A General Matrix Variable Optimization Framework for MIMO Assisted Wireless Communications

Chengwen Xing, Yihan Li, Shiqi Gong, Jianping An, Sheng Chen, *Fellow, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Complex matrix derivatives play an important role in matrix optimization, since they form a theoretical basis for the Karush-Kuhn-Tucker (KKT) conditions associated with matrix variables. We commence with a comprehensive discussion of complex matrix derivatives. First, some fundamental conclusions are presented for deriving the optimal structures of matrix variables from complex matrix derivatives. Then, some restrictions are imposed on complex matrix derivatives for ensuring that the resultant first order equations in the KKT conditions exploit symmetric properties. Accordingly, a specific family of symmetric matrix equations is proposed and their properties are unveiled. Using these symmetric matrix equations, the optimal structures of matrix variables are directly available, and thereby the original optimization problems can be significantly simplified. In addition, we take into account the positive semidefinite constraints imposed on matrix variables. In order to accommodate the positive semidefinitness of matrix variables, we introduce a matrix transformation technique by leveraging the symmetric matrix equations, which can dramatically simplify the KKT conditions based analysis albeit at the expense of destroying convexity. Moreover, this matrix transformation technique is valuable in practice, since it offers a more efficient means of computing the optimal solution based on the optimal structures derived directly from the KKT conditions.**

***Index Terms*—Matrix variable optimization, complex matrix derivatives, Karush-Kuhn-Tucker conditions, matrix symmetric structures, matrix variable transformation.**

I. Introduction

Matrix optimization plays an essential role in wireless system designs [1]–[4]. For example, multiple-input multiple-output (MIMO) system optimization, including both transmit precoder (TPC) matrix optimization and equalizer matrix optimization, is an important design issue in wireless systems [1], [2], [4]–[8]. Complex matrix derivatives provide efficient tools for solving these advanced matrix optimization problems [9]. In view of the importance of complex matrix derivatives, there is already some literature, in which many insightful results are tabulated for readers to look up the required ones [9]–[13]. With the evolution of wireless technologies, many new matrix optimization problems are emerging subject to new constraints. Solving these problems generally requires efficient mathematical optimization tools, which which is the core of our work.

Because the matrix variables are typically high-dimensional, the key issue in matrix optimization is how to reduce the computational complexity. A popular strategy is based on the Karush-Kuhn-Tucker (KKT) conditions, which constitute necessary conditions for finding optimal solutions [14]. Specifically, for convex optimization problems, the solutions that satisfy the KKT conditions are the optimal solutions, since the KKT conditions act as sufficient and necessary conditions for this case. Even for many non-convex problems, all the solutions derived from the KKT conditions have a common structure, which is also the analytical structure of the optimal solution. Therefore, the KKT conditions having beneficial mathematical tractability play a vital role in solving general optimization problems [3], [15]. The second strategy is that of leveraging majorization theory, which relies on matrix inequalities associated with the diagonal elements of Hermitian matrices [2]. However, there are strict limitations on the optimization objective functions and constraints when utilizing the majorization theory based methods. For example, for the classical MIMO transceiver design, the majorization theory is only suitable for handling Schur-convex or Schur-concave objective functions subject to the total power constraint [16], [17]. The third one is referred to as the matrix monotonic optimization framework, which exploits the monotonicity of the positive semidefinite matrix cone to derive the optimal structures of matrix variables [1], [18], [19]. Unfortunately, the matrix monotonic optimization framework is not applicable to many practical applications, such as the optimization problems where multiple constraints conflict with each other.

Among the three strategies, the KKT conditions based strategy may be declared to be the prime technique of matrix variable optimization, given its appealing mathematical simplicity [3], [15], which has also been well studied by wireless researchers. Nevertheless, in general, the KKT conditions are only necessary conditions for the optimal solutions. The complex matrix derivative is a fundamental tool conceived

Manuscript received 4 February 2023; revised 8 July 2023; accepted 4 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62101614 and Grant 62231004. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1, EP/X01228X/1 and EP/X01228X/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). The associate editor coordinating the review of this article and approving it for publication was Cao, Xuanyu. *(Corresponding author: Shiqi Gong.)*

C. Xing and Y. Li are with School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (E-mails: chengwenxing@ieee.org, yihanli.bit@gmail.com).

S. Gong and J. An are with the School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China (E-mails:gsqyx@163.com, an@bit.edu.cn).

S. Chen and L. Hanzo are with School of Electronics and Computer Science, University of Southampton, U.K. (E-mails: sqc@ecs.soton.ac.uk, lh@ecs.soton.ac.uk).

TABLE I
COMPARISON BETWEEN OUR WORK AND THE EXISTING LITERATURE

		[1]	[2]	[3]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	Proposed
KKT conditions	MSE				✓							✓
	Capacity			✓							✓	✓
	QoS				✓					✓		✓
	Data rate				✓					✓		✓
	Robust rate											✓
	Robust MSE											✓
Majorization theory	MSE		✓			✓						✓
	SINR		✓									✓
	Rate						✓					✓
	BER		✓									✓
Matrix monotonic framework	MSE	✓						✓	✓			✓
	Capacity	✓						✓	✓			✓
	BER							✓	✓			✓

for deriving the KKT conditions, which usually has close relationships with the specific structures of matrix variables. A number of papers and textbooks have been published on complex matrix derivatives [9], [13]. Building upon these works, the complex matrix derivatives for the KKT conditions associated with matrix variables can be derived.

The studies [3], [15] have shown the advantages of the KKT conditions based algorithms, in which complex matrix derivatives act as a theoretical basis for formulating the KKT conditions. Generally, there exist equivalent variants of a real-valued function based on matrix manipulations, thereby leading to complex matrix derivatives in a diverse range of mathematical formulas. Motivated by this fact, the KKT conditions based methods usually rely on a case-by-case implementation. Specifically, it was found in [3] that the optimal water-filling structure of the positive semidefinite transmit signal covariance matrix directly accrues from the KKT conditions. In [21], the optimal structures of matrix variables were derived from KKT conditions utilizing uplink-downlink duality. However, the authors of [15] pointed out that the optimal solution structure is difficult to directly obtain from the KKT conditions, since the corresponding optimization problem is non-convex. Fortunately, these KKT conditions can be simplified by exploiting the fact that when a matrix multiplied by a diagonal matrix is diagonal, the original matrix must also be diagonal. Based on this, the optimal closed-form linear TPC/equalizer was derived and also proved to be able to diagonalize the MIMO channel into its eigen sub-channels. In the face of the ever increasing requirements for wireless communications, the performance optimization problems can be diverse. Considering the existence of multiple objective functions, a general framework of KKT conditions based matrix variable optimization is investigated in this paper. Our novel contributions are contrasted to the above-mentioned literature, which can be seen at a glance in Table I and are further detailed as follows:

- We provide a number of fundamental conclusions regarding complex matrix derivatives for wireless communications. It is noted that there are several mathematical formulations for traditional complex matrix derivatives. In contrast to ignoring some strict restrictions on complex matrix derivatives in some classical textbooks, we additionally define symmetric complex matrix derivative operators for reasons of mathematical rigour. Therefore, our provided fundamental conclusions form a firm basis for the successive theoretical analysis and mathematical derivations.
- Inspired by matrix Hermitian symmetry, we propose the concept of symmetric matrix equations and some important conclusions to simplify the related theoretical analysis. Considering several typical matrix optimization problems in MIMO systems, we derive the corresponding symmetric matrix equations from their KKT conditions. Based on this, the optimal structures of matrix variables are available and thereby the original optimization problems can be significantly simplified.
- For practical wireless communications, we consider the positive semi-definite constraints imposed on matrix variables, which substantially affect the resultant complex matrix derivatives. Therefore, we introduce an efficient variable transformation technique to transform the positive semi-definite matrices into matrices associated with independent variables. It is shown that this transformation technique can automatically satisfy the matrix rank constraints. Although the variable transformation may destroy the convexity of the original optimization problems, it can simplify the derivation of the optimal structures of matrix variables. Consequently, the optimization problems can be drastically simplified.

This paper is organized as follows. In Section II, we present fundamental definitions and results concerning complex matrix derivatives, which provide the theoretical basis for our work. In Section III, we derive some fundamental results for the family of the symmetric matrix equations, which form the basis of the KKT conditions based methods harnessed for deriving the optimal structures of matrix variables. Moreover, a number of specific applications are presented in this section. In Section IV, the positive semidefinite constraints are investigated, and the corresponding symmetric matrix equations are derived. In Section V, our numerical results are presented in support of our conclusions offered in Section VI.

Notations: We use regular letters for scalars, lowercase and uppercase boldface letters for vectors and matrices, respectively. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{-1}$ represent the conjugate, transpose, Hermitian and inverse operators, respectively. $\mathbf{0}$ denote the zero matrix and $\boldsymbol{I}$ denote the identity matrix. For a complex matrix $\boldsymbol{X}$, $\boldsymbol{X}_{\mathrm{R}}$ and $\boldsymbol{X}_{\mathrm{I}}$ denote its real and imaginary parts, respectively. $[\boldsymbol{X}]_{i,j}$ denote the (i,j)-th element of $\boldsymbol{X}$. $|\boldsymbol{X}|$ and $\mathrm{Tr}(\boldsymbol{X})$ represent the determinant and trace of $\boldsymbol{X}$, respectively. $\lambda_i(\boldsymbol{X})$ denote the ith largest eigenvalue of $\boldsymbol{X}$. $\boldsymbol{X} \succeq \mathbf{0}$ means that the matrix $\boldsymbol{X}$ is positive semidefinite and $(x)^+ = \max\{0, x\}$.

TABLE II
SUMMARY OF MAIN SYMBOLS

Symbols	Denotations	Symbols	Denotations
$\boldsymbol{s}$	The transmitted signal	$\boldsymbol{\Omega}_n$	The weighting matrix associated with the per-antenna power constraints
$\boldsymbol{n}$	The additive noise	$\boldsymbol{\Omega}$	The equivalent weighting matrix associated with the sum power constraints
$\boldsymbol{F}$	The precoding matrix	$\boldsymbol{\Pi}$	The covariance matrix of the additive noise
$\boldsymbol{H}$	The channel matrix	$\boldsymbol{\Lambda}$	The diagonal matrix of SVD
$\boldsymbol{Q}$	The covariance matrix of the transmitted signal	μ	The Lagrange multiplier associated with the sum power constraints

II. Fundamental Results on Complex Matrix Derivatives

Optimization relying on matrix variables is more general than based on vector or scalar variables, but it subsumes both vector and scalar optimization as its special cases. Secondly, complex-valued matrix derivatives play an important role in wireless communications. From the engineering point of view, the phases of signals must be taken into account. Therefore, the matrices involved in wireless communications are usually defined as complex matrices. However, new applications are continuously emerging, which involve new complex matrix variable optimization subject to specific structural constraints. For example, in hybrid beamforming optimization or transceiver design, the complex-valued analog matrix is a special matrix variable with each element subject to constant modulus constraints. Another example is a reconfigurable intelligent surface aided MIMO system, where the complex reflecting matrix is a diagonal matrix having diagonal elements under constant modulus constraints. It is therefore necessary to investigate complex matrix derivative operators under practical structural constraints. This is a widely open issue.

Let the optimization objective function (OF) f be a real-valued scalar function of the complex matrix variable $\boldsymbol{X}=\boldsymbol{X}_{\mathrm{R}}+\mathsf{j}\boldsymbol{X}_{\mathrm{I}}$, where $\mathsf{j}=\sqrt{-1}$. Then the following fundamental linear complex matrix derivative operators may be defined [9], [12], [13].

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}=\frac{1}{2}\left(\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{\mathrm{R}}}-\mathsf{j}\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{\mathrm{I}}}\right),$$
$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^*}=\frac{1}{2}\left(\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{\mathrm{R}}}+\mathsf{j}\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{\mathrm{I}}}\right). \quad (1)$$

The real matrix derivative operator involved in (1) is defined as follows [9], [12], [13]

$$\frac{\partial f(\boldsymbol{Z})}{\partial \boldsymbol{Z}}=\begin{bmatrix}\frac{\partial f(\boldsymbol{Z})}{\partial [\boldsymbol{Z}]_{1,1}} & \cdots & \frac{\partial f(\boldsymbol{Z})}{\partial [\boldsymbol{Z}]_{1,N_{\mathrm{c}}}}\\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{Z})}{\partial [\boldsymbol{Z}]_{N_{\mathrm{r}},1}} & \cdots & \frac{\partial f(\boldsymbol{Z})}{\partial [\boldsymbol{Z}]_{N_{\mathrm{r}},N_{\mathrm{c}}}}\end{bmatrix}, \quad (2)$$

where $\boldsymbol{Z}\in\mathbb{R}^{N_{\mathrm{r}}\times N_{\mathrm{c}}}$ is a real matrix. When $\boldsymbol{Z}$ is a symmetric matrix, the right hand side of (2) must also be a symmetric matrix.

