\mathbf{d}(\hat{\mathbf{F}}_1^H\hat{\mathbf{F}}_1)$ is majorized by $\boldsymbol{\lambda}(\hat{\mathbf{F}}_1^H\hat{\mathbf{F}}_1)$ [6], [38], by exploring the Schur-convexity and componentwise nondecreasingness of $\psi_n(\mathbf{x})$, we have $\psi_n(\boldsymbol{\lambda}(\hat{\mathbf{D}}_1^H\hat{\mathbf{D}}_1)) \leq \psi_n(\boldsymbol{\lambda}(\hat{\mathbf{F}}_1^H\hat{\mathbf{F}}_1))$, so $\hat{\mathbf{D}}_1$ is feasible too. Consequently, we have proved the optimality of the diagonal structure for $\hat{\mathbf{F}}_1$, which can be achieved by setting $\mathbf{U}_f = \mathbf{V}_w$ and $\mathbf{V}_f = \mathbf{U}_w$, resulting in $\hat{\mathbf{F}}_1 = \mathbf{\Lambda}_f$. ■

Remark 1: The conditions in Theorem 1 are satisfied by the sum and maximum power constraints as well as their combination (intersection of $\mathcal{F}_1$ and $\mathcal{F}_2$), since $\mathrm{Tr}(\mathbf{F}\mathbf{F}^H) = \mathrm{Tr}(\mathbf{F}^H\mathbf{F})$ and $\lambda_{\max}(\mathbf{F}\mathbf{F}^H) = \lambda_{\max}(\mathbf{F}^H\mathbf{F})$ are both Schur-convex functions of the eigenvalues of $\mathbf{F}^H\mathbf{F}$ [6], [38]. Therefore, in the most common case, the optimal transmit directions (i.e., $\mathbf{U}_f$) are the right singular vectors of the equivalent channel $\mathbf{W}$ (i.e., $\mathbf{V}_w$). Denote the SVD of $\mathbf{H}$ by $\mathbf{H} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^H$ with the largest L singular values $\gamma_1 \geq \cdots \geq \gamma_L$ in decreasing order, and the SVD of $\mathbf{G}$ by $\mathbf{G}(\mathbf{H}) = \mathbf{U}_g\mathbf{\Sigma}_g\mathbf{V}_g^H$ with the singular values $\{g_i\}_{i=1}^L$. When the receiver uses no equalizer $\mathbf{G}(\mathbf{H}) = \mathbf{I}(L = M)$, the MF $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H(L = N)$, the ZF equalizer $\mathbf{G}(\mathbf{H}) = \mathbf{H}^\dagger(L = N)$, or the MMSE equalizer $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H(\sigma_n^2\mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1}(L = N)$, we have $\mathbf{V}_w = \mathbf{V}_h$ so that the optimal transmit directions are the right singular vectors of the channel $\mathbf{H}$, which is consistent with the results in [3]–[7]. Using Theorem 1, the matrix-valued problem (7) can be simplified to a scalar power allocation problem whose closed-form solution will be offered subsequently.

B. Optimal Power Allocation

Provided the conditions in Theorem 1 are satisfied, the problem (7) reduces to

$$\underset{\mathbf{f}\in\mathcal{L}}{\text{minimize}} \sum_{i=1}^{L}(w_i f_i - 1)^2 \tag{14}$$

where $\mathcal{L} = \{\mathbf{f} : \psi_n(f_1^2, \ldots, f_L^2) \leq P_n, \forall n\}$. Due to $w_i \geq 0$, the solution to (14) must be nonnegative, thus satisfying the nonnegativity of singular values. Since (14) is quite a simple problem, we write down directly its closed-form solution under different power constraints satisfying the conditions in Theorem 1 as follows.

1) *No Power Constraint* ($\mathcal{F} = \mathbb{C}^{N\times L}$): In this case, (7) is a simple least square (LS) problem leading to the well known solution $\mathbf{F}^\star = \mathbf{W}^\dagger$, implying that $f_i^\star = 1/w_i$.
2) *Maximum Power Constraint* ($\mathcal{F} = \mathcal{F}_2$): Given $\mathcal{L} = \{\mathbf{f} : f_i^2 \leq P_m, \forall i\}$, (14) decouples into L separate subproblems, each admitting the solution $f_i^\star = \min\{1/w_i, \sqrt{P_m}\}$.
3) *Sum Power Constraint* ($\mathcal{F} = \mathcal{F}_1$): Given $\mathcal{L} = \{\mathbf{f} : \sum_{i=1}^L f_i^2 \leq P_s\}$, the solution to (14) is $f_i^\star = w_i/(w_i^2 + \eta)$ with the smallest $\eta \geq 0$ satisfying $\sum_{i=1}^L (f_i^\star)^2 \leq P_s$.

Remark 2: Fundamentally, with a pre-fixed receiver, (7) is similar to a LS problem, and tends to invert the equivalent channel $\mathbf{W}$. Therefore, the optimal power allocation is mainly determined by the inverse singular values of the equivalent channel, suggesting that the MMSE precoder, similar to linear equalizers, may suffer from the singularity of the channel because there may not be enough transmit power to compensate for the singular channel. When the receiver adopts the common linear equalizers mentioned in Remark 1, the correspondence between the singular values of the equivalent channel $\mathbf{W}$ and the channel $\mathbf{H}$, i.e., $\{w_i\}$ and $\{\gamma_i\}$, is given by: no equalizer $w_i = \gamma_i$, the MF $w_i = \gamma_i^2$, the ZF equalizer $w_i = 1$, the MMSE equalizer $w_i = \gamma_i^2/(\sigma_n^2 + \gamma_i^2)$.

IV. Statistical CSI

The formulations (8) and (9), similar to the perfect CSI case, are also convex problems given that $\mathcal{F}$ is a convex set, and thus can be efficiently solved. Interestingly, it is has been shown in [8]–[16] that, with perfect CSIR and statistical CSIT of mean feedback, the transmit directions of the optimal precoder are the right singular vectors of the channel mean under various criteria. In this section, we will show that the same transmit directions are also optimal for the MMSE precoder with a pre-fixed receiver that has either perfect or imperfect statistical CSIR. Then, we derive the closed-form solutions to the resulting power allocation problems.

A. Optimal Transmit Directions

Denote the SVD of the channel mean by $\bar{\mathbf{H}} = \mathbf{U}_{\bar{h}}\mathbf{\Sigma}_{\bar{h}}\mathbf{V}_{\bar{h}}^H$ with the largest L singular values $\bar{\gamma}_1 \geq \cdots \geq \bar{\gamma}_L$ in decreasing order. Let $\bar{\mathbf{M}} \triangleq \mathbf{G}(\bar{\mathbf{H}})\bar{\mathbf{H}}$ where $\mathbf{G}(\bar{\mathbf{H}})$ is the result of replacing $\mathbf{H}$ by $\bar{\mathbf{H}}$ in $\mathbf{G}(\mathbf{H})$, and denote its SVD by $\bar{\mathbf{M}} = \mathbf{U}_{\bar{m}}\mathbf{\Sigma}_{\bar{m}}\mathbf{V}_{\bar{m}}^H$. Let $\bar{\mathbf{W}} \triangleq \bar{\mathbf{G}}(\bar{\mathbf{H}})\bar{\mathbf{H}}$ and denote its SVD by $\bar{\mathbf{W}} = \mathbf{U}_{\bar{w}}\mathbf{\Sigma}_{\bar{w}}\mathbf{V}_{\bar{w}}^H$ with the largest L singular values $\bar{w}_1 \geq \cdots \geq \bar{w}_L$ in decreasing order.

Theorem 2: Let $\mathcal{F} = \{\mathbf{F} : \psi_n(\boldsymbol{\lambda}(\mathbf{F}^H\mathbf{F})) \leq P_n, \forall n\}$, where each $\psi_n(\mathbf{x})$ is a Schur-convex function and componentwise nondecreasing. Let $\mathbf{G}(\mathbf{H}) = \mathbf{I}, \mathbf{H}^H, \mathbf{H}^\dagger$, or $\mathbf{H}^H(\sigma_n^2\mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1}$, and $\bar{\mathbf{G}}(\bar{\mathbf{H}})$ be a function of $\bar{\mathbf{H}}$. Then, $\mathbf{U}_f = \mathbf{V}_{\bar{m}}$ and $\mathbf{V}_f = \mathbf{U}_{\bar{m}}$ are optimal for the problem (8), and $\mathbf{U}_f = \mathbf{V}_{\bar{w}}$ and $\mathbf{V}_f = \mathbf{U}_{\bar{w}}$ are optimal for the problem (9).

Proof: We consider first the problem (8). It is easy to see that $\mathbf{V}_{\bar{m}} = \mathbf{V}_{\bar{h}}$ and $\mathbf{U}_{\bar{m}} = \mathbf{U}_{\bar{h}}$ for $\mathbf{G}(\mathbf{H}) = \mathbf{I}$, and $\mathbf{V}_{\bar{m}} = \mathbf{U}_{\bar{m}} = \mathbf{V}_{\bar{h}}$ for $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H, \mathbf{H}^\dagger$, or $\mathbf{H}^H(\sigma_n^2\mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1}$. When $\mathbf{G}(\mathbf{H}) = \mathbf{I}$, using the statistical model (3) and (4), the objective in (8) can be expressed as

$$E\left\{\|\mathbf{H}\mathbf{F} - \mathbf{I}\|_F^2\right\} = \|\bar{\mathbf{H}}\mathbf{F} - \mathbf{I}\|_F^2 + M\|\mathbf{F}\|_F^2. \tag{15}$$

Then, following the Proof of Theorem 1, one can easily see that $\mathbf{U}_f = \mathbf{V}_{\bar{h}}$ and $\mathbf{V}_f = \mathbf{U}_{\bar{h}}$ are optimal for (8). When $\mathbf{G}(\mathbf{H}) = \mathbf{H}^\dagger$, by defining $\hat{\mathbf{F}} \triangleq \mathbf{V}_{\bar{h}}^H\mathbf{F}\mathbf{V}_{\bar{h}}$, the objective of (8) becomes $\|\mathbf{\Lambda}_r\hat{\mathbf{F}} - \mathbf{I}\|_F^2$ where $\mathbf{\Lambda}_r = \mathrm{diag}\{\mathbf{I}_r, \mathbf{0}\}$ and $r = \mathrm{rank}(\bar{\mathbf{H}})$. Then, from (13), the optimal $\hat{\mathbf{F}}$ has a diagonal structure, which can be achieved by setting $\mathbf{U}_f = \mathbf{V}_f = \mathbf{V}_{\bar{h}}$. When $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H$ or $\mathbf{H}^H(\sigma_n^2\mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1}$, however, the proof is not so straightforward because there is no explicit expression of $E\{\|\mathbf{G}(\mathbf{H})\mathbf{H}\mathbf{F} - \mathbf{I}\|_F^2\}$. We can find the optimal transmit directions by using the following lemma.

Lemma 1 [21]: Let $\mathbf{J} \in \mathbb{R}^{m\times m}$ be a diagonal matrix with the diagonal elements being ± 1. There are $K = 2^m$ different such matrices indexed from $k = 1$ to K. Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be an arbitrary matrix and $\mathbf{D}_\mathbf{A}$ be a diagonal matrix such that $[\mathbf{D}_\mathbf{A}]_{ii} = [\mathbf{A}]_{ii}, \forall i$. Then, $\mathbf{D}_\mathbf{A} = (1/K)\sum_{k=1}^K \mathbf{J}_k\mathbf{A}\mathbf{J}_k$.

Now, for $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H$, by using $\hat{\mathbf{F}} \triangleq \mathbf{V}_{\bar{h}}^H \mathbf{F} \mathbf{V}_{\bar{h}}$, the objective of (8) becomes

$$\begin{aligned} &E\left\{\|\mathbf{G}(\mathbf{H})\mathbf{H}\mathbf{F} - \mathbf{I}\|_F^2\right\} \\ &= E\{\|\mathbf{H}^H\mathbf{H}\mathbf{V}_{\bar{h}}\hat{\mathbf{F}}\mathbf{V}_{\bar{h}}^H - \mathbf{I}\|_F^2\} \\ &= E\{\|\mathbf{V}_{\bar{h}}^H(\bar{\mathbf{H}}+\mathbf{E})^H(\bar{\mathbf{H}}+\mathbf{E})\mathbf{V}_{\bar{h}}\hat{\mathbf{F}} - \mathbf{I}\|_F^2\} \\ &= E\{\|\left(\boldsymbol{\Sigma}_{\bar{h}} + \mathbf{U}_{\bar{h}}^H\mathbf{E}\mathbf{V}_{\bar{h}}\right)^H \\ &\quad \times \left(\boldsymbol{\Sigma}_{\bar{h}} + \mathbf{U}_{\bar{h}}^H\mathbf{E}\mathbf{V}_{\bar{h}}\right)\hat{\mathbf{F}} - \mathbf{I}\|_F^2\} \\ &= E\{\|(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})^H(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})\hat{\mathbf{F}} - \mathbf{I}\|_F^2\} \\ &= S(\hat{\mathbf{F}}) \end{aligned} \tag{16}$$

where we use the fact that $\mathbf{U}_{\bar{h}}^H\mathbf{E}\mathbf{V}_{\bar{h}}$ has the same distribution as $\mathbf{E}$. Let $\mathbf{J}_N \in \mathbb{R}^{N\times N}$ be a diagonal matrix with the diagonal elements being ± 1, and $\mathbf{J}_M = \mathrm{diag}\{\mathbf{J}_N, \mathbf{J}_{M-N}\} \in \mathbb{R}^{M\times M}$ where $\mathbf{J}_{M-N} \in \mathbb{R}^{(M-N)\times(M-N)}$ is also a diagonal matrix with the diagonal elements being ± 1. It follows that

$$\begin{aligned} &S(\mathbf{J}_N\hat{\mathbf{F}}\mathbf{J}_N) \\ &= E\{\|(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})^H(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})\mathbf{J}_N\hat{\mathbf{F}}\mathbf{J}_N - \mathbf{J}_N^2\|_F^2\} \\ &= E\left\{\|(\boldsymbol{\Sigma}_{\bar{h}}\mathbf{J}_N + \mathbf{E}\mathbf{J}_N)^H(\boldsymbol{\Sigma}_{\bar{h}}\mathbf{J}_N + \mathbf{E}\mathbf{J}_N)\hat{\mathbf{F}} - \mathbf{I}\|_F^2\right\} \\ &= E\left\{\|(\mathbf{J}_M\boldsymbol{\Sigma}_{\bar{h}} + \mathbf{J}_M\mathbf{E})^H(\mathbf{J}_M\boldsymbol{\Sigma}_{\bar{h}} + \mathbf{J}_M\mathbf{E})\hat{\mathbf{F}} - \mathbf{I}\|_F^2\right\} \\ &= E\left\{\|(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})^H(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})\hat{\mathbf{F}} - \mathbf{I}\|_F^2\right\} \\ &= S(\hat{\mathbf{F}}) \end{aligned} \tag{17}$$

where we use the property that $\mathbf{J}$ is an orthogonal diagonal matrix, implying that $\mathbf{J}^{-1} = \mathbf{J}$, $\mathbf{J}^2 = \mathbf{I}$, $\boldsymbol{\Sigma}_{\bar{h}}\mathbf{J}_N = \mathbf{J}_M\boldsymbol{\Sigma}_{\bar{h}}$, and $\mathbf{E}$, $\mathbf{E}\mathbf{J}_N$, $\mathbf{J}_M\mathbf{E}$ share the same distribution. Therefore, $S(\hat{\mathbf{F}})$ is invariant with respect to the transformation $\hat{\mathbf{F}} \to \mathbf{J}\hat{\mathbf{F}}\mathbf{J}$ (the subscript N is suppressed). Since $S(\hat{\mathbf{F}})$ is a convex function, we have

$$S(\hat{\mathbf{F}}) = \frac{1}{K}\sum_{k=1}^{K} S(\mathbf{J}_k\hat{\mathbf{F}}\mathbf{J}_k) \geq S\left(\frac{1}{K}\sum_{k=1}^{K}\mathbf{J}_k\hat{\mathbf{F}}\mathbf{J}_k\right) = S(\mathbf{D}_{\hat{\mathbf{F}}}) \tag{18}$$

where, from Lemma 1, $\mathbf{D}_{\hat{\mathbf{F}}}$ is a diagonal matrix such that $[\mathbf{D}_{\hat{\mathbf{F}}}]_{ii} = [\hat{\mathbf{F}}]_{ii}$, $\forall i$. Hence, $\mathbf{D}_{\hat{\mathbf{F}}}$ leads to a smaller objective value than $\hat{\mathbf{F}}$. Moreover, from the Proof of Theorem 1, if $\hat{\mathbf{F}}$ is feasible, so is $\mathbf{D}_{\hat{\mathbf{F}}}$. Consequently, the optimal $\hat{\mathbf{F}}$ should be a diagonal matrix that can be achieved by setting $\mathbf{U}_f = \mathbf{V}_f = \mathbf{V}_{\bar{h}}$, leading to $\mathbf{D}_{\hat{\mathbf{F}}} = \boldsymbol{\Lambda}_f$. For $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H(\sigma_n^2\mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1}$, one can proceed in a similar way.

