Matrix-Monotonic Optimization – Part I: Single-Variable Optimization

Chengwen Xing, *Member, IEEE*, Shuai Wang, *Member, IEEE*, Sheng Chen, *Fellow, IEEE*, Shaodan Ma, H. Vincent Poor, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Matrix-monotonic optimization exploits the monotonic nature of positive semi-definite matrices to derive optimal diagonalizable structures for the matrix variables of matrix-variable optimization problems. Based on the optimal structures derived, the associated optimization problems can be substantially simplified and underlying physical insights can also be revealed. In our work, a comprehensive framework of the applications of matrix-monotonic optimization to multiple-input multiple-output (MIMO) transceiver design is provided for a series of specific performance metrics under various linear constraints. This framework consists of two parts, i.e., Part-I for single-variable optimization and Part-II for multi-variable optimization. In this paper, single-variable matrix-monotonic optimization is investigated under various power constraints and various types of channel state information (CSI) condition. Specifically, three cases are investigated: 1) both the transmitter and receiver have imperfect CSI; 2) perfect CSI is available at the receiver but the transmitter has no CSI; 3) perfect CSI is available at the receiver but the channel estimation error at the transmitter is norm-bounded. In all three cases, the matrix-monotonic optimization framework can be used for deriving the optimal structures of the optimal matrix variables.**

***Index Terms*—Matrix-monotonic optimization, majorization theory, optimal structures, transceiver optimization.**

Manuscript received May 5, 2020; revised August 27, 2020 and September 29, 2020; accepted October 22, 2020. Date of publication November 11, 2020; date of current version February 1, 2021. The associate editor coordinating the review of this article and approving it for publication was Prof. Stefano Tomasin. The work of Chengwen Xing was supported in part by the National Natural Science Foundation of China under Grants 61671058, 61722104, and 61620106001, and in part by Ericsson. The work of Shaodan Ma was supported in part by the Science and Technology Development Fund, Macau SAR (File no. 0036/2019/A1 and File no. SKL-IOTSC2018-2020), and in part by the Research Committee of University of Macau under Grant MYRG2018-00156-FST. The work of H. Vincent Poor was supported by the U.S. National Science Foundation under Grant CCF-1908308. The work of Lajos Hanzo was supported in part by the Engineering and Physical Sciences Research Council projects EP/N004558/1, EP/P034284/1, EP/P034284/1, EP/P003990/1 (COALESCE), of the Royal Society's Global Challenges Research Fund Grant and in part by the European Research Council's Advanced Fellow Grant QuantCom. *(Corresponding author: Shuai Wang.)*

Chengwen Xing is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with the Department of Electrical and Computer Engineering, University of Macau, Macao, S.A.R. 999078, China (e-mail: xingchengwen@gmail.com).

Shuai Wang is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: swang@bit.edu.cn).

Sheng Chen is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K., and also with the King Abdulaziz University, Jeddah 21422, Saudi Arabia (e-mail: sqc@ecs.soton.ac.uk).

Shaodan Ma is with the State Key Laboratory of Internet of Things for Smart City, Department of Electrical and Computer Engineering, University of Macau, Macao, S.A.R. 999078, China (e-mail: shaodanma@umac.mo).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Lajos Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton 430205, U.K. (e-mail: lh@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/TSP.2020.3037513

I. Motivations

ANTENNA arrays are widely employed for improving the bandwidth- and/or the power-efficiency, resulting in the concept of multiple-input multiple-output (MIMO) systems [1]–[9]. Transceiver optimization is of critical importance for fulfilling the potential of MIMO communication systems [7]–[10]. MIMO transceiver optimization hinges on numerous factors, including their implementation issues, the availability of channel state information (CSI) and their system architectures. More specifically, MIMO transceivers can be classified into linear transceivers [8], [9] and nonlinear transceivers [11]–[13]. According to the different levels of CSI knowledge, MIMO transceiver designs can be classified into designs relying on perfect CSI [4]–[6] and designs having partial CSI [15]–[19]. Finally, according to the system architecture, transceiver optimization can be used for point-to-point systems [10], [20], for multi-user (MU) MIMO systems [21], for distributed MIMO systems [22], [23], and for cooperative MIMO systems [24], [25].

In all the above-mentioned multiple antenna aided systems, the corresponding optimization variables become matrix variables [26]. As a result, optimization relying on matrix variables plays an important role in MIMO systems [27]. Optimization relying on matrix variables is generally very challenging and such problems are much more difficult to solve than their counterparts with vector variables or scalar variables, because matrix variable based optimization usually involves complex matrix operations, such as the calculation of the determinants, inverses, matrix decompositions and so on. Furthermore, because of their spatial multiplexing gains, MIMO systems are capable of supporting multiple data streams. This fact makes transceiver optimization problems inherently multi-objective optimization problems. For example, given a limited transmit power, any specific transceiver optimization strikes a tradeoff between the performance of different data streams. This is the reason why there exists a rich body of work addressing various different MIMO transceiver designs [9], [10].

Any transceiver optimization problem hinges on the fundamental elements of the objective function and the specific optimization tools used for finding the extremities of the objective function. The more components the objective function has, the larger the search space becomes, which often makes a full search unrealistic. A third related component is constituted by the constraints. The most widely used objective functions or performance metrics of MIMO transceiver optimization include the classic mean square error (MSE) minimization, signal to interference plus noise ratio (SINR) maximization or mutual information maximization, bit error rate (BER) minimization, etc, [9]. Different performance metrics reflect different design preferences and different tradeoffs among the transmitted data streams [10]. Transceiver optimization problems using different performance metrics imposes different degrees of difficulty to solve. Furthermore, different objective functions also correspond to different implementation strategies resulting in, for example, linear transceivers, nonlinear transceivers using Tomlinson-Harashima precoding (THP) or decision feedback equalizer (DFE) etc. [12]–[14], [26]. Suffice to say that the specific choice of the objective function has a more substantial impact on the overall MIMO design than that of the tools used for optimizing it.

On the other hand, there are many different types of power constraints, such as the sum power constraint [26], per-antenna power constraint [28]–[34], shaping constraint [35], [36], joint power constraints [37], cognitive constraint [34], etc. The most widely used power constraint is the sum power constraint requiring the sum of the powers at all the transmit antennas to be lower than a threshold. In communication systems, usually each antenna has its own amplifier [21]. Therefore, the per-antenna power constraint is more practical than the sum power constraint. However, the per-antenna power constraint is more challenging to consider than the sum power constraint [21], [29]–[31], [34]. The existing literature has revealed that if different transmit antennas have the same statistics, the performance gain of considering the more challenging per-antenna power constraint based design over using the simpler sum power constraint design is negligible [32]. Thus, under the scenario of similar statistics for different transmit antennas, the sum power constraint is an effective modeling technique. It is worth noting however that in some cases, as in distributed antenna systems or heterogeneous networks, different antennas have significantly different statistics, and thus the per-antenna power constraint cannot be replaced by the sum power constraint without a significant performance loss [32], [34]. Moreover, considering other practical constraints, such as signal variances or the peak-to-average-ratio, joint power constraints or other types of constraints have to be taken into account [37].

It can be readily seen from the existing literature [10], [26], [34] that the underlying design principles for various transceiver optimization problems are almost the same. Generally, the main idea is taking advantage of the specific structure of the underlying optimization problem to simplify the transceiver optimization. Optimization theory plays an important role in MIMO transceiver optimization, and in the past decade many elegant results have been derived based on convex optimization theory [23], [27]. Deriving optimal structures is critical in transceiver optimization [1], [6], [9]. Clearly, a general-purpose optimal structure that can cover every MIMO transceiver optimization problems does not exist, and most research has been focused on finding an optimal diagonalizable structure for MIMO transceiver optimization. This is because based on the optimal diagonalizable structures of the MIMO transceivers, the corresponding optimization problems can be substantially simplified and deep underlying physical insights can also be revealed [1], [6], [9].

Again, the optimization variables of MIMO transceiver designs are generally matrix variables. Matrix-monotonic optimization exploits the monotonic nature of positive semi-definite matrices to derive optimal structures of the matrix variables in the underlying optimization problems [26], [34], [36]. Based on matrix-monotonic optimization, the matrix variables can be substantially simplified into vector variables. The optimal structures delivered by matrix-monotonic optimization, therefore, greatly simplify complicated MIMO transceiver designs and make the underlying physical interpretation more transparent. From a matrix-monotonic optimization perspective, MIMO transceiver optimization problems relying on different objective functions and power constraints can be unified and, therefore, their associated optimal structures can be derived using the same matrix-monotonic optimization tool. Explicitly, matrix-monotonic optimization is a powerful mathematical tool conceived for solving challenging matrix-variable transceiver optimization problems.

This paper offers a comprehensive and novel matrix-monotonic optimization framework for a series of specific performance metrics under linear constraints in the context of MIMO transceiver optimization. Explicitly, matrix monotonic optimization problems with various levels of CSI are investigated in depth. Our main contributions are listed as follows.

- In contrast to [26] with only simple sum power constraint, the framework of matrix-monotonic optimization investigated in this treatise is subjected to diverse power constraints, including the sum power constraint, multiple weighted power constraints, joint power constraints and shaping constraints. In other words, the framework investigated in this paper subsumes the solutions in [26] and several other MIMO transceiver optimization solutions as its special cases.
- In contrast to [9] and [11], where the linear and nonlinear transceiver designs are investigated separately under only the sum power constraint, the framework proposed in this paper unifies the families of linear and nonlinear MIMO transceiver optimization under the sum power constraint, shaping constraint, joint power constraints and multiple weighted power constraints.
- Moreover, robust MIMO transceiver optimization relying on partial CSI under various power constraints is investigated based on the matrix-monotonic optimization framework. Specifically, the following three cases are investigated:
  1) Both the transmitter and receiver have only imperfect CSI,
  2) The receiver has perfect CSI but the transmitter has only channel statistics,
  3) The receiver has perfect CSI but the channel estimate available at the transmitter is subject to a certain norm-bounded error.

  Although imperfect CSI makes the MIMO transceiver optimization more complex and challenging, the proposed matrix-monotonic optimization framework is still capable of deriving the underlying optimal structures.

The remainder of this paper is organized as follows. In Section II, we present the fundamentals of the matrix-monotonic optimization framework. Then Section III investigates classic Bayesian robust matrix-monotonic optimization for robust transceiver design when the channel estimation errors are Gaussian distributed. In Section IV, stochastic robust matrix-monotonic optimization is investigated for MIMO transceiver optimization where the receiver has perfect CSI but the transmitter knows only the channel statistics. Section V is devoted to worst-case matrix-monotonic optimization, which focuses on transceiver optimization in the face of norm-bounded channel estimation errors.

Notation: The following notational conventions are adopted throughout our discussions. The normal-faced letters denote scalars, while bold-faced lower-case and upper-case letters denote vectors and matrices, respectively. $\boldsymbol{Z}^{\mathrm{H}}$, $\mathrm{Tr}(\boldsymbol{Z})$ and $|\boldsymbol{Z}|$ denote the Hermitian transpose, trace and determinant of complex matrix $\boldsymbol{Z}$, respectively. Statistical expectation is denoted by $\mathbb{E}\{\cdot\}$, and $a^{+}=\max\{0,a\}$, while $(\cdot)^{\mathrm{T}}$ denotes the vector/matrix transpose operator. $\boldsymbol{Z}^{\frac{1}{2}}$ is the Hermitian square root of $\boldsymbol{Z}$ which is positive semi-definite. The ith largest eigenvalue of $\boldsymbol{Z}$ is denoted by $\lambda_i(\boldsymbol{Z})$, and the ith-row and jth-column element of $\boldsymbol{Z}$ is denoted by $[\boldsymbol{Z}]_{i,j}$, while $\boldsymbol{d}[\boldsymbol{Z}]$ denotes the vector consisting of the diagonal elements of $\boldsymbol{Z}$ and $\mathrm{diag}\{\{\boldsymbol{A}_k\}_{k=1}^{K}\}$ denotes the block diagonal matrix whose diagonal sub-matrices are $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K$. The symbol $\boldsymbol{d}^2[\boldsymbol{Z}]$ denotes the vector consisting of the squared moduli of the diagonal elements of $\boldsymbol{Z}$. Additionally, the ith element of a vector $\boldsymbol{z}$ is denoted by $[\boldsymbol{z}]_i$. The identity matrix of appropriate dimension is denoted by $\boldsymbol{I}$, and $\otimes$ is the Kronecker product. In this paper, $\boldsymbol{\Lambda}$ always denotes a diagonal matrix, and the expressions $\boldsymbol{\Lambda}\searrow$ and $\boldsymbol{\Lambda}\nearrow$ represent a rectangular or square diagonal matrix with the diagonal elements in descending order and ascending order, respectively.

II. Fundamentals of Matrix-Monotonic Optimization

An optimization problem with a real-valued objective function $f_0(\cdot)$ that depends on a complex matrix variable $\boldsymbol{X}$ is generally formulated as

$$\begin{aligned}\min_{\boldsymbol{X}\in\mathcal{C}}\ & f_0(\boldsymbol{X}),\\ \text{s.t.}\ & \psi_i(\boldsymbol{X})\le 0, 1\le i\le I,\end{aligned}\tag{1}$$

where $\psi_i(\cdot)$, $1\le i\le I$, are the constraint functions and $\mathcal{C}$ denotes the complex matrix set. A wide range of optimization problems can be cast in this optimization framework, including the classic MIMO transceiver optimization [10], training designs [26], MIMO radar waveform optimization [26], etc. In order to analyze the properties of this generic optimization problem, we first discuss two of its basic components, namely, the objective function and the constraints, separately.

A. Objective Functions

The objective function reflects the cost or utility of the optimization problem. In this paper, all the optimization problems discussed are formulated with the objective of minimizing a cost function. Let us now discuss the commonly used objective functions, listed in Table I. For transceiver optimization, the mutual information is one of the most important performance metrics. For training optimization, the mutual information is also an important performance metric as it reflects the correlation between the estimated parameters and the true parameters. In these cases, the objective function is given by **Obj. 1** [38], where $\boldsymbol{\Pi}$ and $\boldsymbol{\Phi}$ are constant positive semi-definite matrices which have different physical meanings for different systems. The MSE is another important performance metric for transceiver or training optimization, which reflects how accurately a signal can be recovered rather than how much information can be transmitted. For the optimization problem of sum MSE minimization, the objective function is given in the form of **Obj. 2** [38].

Generally, the MSE formulation for linear transceiver optimization is determined by the specific signal model considered. For example, in a dual-hop AF MIMO relaying network, the MSE minimization has **Obj. 3** [39], where α is a positive scalar and $\boldsymbol{A}$ is a constant complex matrix. Similarly, the mutual information maximization for a dual-hop AF MIMO relaying network aims at minimizing the objective function **Obj. 4** [39][1]. For linear transceiver optimization, to realize different levels of fairness between different transmitted data streams, a general objective function can be formulated as an additively Schur-convex function [9] or additively Schur-concave function [9] of the diagonal elements of the MSE matrix, which are given by **Obj. 5.1** and **Obj. 5.2** [10], respectively. The additively Schur-convex function $f_{\text{A-Schur}}^{\text{convex}}(\cdot)$ and the additively Schur-concave function $f_{\text{A-Schur}}^{\text{concave}}(\cdot)$ represent different levels of fairness among the diagonal elements of the data MSE matrix. In addition, $f_{\text{A-Schur}}^{\text{convex}}(\cdot)$ and $f_{\text{A-Schur}}^{\text{concave}}(\cdot)$ are both increasing functions with respect to the vector variables.

When nonlinear transceivers are chosen for improving the BER performance at the cost of increased complexity, e.g., THP or DFE, the objective functions of the transceiver optimization can be formulated as a multiplicative Schur-convex function or a multiplicative Schur-concave function of the vector consisting of the squared diagonal elements of the Cholesky-decomposition triangular matrix of the MSE matrix, that is, **Obj. 6.1** and **Obj. 6.2** [26], respectively, where $\boldsymbol{L}$ is a lower triangular matrix. The multiplicatively Schur-convex function $f_{\text{M-Schur}}^{\text{convex}}(\cdot)$ and the multiplicatively Schur-concave function $f_{\text{M-Schur}}^{\text{concave}}(\cdot)$ reflect the different levels of fairness among the different data streams, i.e., different tradeoffs among the performance of different data steams [26]. In addition, $f_{\text{M-Schur}}^{\text{convex}}(\cdot)$ and $f_{\text{M-Schur}}^{\text{concave}}(\cdot)$ are both increasing functions with respect to the vector variables.