Highlight 1. *When some structural constraints are imposed on $\boldsymbol{Z}$, e.g., $\boldsymbol{Z}$ is a diagonal matrix, the definition given in (2) becomes meaningless.*

It should be highlighted that for some intermediate steps, the complex matrix derivations with respect to $\boldsymbol{X}$ itself and with respect to its conjugate $\boldsymbol{X}^*$ are different. For example, the following complex matrix operators are defined in the classic textbooks [9], [12]

$$\frac{\partial \mathrm{Tr}(\boldsymbol{X}^{\mathrm{H}})}{\partial \boldsymbol{X}}=\boldsymbol{0},\ \frac{\partial \mathrm{Tr}(\boldsymbol{X}^{\mathrm{H}})}{\partial \boldsymbol{X}^*}=\boldsymbol{I}. \quad (3)$$

It is worth noting that the role of complex matrix derivatives is to find extreme values. From a mathematical viewpoint, it is meaningless to argue that a complex number is larger or smaller than another complex number. Therefore, setting the complex derivatives of a complex valued function is totally meaningless. In the classic textbook [12], the following complex matrix derivative operators were defined:

$$\frac{\partial \mathrm{Tr}(\boldsymbol{W}\boldsymbol{X})}{\partial \boldsymbol{X}}=\boldsymbol{W}^{\mathrm{T}},\ \frac{\partial \mathrm{Tr}(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}})}{\partial \boldsymbol{X}}=\boldsymbol{0}, \quad (4)$$

and

$$\frac{\partial \mathrm{Tr}(\boldsymbol{W}\boldsymbol{X})}{\partial \boldsymbol{X}^*}=\boldsymbol{0},\ \frac{\partial \mathrm{Tr}(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}})}{\partial \boldsymbol{X}^*}=\boldsymbol{W}^{\mathrm{H}}, \quad (5)$$

where $\boldsymbol{X}$ is a complex matrix variable and $\boldsymbol{W}$ is a complex matrix of appropriate dimension. It is worth noting that $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{X})$ can be a complex-valued function. In this case, the matrix derivative operators of (4) and (5) would become meaningless from a mathematical viewpoint. However, in real-world applications $\mathrm{Tr}(\boldsymbol{W}\boldsymbol{X})$ and $\mathrm{Tr}(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}})$ usually appear in together. Thus, it is more meaningful to define the following complex matrix derivative operators

$$\frac{\partial\big[\mathrm{Tr}(\boldsymbol{W}\boldsymbol{X})+\mathrm{Tr}\big(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\big)\big]}{\partial \boldsymbol{X}}=\boldsymbol{W}^{\mathrm{T}},$$
$$\frac{\partial\big[\mathrm{Tr}(\boldsymbol{W}\boldsymbol{X})+\mathrm{Tr}\big(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\big)\big]}{\partial \boldsymbol{X}^*}=\boldsymbol{W}^{\mathrm{H}}, \quad (6)$$

instead of the previous operators of (4) and (5).

Highlight 2. *The complex matrix derivative can be set either with respect to a complex matrix variable $\boldsymbol{X}$ itself or to its conjugate $\boldsymbol{X}^*$. For these two matrix derivatives, the resultant KKT conditions are exactly the same.*

The key is that regardless of which complex matrix derivation is used, the resultant KKT conditions must be the same. In the following, two quadratic derivative operators are defined

$$\frac{\partial \mathrm{Tr}(\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^{\mathrm{H}})}{\partial \boldsymbol{X}^*}=\boldsymbol{X}\boldsymbol{W},\quad \frac{\partial \mathrm{Tr}(\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^{\mathrm{H}})}{\partial \boldsymbol{X}}=(\boldsymbol{W}\boldsymbol{X}^{\mathrm{H}})^{\mathrm{T}}, \quad (7)$$

where $\boldsymbol{W}$ must be a Hermitian matrix, i.e., $\boldsymbol{W}=\boldsymbol{W}^{\mathrm{H}}$. If $\boldsymbol{W}$ is indeed a square matrix but not a Hermitian matrix, the definition in (7) becomes meaningless and the following operator can be defined

$$\frac{\partial[\mathrm{Tr}(\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^{\mathrm{H}})+\mathrm{Tr}(\boldsymbol{X}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}})]}{\partial \boldsymbol{X}^*}=\boldsymbol{X}(\boldsymbol{W}+\boldsymbol{W}^{\mathrm{H}}), \quad (8)$$

in which the term $(\boldsymbol{W}+\boldsymbol{W}^{\mathrm{H}})$ is definitely a Hermitian matrix.

The complex matrix derivative based methods are much more straightforward than the existing traditional signal processing techniques, and they can substantially simplify the optimization [22]. Upon considering beamforming as an example, the corresponding beamforming optimization problem is formulated as [22]

$$\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2}, \quad \text{s.t. } \boldsymbol{w}^{\mathrm{H}}\boldsymbol{w} \leq P. \tag{9}$$

The first order equation (FOEqu) of its KKT conditions may be formulated as

$$\frac{(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2)\boldsymbol{A}\boldsymbol{w}-(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{w})\boldsymbol{B}\boldsymbol{w}}{(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2)^2} = \lambda\boldsymbol{w}. \tag{10}$$

Based on (10) and exploiting that $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{w} = P$, the Lagrange multiplier can be derived as

$$\lambda = \frac{\sigma^2}{P}\frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{w}}{(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2)^2}. \tag{11}$$

Upon substituting (11) into the FOEqu (10), the following matrix equality holds

$$\boldsymbol{A}\boldsymbol{w} = \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2}\left(\boldsymbol{B}+\frac{\sigma^2}{P}\boldsymbol{I}\right)\boldsymbol{w}. \tag{12}$$

It may then be concluded that the optimal beamforming vector $\boldsymbol{w}$ equals to

$$\boldsymbol{w}_{\mathrm{opt}} = \mathcal{P}\left(\left(\boldsymbol{B}+\frac{\sigma^2}{P}\boldsymbol{I}\right)^{-1}\boldsymbol{A}\right), \tag{13}$$

where the operator $\mathcal{P}(\boldsymbol{M})$ denotes the principal eigenvector of $\boldsymbol{M}$ [23]. Let us now consider a more complex beamforming optimization problem under per-antenna power constraints, which is formulated as [22]

$$\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2}, \quad \text{s.t. } [\boldsymbol{w}\boldsymbol{w}^{\mathrm{H}}]_{n,n} \leq P_n. \tag{14}$$

The FOEqu of its KKT conditions is given by

$$\begin{aligned}\frac{(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2)\boldsymbol{A}\boldsymbol{w}-(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{w})\boldsymbol{B}\boldsymbol{w}}{(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{w}+\sigma^2)^2} &= \sum_n \lambda_n\boldsymbol{\Omega}_n\boldsymbol{w} \\ &= \lambda\sum_n \frac{\lambda_n}{\lambda}\boldsymbol{\Omega}_n\boldsymbol{w}. \end{aligned} \tag{15}$$

Let us now define $\frac{\lambda_n}{\lambda} \triangleq \widetilde{\lambda}_n$ and $\sum_n \widetilde{\lambda}_n\boldsymbol{\Omega}_n \triangleq \boldsymbol{\Omega}$, where the value of λ is chosen for ensuring that the optimal beamformer $\boldsymbol{w}$ satisfies $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{w} = \sum_n P_n$. Hence, the optimal $\boldsymbol{w}$ is given by

$$\boldsymbol{w}_{\mathrm{opt}} = \mathcal{P}\left(\left(\boldsymbol{B}+\frac{\sigma^2}{\sum_n P_n}\sum_n\boldsymbol{\Omega}\right)^{-1}\boldsymbol{A}\right), \tag{16}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix. If all the diagonal elements of $\boldsymbol{\Omega}$ are identical, the optimal solutions of the above two optimization problems, namely, (13) and (16), are the same. The values of $\{\widetilde{\lambda}_n\}$ can be computed using the subgradient algorithm, as proposed in [18]. It may then be concluded that the KKT conditions based method is a powerful counterpart of the family of numerical optimization methods relying on optimization software toolboxes.

Highlight 3. *In the traditional definition of complex matrix derivatives, there is no restriction imposed on the structures of the matrix variable. The elements of the matrix variable are independent variables.*

Conclusion 1. *Complex matrix derivatives may also be interpreted as concise and elegant expressions for multiple variables' derivatives. This kind of expression can substantially simplify the derivation and analysis processes. Furthermore, it is always possible to use separate real and imaginary parts to define the corresponding real matrices' derivatives.*

In the above definitions of this section, no structural constraints are imposed on the matrix variable $\boldsymbol{X}$. If there are some specific structural constraints, the results given above may become incorrect. For example, when $\boldsymbol{X}$ is a symmetric real matrix, i.e., $\boldsymbol{X} = \boldsymbol{X}^{\mathrm{T}}$, but $\boldsymbol{W}$ is not a symmetric real matrix, we have

$$\frac{\partial \mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{1}{2}(\boldsymbol{W}+\boldsymbol{W}^{\mathrm{T}}). \tag{17}$$

This example shows that a matrix specific structure significantly influences the matrix derivatives. Moreover, for the same function of complex matrix, there are more than one mathematical formulae for complex matrix derivatives. Here, the MIMO channel capacity may be considered as an example, which satisfies the following equalities

$$C = \log\left|\boldsymbol{I}+\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right| = \log\left|\boldsymbol{I}+\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right|, \tag{18}$$

where $\boldsymbol{F}$ is the linear TPC at the source and $\boldsymbol{H}$ is a constant channel matrix. Therefore, we have the following complex matrix derivative operators

$$\frac{\partial \log\left|\boldsymbol{I}+\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right|}{\partial \boldsymbol{F}^*} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\big(\boldsymbol{I}+\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\big)^{-1}\boldsymbol{F}, \tag{19}$$

$$\frac{\partial \log\left|\boldsymbol{I}+\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right|}{\partial \boldsymbol{F}^*} = \boldsymbol{H}^{\mathrm{H}}\big(\boldsymbol{I}+\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\big)^{-1}\boldsymbol{H}\boldsymbol{F}. \tag{20}$$

It is plausible that the complex matrix derivative operators in (19) and (20) must be equal to each other, since they are derived from the same OF. This conclusion is always true, because the following equality always holds:

$$\big(\boldsymbol{I}+\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\big)\boldsymbol{H} = \boldsymbol{H}\big(\boldsymbol{I}+\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\big). \tag{21}$$

The mathematical formulation of (20) is better than that of (19). Explicitly the formulation in (19) is not suitable for analysis, since the matrix $\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ may not have an eigenvalue decomposition (EVD) for a general matrix $\boldsymbol{F}$ [23].

Symmetry is an important property that can be exploited to derive the optimal structure of $\boldsymbol{F}$. The following section provides important design guidelines, based on which the optimal structure of matrix variables can be derived.

III. Matrix Symmetric Equations

In this section, several important design guidelines are provided for complex matrix derivatives to guarantee having a symmetric structure. First, we introduce the following matrix equality for the pair of Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of appropriate dimensions:

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{\Phi}, \tag{22}$$

where $\mathbf{\Phi}$ is also a Hermitian matrix. This kind of matrix equation is referred to as matrix symmetric equation. We will show that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\mathbf{\Phi}$ have the same unitary EVD matrix.

Lemma 1. *For a pair of Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ satisfying $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$ and $\boldsymbol{B}^{\mathrm{H}} = \boldsymbol{B}$, both $\boldsymbol{A}$ and $\boldsymbol{B}$ as well as $\boldsymbol{AB}$ have the same unitary EVD matrix, provided that $\boldsymbol{AB}$ is a Hermitian matrix, i.e., $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{AB}$.*

It should be pointed out that for a Hermitian matrix, its EVD is not unique. In other words, a Hermitian matrix can have multiple unitary EVD matrices. The statement in **Lemma 1** means that there exists at least one unitary matrix that is also the unitary EVD matrix of $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{AB}$. This result is given in **Theorem 4.1.6** of [23]. However, there is no requirement that $\boldsymbol{A}$ and $\boldsymbol{B}$ should be positive definite here.

Lemma 2. *For complex matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of appropriate dimensions, when $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{BA}$ and $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ have the same unitary EVD matrix, $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$ and $\boldsymbol{AA}^{\mathrm{H}}$ have the same unitary EVD matrix.*

Proof. The proof is given in Appendix A. □

Conclusion 2. *For a pair of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of appropriate dimensions, when $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AB}$ and $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$ have the same unitary EVD matrix, the right unitary matrix in singular value decomposition (SVD) of $\boldsymbol{A}$ and the left unitary matrix in SVD of $\boldsymbol{B}$ are the same.*

Proof. **Conclusion 2** is a direct result of **Lemma 2**. □

Conclusion 3. *For a pair of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of appropriate dimensions and a positive definite matrix $\mathbf{\Phi}$, when $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AB}$ and $\boldsymbol{B}^{\mathrm{H}}\mathbf{\Phi}\boldsymbol{B}$ have the same unitary EVD matrix, there exists a unitary matrix that is simultaneously the right unitary SVD matrix of $\boldsymbol{A}\mathbf{\Phi}^{-\frac{1}{2}}$ and the left unitary SVD matrix of $\mathbf{\Phi}^{\frac{1}{2}}\boldsymbol{B}$. In other words, $\mathbf{\Phi}^{-\frac{1}{2}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\mathbf{\Phi}^{-\frac{1}{2}}$ and $\mathbf{\Phi}^{\frac{1}{2}}\boldsymbol{BB}^{\mathrm{H}}\mathbf{\Phi}^{\frac{1}{2}}$ have the same unitary EVD matrix.*

Proof. **Conclusion 3** can be inferred upon replacing $\boldsymbol{B}$ and $\boldsymbol{A}$ in **Conclusion 2** by $\tilde{\boldsymbol{B}} = \mathbf{\Phi}^{\frac{1}{2}}\boldsymbol{B}$ and $\tilde{\boldsymbol{A}} = \boldsymbol{A}\mathbf{\Phi}^{-\frac{1}{2}}$, respectively. □

Conclusion 4. *For an arbitrary complex matrix $\boldsymbol{A}$, $\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{AA}^{\mathrm{H}} + \boldsymbol{I})^k\boldsymbol{A}$ and $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ have the same unitary EVD matrix. The scalar k can be an arbitrary real number.*

