Joint Design of Multiple Non-Regenerative MIMO Relaying Matrices With Power Constraints

Chao Zhao, *Student Member, IEEE*, and Benoît Champagne, *Senior Member, IEEE*

Abstract—This paper investigates the joint design of multiple non-regenerative multiple-input multiple-output (MIMO) relaying matrices, with the purpose of minimizing the mean square error (MSE) between the transmitted signals from the source and the received signals at the destination. Two types of constraints on the transmit power of the relays are considered separately: 1) a weighted sum power constraint, and 2) per-relay power constraints. As opposed to using general-purpose interior-point methods, we exploit the inherent structure of the problems to develop more efficient algorithms. Under the weighted sum power constraint, the optimal solution is expressed as a function of a Lagrangian parameter. By introducing a complex scaling factor at the destination, we derive a closed-form expression for this parameter, thereby avoiding the need to solve an implicit nonlinear equation numerically. Under the per-relay power constraints, the optimal solution is the same as that under the weighted sum power constraint if particular weights are chosen. We then propose an iterative power balancing algorithm to compute these weights. In addition, under both types of constraints, we investigate the joint design of a MIMO equalizer at the destination and the relaying matrices, using block coordinate descent or steepest descent. The bit-error rate (BER) simulation results demonstrate that all the proposed designs, under either type of constraints, with or without the equalizer, perform much better than previous methods.

Index Terms—Distributed array gain, MIMO, MMSE, non-regenerative, power balancing, relay.

I. INTRODUCTION

MULTIPLE-INPUT MULTIPLE-OUTPUT (MIMO) wireless relaying can increase system throughput, overcome shadowing and expand network coverage more efficiently than its single-antenna counterpart [1]. The multi-antenna relays can either decode and re-encode information bits, or simply apply linear processing matrices to the received baseband signals before retransmitting them. These approaches are known as decode-and-forward (DF) and amplify-and-forward (AF), respectively. This paper is concerned with the latter *non-regenerative* approach, which benefits from shorter processing delays, lower complexity and better security [2]–[8].

Communication between a source and a destination can be assisted by either a single or multiple relays [2]. These configurations are referred to as one-source—one-relay—one-destination (1S-1R-1D) and one-source—multiple-relays—one-destination (1S-MR-1D) hereafter. The optimal relaying matrices for 1S-1R-1D configurations are well established for a wide variety of criteria when the transmit power of the relay is constrained [3]–[6]. These matrices share a common singular value decomposition SVD structure which diagonalizes the backward and forward channels. This framework, however, cannot be extended to the joint design of multiple relaying matrices for 1S-MR-1D systems. Indeed, since the relays can only process their own signals, the compound AF matrix has to be *block-diagonal*.

The essential feature of an appropriate relaying strategy is that the signals from different relays should be coherently combined at the destination, thereby leading to a *distributed array gain* [1], [7], [9]. In this regard, some strategies have been proposed that "borrow" ideas from MIMO transceiver design, including matched filtering (MF), zero-forcing (ZF), linear minimum mean square error (MMSE) [7], QR decomposition [8] and the hybrid relaying framework [10]. These heuristic methods, although structurally constrained, were shown to perform much better than simplistic AF which only amplifies the signals. A more comprehensive approach is to formulate the collaborative design of the relaying matrices as optimization problems with power constraints [11]–[16]. The objective can be to maximize the achievable rate [11] or to minimize the mean square error (MSE) [13], [14]. However, most works rely on numerical algorithms such as gradient descent [13], bisection [14] and iterative schemes [13]–[15] to obtain the optimal solution. This lack of closed-form expressions leads to high implementation complexity, which in turn limits the potential applicability of these methods. For completeness, it is worth mentioning that explicit formulas were derived in [17]–[19] when the power constraints are enforced on the signals *received* at the destination. However, these results do not carry over to the case when the constraints are imposed on the *transmit* power of the relays [16].

In this paper, we concentrate on the similar problems of designing the multiple relaying matrices, with the purpose of minimizing the MSE between the input and output signals. Two types of constraints on the transmit power of the relays are considered separately: 1) a weighted sum power constraint which was not investigated before, and 2) per-relay power constraints. The problems are first recast as standard quadratically constrained quadratic programs (QCQPs) through vectorization. As opposed to using general-purpose interior-point methods [20], we exploit the inherent structures of the problems to develop more efficient algorithms. Under the weighted sum power

constraint, the optimal solution is expressed as an explicit function of a Lagrangian parameter. By introducing a complex scaling factor at the destination, we derive a closed-form expression for this parameter, thereby overcoming the hurdle of solving an implicit nonlinear equation. Under the per-relay power constraints, the optimal solution is the same as that under the weighted sum power constraint if a particular set of weights is chosen. We then propose a simple iterative power balancing algorithm to compute these weights efficiently. In addition, under both types of constraints, we investigate the joint design of a multiple-input multiple-output (MIMO) equalizer at the destination and the relaying matrices, using block coordinate descent or steepest descent. The bit-error rate (BER) simulation results demonstrate better performance for the proposed MMSE-based relaying strategies, under either type of constraints, with or without the equalizer, than previous methods.

Our work provides new insights into the design of non-regenerative 1S-MR-1D systems. Firstly, we point out the possible non-uniqueness of the solution to the first-order necessary condition, which was overlooked in [12], [14], [15]. Moreover, it is not legitimate to simply choose the minimum-norm solution, unless the vectorization is done on specific transformations of the relaying matrices instead of these matrices themselves. Secondly, the optimal design does not require global channel state information (CSI) availability: each relay only needs to know its own backward and forward channel, together with a little additional information. Thirdly, under the weighted sum power constraint, the optimal strategy tends to allocate more power to those relays with better source-relay links or worse relay-destination links. Lastly, under the per-relay power constraints, the optimal strategy sometimes does not use the maximum power at some relays. Forcing equality in the per-relay power constraints as in [13] would result in loss of optimality. Another interesting point is that, no matter how low the signal-to-noise ratio (SNR) is at a particular relay, this relay does not have to be turned off completely.