TABLE I
The Objective Functions

Index	Objective function
Obj. 1	$-\log\left\|\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\boldsymbol{\Phi}\right\|$
Obj. 2	$\mathrm{Tr}\left(\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\boldsymbol{\Phi}\right)^{-1}\right)$
Obj. 3	$\mathrm{Tr}\left(\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\alpha\boldsymbol{I}\right)^{-1}\boldsymbol{A}\right)$
Obj. 4	$\log\left\|\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\alpha\boldsymbol{I}\right)^{-1}\boldsymbol{A}+\boldsymbol{\Phi}\right\|$
Obj. 5.1	$f_{\text{A-Schur}}^{\text{Convex}}\left(\boldsymbol{d}\left[\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\alpha\boldsymbol{I}\right)^{-1}\right]\right)$
Obj. 5.2	$f_{\text{A-Schur}}^{\text{Concave}}\left(\boldsymbol{d}\left[\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\alpha\boldsymbol{I}\right)^{-1}\right]\right)$
Obj. 6.1	$f_{\text{M-Schur}}^{\text{Convex}}\left(\boldsymbol{d}^2[\boldsymbol{L}]\right)$, $\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\alpha\boldsymbol{I}\right)^{-1}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}$
Obj. 6.2	$f_{\text{M-Schur}}^{\text{Concave}}\left(\boldsymbol{d}^2[\boldsymbol{L}]\right)$, $\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\alpha\boldsymbol{I}\right)^{-1}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}$
Obj. 7	$-\log\left\|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\boldsymbol{A}+\boldsymbol{\Phi}\right\|$
Obj. 8	$\mathrm{Tr}\left(\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\boldsymbol{A}+\alpha\boldsymbol{I}\right)^{-1}\right)$
Obj. 9	$\mathrm{Tr}\left(\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}+\boldsymbol{\Phi}\right)^{-1}\boldsymbol{A}\right)$
Obj. 10	$-\log\left\|\boldsymbol{\Phi}\otimes\boldsymbol{\Sigma}_1+\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\right)\otimes\boldsymbol{\Sigma}_3\right\|$
Obj. 11	$-\log\left\|\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Phi}+\boldsymbol{\Sigma}_2\otimes\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\right)\right\|$
Obj. 12	$\mathrm{Tr}\left(\left(\boldsymbol{\Phi}\otimes\boldsymbol{\Sigma}_1+\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\right)\otimes\boldsymbol{\Sigma}_2\right)^{-1}\right)$
Obj. 13	$\mathrm{Tr}\left(\left(\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Phi}+\boldsymbol{\Sigma}_2\otimes\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\right)\right)^{-1}\right)$
Obj. 14	$\mathrm{Tr}\left(\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)\otimes\boldsymbol{\Sigma}_1\left(\boldsymbol{I}+\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\right)\otimes\boldsymbol{\Sigma}_2\right)^{-1}\right)$
Obj. 15	$\mathrm{Tr}\left(\boldsymbol{\Sigma}_1\otimes\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)\left(\boldsymbol{I}+\boldsymbol{\Sigma}_2\otimes\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\right)\right)^{-1}\right)$

[1]This conclusion is based on the fact that maximizing mutual information is equivalent to minimizing the determinant of the MSE matrix [26].

In wireless communication designs, even for the same system or the same optimization problem, the mathematical formulae are not unique. More specifically, for the mutual information maximization, we have the alternative objective function **Obj. 7** [26]. Similarly, the sum MSE minimization has the alternative objective function **Obj. 8** [26]. Moreover, the weighted MSE minimization can be considered as a general extension of the sum MSE minimization by introducing a weighting matrix, which has the objective function **Obj. 9**.

As discussed in the existing literature, some MIMO system optimization problems may involve Kronecker products due to $\mathrm{vec}(\cdot)$ operations [26]. The optimization problems relying on Kronecker product usually look very complicated. In this paper, the pair of optimization problems relying on either the matrix determinant or on the matrix trace are discussed that involve Kronecker products. Based on **Obj. 1**, we have the extended Kronecker structured objective function **Obj. 10**, which is equivalent to **Obj. 11** [26]. It can readily be seen that with the choice of $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, **Obj. 10** and **Obj. 11** are equivalent to **Obj. 1**. In this paper, we also consider a more general case in which $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ have the same eigenvalue decomposition (EVD) unitary matrix. Under this assumption and based on **Obj. 2**, we have the extended Kronecker structured objective function **Obj. 12**, which is equivalent to **Obj. 13**. Similarly, based on **Obj. 3**, we have the objective function **Obj. 14**, which is also equivalent to **Obj. 15**. In our following discussions involving **Obj. 10** to **Obj. 15**, it is always assumed that $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ have the same EVD unitary matrix.

B. Constraint Functions

In practical communication system designs, typically the associated optimization problems have constraints, and these constraints have different physical meanings for different communication systems.

The most natural constraints are the power constraints, since practical amplifiers have certain maximum transmit power thresholds. The simplest power constraint is the sum power constraint, which can be expressed as

$$\textbf{Constraint1}: \ \mathrm{Tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right) \leq P. \tag{2}$$

For the sum power constraint, the optimization problems associated with training sequence designs or transceiver designs are subjected to the constraint of the sum power of all the transmit antennas. In practical systems, each antenna has its own power amplifier and, therefore, the per-antenna power constraints or individual power constraints provide a more reasonable power constraint model, which is expressed as

$$\textbf{Constraint2}: \ \left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right]_{n,n} \leq P_n, \ \ n = 1, \ldots, N, \tag{3}$$

where we have assumed that the number of transmit antennas is N and the matrix variable $\boldsymbol{X}$ has N rows. The per-antenna power constraint (3) may be more practical but it does not include the sum power constraint (2) as its special case.

In sophisticated communication networks, the constraints are not limited to reflect the maximum power constraints at the transmit antennas for the desired signal but they also reflect many other constraints such as the interference constraints between adjacent links. A more general power constraint is the following one having multiple weighted components [38]

$$\textbf{Constraint3}: \ \mathrm{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right) \leq P_i, \ \ i = 1, \ldots, I, \tag{4}$$

where I is the number of weighted power constraints. **Constraint 3** is more general than **Constraint 1** and **Constraint 2**. The constraint model (4) includes the sum power constraint (2) and per-antenna power constraint (3) as its special cases. Specifically, by choosing $I = 1$ and $\boldsymbol{\Omega}_1 = \boldsymbol{I}$, this power constraint model becomes the sum power constraint (2). Furthermore, when $I = N$ and $\boldsymbol{\Omega}_i$ is the matrix whose ith diagonal element is one and all the other elements are zeros, this model is exactly the per-antenna power constraint (3).

In order to avoid or control the interference, it is expected to be cast to the null space of the desired signals, hence the signal and interference become orthogonal to each other. In order to achieve this, constraints can be imposed on the covariance matrix of the transmitted signal, which are referred to as spectral mask constraints [35]. A classic example is the shaping constraint, which is formulated as the following matrix inequality [35], [36]

$$\textbf{Constraint4}: \ \boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} \preceq \boldsymbol{R}_{\mathrm{s}}. \tag{5}$$

From matrix inequality theory, this constraint is equivalent to [40, 471]

$$\mathrm{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right) \leq \mathrm{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{R}_{\mathrm{s}}\right), \tag{6}$$

for any positive semi-definite matrix $\boldsymbol{\Omega}_i$. Based on this fact, we can argue that the shaping constraint represents a special case of the multiple weighted power constraint. A simplified version of **Constraint 4** is the constraint imposed on the eigenvalues of the covariance matrix $\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}$ formulated as

$$\textbf{Constraint5}: \ \lambda_i\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right) \leq \tau_i. \tag{7}$$

A widely used eigenvalue constraint is the constraint on the maximum eigenvalue, $\lambda_1(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}) \leq \tau_1$, which is equivalent to [36]

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} \preceq \tau_1\boldsymbol{I}. \tag{8}$$

This constraint can be used together with the sum power constraint to limit the transmitter's peak power. This is because most of the existing power constraints are based on statistical averages, while from a practical implementation perspective, the power constraint is an instantaneous constraint instead of being an average one [37]. This kind of combined power constraint is termed as the joint power constraint, which is expressed as [36]

$$\textbf{Constraint 6}: \mathrm{Tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right) \leq P, \boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} \preceq \tau_1\boldsymbol{I}. \tag{9}$$

In cognitive radio communications, the interference imposed by the secondary user on the primary user must be smaller than a threshold and this constraint can be written in the following form

$$\textbf{Constraint 7}: \ \mathrm{Tr}\left(\boldsymbol{H}_{\mathrm{c}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{c}}^{\mathrm{H}}\right) \leq \tau_{\mathrm{C}}, \tag{10}$$

where $\boldsymbol{H}_{\mathrm{c}}$ is the channel matrix between the secondary user and primary user, while τ_{C} is the interference threshold. This kind of constraint is also a special case of **Constraint 3**.

In summary, all the power constraint models discussed above represent the different physical constraints on the covariance matrix of the transmit signal, which equals $\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}$. These constraints shape the positive semidefinite covariance matrix. For the simplest sum power constraint model, the sum of the eigenvalues of the covariance matrix has to be smaller than a threshold. For the multiple weighted power constraint model, the eigenvalues of the covariance matrices are constrained in the polyhedron region constructed by the multiple weighting matrices. In this case, except for the restrictions on the eigenvalues, the constraints also restrict the unitary matrix in the eigenvalue decomposition of the covariance matrix. Moreover, for the joint power constraint model the sum of the eigenvalues and the maximum eigenvalue are simultaneously smaller than the predefined thresholds. The upper-bound on the maximum eigenvalue significantly impacts the power allocations on the eigenchannels. For example, some subchannels that are allocated zero power for the sum power constraint will be assigned non-zero powers for the joint power constraints.

Before turning our attention to the optimization problem (1), two fundamental definitions are first introduced.

Definition 1: A constraint $\psi(\boldsymbol{X}) \leq 0$ is a left unitary invariant constraint if we have

$$\psi\left(\boldsymbol{Q}_{\mathrm{L}}\boldsymbol{X}\right) = \psi(\boldsymbol{X}), \tag{11}$$

where $\boldsymbol{Q}_{\mathrm{L}}$ is an arbitrary unitary matrix.

Definition 2: A constraint $\psi(\boldsymbol{X}) \leq 0$ is a right unitarily-invariant constraint if we have

$$\psi\left(\boldsymbol{X}\boldsymbol{Q}_{\mathrm{R}}\right) = \psi(\boldsymbol{X}), \tag{12}$$

where $\boldsymbol{Q}_{\mathrm{R}}$ is an arbitrary unitary matrix.

It is worth noting that all the constraints discussed above are right unitarily-invariant. Specifically, in **Constraints 1** to **7**, after replacing $\boldsymbol{X}$ by $\boldsymbol{X}\boldsymbol{Q}_{\mathrm{R}}$ it can be concluded that these constraints do not change. Therefore, we can focus our attention on the family of right unitarily-invariant constraints only. In particular, we will focus our attention on the shaping constraint, joint power constraints and multiple weighted power constraints.

C. Matrix-Monotonic Optimization

Based on the above discussions, with the objective functions in Table I, the generic optimization problem of MIMO systems can be formulated as

$$\mathbf{Opt.\,1.1}: \min_{\boldsymbol{X}} f\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{X}\right), \text{s.t. } \psi_i(\boldsymbol{X}) \leq 0,\ 1 \leq i \leq I. \tag{13}$$

The function $f(\cdot)$ is matrix monotone decreasing function [27], [41], [42]. Since the constraints are right unitarily-invariant, we introduce the auxiliary matrix variable $\boldsymbol{F}$ and express the original matrix variable $\boldsymbol{X}$ as

$$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}, \tag{14}$$

where $\boldsymbol{Q}_{\boldsymbol{X}}$ is an arbitrary unitary matrix. Based on (14), the optimization problem (13) can be reformulated as

$$\begin{aligned} \min_{\boldsymbol{F},\boldsymbol{Q}_{\boldsymbol{X}}} \quad & f\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right), \\ \text{s.t.} \quad & \psi_j\left(\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right) = \psi_i\left(\boldsymbol{F}\right) \leq 0,\ 1 \leq i \leq I, \end{aligned} \tag{15}$$

where the specific objective functions are given in the left column of Table II. Note that the constraints do not depend on $\boldsymbol{Q}_{\boldsymbol{X}}$. Therefore, the optimal $\boldsymbol{Q}_{\boldsymbol{X}}$ is independent of the constraints.

1) Optimization of $\boldsymbol{Q}_{\boldsymbol{X}}$: Generally, there are two basic approaches to optimize $\boldsymbol{Q}_{\boldsymbol{X}}$. The first one is based on the basic matrix inequality and the other is based on majorization theory.

Basic Matrix Inequalities Typically, the extreme values of basic matrix operations e.g., trace, determinant, etc., are functions of the eigenvalues of the matrices involved. Given the positive semi-definite matrices $\boldsymbol{C} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{D} \in \mathbb{C}^{N\times N}$, we consider the following EVDs

$$\boldsymbol{C} = \boldsymbol{U}_{\boldsymbol{C}}\boldsymbol{\Lambda}_{\boldsymbol{C}}\boldsymbol{U}_{\boldsymbol{C}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{C}} \searrow, \tag{16}$$

$$\boldsymbol{D} = \boldsymbol{U}_{\boldsymbol{D}}\boldsymbol{\Lambda}_{\boldsymbol{D}}\boldsymbol{U}_{\boldsymbol{D}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{D}} \searrow, \tag{17}$$

$$\boldsymbol{D} = \bar{\boldsymbol{U}}_{\boldsymbol{D}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{D}}\bar{\boldsymbol{U}}_{\boldsymbol{D}}^{\mathrm{H}} \text{ with } \bar{\boldsymbol{\Lambda}}_{\boldsymbol{D}} \nearrow, \tag{18}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{D}}$ and $\bar{\boldsymbol{\Lambda}}_{\boldsymbol{D}}$ consist of the eigenvalues of $\boldsymbol{D}$ arranged in descending order and ascending order, while $\boldsymbol{U}_{\boldsymbol{D}}$ and $\bar{\boldsymbol{U}}_{\boldsymbol{D}}$ contain the corresponding eigenvectors of $\boldsymbol{D}$, respectively. Then we have the four basic matrix inequalities, ranging from (19) to (22), shown at the bottom of this page. Furthermore, in both **Matrix Inequality 1** [43, P340, P341] and **Matrix Inequality 2** [26, Appendix A], the left equality holds when $\boldsymbol{U}_{\boldsymbol{C}} = \bar{\boldsymbol{U}}_{\boldsymbol{D}}$, and the right equality holds when $\boldsymbol{U}_{\boldsymbol{C}} = \boldsymbol{U}_{\boldsymbol{D}}$; while in both **Matrix Inequality 3** [43, P333, P334] and **Matrix Inequality 4**, the left equality holds when $\boldsymbol{U}_{\boldsymbol{C}} = \boldsymbol{U}_{\boldsymbol{D}}$, and the right equality holds when $\boldsymbol{U}_{\boldsymbol{C}} = \bar{\boldsymbol{U}}_{\boldsymbol{D}}$ [26].

Majorization Theory Majorization theory constitutes an important branch of matrix equality theory [27], [43]. We have the following two important definitions.

Definition 3 ([43]): For two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, $\boldsymbol{x}$ is said to be majorized by $\boldsymbol{y}$, denoted as $\boldsymbol{x} \prec \boldsymbol{y}$, when the following inequalities are satisfied: $\bigcirc_{i=1}^{k}[\boldsymbol{x}]_i \leq \bigcirc_{i=1}^{k}[\boldsymbol{y}]_i$, for $1 \leq k \leq N-1$, and $\bigcirc_{i=1}^{N}[\boldsymbol{x}]_i = \bigcirc_{i=1}^{N}[\boldsymbol{y}]_i$, where $\bigcirc$ denotes a mathematical operator.

In the following, we only consider the addition and product operators of $\bigcirc = \sum$ and $\bigcirc = \prod$.

Definition 4 ([43]): A real-valued function $\phi: \mathbb{R}^N \to \mathbb{R}$ is additively or multiplicatively Schur-convex for any $\boldsymbol{x}, \boldsymbol{y}$ in the feasible set, $\boldsymbol{x} \prec \boldsymbol{y} \to \phi(\boldsymbol{x}) \leq \phi(\boldsymbol{y})$. On the other hand, ϕ is additively or multiplicatively Schur-concave when $\boldsymbol{x} \prec \boldsymbol{y} \to \phi(\boldsymbol{x}) \geq \phi(\boldsymbol{y})$.