Proof. Upon expressing the SVD of matrix $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{U}\mathbf{\Lambda}\boldsymbol{V}^{\mathrm{H}}$, we have $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{V}\mathbf{\Lambda}^{\mathrm{H}}\mathbf{\Lambda}\boldsymbol{V}^{\mathrm{H}}$ and $\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{AA}^{\mathrm{H}} + \boldsymbol{I})^k\boldsymbol{A} = \boldsymbol{V}\mathbf{\Lambda}^{\mathrm{H}}(\mathbf{\Lambda}\mathbf{\Lambda}^{\mathrm{H}} + \boldsymbol{I})^k\mathbf{\Lambda}\boldsymbol{V}^{\mathrm{H}}$. □

Highlight 4. *The EVD is not unique for a Hermitian matrix since the eigenvalues can be arranged in different orders. For two $N \times N$ positive semidefinite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, the following matrix inequalities can be exploited to choose eigenvalues pairing [1]*

Matrix Inequ. 1:
$$\log|\boldsymbol{I} + \boldsymbol{AB}| \le \sum_{i=1}^{N} \log\left[1+\lambda_i(\boldsymbol{A})\lambda_i(\boldsymbol{B})\right], \tag{23}$$

Matrix Inequ. 2:
$$\log|\boldsymbol{A}+\boldsymbol{B}| \le \sum_{i=1}^{N} \log\left[\lambda_{N-i+1}(\boldsymbol{A})+\lambda_i(\boldsymbol{B})\right], \tag{24}$$

Matrix Inequ. 3:
$$\mathrm{Tr}(\boldsymbol{AB}) \ge \sum_{i=1}^{N} \left[\lambda_{N-i+1}(\boldsymbol{A})\lambda_i(\boldsymbol{B})\right], \tag{25}$$

Matrix Inequ. 4:
$$\mathrm{Tr}\left[(\boldsymbol{A}+\boldsymbol{B})^{-1}\right] \ge \sum_{i=1}^{N}\left[\lambda_{N-i+1}(\boldsymbol{A}) + \lambda_i(\boldsymbol{B})\right]^{-1}. \tag{26}$$

Having a symmetric matrix structure is an important characteristic that can be exploited to derive the optimal solution. In the following subsection, a number of specific matrix variable optimization examples are given to illustrate how to derive the optimal structures of the matrix variables based on symmetric matrix equations.

A. Optimization for MIMO Communications

We mainly consider a general MIMO downlink communication scenario, where the BS and the user are equipped with N_t and N_r antennas, respectively. Then the signal received at the user can be expressed as

$$\boldsymbol{y} = \boldsymbol{HFs} + \boldsymbol{n}, \tag{27}$$

where $\boldsymbol{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix, $\boldsymbol{F}$ represents the precoding matrix, $\boldsymbol{s}$ is the transmitted signal satisfying $\mathbb{E}[\boldsymbol{ss}^{\mathrm{H}}] = \boldsymbol{I}$, and $\boldsymbol{n}$ is the additive noise vector at the user. Based on (27), the capacity of the system can be expressed as

$$C = \log\left|\mathbf{\Pi} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF}\right|, \tag{28}$$

and the mean square error (MSE) can be expressed as

$$MSE = \mathrm{Tr}\big((\mathbf{\Pi} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF})^{-1}\big), \tag{29}$$

where $\mathbf{\Pi}$ satisfying $\mathbf{\Pi} = \mathbb{E}[\boldsymbol{nn}^{\mathrm{H}}]$ is the covariance matrix of the additive noise. We consider the following optimization problem relying on a combined OF under practical per-antenna power constraints.

$$\begin{aligned} \mathbf{P1}: \max_{\boldsymbol{F}} \quad & \gamma_1 \log\left|\mathbf{\Pi} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF}\right| \\ & - \gamma_2 \mathrm{Tr}\big((\mathbf{\Pi} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF})^{-1}\big) \\ \mathrm{s.t.} \quad & \mathrm{Tr}\big(\mathbf{\Omega}_n\boldsymbol{FF}^{\mathrm{H}}\big) \le P_n, \ 1 \le n \le N_t, \end{aligned} \tag{30}$$

where γ_1 and γ_2 are the weighting factors, and P_n denotes the maximum transmit power of the n-th antenna at the BS.

To obtain the optimal structure of $\boldsymbol{F}$, we firstly derive the FOEqu of the KKT conditions associated with **P1**, given by

$$\begin{aligned} \mathbf{FOEqu1}: & \ \gamma_1\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF}(\mathbf{\Pi} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF})^{-1} \\ & + \gamma_2\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF}(\mathbf{\Pi} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF})^{-2} = \mu\mathbf{\Omega}\boldsymbol{F}, \\ & \mathbf{\Omega} \triangleq \sum_{n=1}^{N_t}\frac{\lambda_n}{\mu}\mathbf{\Omega}_n, \end{aligned} \tag{31}$$

where λ_n is the Lagrange multiplier associated with the n-th weighted power constraint. Based on **FOEqu1** (31), we have the following matrix symmetric equation (MSEqu)

$$\mathbf{MSEqu1}: \ \underbrace{\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{HF}}_{\boldsymbol{A}}$$

Algorithm 1: Proposed algorithm for solving **P1**

Input : Initial parameters $\gamma_1, \gamma_2, \mathbf{\Pi}, \boldsymbol{H}, \{\mathbf{\Omega}_k\}, \{P_n\}$.

Initialize $\{\lambda_n\}$ and set $t=1$.
repeat
  Calculate $\mu = (\sum_{n=1}^{N_t} \lambda_n^{(t)} P_n)/P$.
  Derive $\mathbf{\Omega} = (\sum_{n=1}^{N_t} \lambda_n^{(t)} \mathbf{\Omega}_n)/\mu$.
  Obtain $\boldsymbol{U_H}$ and $\boldsymbol{U_\Pi}$ from (34).
  Derive $\mathbf{\Lambda}_{\boldsymbol{F}}$ by the water-filling method.
  Derive the optimal $\boldsymbol{F}$ from (33).
  Update $\lambda_n^{(t+1)} = [\lambda_n^{(t)} + \rho(\mathrm{Tr}(\mathbf{\Omega}_n \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) - P_n)]^+$ with the decreasing step size ρ.
  Set $t = t+1$.
until $|\mu(\mathrm{Tr}(\mathbf{\Omega}_n \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) - P_n)| \le \epsilon, \forall n$;

Output: The optimal precoder $\boldsymbol{F}$.

$$\times \underbrace{\left(\gamma_1(\mathbf{\Pi}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}+\gamma_2(\mathbf{\Pi}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-2}\right)}_{\boldsymbol{B}}$$
$$=\mu \boldsymbol{F}^{\mathrm{H}}\mathbf{\Omega}\boldsymbol{F}. \tag{32}$$

According to **Lemma 1**, $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\mathbf{\Omega}\boldsymbol{F}$ have the same unitary EVD matrix, and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\mathbf{\Pi}$ have the same unitary EVD matrix. Hence it can be concluded from **Conclusion 3** that the optimal solution of $\boldsymbol{F}$ for **P1** satisfies the following structure:

$$\boldsymbol{F}_{\mathrm{opt}} = \mathbf{\Omega}^{-\frac{1}{2}} \boldsymbol{U_H} \mathbf{\Lambda}_{\boldsymbol{F}} \boldsymbol{U}_{\mathbf{\Pi}}^{\mathrm{H}}, \tag{33}$$

where the unitary matrices $\boldsymbol{U_H}$ and $\boldsymbol{U_\Pi}$ are defined based on the following EVDs

$$\mathbf{\Omega}^{-\frac{1}{2}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\mathbf{\Omega}^{-\frac{1}{2}} = \boldsymbol{U_H}\mathbf{\Lambda}_{\boldsymbol{H}}\boldsymbol{U}_{\boldsymbol{H}}^{\mathrm{H}},\ \mathbf{\Pi} = \boldsymbol{U}_{\mathbf{\Pi}}\mathbf{\Lambda}_{\mathbf{\Pi}}\boldsymbol{U}_{\mathbf{\Pi}}^{\mathrm{H}}. \tag{34}$$

As there are two EVDs, an eigenvalue pairing problem naturally exists based on **Highlight 4**. Specifically, when $\gamma_1 \ge 0$ and $\gamma_2 \ge 0$, it can be concluded that based on **MSEqu1**, the eigenvalues of the two EVDs in (34) are in the reverse order. Moreover, for the general case, when $\gamma_1 \ge 0$ and $\gamma_2 \ge 0$ are not guaranteed, we can fix the eigenvalue ordering of the first EVD and then perform an exhaustive search to find the optimal eigenvalue ordering of the second EVD. The proposed algorithm based on the optimal structure is summarized in **Algorithm 1**.

Highlight 5. *It is worth noting that when $\gamma_1 = 0$, **P1** reduces to the MSE minimization problem. Whereas when $\gamma_2 = 0$, **P1** becomes the classical MIMO capacity maximization problem. In particular, when $\gamma_1 \ge 0$ and $\gamma_2 \ge 0$, **P1** is convex. Nonetheless, γ_1 and γ_2 are not limited to nonnegative values. Therefore, the optimal structure derived based on the KKT conditions is also applicable to the case, when **P1** is nonconvex or does not satisfy the matrix-monotonic property [18].*

Highlight 6. *Under the assumption of imperfect CSI, we may derive the expectation of problem **P1** with respect to the channel error, which is ultimately formulated as [16]*

$$\begin{aligned}\max_{\boldsymbol{F}} \quad & \gamma_1 \log\left|\boldsymbol{I} + \boldsymbol{R}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\right| \\ & - \gamma_2 \mathrm{Tr}\big((\boldsymbol{I} + \boldsymbol{R}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}})^{-1}\big), \\ \text{s.t.} \quad & \mathrm{Tr}\big(\mathbf{\Omega}_n \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\big) \le P_n,\ 1 \le n \le N_t, \end{aligned} \tag{35}$$

where $\widehat{\boldsymbol{H}}$ denotes the estimated channel matrix, and $\boldsymbol{R} = \mathbf{\Pi} + \mathrm{Tr}(\boldsymbol{R}_{\mathrm{T}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}})\boldsymbol{R}_{\mathrm{R}}$. $\boldsymbol{R}_{\mathrm{R}}$ and $\boldsymbol{R}_{\mathrm{T}}$ denote the receive and transmit spatial correlation matrices, respectively. Similarly, we can derive the optimal structure of $\boldsymbol{F}$ as follows

$$\boldsymbol{F}_{\mathrm{opt}} = \mathbf{\Gamma}^{-\frac{1}{2}}\boldsymbol{U_R}\mathbf{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\mathrm{DFT}}^{\mathrm{H}}, \tag{36}$$

where the positive semidefinite matrix $\mathbf{\Gamma}$ satisfies

$$\begin{aligned}\mathbf{\Gamma} = & \mu\mathbf{\Omega} + \gamma_1 \mathrm{Tr}\left(\boldsymbol{R}^{-1}\boldsymbol{R}_{\mathrm{R}} - (\boldsymbol{R} + \widehat{\boldsymbol{H}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}})^{-1}\boldsymbol{R}_{\mathrm{R}}\right)\boldsymbol{R}_{\mathrm{T}} \\ & + \gamma_2 \mathrm{Tr}\left((\boldsymbol{I} - \boldsymbol{D})\boldsymbol{D}\boldsymbol{R}^{-1}\boldsymbol{R}_{\mathrm{R}}\right)\boldsymbol{R}_{\mathrm{T}}. \\ & \text{with}\ \ \boldsymbol{D} = (\boldsymbol{I} + \boldsymbol{R}^{-1}\widehat{\boldsymbol{H}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{H}}^{\mathrm{H}})^{-1}\end{aligned} \tag{37}$$

and $\boldsymbol{U_R}$ is the unitary matrix defined by the following EVD.

$$\mathbf{\Gamma}^{-\frac{1}{2}}\widehat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{R}^{-1}\widehat{\boldsymbol{H}}\mathbf{\Gamma}^{-\frac{1}{2}} = \boldsymbol{U_R}\mathbf{\Lambda}_{\boldsymbol{R}}\boldsymbol{U}_{\boldsymbol{R}}^{\mathrm{H}}. \tag{38}$$

B. Other Specific Examples

1) Min-Max Diagonal Element of Matrix Inversion: First, we investigate the optimization problem of minimizing the maximum diagonal element of a matrix inverse under multiple weighted power constraints. The detailed mathematical formulation is elaborated as follows.

$$\begin{aligned}\textbf{P2}: \min_{\boldsymbol{F}} \max_{1\le n\le N} \quad & \left[\left(\boldsymbol{I} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\right]_{n,n} \\ \text{s.t.} \quad & \mathrm{Tr}\big(\mathbf{\Omega}_k\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\big) \le P_k,\quad 1 \le k \le K.\end{aligned} \tag{39}$$

The min-max optimization problem considered in [2] is a special case of **P2**, which aims to minimize the MSE for the worst user. The traditional complex matrix derivative operator is difficult to apply to the max-min OF. In order to overcome this difficulty, an auxiliary variable t is introduced and the optimization problem (39) is reformulated as

$$\begin{aligned}\min_{t,\boldsymbol{F}} \quad & t \\ \text{s.t.} \quad & \left[\left(\boldsymbol{I} + \boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\right]_{n,n} \le t,\ \forall n, \\ & \mathrm{Tr}\big(\mathbf{\Omega}_k\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\big) \le P_k,\quad 1 \le k \le K.\end{aligned} \tag{40}$$

The FOEqu of the KKT conditions for the optimization problem (40) is given by

$$\begin{aligned}\textbf{FOEqu2}: & \ \alpha\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1} \\ & = \mu\mathbf{\Omega}\boldsymbol{F},\ \mathbf{\Omega} = \sum_{k=1}^{K}\frac{\lambda_k}{\mu}\mathbf{\Omega}_k,\end{aligned} \tag{41}$$

where α is the common Lagrange multiplier associated with the first N constraints of the problem (40). The derivation is given in Appendix B. Note that we can prove that all the Lagrange multipliers for the first N constraints are equal. From **FOEqu2** (41), we have the following matrix symmetric equation (MSEqu)

$$\begin{aligned}\textbf{MSEqu2}: & \\ & \underbrace{\alpha\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}}_{\boldsymbol{A}}\underbrace{\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}}_{\boldsymbol{B}} \\ & = \mu\boldsymbol{F}^{\mathrm{H}}\mathbf{\Omega}\boldsymbol{F}.\end{aligned} \tag{42}$$

Based on both (42) and **Lemma 1**, it can be concluded that $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ have the same EVD unitary matrix. Then based on **Conclusion 3**, the optimal $\boldsymbol{F}$ satisfies the structure:

$$\boldsymbol{F}_{\mathrm{opt}}=\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\boldsymbol{H}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\boldsymbol{F}}^{\mathrm{H}}, \tag{43}$$

where $\boldsymbol{U}_{\boldsymbol{H}}$ is the right SVD unitary matrix of $\boldsymbol{H}\boldsymbol{\Omega}^{-\frac{1}{2}}$ and the unitary matrix $\boldsymbol{U}_{\boldsymbol{F}}$ is still unknown. Furthermore, according to the KKT conditions of (40), the following equalities must hold

$$\left[\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\right]_{n,n}=t,\ \forall n. \tag{44}$$

Consequently, the optimal $\boldsymbol{U}_{\boldsymbol{F}}$ can be chosen as a DFT matrix [23] and thus the optimal $\boldsymbol{F}$ satisfies the following structure

$$\boldsymbol{F}_{\mathrm{opt}}=\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\boldsymbol{H}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\mathrm{DFT}}^{\mathrm{H}}. \tag{45}$$

It is seen that for the min-max design, the method of KKT conditions based on symmetric matrix equations is more straightforward than that based on majorization theory [2].