The organization of this paper is as follows: Section II introduces the system model and formulates the mathematical problem. Section III derives the closed-form optimal relaying matrices under the weighted sum power constraint. Section IV studies the per-relay power constraints. The joint design of the relaying matrices and the MIMO equalizer is discussed in Section V. Section VI covers the implementation issues and computational complexity. Numerical results are presented in Section VII followed by a brief conclusion in Section VIII.

The following notations are used: italic, boldface lowercase and boldface uppercase letters represent scalars, vectors and matrices; superscripts $^-$, ^{T}, ^{H} and † denote conjugate, transpose, Hermitian transpose and Moore-Penrose pseudo-inverse, respectively; $\mathrm{tr}()$ refers to the trace of a matrix; $\|\cdot\|_2$ $(\|\cdot\|_F)$ stands for the Euclidean (or Frobenius) norm of a vector; $\mathrm{col}()$ stacks many column vectors into a single vector, $\mathrm{vec}()$ stacks the columns of a matrix into a vector and $\mathrm{unvec}()$ is its inverse operator; $\mathrm{diag}()$ forms a diagonal (or block-diagonal) matrix from multiple scalars (or matrices); $\otimes$ represents the Kronecker product; $\mathbf{I}_n$ is an identity matrix of dimension n; $\mathrm{E}\{\}$ refers to mathematical expectation; $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers; $\mathcal{R}()$ and $\mathcal{N}()$ are the column space and the null space of a matrix; $\dim()$ is the dimension of a space.
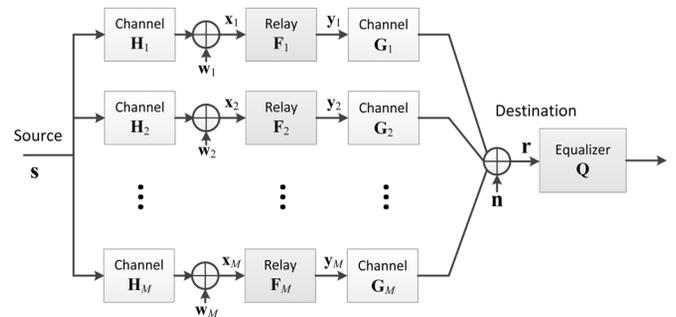
Fig. 1.   System model of 1S-MR-1D.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

In the 1S-MR-1D system model depicted by Fig. 1, a multi-antenna source is sending symbols to a multi-antenna destination with the aid of multiple multi-antenna relays. The transmitted signals propagate through the backward channels between the source and the relays. These signals are processed at the individual relays and propagate through the forward channels to the destination. The relays work in a half-duplex mode: their antennas are used for either transmitting or receiving during different time slots. We neglect the presence of a possible direct source-to-destination link, which is typically hindered by high levels of attenuation.

We assume that the channels undergo frequency-flat block-fading [21]. The source does not have access to the CSI; each relay knows its own backward and forward channels and a little additional shared information; the destination may need an equalizer matrix. Once knowing the structures of the optimal solution, we shall be able to discuss this topic in detail (cf. Section VI). The channel matrices have to be estimated timely and accurately, which is an important topic in its own right. For more details, we refer the reader to [22]–[24] and the references therein.

The bandpass signals and channels are modeled in terms of their discrete-time complex baseband counterparts. The numbers of antennas at the source, relays and destination are respectively denoted by N_S, N_R and N_D.[1] The source signal $\mathbf{s} \in \mathbb{C}^{N_S \times 1}$ consists of N_S statistically independent symbol streams. It is assumed to have zero mean and a full-rank covariance matrix $\mathbf{R}_s \triangleq \mathrm{E}\{\mathbf{s}\mathbf{s}^H\}$. The received signal at the kth relay, $\mathbf{x}_k \in \mathbb{C}^{N_R \times 1}$, can be expressed as

$$\mathbf{x}_k = \mathbf{H}_k \mathbf{s} + \mathbf{w}_k, \quad \forall\, 1 \le k \le M, \tag{1}$$

where $\mathbf{H}_k \in \mathbb{C}^{N_R \times N_S}$ is the backward channel matrix from the source to the kth relay, and $\mathbf{w}_k \in \mathbb{C}^{N_R \times 1}$ is an additive noise term modeled as a circularly symmetric complex Gaussian random vector with zero mean and full-rank covariance matrix $\mathbf{R}_{w_k} \triangleq \mathrm{E}\{\mathbf{w}_k \mathbf{w}_k^H\}$. The random vectors $\mathbf{w}_1, \ldots, \mathbf{w}_M$ and $\mathbf{s}$ are statistically independent.

The kth relay retransmits its noisy signal $\mathbf{x}_k$ as

$$\mathbf{y}_k = \mathbf{F}_k \mathbf{x}_k, \quad \forall\, 1 \le k \le M, \tag{2}$$

[1]For notational simplicity, each relay is equipped with the same number of antennas; however, generalization to different numbers of antennas at the relays, i.e., $N_{R,k}$, is straightforward.