Finally, we consider the problem (9), whose objective can be explicitly written as

$$\begin{aligned} &E\left\{\|\bar{\mathbf{G}}(\bar{\mathbf{H}})\mathbf{H}\mathbf{F} - \mathbf{I}\|_F^2\right\} \\ &\qquad = \|\bar{\mathbf{G}}(\bar{\mathbf{H}})\bar{\mathbf{H}}\mathbf{F} - \mathbf{I}\|_F^2 + \|\bar{\mathbf{G}}(\bar{\mathbf{H}})\|_F^2\|\mathbf{F}\|_F^2. \end{aligned} \tag{19}$$

Therefore, following the Proof of Theorem 1 again, one can easily find that $\mathbf{U}_f = \mathbf{V}_{\bar{w}}$ and $\mathbf{V}_f = \mathbf{U}_{\bar{w}}$ are optimal. ■

B. Optimal Power Allocation

Provided the conditions in Theorem 2 are satisfied, both (8) and (9) can be simplified to a scalar power allocation problem. The problem (9), as well as (8) when $\mathbf{G}(\mathbf{H}) = \mathbf{I}$, reduces to

$$\underset{\mathbf{f}\in\mathcal{L}}{\mathrm{minimize}} \sum_{i=1}^{L}\left((a_i f_i - 1)^2 + b f_i^2\right) \tag{20}$$

where $a_i = \bar{w}_i$, $b = \|\mathbf{G}(\bar{\mathbf{H}})\|_F^2$ for (9), and $a_i = \bar{\gamma}_i$, $b = M$ for (8). When $\mathbf{G}(\mathbf{H}) = \mathbf{H}^\dagger$, (8) reduces to

$$\underset{\mathbf{f}\in\mathcal{L}}{\mathrm{minimize}} \sum_{i=1}^{r}(f_i - 1)^2 \tag{21}$$

where $r = \mathrm{rank}(\bar{\mathbf{H}})$. When $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H$ or $\mathbf{H}^H(\sigma_n^2\mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1}$, the simplified problem of (8) is

$$\underset{\mathbf{f}\in\mathcal{L}}{\mathrm{minimize}}\ E\left\{\|\mathbf{R}(\mathbf{E})\boldsymbol{\Lambda}_f - \mathbf{I}\|_F^2\right\} = \sum_{i=1}^{L}\left(a_i f_i^2 - 2b_i f_i + 1\right) \tag{22}$$

where $\mathbf{R}(\mathbf{E}) = (\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})^H(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})$ for $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H$, and $\mathbf{R}(\mathbf{E}) = (\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})^H(\sigma_n^2\mathbf{I}+(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})^H)^{-1}(\boldsymbol{\Sigma}_{\bar{h}}+\mathbf{E})$ for $\mathbf{G}(\mathbf{H}) = \mathbf{H}^H(\sigma_n^2\mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1}$, and $a_i = [E\{\mathbf{R}^2(\mathbf{E})\}]_{i,i}$, $b_i = [E\{\mathbf{R}(\mathbf{E})\}]_{i,i}$. The closed-form solutions to the above problems are given in the following.

1) *No Power Constraint* ($\mathcal{F} = \mathbb{C}^{N\times L}$): The solution to (20) is $f_i^\star = a_i/(a_i^2 + b)$; the solution to (21) is $f_i^\star = 1$ for $i \leq r$ and $f_i^\star = 0$ for $i > r$; the solution to (22) is $f_i^\star = b_i/a_i$.
2) *Maximum Power Constraint* ($\mathcal{F} = \mathcal{F}_2$): Given $\mathcal{L} = \{\mathbf{f} : f_i^2 \leq P_m,\ \forall i\}$, the solution to (20) is $f_i^\star = \min\{a_i/(a_i^2 + b), \sqrt{P_m}\}$; the solution to (21) is $f_i^\star = \min\{1, \sqrt{P_m}\}$ for $i \leq r$ and $f_i^\star = 0$ for $i > r$; the solution to (22) is $f_i^\star = \min\{b_i/a_i, \sqrt{P_m}\}$.
3) *Sum Power Constraint* ($\mathcal{F} = \mathcal{F}_1$): Given $\mathcal{L} = \{\mathbf{f} : \sum_{i=1}^{L} f_i^2 \leq P_s\}$, the solution to (20) is $f_i^\star = a_i/(a_i^2 + b + \eta)$ with the smallest $\eta \geq 0$ satisfying $\sum_{i=1}^{L}(f_i^\star)^2 \leq P_s$; the solution to (21) is $f_i^\star = \min(1, P_s/r)$ for $i \leq r$ and $f_i^\star = 0$ for $i > r$; the solution to (22) is $f_i^\star = b_i/(a_i + \eta)$ the smallest $\eta \geq 0$ satisfying $\sum_{i=1}^{L}(f_i^\star)^2 \leq P_s$.

V. Worst-Case Design

The minimax problem (10), by using the channel error definition in (5), can be written as

$$\min_{\mathbf{F}\in\mathcal{F}}\max_{\boldsymbol{\Delta}\in\mathcal{E}} \|\hat{\mathbf{G}}(\hat{\mathbf{H}})(\hat{\mathbf{H}} + \boldsymbol{\Delta})\mathbf{F} - \mathbf{I}\|_F^2 \tag{23}$$

which corresponds to the situation where both the transmitter and the receiver are subject to deterministic imperfect CSI [26]–[28]. Note that the worst-case robust MMSE precoder design with perfect CSIR, i.e., replacing $\hat{\mathbf{G}}(\hat{\mathbf{H}})$ by $\mathbf{G}(\mathbf{H})$ in (23), is still an open problem. As a special case of our work, [22] has considered a spherical uncertainty region ($\mathbf{T} = \mathbf{I}$) without any power constraint ($\mathcal{F} = \mathbb{C}^{N\times L}$).

It is not difficult to see that the objective value of (23) is upper bounded by L, which is achieved by $\boldsymbol{\Delta} = -\hat{\mathbf{H}}$ when $\varepsilon \geq \|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$. This means that if the channel uncertainty is too large, then there is no guarantee of any performance in the worst channel. Therefore, the worst-case design is meaningful for $\varepsilon < \|\hat{\mathbf{H}}\|_F^{\mathbf{T}}$, a reasonable assumption in practice since large channel uncertainty suggests that the quality of CSI is too poor to be used.

Different from the cases of perfect and statistical CSI, the problem (23), at first glance, is not convex and cannot be easily solved. We will first show that (23) can be equivalently transformed into a convex problem or even further an SDP [34] that can be efficiently solved. In light of the optimality of the eigenmode transmission for perfect and statistical CSI, one may wonder whether this favorable property still exists in the worst-case design. In this section, we will prove the most important result in this paper: for the worst-case design, the optimal transmit directions are the right singular vectors of the nominal channel under some mild conditions.

A. Convex Reformulation

We first show that, under a general power constraint, the minimax problem (23) can be equivalently transformed into a convex problem or even further an SDP [34] that can be efficiently solved by some numerical methods, for example the interior point method [33].

Proposition 1: Let $\hat{\mathbf{G}}(\hat{\mathbf{H}})$ be a function of $\hat{\mathbf{H}}$. Then, the minimax problem (23) is equivalent to the following problem:

$$\begin{aligned}\underset{\mathbf{F},\lambda,t}{\text{minimize}}\quad & t\\ \text{subject to}\quad & \mathbf{F}\in\mathcal{F}\\ & \begin{bmatrix} t-\lambda & \mathbf{y}^H & \mathbf{0}\\ \mathbf{y} & \mathbf{I} & \varepsilon\mathbf{Z}\\ \mathbf{0} & \varepsilon\mathbf{Z}^H & \lambda\mathbf{I}\end{bmatrix}\succeq 0\end{aligned}\tag{24}$$

where $\mathbf{y}\triangleq\text{vec}(\hat{\mathbf{G}}(\hat{\mathbf{H}})\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})$ and $\mathbf{Z}\triangleq\mathbf{F}^T\mathbf{T}^{-1/2}\otimes\hat{\mathbf{G}}(\hat{\mathbf{H}})$.

Proof: We first introduce some useful lemmas.

Lemma 2 (Schur's Complement [39]): Let

$$\mathbf{M}=\begin{bmatrix}\mathbf{A} & \mathbf{B}^H\\ \mathbf{B} & \mathbf{C}\end{bmatrix}$$

be a Hermitian matrix. Then, $\mathbf{M}\succeq 0$ if and only if $\mathbf{A}-\mathbf{B}^H\mathbf{C}^{-1}\mathbf{B}\succeq 0$ (assuming $\mathbf{C}\succ 0$), or $\mathbf{C}-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H\succeq 0$ (assuming $\mathbf{A}\succ 0$).

Lemma 3 [21]: Given matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ with $\mathbf{A}=\mathbf{A}^H$, then

$$\mathbf{A}\succeq\mathbf{B}^H\mathbf{X}\mathbf{C}+\mathbf{C}^H\mathbf{X}^H\mathbf{B},\quad\forall\mathbf{X}:\|\mathbf{X}\|_2\le\varepsilon$$

if and only if there exists λ such that

$$\begin{bmatrix}\mathbf{A}-\lambda\mathbf{C}^H\mathbf{C} & -\varepsilon\mathbf{B}^H\\ -\varepsilon\mathbf{B} & \lambda\mathbf{I}\end{bmatrix}\succeq 0.$$

The minimax problem (23) is equivalent to

$$\begin{aligned}\underset{\mathbf{F}\in\mathcal{F},t}{\text{minimize}}\quad & t\\ \text{subject to}\quad & \|\hat{\mathbf{G}}(\hat{\mathbf{H}})(\hat{\mathbf{H}}+\hat{\boldsymbol{\Delta}}\mathbf{T}^{-1/2})\mathbf{F}-\mathbf{I}\|_F^2\le t,\\ & \forall\hat{\boldsymbol{\Delta}}:\|\hat{\boldsymbol{\Delta}}\|_F\le\varepsilon\end{aligned}\tag{25}$$

where $\hat{\boldsymbol{\Delta}}\triangleq\boldsymbol{\Delta}\mathbf{T}^{1/2}$ and

$$\begin{aligned}&\|\hat{\mathbf{G}}(\hat{\mathbf{H}})(\hat{\mathbf{H}}+\hat{\boldsymbol{\Delta}}\mathbf{T}^{-1/2})\mathbf{F}-\mathbf{I}\|_F^2\\ &=\|\text{vec}(\hat{\mathbf{G}}(\hat{\mathbf{H}})(\hat{\mathbf{H}}+\hat{\boldsymbol{\Delta}}\mathbf{T}^{-1/2})\mathbf{F}-\mathbf{I})\|^2\\ &=\|\text{vec}(\hat{\mathbf{G}}(\hat{\mathbf{H}})\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})+\text{vec}(\hat{\mathbf{G}}(\hat{\mathbf{H}})\hat{\boldsymbol{\Delta}}\mathbf{T}^{-1/2}\mathbf{F})\|^2\\ &=\|\mathbf{y}+\mathbf{Z}\boldsymbol{\delta}\|^2=(\mathbf{y}+\mathbf{Z}\boldsymbol{\delta})^H(\mathbf{y}+\mathbf{Z}\boldsymbol{\delta})\end{aligned}\tag{26}$$

with $\mathbf{y}$ and $\mathbf{Z}$ defined in Proposition 1 and $\boldsymbol{\delta}\triangleq\text{vec}(\hat{\boldsymbol{\Delta}})$. From Lemma 2, the constraint in (25) can be rewritten as

$$\begin{bmatrix} t & (\mathbf{y}+\mathbf{Z}\boldsymbol{\delta})^H\\ \mathbf{y}+\mathbf{Z}\boldsymbol{\delta} & \mathbf{I}\end{bmatrix}\succeq 0,\quad\forall\boldsymbol{\delta}:\|\boldsymbol{\delta}\|\le\varepsilon\tag{27}$$

which can be alternatively expressed as

$$\mathbf{A}\succeq\mathbf{B}^H\boldsymbol{\delta}\mathbf{c}+\mathbf{c}^H\boldsymbol{\delta}^H\mathbf{B},\quad\forall\boldsymbol{\delta}:\|\boldsymbol{\delta}\|\le\varepsilon\tag{28}$$

where

$$\mathbf{A}\triangleq\begin{bmatrix} t & \mathbf{y}^H\\ \mathbf{y} & \mathbf{I}\end{bmatrix},\quad\mathbf{B}\triangleq\begin{bmatrix}\mathbf{0} & -\mathbf{Z}^H\end{bmatrix},\quad\mathbf{c}\triangleq[1\quad\mathbf{0}^T].$$

From Lemma 3, (28) holds if and only if there exists λ such that

$$\begin{bmatrix} t-\lambda & \mathbf{y}^H & \mathbf{0}\\ \mathbf{y} & \mathbf{I} & \varepsilon\mathbf{Z}\\ \mathbf{0} & \varepsilon\mathbf{Z}^H & \lambda\mathbf{I}\end{bmatrix}\succeq 0\tag{29}$$

indicating the equivalence between (25) and (24). ■

Remark 3: Clearly, if $\mathcal{F}$ is a convex set, then (24) is a convex problem. Furthermore, when $\mathcal{F}$ equals $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ or $\mathcal{F}_4$, or any combination (intersection) of them, (24) can be easily transformed into an SDP [34], a very tractable form of convex optimization. Note that a similar convex formulation was independently obtained in [27] and [28], which only focused on solving problems through SDPs. For us, the reformulation (24) will serve as an intermediate step to find the optimal transmit directions for the worst-case robust MMSE precoder.