2) Optimization of Matrix Inversion: For completeness, we also investigate the following optimization problem with the OF in the form of summing up the diagonal elements of matrix inversion

$$\begin{aligned}\textbf{P3}:\ \ &\max_{\boldsymbol{F}}\ \ \mathrm{Tr}\left(\left((\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}+\boldsymbol{\Psi}\right)^{-1}\right)\\ &\ \mathrm{s.t.}\ \ \ \mathrm{Tr}\big(\boldsymbol{\Omega}_k\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\big)\le P_k,\ \ 1\le k\le K.\end{aligned} \tag{46}$$

The FOEqu of the KKT conditions for **P3** and the associated MSEqu are given respectively by

$$\begin{aligned}&\textbf{FOEqu3}:\\ &\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}\big((\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}+\boldsymbol{\Psi}\big)^{-2}\\ &\times(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}=\mu\boldsymbol{\Omega}\boldsymbol{F},\end{aligned} \tag{47}$$

$$\begin{aligned}&\textbf{MSEqu3}:\ \ \underbrace{\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}}_{\boldsymbol{A}}\times\\ &\underbrace{(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}\big((\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}+\boldsymbol{\Psi}\big)^{-2}(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}}_{\boldsymbol{B}}\\ &=\mu\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}.\end{aligned} \tag{48}$$

According to **Lemma 1**, $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ have the same unitary EVD matrix. Moreover, $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{\Psi}$ have the same unitary EVD matrix. As such, the optimal $\boldsymbol{F}$ for the optimization problem **P3** of (46) satisfies the following structure

$$\boldsymbol{F}_{\mathrm{opt}}=\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\boldsymbol{H}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\boldsymbol{\Psi}}^{\mathrm{H}}, \tag{49}$$

where $\boldsymbol{U}_{\boldsymbol{H}}$ and $\boldsymbol{U}_{\boldsymbol{\Psi}}$ are the unitary matrices defined by the following EVDs

$$\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{\Omega}^{-\frac{1}{2}}=\boldsymbol{U}_{\boldsymbol{H}}\boldsymbol{\Lambda}_{\boldsymbol{H}}\boldsymbol{U}_{\boldsymbol{H}}^{\mathrm{H}},\ \boldsymbol{\Psi}=\boldsymbol{U}_{\boldsymbol{\Psi}}\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}\boldsymbol{U}_{\boldsymbol{\Psi}}^{\mathrm{H}}. \tag{50}$$

As there are two EVDs, an eigenvalue pairing problem naturally exists. According to **Highlight 4**, the eigenvalues of the two matrices given in (50) are in the same order.

3) Optimization of Combined Functions: Similarly, another optimization of combined OFs is expressed as

$$\begin{aligned}\textbf{P4}:\max_{\boldsymbol{F}}\ \ &\gamma_1\log\left|\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right|\\ &-\gamma_2\mathrm{Tr}\big(\boldsymbol{W}(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}\big)\\ \mathrm{s.t.}\ \ &\mathrm{Tr}(\boldsymbol{\Omega}_k\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}})\le P_k,\ 1\le k\le K,\end{aligned} \tag{51}$$

with weighting factors γ_1 and γ_2. Sepcifically, when $\gamma_1=0$, **P4** reduces to the weighted MSE minimization problem. Whereas when $\gamma_2=0$, **P4** becomes the classical MIMO capacity maximization problem. Similar to **P3**, γ_1 and γ_2 are not limited to nonnegative values. The FOEqu of the KKT conditions and the corresponding MSEqu are then respectively shown in (52) and (53) at the top of the next page. According to **Lemma 1**, $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ have the same unitary EVD matrix. Furthermore, $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{W}$ have the same unitary EVD matrix. Based on **Conclusion 3**, the optimal structure of $\boldsymbol{F}$ for **P4** is given by

$$\boldsymbol{F}_{\mathrm{opt}}=\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\boldsymbol{H}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\boldsymbol{W}}^{\mathrm{H}}, \tag{54}$$

where the unitary matrices $\boldsymbol{U}_{\boldsymbol{H}}$ and $\boldsymbol{U}_{\boldsymbol{W}}$ are defined by the following EVDs

$$\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{\Omega}^{-\frac{1}{2}}=\boldsymbol{U}_{\boldsymbol{H}}\boldsymbol{\Lambda}_{\boldsymbol{H}}\boldsymbol{U}_{\boldsymbol{H}}^{\mathrm{H}},\ \boldsymbol{W}=\boldsymbol{U}_{\boldsymbol{W}}\boldsymbol{\Lambda}_{\boldsymbol{W}}\boldsymbol{U}_{\boldsymbol{W}}^{\mathrm{H}}. \tag{55}$$

When $\gamma_1\ge 0$ and $\gamma_2\ge 0$, it can be concluded that the eigenvalue ordering of the two EVDs in (55) are in the same order. Furthermore, for the general case, when $\gamma_1\ge 0$ and $\gamma_2\ge 0$ are not guaranteed, again we can fix the the eigenvalue ordering of the first EVD and then perform an exhaustive search to find the optimal eigenvalue ordering of the second EVD.

4) Capacity Maximization of Dual-Hop MIMO: The capacity maximization problem for the dual-hop amplify-and-forward MIMO system is formulated as [24]

$$\begin{aligned}\textbf{P5}:\max_{\boldsymbol{F}}\ \ &\log\frac{|\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{H}_1\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I}|}{|\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I}|}\\ \mathrm{s.t.}\ \ &\mathrm{Tr}\big(\boldsymbol{F}(\boldsymbol{H}_1\boldsymbol{H}_1^{\mathrm{H}}+\sigma_1^2\boldsymbol{I})\boldsymbol{F}^{\mathrm{H}}\big)\le P,\end{aligned} \tag{56}$$

where $\boldsymbol{H}_2$ and $\boldsymbol{H}_1$ are the channel matrices in the second hop and the first hop, respectively, while $\boldsymbol{F}$ is the relay forwarding matrix. The FOEqu with respect to $\boldsymbol{F}$ is given by

$$\begin{aligned}&\textbf{FOEqu5}:\\ &\boldsymbol{H}_2^{\mathrm{H}}(\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I})^{-1}\boldsymbol{H}_2\boldsymbol{F}\\ &-\boldsymbol{H}_2^{\mathrm{H}}(\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I})^{-1}\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{\Phi}^{-1}=\mu\boldsymbol{F},\end{aligned} \tag{57}$$

where $\boldsymbol{\Phi}=\boldsymbol{H}_1\boldsymbol{H}_1^{\mathrm{H}}+\sigma_1^2\boldsymbol{I}$. Based on (57), the following MSEqu holds

$$\begin{aligned}\textbf{MSEqu5}:\ \ &\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\big(\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I}\big)^{-1}\boldsymbol{H}_2\boldsymbol{F}\\ &-\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\big(\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I}\big)^{-1}\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{\Phi}^{-1}=\mu\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}.\end{aligned} \tag{58}$$

According to **MSEqu5**, $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\big(\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I}\big)^{-1}\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}$ is a Hermitian matrix, and we conclude that $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\boldsymbol{H}_2\boldsymbol{F}$ as well as $\boldsymbol{\Phi}$ have the same unitary EVD matrix. Similarly, $\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\boldsymbol{H}_2\boldsymbol{F}$ also have the same

$$\textbf{FOEqu4}:$$
$$\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\left(\gamma_1(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}+\gamma_2(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}\boldsymbol{W}(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}\right)=\mu\boldsymbol{\Omega}\boldsymbol{F}, \tag{52}$$

$$\textbf{MSEqu4}:\ \underbrace{\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}}_{\boldsymbol{A}}$$
$$\times\underbrace{\left(\gamma_1(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}+\gamma_2(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}\boldsymbol{W}(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F})^{-1}\right)}_{\boldsymbol{B}}=\mu\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}. \tag{53}$$

unitary EVD matrix. As a result, the optimal relay forwarding matrix $\boldsymbol{F}$ satisfies the following structure

$$\boldsymbol{F}=\boldsymbol{V}_{\boldsymbol{H}_2}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\boldsymbol{H}_1}^{\mathrm{H}}, \tag{59}$$

where the unitary matrices $\boldsymbol{V}_{\boldsymbol{H}_2}$ and $\boldsymbol{U}_{\boldsymbol{H}_1}$ are defined based on the following EVDs

$$\boldsymbol{H}_1=\boldsymbol{U}_{\boldsymbol{H}_1}\boldsymbol{\Lambda}_{\boldsymbol{H}_1}\boldsymbol{V}_{\boldsymbol{H}_1}^{\mathrm{H}},\quad \boldsymbol{H}_2=\boldsymbol{U}_{\boldsymbol{H}_2}\boldsymbol{\Lambda}_{\boldsymbol{H}_2}\boldsymbol{V}_{\boldsymbol{H}_2}^{\mathrm{H}}. \tag{60}$$

It follows from **MSEqu5** that the eigenvalues of $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}$ are in the same order. Therefore, the eigenvalues of the two EVDs in (60) are in the same order.

It can be seen that using symmetric matrix equation, the KKT conditions based method is much more straightforward than the derivations given in [24].

5) Two alternative BER minimization examples: Bit error rate (BER) is an important performance metric for MIMO transceiver optimization, which reflects the reliability of data transmission. However, since an analytical expression of BER may be not readily accessible, we instead consider an alternative BER performance metric, i.e., the so-called MSE metric, for guaranteeing the reliable data transmission. This approach has been widely adopted in the existing MIMO related literature [15]. For example, two alternative BER minimization examples are given below. Firstly, in terms of the transmitted signal detection at high SNRs, the alternative BER minimization (i.e., sum MSE minimization) problem can be relaxed as

$$\begin{aligned}\textbf{P6}:\ &\min_{\boldsymbol{F}}\ \ \mathrm{Tr}\big((\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F})^{-1}\big)\\ &\ \mathrm{s.t.}\ \ \mathrm{Tr}\big(\boldsymbol{\Omega}_k\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\big)\le P_k,\ 1\le k\le K.\end{aligned} \tag{61}$$

The FOEqu of the KKT conditions for **P6** is given by

$$\textbf{FOEqu6}:\ \boldsymbol{\Pi}\boldsymbol{F}\big(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\big)^{-2}=\mu\boldsymbol{\Omega}\boldsymbol{F}, \tag{62}$$

based on which the following MSEqu is obtained

$$\textbf{MSEqu6}:\ \big(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}\big)^{-1}=\mu\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}. \tag{63}$$

It can be concluded that $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ have the same unitary EVD matrix. Based on **Conclusion 3**, the optimal $\boldsymbol{F}$ satisfies the following structure

$$\boldsymbol{F}_{\mathrm{opt}}=\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}}, \tag{64}$$

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

$$\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Pi}\boldsymbol{\Omega}^{-\frac{1}{2}}=\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{\Pi}}}\boldsymbol{U}_{\widetilde{\boldsymbol{\Pi}}}^{\mathrm{H}}, \tag{65}$$

and $\boldsymbol{U}_{\mathrm{Arb}}$ is an arbitrary unitary matrix of appropriate dimensions.