where $\mathbf{F}_k \in \mathbb{C}^{N_R \times N_R}$ is the corresponding non-regenerative MIMO relaying matrix. The signal received at the destination, denoted by $\mathbf{r} \in \mathbb{C}^{N_D \times 1}$, takes the form of

$$
\begin{aligned}
\mathbf{r} &= \sum_{k=1}^{M} \mathbf{G}_k \mathbf{y}_k + \mathbf{n} \\
&= \sum_{k=1}^{M} \mathbf{G}_k \mathbf{F}_k \mathbf{H}_k \mathbf{s} + \sum_{k=1}^{M} \mathbf{G}_k \mathbf{F}_k \mathbf{w}_k + \mathbf{n},
\end{aligned}
\tag{3}
$$

where $\mathbf{G}_k \in \mathbb{C}^{N_D \times N_R}$ is the forward channel matrix from the kth relay to the destination and $\mathbf{n} \in \mathbb{C}^{N_D \times 1}$ is the noise induced at the destination receiver. This term is independent from $\mathbf{s}$ and $\{\mathbf{w}_k\}$ and also modeled as a circularly symmetric complex Gaussian random vector, with zero mean and covariance $\mathbf{R}_n \triangleq \mathrm{E}\{\mathbf{n}\mathbf{n}^H\}$. The destination may apply a linear MIMO equalizer (combiner) $\mathbf{Q} \in \mathbb{C}^{N_S \times N_D}$, resulting in

$$
\hat{\mathbf{r}} = \mathbf{Q}\mathbf{r}.
\tag{4}
$$

The above signal model can also be expressed in a compact block-diagonal form, *viz.*,

$$
\hat{\mathbf{r}} = \mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{H}\mathbf{s} + \mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{w} + \mathbf{Q}\mathbf{n},
\tag{5}
$$

where we define $\mathbf{G} \triangleq [\mathbf{G}_1, \ldots, \mathbf{G}_M]$, $\mathbf{H} \triangleq [\mathbf{H}_1^H, \ldots, \mathbf{H}_M^H]^H$, $\mathbf{F} \triangleq \mathrm{diag}(\mathbf{F}_1, \ldots, \mathbf{F}_M)$ and $\mathbf{w} \triangleq \mathrm{col}(\mathbf{w}_1, \ldots, \mathbf{w}_M)$ with $\mathbf{R}_w \triangleq \mathrm{E}\{\mathbf{w}\mathbf{w}^H\} = \mathrm{diag}\{\mathbf{R}_{w_1}, \ldots, \mathbf{R}_{w_M}\}$. If $M = 1$, this signal model reduces to the 1S-1R-1D case.

In addition, with proper refinement, the mathematical model in this paper is applicable to a much broader scope such as 1S-MR-1D systems with broadband transmission [25], distributed relaying systems, and multiuser multi-relay systems.

B. Problem Formulation

The major goal is to design the relay processing matrices $\{\mathbf{F}_k\}$, so that the distortion between the output $\hat{\mathbf{r}}$ and the input $\mathbf{s}$ is minimized. Our choice of the objective function, for practical reasons, is the MSE:

$$
\begin{aligned}
\mathrm{MSE}&(\mathbf{F}, \mathbf{Q}) \\
&\triangleq \mathrm{E}_{\mathbf{s},\mathbf{w},\mathbf{n}}\{\|\hat{\mathbf{r}} - \mathbf{s}\|^2\} \\
&= \mathrm{tr}((\mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{H} - \mathbf{I})\mathbf{R}_s(\mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{H} - \mathbf{I})^H) \\
&\quad + \mathrm{tr}(\mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{R}_w \mathbf{F}^H \mathbf{G}^H \mathbf{Q}^H) + \mathrm{tr}(\mathbf{Q}\mathbf{R}_n \mathbf{Q}^H).
\end{aligned}
\tag{6}
$$

Although BER performance also depends on nonlinear components such as channel coding, space-time coding, interleaving and constellation mapping, the MSE serves as a good performance indicator and is more mathematically tractable [26].

Two types of power constraints are separately imposed on the relays. The first is the weighted sum power constraint

$$
\sum_{k=1}^{M} w_k \mathrm{tr}\left(\mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H\right) \leq P_r,
\tag{7}
$$

where $\mathbf{R}_{x_k} \triangleq \mathrm{E}\{\mathbf{x}_k \mathbf{x}_k^H\} = \mathbf{H}_k \mathbf{R}_s \mathbf{H}_k^H + \mathbf{R}_{w_k}$, and $w_k \geq 0$ for $1 \leq k \leq M$ are the weights assigned to different relays. The other type is the per-relay power constraints, i.e., each relay has its own power budget, expressed as

$$
\mathrm{tr}\left(\mathbf{R}_{y_k}\right) = \mathrm{tr}\left(\mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H\right) \leq P_k, \quad \forall 1 \leq k \leq M,
\tag{8}
$$

where $\mathbf{R}_{y_k} \triangleq \mathrm{E}\{\mathbf{y}_k \mathbf{y}_k^H\} = \mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H$.

To simplify the mathematical development, it is convenient to vectorize the relaying matrices. To this end, we define

$$
\mathbf{f}_k \triangleq \mathrm{vec}\left(\mathbf{F}_k \mathbf{R}_{x_k}^{1/2}\right), \quad 1 \leq k \leq M,
\tag{9}
$$

and $\mathbf{f} \triangleq \mathrm{col}(\mathbf{f}_1, \ldots, \mathbf{f}_M)$. The reason for this definition, instead of $\mathrm{vec}(\mathbf{F}_k)$, is that the square of the 2-norm of $\mathbf{f}_k$ is equal to the transmit power of the kth relay, *viz.*,

$$
\|\mathbf{f}_k\|_2^2 = \mathbf{f}_k^H \mathbf{f}_k = \mathrm{tr}\left(\mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H\right).
\tag{10}
$$

As shown later, this will bring much convenience. It is straightforward to invert (9) as $\mathbf{F}_k = \mathrm{unvec}(\mathbf{f}_k)\mathbf{R}_{x_k}^{-1/2}$. For notational simplicity, we also define the matrices

$$
\mathbf{T}_k \triangleq \left(\mathbf{H}_k^H \mathbf{R}_{x_k}^{-1/2}\right)^T \otimes \mathbf{G}_k^H \mathbf{Q}^H,
\tag{11a}
$$

$$
\mathbf{S}_k \triangleq \left(\mathbf{R}_{x_k}^{-1/2} \mathbf{R}_{w_k} \mathbf{R}_{x_k}^{-1/2}\right)^T \otimes \left(\mathbf{G}_k^H \mathbf{Q}^H \mathbf{Q} \mathbf{G}_k\right),
\tag{11b}
$$

for $1 \leq k \leq M$. These matrices serve as the building blocks for the following matrices and vectors:

$$
\mathbf{T} \triangleq \left[\mathbf{T}_1^T, \ldots, \mathbf{T}_M^T\right]^T,
\tag{12a}
$$

$$
\mathbf{S} \triangleq \mathrm{diag}(\mathbf{S}_1, \ldots, \mathbf{S}_M),
\tag{12b}
$$

$$
\mathbf{\Phi} \triangleq \mathbf{T}\left(\mathbf{R}_s^T \otimes \mathbf{I}_{N_S}\right)\mathbf{T}^H + \mathbf{S},
\tag{12c}
$$

$$
\mathbf{b} \triangleq \mathbf{T}\left(\mathbf{R}_s^T \otimes \mathbf{I}_{N_S}\right)\mathrm{vec}(\mathbf{I}_{N_S}).
\tag{12d}
$$