B. Optimal Transmit Directions

Denote the SVDs of $\mathbf{F}$ and $\hat{\mathbf{G}}(\hat{\mathbf{H}})$ by $\mathbf{F}=\mathbf{U}_f\boldsymbol{\Sigma}_f\mathbf{V}_f^H$ and $\hat{\mathbf{G}}(\hat{\mathbf{H}})=\mathbf{U}_g\boldsymbol{\Sigma}_g\mathbf{V}_g^H$ with $\boldsymbol{\Sigma}_f=[\boldsymbol{\Lambda}_f\ \mathbf{0}]^T$ and $\boldsymbol{\Sigma}_g=[\boldsymbol{\Lambda}_g\ \mathbf{0}]$ where the diagonal matrices $\boldsymbol{\Lambda}_f,\boldsymbol{\Lambda}_g\in\mathbb{R}^{L\times L}$ contain the singular values $\{f_i\}_{i=1}^L$ and $\{g_i\}_{i=1}^L$, respectively. Denote the SVD of $\hat{\mathbf{H}}$ by $\hat{\mathbf{H}}=\mathbf{U}_{\hat{h}}\boldsymbol{\Sigma}_{\hat{h}}\mathbf{V}_{\hat{h}}^H$ and let $\boldsymbol{\Lambda}_{\hat{h}}\in\mathbb{R}^{L\times L}$ be a diagonal matrix containing the largest L singular values $\hat{\gamma}_1\ge\cdots\ge\hat{\gamma}_L$ in decreasing order. Denote the eigenvalue decomposition (EVD) of $\mathbf{T}$ by $\mathbf{T}=\mathbf{U}_t\boldsymbol{\Lambda}_t\mathbf{U}_t^H$ with the eigenvalues $\tau_1^2\ge\cdots\ge\tau_N^2$ in decreasing order and $\boldsymbol{\Lambda}_t=\text{diag}\{\boldsymbol{\Lambda}_{t,1},\boldsymbol{\Lambda}_{t,2}\}$ with $\boldsymbol{\Lambda}_{t,1}\in\mathbb{R}^{L\times L}$ and $\boldsymbol{\Lambda}_{t,2}\in\mathbb{R}^{(N-L)\times(N-L)}$.

Theorem 3: Let $\mathcal{F}=\{\mathbf{F}:\psi_n(\boldsymbol{\lambda}(\mathbf{F}^H\mathbf{F}))\le P_n,\forall n\}$, where each $\psi_n(\mathbf{x})$ is a Schur-convex function and componentwise nondecreasing, and $\hat{\mathbf{G}}(\hat{\mathbf{H}})$ be a function of $\hat{\mathbf{H}}$. Let $\mathbf{V}_g=\mathbf{U}_{\hat{h}}$ and $\mathbf{U}_t=\mathbf{V}_{\hat{h}}$. Then, $\mathbf{U}_f=\mathbf{V}_{\hat{h}}$ and $\mathbf{V}_f=\mathbf{U}_g$ are optimal for the minimax problem (23).

Proof: Through defining $\hat{\boldsymbol{\Delta}}\triangleq\boldsymbol{\Delta}\mathbf{T}^{1/2}$ and $\hat{\mathbf{F}}\triangleq\mathbf{V}_{\hat{h}}^H\mathbf{F}\mathbf{U}_g$, and using $\mathbf{V}_g=\mathbf{U}_{\hat{h}}$ and $\mathbf{U}_t=\mathbf{V}_{\hat{h}}$, the minimax problem (23) can be expressed as

$$\min_{\hat{\mathbf{F}}\in\mathcal{F}}\max_{\|\hat{\boldsymbol{\Delta}}\|_F\le\varepsilon}\|\boldsymbol{\Sigma}_g(\boldsymbol{\Sigma}_{\hat{h}}+\mathbf{U}_{\hat{h}}^H\hat{\boldsymbol{\Delta}}\mathbf{V}_{\hat{h}}\boldsymbol{\Lambda}_t^{-1/2})\hat{\mathbf{F}}-\mathbf{I}\|_F^2\tag{30}$$

which, by introducing $\tilde{\mathbf{\Delta}} \triangleq \mathbf{U}_{\hat{h}}^H \hat{\mathbf{\Delta}} \mathbf{V}_{\hat{h}}$, amounts to

$$\min_{\hat{\mathbf{F}} \in \mathcal{F}} \max_{\|\tilde{\mathbf{\Delta}}\|_F \le \varepsilon} \|\mathbf{\Sigma}_g(\mathbf{\Sigma}_{\hat{h}} + \tilde{\mathbf{\Delta}}\mathbf{\Lambda}_t^{-1/2})\hat{\mathbf{F}} - \mathbf{I}\|_F^2. \quad (31)$$

Following the proof of Proposition 1, (31) can also be equivalently transformed into the form of (24) with $\mathbf{y} \triangleq \mathrm{vec}(\mathbf{\Sigma}_g \mathbf{\Sigma}_{\hat{h}} \hat{\mathbf{F}} - \mathbf{I})$ and $\mathbf{Z} \triangleq \hat{\mathbf{F}}^T \mathbf{\Lambda}_t^{-1/2} \otimes \mathbf{\Sigma}_g$.

Divide $\hat{\mathbf{F}}$ into two parts as $\hat{\mathbf{F}} = [\hat{\mathbf{F}}_1^T \ \hat{\mathbf{F}}_2^T]^T$ with $\hat{\mathbf{F}}_1 \in \mathbb{C}^{L\times L}$ and $\hat{\mathbf{F}}_2 \in \mathbb{C}^{(N-L)\times L}$. Then, it follows that $\mathbf{\Sigma}_g \mathbf{\Sigma}_{\hat{h}} \hat{\mathbf{F}} - \mathbf{I} = \mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}} \hat{\mathbf{F}}_1 - \mathbf{I}$ and $\hat{\mathbf{F}}^T \mathbf{\Lambda}_t^{-1/2} \otimes \mathbf{\Sigma}_g = [\hat{\mathbf{F}}_1^T \mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g \ \hat{\mathbf{F}}_2^T \mathbf{\Lambda}_{t,2}^{-1/2} \otimes \mathbf{\Sigma}_g]$. Consequently, the equivalent form of (24) can be explicitly written as (32), shown at the bottom of the page. Observe that if $\hat{\mathbf{F}}$ satisfies the linear matrix inequality (LMI) in (32), so does $\tilde{\mathbf{F}} \triangleq [\hat{\mathbf{F}}_1^T \ \mathbf{0}]^T$. Following the Proof of Theorem 1, it can be shown that $\tilde{\mathbf{F}} \in \mathcal{F}$, which implies that we can set $\hat{\mathbf{F}}_2 = \mathbf{0}$ without losing any optimality and focus on the following problem:

$$\begin{aligned} &\underset{\hat{\mathbf{F}}_1, \lambda, t}{\text{minimize}} \quad t \\ &\text{subject to} \quad \psi_n(\boldsymbol{\lambda}(\hat{\mathbf{F}}_1^H \hat{\mathbf{F}}_1)) \le P_n, \quad \forall n \\ &\qquad \begin{bmatrix} t-\lambda & \hat{\mathbf{y}}^H & \mathbf{0} \\ \hat{\mathbf{y}} & \mathbf{I} & \varepsilon\hat{\mathbf{Z}} \\ \mathbf{0} & \varepsilon\hat{\mathbf{Z}}^H & \lambda\mathbf{I} \end{bmatrix} \succeq 0. \end{aligned} \quad (33)$$

where $\hat{\mathbf{y}} \triangleq \mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}} \hat{\mathbf{F}}_1 - \mathbf{I})$ and $\hat{\mathbf{Z}} = \hat{\mathbf{F}}_1^T \mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g$.

Now we show that the optimal $\hat{\mathbf{F}}_1$ can be diagonal. Let $\mathbf{J}_L \in \mathbb{R}^{L\times L}$ be a diagonal matrix with the diagonal elements being ± 1, and $\mathbf{J}_M = \mathrm{diag}\{\mathbf{J}_L, \mathbf{J}_{M-L}\} \in \mathbb{R}^{M\times M}$ where $\mathbf{J}_{M-L} \in \mathbb{R}^{(M-L)\times(M-L)}$ is also a diagonal matrix with the diagonal elements being ± 1. Replacing $\hat{\mathbf{F}}_1$ by $\mathbf{J}_L \hat{\mathbf{F}}_1 \mathbf{J}_L$ in $\hat{\mathbf{y}}$ and $\hat{\mathbf{Z}}$ leads to

$$\begin{aligned} \hat{\mathbf{y}}_{\mathbf{J}} &= \mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}} \mathbf{J}_L \hat{\mathbf{F}}_1 \mathbf{J}_L - \mathbf{I}) \\ &= \mathrm{vec}(\mathbf{J}_L(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}} \hat{\mathbf{F}}_1 - \mathbf{I})\mathbf{J}_L) \\ &= (\mathbf{J}_L \otimes \mathbf{J}_L)\mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}} \hat{\mathbf{F}}_1 - \mathbf{I}) \\ &= (\mathbf{J}_L \otimes \mathbf{J}_L)\hat{\mathbf{y}}, \end{aligned} \quad (34)$$

and

$$\begin{aligned} \hat{\mathbf{Z}}_{\mathbf{J}} &= \mathbf{J}_L \hat{\mathbf{F}}_1^T \mathbf{J}_L \mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g \\ &= \mathbf{J}_L \hat{\mathbf{F}}_1^T \mathbf{\Lambda}_{t,1}^{-1/2} \mathbf{J}_L \otimes \mathbf{J}_L^2 \mathbf{\Sigma}_g \\ &= \mathbf{J}_L \hat{\mathbf{F}}_1^T \mathbf{\Lambda}_{t,1}^{-1/2} \mathbf{J}_L \otimes \mathbf{J}_L \mathbf{\Sigma}_g \mathbf{J}_M \\ &= (\mathbf{J}_L \otimes \mathbf{J}_L)(\hat{\mathbf{F}}_1^T \mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g)(\mathbf{J}_L \otimes \mathbf{J}_M) \\ &= (\mathbf{J}_L \otimes \mathbf{J}_L)\hat{\mathbf{Z}}(\mathbf{J}_L \otimes \mathbf{J}_M), \end{aligned} \quad (35)$$

where we use the properties that $\mathbf{J}$ is a diagonal matrix, $\mathbf{J}^2 = \mathbf{I}$ and $\mathbf{J}_L \mathbf{\Sigma}_g = \mathbf{\Sigma}_g \mathbf{J}_M$. Hence, the LMI in (33) amounts to

$$\begin{aligned} &\begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_L \otimes \mathbf{J}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_L \otimes \mathbf{J}_M \end{bmatrix} \begin{bmatrix} t-\lambda & \hat{\mathbf{y}}^H & \mathbf{0} \\ \hat{\mathbf{y}} & \mathbf{I} & \varepsilon\hat{\mathbf{Z}} \\ \mathbf{0} & \varepsilon\hat{\mathbf{Z}}^H & \lambda\mathbf{I} \end{bmatrix} \\ &\times \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_L \otimes \mathbf{J}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_L \otimes \mathbf{J}_M \end{bmatrix} \\ &= \begin{bmatrix} t-\lambda & \hat{\mathbf{y}}_{\mathbf{J}}^H & \mathbf{0} \\ \hat{\mathbf{y}}_{\mathbf{J}} & \mathbf{I} & \varepsilon\hat{\mathbf{Z}}_{\mathbf{J}} \\ \mathbf{0} & \varepsilon\hat{\mathbf{Z}}_{\mathbf{J}}^H & \lambda\mathbf{I} \end{bmatrix} \succeq 0 \end{aligned} \quad (36)$$

indicating that if $\hat{\mathbf{F}}_1$ satisfies the LMI in (33), so does $\mathbf{J}\hat{\mathbf{F}}_1\mathbf{J}$ (the subscript L is suppressed). Since the feasible set defined by a LMI is convex, given any feasible $\hat{\mathbf{F}}_1$, the convex combination $\mathbf{D}_{\hat{\mathbf{F}}_1} = (1/K)\sum_{k=1}^K \mathbf{J}_k \hat{\mathbf{F}}_1 \mathbf{J}_k$ is also inside this set, where $\mathbf{D}_{\hat{\mathbf{F}}_1}$ is a diagonal matrix such that $[\mathbf{D}_{\hat{\mathbf{F}}_1}]_{ii} = [\hat{\mathbf{F}}_1]_{ii}, \forall i$ from Lemma 1. Moreover, from the Proof of Theorem 1, we know $\psi_n(\boldsymbol{\lambda}(\mathbf{D}_{\hat{\mathbf{F}}_1}^H \mathbf{D}_{\hat{\mathbf{F}}_1})) \le \psi_n(\boldsymbol{\lambda}(\hat{\mathbf{F}}_1^H \hat{\mathbf{F}}_1))$, so $\mathbf{D}_{\hat{\mathbf{F}}_1}$ is feasible as well. Therefore, the optimal $\hat{\mathbf{F}}_1$ has a diagonal structure that can be achieved by setting $\mathbf{U}_f = \mathbf{V}_{\hat{h}}$ and $\mathbf{V}_f = \mathbf{U}_g$, leading to $\hat{\mathbf{F}}_1 = \mathbf{\Lambda}_f$. ■

Remark 4: The conditions on the power constraint and uncertainty region in Theorem 3 are satisfied in the most common situation—the sum and maximum power constraints with a spherical uncertainty region [18]–[28]. The condition $\mathbf{V}_g = \mathbf{U}_{\hat{h}}$ is satisfied by using common linear equalizers at the receiver, e.g., no equalizer $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \mathbf{I}$, the MF $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}^H$, the ZF equalizer $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}^\dagger$, the MMSE equalizer $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}^H(\sigma_n^2\mathbf{I} + \hat{\mathbf{H}}\hat{\mathbf{H}}^H)^{-1}$. Consequently, Theorem 3 indicates that the eigenmode transmission (over the nominal channel) is still optimal for the worst-case design, hence complying with the cases of perfect and statistical CSI. Interestingly, it has been found recently in [35] that maximizing the worst-case received SNR leads to the same optimal transmit directions.