Secondly, from the perspective of channel estimation at high SNRs, the relaxed sum MSE minimization problem is expressed as [25]

$$\begin{aligned}\textbf{P7}\ \ &\min_{\boldsymbol{F}}\ \ \mathrm{Tr}\big((\boldsymbol{F}\boldsymbol{\Pi}\boldsymbol{F}^{\mathrm{H}})^{-1}\big)\\ &\ \mathrm{s.t.}\ \ \mathrm{Tr}\big(\boldsymbol{\Omega}_k\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\big)\le P_k,\ 1\le k\le K.\end{aligned} \tag{66}$$

Note that **P6** and **P7** are significantly different, because of the position of the Hermitian operation in the OF. The FOEqu of the KKT conditions for **P7** is given by

$$\textbf{FOEqu7}:\ \big(\boldsymbol{F}\boldsymbol{\Pi}\boldsymbol{F}^{\mathrm{H}}\big)^{-2}\boldsymbol{F}\boldsymbol{\Pi}=\mu\boldsymbol{\Omega}\boldsymbol{F}, \tag{67}$$

based on which the following MSEqu holds

$$\textbf{MSEqu7}:\ \big(\boldsymbol{F}\boldsymbol{\Pi}\boldsymbol{F}^{\mathrm{H}}\big)^{-1}=\mu\boldsymbol{\Omega}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}. \tag{68}$$

We conclude that $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ and $\boldsymbol{\Omega}$ have the same unitary EVD matrix, and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Pi}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ have the same unitary EVD matrix. Recalling **Conclusion 3** yields the optimal structure of $\boldsymbol{F}$

$$\boldsymbol{F}_{\mathrm{opt}}=\boldsymbol{U}_{\boldsymbol{\Omega}}\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{U}_{\boldsymbol{\Pi}}^{\mathrm{H}}, \tag{69}$$

where the unitary matrices $\boldsymbol{U}_{\boldsymbol{\Pi}}$ and $\boldsymbol{U}_{\boldsymbol{\Omega}}$ are defined using the following EVDs

$$\boldsymbol{\Pi}=\boldsymbol{U}_{\boldsymbol{\Pi}}\boldsymbol{\Lambda}_{\boldsymbol{\Pi}}\boldsymbol{U}_{\boldsymbol{\Pi}}^{\mathrm{H}},\quad \boldsymbol{\Omega}=\boldsymbol{U}_{\boldsymbol{\Omega}}\boldsymbol{\Lambda}_{\boldsymbol{\Omega}}\boldsymbol{U}_{\boldsymbol{\Omega}}^{\mathrm{H}}. \tag{70}$$

As there are two EVDs, there is an eigenvalue pairing issue according to **Highlight 4**. From **MSEqu7**, the optimal eigenvalues involved in $\boldsymbol{\Lambda}_{\boldsymbol{\Pi}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Omega}}$ in (70) are in the reverse order.

C. EVD Pairing Results

Based on the symmetric matrix equations derived for the KKT conditions based method, the optimal structure of the matrix variable can be efficiently obtained. However, since the eigenvalues of a Hermitian matrix can be arranged in different orders, the optimal matrix variable cannot be uniquely determined from the optimal structure. In order to overcome this issue, an important conclusion is given in the following.

Conclusion 5. *For the OF that is a matrix monotonic function with respect to $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ under multiple weighted power constraints, i.e., when $\mathrm{Tr}(\boldsymbol{\Omega}_k\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}})\le P_k$, if $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ have the same unitary EVD matrix, we can conclude that the eigenvalues of the matrices $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}$ and $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{F}$ are in the same order.*

IV. Positive Semidefinite Constraints

In many MIMO optimization problems, the covariance matrix $\boldsymbol{Q}$ of the transmit signal is the complex matrix variable, which is positive semidefinite [26]–[29]. In contrast to an unconstrained complex matrix variable, there are two constraints imposed on $\boldsymbol{Q}$: 1) it is conjugate symmetric, and 2) its eigenvalues are all nonnegative. Based on the definition of the complex matrix derivative [12], when the OF is a real valued function, the matrix derivative result is a Hermitian matrix of the same dimension. Motivated by this fact, we have the following equality for the corresponding complex matrix derivative operators

$$\left(\frac{\partial f(\boldsymbol{Q})}{\partial \boldsymbol{Q}}\right)^{\mathrm{T}}=\frac{\partial f(\boldsymbol{Q})}{\partial \boldsymbol{Q}^{*}}. \tag{71}$$

For example, based on the complex matrix derivative and on the fact that $\boldsymbol{Q}$ is positive semidefinite, we have the following complex matrix derivative operators

$$\frac{\partial \mathrm{Tr}(\boldsymbol{W}\boldsymbol{Q})}{\partial \boldsymbol{Q}^*} = \frac{\partial \mathrm{Tr}\big(\boldsymbol{W}\boldsymbol{Q}^{\mathrm{H}}\big)}{\partial \boldsymbol{Q}^*} = \boldsymbol{W}. \tag{72}$$

It can be seen that the matrix derivative operators of (72) are significantly different from those of (3). Here the matrix $\boldsymbol{W}$ must be a Hermitian matrix. Otherwise, this derivation definition is meaningless. In other words, there is no definition for the complex matrix derivative in this case. In the following, we consider the derivative of a quadratic function with respect to $\boldsymbol{Q}$

$$\frac{\partial \mathrm{Tr}(\boldsymbol{Q}\boldsymbol{W}\boldsymbol{Q}^{\mathrm{H}})}{\partial \boldsymbol{Q}^*} = \boldsymbol{Q}\boldsymbol{W} + \boldsymbol{W}\boldsymbol{Q}, \tag{73}$$

where $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{H}}$ and $\boldsymbol{Q} = \boldsymbol{Q}^{\mathrm{H}}$. It is plausible that the right-hand side of (73) is also a Hermitian matrix. It is worth pointing out that this result is significantly different from the traditional result derived without specific structural constraints.

A. Positive Semidefinite Matrix Variable

Moreover, as $\boldsymbol{Q}$ is a positive semidefinite matrix, the complex matrix derivative must be defined on the set of positive semidefinite matrices. Unfortunately, the existing definitions of complex matrix derivatives never consider this. In other words, there is a relaxation when performing complex matrix derivatives in the classical literature, and if the final solution is positive semidefinite, then there will be no loss. This is usually guaranteed by the mathematical formula of the OF.

1) Lagrange Multiplier Method: It is worth noting that there are certain specific structural constraints that cannot be imposed on the matrix derivative results, but they are reflected by Lagrange multipliers. Considering the positive semidefinite constraint as an example, i.e., $\boldsymbol{Q} \succeq \boldsymbol{0}$, the physical meaning of this inequality is that all the eigenvalues of $\boldsymbol{Q}$ are nonnegative. Then a matrix-valued Lagrange multiplier associated with this positive semidefinite constraint may be introduced [21]

$$\boldsymbol{\Psi}\boldsymbol{Q} = \boldsymbol{0}, \quad \boldsymbol{\Psi} \succeq \boldsymbol{0}. \tag{74}$$

In [27], the Lagrange multiplier for the positive semidefinite matrix $\boldsymbol{Q}$ with rank constraints is also given. In the following, we take a variant of **P3** as an example:

$$\begin{aligned} \mathbf{P8}: \max_{\boldsymbol{Q}} \quad & \gamma_1 \log\big|\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}}\big| - \gamma_2 \mathrm{Tr}\big((\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}})^{-1}\big) \\ \text{s.t.} \quad & \mathrm{Tr}(\boldsymbol{\Omega}_k \boldsymbol{Q}) \le P_k, \ 1 \le k \le K, \ \boldsymbol{Q} \succeq \boldsymbol{0}. \end{aligned} \tag{75}$$

The optimization problem **P8** is convex with respect to the positive semidefinite matrix variable $\boldsymbol{Q}$. It is straightforward to see that the KKT conditions of **P8** are the necessary and sufficient conditions for the optimal solution. In the following, a short discussion is given to show how to derive the optimal solution based on the KKT conditions. More details can be found in [27]. The FOEqu of the KKT conditions for problem **P8** is given by

$$\begin{aligned} \mathbf{FOEqu8}: \boldsymbol{H}^{\mathrm{H}} \Big(\gamma_1 \big(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}}\big)^{-1} + \gamma_2 \big(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}}\big)^{-2}\Big) \boldsymbol{H} \\ = \boldsymbol{\Omega} - \boldsymbol{\Psi}, \end{aligned} \tag{76}$$

based on which the following MSEqu holds

$$\begin{aligned} \mathbf{MSEqu8}: \boldsymbol{Q}^{\frac{1}{2}} \boldsymbol{H}^{\mathrm{H}} \big(\gamma_1 \big(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}}\big)^{-1} \\ + \gamma_2 \big(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}}\big)^{-2}\big) \boldsymbol{H}\boldsymbol{Q}^{\frac{1}{2}} = \boldsymbol{Q}^{\frac{1}{2}} \boldsymbol{\Omega} \boldsymbol{Q}^{\frac{1}{2}}. \end{aligned} \tag{77}$$

It is plausible that the two matrices $\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{Q}^{\frac{1}{2}}$ and $\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{Q}^{\frac{1}{2}}$ have the same unitary EVD matrix. Then based on **Conclusion 3**, the optimal structure of $\boldsymbol{Q}$ can be obtained as

$$\boldsymbol{Q}_{\mathrm{opt}} = \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{U}_{\boldsymbol{H}} \boldsymbol{\Lambda}_{\boldsymbol{Q}} \boldsymbol{U}_{\boldsymbol{H}}^{\mathrm{H}} \boldsymbol{\Omega}^{-\frac{1}{2}}, \tag{78}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{Q}}$ is a diagonal matrix and the unitary matrix $\boldsymbol{U}_{\boldsymbol{H}}$ is derived from the following EVD

$$\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \boldsymbol{\Omega}^{-\frac{1}{2}} = \boldsymbol{U}_{\boldsymbol{H}} \boldsymbol{\Lambda}_{\boldsymbol{H}} \boldsymbol{U}_{\boldsymbol{H}}^{\mathrm{H}}, \tag{79}$$

where the eigenvalues involved in $\boldsymbol{\Lambda}_{\boldsymbol{H}}$ are arranged in a non-increasing order.

Similar to the sum MSE minimization, in the following, we consider another general optimization problem associated with the OF in the form of matrix inversion

$$\begin{aligned} \mathbf{P9} \quad \min_{\boldsymbol{Q}} \quad & \mathrm{Tr}\big[(\boldsymbol{I} + \boldsymbol{N}\boldsymbol{Q}\boldsymbol{N}^{\mathrm{H}})^{-K}\big] \\ \text{s.t.} \quad & \mathrm{Tr}(\boldsymbol{\Omega}_n \boldsymbol{Q}) \le P_n, \ 1 \le n \le N, \ \boldsymbol{Q} \succeq \boldsymbol{0}. \end{aligned} \tag{80}$$

The FOEqu of the KKT conditions for **P9** and the associated MSEqu are given respectively by

$$\mathbf{FOEqu9}: \boldsymbol{N}^{\mathrm{H}} (\boldsymbol{I} + \boldsymbol{N}\boldsymbol{Q}\boldsymbol{N}^{\mathrm{H}})^{-K-1} \boldsymbol{N} = \mu \boldsymbol{\Omega} - \boldsymbol{\Psi}, \tag{81}$$

$$\mathbf{MSEqu9}: \boldsymbol{Q}^{\frac{1}{2}} \boldsymbol{N}^{\mathrm{H}} (\boldsymbol{I} + \boldsymbol{N}\boldsymbol{Q}\boldsymbol{N}^{\mathrm{H}})^{-K-1} \boldsymbol{N} \boldsymbol{Q}^{\frac{1}{2}} = \mu \boldsymbol{Q}^{\frac{1}{2}} \boldsymbol{\Omega} \boldsymbol{Q}^{\frac{1}{2}}. \tag{82}$$

It may be readily shown that $\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{Q}^{\frac{1}{2}}$ and $\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{N}^{\mathrm{H}}\boldsymbol{N}\boldsymbol{Q}^{\frac{1}{2}}$ have the same unitary EVD matrix. Then based on **Conclusion 3**, the optimal structure of $\boldsymbol{Q}$ is derived as

$$\boldsymbol{Q}_{\mathrm{opt}} = \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{U}_{\boldsymbol{N}} \boldsymbol{\Lambda}_{\boldsymbol{Q}} \boldsymbol{U}_{\boldsymbol{N}}^{\mathrm{H}} \boldsymbol{\Omega}^{-\frac{1}{2}}, \tag{83}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{Q}}$ is a diagonal matrix and the unitary matrix $\boldsymbol{U}_{\boldsymbol{N}}$ is defined in the following EVD

$$\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{N}^{\mathrm{H}} \boldsymbol{N} \boldsymbol{\Omega}^{-\frac{1}{2}} = \boldsymbol{U}_{\boldsymbol{N}} \boldsymbol{\Lambda}_{\boldsymbol{N}} \boldsymbol{U}_{\boldsymbol{N}}^{\mathrm{H}}, \tag{84}$$

where the eigenvalues are sorted in a non-increasing order.

2) Relaxation-Based Methods: Another popular approach to deal with the positive semidefinite constraint is utilizing the relaxation-based method. For example, a MIMO training optimization problem associated with the least square (LS) estimator is formulated as [25]

$$\begin{aligned} \mathbf{P10}: \quad \min_{\boldsymbol{Q}} \quad & \mathrm{Tr}\big(\boldsymbol{W}\boldsymbol{Q}^{-1}\big) \\ \text{s.t.} \quad & \mathrm{Tr}(\boldsymbol{Q}) \le P, \ \boldsymbol{Q} \succ \boldsymbol{0}, \end{aligned} \tag{85}$$

where both the matrix variable $\boldsymbol{Q}$ and the weighting matrix $\boldsymbol{W}$ are positive definite. Using the relaxation-based method, the positive definite constraint is firstly relaxed and thus the original training optimization problem is simplified as

$$\begin{aligned} \min_{\boldsymbol{Q}} \quad & \mathrm{Tr}\big(\boldsymbol{W}\boldsymbol{Q}^{-1}\big) \\ \text{s.t.} \quad & \mathrm{Tr}(\boldsymbol{Q}) \le P. \end{aligned} \tag{86}$$

Accordingly, the KKT conditions of the problem (86) are given by

$$\boldsymbol{Q}^{-1}\boldsymbol{W}\boldsymbol{Q}^{-1}=\lambda\boldsymbol{I},\ \lambda\geq 0,\ \lambda(\mathrm{Tr}(\boldsymbol{Q})-P)=0,\ \mathrm{Tr}(\boldsymbol{Q})\leq P. \tag{87}$$

The optimal solution satisfying the above KKT conditions is unique, and is given by [25]

$$\boldsymbol{Q}=\frac{P}{\mathrm{Tr}(\boldsymbol{W}^{\frac{1}{2}})}\boldsymbol{W}^{\frac{1}{2}}. \tag{88}$$

The solution (88) is found to be positive definite, and thus it is also the optimal solution of the original optimization problem (85). This relaxation-based method enjoys simplicity, but a double check is needed for the derived solution. Moreover, the main weakness of this kind of algorithm is that there are strict requirements on the optimization problems considered.