With the above notations, the objective function in (6) becomes a quadratic function of the vector $\mathbf{f}$:

$$
\begin{aligned}
\mathrm{MSE}(\mathbf{f}, \mathbf{Q}) = \mathbf{f}^H \mathbf{\Phi} \mathbf{f} - \mathbf{f}^H \mathbf{b} - \mathbf{b}^H \mathbf{f} \\
+ \mathrm{tr}(\mathbf{R}_s) + \mathrm{tr}(\mathbf{Q}\mathbf{R}_n \mathbf{Q}^H),
\end{aligned}
\tag{13}
$$

where we have used the following properties [27]

$$
\mathrm{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B}),
\tag{14a}
$$

$$
(\mathbf{A}\mathbf{B}) \otimes (\mathbf{C}\mathbf{D}) = (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}),
\tag{14b}
$$

$$
\mathrm{tr}(\mathbf{A}^T \mathbf{Y}^T \mathbf{B}\mathbf{X}) = [\mathrm{vec}(\mathbf{Y})]^T(\mathbf{A} \otimes \mathbf{B})\mathrm{vec}(\mathbf{X}).
\tag{14c}
$$

Hereafter, we may denote the arguments of the function $\mathrm{MSE}(\mathbf{f}, \mathbf{Q})$ in (13) differently, to emphasize its dependence on certain variables, vectors or matrices.

The power constraints are also represented in terms of $\mathbf{f}$. For convenience, define $\mathbf{I}_{(k)}$ as

$$
\mathbf{I}_{(k)} \triangleq \mathrm{diag}\left(\mathbf{0}_{N_R^2}, \ldots, \mathbf{I}_{N_R^2}, \ldots, \mathbf{0}_{N_R^2}\right),
\tag{15}
$$

where $\mathbf{I}_{N_R^2}$ is in the kth diagonal sub-block. A weighted sum of such matrices is also defined: $\mathbf{I}_{\mathrm{sum}} \triangleq \sum_{k=1}^{M} w_k \mathbf{I}_{(k)}$. Then, the weighted sum power constraint in (7) becomes

$$
\mathbf{f}^H \mathbf{I}_{\mathrm{sum}} \mathbf{f} \leq P_r,
\tag{16}
$$

and the per-relay power constraints in (8) would be

$$
\mathbf{f}^H \mathbf{I}_{(k)} \mathbf{f} \leq P_k, \quad \forall 1 \leq k \leq M.
\tag{17}
$$

The above formulated optimization problems are flexible with respect to the equalizer $\mathbf{Q}$—it can be either pre-determined, or jointly designed with $\{\mathbf{F}_k\}$ (or equivalently $\mathbf{f}$). If $\mathbf{Q}$ is fixed, the problems under both types of constraints are standard QCQPs that can be solved by general-purpose interior-point methods [20], [28]. In Sections III and IV, however, the proper

exploitation of the sparse structures of $\mathbf{\Phi}$, $\mathbf{b}$ and $\mathbf{I}_{(k)}$ leads to more efficient algorithms and in some cases closed-form expressions for the optimal solution. For the joint design of $\mathbf{Q}$ and $\mathbf{f}$, we shall propose two algorithms in Section V, both of which rely upon the results from Sections III and IV.

III. THE WEIGHTED SUM POWER CONSTRAINT

In this section, we assume $\mathbf{Q}$ is fixed and derive the optimal $\mathbf{f}^\star$ under the weighted sum power constraint. The first step is to establish the optimality conditions through the framework of Lagrangian duality. Then, the optimal solution $\mathbf{f}^\star$ is expressed as an explicit function of a Lagrangian parameter. By introducing a complex scaling factor in $\mathbf{Q}$, we derive a closed-form expression for this parameter.

A. Optimality Conditions

Most constrained optimization problems are solved through Lagrangian duality [20], [28]. The starting step is to define the Lagrangian function as

$$L(\mathbf{f}, \lambda) \triangleq \mathrm{MSE}(\mathbf{f}) + \lambda(\mathbf{f}^H \mathbf{I}_{\mathrm{sum}} \mathbf{f} - P_r), \tag{18}$$

where the dual variable satisfies $\lambda \geq 0$. The infimum of (18) over $\mathbf{f} \in \mathbb{C}^{M N_R^2 \times 1}$ is defined as

$$D(\lambda) \triangleq \inf_{\mathbf{f}} L(\mathbf{f}, \lambda). \tag{19}$$

The dual problem is defined as:

$$\begin{aligned} \text{maximize} \quad & D(\lambda) \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned} \tag{20}$$