VI. Optimal Power Allocation for Worst-Case Design

In this section, the optimal power allocation for the worst-case robust design will be derived. First, by using the optimal transmit directions found in the previous section, the matrix-valued precoder design can be simplified to a scalar power allocation problem as follows.

$$\begin{aligned} &\underset{\hat{\mathbf{F}}, \lambda, t}{\text{minimize}} \quad t \\ &\text{subject to} \quad \hat{\mathbf{F}} \in \mathcal{F} \\ &\begin{bmatrix} t-\lambda & \mathrm{vec}^H(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}} \hat{\mathbf{F}}_1 - \mathbf{I}) & \mathbf{0} & \mathbf{0} \\ \mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}} \hat{\mathbf{F}}_1 - \mathbf{I}) & \mathbf{I} & \varepsilon\hat{\mathbf{F}}_1^T \mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g & \varepsilon\hat{\mathbf{F}}_2^T \mathbf{\Lambda}_{t,2}^{-1/2} \otimes \mathbf{\Sigma}_g \\ \mathbf{0} & (\varepsilon\hat{\mathbf{F}}_1^T \mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g)^H & \lambda\mathbf{I} & \mathbf{0} \\ \mathbf{0} & (\varepsilon\hat{\mathbf{F}}_2^T \mathbf{\Lambda}_{t,2}^{-1/2} \otimes \mathbf{\Sigma}_g)^H & \mathbf{0} & \lambda\mathbf{I} \end{bmatrix} \succeq 0. \end{aligned} \quad (32)$$

Proposition 2: Provided the conditions in Theorem 3 are satisfied, (23) and (24) can be simplified to the following problem:

$$\begin{aligned}\underset{\mathbf{f},\mu}{\text{minimize}} \quad & \sum_{i=1}^{L} q_i(f_i,\mu) + \varepsilon^2\mu \\ \text{subject to} \quad & f_i^2 \le \mu\tau_i^2/g_{\max}^2, \quad i=1,\ldots,L \\ & \psi_n(f_1^2,\ldots,f_L^2) \le P_n, \quad \forall n \end{aligned} \tag{37}$$

where

$$q_i(f_i,\mu) = \frac{\mu\tau_i^2(\hat{w}_i f_i - 1)^2}{\mu\tau_i^2 - g_i^2 f_i^2} \tag{38}$$

is jointly convex in (f_i,μ) and defined as its limit[2] on the boundary of $\mu\tau_i^2 = g_i^2 f_i^2$ for $\mu > 0$, and each $\psi_n(\mathbf{x})$ is a Schur-convex function and componentwise nondecreasing, and $\hat{w}_i \triangleq g_i\hat{\gamma}_i$, $g_{\max} \triangleq \max_i\{g_i\}$.

Proof: See Appendix A. ■

This means that, if in addition each $\psi_n(f_1^2,\ldots,f_L^2)$ is a convex function of $\mathbf{f}$, then (37) is a convex problem and can be numerically solved very efficiently. But we will go deeper by showing that in the case of no power constraint, there exists a closed-form solution to (37); for the sum and maximum power constraints, an efficient method is available to solve (37).

A. No Power Constraint

Theorem 4: Let $P_n = \infty, \forall n$, and $g_i = g$, $i = 1,\ldots,L$, and $\{\hat{w}_i\}_{i=1}^L$ are ordered decreasingly. Then, the solution to the problem (37) is

$$f_i^\star = \max\left\{\frac{1}{\hat{w}_i}, \frac{\hat{w}_i\tau_i^2}{\hat{w}_k^2\tau_k^2}\right\} \tag{39}$$

where $1/(\hat{w}_0^2\tau_0^2) \triangleq 0$, and $k \in \{0,\ldots,L\}$ such that $\varepsilon^2 \in [\beta_{k+1},\beta_k)$ with $\beta_0 \triangleq \infty$, $\beta_{L+1} \triangleq 0$ and $\beta_j \triangleq \sum_{i=j}^L \hat{\gamma}_i^2\tau_i^2$, $j = 1,\ldots,L$. The optimum value of (37) is $L - k + g^2(\varepsilon^2 - \beta_{k+1})/(\hat{w}_k^2\tau_k^2)$.

Proof: See Appendix B. ■

Corollary 1: Provided the conditions in Theorems 3 and 4 are satisfied, the worst-case robust precoder is $\hat{\mathbf{W}}^\dagger$ if $\varepsilon^2 < \hat{\gamma}_L^2\tau_L^2$, where $\hat{\mathbf{W}} \triangleq \hat{\mathbf{G}}(\hat{\mathbf{H}})\hat{\mathbf{H}}$ is the equivalent nominal channel.

Remark 5: One situation, where the conditions in Theorems 3 and 4 are satisfied, is that there is no power constraint at the transmitter and no equalizer at the receiver. It seems that the assumption of no transmit power constraint makes Theorem 4 of little practical use. However, one will see that (39) can be used to guess a good initial point for the algorithm to solve (37) under the sum and maximum power constraints. Note that a similar solution was also obtained in [22] that considered only a special case of our framework.

[2]The limit of $q_i(f_i,\mu)$, as $(f_i,\mu) \to (x,y)$ from the interior of $\{(f_i,\mu) : f_i^2 \le \mu\tau_i^2/g_i^2\}$, is 0 when $y = g_i^2x^2/\tau_i^2$, $x = 1/\hat{w}_i$; and is ∞ when $y = g_i^2x^2/\tau_i^2$, $x \ne 1/\hat{w}_i$ and $x \ne 0$. If $\mu = 0$, then $f_i = 0, \forall i$, which cannot happen when $\varepsilon < \|\hat{\mathbf{H}}\|_F^{\mathrm{T}}$.

B. Sum and Maximum Power Constraints

With the sum and maximum power constraints, the power allocation problem is

$$\begin{aligned}\underset{\mathbf{p},\mu}{\text{minimize}} \quad & \sum_{i=1}^{L} q_i(f_i,\mu) + \mu\varepsilon^2 \\ \text{subject to} \quad & f_i^2 \le \mu\tau_i^2/g_{\max}^2, \quad f_i^2 \le P_m, \quad i=1,\ldots,L \\ & \sum_{i=1}^{L} f_i^2 \le P_s \end{aligned} \tag{40}$$

which is much more difficult and has no closed-form solution. The main difficulty in (40) lies in the coupling variable μ and the coupling constraint $\sum_{i=1}^L f_i^2 \le P_s$. That is, if we fix μ and remove the constraint $\sum_{i=1}^L f_i^2 \le P_s$, (40) will decouple into separate subproblems, each containing only one variable f_i with a straightforward solution. Upon this observation, we will propose an efficient algorithm based on decomposition methods for convex optimization [36], [37] to solve (40). Particularly, among various decomposition methods, we use the so-called primal-primal decomposition method [36], [37], which is suitable for our problem.

First, by introducing the auxiliary variables $\{p_i\}_{i=1}^L$, the problem (40) is equivalent to

$$\begin{aligned}\underset{\mathbf{f},\mu,\mathbf{p}}{\text{minimize}} \quad & \sum_{i=1}^{L} q_i(f_i,\mu) + \mu\varepsilon^2 \\ \text{subject to} \quad & f_i^2 \le \mu\tau_i^2/g_{\max}^2,\ f_i^2 \le p_i,\ i=1,\ldots,L \\ & 0 \le p_i \le P_m,\ i=1,\ldots,L \\ & \sum_{i=1}^{L} p_i \le P_s. \end{aligned} \tag{41}$$

Given $\mathbf{p}$ and μ, at the lowest level (the third level) are the decoupled subproblems, one for each i as

$$\begin{aligned}\underset{f_i}{\text{minimize}} \quad & q_i(f_i,\mu) \\ \text{subject to} \quad & f_i^2 \le \mu\tau_i^2/g_{\max}^2,\ f_i^2 \le p_i \end{aligned} \tag{42}$$

which admits a closed-form solution (see Proposition 3). Next, denote the optimum value of (42), as a function of $\mathbf{p}$ and μ, by $q_i^\star(\mathbf{p},\mu)$. For fixed $\mathbf{p}$, at the middle level (the second level) is the problem

$$\underset{\mu\ge 0}{\text{minimize}}\ h(\mathbf{p},\mu) = \sum_{i=1}^{L} q_i^\star(\mathbf{p},\mu) + \mu\varepsilon^2 \tag{43}$$

whose solution can be found through the bisection method[3] that only needs a subgradient [32] of $h(\mathbf{p},\mu)$ with respect to μ. Finally, denote the optimum value of (43), as a function of $\mathbf{p}$, by $h^\star(\mathbf{p})$. At the top level is the master problem

$$\begin{aligned}\underset{\mathbf{p}}{\text{minimize}} \quad & h^\star(\mathbf{p}) \\ \text{subject to} \quad & \sum_{i=1}^{L} p_i \le P_s, \quad 0 \le p_i \le P_m, \quad i=1,\ldots,L \end{aligned} \tag{44}$$

[3]Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a convex function with a minimizer $x^\star$ in the feasible set $\mathcal{X}$, and $s(x)$ be a subgradient of f evaluated at x. Suppose that $x^\star \in [a,b] \subseteq \mathcal{X}$ and let $c = (a+b)/2$. If $s(c) < 0$, then $x^\star \in [c,b]$, otherwise $x^\star \in [a,c]$.

whose solution can be found by the subgradient projection method. Specifically, a sequence of feasible points $\{\mathbf{p}(t)\}$, indexed by t, will be generated via

$$\mathbf{p}(t+1) = [\mathbf{p}(t) - \alpha(t)\mathbf{s}(\mathbf{p}(t))]_{\mathcal{P}} \tag{45}$$

where $\mathbf{s}(\mathbf{p}(t))$ is a subgradient of $h^\star(\mathbf{p})$ at $\mathbf{p}(t)$, $\alpha(t)$ is a positive scalar stepsize, and $[\cdot]_{\mathcal{P}}$ denotes the projection onto the a truncated simplex feasible set $\mathcal{P} \triangleq \{\mathbf{p} : 0 \leq \mathbf{p} \leq P_m\mathbf{1}, \mathbf{1}^T\mathbf{p} \leq P_s\}$. With properly chosen stepsize, for example a diminishing stepsize rule $\alpha(t) = \alpha(0)(1+d)/(t+d)$, where $\alpha(0) \in (0,1]$ is the initial stepsize and d is a fixed nonnegative number, the sequence of $\{\mathbf{p}(t)\}$ is guaranteed to converge to the optimal solution (assuming bounded subgradients). It was recently found that the projection onto a truncated simplex has a simple water-filling form [40]–[42]. In particular, if $\mathbf{p} = [\mathbf{x}]_{\mathcal{P}}$, then

$$p_i = [x_i - \eta]_0^{P_m}, \quad \forall i \tag{46}$$

where the waterlevel $\eta \geq 0$ is the minimum value such that $\sum_{i=1}^L p_i \leq P_s$.

Proposition 3: The solution to the problem (42) is given by

$$f_i^\star = \min\left\{\sqrt{p_i}, \frac{1}{\hat{w}_i}, \frac{\mu\hat{w}_i\tau_i^2}{g_i^2}, \frac{\sqrt{\mu}\tau_i}{g_{\max}}\right\} \tag{47}$$

and a subgradient of $h(\mathbf{p},\mu)$ with respect to μ is

$$s(\mu) = \sum_{i=1}^L s_i(\mu) + \varepsilon^2 \tag{48}$$

where $s_i(\mu)$ is given in (49) at the bottom of the page, and a subgradient of $h^\star(\mathbf{p})$ is $\mathbf{s}(\mathbf{p})$ given in

$$s_i(\mathbf{p}) = \begin{cases} 0, & \text{for } f_i^\star \neq \sqrt{p_i} \\ -\frac{\mu^\star\tau_i^2(1-\hat{w}_i\sqrt{p_i})(\mu^\star\hat{w}_i\tau_i^2-g_i^2\sqrt{p_i})}{\sqrt{p_i}(\mu^\star\tau_i^2-g_i^2p_i)^2}, & \text{for } f_i^\star = \sqrt{p_i} \neq \frac{\sqrt{\mu^\star}\tau_i}{g_i} \\ -\frac{\hat{w}_i}{\sqrt{p_i}(1+\hat{w}_i\sqrt{p_i})^2}, & \text{for } f_i^\star = \sqrt{p_i} = \frac{\sqrt{\mu^\star}\tau_i}{g_i}. \end{cases} \tag{50}$$

and $\mu^\star$ is the solution to (43).

Proof: See Appendix C. ■

In addition to Proposition 3, two more things are needed. One is an initial interval for the bisection method to update μ in (43). Noticing that the objective value of the minimax problem (23) is upper bounded by L, so is the objective value of (41). Since $q_i(f_i) \geq 0$, it follows that $\mu^\star\varepsilon^2 \leq L$, meaning that $\mu^\star \in [0, L/\varepsilon^2]$. The other thing is an initial point for the subgradient method to update $\mathbf{p}$ in (44). Theoretically speaking, any feasible $\mathbf{p}$ could be an initial point, but a good initial point that is close to the optimal point may remarkably accelerate the convergence. Such a good initial point can be obtained by properly scaling and bounding (39). Finally, we summarize the above in Algorithm 1.

Algorithm 1: For Solving the Problem (40)

1: set the precision ϵ; choose an initial point $\mathbf{p}(0)$; $t = 0$;
2: **repeat**
3: $a = 0$; $b = L/\varepsilon^2$; $\mu = (a+b)/2$;
4: **while** $b - a > \epsilon$ **do**
5: compute $f_i^\star$, $i = 1, \ldots, L$, from (47);
6: compute $s(\mu)$ from (48)–(49);
7: if $s(\mu) < 0$, then $a = \mu$, else $b = \mu$; $\mu = (a+b)/2$;
8: **end while**
9: compute $\mathbf{s}(\mathbf{p}(t))$ from (50);
10: $\mathbf{p}(t+1) = [\mathbf{p}(t) - \alpha(t)\mathbf{s}(\mathbf{p}(t))]_{\mathcal{P}}$ using (46);
11: $t = t + 1$;
12: **until** $|q^\star(\mathbf{p}(t)) - q^\star(\mathbf{p}(t-1))| \leq \epsilon$

VII. Least Favorable Channels

To evaluate the robustness of a precoder, one needs to know what is the least favorable or worst channel [22], [28], [43] for this precoder, which in general differs from the worst channel for another precoder. The worst channel is meaningful not only to the robust design, but also to the non-robust design that simply regards the nominal channel as the actual channel, i.e., substituting $\mathbf{G}(\mathbf{H})\mathbf{H}$ with $\hat{\mathbf{G}}(\hat{\mathbf{H}})\hat{\mathbf{H}}$ in (7). In the following, we use $\hat{\mathbf{G}}$ to represent $\hat{\mathbf{G}}(\hat{\mathbf{H}})$ for simplicity. Mathematically, assuming that the precoder $\mathbf{F}$ has been determined, the worst channel error is the solution to the following problem:

$$\underset{\|\boldsymbol{\Delta}\|_F^{\mathbf{T}}=\varepsilon}{\text{maximize}} \ \|\hat{\mathbf{G}}(\hat{\mathbf{H}} + \boldsymbol{\Delta})\mathbf{F} - \mathbf{I}\|_F^2 \tag{51}$$

where the constraint $\|\boldsymbol{\Delta}\|_F^{\mathbf{T}} \leq \varepsilon$ is replaced by $\|\boldsymbol{\Delta}\|_F^{\mathbf{T}} = \varepsilon$, since the maximum of a convex function is achieved on the boundary of the convex feasible set [44].