Optimal $\boldsymbol{Q}_{\boldsymbol{X}}$ Based on the basic matrix inequalities and majorization theory together with the following EVDs (23) to (26) and the singular value decomposition (SVD) (27)

$$\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F} = \boldsymbol{U}_{\boldsymbol{F}\boldsymbol{\Pi}\boldsymbol{F}}\boldsymbol{\Lambda}_{\boldsymbol{F}\boldsymbol{\Pi}\boldsymbol{F}}\boldsymbol{U}_{\boldsymbol{F}\boldsymbol{\Pi}\boldsymbol{F}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{F}\boldsymbol{\Pi}\boldsymbol{F}} \searrow, \tag{23}$$

$$\boldsymbol{\Phi} = \boldsymbol{U}_{\boldsymbol{\Phi}}\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}\boldsymbol{U}_{\boldsymbol{\Phi}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{\Phi}} \searrow, \tag{24}$$

$$\boldsymbol{\Phi} = \bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Phi}}\bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}^{\mathrm{H}} \text{ with } \bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Phi}} \nearrow, \tag{25}$$

$$\boldsymbol{A}\boldsymbol{\Phi}^{-1}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{U}_{\boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{A}}\boldsymbol{U}_{\boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{A}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{A}} \searrow, \tag{26}$$

$$\boldsymbol{A} = \boldsymbol{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{V}_{\boldsymbol{A}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{A}} \searrow, \tag{27}$$

the optimal unitary matrices $\boldsymbol{Q}_{\boldsymbol{X}}$ corresponding to the various objective functions can be derived and they are listed in the right

$$\mathbf{Matrix\ Inequality\ 1}: \sum\nolimits_{i=1}^{N} \lambda_{i-1+N}(\boldsymbol{C})\lambda_i(\boldsymbol{D}) \leq \mathrm{Tr}(\boldsymbol{C}\boldsymbol{D}) \leq \sum\nolimits_{i=1}^{N} \lambda_i(\boldsymbol{C})\lambda_i(\boldsymbol{D}), \tag{19}$$

$$\mathbf{Matrix\ Inequality\ 2}: \sum\nolimits_{i=1}^{N} \left(\lambda_{i-1+N}(\boldsymbol{C}) + \lambda_i(\boldsymbol{D})\right)^{-1} \leq \mathrm{Tr}\left((\boldsymbol{C}+\boldsymbol{D})^{-1}\right) \leq \sum\nolimits_{i=1}^{N} \left(\lambda_i(\boldsymbol{C}) + \lambda_i(\boldsymbol{D})\right)^{-1}, \tag{20}$$

$$\mathbf{Matrix\ Inequality\ 3}: \prod\nolimits_{i=1}^{N} \left(\lambda_i(\boldsymbol{C}) + \lambda_i(\boldsymbol{D})\right) \leq |\boldsymbol{C}+\boldsymbol{D}| \leq \prod\nolimits_{i=1}^{N} \left(\lambda_{i-1+N}(\boldsymbol{C}) + \lambda_i(\boldsymbol{D})\right), \tag{21}$$

$$\mathbf{Matrix\ Inequality\ 4}: \prod\nolimits_{i=1}^{N} \left(\lambda_{i-1+N}(\boldsymbol{C})\lambda_i(\boldsymbol{D}) + 1\right) \leq |\boldsymbol{C}\boldsymbol{D}+\boldsymbol{I}| \leq \prod\nolimits_{i=1}^{N} \left(\lambda_i(\boldsymbol{C})\lambda_i(\boldsymbol{D}) + 1\right). \tag{22}$$

TABLE II
THE OBJECTIVE FUNCTIONS AND THE OPTIMAL UNITARY MATRICES $\boldsymbol{Q_X}$

Index	Objective function	Optimum $\boldsymbol{Q_X}$
Obj. 1	$-\log\left\|\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\boldsymbol{\Phi}\right\|$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}^{\mathrm{H}}$
Obj. 2	$\mathrm{Tr}\left(\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\boldsymbol{\Phi}\right)^{-1}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}^{\mathrm{H}}$
Obj. 3	$\mathrm{Tr}\left(\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\alpha\boldsymbol{I}\right)^{-1}\boldsymbol{A}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\boldsymbol{A}}^{\mathrm{H}}$
Obj. 4	$\log\left\|\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\alpha\boldsymbol{I}\right)^{-1}\boldsymbol{A}+\boldsymbol{\Phi}\right\|$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\boldsymbol{A\Phi A}}^{\mathrm{H}}$
Obj. 5.1	$f_{\text{A-Schur}}^{\text{Convex}}\left(\boldsymbol{d}\left[\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\alpha\boldsymbol{I}\right)^{-1}\right]\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\mathrm{DFT}}^{\mathrm{H}}$
Obj. 5.2	$f_{\text{A-Schur}}^{\text{Concave}}\left(\boldsymbol{d}\left[\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\alpha\boldsymbol{I}\right)^{-1}\right]\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}$
Obj. 6.1	$f_{\text{M-Schur}}^{\text{Convex}}\left(\boldsymbol{d}^2[\boldsymbol{L}]\right)$ with $\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\alpha\boldsymbol{I}\right)^{-1}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\mathrm{GMD}}^{\mathrm{H}}$
Obj. 6.2	$f_{\text{M-Schur}}^{\text{Concave}}\left(\boldsymbol{d}^2[\boldsymbol{L}]\right)$ with $\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\alpha\boldsymbol{I}\right)^{-1}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}$
Obj. 7	$-\log\left\|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\boldsymbol{A}+\boldsymbol{\Phi}\right\|$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\boldsymbol{A\Phi A}}^{\mathrm{H}}$
Obj. 8	$\mathrm{Tr}\left(\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\boldsymbol{A}+\alpha\boldsymbol{I}\right)^{-1}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\boldsymbol{A}}^{\mathrm{H}}$ (High SNR)
Obj. 9	$\mathrm{Tr}\left(\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\boldsymbol{\Phi}\right)^{-1}\boldsymbol{A}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\boldsymbol{A}}^{\mathrm{H}}$ (High SNR)
Obj. 10	$-\log\left\|\boldsymbol{\Phi}\otimes\boldsymbol{\Sigma}_1+\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)\otimes\boldsymbol{\Sigma}_2\right\|$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}^{\mathrm{H}}$
Obj. 11	$-\log\left\|\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Phi}+\boldsymbol{\Sigma}_2\otimes\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)\right\|$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}^{\mathrm{H}}$
Obj. 12	$\mathrm{Tr}\left(\left(\boldsymbol{\Phi}\otimes\boldsymbol{\Sigma}_1+\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)\otimes\boldsymbol{\Sigma}_2\right)^{-1}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}^{\mathrm{H}}$
Obj. 13	$\mathrm{Tr}\left(\left(\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Phi}+\boldsymbol{\Sigma}_2\otimes\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)\right)^{-1}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\bar{\boldsymbol{U}}_{\boldsymbol{\Phi}}^{\mathrm{H}}$
Obj. 14	$\mathrm{Tr}\left(\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)\otimes\boldsymbol{\Sigma}_1\left(\boldsymbol{I}+\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)\otimes\boldsymbol{\Sigma}_2\right)^{-1}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\boldsymbol{A}}^{\mathrm{H}}$
Obj. 15	$\mathrm{Tr}\left(\boldsymbol{\Sigma}_1\otimes\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)\left(\boldsymbol{I}+\boldsymbol{\Sigma}_2\otimes\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)\right)^{-1}\right)$	$\boldsymbol{U}_{\boldsymbol{F\Pi F}}\boldsymbol{U}_{\boldsymbol{A}}^{\mathrm{H}}$

column of Table II[2]. In the SVD (27), $\boldsymbol{\Lambda}_{\boldsymbol{A}}$ contains the singular values of $\boldsymbol{A}$, while $\boldsymbol{U}_{\boldsymbol{A}}$ and $\boldsymbol{V}_{\boldsymbol{A}}$ are the corresponding left and right unitary matrices, respectively.

In Table II, the unitary $\boldsymbol{U}_{\mathrm{DFT}}$ for **Obj. 5.1** is a discrete Fourier transform (DFT) matrix, and $\boldsymbol{U}_{\mathrm{GMD}}$ for **Obj. 6.1** is the unitary matrix that makes the diagonal elements of $\boldsymbol{L}$ identical, that is, $\boldsymbol{U}_{\mathrm{GMD}}$ is the right unitary matrix of the geometric mean decomposition (GMD) of $(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}+\alpha\boldsymbol{I})^{-0.5}$. It is also worth highlighting that for **Obj. 8** and **Obj. 9**, in general, the closed-form optimal $\boldsymbol{Q}_{\boldsymbol{X}}$ cannot be derived, and only the approximated optimal solutions can be obtained at high signal-to-noise ratio (SNR) conditions.

2) Optimization of $\boldsymbol{F}$: For **Opt. 1.1**, given the optimal $\boldsymbol{Q}_{\boldsymbol{X}}$ in Table II, the objective functions in Table II are monotonically decreasing functions with respect to the eigenvalues of $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}$. Therefore, the optimal solutions of $\boldsymbol{F}$ fall in the Pareto optimal solution set of the following multi-objective optimization problem [26]

$$\mathbf{Opt.\ 1.2}:\max_{\boldsymbol{F}}\lambda\left(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\right),\ \text{s.t.}\psi_j(\boldsymbol{F})\leq 0,\ 1\leq j\leq I,\tag{28}$$

where $\lambda(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F})=[\lambda_1(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F})\cdots\lambda_N(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F})]^{\mathrm{T}}$. Clearly, the optimal structure of $\boldsymbol{F}$ depends on both the objective function and on the constraints. As discussed in [26], deriving the optimal structure of $\boldsymbol{F}$ for **Opt. 1.2** corresponds to deriving the optimal structures of $\boldsymbol{F}$ for **Opt. 1.1** for various objectives functions, including **Obj. 1** to **Obj. 15**.

Since $\psi_j(\boldsymbol{F})$ is right unitarily-invariant, **Opt. 1.2** is equivalent to the following matrix-monotonic optimization problem

$$\mathbf{Opt.\ 1.3}:\max_{\boldsymbol{F}}\quad\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F},\ \text{s.t.}\psi_j(\boldsymbol{F})\leq 0,\ 1\leq j\leq I.\tag{29}$$

Generally, matrix-monotonic optimization maximizes a positive semi-definite matrix under certain power constraints. The fundamental idea of matrix-monotonic optimization is to extend the objective functions and solution sets to get more freedoms in return that can be exploited to simplify the analysis. The optimal solutions of **Opt. 1.1** for the objective functions **Obj. 1** to **Obj. 15** are all in the Pareto optimal solution set of **Opt. 1.3**. Since matrix-monotonic optimization derives the common structure of the Pareto optimal solution set of **Opt. 1.3**, the common optimal structures derived are exactly the structures of the optimal solutions of **Opt. 1.1**. By taking advantage of these optimal structures, **Opt. 1.1** can be substantially simplified.

Interestingly, $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}$ can be interpreted as a matrix version SNR [26]. Thus, based on **Opt. 1.3** it can be concluded that various MIMO transceiver optimization problems maximize this matrix version SNR. When there are multiple data streams, maximizing the matrix version SNR inherently constitutes a multi-objective optimization problem. In addition, each unitary matrix $\boldsymbol{Q}_{\boldsymbol{X}}$ corresponds to a specific implementation scheme. The focus of matrix-monotonic optimization is how to maximize the positive semi-definite matrix $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}$ under certain constraints. Different objective functions realize different tradeoffs among the multiple data streams, and matrix-monotonic optimization is

[2]Note that the solutions of **Obj. 8** and **Obj. 9** are derived based on high SNR approximation as the effects of $\alpha\boldsymbol{I}$ and $\boldsymbol{\Phi}$ are neglected.

a powerful tool that unifies the different constrained optimization problems with various objective functions. Specifically, based on matrix-monotonic optimization, the common properties of these objective functions are revealed, which are reflected on the optimal diagonalizable structures.

These structures can transform complex optimization problems relying on matrix variables into much simpler ones with only vector variables. Thus case-by-case investigations for different objective functions are avoided. Since the optimal structure of $\boldsymbol{F}$ also depends on the specific form of the constraints, in the following, three right unitary invariant constraints are investigated, namely, the shaping constraint [36], joint power constraint [36] and multiple weighted power constraints [34].

Shaping Constraint For the shaping constraint, i.e., **Constraint 4**, **Opt. 1.3** becomes the following optimization problem [36]

$$\mathbf{Opt.\,1.4}: \max_{\boldsymbol{F}} \boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}, \mathrm{s.t.} \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} \preceq \boldsymbol{R}_{\mathrm{s}}. \tag{30}$$

The following lemma reveals the optimal structure of $\boldsymbol{F}$ for **Opt. 1.4** with the shaping constraint.

Lemma 1: When $\boldsymbol{R}_{\mathrm{s}}$ is attainable, i.e. the rank of $\boldsymbol{R}_{\mathrm{s}}$ is not higher than the number of columns and the number of rows in $\boldsymbol{F}$, the optimal solution $\boldsymbol{F}_{\mathrm{opt}}$ of **Opt. 1.4** is a square root of $\boldsymbol{R}_{\mathrm{s}}$, i.e., $\boldsymbol{F}_{\mathrm{opt}}\boldsymbol{F}_{\mathrm{opt}}^{\mathrm{H}} = \boldsymbol{R}_{\mathrm{s}}$.

Proof: Since the shaping constraint in **Opt. 1.4** is right unitarily-invariant for $\boldsymbol{F}$, the objective is equivalent to maximizing $\lambda(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F})$, which is in turn equivalent to maximizing $\lambda(\boldsymbol{\Pi}^{1/2}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}^{1/2})$. As $\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} \preceq \boldsymbol{R}_{\mathrm{s}}$, it can be concluded that $\lambda(\boldsymbol{\Pi}^{1/2}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}^{1/2}) \preceq \lambda(\boldsymbol{\Pi}^{1/2}\boldsymbol{R}_{\mathrm{s}}\boldsymbol{\Pi}^{1/2})$, in which the equality holds when $\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} = \boldsymbol{R}_{\mathrm{s}}$. When the rank of $\boldsymbol{R}_{\mathrm{s}}$ is not higher than the number of columns and the number of rows in $\boldsymbol{F}$, the optimal solution $\boldsymbol{F}_{\mathrm{opt}}$ is a square root of $\boldsymbol{R}_{\mathrm{s}}$. It is worth noting that the square roots of $\boldsymbol{R}_{\mathrm{s}}$ are not unique. There are many square roots of $\boldsymbol{R}_{\mathrm{s}}$, however the different square roots have the same performance. We can choose an arbitrary square root of $\boldsymbol{R}_{\mathrm{s}}$ without performance loss. ∎

Joint Power Constraint Under the joint power constraint, **Constraint 6**, **Opt. 1.3** can be rewritten as

$$\mathbf{Opt.\,1.5}\ \max_{\boldsymbol{F}} \boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}, \mathrm{s.t.} \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right) \leq P, \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} \preceq \tau\boldsymbol{I}. \tag{31}$$

The optimal solution $\boldsymbol{F}_{\mathrm{opt}}$ for **Opt. 1.5** is given in Lemma 2.

Lemma 2: For **Opt. 1.5** with the joint power constraint, the Pareto optimal solutions satisfy the following structure

$$\boldsymbol{F}_{\mathrm{opt}} = \boldsymbol{U}_{\boldsymbol{\Pi}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}}, \tag{32}$$

where the unitary matrix $\boldsymbol{U}_{\boldsymbol{\Pi}}$ is specified by the EVD

$$\boldsymbol{\Pi} = \boldsymbol{U}_{\boldsymbol{\Pi}}\boldsymbol{\Lambda}_{\boldsymbol{\Pi}}\boldsymbol{U}_{\boldsymbol{\Pi}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{\Pi}} \searrow, \tag{33}$$

every diagonal element of the rectangular diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}}$ is smaller than $\sqrt{\tau}$, and $\boldsymbol{U}_{\mathrm{Arb}}$ is an arbitrary unitary matrix having the appropriate dimension.

Proof: The proof is given in Appendix A. ∎

Remark 1: For the optimization problem only under the sum power constraint, the optimal structure for $\boldsymbol{F}_{\mathrm{opt}}$ is also specified by (32), where the sum of the diagonal elements of $\boldsymbol{\Lambda}_{\boldsymbol{F}}$ is no larger than P.

Multiple Weighted Power Constraints Under the multiple weighted power constraints, **Opt. 1.3** becomes

$$\mathbf{Opt.\,1.6}: \max_{\boldsymbol{F}} \boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}, \mathrm{s.t.} \mathrm{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right) \leq P_i, 1 \leq i \leq I. \tag{34}$$

Algorithm 1: The Sub-Gradient Algorithm for Finding the Weighting Factors $\alpha_i, \forall i$.