Let the solution of the primal problem be $\mathbf{f}^\star$ and $p^\star \triangleq \mathrm{MSE}(\mathbf{f}^\star)$; let the solution of the dual problem be $\lambda^\star$ and $d^\star \triangleq D(\lambda^\star)$. For a convex primal problem, strong duality holds (i.e., the duality gap $p^\star - d^\star$ is zero) if the Slater's condition is satisfied ([28], p. 226). Here, the primal problem is convex ($\mathbf{\Phi}$ and $\mathbf{I}_{\mathrm{sum}}$ are both positive semidefinite) and the Slater's condition is always satisfied ($\mathbf{f} = \mathbf{0}$ is strictly feasible: $0 < P_r$). Henceforth, the Karush-Kuhn-Tucker (KKT) conditions are *necessary and sufficient* for the optimal primal-dual pair $(\mathbf{f}^\star, \lambda^\star)$, *viz.*,

$$\nabla_{\bar{\mathbf{f}}} L(\mathbf{f}, \lambda^\star)|_{\mathbf{f} = \mathbf{f}^\star} = \mathbf{0}, \tag{21a}$$

$$\mathbf{f}^{\star H} \mathbf{I}_{\mathrm{sum}} \mathbf{f}^\star - P_r \leq 0, \tag{21b}$$

$$\lambda^\star \geq 0, \tag{21c}$$

$$\lambda^\star (\mathbf{f}^{\star H} \mathbf{I}_{\mathrm{sum}} \mathbf{f}^\star - P_r) = 0. \tag{21d}$$

Among these four conditions, the first-order necessary condition (21a), which we referred to as the *stationarity condition* from now on, determines the analytical form for the optimal solution; the complementary slackness condition (21d) serves as the key to computing the value of $\lambda^\star$.

B. The Solution Space of the Stationarity Condition

The stationarity condition (21a) can be rewritten as the following linear equation

$$\mathbf{\Psi} \mathbf{f}^\star = \mathbf{b}, \tag{22}$$

in which $\mathbf{\Psi} \triangleq \mathbf{\Phi} + \lambda^\star \mathbf{I}_{\mathrm{sum}}$ is a function of $\lambda^\star$. Three questions can be asked about (22): Does the solution always exist? If yes, is it unique? If non-unique, what is the general solution form? The following theorem answers the first question:

Theorem 1: The stationarity equation (22) has at least one solution for any $\lambda^\star \geq 0$ and $w_k \geq 0$, that is: $\mathbf{b} \in \mathcal{R}(\mathbf{\Psi})$.

Proof: We proceed by contradiction. If no solution exists, $\mathbf{b} \notin \mathcal{R}(\mathbf{\Psi})$. Let $\mathbf{b}_0 = \mathbf{\Psi} \mathbf{f}_0$ be the orthogonal projection of $\mathbf{b}$ onto $\mathcal{R}(\mathbf{\Psi})$ and then the non-zero residual $\mathbf{b}_\perp$ would satisfy

$$\mathbf{b}_\perp \triangleq \mathbf{b} - \mathbf{b}_0 = \mathbf{b} - \mathbf{\Psi} \mathbf{f}_0 \in \mathcal{N}(\mathbf{\Psi}^H) = \mathcal{N}(\mathbf{\Psi}).$$

The value of the Lagrangian function (18) evaluated at $\mathbf{f} = \mathbf{f}_0 + \alpha \mathbf{b}_\perp$ would be

$$\begin{aligned} L(\mathbf{f}_0 &+ \alpha \mathbf{b}_\perp, \lambda^\star) \\ &= -\mathbf{f}_0^H \mathbf{\Psi} \mathbf{f}_0 - \mathbf{b}_\perp^H \mathbf{f}_0 - \mathbf{f}_0^H \mathbf{b}_\perp - 2\alpha \|\mathbf{b}_\perp\|^2 \\ &\quad + \mathrm{tr}(\mathbf{R}_s) + \mathrm{tr}(\mathbf{Q} \mathbf{R}_n \mathbf{Q}^H) - \lambda^\star P_r. \end{aligned}$$

As $\alpha \to +\infty$, $L(\mathbf{f}_0 + \alpha \mathbf{b}_\perp, \lambda^\star) \to -\infty$ because $\|\mathbf{b}_\perp\|^2 > 0$. However, all terms in the definition of $L(\mathbf{f}, \lambda^\star)$ in (18) are nonnegative except for the constant $-\lambda^\star P_r$. This means $L(\mathbf{f}_0 + \alpha \mathbf{b}_\perp, \lambda^\star) \geq -\lambda^\star P_r$, which leads to a contradiction. ∎

The second question is whether the solution is unique. This is true, if and only if, $\mathbf{\Psi} \mathbf{f} = \mathbf{0}$ does not have a non-zero solution, or equivalently $\mathcal{N}(\mathbf{\Psi}) = \{\mathbf{0}\}$. The following propositions establish some facts on $\mathcal{N}(\mathbf{\Phi})$ and $\mathcal{N}(\mathbf{\Psi})$, respectively:

Proposition 2:
1) If $\mathbf{f}_k^\perp \in \mathbb{C}^{N_R^2 \times 1}$ is in the null space of $\mathbf{I}_{N_R} \otimes \mathbf{Q} \mathbf{G}_k$, the vector $\mathbf{f}_{(k)}^\perp \triangleq \mathrm{col}(\mathbf{0}, \ldots, \mathbf{f}_k^\perp, \ldots, \mathbf{0}) \in \mathbb{C}^{M N_R^2 \times 1}$ must be in the null space of $\mathbf{\Phi}$.
2) For any k, all such $\mathbf{f}_{(k)}^\perp$ together span a subspace $\mathcal{F}_k$ with dimension $N_R^2 - N_R \mathrm{rank}(\mathbf{Q} \mathbf{G}_k)$.
3) For $k \neq l$, $\mathcal{F}_k$ and $\mathcal{F}_l$ are orthogonal to each other.
4) The vector $\mathbf{b}$ is orthogonal to all $\mathcal{F}_k$ for $1 \leq k \leq M$.