Theorem 5: $\boldsymbol{\Delta}$ is a solution to the problem (51) if and only if there exists μ such that

$$\mu\boldsymbol{\Delta}\mathbf{T} - \hat{\mathbf{G}}^H\hat{\mathbf{G}}\boldsymbol{\Delta}\mathbf{F}\mathbf{F}^H = \hat{\mathbf{G}}^H(\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F} - \mathbf{I})\mathbf{F}^H \tag{52}$$

$$\mu \geq \lambda_{\max}(\hat{\mathbf{G}}\hat{\mathbf{G}}^H)\lambda_{\max}(\mathbf{F}^H\mathbf{T}^{-1}\mathbf{F}) \tag{53}$$

$$\mathrm{Tr}(\boldsymbol{\Delta}\mathbf{T}\boldsymbol{\Delta}^H) = \varepsilon^2. \tag{54}$$

$$s_i(\mu) = \begin{cases} -\frac{\tau_i^2g_i^2(f_i^\star)^2(\hat{w}_if_i^\star-1)^2}{(\mu\tau_i^2-g_i^2(f_i^\star)^2)^2}, & \text{for } f_i^\star \neq \frac{\sqrt{\mu}\tau_i}{g_{\max}} \text{ and } f_i^\star \neq \frac{\sqrt{\mu}\tau_i}{g_i} \\ 0, & \text{for } f_i^\star \neq \frac{\sqrt{\mu}\tau_i}{g_{\max}} \text{ and } f_i^\star = \frac{\sqrt{\mu}\tau_i}{g_i} \\ -\frac{\tau_i^2g_{\max}(g_{\max}-\sqrt{\mu}\hat{w}_i\tau_i)(\sqrt{\mu}\hat{w}_i\tau_ig_{\max}-g_i^2)}{\mu\tau_i^2(g_{\max}^2-g_i^2)^2}, & \text{for } f_i^\star = \frac{\sqrt{\mu}\tau_i}{g_{\max}} \neq \frac{\sqrt{\mu}\tau_i}{g_i} \\ -\frac{\hat{w}_i^2\tau_i^2}{4g_{\max}^2}, & \text{for } f_i^\star = \frac{\sqrt{\mu}\tau_i}{g_{\max}} = \frac{\sqrt{\mu}\tau_i}{g_i}. \end{cases} \tag{49}$$

Proof: We will use the following lemmas.

Lemma 4 (The Sub-Region Problem [45], [46]): Consider the following quadratic minimization problem:

$$\underset{\|\mathbf{x}\|=s}{\text{minimize}}\ \mathbf{x}^H\mathbf{A}\mathbf{x} - 2\text{Re}\{\mathbf{a}^H\mathbf{x}\}$$

where s is a positive scalar, $\mathbf{a},\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{A} \in \mathbb{C}^{n\times n}$ is a Hermitian matrix. Then, $\mathbf{x}$ is a global minimizer if and only if there exists μ such that

$$(\mathbf{A}+\mu\mathbf{I})\mathbf{x} = \mathbf{a},\ \mathbf{A}+\mu\mathbf{I} \succeq 0,\ \mathbf{x}^H\mathbf{x} = s^2.$$

Lemma 5 [47, Proposition 7.1.10]: Let λ be an eigenvalue of a square matrix $\mathbf{A}$, and $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ corresponding to λ. Let τ be an eigenvalue of a square matrix $\mathbf{B}$, and $\mathbf{y}$ be an eigenvector of $\mathbf{B}$ corresponding to τ. Then, $\lambda\tau$ is an eigenvalue of $\mathbf{A}\otimes\mathbf{B}$, and $\mathbf{x}\otimes\mathbf{y}$ is an eigenvector of $\mathbf{A}\otimes\mathbf{B}$ corresponding to $\lambda\tau$.

The objective of (51), by introducing $\hat{\mathbf{\Delta}} \triangleq \mathbf{\Delta}\mathbf{T}^{1/2}$, can be expanded as

$$\begin{aligned}\|\hat{\mathbf{G}}(\hat{\mathbf{H}}+\mathbf{\Delta})\mathbf{F}-\mathbf{I}\|_F^2 &= \text{Tr}\{\hat{\mathbf{G}}^H(\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})\mathbf{F}^H\mathbf{T}^{-1/2}\hat{\mathbf{\Delta}}^H\}\\ &\quad + \text{Tr}\{\hat{\mathbf{\Delta}}\mathbf{T}^{-1/2}\mathbf{F}(\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})^H\hat{\mathbf{G}}\}\\ &\quad + \text{Tr}(\hat{\mathbf{\Delta}}\mathbf{T}^{-1/2}\mathbf{F}\mathbf{F}^H\mathbf{T}^{-1/2}+\hat{\mathbf{\Delta}}^H\hat{\mathbf{G}}^H\hat{\mathbf{G}})\\ &\quad + \|\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F}-\mathbf{I}\|_F^2 \end{aligned}\tag{55}$$

where the last term does not contain $\hat{\mathbf{\Delta}}$. Defining $\mathbf{b} \triangleq \text{vec}(\hat{\mathbf{G}}^H(\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})\mathbf{F}^H\mathbf{T}^{-1/2})$, $\mathbf{B} \triangleq \mathbf{T}^{-1/2}\mathbf{F}\mathbf{F}^H\mathbf{T}^{-1/2}$, $\mathbf{C} \triangleq \hat{\mathbf{G}}^H\hat{\mathbf{G}}$, and $\boldsymbol{\delta} \triangleq \text{vec}(\hat{\mathbf{\Delta}})$, then it is not difficult to verify that[4]

$$\begin{aligned}&\text{Tr}\{\hat{\mathbf{G}}^H(\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})\mathbf{F}^H\mathbf{T}^{-1/2}\hat{\mathbf{\Delta}}^H\}\\ &\quad + \text{Tr}\{\hat{\mathbf{\Delta}}\mathbf{T}^{-1/2}\mathbf{F}(\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})^H\hat{\mathbf{G}}\} = 2\text{Re}\{\mathbf{b}^H\boldsymbol{\delta}\} \end{aligned}\tag{56}$$

$$\text{Tr}(\hat{\mathbf{\Delta}}\mathbf{T}^{-1/2}\mathbf{F}\mathbf{F}^H\mathbf{T}^{-1/2}\hat{\mathbf{\Delta}}^H\hat{\mathbf{G}}^H\hat{\mathbf{G}}) = \boldsymbol{\delta}^H(\mathbf{B}^T\otimes\mathbf{C})\boldsymbol{\delta}. \tag{57}$$

Hence, the problem (51) amounts to

$$\underset{\|\boldsymbol{\delta}\|=\varepsilon}{\text{minimize}}\ \boldsymbol{\delta}^T(-\mathbf{B}^T\otimes\mathbf{C})\boldsymbol{\delta} - 2\text{Re}\{\mathbf{b}^H\boldsymbol{\delta}\} \tag{58}$$

whose solution, from Lemma 4, is characterized by

$$(-\mathbf{B}^T\otimes\mathbf{C}+\mu\mathbf{I})\boldsymbol{\delta} = \mathbf{b},\ -\mathbf{B}^T\otimes\mathbf{C}+\mu\mathbf{I}\succeq 0,\ \boldsymbol{\delta}^H\boldsymbol{\delta} = \varepsilon^2 \tag{59}$$

or equivalently

$$\mu\hat{\mathbf{\Delta}} - \mathbf{C}\hat{\mathbf{\Delta}}\mathbf{B} = \hat{\mathbf{G}}^H(\hat{\mathbf{G}}\hat{\mathbf{H}}\mathbf{F}-\mathbf{I})\mathbf{F}^H\mathbf{T}^{-1/2} \tag{60}$$

$$\mu \geq \lambda_{\max}(\mathbf{B}^T\otimes\mathbf{C}) \tag{61}$$

$$\text{Tr}(\hat{\mathbf{\Delta}}\hat{\mathbf{\Delta}}^H) = \varepsilon^2. \tag{62}$$

Now that $\mathbf{B}\succeq 0$ and $\mathbf{C}\succeq 0$, from Lemma 5, it follows that $\lambda_{\max}(\mathbf{B}^T\otimes\mathbf{C}) = \lambda_{\max}(\mathbf{B}^T)\lambda_{\max}(\mathbf{C}) = \lambda_{\max}(\mathbf{B})\lambda_{\max}(\mathbf{C})$. Then, (60)–(62) can be easily rewritten as (52)–(54). ■

Corollary 2: Let the conditions in Theorem 3 be satisfied. Then, the solution to the problem (51) is $\mathbf{\Delta} = \mathbf{U}_{\hat{h}}\mathbf{X}\mathbf{\Lambda}_t^{-1/2}\mathbf{V}_{\hat{h}}^H$ with $\mathbf{X} = \text{diag}\{\mathbf{Y},\mathbf{0}\}$ and $\mathbf{Y}\in\mathbb{R}^{L\times L}$, where the ($i$th, jth) element of $\mathbf{Y}\in\mathbb{R}^{L\times L}$ is denoted by $\delta_{i,j}$, if and only if there exists μ such that

$$\left(\mu - g_i^2f_i^2/\tau_i^2\right)\delta_{i,i} = (\hat{w}_if_i-1)g_if_i/\tau_i, \quad \text{for } i=j \tag{63}$$

$$(\mu - g_i^2f_j^2/\tau_j^2)\delta_{i,j} = 0, \quad \text{for } i\neq j$$

$$\mu \geq \max_i g_i^2 \max_j f_j^2/\tau_j^2 \tag{64}$$

$$\sum_{i,j=1}^{L}\delta_{i,j}^2 = \varepsilon^2 \tag{65}$$

where $\hat{w}_i \triangleq g_i\hat{\gamma}_i$. The optimum value of (51) is then given by

$$\sum_{i=1}^{L}(\delta_{i,i}g_if_i/\tau_i + \hat{w}_if_i - 1)^2 + \sum_{i\neq j}(\delta_{i,j}g_if_j/\tau_j)^2. \tag{66}$$

In general, the worst channel error that satisfies the conditions in Theorem 5 or Corollary 2 may not be unique. Here, we provide one solution satisfying the conditions (63)–(65). Define $i_{\max} \triangleq \arg\max_i g_i^2$, $j_{\max} \triangleq \arg\max_j f_j^2/\tau_j^2$ and $\omega_{\max}^2 \triangleq \max_j f_j^2/\tau_j^2$. Let $\mathbf{D}\triangleq \text{diag}\ \{(\hat{w}_if_i-1)g_if_i/\tau_i\}_{i=1}^L$, $\mathcal{I} \triangleq \{i : i=1,\ldots,L,\ i\neq i_{\max}$ if $i_{\max}=j_{\max}$ and $\mathbf{D}_{i_{\max}j_{\max}}=0\}$, and

$$\varphi(\mu) = \sum_{i\in\mathcal{I}}\frac{(\hat{w}_if_i-1)^2g_i^2f_i^2\tau_i^2}{\left(\mu\tau_i^2 - g_i^2f_i^2\right)^2}. \tag{67}$$

Then, we can choose as follows:

1) $\mathbf{D}_{i_{\max}j_{\max}} = 0$: If $\varphi(g_{\max}^2\omega_{\max}^2) \leq \varepsilon^2$, then $\mu = g_{\max}^2\omega_{\max}^2$; if $\varphi(g_{\max}^2\omega_{\max}^2) > \varepsilon^2$, then μ is the root of the equation $\varphi(\mu)=\varepsilon^2$. So

$$\delta_{i,j} = \begin{cases}\frac{(\hat{w}_if_i-1)g_if_i\tau_i}{\mu\tau_i^2-g_i^2f_i^2}, & \text{for } i=j, i\in\mathcal{I}\\ -\sqrt{\varepsilon^2-\varphi(\mu)}, & \text{for } i=i_{\max}, j=j_{\max}\\ 0, & \text{otherwise.}\end{cases} \tag{68}$$

2) $\mathbf{D}_{i_{\max}j_{\max}} \neq 0$: Then, μ should be the root of the equation $\varphi(\mu)=\varepsilon^2$. So

$$\delta_{i,j} = \begin{cases}\frac{(\hat{w}_if_i-1)g_if_i\tau_i}{\mu\tau_i^2-g_i^2f_i^2}, & \text{for } i=j, i\in\mathcal{I}\\ 0, & \text{otherwise.}\end{cases} \tag{69}$$

Note that the root of $\varphi(\mu)=\varepsilon^2$ can be found via the bisection method with an initial interval $[\mu_l,\mu_u]$ where $\mu_l = g_{\max}^2\omega_{\max}^2$ and

$$\mu_u = \frac{g_{\max}\omega_{\max}}{\varepsilon}\sqrt{\sum_{i\in\mathcal{I}}(\hat{w}_if_i-1)^2} + g_{\max}^2\omega_{\max}^2. \tag{70}$$

[4] $\text{Tr}(\mathbf{A}^T\mathbf{B}) = (\text{vec}(\mathbf{A}))^T\text{vec}(\mathbf{B})$ and $\text{Tr}(\mathbf{ABCD}) = (\text{vec}(\mathbf{D}))^T(\mathbf{A}\otimes\mathbf{C}^T)\text{vec}(\mathbf{B}^T)$.

VIII. Numerical Results

This section demonstrates the effect of the robust MMSE precoders through several numerical examples. Considering the space limitation and that the focus of this paper is on the worst-case robust design, we only show the numerical results of the worst-case design in the most common situation—the sum power constraint ($\mathcal{F} = \mathcal{F}_1$) and the spherical channel uncertainty region ($\mathbf{T} = \mathbf{I}$). To take different channels into account, the elements of the nominal channel $\hat{\mathbf{H}}$ are randomly generated according to zero-mean, unit-variance, i.i.d. Gaussian distributions. The worst-case robust precoder is compared with the non-robust precoder that regards the nominal channel $\hat{\mathbf{H}}$ as the actual channel and is given in Section III. The importance of robustness lies in guaranteeing a performance level for any channel realization in the uncertainty region, hence embodied by the worst-case behavior. Therefore, the performance is demonstrated by the average worst-case MSE and symbol error rate (SER), i.e., the MSE and SER of a (robust or non-robust) precoder in its worst channel averaged over the nominal channel $\hat{\mathbf{H}}$, where the worst channel for a given precoder can be found in Section VII. The radius of the channel uncertainty is set to be $\varepsilon^2 = s\|\hat{\mathbf{H}}\|_F^2$ with $s \in [0,1)$.