1: Initialization: set iteration index as $I_{\mathrm{ite}} = 0$;
set the maximum iteration number $I_{\max}$;
randomly set initial weighting parameters $\alpha_i^{(0)}$, $\forall\, i = 1, \ldots, I$;
2: **repeat**
3: Solve the problem (34) to obtain $\boldsymbol{F}^{(I_{\mathrm{ite}})}$ given $\alpha_i^{(I_{\mathrm{ite}})}$;
4: Define the step size $t_{I_{\mathrm{ite}}} = \frac{c}{a + I_{\mathrm{ite}} \cdot b}$, $\{a, b, c\} > 0$;
5: Update
$\alpha_i^{(I_{\mathrm{ite}}+1)} = [\alpha_i^{(I_{\mathrm{ite}})} + t_{I_{\mathrm{ite}}}(\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}^{(I_{\mathrm{ite}})}(\boldsymbol{F}^{(I_{\mathrm{ite}})})^H) - P_i)]^+$, $\forall i$;
6: Update $I_{\mathrm{ite}} = I_{\mathrm{ite}} + 1$
7: **until**
$\alpha_i^{(I_{\mathrm{ite}}-1)}(\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}^{(I_{\mathrm{ite}}-1)}(\boldsymbol{F}^{(I_{\mathrm{ite}}-1)})^H) - P_i) \leq \varepsilon_i, \forall i$ or $I_{ite} \leq I_{\max}$, where $\varepsilon_i > 0, \forall i$ is sufficiently small.
8: **return** $\alpha_i, \boldsymbol{F}^{I_{ite}}, \forall\, i = 1, \ldots, I$.

Note that the weighted power constraints are convex [23] and include both the sum power constraint and per-antenna power constraints as its special cases. The optimal solution $\boldsymbol{F}_{\mathrm{opt}}$ for **Opt. 1.6** is given in Lemma 3.

Lemma 3: The Pareto optimal solutions of **Opt. 1.6** satisfy the following structure

$$\boldsymbol{F}_{\mathrm{opt}} = \boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}}, \tag{35}$$

where $\boldsymbol{U}_{\mathrm{Arb}}$ is an arbitrary unitary matrix of appropriate dimension, $\boldsymbol{\Omega} = \sum_{i=1}^{I} \alpha_i\boldsymbol{\Omega}_i$, the nonnegative scalars α_i are the weighting factors that ensure that the constraints $\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) \leq P_i$ hold and they can be computed by classic sub-gradient methods, while the unitary matrix $\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}$ is specified by the EVD

$$\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}} = \boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{\Pi}}} \searrow. \tag{36}$$

Proof: See Appendix B. ∎

Specific Applications Three specific applications are given for each lemma. In wireline communications relying on the ubiquitous digital subscriber lines (DSL), the shaping constraint, i.e., spectral mask constraint, is the most important constraint used for limiting the crosstalk by forcing the users/services to have zero power outside their predefined spectral ranges [35]. In order to impose a maximum transmit power limit in the different transmit directions, the joint power constraint can be used [37]. For per-antenna power constraints, the most representative application example is the beamforming design of C-RAN, where the signals are transmitted from distributed antennas [34].

D. Advantages of Matrix-Monotonic Optimization

Matrix-monotonic optimization theory can simplify the optimization problem relying on matrix variables into a much simpler one manipulating only vector variables. Using matrix-monotonic optimization, for example, the optimal structure of the matrix variable $\boldsymbol{F}$ can be derived and the remaining optimization problem becomes a much simpler one that optimizes the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}}$. For the various objective functions and constraints discussed previously, the optimal solutions of the diagonal elements of the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}}$ are in fact diverse variants of classic water-filling solutions [46], which can be readily obtained based on the corresponding Karush-Kuhn-Tucker (KKT) conditions [45, P244].

In the existing literature, MIMO transceiver optimization problems are unified in the framework based on majorization theory [10]. Our work is different from this existing framework in two perspectives. Firstly, in [10], linear and nonlinear transceiver optimization is considered separately. In our work, they are considered in the same framework. Additionally, in our work, more objective functions are considered. More importantly, the shaping constraint, joint power constraint and multiple weighted power constraints are considered in our work instead of merely the sum power constraint.

For the multiple weighted power constraints, to the best of our knowledge, all the existing works are based on the KKT conditions. There are several limitations for these existing works. Firstly, this method is only applicable to mutual information maximization and MSE minimization. It cannot be used for more general objective functions. The method is not applicable for example to more complex systems, such as multi-hop AF MIMO relaying systems. Moreover, the KKT condition based methods also suffer from serious weaknesses due to the fact that the KKT conditions are only necessary conditions for the optimal solutions. As discussed in [44], the so-called turning-off effect and ambiguity effect usually perturb the KKT conditions based methods when deriving the optimal solutions. To overcome this problem, a widely used method is to consider the covariance matrix as a new variable in order to exploit its hidden convex nature. Unfortunately, the cost of adopting this approach is that the rank constraint has to be relaxed first. By contrast, our matrix-monotonic optimization framework does not suffer from these problems and has much wider applications.

III. Bayes Robust Matrix-Monotonic Optimization

In wireless communication systems, the channel parameters have to be estimated. However, due to the uncertainty introduced both by noise and the time-varying nature of wireless channels, channel estimation errors inevitably exist [17], where the true channel matrix $\boldsymbol{H}$ can be expressed by the following Kronecker formula [18], [20]

$$\boldsymbol{H} = \widehat{\boldsymbol{H}} + \boldsymbol{H}_{\mathrm{W}}\boldsymbol{\Psi}^{\frac{1}{2}}, \tag{37}$$

Here $\widehat{\boldsymbol{H}}$ is the estimated channel matrix and $\boldsymbol{H}_{\mathrm{W}}\boldsymbol{\Psi}^{\frac{1}{2}}$ is the channel estimation error, in which the elements of $\boldsymbol{H}_{\mathrm{W}}$ obey the independent and identical complex Gaussian distribution $\mathcal{CN}(0,1)$ and the covariance matrix $\boldsymbol{\Psi}$ of the channel estimate is a function of both the training sequence and of the channel estimator [18], [20]. It is worth noting that in this section we focus our attention on the robust transceiver design for the scenario, where both the source and destination have imperfect CSI. Based on (37), for Bayes robust transceiver optimization, the matrix $\boldsymbol{\Pi}$ in the matrix-monotonic optimization can be expressed as [26]

$$\boldsymbol{\Pi} = \widehat{\boldsymbol{H}}^{\mathrm{H}}\left(\sigma_{\mathrm{n}}^2\boldsymbol{I} + \mathrm{Tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Psi}\right)\boldsymbol{I}\right)^{-1}\widehat{\boldsymbol{H}}, \tag{38}$$

where σ_{n}^2 is the additive white noise power in the data transmission.

As a result, the generic Bayes robust matrix-variable optimization can be formulated as [26]

$$\begin{aligned}\textbf{Opt. 2.1}:\ &\min_{\boldsymbol{X}}\ f\left(\boldsymbol{X}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{X}\right),\\ &\ \mathrm{s.t.}\ \ \boldsymbol{K}_{\mathrm{n}} = \sigma_{\mathrm{n}}^2\boldsymbol{I} + \mathrm{Tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Psi}\right)\boldsymbol{I},\\ &\qquad\ \psi_i(\boldsymbol{X}) \le 0,\ 1 \le i \le I.\end{aligned} \tag{39}$$

As discussed in [26], after introducing the transformation $\boldsymbol{X} = \boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}$ and recalling that the constraints $\psi_i(\cdot)$ are right unitarily-invariant, **Opt. 2.1** is transferred equivalently to the following matrix-monotonic optimization problem:

$$\begin{aligned}\textbf{Opt. 2.2}:\ &\max_{\boldsymbol{F}}\ \boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F},\\ &\ \mathrm{s.t.}\ \ \boldsymbol{K}_{\mathrm{n}} = \sigma_{\mathrm{n}}^2\boldsymbol{I} + \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\boldsymbol{I},\\ &\qquad\ \psi_i(\boldsymbol{F}) \le 0, 1 \le i \le I.\end{aligned} \tag{40}$$

Here the matrix $\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F}$ can be regarded as an extended SNR matrix in the presence of channel estimation errors, and this kind of matrix-monotonic optimization is termed as robust matrix-monotonic optimization in [26]. In the following, we discuss the optimal solutions of this robust matrix-monotonic optimization problem under specific power constraints.

1) Shaping Constraint: Consider the shaping constraint of

$$\psi_1(\boldsymbol{F}) = \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} - \boldsymbol{R}_{\mathrm{s}}. \tag{41}$$

As proved in Appendix C, for the general case of $\boldsymbol{\Psi} \not\propto \boldsymbol{I}$, a suboptimal solution for **Opt. 2.2** which maximizes a lower bound of the objective of **Opt. 2.2** is given by Lemma 1. When $\boldsymbol{\Psi} = \boldsymbol{0}$, the lower bound is tight and the solution given in Lemma 1 is exactly the Pareto optimal solution of **Opt. 2.2**.

2) Joint Power Constraint: Next consider the joint power constraint specified by

$$\psi_1(\boldsymbol{F}) = \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right) - P,\ \psi_2(\boldsymbol{F}) = \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} - \tau\boldsymbol{I}. \tag{42}$$

For the perfect CSI case associated with $\boldsymbol{\Psi} = \boldsymbol{0}$, the Pareto optimal solutions of **Opt. 2.2** are specified by Lemma 2. When $\boldsymbol{\Psi} \propto \boldsymbol{I}$ and $\psi_1(\boldsymbol{F}) \le 0$ is active at the optimal solutions $\boldsymbol{F}_{\mathrm{opt}}$, the Pareto optimal solutions of **Opt. 2.2** also satisfy the structure given in Lemma 2, since in this case $\boldsymbol{K}_{\mathrm{n}}$ is constant. As proved in Appendix C, for the general case $\boldsymbol{\Psi} \not\propto \boldsymbol{I}$, the suboptimal solution that maximizes a lower bound of the objective of **Opt. 2.2** satisfies the following structure

$$\boldsymbol{F} = \frac{\sigma_{\mathrm{n}}\widetilde{\boldsymbol{\Psi}}^{-\frac{1}{2}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}}}{\left(1 - \mathrm{Tr}\left(\widetilde{\boldsymbol{\Psi}}^{-\frac{1}{2}}\boldsymbol{\Psi}\widetilde{\boldsymbol{\Psi}}^{-\frac{1}{2}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}^{\mathrm{H}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}^{\mathrm{H}}\right)\right)^{\frac{1}{2}}}, \tag{43}$$

where $\widetilde{\boldsymbol{\Psi}} = \sigma_{\mathrm{n}}^2\boldsymbol{I} + P\boldsymbol{\Psi}$ and the unitary matrix $\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}$ is defined based on the following SVD

$$\widehat{\boldsymbol{H}}\left(\sigma_{\mathrm{n}}^2\boldsymbol{I} + P\boldsymbol{\Psi}\right)^{-\frac{1}{2}} = \boldsymbol{U}_{\widetilde{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{H}}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}^{H},\ \text{with},\ \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{H}}} \searrow. \tag{44}$$

The diagonal elements of the rectangular diagonal matrix $\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}_k}$ are smaller than $\sqrt{\tau(\sigma_{\mathrm{n}}^2 + P\lambda_{\min}(\boldsymbol{\Psi}))/(\sigma_{\mathrm{n}}^2 + P\lambda_{\max}(\boldsymbol{\Psi}))}$.

3) Multiple Weighted Power Constraints: When multiple weighted power constraints are used, we have

$$\psi_i(\boldsymbol{F}) = \mathrm{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right) - P_i,\ 1 \le i \le I. \tag{45}$$

From $\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) \le P_i$, it is readily seen that the following inequality holds

$$\begin{aligned}\mathrm{Tr}\left(\left(\sigma_{\mathrm{n}}^2\boldsymbol{\Omega}_i + P_i\boldsymbol{\Psi}\right)\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right) &= \mathrm{Tr}\left(\sigma_{\mathrm{n}}^2\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right)\\ &\quad + P_i\mathrm{Tr}\left(\boldsymbol{\Psi}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right)\\ &\le \sigma_{\mathrm{n}}^2P_i + P_i\mathrm{Tr}\left(\boldsymbol{\Psi}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right).\end{aligned} \tag{46}$$

Hence $\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) \leq P_i$ is equivalent to

$$\frac{\mathrm{Tr}\left[\left(\sigma_{\mathrm{n}}^2\boldsymbol{\Omega}_i + P_i\boldsymbol{\Psi}\right)\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right]}{\sigma_{\mathrm{n}}^2 + \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)} \leq P_i. \tag{47}$$

As a result, the robust matrix-monotonic optimization problem (40) of Bayes is equivalent to the following problem

$$\begin{aligned}\mathbf{Opt.\,2.3}:\ \max_{\boldsymbol{F}}\ & \boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F},\\ \text{s.t. } & \boldsymbol{K}_{\mathrm{n}} = \sigma_{\mathrm{n}}^2\boldsymbol{I} + \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\boldsymbol{I},\\ & \frac{\mathrm{Tr}\left(\left(\sigma_{\mathrm{n}}^2\boldsymbol{\Omega}_i + P_i\boldsymbol{\Psi}\right)\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right)}{\sigma_{\mathrm{n}}^2 + \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)} \leq P_i, 1 \leq i \leq I.\end{aligned} \tag{48}$$

By defining the auxiliary matrix variable

$$\bar{\boldsymbol{F}} = \left[\sigma_{\mathrm{n}}^2 + \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\right]^{-\frac{1}{2}}\boldsymbol{F}, \tag{49}$$

the optimization problem (48) can be simplified to:

$$\begin{aligned}\mathbf{Opt.\,2.4}:\ \max_{\bar{\boldsymbol{F}}}\ & \bar{\boldsymbol{F}}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\widehat{\boldsymbol{H}}\bar{\boldsymbol{F}},\\ \text{s.t.}\quad & \mathrm{Tr}\left(\left(\sigma_{\mathrm{n}}^2\boldsymbol{\Omega}_i + P_i\boldsymbol{\Psi}\right)\bar{\boldsymbol{F}}\bar{\boldsymbol{F}}^{\mathrm{H}}\right) \leq P_i,\ 1 \leq i \leq I.\end{aligned} \tag{50}$$

Similar to the proof of Lemma 3, specifically to (92) in Appendix B, the above optimization problem is equivalent to

$$\mathbf{Opt.\,2.5}:\max_{\bar{\boldsymbol{F}}}\bar{\boldsymbol{F}}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\widehat{\boldsymbol{H}}\bar{\boldsymbol{F}},\ \text{s.t. } \mathrm{Tr}\left(\bar{\boldsymbol{\Omega}}\bar{\boldsymbol{F}}\bar{\boldsymbol{F}}^{\mathrm{H}}\right) \leq \sum_{i=1}^{I} P_i, \tag{51}$$

where

$$\bar{\boldsymbol{\Omega}} = \sum_{i=1}^{I}\alpha_i\left(\sigma_{\mathrm{n}}^2\boldsymbol{\Omega}_i + P_i\boldsymbol{\Psi}\right). \tag{52}$$

According to Lemma 3, the Pareto optimal solutions $\bar{\boldsymbol{F}}_{\mathrm{opt}}$ of **Opt. 2.5** satisfy the following structure

$$\bar{\boldsymbol{F}}_{\mathrm{opt}} = \bar{\boldsymbol{\Omega}}^{-\frac{1}{2}}\boldsymbol{V}_{\boldsymbol{\mathcal{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}}, \tag{53}$$

where the unitary matrix $\boldsymbol{V}_{\boldsymbol{\mathcal{H}}}$ is specified by the SVD of:

$$\widehat{\boldsymbol{H}}\bar{\boldsymbol{\Omega}}^{-\frac{1}{2}} = \boldsymbol{U}_{\boldsymbol{\mathcal{H}}}\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{H}}}\boldsymbol{V}_{\boldsymbol{\mathcal{H}}}^{\mathrm{H}}. \tag{54}$$

From (49), we have $[\sigma_{\mathrm{n}}^2 + \mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi})]^{\frac{1}{2}}\bar{\boldsymbol{F}} = \boldsymbol{F}$ and based on this conclusion we have the following equation

$$\left[\sigma_{\mathrm{n}}^2 + \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\right]\mathrm{Tr}\left(\boldsymbol{\Psi}\bar{\boldsymbol{F}}\bar{\boldsymbol{F}}^{\mathrm{H}}\right) + \sigma_{\mathrm{n}}^2 = \mathrm{Tr}\left(\boldsymbol{\Psi}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right) + \sigma_{\mathrm{n}}^2. \tag{55}$$

This yields

$$\sigma_{\mathrm{n}}^2 + \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right) = \sigma_{\mathrm{n}}^2/\left(1 - \mathrm{Tr}\left(\boldsymbol{\Psi}\bar{\boldsymbol{F}}\bar{\boldsymbol{F}}^{\mathrm{H}}\right)\right). \tag{56}$$

Thus, given the Pareto optimal $\bar{\boldsymbol{F}}_{\mathrm{opt}}$, the Pareto optimal $\boldsymbol{F}_{\mathrm{opt}}$ is expressed as

$$\boldsymbol{F}_{\mathrm{opt}} = \sqrt{\sigma_{\mathrm{n}}^2/[1 - \mathrm{Tr}\left(\boldsymbol{\Psi}\bar{\boldsymbol{F}}_{\mathrm{opt}}\bar{\boldsymbol{F}}_{\mathrm{opt}}^{\mathrm{H}}\right)]}\,\bar{\boldsymbol{F}}_{\mathrm{opt}}. \tag{57}$$

Given (57) and (53), we arrive at the following lemma.