Highlight 7. *The complex matrix derivative for a real-valued function with respect to a positive semidefinite matrix variable must be a Hermitian matrix. The semi-positivity is guaranteed by the fact that the optimal OF value occurs at the set of the positive semidefinite matrix variables.*

B. Variable Transformation for Single-matrix Variable

In order to overcome the challenges imposed by specific structural constraints, an effective technique is to transform the original complex matrix variables. For convex optimization problems, it is possible to derive the optimal solutions purely based on the KKT conditions. However, when the problem considered is nonconvex, some variable transformation techniques can be adopted to simplify the derivation of the optimal structures of matrix variables. As such, the matrix derivative operation involved can be substantially simplified.

In order to address the positive semidefinite constraint, such as $\boldsymbol{Q}\succeq\boldsymbol{0}$, a feasible strategy is to exploit the equality

$$\boldsymbol{Q}=\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}, \tag{89}$$

where $\widetilde{\boldsymbol{F}}$ is a general matrix without structural constraints. The benefit of this strategy is twofold. On the one hand, no structural constraint is imposed on $\widetilde{\boldsymbol{F}}$ and thus the complex matrix derivative with respect to $\widetilde{\boldsymbol{F}}$ is much easier to derive. On the other hand, the rank constraint can be taken into account by adjusting the number of columns in $\widetilde{\boldsymbol{F}}$. Unfortunately, this strategy destroys convexity, where the KKT conditions constitute necessary but not sufficient conditions for the optimal solution. However in many non-convex optimization problems, all the solutions that satisfy the KKT conditions have a uniform structure, which is thus the optimal structure of the matrix variable. It should be emphasized that having optimal structures are of great importance, which can simplify the original optimization problem.

For example, using the matrix variable transformation (89), **P9** can be transferred into

$$\begin{aligned}\textbf{P11}:\ &\min_{\widetilde{\boldsymbol{F}}}\ \ \mathrm{Tr}\big[(\boldsymbol{I}+\boldsymbol{N}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{N}^{\mathrm{H}})^{-K}\big]\\ &\ \mathrm{s.t.}\ \ \ \mathrm{Tr}\big(\boldsymbol{\Omega}_n\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\big)\leq P_n,\ 1\leq n\leq N.\end{aligned} \tag{90}$$

The FOEqu of the KKT conditions of the problem **P11** is given by

$$\textbf{FOEqu11}:\ K\boldsymbol{N}^{\mathrm{H}}\big(\boldsymbol{I}+\boldsymbol{N}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{N}^{\mathrm{H}}\big)^{-K-1}\boldsymbol{N}\widetilde{\boldsymbol{F}}=\mu\boldsymbol{\Omega}\widetilde{\boldsymbol{F}}, \tag{91}$$

based on which the following MSEqu holds

$$\begin{aligned}\textbf{MSEqu 11:}\ &K\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{N}^{\mathrm{H}}\big(\boldsymbol{I}+\boldsymbol{N}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{N}^{\mathrm{H}}\big)^{-K-1}\boldsymbol{N}\widetilde{\boldsymbol{F}}\\ &=\mu\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{\Omega}\widetilde{\boldsymbol{F}}.\end{aligned} \tag{92}$$

It is plausible that $\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{\Omega}\widetilde{\boldsymbol{F}}$ and $\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{N}^{\mathrm{H}}\boldsymbol{N}\widetilde{\boldsymbol{F}}$ have the same unitary EVD matrix. Then based on **Conclusion 3**, the optimal structure of $\widetilde{\boldsymbol{F}}$ can be derived as

$$\widetilde{\boldsymbol{F}}_{\mathrm{opt}}=\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{U}_{\boldsymbol{N}}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}\boldsymbol{V}_{\mathrm{Arb}}^{\mathrm{H}}, \tag{93}$$

where $\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}}$ is a diagonal matrix and the unitary matrix $\boldsymbol{U}_{\boldsymbol{N}}$ is defined based on the following EVD with the eigenvalues arranged in a non-increasing order

$$\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{N}^{\mathrm{H}}\boldsymbol{N}\boldsymbol{\Omega}^{-\frac{1}{2}}=\boldsymbol{U}_{\boldsymbol{N}}\boldsymbol{\Lambda}_{\boldsymbol{N}}\boldsymbol{U}_{\boldsymbol{N}}^{\mathrm{H}}. \tag{94}$$

Likewise, based on $\boldsymbol{Q}=\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}$, the problem **P10** can be transferred into the following one

$$\begin{aligned}\textbf{P12}:\ &\min_{\widetilde{\boldsymbol{F}}}\ \ \mathrm{Tr}\big(\boldsymbol{W}\big(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\big)^{-1}\big)\\ &\ \mathrm{s.t.}\ \ \ \mathrm{Tr}\big(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\big)\leq P.\end{aligned} \tag{95}$$

The FOEqu of the KKT conditions for the optimization **P12** is given by

$$\textbf{FOEqu12}:\ \big(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\big)^{-1}\boldsymbol{W}\big(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\big)^{-1}\widetilde{\boldsymbol{F}}=\lambda\widetilde{\boldsymbol{F}}, \tag{96}$$

where λ is the Lagrange multiplier for the power constraint. It follows from **FOEqu12** that

$$\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}=\sqrt{\frac{1}{\lambda}}\boldsymbol{W}^{\frac{1}{2}}. \tag{97}$$

Clearly, $\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}$ is the same as the optimal $\boldsymbol{Q}$ derived in [25]. This example shows that leveraging the variable transformation technique beneficially simplifies the derivation of the optimal solution.

C. Variable Transformation for Multi-matrix Variables

1) Multi-user MIMO Capacity Maximization: The capacity maximization problem of uplink multi-user MIMO communications can be formulated as [21]

$$\begin{aligned}\textbf{P13}:\ &\max_{\{\boldsymbol{Q}_n\}_{n=1}^N}\ \ \log\left|\boldsymbol{\Pi}+\sum_{n=1}^{N}\boldsymbol{H}_n\boldsymbol{Q}_n\boldsymbol{H}_n^{\mathrm{H}}\right|\\ &\ \ \mathrm{s.t.}\ \ \mathrm{Tr}\left(\sum_{n=1}^{N}\boldsymbol{\Omega}_{n,m,k}\boldsymbol{Q}_n\right)\leq P_{m,k},\ \forall m,k,\\ &\ \ \ \ \ \ \ \ \boldsymbol{Q}_n\succeq\boldsymbol{0},\ 1\leq n\leq N,\end{aligned} \tag{98}$$

which can be equivalently transferred into the following problem by utilizing $\boldsymbol{Q}_n=\widetilde{\boldsymbol{F}}_n\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}$

$$\textbf{P14}:\ \max_{\{\widetilde{\boldsymbol{F}}_n\}_{n=1}^N}\ \ \log\left|\boldsymbol{\Pi}+\sum_{n=1}^{N}\boldsymbol{H}_n\widetilde{\boldsymbol{F}}_n\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\boldsymbol{H}_n^{\mathrm{H}}\right|$$

$$\text{s.t.}\ \ \mathrm{Tr}\left(\sum_{n=1}^{N}\boldsymbol{\Omega}_{n,m,k}\widetilde{\boldsymbol{F}}_n\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\right)\leq P_{m,k},\ \forall m,k, \quad (99)$$

where $\boldsymbol{H}_n$ denotes the channel between the n-th user and the base station. The FOEqu of the KKT conditions for the optimization **P14** is then written as

FOEqu14 :

$$\boldsymbol{H}_n^{\mathrm{H}}\left(\boldsymbol{\Pi}+\sum_{n=1}^{N}\boldsymbol{H}_n\widetilde{\boldsymbol{F}}_n\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\boldsymbol{H}_n^{\mathrm{H}}\right)^{-1}\boldsymbol{H}_n\widetilde{\boldsymbol{F}}_n=\mu_n\boldsymbol{\Omega}_n\widetilde{\boldsymbol{F}}_n$$
$$\boldsymbol{\Omega}_n=\sum_k\sum_m\frac{\lambda_{m,k}}{\mu_n}\boldsymbol{\Omega}_{n,m,k},\ 1\leq n\leq N. \quad (100)$$

The scalars $\lambda_{m,k}$ are the Lagrange multipliers associated with the power constraints in **P14**. Then we have the following MSEqu.

MSEqu14 :

$$\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\boldsymbol{H}_n^{\mathrm{H}}\boldsymbol{\Sigma}_n^{-\frac{1}{2}}\left(\boldsymbol{I}+\boldsymbol{\Sigma}_n^{-\frac{1}{2}}\boldsymbol{H}_n\widetilde{\boldsymbol{F}}_n\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\boldsymbol{H}_n^{\mathrm{H}}\boldsymbol{\Sigma}_n^{-\frac{1}{2}}\right)^{-1}\boldsymbol{\Sigma}_n^{-\frac{1}{2}}\boldsymbol{H}_n\widetilde{\boldsymbol{F}}_n$$
$$=\mu_n\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\boldsymbol{\Omega}_n\widetilde{\boldsymbol{F}}_n,$$
$$\boldsymbol{\Sigma}_n=\sum_{j\neq n}\boldsymbol{H}_j\widetilde{\boldsymbol{F}}_j\widetilde{\boldsymbol{F}}_j^{\mathrm{H}}\boldsymbol{H}_j^{\mathrm{H}}+\boldsymbol{\Pi},\ 1\leq n\leq N. \quad (101)$$

From **MSEqu14**, we conclude that $\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\boldsymbol{H}_n^{\mathrm{H}}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{H}_n^{\mathrm{H}}\widetilde{\boldsymbol{F}}_n$ and $\widetilde{\boldsymbol{F}}_n^{\mathrm{H}}\boldsymbol{\Omega}_n\widetilde{\boldsymbol{F}}_n$ have the same unitary EVD matrix. By recalling **Conclusion 3**, the optimal $\widetilde{\boldsymbol{F}}_{\mathrm{opt},n}$ satisfies the following structure [18]

$$\widetilde{\boldsymbol{F}}_{\mathrm{opt},n}=\boldsymbol{\Omega}_n^{-\frac{1}{2}}\boldsymbol{U}_{\boldsymbol{H}_n}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}_n}\boldsymbol{V}_{\mathrm{Arb},n}^{\mathrm{H}},\ \forall n, \quad (102)$$

where the unitary matrix $\boldsymbol{U}_{\boldsymbol{H}_n}$ is defined according to the following EVD

$$\boldsymbol{\Omega}_n^{-\frac{1}{2}}\boldsymbol{H}_n^{\mathrm{H}}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{H}_n\boldsymbol{\Omega}_n^{-\frac{1}{2}}=\boldsymbol{U}_{\boldsymbol{H}_n}\boldsymbol{\Lambda}_{\boldsymbol{H}_n}\boldsymbol{U}_{\boldsymbol{H}_n}^{\mathrm{H}}. \quad (103)$$

Clearly, using the variable transformation technique, the analytical structures of the optimal solutions for the more complex multi-variable optimization are still available.

2) Dual-hop MIMO Optimization: Next, we investigate the transmit precoder optimization of a dual-hop amplify-and-forward MIMO relaying system having two matrix variables, i.e., the transmit covariance matrix $\boldsymbol{Q}_1$ at the source and the forwarding matrix $\boldsymbol{F}_2$ at the relay [24]. Specifically, using $\boldsymbol{Q}_1=\widetilde{\boldsymbol{F}}_1\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}$, the resultant matrix variable optimization problem is expressed as problem **P15**, shown at the top of the next page. In **P15**, the precoder at the source and the forwarding matrix at the relay are jointly optimized. Observe that **P15** can be obtained by replacing $\boldsymbol{H}_1$ and $\boldsymbol{F}$ in the problem **P5** with $\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1$ and $\boldsymbol{F}_2$, respectively. Based on the results for **P5**, we directly infer that $\boldsymbol{F}_2\boldsymbol{F}_2^{\mathrm{H}}$ and $\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\boldsymbol{H}_2\boldsymbol{F}_2$ have the same unitary EVD matrix, and $\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2$ and $\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}\boldsymbol{H}_1^{\mathrm{H}}$ also have the same unitary EVD matrix. In addition, the FOEqu of the KKT condition for **P15** with respect to $\widetilde{\boldsymbol{F}}_1$ is given by

$$\textbf{FOEqu15}:\ \boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1\big(\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1+\boldsymbol{I}\big)^{-1}$$
$$=\lambda_2\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1+\lambda_2\widetilde{\boldsymbol{F}}_1,$$
$$\boldsymbol{\Sigma}=\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}(\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}_2\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}+\sigma_2^2\boldsymbol{I})^{-1}\boldsymbol{H}_2\boldsymbol{F}_2. \quad (105)$$

From (105), the following MSEqu holds

$$\textbf{MSEqu15}:\ \widetilde{\boldsymbol{F}}_1^{\mathrm{H}}\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1\big(\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1+\boldsymbol{I}\big)^{-1}$$
$$=\lambda_2\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1+\lambda_2\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}\widetilde{\boldsymbol{F}}_1. \quad (106)$$

It can then be concluded that $\widetilde{\boldsymbol{F}}_1\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}$ and $\widetilde{\boldsymbol{F}}_1^{\mathrm{H}}\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1$ have the same unitary EVD matrix. Finally, the optimal structures of $\widetilde{\boldsymbol{F}}_1$ and $\boldsymbol{F}_2$ are derived as

$$\widetilde{\boldsymbol{F}}_1=\boldsymbol{V}_{\boldsymbol{H}_1}\boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}_1}\boldsymbol{U}_{\mathrm{Arb}}^{\mathrm{H}},\ \boldsymbol{F}_2=\boldsymbol{V}_{\boldsymbol{H}_2}\boldsymbol{\Lambda}_{\boldsymbol{F}_2}\boldsymbol{U}_{\boldsymbol{H}_2}^{\mathrm{H}}. \quad (107)$$

D. Discussions and Results

Based on the aforementioned results, the most important advantage of using the matrix variable $\boldsymbol{Q}$ ($\boldsymbol{Q}\succeq\boldsymbol{0}$) as the optimization variable is that the convexity of the optimization problem considered is guaranteed. Therefore, the optimal solutions of the original optimization problems can be directly derived from the KKT conditions. For example, the popular water-filling solutions are available for MIMO communications from the KKT conditions of **P13** [27], but the complex matrix derivative with respect to $\boldsymbol{Q}$ is complicated, making it hard to derive the corresponding KKT conditions.