Note that, for a pre-fixed receiver, a scaling factor is required to appropriately scale the amplitude of the equalizer at different SNRs, otherwise the term $\sigma_n^2\|\hat{\mathbf{G}}(\hat{\mathbf{H}})\|_F^2$ in the MSE of (2) will not change as the transmit power varies. Therefore, instead of $\hat{\mathbf{G}}(\hat{\mathbf{H}})$, we use $\hat{\mathbf{G}}(\hat{\mathbf{H}})' = \beta\hat{\mathbf{G}}(\hat{\mathbf{H}})$ with a scaling factor $\beta \geq 0$ as in [30]. For perfect CSI, β can be jointly optimized with the precoder as

$$\underset{\mathbf{F}\in\mathcal{F}_1,\beta\geq 0}{\text{minimize}} \ \|\beta\hat{\mathbf{G}}(\hat{\mathbf{H}})\hat{\mathbf{H}}\mathbf{F}-\mathbf{I}\|_F^2 + \sigma_n^2\|\beta\hat{\mathbf{G}}(\hat{\mathbf{H}})\|_F^2 \tag{71}$$

which, by using Theorem 1, can be simplified to

$$\underset{\sum_{i=1}^L f_i^2\leq P_s,\beta\geq 0}{\text{minimize}} \ \sum_{i=1}^{L}(\beta\hat{w}_i f_i - 1)^2 + \beta^2 c \tag{72}$$

where $\hat{w}_i \triangleq g_i\hat{\gamma}_i$ and $c \triangleq \sigma_n^2\sum_{i=1}^L g_i^2 = \sigma_n^2\|\hat{\mathbf{G}}(\hat{\mathbf{H}})\|_F^2$. For fixed β, the optimal $\{f_i\}_{i=1}^L$ are given in Section III-B with w_i replaced by $\beta\hat{w}_i$, which further simplifies (72) to

$$\underset{\beta\geq 0}{\text{minimize}} \ \sum_{i=1}^{L}\frac{\eta^2}{(\beta^2 w_i^2+\eta)^2} + \beta^2 c \tag{73}$$

where $\eta \geq 0$ is the smallest value such that $\sum_{i=1}^L \beta^2 w_i^2/(\beta^2 w_i^2+\eta)^2 \leq P_s$. Although (73) is not a convex problem, the optimal β can be found through a line search. However, when it comes to the worst-case design, it is extremely difficult to jointly optimize $\mathbf{F}$ and β, so we use β that is optimal in the case of perfect CSI for the worst-case design. It should be pointed out that this scaling factor is not optimal for the worst-case design that could thus achieve a better performance.

Figs. 1 and 2 show the worst-case MSEs versus SNR for different values of s, i.e., the size of the channel uncertainty region, in the cases of no equalizer and MMSE equalizer, respectively. The numbers of transmit and receive antennas, and symbols in the transmit signal vector are set to be $N = M = L = 2$. As can be observed, the robust precoder always outperforms the non-robust precoder by providing the lowest worst-case MSE indifferent to what kind of equalizers is used. The gap between the robust and non-robust precoders increases rapidly as the channel uncertainty raises. This phenomenon can be more evidently observed from Fig. 3 that displays the worst-case MSE versus the size of the uncertainty region at $\mathrm{SNR} = 20$ dB. Compared to the robust precoder, the non-robust precoder is quite sensitive to the channel uncertainty, even a small increase of which may dramatically enlarge the worse-case MSE.

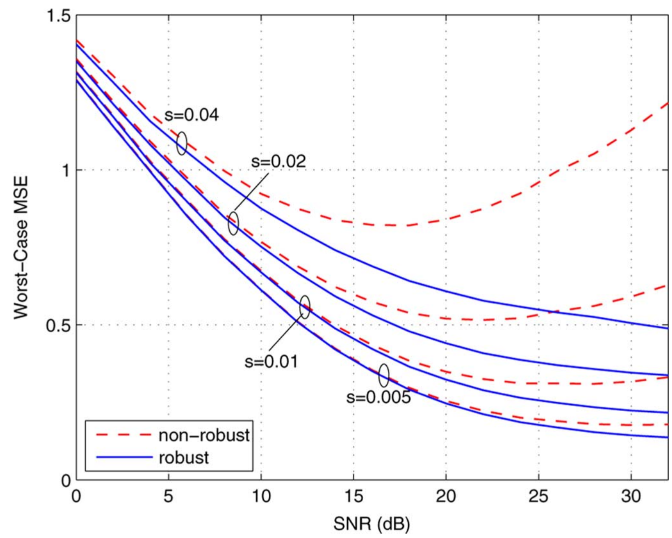

Fig. 1. Worst-case MSE versus SNR for different values of s with $N = M = L = 2$. The receiver uses no equalizer.

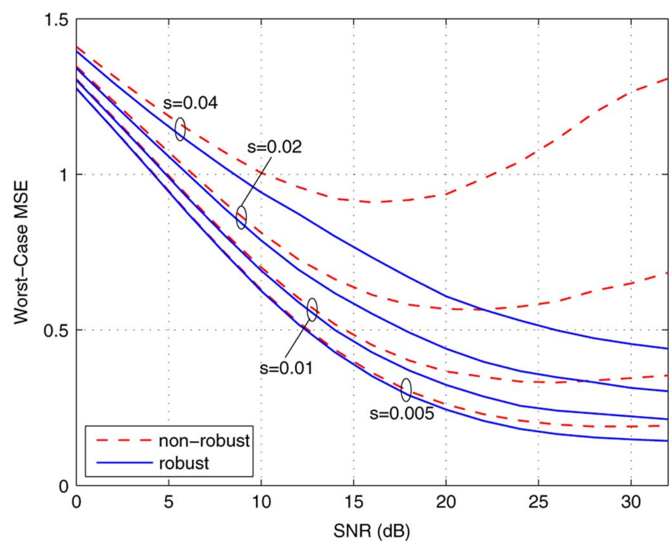

Fig. 2. Worst-case MSE versus SNR for different values of s with $N = M = L = 2$. The receiver uses the MMSE equalizer.

Another interesting phenomenon is that, among all four common equalizers, i.e., no equalizer $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \mathbf{I}$, the MF $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}^H$, the ZF equalizer $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}^\dagger$, and the MMSE equalizer $\hat{\mathbf{G}}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}^H(\sigma_n^2\mathbf{I}+\hat{\mathbf{H}}\hat{\mathbf{H}}^H)^{-1}$, the precoder with no equalizer may achieve nearly the best worst-case performance in some cases (e.g., at low or moderate SNR). Therefore, for the optimal MMSE precoder with a pre-fixed receiver using one of

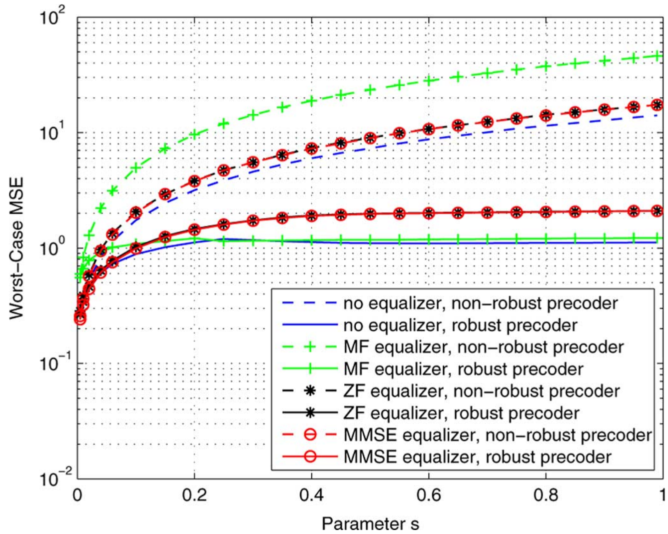

Fig. 3. Worst-case MSE versus parameter s at $\mathrm{SNR} = 20$ dB with $N = M = L = 2$.

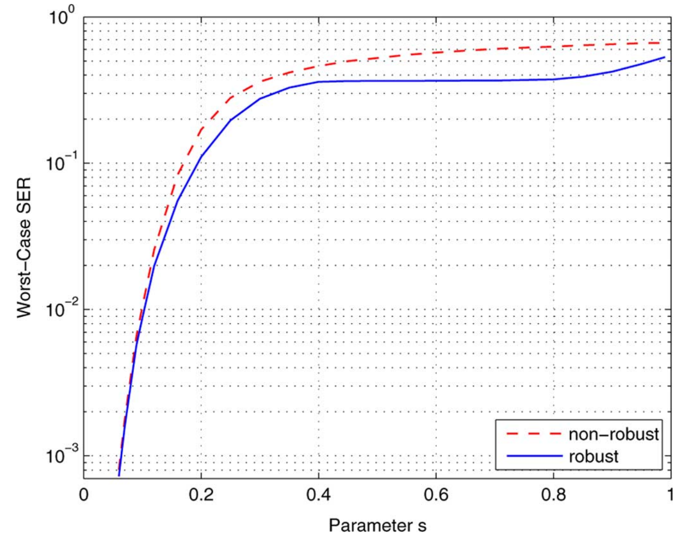

Fig. 5. Worst-case SER versus parameter s at $\mathrm{SNR} = 20$ dB with $N = 8$ and $M = L = 2$.

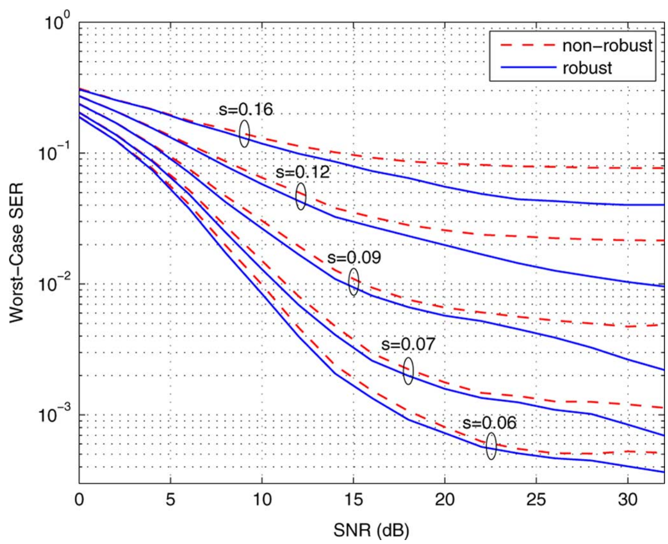

Fig. 4. Worst-case SER versus SNR for different values of s with $N = 8$ and $M = L = 2$.

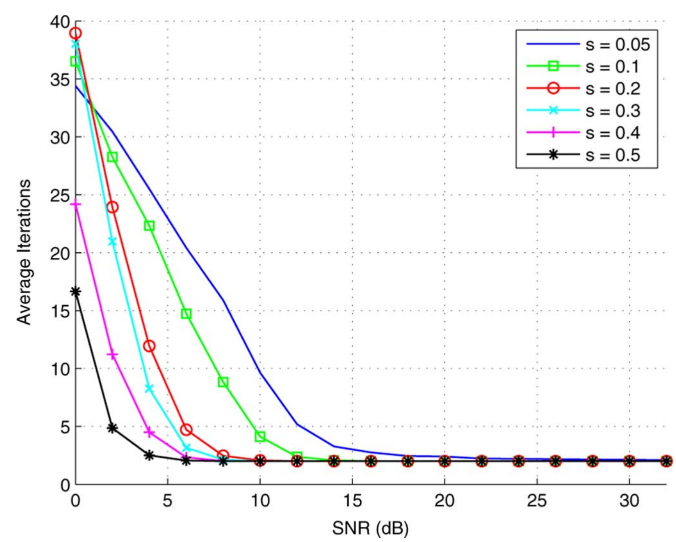

Fig. 6. Average iteration of Algorithm 1 versus SNR for different values of s with $\epsilon = 10^{-5}$ and $N = M = L = 4$.

these equalizers, a good choice that balances performance and complexity could be using no equalizer at the receiver. Due to the space limitation, we cannot show simulation results of all four common equalizers, so in the following we focus on the case of no equalizer.

For no equalizer and under the setting of $N = 8, M = L = 2$ and the QPSK modulation, Fig. 4 depicts the worst-case SER versus SNR for different values of s, while Fig. 5 shows the relation between the worst-case SER and the size of the uncertainty region at $\mathrm{SNR} = 20$ dB. Although the robust precoder is based on the MMSE criterion, its superiority to the non-robust precoder still holds in terms of the worst-case SER, especially for large channel uncertainty.

Finally, the convergence property of Algorithm 1, which iteratively solves the power allocation problem (40), is investigated. In Fig. 6, the average number of iterations to reach a precision $\epsilon = 10^{-5}$ is shown at different SNRs for different sizes of the uncertainty region, where we set $N = M = L = 4$ and there is no equalizer. As a sub-gradient method, Algorithm 1 converges quite fast with the properly chosen stepsize, especially at high SNRs. This is mainly because we can use (39), i.e., the closed-form solution to (40) when there is no power constraint, to guess a good initial point.

IX. Conclusion

We have designed the robust MMSE precoders with a pre-fixed receiver for different types of CSI in MIMO channels. All formulated problems are or can be equivalently transformed into convex problems having globally optimal solutions that can be efficiently found. As the most important result, we have proved that, under some mild conditions, the optimal transmit directions of the precoder are the right singular vectors of the channel, the channel mean, and the nominal channel for perfect CSI, statistical CSI, and the worst-case design, respectively. Consequently, the matrix-valued problems can be simplified to scalar power allocation problems without losing any optimality. Then, we have provided the closed-form solutions to the power allocation problems for perfect and statistical CSI, and proposed an efficient algorithm to solve the power allocation problem for the worst-case design. Finally, the worst channels for both robust and non-robust MMSE precoders have been derived.