Lemma 4: The Pareto optimal solutions $\boldsymbol{F}_{\mathrm{opt}}$ of $\mathbf{Opt.\,2.2}$ under the multiple weighted power constraints satisfy the following structure

$$\boldsymbol{F}_{\mathrm{opt}} = \frac{\sigma_{\mathrm{n}}\bar{\boldsymbol{\Omega}}^{-\frac{1}{2}}\boldsymbol{V}_{\boldsymbol{\mathcal{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}}}{\left[1 - \mathrm{Tr}\left(\bar{\boldsymbol{\Omega}}^{-\frac{1}{2}}\boldsymbol{\Psi}\bar{\boldsymbol{\Omega}}^{-\frac{1}{2}}\boldsymbol{V}_{\boldsymbol{\mathcal{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}^{\mathrm{H}}\boldsymbol{V}_{\boldsymbol{\mathcal{H}}}^{\mathrm{H}}\right)\right]^{\frac{1}{2}}}. \tag{58}$$

The robust optimal structure under the multiple weighted power constraints given in Lemma 4 is significantly different from the previous robust solutions under designed the sum-power constraints [26] and for the transceiver designs relying on perfect CSI under the per-antenna power constraints [34]. In the traditional robust transceivers designed under the sum power constraint, there is no restriction on the unitary matrix in the eigenvalue decomposition of the covariance matrix $\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}$. For the multiple weighted power constraints, there is a restriction on the unitary matrix in the eigenvalue decomposition. Then a new rotation matrix $\bar{\boldsymbol{\Omega}}^{-\frac{1}{2}}$ is needed, based on which the precoder matrix can align the direction with the space constructed by the multiple weighting matrices.

IV. Stochastically Robust Matrix-Monotonic Optimization

When the CSI knowledge at the receiver (CSIR) is perfect, but only statistical CSI is available, at the transmitter (CSIT) the corresponding stochastically robust matrix-monotonic optimization can be formulated as [15]

$$\begin{aligned}\mathbf{Opt.\,3.1}:\ \min_{\boldsymbol{X}}\ & \mathbb{E}_{\boldsymbol{H}}\left\{f\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{n}}^{-1}\boldsymbol{H}\boldsymbol{X}\right)\right\},\\ \text{s.t. } & \psi_i(\boldsymbol{X}) \leq 0,\ 1 \leq i \leq I,\end{aligned} \tag{59}$$

where $\boldsymbol{R}_{\mathrm{n}}$ is the noise covariance matrix. For simplicity, we mainly consider $\boldsymbol{R}_{\mathrm{n}} = \sigma_{\mathrm{n}}^2\boldsymbol{I}$ as in Section III. For this kind of optimization problems, the objective function is an average value over the distribution of the channel matrix $\boldsymbol{H}$ modeled by

$$\boldsymbol{H} = \widehat{\boldsymbol{H}} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\mathrm{W}}\boldsymbol{\Psi}^{\frac{1}{2}}, \tag{60}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ are the row and column correlation matrices, respectively. For MIMO systems, $\boldsymbol{\Sigma}$ is the spatial correlation matrix of the receiver antenna array, while $\boldsymbol{\Psi}$ is the spatial correlation matrix of the transmitter antenna array. Since the constraints are right unitarily-invariant, **Opt. 3.1** can be expressed as

$$\begin{aligned}\mathbf{Opt.\,3.2}:\ \min_{\boldsymbol{F}}\ & \mathbb{E}_{\boldsymbol{H}}\left\{f\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{n}}^{-1}\boldsymbol{H}\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)\right\},\\ \text{s.t. } & \psi_i(\boldsymbol{F}) \leq 0,\ 1 \leq i \leq I.\end{aligned} \tag{61}$$

The stochastically robust matrix-monotonic optimization naturally aims at optimizing the distribution of the random matrix $\boldsymbol{H}\boldsymbol{F}$, based on the channel model (60). Therefore, **Opt. 3.2** can be rewritten as **Opt. 3.3** of (63) shown at the bottom of the next page, where $p(\boldsymbol{H}_{\mathrm{W}})$ is the probability density function (PDF) of $\boldsymbol{H}_{\mathrm{W}}$. As pointed out in [17], the analytical expression of the average value of an arbitrary objective function $f(\cdot)$ in **Opt. 3.3** is impossible to obtain, which thus makes **Opt. 3.3** difficult to address. An alternative scheme is to consider the average matrix in the objective function $f(\cdot)$ and the corresponding optimization problem is **Opt. 3.4** given in (64) shown at the bottom of the next page. Taking

$$\boldsymbol{\Pi} = \mathbb{E}\left\{\left(\widehat{\boldsymbol{H}} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\mathrm{W}}\boldsymbol{\Psi}^{\frac{1}{2}}\right)^{\mathrm{H}}\boldsymbol{R}_{\mathrm{n}}^{-1}\left(\widehat{\boldsymbol{H}} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\mathrm{W}}\boldsymbol{\Psi}^{\frac{1}{2}}\right)\right\} \tag{62}$$

the optimal solutions of $\boldsymbol{Q}_{\boldsymbol{X}}$ are given in Table II. Based on the optimal solutions of $\boldsymbol{Q}_{\boldsymbol{X}}$, similar to **Opt. 1.2**, the optimal solutions $\boldsymbol{F}$ of **Opt. 3.4** fall in the Pareto optimal solution set of **Opt. 3.5** in (65) shown at the bottom of the next page. It is obvious that the key to the optimization of **Opt. 3.5** is to maximize the eigenvalues of $\mathbb{E}[\boldsymbol{F}^{\mathrm{H}}(\boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{H}_{\mathrm{W}}^{\mathrm{H}}\boldsymbol{\Sigma}^{\frac{1}{2}} + \widehat{\boldsymbol{H}}^{\mathrm{H}})\boldsymbol{R}_{\mathrm{n}}^{-1}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\mathrm{W}}\boldsymbol{\Psi}^{\frac{1}{2}} + \widehat{\boldsymbol{H}})\boldsymbol{F}] =$

$\boldsymbol{F}^{\rm H}(\widehat{\boldsymbol{H}}^{\rm H}\boldsymbol{R}_{\rm n}^{-1}\widehat{\boldsymbol{H}}+{\rm Tr}(\boldsymbol{\Sigma}\boldsymbol{R}_{\rm n}^{-1})\boldsymbol{\Psi})\boldsymbol{F}$. Hence, this stochastically robust optimization problem **Opt. 3.5** is equivalent to

$$\begin{aligned}\textbf{Opt. 3.6}: &\max_{\boldsymbol{F}} \ \lambda\left(\boldsymbol{F}^{\rm H}(\widehat{\boldsymbol{H}}^{\rm H}\boldsymbol{R}_{\rm n}^{-1}\widehat{\boldsymbol{H}}+{\rm Tr}(\boldsymbol{\Sigma}\boldsymbol{R}_{\rm n}^{-1})\boldsymbol{\Psi})\boldsymbol{F}\right),\\ &{\rm s.t.} \ \ \psi_i(\boldsymbol{F})\leq 0, 1\leq i\leq I.\end{aligned} \tag{66}$$

As discussed previously in Section II-C, the above multi-objective optimization problem is equivalent to the following matrix-monotonic optimization problem

$$\begin{aligned}\textbf{Opt. 3.7}: &\max_{\boldsymbol{F}} \ \boldsymbol{F}^{\rm H}(\widehat{\boldsymbol{H}}^{\rm H}\boldsymbol{R}_{\rm n}^{-1}\widehat{\boldsymbol{H}}+{\rm Tr}(\boldsymbol{\Sigma}\boldsymbol{R}_{\rm n}^{-1})\boldsymbol{\Psi})\boldsymbol{F},\\ &{\rm s.t.} \ \ \psi_i(\boldsymbol{F})\leq 0, 1\leq i\leq I.\end{aligned} \tag{67}$$

Again, we discuss the Pareto optimal solutions of this stochastically robust matrix-monotonic optimization problem under three specific power constraints, respectively.

1) Shaping Constraint: Under the shaping constraint of $\psi(\boldsymbol{F})=\boldsymbol{F}\boldsymbol{F}^{\rm H}-\boldsymbol{R}_{\rm s}$, the optimal solution $\boldsymbol{F}_{\rm opt}$ to **Opt. 3.7** is specified by Lemma 1. Specifically, when the rank of $\boldsymbol{R}_{\rm s}$ is not higher than the number of columns and the number of rows in $\boldsymbol{F}$, $\boldsymbol{F}_{\rm opt}$ is a square root of $\boldsymbol{R}_{\rm s}$.

2) Joint Power Constraint: Clearly, under the joint power constraint (42), **Opt. 3.7** is identical to **Opt. 1.5** with

$$\boldsymbol{\Pi}=\widehat{\boldsymbol{H}}^{\rm H}\boldsymbol{R}_{\rm n}^{-1}\widehat{\boldsymbol{H}}+{\rm Tr}(\boldsymbol{\Sigma}\boldsymbol{R}_{\rm n}^{-1})\boldsymbol{\Psi}. \tag{68}$$

Therefore, the Pareto optimal solutions $\boldsymbol{F}_{\rm opt}$ of **Opt. 3.7** under the joint power constraint are defined exactly in Lemma 2 by simply replacing $\boldsymbol{\Pi}$ in (32) and (33) with (68).

3) Multiple Weighted Power Constraints: Obviously, under the multiple weighted power constraints (45), the Pareto optimal solutions $\boldsymbol{F}_{\rm opt}$ of **Opt. 3.7** are specified by Lemma 3, where $\boldsymbol{\Pi}$ should be replaced by (68).

V. Worst Case Robust Matrix-Monotonic Optimization

In this section, we consider the norm-bounded CSI error, i.e., $\|\Delta\boldsymbol{H}\|_2\leq\gamma$ with $\|\cdot\|_2$ denoting the matrix Spectral norm, and adopt the worst-case (min-max) criterion as a figure-of-merit for robust designs [19]. Hereafter, the spectral norm is adopted since it can act as both lower and upper bounds of the widely adopted Frobenius and Nuclear norms and is generally tractable [19], [47]. For example, we have $\|\cdot\|_2\leq\|\cdot\|_F\leq\sqrt{{\rm rank}(\cdot)}\|\cdot\|_2$, implying that the spectral norm constrained error can also provide valuable insights for Frobenius norm constrained case. Moreover, for the same error size, the spectral norm generally covers the largest error region.

Let us denote the estimated channel matrix and channel error matrix by $\widehat{\boldsymbol{H}}$ and $\Delta\boldsymbol{H}$. Given the norm-bounded CSI error $\|\Delta\boldsymbol{H}\|_2\leq\gamma$, substituting $\boldsymbol{H}=\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}$ and $\boldsymbol{R}_{\rm n}=\sigma_{\rm n}^2\boldsymbol{I}$ into **Opt. 1.3**, the robust matrix-monotonic optimization problem under Spectral norm bounded CSI error can be formulated as

$$\begin{aligned}\textbf{Opt. 4.1}: &\max_{\boldsymbol{F}}\min_{\Delta\boldsymbol{H}}\sigma_{\rm n}^{-2}\boldsymbol{F}^{\rm H}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}\right)^{\rm H}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}\right)\boldsymbol{F},\\ &{\rm s.t.} \ \ \psi_i(\boldsymbol{F})\leq 0, 1\leq i\leq I, \|\Delta\boldsymbol{H}\|_2\leq\gamma.\end{aligned} \tag{69}$$

The Spectral norm is unitarily-invariant, which means that for the arbitrary $\|\boldsymbol{E}\|_2\leq\epsilon_s$, it yields $\|\boldsymbol{U}\boldsymbol{E}\boldsymbol{V}\|_2\leq\epsilon_s$ given any unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V}$. Based on the following matrix inequality [40, P471]

$$(\sigma_i(\boldsymbol{B})-\sigma_1(\boldsymbol{C}))^+\leq\sigma_i(\boldsymbol{B}+\boldsymbol{C}) \tag{70}$$

we readily conclude that

$$\sigma_i(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H})\geq\left(\sigma_i(\widehat{\boldsymbol{H}})-\sigma_1(\Delta\boldsymbol{H})\right)^+. \tag{71}$$

Therefore, we have the following eigenvalue inequality

$$\begin{aligned}&\lambda\left(\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}\right)^{\rm H}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}\right)\right)\\ &\succeq\lambda\left(\left(\widehat{\boldsymbol{H}}-\boldsymbol{U}_{\widehat{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\Delta\boldsymbol{H}}\boldsymbol{V}_{\widehat{\boldsymbol{H}}}^{H}\right)^{\rm H}\left(\widehat{\boldsymbol{H}}-\boldsymbol{U}_{\widehat{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\Delta\boldsymbol{H}}\boldsymbol{V}_{\widehat{\boldsymbol{H}}}^{H}\right)\right)\end{aligned} \tag{72}$$

where $\boldsymbol{U}_{\widehat{\boldsymbol{H}}}$ and $\boldsymbol{V}_{\widehat{\boldsymbol{H}}}$ are derived from the following SVD

$$\widehat{\boldsymbol{H}}=\boldsymbol{U}_{\widehat{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widehat{\boldsymbol{H}}}\boldsymbol{V}_{\widehat{\boldsymbol{H}}}^{H}, \text{with}, \boldsymbol{\Lambda}_{\widehat{\boldsymbol{H}}}\searrow. \tag{73}$$

The diagonal matrix $\boldsymbol{\Lambda}_{\Delta\boldsymbol{H}}$ equals

$$[\boldsymbol{\Lambda}_{\Delta\boldsymbol{H}}]_{i,i}=\min\left([\boldsymbol{\Lambda}_{\widehat{\boldsymbol{H}}}]_{i,i},\gamma\right),\forall i. \tag{74}$$

Then there exists a unitary matrix $\boldsymbol{Q}$ for which the following matrix inequality hold

$$\begin{aligned}&\boldsymbol{Q}^{\rm H}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}\right)^{\rm H}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}\right)\boldsymbol{Q}\\ &\succeq\left(\widehat{\boldsymbol{H}}-\boldsymbol{U}_{\widehat{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\Delta\boldsymbol{H}}\boldsymbol{V}_{\widehat{\boldsymbol{H}}}^{H}\right)^{\rm H}\left(\widehat{\boldsymbol{H}}-\boldsymbol{U}_{\widehat{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\Delta\boldsymbol{H}}\boldsymbol{V}_{\widehat{\boldsymbol{H}}}^{H}\right).\end{aligned} \tag{75}$$

Based on (75), when the constraint functions $\psi_i(\boldsymbol{F})$'s in **Opt. 4.1** are left unitarily invariant, the worst-case $\Delta\boldsymbol{H}$ for **Opt. 4.1** is

$$\Delta\boldsymbol{H}_{\rm worst}=\boldsymbol{U}_{\widehat{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\Delta\boldsymbol{H}}\boldsymbol{V}_{\widehat{\boldsymbol{H}}}^{H}. \tag{76}$$