On the other hand, using the transformed matrix variable $\widetilde{\boldsymbol{F}}$ ($\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}=\boldsymbol{Q}$) as the optimization variable, the corresponding optimization problem is no longer convex and the corresponding KKT conditions are only necessary conditions for the optimal solution. From a theoretical viewpoint, the KKT condition based methods suffer from both "turning off" effects and "permutation ambiguity" effects [30]. For example, based on the KKT conditions of **P15**, the water-filling solution cannot be derived directly. However, the complex matrix derivative with respect to $\widetilde{\boldsymbol{F}}$ becomes simple, making the inference of the optimal structure from the KKT conditions simpler. Moreover, in many non-convex optimization problems, all the solutions that satisfy the KKT conditions have a uniform structure, which is thus the optimal structure of the matrix variable considered. Therefore, based on the derived optimal structures, the original optimization problems can be beneficially simplified. It can be concluded that the benefits of the variable transformation often outweigh its drawbacks.

V. Numerical Results

In this section, we present numerical evidence to support our conclusions. Specifically, we consider a point-to-point MIMO system, where both the transmitter and the receiver are equipped with 6 antennas. The MIMO channel is considered to obey Rayleigh fading, denoted by $\boldsymbol{H}$ or $\boldsymbol{N}$. Unless otherwise stated, $\boldsymbol{F}$ and $\boldsymbol{Q}$ represent the transmit precoding matrix and the transmit covariance matrix, respectively. The maximum per-antenna power P_k is assumed to be the same for all the users and the power weighting matrix $\boldsymbol{\Omega}_k$ is set to a diagonal matrix whose k-th diagonal element is one and all the other elements are zero. All the simulation results are obtained by averaging over 100 random channel realizations.

First we consider the maximization problem **P3**, where we set $\gamma_1=\gamma_2=1$ and $\boldsymbol{\Pi}=\boldsymbol{I}$. We compare the performance of three solutions, namely, 1) Our derived **optimal** solution

$$\textbf{P15}: \max_{\boldsymbol{F}_2,\widetilde{\boldsymbol{F}}_1} \log \frac{\left|\boldsymbol{H}_2\boldsymbol{F}_2\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1\widetilde{\boldsymbol{F}}_1^{\rm H}\boldsymbol{H}_1^{\rm H}\boldsymbol{F}_2^{\rm H}\boldsymbol{H}_2^{\rm H}+\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}_2\boldsymbol{F}_2^{\rm H}\boldsymbol{H}_2^{\rm H}+\sigma_2^2\boldsymbol{I}\right|}{\left|\sigma_1^2\boldsymbol{H}_2\boldsymbol{F}_2\boldsymbol{F}_2^{\rm H}\boldsymbol{H}_2^{\rm H}+\sigma_2^2\boldsymbol{I}\right|}$$
$$\text{s.t. } \operatorname{Tr}\big(\boldsymbol{F}_2(\boldsymbol{H}_1\widetilde{\boldsymbol{F}}_1\widetilde{\boldsymbol{F}}_1^{\rm H}\boldsymbol{H}_1^{\rm H}+\sigma_1^2\boldsymbol{I})\boldsymbol{F}_2^{\rm H}\big)\le P_2,\ \operatorname{Tr}\big(\widetilde{\boldsymbol{F}}_1\widetilde{\boldsymbol{F}}_1^{\rm H}\big)\le P_1. \tag{104}$$

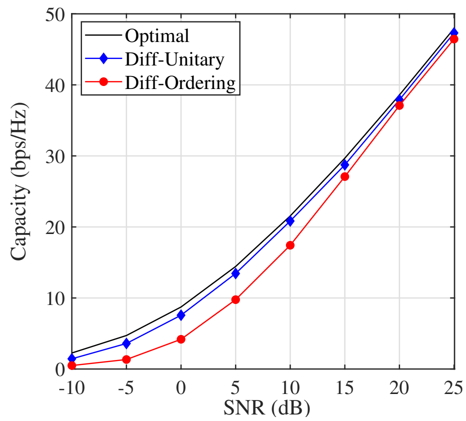


Fig. 1. The achievable performance as the functions of SNR obtained by three different solutions for the optimization problem **P3** under the weighted power constraints.

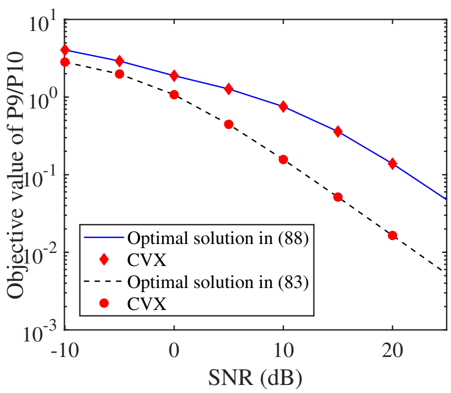


Fig. 2. Comparison between the optimal solution (83) using the Lagrange multiplier method and the numerical solution using CVX toolbox for **P9** as well as between the optimal solution (88) using the relaxation-based method and the numerical solution using CVX toolbox for **P10**.

to the optimal structure (33) in which the unitary matrices $\boldsymbol{U}_{\boldsymbol{H}}$ and $\boldsymbol{U}_{\boldsymbol{\Pi}}$ are defined in (34) and the diagonal elements of $\boldsymbol{\Lambda}_{\boldsymbol{H}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Pi}}$ in (34) are sorted in the reverse order; 2) **Diff-Unitary**: The unitary matrices $\boldsymbol{U}_{\boldsymbol{H}}$ and $\boldsymbol{U}_{\boldsymbol{\Pi}}$ in (33) are chosen as random unitary matrices; and 3) **Diff-Ordering**: The diagonal elements of $\boldsymbol{\Lambda}_{\boldsymbol{H}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Pi}}$ in (34) are sorted in the same order. Fig. 1 shows the OF values achieved by the three schemes as the functions of SNR. It can be seen from Fig. 1 that our **optimal** solution achieves the best performance, which demonstrates both the optimality of the derived unitary matrices $\boldsymbol{U}_{\boldsymbol{H}}$ and $\boldsymbol{U}_{\boldsymbol{\Pi}}$ and the fact that the eigenvalues involved in $\boldsymbol{\Lambda}_{\boldsymbol{H}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Pi}}$ must be in the reverse order.

Next we consider the general optimization problem **P9** and the LS estimation problem **P10** under the positive semidefinite matrix constraint. The positive definite weighting matrix $\boldsymbol{W}$ is set to $\boldsymbol{I}$. Notes that when we set $K=1$, **P9** actually becomes MSE minimization. It is worth noting that both the optimization problems are convex, and the globally optimal solutions can be derived numerically using the CVX toolbox [31]. Recall from Subsection IV-A that the Lagrange

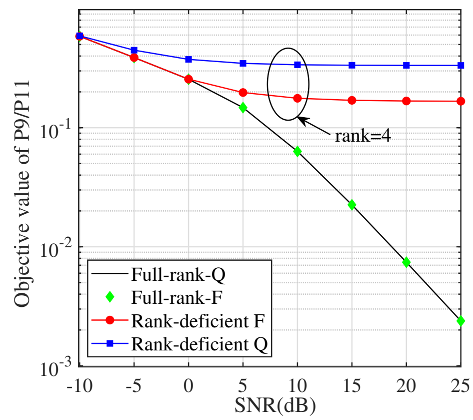


Fig. 3. Comparison of the objective values of problem **P9/P11** under different rank constraints of $\{\widetilde{\boldsymbol{F}},\boldsymbol{Q}\}$ for the single-variable case.

multiplier method and the relaxation-based method can be utilized to solve the problems **P9** and **P10**, respectively, and the corresponding optimal structures of $\boldsymbol{Q}$ are obtained as (83) and (88), respectively. Fig. 2 confirms that the two proposed optimal structures given in (83) and (88) are capable of achieving the same performance as their corresponding numerical CVX solutions for the optimization problems **P9** and **P10**, respectively. This verifies the global optimality of both these obtained structures.

As discussed in Subsection IV-B, using the transformation of the positive semidefinite matrix variable $\boldsymbol{Q}=\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\rm H}$ of (89), the minimization problem **P9** is converted to the minimization problem **P11**, where the rank constraint of $\boldsymbol{Q}$ is taken into account by adjusting the number of columns in $\widetilde{\boldsymbol{F}}$. Fig. 3 plots the achievable OF values of the problems **P9** and **P11** under different rank constraints. As expected, when both the positive semidefinite matrix $\boldsymbol{Q}$ and the transformed matrix variable $\widetilde{\boldsymbol{F}}$ are full-rank, the same optimal performances are achieved by the optimal structures of $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{F}}$. However, it follows from Fig. 3 that for a rank-deficient $\boldsymbol{Q}$, there is a significant performance gap between the optimal OF values achieved by the optimal structures of $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{F}}$. More specifically, the rank-deficient $\boldsymbol{F}$ attains better performance than the rank-deficient $\boldsymbol{Q}$ of the same rank. This is because the optimal solution $\boldsymbol{Q}$ obtained by directly solving the problem **P9** may not satisfy the rank constraint. Therefore, a feasible $\boldsymbol{Q}$ can only be derived using a relaxation method, such as the Gaussian randomization method, which causes a further degradation of the system performance.

Similarly, using the transformation of the positive semidefinite matrix variables $\boldsymbol{Q}_n=\widetilde{\boldsymbol{F}}_n\widetilde{\boldsymbol{F}}_n^{\rm H}$, the maximization problem **P13** is converted to the maximization problem **P14**, where the rank constraints of $\boldsymbol{Q}_n$ are taken into account by adjusting the number of columns in $\boldsymbol{F}_n$. Fig. 4 plots the achievable OF values of the maximization problems **P13** and **P14** under different rank constraints. As expected, when the positive semidefinite matrices $\boldsymbol{Q}_n$ are of full-rank, the optimal structures of $\boldsymbol{Q}_n$ achieve the same optimal performance as that of the full-

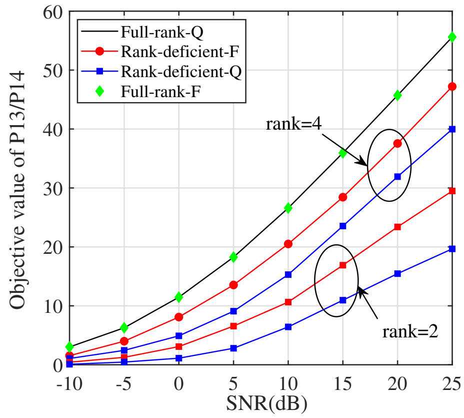


Fig. 4. Comparison of the objective values of problem **P13/P14** under different rank constraints of $\{\widetilde{\boldsymbol{F}}_n, \boldsymbol{Q}_n\}$ for the multi-variable case.

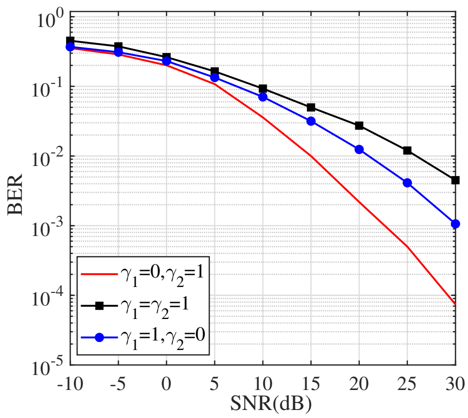


Fig. 5. Comparison of the BER performance achieved by solving problem **P3** with different parameter values.

rank $\boldsymbol{F}_n$. However, for rank-deficient $\boldsymbol{Q}_n$, there is a significant performance gap between the optimal OF values achieved by $\boldsymbol{F}_n$ and $\boldsymbol{Q}_n$, which also becomes larger when lower-rank $\boldsymbol{Q}_n$ are considered. Specifically, the rank-deficient $\boldsymbol{F}_n$ attain better performance than the rank-deficient $\boldsymbol{Q}_n$ of the same rank. This is again because the optimal solutions $\boldsymbol{Q}_n$ obtained by directly solving the maximization problem **P13** may not satisfy the rank constraint. Therefore, a feasible $\boldsymbol{Q}_n$ can only be derived using a relaxation method such as the Gaussian randomization method, which causes a further degradation of the system performance.