Appendix A
Proof of Proposition 2

The proof starts form the equivalent formulation (33). Using $\mathbf{U}_t = \mathbf{V}_{\hat{h}}$, $\mathbf{V}_g = \mathbf{U}_{\hat{h}}$, $\mathbf{U}_f = \mathbf{V}_{\hat{h}}$ and $\mathbf{V}_f = \mathbf{U}_g$, the LMI in (33) reduces to (74) at the bottom of the page, which, via proper row and column permutations, can be reformulated into $\text{diag}\{\mathbf{A}, \mathbf{B}, \lambda\mathbf{I}\} \succeq 0$, where $\lambda\mathbf{I}$ accounts for the zero part in $\mathbf{\Sigma}_g$, and

$$\mathbf{A} = \begin{bmatrix} t-\lambda & \mathbf{0} & \mathbf{c}^T \\ \mathbf{0} & \lambda\mathbf{I} & \mathbf{D} \\ \mathbf{c} & \mathbf{D} & \mathbf{I} \end{bmatrix},$$
$$\mathbf{B} = \text{diag}\left\{ \begin{bmatrix} 1 & \varepsilon g_i f_j/\tau_j \\ \varepsilon g_i f_j/\tau_j & \lambda \end{bmatrix} \right\}_{i\neq j},$$

with $\mathbf{c} = [g_1\hat{\gamma}_1 f_1 - 1, \ldots, g_L\hat{\gamma}_L f_L - 1]^T$, $\mathbf{D} = \text{diag}\{\varepsilon g_1 f_1/\tau_1, \ldots, \varepsilon g_L f_L/\tau_L\}$. Therefore, the constraint (74) is equivalent to $\mathbf{A} \succeq 0$ and $\mathbf{B} \succeq 0$. While $\mathbf{B} \succeq 0$ amounts to

$$\begin{bmatrix} 1 & \varepsilon g_i f_j/\tau_j \\ \varepsilon g_i f_j/\tau_j & \lambda \end{bmatrix} \succeq 0,\ i \neq j \tag{75}$$

which is equivalent to $\lambda \geq \varepsilon^2 g_i^2 f_j^2/\tau_j^2$ for $i \neq j$, $\mathbf{A} \succeq 0$, from Lemma 2, amounts to

$$\begin{bmatrix} t-\lambda & \mathbf{0} \\ \mathbf{0} & \lambda\mathbf{I} \end{bmatrix} - \begin{bmatrix} \mathbf{c}^T \\ \mathbf{D} \end{bmatrix} [\mathbf{c} \quad \mathbf{D}] = \begin{bmatrix} t-\lambda-\mathbf{c}^T\mathbf{c} & -\mathbf{c}^T\mathbf{D} \\ -\mathbf{D}\mathbf{c} & \lambda\mathbf{I}-\mathbf{D}^2 \end{bmatrix} \succeq 0 \tag{76}$$

implying $\lambda\mathbf{I} \succeq \mathbf{D}^2$ or equally $\lambda \geq \varepsilon^2 g_i^2 f_i^2/\tau_i^2, \forall i$. Therefore, (33) can be simplified to the following problem:

$$\begin{aligned} \underset{\mathbf{f},\lambda,t}{\text{minimize}} \quad & t \\ \text{subject to} \quad & \begin{bmatrix} t-\lambda-\mathbf{c}^T\mathbf{c} & -\mathbf{c}^T\mathbf{D} \\ -\mathbf{D}\mathbf{c} & \lambda\mathbf{I}-\mathbf{D}^2 \end{bmatrix} \succeq 0 \\ & \lambda \geq \varepsilon^2 g_i^2 f_j^2/\tau_j^2,\ i,j = 1,\ldots,L \\ & \psi_n(f_1^2,\ldots,f_L^2) \leq P_n,\ \forall n. \end{aligned} \tag{77}$$

Assuming for the moment that $\lambda > \varepsilon^2 g_i^2 f_i^2/\tau_i^2, \forall i$, then $\lambda\mathbf{I}-\mathbf{D}^2$ is invertible. Using Lemma 2 again, (76) is equivalent to

$$t-\lambda-\mathbf{c}^T\mathbf{c}-\mathbf{c}^T\mathbf{D}(\lambda\mathbf{I}-\mathbf{D}^2)^{-1}\mathbf{D}\mathbf{c} = t-\lambda-\varphi_i(f_i,\lambda) \geq 0, \tag{78}$$

where

$$\varphi_i(f_i,\lambda) = \frac{\lambda(g_i\hat{\gamma}_i f_i - 1)^2}{\lambda - \varepsilon^2 g_i^2 f_i^2/\tau_i^2}. \tag{79}$$

Hence, the problem (77) is equivalent to

$$\begin{aligned} \underset{\mathbf{f},\lambda}{\text{minimize}} \quad & \sum_{i=1}^{L} \varphi_i(f_i,\lambda) + \lambda \\ \text{subject to} \quad & \lambda > \varepsilon^2 g_i^2 f_i^2/\tau_i^2,\ i = 1,\ldots,L \\ & \lambda \geq \varepsilon^2 g_i^2 f_j^2/\tau_j^2,\ i \neq j \\ & \psi_n(f_1^2,\ldots,f_L^2) \leq P_n,\ \forall n. \end{aligned} \tag{80}$$

Now, assume without loss of generality (w.l.o.g.) that $\lambda = \varepsilon^2 g_m^2 f_m^2/\tau_m^2$ for some m. Then, by choosing the 1 st and $(m+1)$th rows and columns of (76), we have (81) at the bottom of the page, or equivalently $(g_m\hat{\gamma}_m f_m - 1)^2 \varepsilon^2 g_m^2 f_m^2/\tau_m^2 \leq 0$, which implies that either $f_m = 1/(g_m\hat{\gamma}_m)$ or $f_m = 0$. If $f_m = 0$, then $\lambda = 0$ and $f_i = 0, \forall i$. As mentioned at the beginning of Section V, $f_i = 0, \forall i$, can happen only when $\varepsilon \geq ||\hat{\mathbf{H}}||_F^{\mathbf{T}}$. With the assumption that $\varepsilon < ||\hat{\mathbf{H}}||_F^{\mathbf{T}}$, $\lambda = \varepsilon^2 g_m^2 f_m^2/\tau_m^2$ implies only $f_m = 1/(g_m\hat{\gamma}_m)$. Then, one can remove the $(m+1)$th row and column of (76), leading to

$$\begin{bmatrix} t-\lambda-(\mathbf{c})_{-m}^T(\mathbf{c})_{-m} & -(\mathbf{D}\mathbf{c})_{-m}^T \\ -(\mathbf{D}\mathbf{c})_{-m} & (\lambda\mathbf{I}-\mathbf{D}^2)_{-m} \end{bmatrix} \succeq 0 \tag{82}$$

where $(\mathbf{x})_{-m}$ is the result of deleting the mth element of a vector $\mathbf{x}$, and $(\mathbf{X})_{-m}$ is the result of deleting the mth row and column of a square matrix $\mathbf{X}$. Hence, in the case of $\lambda = \varepsilon^2 g_m^2 f_m^2/\tau_m^2$, following (78), (33) is equivalent to

$$\begin{aligned} \underset{\mathbf{f},\lambda}{\text{minimize}} \quad & \sum_{i\neq m} \varphi_i(f_i,\lambda) + \lambda \\ \text{subject to} \quad & \lambda \geq \varepsilon^2 g_i^2 f_j^2/\tau_j^2,\ i,j = 1,\ldots,L \\ & \psi_n(f_1^2,\ldots,f_L^2) \leq P_n,\ \forall n. \end{aligned} \tag{83}$$

Note that, as $(f_m,\lambda) \to (x,y)$ from the interior of $\{(f_m,\lambda) : \lambda \geq \varepsilon^2 g_m^2 f_m^2/\tau_m^2\}$, the limit of $\varphi_m(f_m,\lambda)$ is

$$\begin{bmatrix} t-\lambda & \text{vec}^T(\mathbf{\Lambda}_g\mathbf{\Lambda}_{\hat{h}}\mathbf{\Lambda}_f - \mathbf{I}) & \mathbf{0} \\ \text{vec}(\mathbf{\Lambda}_g\mathbf{\Lambda}_{\hat{h}}\mathbf{\Lambda}_f - \mathbf{I}) & \mathbf{I} & \varepsilon\mathbf{\Lambda}_f^T\mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g \\ \mathbf{0} & (\varepsilon\mathbf{\Lambda}_f^T\mathbf{\Lambda}_{t,1}^{-1/2} \otimes \mathbf{\Sigma}_g)^T & \lambda\mathbf{I} \end{bmatrix} \succeq 0 \tag{74}$$

$$\begin{bmatrix} t-\lambda-\mathbf{c}^T\mathbf{c} & -(g_m\hat{\gamma}_m f_m - 1)\varepsilon g_m f_m/\tau_m \\ -(g_m\hat{\gamma}_m f_m - 1)\varepsilon g_m f_m/\tau_m & 0 \end{bmatrix} \succeq 0 \tag{81}$$

0 if $x = 1/(g_m\hat{\gamma}_m)$ and $y = \varepsilon^2 g_m^2 x^2/\tau_m^2$; and is ∞ if $x \neq 1/(g_m\hat{\gamma}_m)$ and $y = \varepsilon^2 g_m^2 x^2/\tau_m^2$. Therefore, by defining each $\varphi_i(f_i, \lambda)$ on the boundary of $\lambda = \varepsilon^2 g_i^2 f_i^2/\tau_i^2$ as its limit, we can extend (80) to

$$\begin{aligned} \underset{\mathbf{f},\lambda}{\text{minimize}} \quad & \sum_{i=1}^{L} \varphi_i(f_i, \lambda) + \lambda \\ \text{subject to} \quad & \lambda \geq \varepsilon^2 g_i^2 f_j^2/\tau_j^2,\ i,j = 1,\ldots,L \\ & \psi_n(f_1^2,\ldots,f_L^2) \leq P_n,\ \forall n \end{aligned} \tag{84}$$

so that (77) is equivalent to (84) in both cases that $\lambda > \varepsilon^2 g_i^2 f_i^2/\tau_i^2, \forall i$, and $\lambda = \varepsilon^2 g_m^2 f_m^2/\tau_m^2$ for some m. Letting $\mu \triangleq \lambda/\varepsilon^2$, (84) can be easily rewritten as (37).

Finally, we prove the convexity of $q_i(f_i,\mu)$ in (38) by considering the following function:

$$\phi(\delta_i, f_i, \mu) = ((\hat{\gamma}_i + \delta_i)g_i f_i - 1)^2 - \mu\tau_i^2\delta_i^2 \tag{85}$$

which is convex in (f_i,μ) for fixed δ_i. Since $\partial^2\phi(\delta_i, f_i, \mu)/\partial\delta_i^2 = 2(g_i^2 f_i^2 - \mu\tau_i^2)$, $\phi(\delta_i, f_i, \mu)$ is concave in δ_i for fixed (f_i,μ) under the condition $\mu \geq g_i^2 f_i^2/\tau_i^2$. The maximizer of $\phi(\delta_i, f_i, \mu)$, for fixed (f_i,μ), is given by

$$\delta_i^\star = \frac{(g_i\hat{\gamma}_i f_i - 1)g_i f_i}{\mu\tau_i^2 - g_i^2 f_i^2} \tag{86}$$

leading to the maximum value

$$\phi(\delta_i^\star, f_i, \mu) = \frac{\mu\tau_i^2(g_i\hat{\gamma}_i f_i - 1)^2}{\mu\tau_i^2 - g_i^2 f_i^2} = q_i(f_i, \mu). \tag{87}$$

Since maximization preserves convexity [32], $q_i(f_i,\mu)$ is convex in (f_i,μ).

Appendix B
Proof of Theorem 4

When there is no power constraint, by introducing $x_i \triangleq f_i g/\tau_i$, the problem (37) becomes

$$\begin{aligned} \underset{\mathbf{x},\mu}{\text{minimize}} \quad & \sum_{i=1}^{L} \frac{\mu(\hat{\gamma}_i\tau_i x_i - 1)^2}{\mu - x_i^2} + \mu\varepsilon^2 \\ \text{subject to} \quad & x_i^2 \leq \mu,\ i = 1,\ldots,L. \end{aligned} \tag{88}$$

To solve it, we first fix μ so that (88) decouples into separate subproblems, each for a variable x_i, $i = 1,\ldots,L$. Then, it is easy to find the optimal x_i as

$$x_i^\star = \begin{cases} \frac{1}{\hat{\gamma}_i\tau_i}, & \text{if } \frac{1}{\hat{\gamma}_i^2\tau_i^2} \leq \mu \\ \mu\hat{\gamma}_i\tau_i, & \text{if } \frac{1}{\hat{\gamma}_i^2\tau_i^2} > \mu. \end{cases} \tag{89}$$

Defining $\alpha_0 \triangleq 0$, $\alpha_{L+1} \triangleq \infty$ and $\alpha_j \triangleq (1)/(\hat{\gamma}_j^2\tau_j^2)$, $j = 1,\ldots,L$, then the region $[0,\infty)$ can be divided into $L+1$ consecutive intervals $[\alpha_j, \alpha_{j+1})$, $j = 0,\ldots,L$. So (89) can be rewritten as

$$x_i^\star = \begin{cases} \frac{1}{\hat{\gamma}_i\tau_i}, & \text{for } i \leq m \\ \mu\hat{\gamma}_i\tau_i, & \text{for } i > m \end{cases} \tag{90}$$

where $m \in \{0,\ldots,L\}$ such that $\mu \in [\alpha_m, \alpha_{m+1})$.

Now we search the optimal μ by substituting (90) back into (88), and the resulting objective function for a specific m is

$$\begin{aligned} h_m(\mu) &= \sum_{i=m+1}^{L} \left(1 - \mu\hat{\gamma}_i^2\tau_i^2\right) + \mu\varepsilon^2 \\ &= L - m + \left(\varepsilon^2 - \sum_{i=m+1}^{L} \hat{\gamma}_i^2\tau_i^2\right)\mu \end{aligned} \tag{91}$$

which is linear in μ. Consequently, the objective is a piecewise linear function as

$$h(\mu) = \begin{cases} h_m(\mu), & \text{if } \mu \in [\alpha_m, \alpha_{m+1}),\ m = 0,\ldots,L \\ 0, & \text{otherwise} \end{cases} \tag{92}$$

and the corresponding problem becomes

$$\underset{\mu \geq 0}{\text{minimize}}\ h(\mu). \tag{93}$$

Defining $\beta_0 \triangleq \infty$, $\beta_{L+1} \triangleq 0$ and $\beta_j \triangleq \sum_{i=j}^{L} \hat{\gamma}_i^2\tau_i^2$, $j = 1,\ldots,L$, then the region $[0,\infty)$ can be divided into $L+1$ consecutive intervals $[\beta_{j+1}, \beta_j)$, $j = 0,\ldots,L$. Assuming that $\varepsilon^2 \in [\beta_{k+1}, \beta_k)$ for some k, we will show that the optimal μ must be within $[\alpha_k, \alpha_{k+1})$ and further is equal to α_k. For $m < k$, the slope of $h_m(\mu)$ is negative due to $\varepsilon^2 < \beta_{m+1}$, so the minimum of $h_m(\mu)$ is achieved at the limit $\mu \to \alpha_{m+1}$, meaning that the optimal μ is not in $[\alpha_m, \alpha_{m+1})$. For $m \geq k$, the slope of $h_m(\mu)$ is nonnegative, so the minimum of $h_m(\mu)$ is achieved at $\mu = \alpha_m$. However, from

$$\begin{aligned} h_m(\alpha_m) &\\ = L - m + \alpha_m \left(\varepsilon^2 - \beta_{m+1}\right) & \end{aligned} \tag{94}$$

$$\begin{aligned} h_{m+1}(\alpha_{m+1}) &\\ = L - m - 1 + \alpha_{m+1}(\varepsilon^2 - \beta_{m+2}) &\\ = L - m - 1 + \alpha_{m+1}(\varepsilon^2 - \beta_{m+1} + \tau_{m+1}^2\hat{\gamma}_{m+1}^2) &\\ = L - m + \alpha_{m+1}(\varepsilon^2 - \beta_{m+1}) & \end{aligned} \tag{95}$$

we have $h_m(\alpha_m) \leq h_{m+1}(\alpha_{m+1})$, indicating that the optimal μ can only lie in $[\alpha_k, \alpha_{k+1})$ and $\mu^\star = \alpha_k$. The optimum value is given by

$$h(\mu^\star) = h_k(\alpha_k) = L - k + \alpha_k(\varepsilon^2 - \beta_{k+1}). \tag{96}$$

Substituting $\mu^\star = \alpha_k$ back into (90), we have

$$x_i^\star = \begin{cases} \frac{1}{\hat{\gamma}_i\tau_i}, & \text{for } i \leq k \\ \frac{\hat{\gamma}_i\tau_i}{\hat{\gamma}_k^2\tau_k^2}, & \text{for } i > k \end{cases} = \max\left\{\frac{1}{\hat{\gamma}_i\tau_i}, \frac{\hat{\gamma}_i\tau_i}{\hat{\gamma}_k^2\tau_k^2}\right\} \tag{97}$$

so that the optimal f_i can be obtained as $f_i^\star = x_i^\star\tau_i/g$.