That is because when the $\psi_i(\boldsymbol{F})$s in **Opt. 4.1** are left unitarily invariant, such as **Constraint 6:** the joint power constraint or **Constraint 5:** the constraints on the eigenvalues of $\boldsymbol{F}\boldsymbol{F}^{\rm H}$, for any feasible $\boldsymbol{F}$ and an arbitrary unitary matrix $\boldsymbol{Q}$, $\boldsymbol{Q}\boldsymbol{F}$ is also feasible. Thus it is always possible to find $\boldsymbol{Q}$ the matrix inequality in (75) holds. Furthermore, based on the worst-case $\Delta\boldsymbol{H}_{\rm worst}$

$$\textbf{Opt. 3.3}: \min_{\boldsymbol{F},\boldsymbol{Q}_{\boldsymbol{X}}}\int f\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\rm H}\boldsymbol{F}^{\rm H}(\widehat{\boldsymbol{H}}+\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\rm W}\boldsymbol{\Psi}^{\frac{1}{2}})^{\rm H}\boldsymbol{R}_{\rm n}^{-1}(\widehat{\boldsymbol{H}}+\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\rm W}\boldsymbol{\Psi}^{\frac{1}{2}})\boldsymbol{F}\boldsymbol{Q}_{\boldsymbol{X}}\right)p\left(\boldsymbol{H}_{\rm W}\right){\rm d}\boldsymbol{H}_{\rm W}, \ {\rm s.t.}\ \psi_i\left(\boldsymbol{F}\right)\leq 0, 1\leq i\leq I, \tag{63}$$

$$\textbf{Opt. 3.4}: \min_{\boldsymbol{F},\boldsymbol{Q}_{\boldsymbol{X}}} f\left(\boldsymbol{Q}_{\boldsymbol{X}}^{\rm H}\int\boldsymbol{F}^{\rm H}\left(\widehat{\boldsymbol{H}}+\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\rm W}\boldsymbol{\Psi}^{\frac{1}{2}}\right)^{\rm H}\boldsymbol{R}_{\rm n}^{-1}\left(\widehat{\boldsymbol{H}}+\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\rm W}\boldsymbol{\Psi}^{\frac{1}{2}}\right)\boldsymbol{F}p\left(\boldsymbol{H}_{\rm W}\right){\rm d}\boldsymbol{H}_{\rm W}\boldsymbol{Q}_{\boldsymbol{X}}\right), \ {\rm s.t.}\ \psi_i\left(\boldsymbol{F}\right)\leq 0, 1\leq i\leq I. \tag{64}$$

$$\textbf{Opt. 3.5}: \min_{\boldsymbol{F}}\lambda\left(\mathbb{E}\{\boldsymbol{F}^{\rm H}(\boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{H}_{\rm W}^{\rm H}\boldsymbol{\Sigma}^{\frac{1}{2}}+\widehat{\boldsymbol{H}}^{\rm H})\boldsymbol{R}_{\rm n}^{-1}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{H}_{\rm W}\boldsymbol{\Psi}^{\frac{1}{2}}+\widehat{\boldsymbol{H}})\boldsymbol{F}\}\right), \ {\rm s.t.}\ \psi_i\left(\boldsymbol{F}\right)\leq 0, 1\leq i\leq I. \tag{65}$$

in (76), **Opt. 4.1** is rewritten as

$$\begin{aligned}\mathbf{Opt.\,4.2}:\ \max_{\boldsymbol{F}}\ & \sigma_{\mathrm{n}}^{-2}\boldsymbol{F}^{\mathrm{H}}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}_{\mathrm{worst}}\right)^{\mathrm{H}} \\ & \times\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}_{\mathrm{worst}}\right)\boldsymbol{F}, \\ \text{s.t.}\ & \psi_i(\boldsymbol{F})\le 0, 1\le i\le I,\end{aligned} \tag{77}$$

We would like to highlight that all the constraints $\psi_i(\boldsymbol{F})$ in **Opt. 4.1** are right unitarily invariant. The objective function in **Opt. 4.2** is an upper bound of the worst case of the objective function of **Opt. 4.1** and this bound is tight when $\psi_i(\boldsymbol{F})$s are also left unitarily invariant. Specifically, when $\psi_i(\boldsymbol{F})$s in **Opt. 4.1** are right unitarily invariant, the term $(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}_{\mathrm{worst}})^{\mathrm{H}}(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}_{\mathrm{worst}})$ is the worst case of $(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H})^{\mathrm{H}}(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H})$ only when $\Delta\boldsymbol{H}$ is restricted to have the same SVD unitary matrices as $\widehat{\boldsymbol{H}}$.

1) Shaping Constraint: For the shaping constraint (41), the optimal solution $\boldsymbol{F}_{\mathrm{opt}}$ of **Opt. 4.2** is also specified by Lemma 1. That is, when the rank of $\boldsymbol{R}_{\mathrm{s}}$ is not larger than the number of columns and the number of rows in $\boldsymbol{F}$, $\boldsymbol{F}_{\mathrm{opt}}$ of **Opt. 4.2** is a square root of $\boldsymbol{R}_{\mathrm{s}}$.

2) Joint Power Constraint: Under the joint power constraint (42), **Opt. 4.2** is identical to **Opt. 1.5** with

$$\boldsymbol{\Pi}=\sigma_{\mathrm{n}}^{-2}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}_{\mathrm{worst}}\right)^{\mathrm{H}}\left(\widehat{\boldsymbol{H}}-\Delta\boldsymbol{H}_{\mathrm{worst}}\right). \tag{78}$$

As a result, the Pareto optimal solutions $\boldsymbol{F}_{\mathrm{opt}}$ of **Opt. 4.2** under the joint power constraint are defined exactly in Lemma 2 by simply replacing $\boldsymbol{\Pi}$ in (32) and (33) with (78).

3) Multiple Weighted Power Constraints: Given the multiple weighted power constraints of (45), the Pareto optimal solutions of **Opt. 4.2** are specified by Lemma 3, where $\boldsymbol{\Pi}$ should be replaced by (78).

VI. Simulation Results and Discussions

In this section, we take the MSE criterion (**Obj. 2** in Table I) of MIMO systems as a central figure-of-merit to demonstrate the proposed robust designs in Section III with statistically imperfect CSIT (ICSIT) and CSIR (ICSIR), Section IV with statistically imperfect CSIT (ICSIT) and perfect CSIR (PCSIR), and Section V with deterministically imperfect CSIT and CSIR. Specifically, in Section III and Section IV, the sum average MSE is studied to illustrate the influence of imperfect CSIT and/or CSIR on average symbol detection performance. Finally, in Section V, the worst-case MSE is adopted to guarantee the symbol detection performance for all the channel realizations Notice that the proposed robust designs in above Sections all have analytical solutions, and can be reduced to simple power allocation problems with water-filling solutions.

Unless otherwise stated, numerical results are presented for the point-to-point MIMO scenario with the transmitter and receiver equipped with $N_t=4$ and $N_r=4$ antennas, respectively. Moreover, the number of data streams is $L=2$. According to (37) adopted in Section III and Section IV, we assume that the imperfect CSI consists of the estimated term $\widehat{\boldsymbol{H}}$ distributed as $\mathcal{CN}(\boldsymbol{0},(1-\sigma_e^2)\boldsymbol{I}_{N_r}\otimes\boldsymbol{I}_{N_t})$ and the error term $\boldsymbol{H}_W\boldsymbol{\Psi}^{\frac{1}{2}}$, where $\boldsymbol{\Psi}$ is defined by the exponential model, i.e., $[\boldsymbol{\Psi}]_{m,n}=\sigma_e^2p_t^{|m-n|}$ with $p_t=0.5$ and $\sigma_e^2=0.1$, to realize the normalized channel $\mathbb{E}\{[\boldsymbol{H}]_{m,n}[\boldsymbol{H}]_{m,n}^*\}=1,\forall m,n$. Finally, for the worst-case optimization in Section V, the relative error threshold subject to Spectral norm is set as $\gamma=s\|\widehat{\boldsymbol{H}}\|_2$ with $s\in[0,1]$. In addition, for both Section III and Section IV, since the unknown weighting

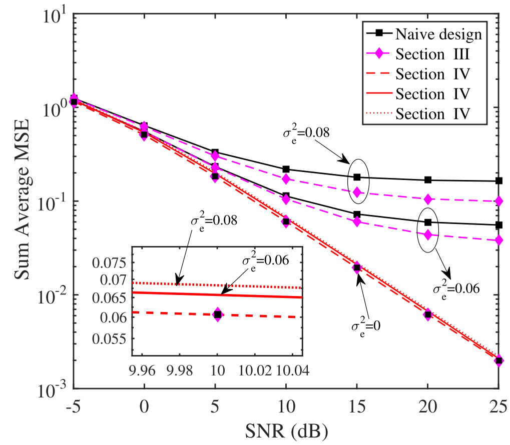


Fig. 1. Sum average MSE versus the estimated channel error σ_e^2 for all studied designs in Section III and Section IV.

factors need to be determined via the sub-gradient method, the per-antenna power constraints, $\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^H)\le P_i,\forall i=1,\cdots N_t$, where $\boldsymbol{\Omega}_i=\mathrm{diag}[\boldsymbol{0}_{1\times i-1},1,\boldsymbol{0}_{1\times N_t-i}],\forall i$ and $P_i=P_t,\forall i$, are mainly studied in the simulations as a special case of multiple weighted power constraints. Finally, for Section V, the joint power constraints $\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^H)\le(L-1)P_t$ and $\boldsymbol{F}\boldsymbol{F}^H\preceq P_t\boldsymbol{I}_{N_t}$ are investigated due to their tractability. Particularly, the per-antenna power limitation $\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^H)\le P_t,\forall i$ can be readily inferred from the shaping power constraint $\boldsymbol{F}\boldsymbol{F}^H\preceq P_t\boldsymbol{I}_{N_t}$. We also define the SNR as $\frac{P_t}{\sigma_{\mathrm{n}}^2}$, where $P_t=1$W and the noise power σ_{n}^2 is varied.

For a comprehensive comparison, we also consider three baselines for the MIMO scenario as follows: For Section III and Section IV, the naive design that simply regards $\widehat{\boldsymbol{H}}$ as a perfect channel estimate of the instantaneous channel $\boldsymbol{H}$ is studied. The optimal solution is derived by solving the problem (51)/(67) with $\boldsymbol{\Psi}=\boldsymbol{0}$, and the sum average MSE is obtained through Monte Carlo experiments. By contrast, for Section V, the nonrobust design is studied by firstly considering $\Delta\boldsymbol{H}=\boldsymbol{0}$ in the problem (69), and then the optimal solution obtained is substituted into the inner minimization of the problem (69) to find the worst-case MSE. Moreover, the ideal case assuming both PCSIR and perfect CSIT (PCSIT) is also considered for all above Sections.

Fig. 1 shows the sum average MSE of all designs studied in Section III and Section IV as the function of the channel error σ_e^2. Clearly, it is observed that when σ_e^2 decreases, the sum average MSE performances of all studied designs improve. Also, the performance gap between the native design and robust design in Section III with ICSIT and ICSIR becomes narrowed. In particular, as σ_e^2, is increased the performance loss of the robust design in Section IV with PCSIR is reduced compared to that in Section III with ICSIR, which further indicates that PCSIR is crucial to realize acceptable average MSE performance.

Fig. 2 shows the worst-case MSE of the proposed robust design in Section V and the other baselines as the function of SNR. Naturally, the ideal design achieves the best worst-case MSE performance, and the proposed robust design is the next. The nonrobust design has the worst performance since the robustness against channel error is not considered. Similarly to the robust design in Section IV with ICSIT and PCSIR, we also find that the slopes of all worst-case designs are nearly identical, and the corresponding worst-case MSE performance is similar,

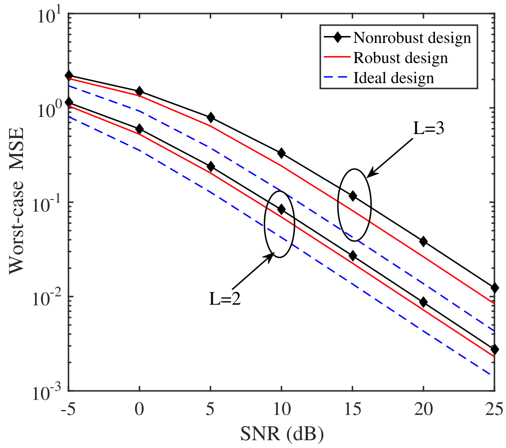


Fig. 2. Worst-case MSE versus SNR for the proposed robust design in Section V and all other baselines.

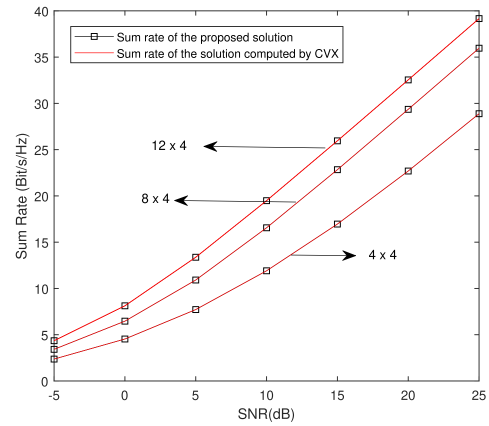


Fig. 4. The performance comparisons between the proposed solution and the solution computed by CVX in terms of sum rate under multiple weighted power constraints.

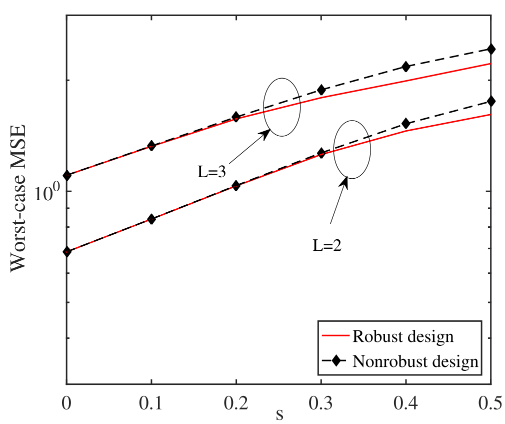


Fig. 3. Worst-case MSE versus the relative error threshold s for the proposed robust and nonrobust designs in Section V.

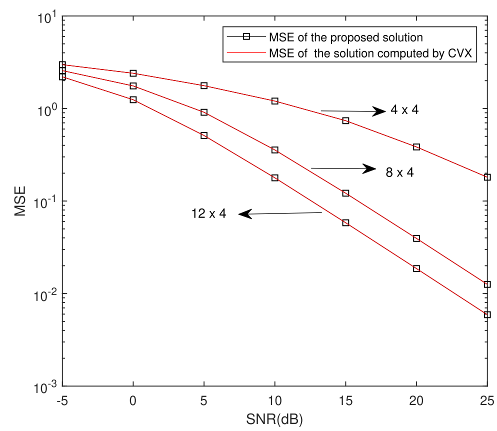


Fig. 5. The performance comparisons between the proposed solution and the solution computed by CVX in terms of sum MSE under multiple weighted power constraints.

especially at high SNRs, because in this context the high transmit power weakens the influence of deterministic channel errors on the achievable MSE. The above conclusions can also be drawn when increasing number of date streams to $L = 3$ is considered. In this context, the performance gap among all studied designs is further enlarged.

Fig. 3 shows the worst-case MSE of the proposed robust design in Section V and the other baselines as the function of the relative error threshold s. As expected, the robust design outperforms the non-robust one, and the performance gain becomes more evident with the increase of both the number of date streams to $L = 3$ and error threshold s.

In order to assess the performance of the proposed solutions under general multiple weighted power constraints, without loss of generality we first build an exponential correlation matrix $\boldsymbol{\Omega}$ with $[\boldsymbol{\Omega}]_{i,j} = 0.3^{|i-j|}$. Based on $\boldsymbol{\Omega}$, a pair of weighting matrices i.e., $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are constructed. Specifically, $\boldsymbol{\Omega}_1$ corresponds to the first two eigenchannels and $\boldsymbol{\Omega}_2$ corresponds to the last two eigenchannels. In addition, the power ratio between the two constraints is 0.6 and 0.4. Moreover, in the simulations the numbers of antennas and data streams are equal to each other. Then both the MSE minimization and sum-rate maximization are convex, hence the problem can be solved by using CVX [48]. It can be concluded from Fig. 4 and Fig. 5 that the proposed solutions have the same performance as that computed by CVX for all the settings investigated. We would like to point out that the numerical algorithms based on CVX have no tangible physical meanings and suffer from high computational complexity as well as from limited scalability.

VII. CONCLUSION

In this paper, a comprehensive framework has been given for matrix-monotonic optimization under various power constraints, including shaping constraints, joint power constraints and multiple weighted power constraints. Matrix-monotonic optimization problems of three different CSI scenarios have been investigated in depth, which are: 1) both transmitter and receiver have imperfect CSI; 2) perfect CSI is available at the receiver but the transmitter has only channel statistics; and 3) perfect CSI is available at the receiver, but the channel estimation error at the transmitter is norm-bounded. In all three cases, the matrix-monotonic optimization framework has been used to derive closed-form structures of the optimal matrix variables, which significantly simplifies the associated optimization problems and reveals a range of underlying physical insights.

Appendix A
Proof of Lemma 2

The Pareto optimal solution set of (31) falls in the optimal solution set of the following optimization problem for all the possible $\boldsymbol{F}_{\text{in}}$ that are in the region of $\text{Tr}(\boldsymbol{F}\boldsymbol{F}^{\text{H}}) \leq P$ and $\boldsymbol{F}\boldsymbol{F}^{\text{H}} \preceq \tau \boldsymbol{I}$:

$$\begin{aligned} \max_{\boldsymbol{F},\alpha} \quad & \alpha \\ \text{s.t.} \quad & \boldsymbol{F}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F} = \alpha \boldsymbol{F}_{\text{in}}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{in}} \\ & \text{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\text{H}}\right) \leq P, \boldsymbol{F}\boldsymbol{F}^{\text{H}} \preceq \tau \boldsymbol{I}. \end{aligned} \tag{79}$$

This conclusion is obvious because for any Pareto optimal solution of (31) $\boldsymbol{F}_{\text{opt}}$, for an $\alpha < 1$, $\boldsymbol{F}_{\text{in}} = \alpha \boldsymbol{F}_{\text{opt}}$ obviously satisfies $\text{Tr}(\boldsymbol{F}_{\text{in}}\boldsymbol{F}_{\text{in}}^{\text{H}}) \leq P$ and $\boldsymbol{F}_{\text{in}}\boldsymbol{F}_{\text{in}}^{\text{H}} \preceq \tau \boldsymbol{I}$. In the following, we will prove that the optimal solutions of (79) own the same structure.