Finally, Fig. 5 shows the BER performance achieved by solving problem **P1** with different parameter values, i.e., 1) $\gamma_1 = 0$ and $\gamma_2 = 1$; 2) $\gamma_1 = 1$ and $\gamma_2 = 0$ and 3) $\gamma_1 = 1$ and $\gamma_2 = 1$, which corresponds to the MSE minimization, the capacity maximization and the hybrid optimization, respectively. It is clear from Fig. 5 that the optimal solution for case 1) achieves the best BER performance among the three considered cases, which demonstrates that the MSE is indeed closely related to the BER. Moreover, we observe that cases 1) and 2) have almost identical BER performance at low SNRs, since the OFs in these two cases degenerate to the same form. Under the assumption of Gaussian distributed received signals, we recall the OF curve associated with the optimal solution in (83) in Fig. 2, and find that the MSE-trend and BER-trend are similar, which also verifies the close relationship between MSE and BER.

VI. Summary and Conclusions

In this paper, we have presented a comprehensive framework of complex matrix derivatives, based on which the KKT conditions of the matrix variable optimization problems considered can be derived directly. Our contribution has been three-fold. In order to facilitate the theoretical analysis, firstly some fundamental conclusions have been presented for complex matrix derivatives, which represent the boundary conditions for the applications of complex matrix derivatives. Secondly, the symmetric properties involved in complex matrix derivatives and the corresponding KKT conditions have been investigated in depth. In addition, an important matrix equation, referred to as symmetric matrix equation, has been proposed in this paper. Using symmetric matrix equations, the optimal structures of matrix variables can be derived, based on which the matrix-variable optimization is substantially simplified. Moreover, a number of specific matrix-variable optimization problems have been discussed in detail. Thirdly, considering positive semidefinite structural constraints imposed on matrix variables, a useful variable transformation technique has been discussed in depth, which can be utilized for simplifying the KKT conditions and thus for deriving the optimal structures more easily. In a nutshell, we have improved the KKT conditions based methods of matrix variable optimization beyond simply trying to derive the optimal solutions purely based on the KKT conditions.

Appendix A
Proof of Lemma 2

Assume that a unitary matrix $\boldsymbol{U}$ is the EVD matrix of both $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{A}$ and $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$. Therefore, it is straightforward to show that $\boldsymbol{A}$ has the following SVD

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{U}^{\mathrm{H}}. \tag{108}$$

Based on this and together with the fact that $\boldsymbol{U}$ is the EVD unitary matrix of $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{A}$ the following two matrix equalities hold

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}_{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Q}\boldsymbol{\Lambda}_{\boldsymbol{A}} = \boldsymbol{\Lambda}, \tag{109}$$

where $\boldsymbol{\Lambda}$ is the diagonal EVD matrix of $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{A}$. It is plausible that in the case that the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{A}}$ is a square full rank matrix, the proof can be completed directly. In general, however, some diagonal elements of $\boldsymbol{\Lambda}_{\boldsymbol{A}}$ may be zero. To accommodate this general case, we define an index set $\mathcal{C}$ as follows

$$[\boldsymbol{\Lambda}_{\boldsymbol{A}}]_{n,n} \neq 0, \ n \in \mathcal{C}. \tag{110}$$

Upon defining $\boldsymbol{Q} = [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_N]$ and $[\boldsymbol{\Lambda}_{\boldsymbol{A}}]_{n,n} = f_n$, the second equality in (109) is equivalent to

$$[\boldsymbol{\Lambda}_{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Q}\boldsymbol{\Lambda}_{\boldsymbol{A}}]_{i,j} = f_i f_j \boldsymbol{q}_i^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{q}_j. \tag{111}$$

Based on (111) as well as constructing $\widetilde{\boldsymbol{Q}} = [\boldsymbol{q}_{(1)}, \cdots, \boldsymbol{q}_{(N\{\mathcal{C}\})}]$ and $\widetilde{\boldsymbol{\Lambda}}_{\boldsymbol{A}} = \mathrm{diag}\{f_{(1)}, \cdots, f_{(N\{\mathcal{C}\})}\}$, where (n) denotes the n-th largest index in $\mathcal{C}$ and $N\{\mathcal{C}\}$ is the total number of indices in $\mathcal{C}$, we have

$$\widetilde{\boldsymbol{\Lambda}}_{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{Q}}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{\Lambda}}_{\boldsymbol{A}} = \widetilde{\boldsymbol{\Lambda}}, \tag{112}$$

where $\widetilde{\boldsymbol{\Lambda}}$ is a diagonal matrix. In (112), $\widetilde{\boldsymbol{\Lambda}}_{\boldsymbol{A}}$ is a full-rank diagonal matrix and thus it is concluded that $\widetilde{\boldsymbol{Q}}$ consists of $N\{\mathcal{C}\}$ eigenvectors of $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$. In other words, the eigenvectors of the EVD of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ corresponding to the nonzero eigenvalues are the eigenvectors of the EVD of $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$. This proves that there exists a unitary EVD matrix of $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$, which is also the unitary EVD matrix of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$.

APPENDIX B
DERIVATION OF (41)

The Lagrange of the optimization (40) is given by [14]

$$\mathcal{L}(\boldsymbol{F},t,\{\alpha_n\},\{\lambda_k\}) = \sum_n \alpha_n \left(\left[\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1} \right]_{n,n} - t \right) + \sum_k \lambda_k (\mathrm{Tr}(\boldsymbol{\Omega}_k \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) - P_k) + t, \tag{113}$$

based on which the KKT conditions of (40) are derived as

$$1 = \sum_n \alpha_n, \ \left[\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1} \right]_{n,n} = t, \tag{114}$$

$$\lambda_k(\mathrm{Tr}(\boldsymbol{\Omega}_k \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) - P_k) = 0, \ \mathrm{Tr}(\boldsymbol{\Omega}_k \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) \leq P_k, \ \alpha_k \geq 0,$$

$$\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\boldsymbol{\Lambda}_\alpha \left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1} = \boldsymbol{\mu}\boldsymbol{\Omega}\boldsymbol{F}.$$

Based on the first two KKT conditions, the Lagrange function can be rewritten as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{F},t,\{\alpha_n\},\{\lambda_k\}) \\ &= t + \frac{1}{N}\mathrm{Tr}\left(\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\right)\sum_n \alpha_n - t\sum_n \alpha_n \\ &\quad + \sum_k \lambda_k (\mathrm{Tr}(\boldsymbol{\Omega}_k \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) - P_k) \\ &= \alpha \mathrm{Tr}\left(\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\right) + \sum_k \lambda_k (\mathrm{Tr}(\boldsymbol{\Omega}_k \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) - P_k). \end{aligned} \tag{115}$$

The first order equation with respect to $\boldsymbol{F}$ is then given by

$$\begin{aligned} &\alpha \boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1}\left(\boldsymbol{I}+\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right)^{-1} \\ &= \mu \boldsymbol{\Omega}\boldsymbol{F}. \end{aligned} \tag{116}$$

REFERENCES

[1] 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.

[2] 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.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans.Commun.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[4] Y. Cai, Q. Shi, B. Champagne and G. Y. Li, "Joint transceiver design for secure downlink communications over an amplify-and-forward MIMO relay," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3691–3704, Sep. 2017.

[5] H. Vaezy, M. J. Omidi, M. M. Naghsh and H. Yanikomeroglu, "Energy efficient transceiver design in MIMO interference channels: The selfish, unselfish, worst-case, and robust methods," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5377–5389, Aug. 2019.

[6] A. K. Dutta, K. V. S. Hari and L. Hanzo, "Linear transceiver design for an amplify-and-forward relay based on the MBER criterion," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3765–3777, Nov. 2014.

[7] T. Endeshaw Bogale, B. K. Chalise and L. Vandendorpe, "Robust transceiver optimization for downlink multiuser MIMO systems," in *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 446–453, Jan. 2011.

[8] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," in *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 214–226, Jan. 2004.

[9] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge University Press: Cambridge, UK, 2011.

[10] A. Hjørungnes, D. Gesbert, and D. P. Palomar, "Unified theory of complex-valued matrix differentiation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2007, pp. 345–348.

[11] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[12] S. Kay, "Fundamentals of statistical signal processing: Estimation theory," *Technometrics*, vol. 37, no. 4, pp. 465–466, Mar. 1993.

[13] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," *arXiv:0906.4835*, 2009.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press: Cambridge, UK, 2004.

[15] 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.

[16] S. Gong, C. Xing, X. Zhao, S. Ma and J. An, "Unified IRS-aided MIMO transceiver designs via majorization theory," *IEEE Trans. Signal Process.*, vol. 69, pp. 3016–3032, 2021.

[17] R. Li, S. Sun, Y. Chen, C. Han and M. Tao, "Ergodic achievable rate analysis and optimization of RIS-assisted millimeter-wave MIMO communication systems," in *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 972–985, Feb. 2023.

[18] C. Xing, *et al.*, "Matrix-monotonic optimization - Part I: Single-variable optimization," *IEEE Trans. Signal Process.*, vol. 69, pp 738–754, 2021.

[19] C. Xing, *et al.*, "Matrix-monotonic optimization - Part II: Multi-variable optimization," *IEEE Trans. Signal Process.*, vol. 69, pp. 179–194, 2021.

[20] A. Liu, Y. Liu, H. Xiang, and W. Luo, "MIMO B-MAC interference network optimization under rate constraints by polite water-filling and duality," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 263–276, Jan. 2011.

[21] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 361–374, Feb. 2006.

[22] 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 Trans. Signal Process.*, vol. 56, no. 9, pp. 4306–4316, Sep. 2008.

[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press. 1985.

[24] Y. Yu and Y. Hua, "Power allocation for a MIMO relay systems with multipe-antenna users," *IEEE Trans. Signal Process.*, vo. 58, no. 5, pp. 2823–2835, May 2010.

[25] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[26] 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.

[27] S. Gong, *et al.*, "A unified MIMO optimization framework relying on the KKT conditions," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7251–7268, Nov. 2021.

[28] L. Zhang, *et al.*, "On Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.

[29] W. Utschick and J. Brehmer, "Monotonic optimization framework for coordinated beamforming in multicell networks," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1899–1909, Apr. 2012.

[30] C. Xing, Y. Jing, and Y. Zhou, "On weighted MSE model for MIMO transceiver optimization," *IEEE Trans. Veh. Tech.*, vol. 66, no. 8, pp. 7072–7085, Aug. 2017.

[31] G. Michael and B. Stephen, *CVX Users' Guide for CVX ver. 1.21* (Apr. 2011). [Online]. Available: http://cvxr.com/

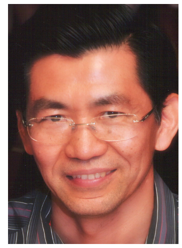
Chengwen Xing (S'08–M'10) received the B.Eng. degree from Xidian University, Xi'an, China, in 2005, and the Ph.D. degree from the University of Hong Kong, Hong Kong, China, in 2010. Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, where he is currently a Full Professor. From September 2012 to December 2012, he was a Visiting Scholar at the University of Macau, Macau SAR, China. His current research interests include machine learning, statistical signal processing, convex optimization, multivariate statistics, and array signal processing.

Yihan Li received the B.S. degree from Beijing Institute of Technology, Beijing, China, in 2021, where she is currently working towards the Ph.D. degree with the School of Information and Electronics. Her research interests include signal processing, near-field communication, and convex optimization.

Shiqi Gong received the B.S. and Ph.D. degrees in electronic engineering from Beijing Institute of Technology, China, in 2014 and 2020, respectively. She is currently an associate professor with the School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing, China. Her research interests are in the area of intelligent reflecting surface, physical-layer security, resource allocation, and convex optimization. She was a recipient of the Best Ph.D. Thesis Award of Beijing Institute of Technology in 2020.

Jianping An (Senior Member, IEEE) received the B.Eng. degree from PLA Information Engineering University, Zhengzhou, China, in 1987, and the Ph.D. degree from the Beijing Institute of Technology (BIT), Beijing, China, in 1996.

From 2010 to 2011, he was a Visiting Professor with the University of California at San Diego, San Diego, CA, USA. He is currently a Professor and the Dean of the School of Cyberspace Science and Technology, BIT. He has authored or coauthored more than 150 journal and conference articles and holds or coholds 37 patents. His current research interests include signal processing theory and algorithms for communication systems. He was the recipient of various awards for his academic achievements and the resultant industrial influences, including the National Award for Scientific and Technological Progress of China in 1997 and the Excellent Young Teacher Award by the China's Ministry of Education in 2000. Since 2010, he has been a Chief Reviewing Expert for the Information Technology Division, National Scientific Foundation of China.



Sheng Chen (IEEE Fellow) received his BEng degree from the East China Petroleum Institute, Dongying, China, in 1982, and his PhD degree from the City University, London, in 1986, both in control engineering. In 2005, he was awarded the higher doctoral degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK. From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has been with the School of Electronics and Computer Science, the University of Southampton, UK, where he holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen's research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, evolutionary computation methods and optimization. He has coauthored over 700 research papers. Professor Chen has 19,100+ Web of Science citations with h-index 60 and 37,300+ Google Scholar citations with h-index 82. Dr. Chen is a Fellow of the United Kingdom Royal Academy of Engineering, a Fellow of Asia-Pacific Artificial Intelligence Association and a Fellow of IET. He is one of the original ISI highly cited researchers in engineering (March 2004).

Lajos Hanzo (FIEEE'04) received Honorary Doctorates from the Technical University of Budapest (2009) and Edinburgh University (2015). He is a Foreign Member of the Hungarian Science-Academy, Fellow of the Royal Academy of Engineering (FREng), of the IET, of EURASIP and holds the IEEE Eric Sumner Technical Field Award.(http://www-mobile.ecs.soton.ac.uk,https://en.wikipedia.org/wiki/Lajos_Hanzo)