Proof: 1) Define $\mathbf{\Phi}_{lk}$ as the (l, k)th sub-block of $\mathbf{\Phi}$. The lth sub-block of $\mathbf{\Phi} \mathbf{f}_{(k)}^\perp$ is

$$[\mathbf{\Phi}_{l1}, \ldots, \mathbf{\Phi}_{lM}] \mathbf{f}_{(k)}^\perp = \mathbf{\Phi}_{lk} \mathbf{f}_k^\perp = \mathbf{X}_k (\mathbf{I}_{N_R} \otimes \mathbf{Q} \mathbf{G}_k) \mathbf{f}_k^\perp = \mathbf{0},$$

where $\mathbf{X}_k = (\mathbf{R}_{x_k}^{-1/2} \mathbf{H}_k \mathbf{R}_s \mathbf{H}_l^H \mathbf{R}_{x_l}^{-1/2})^T \otimes \mathbf{G}_l^H \mathbf{Q}^H$ if $l \neq k$ and $\mathbf{X}_k = \mathbf{I} \otimes \mathbf{G}_l^H \mathbf{Q}^H$ if $l = k$. 2) Because a one-to-one correspondence exists between $\mathbf{f}_k^\perp$ and $\mathbf{f}_{(k)}^\perp$ according to the definition, all $\mathbf{f}_{(k)}^\perp$ together also span a subspace $\mathcal{F}_k$ isomorphic to $\mathcal{N}(\mathbf{I}_{N_R} \otimes \mathbf{Q} \mathbf{G}_k)$. Since the rank of $\mathbf{I}_{N_R} \otimes \mathbf{Q} \mathbf{G}_k$ (with dimension $N_R N_S \times N_R^2$) is $N_R \mathrm{rank}(\mathbf{Q} \mathbf{G}_k)$, the dimension of its null space is $N_R^2 - N_R \mathrm{rank}(\mathbf{Q} \mathbf{G}_k)$, which is also that of $\mathcal{F}_k$. 3) From the definition, $\mathbf{f}_{(k)}^{\perp H} \mathbf{f}_{(l)}^\perp = 0$ always holds for $k \neq l$, and so $\mathcal{F}_k$ and $\mathcal{F}_l$ are orthogonal to each other. 4) According to Theorem 1, $\lambda^\star = 0$ leads to $\mathbf{b} \in \mathcal{R}(\mathbf{\Phi})$. Therefore, $\mathbf{b}$ is orthogonal to $\mathcal{N}(\mathbf{\Phi}^H) = \mathcal{N}(\mathbf{\Phi})$ which includes any $\mathcal{F}_k$ as a subset. ∎

Proposition 3: Let $\mathcal{M} \triangleq \{1, \ldots, M\}$ denote the set of relay indexes and $\mathcal{K} \triangleq \{m_1, \ldots, m_K\} \subseteq \mathcal{M}$ includes all the indexes m_k satisfying $\lambda^\star w_{m_k} = 0$. The null space of $\mathbf{\Psi}$ is the direct sum of $\{\mathcal{F}_{m_k}\}$, that is, $\mathcal{N}(\mathbf{\Psi}) = \mathcal{F}_{m_1} \oplus \cdots \oplus \mathcal{F}_{m_K}$.

Proof: We have $\mathcal{F}_{m_1} \oplus \cdots \oplus \mathcal{F}_{m_K} \subseteq \mathcal{N}(\mathbf{\Psi})$ because any $\mathbf{f}_{(m_k)}^\perp \in \mathcal{F}_{m_k}$ as defined in Proposition 2 would satisfy

$$\mathbf{\Psi} \mathbf{f}_{(m_k)}^\perp = \mathbf{\Phi} \mathbf{f}_{(m_k)}^\perp + \lambda^\star w_{m_k} \mathbf{I}_{(m_k)} \mathbf{f}_{(m_k)}^\perp = \mathbf{0}.$$

To prove equality, we only need to prove that $\mathcal{F}_{m_1} \oplus \cdots \oplus \mathcal{F}_{m_K}$ and $\mathcal{N}(\boldsymbol{\Psi})$ have the same dimension. On the one hand, the subset relation leads to

$$\dim(\mathcal{N}(\boldsymbol{\Psi})) \geq \dim(\mathcal{F}_{\pi_1} \oplus \cdots \oplus \mathcal{F}_{\pi_M}).$$

On the other hand, we have

$$\begin{aligned}
\dim(\mathcal{N}(\boldsymbol{\Psi})) &= MN_R^2 - \operatorname{rank}(\boldsymbol{\Psi}) \\
&\leq MN_R^2 - \operatorname{rank}(\mathbf{S} + \lambda^\star \mathbf{I}_{\text{sum}}) \\
&= MN_R^2 - \sum_{k=1}^{M} \operatorname{rank}(\mathbf{S}_k + \lambda^\star w_k \mathbf{I}_{N_R^2}) \\
&= KN_R^2 - \sum_{k=1}^{K} N_R \times \operatorname{rank}(\mathbf{Q}\mathbf{G}_{m_k}) \\
&= \dim(\mathcal{F}_{m_1} \oplus \cdots \oplus \mathcal{F}_{m_M}),
\end{aligned}$$

where the inequality comes from the fact that $\boldsymbol{\Psi}$ is a sum of positive semidefinite matrices and hence its rank cannot increase by removing $\mathbf{T}(\mathbf{R}_s^T \otimes \mathbf{I})\mathbf{T}^H$. ∎

Proposition 3 indicates that a positive $\lambda^\star w_k$ causes the null space of $\boldsymbol{\Psi}$ to shrink and its column space to expand, by the "amount" of $\mathcal{F}_k$. When $\lambda^\star w_k$ is positive, any vector $\mathbf{f}_{(k)}^\perp \in \mathcal{F}_k$ would satisfy

$$\boldsymbol{\Psi}\mathbf{f}_{(k)}^\perp = \boldsymbol{\Phi}\mathbf{f}_{(k)}^\perp + \sum_{l=1}^{M} \lambda^\star w_l \mathbf{I}_{(l)}\mathbf{f}_{(k)}^\perp = \lambda^\star w_k \mathbf{f}_{(k)}^\perp. \quad (23)$$

This means that $\mathbf{f}_{(k)}^\perp$ is always an eigenvector of $\boldsymbol{\Psi}$ with eigenvalue $\lambda^\star w_k$.