Based on the matrix equality properties that when $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same dimensionality $\boldsymbol{A}^{\text{H}}\boldsymbol{A} = \boldsymbol{B}^{\text{H}}\boldsymbol{B}$ is equivalent to $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{A}$ with $\boldsymbol{U}$ being an unitary matrix [40, P406], the first constraint of (79) is equivalent to

$$\boldsymbol{\Pi}^{1/2}\boldsymbol{F} = \sqrt{\alpha}\boldsymbol{U}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}_{\text{in}} \tag{80}$$

based on which and defining the pseudo inverse of $\boldsymbol{\Pi}^{1/2}$ as $(\boldsymbol{\Pi}^{1/2})^{\dagger}$, we have

$$(\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{\Pi}^{1/2}\boldsymbol{F} = \sqrt{\alpha}(\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}_{\text{in}}. \tag{81}$$

Then we have:

$$\alpha = \frac{\text{Tr}\left[\left((\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}\right)^{\text{H}}(\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}\right]}{\text{Tr}\left[\left((\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}_{\text{in}}\right)^{\text{H}}(\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}_{\text{in}}\right]}. \tag{82}$$

Based on **Matrix Inequality 1**, the numerator of the righthand side of the above equation satisfies

$$\text{Tr}\left[((\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{\Pi}^{1/2}\boldsymbol{F})^{\text{H}}(\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}\right] \leq \sum_j \lambda_j(\boldsymbol{F}\boldsymbol{F}^{\text{H}}) \tag{83}$$

and the denominator satisfies

$$\begin{aligned} & \text{Tr}\left[\left((\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}_{\text{in}}\right)^{\text{H}}(\boldsymbol{\Pi}^{1/2})^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}^{1/2}\boldsymbol{F}_{\text{in}}\right] \\ & \geq \sum_j \frac{\lambda_j\left(\boldsymbol{\Pi}^{1/2}\boldsymbol{F}_{\text{in}}\boldsymbol{F}_{\text{in}}^{\text{H}}\boldsymbol{\Pi}^{1/2}\right)}{\lambda_j(\boldsymbol{\Pi})}. \end{aligned} \tag{84}$$

Based on (83) and (84), α is maximized when $\boldsymbol{F}$ satisfies the following structure

$$\boldsymbol{F} = \boldsymbol{U}_{\boldsymbol{\Pi}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{V}_{\text{in}}^{\text{H}}, \tag{85}$$

where the unitary matrices $\boldsymbol{U}_{\boldsymbol{\Pi}}$ and $\boldsymbol{V}_{\text{in}}$ are defined based on the following EVDs

$$\begin{aligned} \boldsymbol{\Pi} &= \boldsymbol{U}_{\boldsymbol{\Pi}}\boldsymbol{\Lambda}_{\boldsymbol{\Pi}}\boldsymbol{U}_{\boldsymbol{\Pi}}^{\text{H}} \text{ with } \boldsymbol{\Lambda}_{\boldsymbol{\Pi}} \searrow, \\ \boldsymbol{F}_{\text{in}}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{in}} &= \boldsymbol{V}_{\text{in}}\boldsymbol{\Lambda}_{\text{in}}\boldsymbol{V}_{\text{in}}^{\text{H}} \text{ with } \boldsymbol{\Lambda}_{\text{in}} \searrow. \end{aligned} \tag{86}$$

It is worth noting that the final two constraints in the optimization problem (79) only constrain the eigenvalues of $\boldsymbol{F}\boldsymbol{F}^{\text{H}}$. In other words, the final two constraints in (79) only constrain the singular values of $\boldsymbol{F}$. Moreover, it is obvious that the final two constraints in (79) are both right unitarily invariant and left unitarily invariant. The derivations in (83) and (84) are independent of the singular values of of $\boldsymbol{F}$. This means that for any given $\boldsymbol{\Lambda}_{\boldsymbol{F}}$ in (85) the optimal $\boldsymbol{F}$ maximizing α satisfies the structure in (85) without violating the final two constraints in (79). Therefore, it is concluded that the optimal solutions of (79) satisfy the structure in (85) and thus the Pareto optimal solutions of (50) satisfy the structure given by (85). Furthermore, substituting (85) into the objective function of (50), it can be seen that for the Pareto optimal solutions, the value of the unitary matrix $\boldsymbol{V}_{\text{in}}$ in (85) does not affect the optimality of the Paremto optimal solutions. Finally the Pareto optimal solutions of (50) satisfy the following structure

$$\boldsymbol{F}_{\text{opt}} = \boldsymbol{U}_{\boldsymbol{\Pi}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\text{Arb}}^{\text{H}}. \tag{87}$$

Appendix B
Proof of Lemma 3

Any Pareto optimal solution of **Opt. 1.6**, $\boldsymbol{F}_{\text{Pareto}}$, is also a Pareto optimal solution of the following multi-objective optimization problem

$$\min_{\boldsymbol{F}} \left\{\text{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\text{H}}\right)\right\}_{i=1}^{I}, \text{ s.t. } \boldsymbol{F}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F} = \boldsymbol{F}_{\text{Pareto}}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{Pareto}}. \tag{88}$$

This transformation is built on the proof by contradiction. Otherwise we can find a solution better than $\boldsymbol{F}_{\text{Pareto}}$, and this contradicts to the fact that $\boldsymbol{F}_{\text{Pareto}}$ is Pareto optimal.

Since the constraint of (88) is equivalent to $\boldsymbol{U}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F} = \boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_{\text{Pareto}}$, where $\boldsymbol{U}$ is a suitable unitary matrix [40, P406], the optimization problem (88) is equivalent to

$$\min_{\boldsymbol{F}} \left\{\text{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\text{H}}\right)\right\}_{i=1}^{I}, \text{ s.t. } \boldsymbol{U}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F} = \boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_{\text{Pareto}}. \tag{89}$$

In (89), the objective functions are quadratic functions and the constraint is a linear function with respect to $\boldsymbol{F}$, which means that the multi-objective optimization problem (89) is convex [45, P135][3]. Therefore, for any Pareto optimal solution of (89), there exist weights $\alpha_i, 1 \leq i \leq I$, for ensuring that the Pareto optimal solution can be computed via solving the following weighted sum optimization problem [45, P179]

$$\min_{\boldsymbol{F}} \sum_{i=1}^{I} \alpha_i \text{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\text{H}}\right), \text{ s.t. } \boldsymbol{U}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F} = \boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_{\text{Pareto}}. \tag{90}$$

The above conclusion for computing Pareto optimal solution of (89) using weights $\alpha_i, 1 \leq i \leq I$ in (90) are feasible to any unitary matrix $\boldsymbol{U}$. Meanwhile, it is worth noting that $\boldsymbol{F}_{\text{Pareto}}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{Pareto}}$ is equivalent to $\boldsymbol{U}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F} = \boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_{\text{Pareto}}$ where $\boldsymbol{U}$ is a suitable unitary matrix [40, P406]. Thus the whole Pareto optimal solution set of (88) can be achieved via solving the following optimization problem by changing the weights $\alpha_i, 1 \leq i \leq I$,

$$\min_{\boldsymbol{F}} \sum_{i=1}^{I} \alpha_i \text{Tr}\left(\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\text{H}}\right), \text{s.t. } \boldsymbol{F}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F} = \boldsymbol{F}_{\text{Pareto}}^{\text{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{Pareto}}. \tag{91}$$

[3]As $\boldsymbol{\Omega}_k$'s are positive semidefinite, this conclusion can be proved based on the definition of convex function give in [45, P95].

For the optimal solution of (91), $\sum_{i=1}^{I}\alpha_i \mathrm{Tr}(\boldsymbol{\Omega}_i \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}})=P$. Then $\boldsymbol{F}_{\text{Pareto}}$ is a Pareto optimal solution of the following optimization problem

$$\max_{\boldsymbol{F}} \boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}, \ \text{ s.t. } \mathrm{Tr}\left(\sum_{i=1}^{I}\alpha_i\boldsymbol{\Omega}_i\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right)\leq P. \quad (92)$$

This is concluded based on the proof of contradiction.

If $\boldsymbol{F}_{\text{Pareto}}$ is not a Pareto optimal solution of (92), for (92) there will exist $\boldsymbol{F}_1$ which satisfies $\boldsymbol{F}_1^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}_1 \succeq \boldsymbol{F}_{\text{Pareto}}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{Pareto}}$ and $\mathrm{Tr}(\sum_{i=1}^{I}\alpha_i\boldsymbol{\Omega}_i\boldsymbol{F}_1\boldsymbol{F}_1^{\mathrm{H}})\leq P$. For positive semidefinite matrices, $\boldsymbol{A}\succeq\boldsymbol{B}$ implies $\lambda(\boldsymbol{A})\succeq\lambda(\boldsymbol{B})$ [40, P471]. Meanwhile, for two complex matrices $\boldsymbol{C}$ and $\boldsymbol{D}$, $\boldsymbol{C}\boldsymbol{D}$ and $\boldsymbol{D}\boldsymbol{C}$ have the same nonzero eigenvalues. Therefore we have

$$\lambda(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_1\boldsymbol{F}_1^{\mathrm{H}}\boldsymbol{\Pi}^{\frac{1}{2}})\succeq\lambda(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_{\text{Pareto}}\boldsymbol{F}_{\text{Pareto}}^{\mathrm{H}}\boldsymbol{\Pi}^{\frac{1}{2}}), \quad (93)$$

based on which it can be concluded that we can find a matrix $\boldsymbol{F}_2$ which satisfies

$$\begin{aligned}&\boldsymbol{F}_2\boldsymbol{F}_2^{\mathrm{H}}\preceq\boldsymbol{F}_1\boldsymbol{F}_1^{\mathrm{H}},\\&\lambda(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_2\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{\Pi}^{\frac{1}{2}})=\lambda(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{F}_{\text{Pareto}}\boldsymbol{F}_{\text{Pareto}}^{\mathrm{H}}\boldsymbol{\Pi}^{\frac{1}{2}}).\end{aligned} \quad (94)$$

As the weighted power constraint is right unitarily invariant, there will exist a unitary matrix $\boldsymbol{Q}_2$ for which the following equality holds:

$$\boldsymbol{Q}_2^{\mathrm{H}}\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}_2\boldsymbol{Q}_2=\boldsymbol{F}_{\text{Pareto}}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{Pareto}}. \quad (95)$$

Taking $\boldsymbol{F}_2\boldsymbol{Q}_2=\boldsymbol{F}_3$ as a new variable, it is concluded that for (91) $\boldsymbol{F}_3^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}_3=\boldsymbol{F}_{\text{Pareto}}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}_{\text{Pareto}}$ and $\sum_{i=1}^{I}\alpha_i\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}_3\boldsymbol{F}_3^{\mathrm{H}})<\sum_{i=1}^{I}\alpha_i\mathrm{Tr}(\boldsymbol{\Omega}_i\boldsymbol{F}_2\boldsymbol{F}_2^{\mathrm{H}})\leq P$. This contradicts to the previous conclusion. In other words, $\boldsymbol{F}_{\text{Pareto}}$ is a Pareto optimal solution of (92).

In a nutshell, for any Pareto optimal solution of **Opt. 1.6**, there exist the weights α_i, $1\leq i\leq I$, for ensuring that this Pareto optimal solution of **Opt. 1.6** is also the Pareto optimal solution of (92). Therefore, it can be concluded that any Pareto optimal solution of **Opt. 1.6** satisfies the common structures of the Pareto optimal solutions of (92).

Next, we show that the Pareto optimal solutions of (92) own the same diagonalizable structure and thus this structure is also the optimal structure of the Pareto optimal solutions of **Opt. 1.6**. First we define the auxiliary variables

$$\widetilde{\boldsymbol{F}}=\left(\sum_{i=1}^{I}\alpha_i\boldsymbol{\Omega}_i\right)^{\frac{1}{2}}\boldsymbol{F}, \ \boldsymbol{\Omega}=\left(\sum_{i=1}^{I}\alpha_i\boldsymbol{\Omega}_i\right). \quad (96)$$

Then the optimization (92) is transferred into:

$$\max_{\widetilde{\boldsymbol{F}}} \ \widetilde{\boldsymbol{F}}^{\mathrm{H}}\left(\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}, \ \text{ s.t. } \mathrm{Tr}\left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\right)\leq P. \quad (97)$$

The Pareto optimal solution set of (97) consists of the optimal solutions of the following optimization problem for all the possible $\widetilde{\boldsymbol{F}}_{\text{in}}$ that are in the sphere region of $\mathrm{Tr}(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}})\leq P$:

$$\begin{aligned}&\max_{\widetilde{\boldsymbol{F}},\alpha} \ \alpha,\\&\text{s.t. } \widetilde{\boldsymbol{F}}^{\mathrm{H}}\left(\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}=\alpha\widetilde{\boldsymbol{F}}_{\text{in}}^{\mathrm{H}}\left(\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}},\\&\quad\ \mathrm{Tr}\left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\right)\leq P.\end{aligned} \quad (98)$$

The first constraint in (98) is equivalent to

$$\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}=\sqrt{\alpha}\boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}. \quad (99)$$

Using pseudo inverse, we have

$$\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}=\sqrt{\alpha}\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}, \quad (100)$$

based on which we have

$$\begin{aligned}&\left\|\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}\right\|_{\mathrm{F}}^2\\&=\mathrm{Tr}\left([\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}]^{\mathrm{H}}\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}\right)\\&=\alpha\mathrm{Tr}\left([\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}]^{\mathrm{H}}\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}\right)\\&=\alpha\left\|\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}\right\|_{\mathrm{F}}^2.\end{aligned} \quad (101)$$

Therefore, α is expressed as:

$$\alpha=\frac{\left\|\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}\right\|_{\mathrm{F}}^2}{\left\|\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}\right\|_{\mathrm{F}}^2}. \quad (102)$$

Based on **Matrix Inequality 1**, the numerator of (102) satisfies

$$\left\|\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}\right\|_{\mathrm{F}}^2\leq\sum_j\lambda_j\left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\right), \quad (103)$$

while its denominator satisfies

$$\left\|\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)^{\dagger}\boldsymbol{U}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}\right\|_{\mathrm{F}}^2\geq\sum_j\frac{\sigma_j^2\left(\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}\widetilde{\boldsymbol{F}}_{\text{in}}\right)}{\sigma_j^2\left(\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\right)}, \quad (104)$$

where $\sigma_j^2(\boldsymbol{A})$ denotes the jth the both equalities in (103) and (104) hold. For the optimal $\widetilde{\boldsymbol{F}}$ and $\boldsymbol{U}$ together with the fact that for **Opt. 1.6**, the optimal $\boldsymbol{F}$ is right unitary invariant, the optimal $\widetilde{\boldsymbol{F}}$ satisfies the following structure

$$\widetilde{\boldsymbol{F}}=\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\text{Arb}}^{\mathrm{H}}, \quad (105)$$

where the unitary matrix $\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}$ is defined based on the following EVD

$$\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}=\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}^{\mathrm{H}} \text{ with } \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{\Pi}}}\searrow. \quad (106)$$

Based on (105) and the definition of $\widetilde{\boldsymbol{F}}$ in (96)

$$\boldsymbol{F}=\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\text{Arb}}^{\mathrm{H}}. \quad (107)$$

Appendix C
Bayes Robust Matrix-Monotonic Optimization

A. Shaping Constraint

With the shaping constraint, **Opt. 2.2** becomes

$$\begin{aligned}&\max_{\boldsymbol{F}} \ \boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F},\\&\ \text{s.t.} \ \ \boldsymbol{K}_{\mathrm{n}}=\sigma_{\mathrm{n}}^2\boldsymbol{I}+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\boldsymbol{I},\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\preceq\boldsymbol{R}_{\mathrm{s}}.\end{aligned} \quad (108)$$

Note that $\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi})\leq\mathrm{Tr}(\boldsymbol{R}_{\mathrm{s}}\boldsymbol{\Psi})$ and then we have the following matrix inequality

$$\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F}\succeq\frac{\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\widehat{\boldsymbol{H}}\boldsymbol{F}}{\sigma_{\mathrm{n}}^2+\mathrm{Tr}(\boldsymbol{R}_{\mathrm{s}}\boldsymbol{\Psi})}. \quad (109)$$

Replacing the objective in (108) by its lower bound in (109), the following optimization problem is formulated:

$$\max_{\boldsymbol{F}} \frac{\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\widehat{\boldsymbol{H}}\boldsymbol{F}}{\sigma_{\mathrm{n}}^2+\mathrm{Tr}(\boldsymbol{R}_{\mathrm{s}}\boldsymbol{\Psi})}, \text{ s.t. } \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} \preceq \boldsymbol{R}_{\mathrm{s}}. \tag{110}$$

whose Pareto optimal solution is given by Lemma 1. It is obvious that when $\boldsymbol{\Psi}=0$ or $\boldsymbol{\Psi}\propto\boldsymbol{I}$ and $\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}})=\mathrm{Tr}(\boldsymbol{R}_{\mathrm{s}})$ is achievable then the lower bound is tight.