Appendix C
Proof of Proposition 3

Lemma 6 [37]: Consider the following function defined as the optimal value of a minimization problem:

$$f^\star(x) \triangleq \inf_{y: f_i(x,y) \leq 0} f_0(x,y)$$

where $f_0 : \mathbb{R}\times\mathbb{R}\mapsto\mathbb{R}$ is (strictly) convex, $f_i : \mathbb{R}\times\mathbb{R}\mapsto\mathbb{R}$ are convex, the infermum is achieved by $y^\star(x)$, and the strong duality holds for any given x. Let $\mathcal{X} = \{x|f^\star(x) > -\infty\}$ be the domain of $f^\star$. Then, $f^\star(x)$ is (strictly) convex, provided that $f^\star(x) < \infty$ for some $x\in\mathcal{X}$, and a subgradient is given by

$$s(x) = s_{0,x}(x, y^\star(x)) + \mathbf{s}_x(x, y^\star(x))^T\boldsymbol{\lambda}^\star(x),$$

where $s_{0,x}(x, y^\star(x))$ is the subgradient of $f_0(x, y)$ with respect to x, $\mathbf{s}_x(x, y^\star(x))$ is a vector with ith element being a subgradient of $f_i(x, y)$ with respect to x, and $\boldsymbol{\lambda}^\star(x)$ is a vector of the optimal Lagrange multipliers associated with the constraints $f_i(x, y)\le 0, \forall i$.

For fixed $\mathbf{f}$ and μ, it is not difficult to verify that (47) is the solution to (42). Particularly, when $f_i^\star = (\sqrt{\mu}\tau_i)/(g_i)\ (=\sqrt{p_i})$, it must be $f_i^\star = (\sqrt{\mu}\tau_i)/(g_i) = (1)/(\hat{w}_i) = (\mu\hat{w}_i\tau_i^2)/(g_i^2)$ $(=\sqrt{p_i})$. In this case, $q_i(f_i,\mu)$ reduces to

$$q_i(f_i,\mu) = \frac{1-\hat{w}_if_i}{1+\hat{w}_if_i}. \tag{98}$$

Hence, we have

$$\frac{\partial q_i(f_i,\mu)}{\partial f_i} = \begin{cases} \frac{2\mu\tau_i^2(\hat{w}_if_i-1)\left(\mu\hat{w}_i\tau_i^2-g_i^2f_i\right)}{(\mu\tau_i^2-g_i^2f_i^2)^2}, & \text{for } f_i^\star\ne\frac{\sqrt{\mu}\tau_i}{g_i}\\ \frac{-2\hat{w}_i}{(1+\hat{w}_if_i)^2}, & \text{for } f_i^\star=\frac{\sqrt{\mu}\tau_i}{g_i}. \end{cases} \tag{99}$$

After $\mathbf{f}^\star$ is obtained, to find a subgradient of $q_i^\star(\mathbf{p},\mu)$ with respect to μ, according to Lemma 6, one requires the optimal Lagrange multiplier $\alpha_i^\star$ associated with the constraint $f_i^2\le\mu\tau_i^2/g_{\max}^2$. Let $\beta_i^\star$ be the optimal Lagrange multiplier associated with the constraint $f_i^2\le p_i$. Then, $f_i^\star$, $\alpha_i^\star$ and $\beta_i^\star$ satisfy the following KKT conditions [32]:

$$\alpha_{i,j}^\star\ge 0,\ \alpha_i^\star((f_i^\star)^2-\mu\tau_i^2/g_{\max}^2)=0 \tag{100}$$
$$\beta_i^\star\ge 0,\ \beta_i^\star((f_i^\star)^2-p_i)=0 \tag{101}$$
$$\partial q_i(f_i,\mu)/\partial f_i|_{f_i=f_i^\star}+2(\alpha_i^\star+\beta_i^\star)f_i^\star=0. \tag{102}$$

When $f_i^\star\ne(\sqrt{\mu}\tau_i)/(g_{\max})$, it is easily seen from (100) that $\alpha_i^\star=0$. When $f_i^\star=(\sqrt{\mu}\tau_i)/g_{\max}$, we need to discuss two cases.

1) $f_i^\star=\frac{\sqrt{\mu}\tau_i}{g_{\max}}\ne\frac{\sqrt{\mu}\tau_i}{g_i}$: If $\frac{\sqrt{\mu}\tau_i}{g_{\max}}\ne\sqrt{p_i}$, then from (101) $\beta_i^\star=0$, and from (102) and (99)

$$\alpha_i^\star=\frac{g_{\max}^3(g_{\max}-\sqrt{\mu}\hat{w}_i\tau_i)(\sqrt{\mu}\hat{w}_i\tau_ig_{\max}-g_i^2)}{\mu\tau_i^2\left(g_{\max}^2-g_i^2\right)^2}. \tag{103}$$

If $(\sqrt{\mu}\tau_i)/(g_{\max})=\sqrt{p_i}$, it follows from (100)–(102) that

$$\alpha_i^\star+\beta_i^\star=\frac{g_{\max}^3(g_{\max}-\sqrt{\mu}\hat{w}_i\tau_i)(\sqrt{\mu}\hat{w}_i\tau_ig_{\max}-g_i^2)}{\mu\tau_i^2\left(g_{\max}^2-g_i^2\right)^2} \tag{104}$$

and $\alpha_i^\star$, $\beta_i^\star\ge 0$, so we can still choose $\beta_i^\star=0$ and $\alpha_i^\star$ as (103).

2) $f_i^\star=(\sqrt{\mu}\tau_i)/(g_{\max})=(\sqrt{\mu}\tau_i)/(g_i)$: If $f_i^\star=(\sqrt{\mu}\tau_i)/(g_i)=(1)/(\hat{w}_i)\ne\sqrt{p_i}$, then from (101) $\beta_i^\star=0$, and from (102) and (99) $\alpha_i^\star=(\hat{w}_i^2)/(4)$. If $f_i^\star=(\sqrt{\mu}\tau_i)/(g_i)=(1)/(\hat{w}_i)=\sqrt{p_i}$, it follows from (100)–(102) that

$$\alpha_i^\star+\beta_i^\star=\frac{\hat{w}_i^2}{4}, \tag{105}$$

and $\alpha_i^\star$, $\beta_i^\star\ge 0$, so we can choose $\beta_i^\star=0$ and $\alpha_i^\star=((\hat{w}_i^2)/4)$.

Consequently, we have $\alpha_i^\star$ in (106) at the bottom of the page, which, from Lemma 6 , leads to a subgradient of $q_i^\star(\mathbf{p},\mu)$ with respect to μ as

$$s_i(\mu)=\frac{\partial q_i(f_i^\star,\mu)}{\partial\mu}-\frac{\tau_i^2\alpha_i^\star}{g_{\max}^2}=\begin{cases}\frac{\partial q_i(f_i^\star)}{\partial\mu}, & \text{for } f_i^\star\ne\frac{\sqrt{\mu}\tau_i}{g_{\max}}\\ -\frac{\tau_i^2\alpha_i^\star}{g_{\max}^2}, & \text{for } f_i^\star=\frac{\sqrt{\mu}\tau_i}{g_{\max}}.\end{cases} \tag{107}$$

where from (98)

$$\frac{\partial q_i(f_i^\star,\mu)}{\partial\mu}=\begin{cases}-\frac{\tau_i^2g_i^2(f_i^\star)^2(\hat{w}_if_i^\star-1)^2}{(\mu\tau_i^2-g_i^2(f_i^\star)^2)^2}, & \text{for } f_i^\star\ne\frac{\sqrt{\mu}\tau_i}{g_i}\\ 0, & \text{for } f_i^\star=\frac{\sqrt{\mu}\tau_i}{g_i}.\end{cases} \tag{108}$$

Substituting (106) and (108) into (107) leads to (49).

Given $\mathbf{f}^\star$ and $\mu^\star$, to find a subgradient of $h^\star(\mathbf{p})$, one requires the optimal Lagrange multiplier $\beta_i^\star$ associated with the constraint $f_i^2\le p_i$. The KKT conditions, satisfied by $\mu^\star$, $\alpha_i^\star$ and $\beta_i^\star$, are still (100)–(102) except μ is replaced by $\mu^\star$. Through the similar analysis, we can obtain $\beta_i^\star$ as

$$\beta_i^\star=\begin{cases}0, & \text{for } f_i^\star\ne\sqrt{p_i}\\ \frac{\mu^\star\tau_i^2(1-\hat{w}_i\sqrt{p_i})(\mu^\star\hat{w}_i\tau_i^2-g_i^2\sqrt{p_i})}{\sqrt{p_i}(\mu^\star\tau_i^2-g_i^2p_i)^2}, & \text{for } f_i^\star=\sqrt{p_i}\ne\frac{\sqrt{\mu^\star}\tau_i}{g_i}\\ \frac{\hat{w}_i}{\sqrt{p_i}(1+\hat{w}_i\sqrt{p_i})^2}, & \text{for } f_i^\star=\sqrt{p_i}=\frac{\sqrt{\mu^\star}\tau_i}{g_i}.\end{cases} \tag{109}$$

and a subgradient of $h^\star(\mathbf{p})$ is just $-\beta_i^\star$. The proof of Proposition 3 is completed.

$$\alpha_i^\star=\begin{cases}0, & \text{for } f_i^\star\ne\frac{\sqrt{\mu}\tau_i}{g_{\max}}\\ \frac{g_{\max}^3(g_{\max}-\sqrt{\mu}\hat{w}_i\tau_i)(\sqrt{\mu}\hat{w}_i\tau_ig_{\max}-g_i^2)}{\mu\tau_i^2(g_{\max}^2-g_i^2)^2}, & \text{for } f_i^\star=\frac{\sqrt{\mu}\tau_i}{g_{\max}}\ne\frac{\sqrt{\mu}\tau_i}{g_i}\\ \frac{\hat{w}_i^2}{4}, & \text{for } f_i^\star=\frac{\sqrt{\mu}\tau_i}{g_{\max}}=\frac{\sqrt{\mu}\tau_i}{g_i}.\end{cases} \tag{106}$$

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999, (See also a previous version of the paper in AT&T Bell Labs Internal Tech. Memo, June 1995).

[2] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[3] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Redundant filter-bank precoders and equalizers Part 1: Unification and optimal design," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 1988–2006, Jul. 1999.

[4] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[5] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[6] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 4–5, pp. 331–551, 2006.

[7] M. Payaró and D. P. Palomar, "On optimal precoding in linear vector Gaussian channels with arbitrary input distribution," in *Proc. IEEE International Symposium on Information Theory (ISIT'09)*, Seoul, Korea, Jun. 2009.

[8] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[9] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.

[10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[11] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[12] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[13] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 662–671, Mar. 2006.

[14] E. A. Jorswieck, A. Sezgin, H. Boche, and E. Costa, "Multiuser MIMO MAC with statistical CSI and MMSE receiver: Feedback strategies and transmitter optimization," in *Proc. International Wireless Communications and Mobile Computing Conference (IWCMC)*, Vancouver, Canada, Jul. 2006.

[15] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Processing*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[16] A. Soysal and S. Ulukus, *Joint Channel Estimation and Resource Allocation for MIMO Systems—Part I: Single-User Analysis*, vol. 9, no. 2, pp. 624–631, Feb. 2010.

[17] M. Vu and A. Paulraj, "Optimal linear precoders for MIMO wireless correlated channels with nonzero mean in space-time coded systems," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2318–2332, Jun. 2006.

[18] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[19] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[20] R. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[21] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Processing*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[22] Y. Guo and B. C. Levy, "Worst-case MSE precoder design for imperfectly known MIMO communications channels," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2918–2930, Aug. 2005.

[23] A. Pascual-Iserte, D. P. Palomar, A. I. Pérez-Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with partial channel state information based on convex optimization," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 346–360, Jan. 2006.

[24] A. Abdel-Samad, T. N. Davidson, and A. B. Gershman, "Robust transmit eigen beamforming based on imperfect channel state information," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1596–1609, May 2006.

[25] M. Payaró, A. Pascual-Iserte, and M. A. Lagunas, "Robust power allocation designs for multiuser and multiantenna downlink communication systems through convex optimization," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1390–1401, Sep. 2007.

[26] M. B. Shenouda and T. N. Davidson, "On the design of linear transceivers for multiuser systems with channel uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 1015–1024, Aug. 2008.

[27] N. Vucic and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Processing*, vol. 57, no. 2, pp. 714–725, Feb. 2009.

[28] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Processing*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.

[29] B. R. Vojcic and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1346–1355, Oct. 1998.

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

[31] A. Wiesel, Y. C. Eldar, and S. S. (Shitz), "Linear precoding via conic programming for fixed MIMO receivers," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[33] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994, Studies in Applied Mathematics 13.

[34] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.

[35] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Processing*, vol. 57, no. 8, pp. 3086–3100, Aug. 2009.

[36] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[37] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[38] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.

[39] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.

[40] D. P. Palomar, "Convex primal decomposition for multicarrier linear MIMO transceivers," *IEEE Trans. Signal Processing*, vol. 53, no. 12, pp. 4661–4674, Dec. 2005.

[41] D. P. Palomar, M. Bengtsson, and B. Ottersten, "Minimum BER linear transceivers for MIMO channels via primal decomposition," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2866–2882, Aug. 2005.

[42] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory—Part II: Algorithms," *IEEE Trans. Signal Processing*, vol. 55, no. 3, pp. 1250–1267, Mar. 2008.

[43] Y. Guo and B. C. Levy, "Robust MSE equalizer design for MIMO communication systems in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1840–1852, May 2006.

[44] R. T. Rockafellar, *Convex Analysis*, 2nd ed. Princeton, NJ: Princeton Univ. Press, 1970.

[45] F. Rendl and H. Wolkowicz, "A semidefinite framework for trust region subproblems with applications to large scale minimization," *Mathematical Programming*, vol. 77, no. 1, pp. 273–299, Apr. 1997.

[46] C. Fortin and H. Wolkowicz, "The trust region subproblems and semidefinite programming," *Optim. Methods and Software*, vol. 19, no. 1, pp. 41–67, Feb. 2004.

[47] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas With Application to Linear Systems Theory*. Princeton, NJ: Princeton Univ. Press, 2005.

Jiaheng Wang (S'08) received the B.E. and M.S. degrees in electrical engineering from the Southeast University, Nanjing, China, in 2001 and 2006, respectively, and the Ph.D. degree from the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology in 2010. His research interests mainly include optimization in signal processing, wireless communications and networks.

Daniel P. Palomar (S'99–M'03–SM'08) received the Electrical Engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He is an Associate Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. He had previously held several research appointments, namely, at King's College London (KCL), London, U.K.; Technical University of Catalonia (UPC), Barcelona; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona, Spain; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza," Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to signal processing and communications.

Dr. Palomar has been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on Convex Optimization for Signal Processing, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on Game Theory in Communication Systems, the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on Optimization of MIMO Transceivers for Realistic Communication Networks. He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM). He was the General Co-Chair of the 2009 IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). He is a recipient of a 2004–2006 Fulbright Research Fellowship; the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society; the 2002–2003 Best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC); the 2002–2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation; and the 2004 prize for the Best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.