Since the linear equation in (22) is consistent and the null space $\mathcal{N}(\boldsymbol{\Psi})$ has been established in Proposition 2 and 3, we are ready to answer the third question:

Theorem 4: The general solution form of (22) is

$$\mathbf{f}^\star = \boldsymbol{\Psi}^\dagger \mathbf{b} + \sum_{k=1}^{K} \mathbf{f}_{(m_k)}^\perp, \quad (24)$$

in which $\{m_1, \ldots, m_K\}$ is the set of all indexes such that $\lambda^\star w_{m_k} = 0$, and $\mathbf{f}_{(m_k)}^\perp = \operatorname{col}(\mathbf{0}, \ldots, \mathbf{f}_{m_k}^\perp, \ldots, \mathbf{0})$ satisfies $(\mathbf{I}_{N_R} \otimes \mathbf{Q}\mathbf{G}_{m_k})\mathbf{f}_{m_k}^\perp = \mathbf{0}$. The $K + 1$ terms in (24) are orthogonal to each other. The first term, $\boldsymbol{\Psi}^\dagger \mathbf{b}$, is the solution that minimizes the transmit power of each relay simultaneously.

Proof: Since $\mathbf{b} \in \mathcal{R}(\boldsymbol{\Psi})$ and $\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger$ is a projection matrix onto $\mathcal{R}(\boldsymbol{\Psi})$, we have $\boldsymbol{\Psi}(\boldsymbol{\Psi}^\dagger \mathbf{b}) = \mathbf{b}$. The general form in (24) follows immediately because $\mathcal{N}(\boldsymbol{\Psi}) = \mathcal{F}_{m_1} \oplus \cdots \oplus \mathcal{F}_{m_K}$ (*cf.* Proposition 3). Since $\mathcal{R}(\boldsymbol{\Psi})$, $\mathcal{F}_{m_1}, \ldots, \mathcal{F}_{m_K}$ are mutually orthogonal, the $K + 1$ terms in (24) are also orthogonal to each other. Let $\mathbf{f}^\star = \operatorname{col}(\mathbf{f}_1^\star, \ldots, \mathbf{f}_M^\star)$. For those k satisfying $\lambda^\star w_k > 0$, $\mathbf{f}_k^\star$ is unique; for those $k \in \{m_1, \ldots, m_K\}$, the transmit power of the kth relay is

$$\left\| \mathbf{I}_{(k)}\mathbf{f}^\star \right\|_2^2 = \left\| \mathbf{I}_{(k)}\boldsymbol{\Psi}^\dagger \mathbf{b} \right\|_2^2 + \left\| \mathbf{f}_{(k)}^\perp \right\|_2^2,$$

which can only be minimized by setting $\mathbf{f}_{(k)}^\perp = \mathbf{0}$. Therefore among all the solutions, $\boldsymbol{\Psi}^\dagger \mathbf{b}$ minimizes the transmit power of each relay simultaneously. ∎

It appears that immediately after Theorem 1, we could have applied the pseudo-inverse to obtain the solution. This approach, however, would only guarantee that among all the solutions of (22), $\boldsymbol{\Psi}^\dagger \mathbf{b}$ minimizes the *sum* power of the relays,

$\|\mathbf{f}^\star\|_2^2$. With the help of Propositions 2 and 3, we were able to prove a stronger conclusion: $\boldsymbol{\Psi}^\dagger \mathbf{b}$ minimizes the power of *each* relay *simultaneously*. Moreover, if $\mathbf{f}_k$ were defined as $\operatorname{vec}(\mathbf{F}_k)$ instead of $\operatorname{vec}(\mathbf{F}_k \mathbf{R}_k^{1/2})$, taking the pseudo inverse directly would not even minimize the sum power. This is because for Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$ (with $\mathbf{B}$ nonsingular), $(\mathbf{B}\mathbf{A}\mathbf{B})^\dagger \neq \mathbf{B}^{-1}\mathbf{A}^\dagger \mathbf{B}^{-1}$ except under special situations.

The main drawback of computing $\mathbf{f}^\star$ from (24) is that the dimension of $\boldsymbol{\Psi}$, $MN_R^2 \times MN_R^2$, is larger than those of the original matrices. We can use the relationships in (12) to simplify (24) as in the following corollary:

Corollary 5: The minimum-norm solution of (22) can be expressed in an alternative form:

$$\mathbf{f}_k^\star = \left(\mathbf{S}_k + \lambda^\star w_k \mathbf{I}_{N_R^2}\right)^\dagger \mathbf{T}_k \boldsymbol{\Sigma}^{-1} \operatorname{vec}(\mathbf{I}_{N_S}), \quad (25)$$

where $\boldsymbol{\Sigma} \in \mathbb{C}^{N_S^2 \times N_S^2}$ is defined as

$$\boldsymbol{\Sigma} \triangleq \mathbf{R}_s^{-T} \otimes \mathbf{I}_{N_S} + \sum_{l=1}^{M} \mathbf{T}_l^H \left(\mathbf{S}_l + \lambda^\star w_l \mathbf{I}_{N_R^2}\right)^\dagger \mathbf{T}_l. \quad (26)$$

Proof: See Appendix A. ∎

C. Optimal Solution

We now return to the KKT conditions in (21). The complementary slackness in (21d) indicates that either $\lambda^\star = 0$ and the constraint (21b) is inactive, or $\lambda^\star > 0$ and the constraint is tightly satisfied. Define the weighted sum power as a function of λ, *viz.*,

$$g(\lambda) \triangleq \mathbf{b}^H \boldsymbol{\Psi}(\lambda)^\dagger \mathbf{I}_{\text{sum}} \boldsymbol{\Psi}(\lambda)^\dagger \mathbf{b}, \quad (27)$$

where the explicit argument for $\boldsymbol{\Psi}$ is used to emphasize its dependence on λ. If $0 \leq g(0) \leq P_r$, the unconstrained solution satisfies the weighted sum power constraint (21d) and $\lambda^\star = 0$. Otherwise, $g(0) > P_r$ and $\lambda^\star > 0$ should be the implicit solution to the nonlinear equation

$$g(\lambda) = P_r. \quad (28)$$

The following proposition justifies the uniqueness of $\lambda^\star$:

Proposition 6: If $g(0) > 0$, $g(\lambda)$ is a monotonically decreasing function of $\lambda > 0$ with $\lim_{\lambda \to \infty} g(\lambda) = 0$.