B. Joint Power Constraints

Under the joint power constraints, **Opt. 2.2** is written in the following formula

$$\begin{aligned}\max_{\boldsymbol{F}} \;& \boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F},\\ \text{s.t.} \;& \boldsymbol{K}_{\mathrm{n}}=\sigma_{\mathrm{n}}^2\boldsymbol{I}+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\boldsymbol{I},\\ & \mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right)\le P,\; \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\preceq\tau\boldsymbol{I}.\end{aligned} \tag{111}$$

The sum power constraint $\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}})\le P$ is equivalent to the following equality [26]

$$\left(\sigma_{\mathrm{n}}^2+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\right)^{-1}\mathrm{Tr}\left[\left(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}\right)\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right]\le P. \tag{112}$$

Based on (112), the optimization problem (111) is equivalent to the following one

$$\begin{aligned}\max_{\boldsymbol{F}} \;& \boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{K}_{\mathrm{n}}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F},\\ \text{s.t.} \;& \boldsymbol{K}_{\mathrm{n}}=\sigma_{\mathrm{n}}^2\boldsymbol{I}+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\boldsymbol{I},\\ & \frac{\mathrm{Tr}\left[\left(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}\right)\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\right]}{\sigma_{\mathrm{n}}^2+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)}\le P,\; \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\preceq\tau\boldsymbol{I}.\end{aligned} \tag{113}$$

By defining the following matrix variable

$$\widetilde{\boldsymbol{F}}=\left[\sigma_{\mathrm{n}}^2+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)\right]^{-\frac{1}{2}}\left(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}\right)^{\frac{1}{2}}\boldsymbol{F}, \tag{114}$$

the optimization problem (113) can be transferred into the following equivalent one

$$\begin{aligned}\max_{\widetilde{\boldsymbol{F}}} \;& \widetilde{\boldsymbol{F}}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\widehat{\boldsymbol{H}}\widetilde{\boldsymbol{F}},\\ \text{s.t.} \;& \mathrm{Tr}\left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\right)\le P,\; \widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\preceq\tau\frac{\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}}{\sigma_{\mathrm{n}}^2+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)}.\end{aligned} \tag{115}$$

For the final matrix inequality, we have the following lower bound of the righthand side term, i.e.,

$$\tau\frac{\sigma_{\mathrm{n}}^2+P\lambda_{\min}(\boldsymbol{\Psi})}{\sigma_{\mathrm{n}}^2+P\lambda_{\max}(\boldsymbol{\Psi})}\boldsymbol{I}\preceq\tau\frac{\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}}{\sigma_{\mathrm{n}}^2+\mathrm{Tr}\left(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Psi}\right)}, \tag{116}$$

where the equality holds when $\boldsymbol{\Psi}\propto\boldsymbol{I}$. Based on the lower bound in (116), for the Pareto optimal solutions of the following optimization problem, the corresponding objective is a lower bound of that in (113)

$$\begin{aligned}\max_{\widetilde{\boldsymbol{F}}} \;& \widetilde{\boldsymbol{F}}^{\mathrm{H}}\left(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}\right)^{-\frac{1}{2}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\widehat{\boldsymbol{H}}\left(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}\right)^{-\frac{1}{2}}\widetilde{\boldsymbol{F}},\\ \text{s.t.} \;& \mathrm{Tr}\left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\right)\le P,\; \widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\preceq\tau\frac{\sigma_{\mathrm{n}}^2+P\lambda_{\max}(\boldsymbol{\Psi})}{\sigma_{\mathrm{n}}^2+P\lambda_{\min}(\boldsymbol{\Psi})}\boldsymbol{I}.\end{aligned} \tag{117}$$

It is obvious that based on Lemma 2 the Pareto optimal solutions of (117) satisfy the following structure

$$\widetilde{\boldsymbol{F}}=\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}} \tag{118}$$

where the unitary matrix $\boldsymbol{V}_{\widehat{\boldsymbol{H}}}$ is defined based on the SVD

$$\widehat{\boldsymbol{H}}\left(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}\right)^{-\frac{1}{2}}=\boldsymbol{U}_{\widetilde{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{H}}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}^{H}, \text{with}, \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{H}}}\searrow. \tag{119}$$

The diagonal elements of the rectangular diagonal matrix $\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}_k}$ are smaller than $\sqrt{\tau(\sigma_{\mathrm{n}}^2+P\lambda_{\min}(\boldsymbol{\Psi}))/(\sigma_{\mathrm{n}}^2+P\lambda_{\max}(\boldsymbol{\Psi}))}$. Based on the definition in (114), $\boldsymbol{F}$ equals

$$\boldsymbol{F}=\frac{\sigma_{\mathrm{n}}\widetilde{\boldsymbol{\Psi}}^{-\frac{1}{2}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}}}{\left(1-\mathrm{Tr}\left(\widetilde{\boldsymbol{\Psi}}^{-\frac{1}{2}}\boldsymbol{\Psi}\widetilde{\boldsymbol{\Psi}}^{-\frac{1}{2}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}^{\mathrm{H}}\boldsymbol{V}_{\widetilde{\boldsymbol{H}}}^{\mathrm{H}}\right)\right)^{\frac{1}{2}}}, \tag{120}$$

where $\widetilde{\boldsymbol{\Psi}}=\sigma_{\mathrm{n}}^2\boldsymbol{I}+P\boldsymbol{\Psi}$.

References

[1] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[2] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Commun.*, vol. 10, no. 2, pp. 585–595, Nov./Dec. 1999.

[4] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[5] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 239–241, Jun. 2002.

[6] A. Scaglione *et al.*, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[7] A. Feiten, R. Mathar, and S. Hanly, "Eigenvalue-based optimum-power allocation for Gaussian vector channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2304–2309, Jun. 2007.

[8] A. Yadav, M. Juntti, and J. Lilleberg, "Linear precoder design for doubly correlated partially coherent fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3621–3635, Jul. 2014.

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

[10] D. P. Palomar and Y. Jiang, "MIMO transceiver designs via majorization theory," *Foundations and Trends in Commun. and Inf. Theory*, vol. 3, no. 4-5, pp. 331–551, Jun. 2007.

[11] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.

[12] C. Weng, C. Chen, and P. P. Vaidyanathan, "MIMO transceivers with decision feedback and bit loading: Theory and optimization," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1334–1346, Mar. 2010.

[13] C. Liu and P. P. Vaidyanathan, "MIMO broadcast DFE transceivers with QoS constraints: Min-power and max-rate solutions," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5550–5562, Nov. 2013.

[14] C. Weng and P. P. Vaidyanathan, "MIMO transceiver optimization with linear constraints on transmitted signal covariance components," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 458–462, Jan. 2010.

[15] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.

[16] 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.

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

[18] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[19] J. Wang, M. Bengtsson, B. Ottersten, and D. P. Palomar, "Robust MIMO precoding for several classes of channel uncertainty," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3056–3070, Jun. 2013.

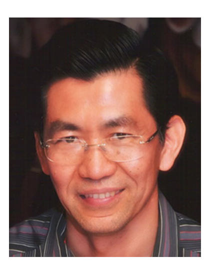
[20] A. Pastore, M. Joham, and J. R. Fonollosa, "A framework for joint design of pilot sequence and linear precoder," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5059–5079, Sep. 2016.
[21] W. Yu and T. Lan, "Transceiver optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.
[22] J. Fang, H. Li, Z. Chen, and Y. Gong, "Joint precoder design for distributed transmission of correlated sources in sensor networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2918–2929, Jun. 2013.
[23] C. Xing, S. Li, Z. Fei, and J. Kuang, "How to understand linear minimum mean square error transceiver design for multiple input multiple output systems from quadratic matrix programming," *IET Commun.*, vol. 7, no. 12, pp. 1231–1242, Aug. 2013.
[24] C. Xing, *et al.*"A general robust linear transceiver design for amplify-and-forward multi-hop MIMO relaying systems," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1196–1209, Mar. 2013.
[25] C. Xing, M. Xia, F. Gao and Y.-C. Wu, "Robust transceiver with Tomlinson-Harashima precoding for amplify-and-forward MIMO relaying systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1370–1382, Sep. 2012.
[26] C. Xing, S. Ma, and Y. Zhou, "Matrix-monotonic optimization for MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 334–348, Jan. 2015.
[27] E. Jorswieck and H. Boche, "Majorization and matrix-monotone functions in wireless communications," *Foundation Trends Commun. Inf. Theory*, vol. 3, no. 6, pp. 553–701, Jul. 2007.
[28] S. X. Wu, Q. Li, A. M. So, and W. Ma, "A stochastic beamformed amplify-and-forward scheme in a multigroup multicast MIMO relay network with per-antenna power constraints," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4973–4986, Jul. 2016.
[29] A. Tölli, M. Codreanu, and M. Juntti, "Linear multiuser MIMO transceiver design with quality of service and per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3049–3055, Jul. 2008.
[30] M. Dong, B. Liang, and Q. Xiao, "Unicast multi-antenna relay beamforming with per-antenna power control: Optimization and duality," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 6076–6090, Dec. 2013.
[31] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, Oct. 2014.
[32] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z.-Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Signal Process.*, vol. 56, no. 9, pp. 4306–4316, Sep. 2008.
[33] M. Vu, "MIMO capacity with per-antenna power constraint," in *Proc. GLOBECOM* Houston, USA, Dec. 5–9, 2011, pp. 1–5.
[34] C. Xing, Y. Ma, Y. Zhou, and F. Gao, "Transceiver optimization for multi-hop communications with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1519–1534, Mar. 2016.
[35] D. P. Palomar, "Unified framework for linear MIMO transceivers with shaping constraints," *IEEE Commun. Lett.*, vol. 8, no. 12, pp. 697–699, Dec. 2004.
[36] C. Xing, F. Gao, and Y. Zhou, "A framework for transceiver designs for multi-hop communications with covariance shaping constraints," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3930–3945, Aug. 2015.
[37] J. Dai, C. Chang, W. Xu, and Z. Ye, "Linear precoder optimization for MIMO systems with joint power constraints," *IEEE Trans. Commun.*, vol. 60, no. 8, pp. 2240–2254, Aug. 2012.
[38] S. Wang, S. Ma, C. Xing, S. Gong, and J. An, "Optimal training design for MIMO systems with general power constraints," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3649–3664, Jul. 2018.
[39] C. Xing, W. Li, S. Ma, Z. Fei, and J. Kuang, "A matrix-field weighted mean-square-error model for MIMO transceiver design," *Commun. Lett.*, vol. 17, no. 8, pp. 1652–1655, Aug. 2013.
[40] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press: Cambridge, UK, 1990.
[41] B. Simon, *Loewner's Theorem on Monotone Matrix Functions (Grundlehren der mathematischen Wissenschaften Book 354).* Switzerland: Springer, 2019.
[42] X. Zhang, *Matrix Inequalities (Lecture Notes in Mathematics).* New York, NY, USA: Springer, 2002.
[43] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York, NY, USA: Academic Press, 1979.
[44] C. Xing, Y. Jing, and Y. Zhou, "On weighted MSE model for MIMO transceiver optimization," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7072–7085, Aug. 2017.
[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.
[46] C. Xing, Y. Jing, S. Wang, S. Ma, and H. V. Poor, "New viewpoint and algorithms for water-filling solutions in wireless communications," *IEEE Trans. Signal Process.*, vol. 68, pp. 1618–1634, Feb. 2020.
[47] S. Gong, S. Wang, C. Xing, S. Ma, and T. Q. S. Quek, "Robust superimposed training optimization for UAV assisted communication systems," *IEEE Trans. Commun.*, vol. 19, no. 3, pp. 1704–1721, Mar. 2020.
[48] M. C. Grant and S. P. Boyd, *The CVX Users' Guide* (Release 2.1) CVX Research, Inc., 2015

Chengwen Xing (Member, IEEE) received the B.Eng. degree from Xidian University, Xi'an, China, in 2005, and the Ph.D. degree from the University of Hong Kong, Hong Kong, in 2010.

Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, where he is currently a Full Professor. From September 2012 to December 2012, he was a Visiting Scholar with the University of Macau. His current research interests include statistical signal processing, convex optimization, multivariate statistics, combinatorial optimization, massive multi-in multi-out systems, and high frequency band communication systems.

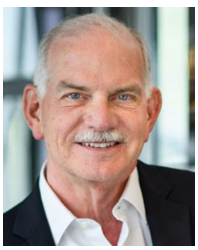
Prof. Xing is an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *KSII Transactions on Internet and Information Systems*, *Transactions on Emerging Telecommunications Technologies*, and *China Communications.*

Shuai Wang received the B.Eng degree (Hons.) from Zhengzhou University, Zhengzhou, China, in 2005, and the Ph.D. degree (Hons.) from the Beijing Institute of Technology, Beijing, China, in 2012, all in communications engineering. From 2010 to 2011, he was a Visiting Ph.D. Student with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K. Since July 2012, he has been with the School of Information Science and Electronics, Beijing Institute of Technology, where he is currently an Associate Professor. He has authored/coauthored more than 20 peer-reviewed articles, mainly in leading IEEE journals or conferences. His research interests include channel estimation, anti-jamming transmission, synchronization techniques, and beamforming. He served/serving as the Principal Investigator for 12 research funds, including two granted by the National Science Foundation of China. He was the recipient of the (Second Class) Scientific and Technical Progress Award granted by the Ministry of Industry and Information Technology of China. He also holds 17 patents. He was also the recipient of the award for outstanding Ph.D. dissertation granted by the Beijing Municipal Education Commission in 2013 with other 49 co-winners, nominated from all the Ph.D. graduates that received their degree in Beijing.



Sheng Chen (Fellow, IEEE) received the B.Eng. degree from the East China Petroleum Institute, Dongying, China, in 1982, and the Ph.D. degree from the City University, London, U.K., in 1986, both in control engineering, and the Doctor of Sciences (D.Sc.) degree from the University of Southampton, Southampton, U.K., in 2005. From 1986 to 1999, he held research and academic appointments with the Universities of Sheffield, Edinburgh, and Portsmouth, all in U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of Professor in Intelligent Systems and Signal Processing. His research interests include neural network and machine learning, adaptive signal processing, and wireless communications, and nonlinear system modelling. He has authored/coauthored more than 700 research papers. He is a Fellow of the United Kingdom Royal Academy of Engineering, a Fellow of IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an original ISI highly cited researcher in engineering (March 2004). He has 15 300+ Web of Science citations with h-index 54, and 30 900+ Google Scholar citations with h-index 75.

Shaodan Ma received the double bachelor's degrees in science and economics, and the master's degree of engineering, both from Nankai University, Tianjin, China, in 1999 and 2002, respectively, and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2006. From 2006 to 2011, she was a Postdoctoral Fellow with the University of Hong Kong. Since August 2011, she has been with the University of Macau, Macao, China, where she is currently an Associate Professor. She was a Visiting Scholar with Princeton University in 2010. Her research interests are in the general areas of signal processing and communications, particularly, transceiver design, resource allocation and performance analysis.



H. Vincent Poor (Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University, Princeton, NJ, USA, in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton University, where he is the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he was a Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, including most recently at Berkeley and Cambridge. His research interests are in the areas of information theory, machine learning and network science, and their applications in wireless networks, energy systems and related fields. Among his publications in these areas is the forthcoming book *Advanced Data Analytics for Power Systems* (Cambridge University Press, 2021).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a Foreign Member of the Chinese Academy of Sciences, the Royal Society and other national and international academies. He was the recipient of the Technical Achievement and Society Awards of the IEEE Signal Processing Society in 2007 and 2011, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, the 2019 ASEE Benjamin Garver Lamme Award, a D.Sc. *honoris causa* from Syracuse University, awarded in 2017, and a D.Eng. *honoris causa* from the University of Waterloo, awarded in 2019.

Lajos Hanzo (Fellow, IEEE), received the master's and Doctorate degrees from the Technical University (TU) of Budapest, Budapest, Hungary, in 1976 and 1983, respectively, and the Doctor of Sciences degree from the University of Southampton, Southampton, U.K., in 2004 and Honorary Doctorates from the TU of Budapest in 2009 and from the University of Edinburgh, Edinburgh, U.K., in 2015. He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served several terms as Governor of both IEEE ComSoc and of VTS. He has authored/coauthored 1900+ contributions at IEEE Xplore, 19 Wiley-IEEE Press books and has helped the fast-track career of 123 Ph.D. students. Over 40 of them are Professors at various stages of their careers in academia and many of them are leading Scientists in the wireless industry. He is Fellow of the Royal Academy of Engineering of the IET and of EURASIP.