Proof: See Appendix B. ∎

Up to now, the optimal solution has been expressed in closed forms (24) or (25), but the dual variable $\lambda^\star$ does not have an explicit formula. Numerical methods such as bisection or Newton's method [20] are necessary to solve the nonlinear equation (28). This, in fact, can be improved by allowing a complex scaling in the equalizer $\mathbf{Q}$. That is, we consider the set $\{\eta^{-1}e^{-j\phi}\mathbf{Q} | \eta > 0, 0 \leq \phi < 2\pi\}$, in which each member is a complex-scaled version of $\mathbf{Q}$. For different (η, ϕ), the optimal $\lambda^\star$ and $\mathbf{f}^\star$, and the corresponding minimum value of the MSE in (13) are also different. We are interested in a single (η_o, ϕ_o) leading to the smallest minimum MSE, so that any other member in the set can be replaced by $\eta_o^{-1}e^{-j\phi_o}\mathbf{Q}$. Interestingly, for this special (η_o, ϕ_o), the optimal $\lambda^\star$, $\mathbf{f}^\star$ and the minimum MSE always have *explicit* formulas, as shown in the following theorem:

Theorem 7: Any equalizer $\mathbf{Q}$ can be replaced by a complex-scaled version $\eta_o^{-1}e^{-j\phi}\mathbf{Q}$ so that:

1) The optimal solution is

$$\mathbf{f}^\star = \eta_o e^{j\phi}(\mathbf{\Phi} + \theta\mathbf{I}_{\text{sum}})^\dagger \mathbf{b}, \tag{29}$$

where $\theta \triangleq \text{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)/P_r$, ϕ can be an arbitrary number in $[0, 2\pi)$, and $\eta_o > 0$ is the unique number satisfying $\mathbf{f}^{\star H}\mathbf{I}_{\text{sum}}\mathbf{f}^\star = P_r$, that is,

$$\eta_o = \sqrt{P_r/\mathbf{b}^H(\mathbf{\Phi} + \theta\mathbf{I}_{\text{sum}})^\dagger\mathbf{I}_{\text{sum}}(\mathbf{\Phi} + \theta\mathbf{I}_{\text{sum}})^\dagger\mathbf{b}}. \tag{30}$$

The optimal duality parameter is $\lambda_o^\star = \theta\eta_o^{-2}$.

2) The minimum MSE with the equalizer $\eta_o^{-1}e^{-j\phi}\mathbf{Q}$, i.e.,

$$\text{MSE}_{\min} = \text{tr}(\mathbf{R}_s) - \mathbf{b}^H(\mathbf{\Phi} + \theta\mathbf{I}_{\text{sum}})^\dagger\mathbf{b} \tag{31a}$$

$$= \text{vec}(\mathbf{I})^H\mathbf{\Sigma}^{-1}\text{vec}(\mathbf{I}), \tag{31b}$$

is always smaller than or equal to that with any other scaled equalizer $\eta^{-1}e^{-j\phi}\mathbf{Q}$ ($\eta > 0, 0 \leq \phi < 2\pi$), including $\mathbf{Q}$ itself.

Proof: With the equalizer $\eta^{-1}e^{-j\phi}\mathbf{Q}$, we rewrite the MSE function in (13) as

$$\text{MSE}(\mathbf{f}, \mathbf{Q}, \eta, \phi) = \eta^{-2}\mathbf{f}^{\star H}\mathbf{\Phi}\mathbf{f}^\star - \eta^{-1}e^{j\phi}\mathbf{f}^{\star H}\mathbf{b}$$
$$-\eta^{-1}e^{-j\phi}\mathbf{b}^H\mathbf{f}^\star + \text{tr}(\mathbf{R}_s) + \eta^{-2}\text{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H). \tag{32}$$

The minimum-norm solution is obtained by replacing $\mathbf{Q}$ in (24) with $\eta^{-1}e^{-j\phi}\mathbf{Q}$:

$$\mathbf{f}^\star = \eta e^{j\phi}(\mathbf{\Phi} + \lambda^\star\eta^2\mathbf{I}_{\text{sum}})^\dagger\mathbf{b}. \tag{33}$$

The duality parameter $\lambda^\star$ should satisfy the KKT conditions in (21b), (21c) and (21d). If the unconstrained solution satisfies

$$\mathbf{f}^{\star H}\mathbf{I}_{\text{sum}}\mathbf{f}^\star|_{\lambda^\star=0} = \eta^2\mathbf{b}^H\mathbf{\Phi}^\dagger\mathbf{I}_{\text{sum}}\mathbf{\Phi}^\dagger\mathbf{b} \leq P_r, \tag{34}$$

or equivalently, $\eta \leq \eta_c \triangleq \sqrt{P_r/(\mathbf{b}^H\mathbf{\Phi}^\dagger\mathbf{I}_{\text{sum}}\mathbf{\Phi}^\dagger\mathbf{b})}$, the constraint (21b) is inactive, i.e., $\lambda^\star = 0$. Substituting (33) into the MSE expression in (32), we have

$$\text{MSE}_{\min}(\eta) = \text{tr}(\mathbf{R}_s) - \mathbf{b}^H\mathbf{\Phi}^\dagger\mathbf{b} + \eta^{-2}\text{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H). \tag{35}$$

If $\eta \geq \eta_c$, the constraint (21b) is tightly satisfied:

$$\eta^2\mathbf{b}^H(\mathbf{\Phi} + \lambda^\star\eta^2\mathbf{I}_{\text{sum}})^\dagger\mathbf{I}_{\text{sum}}(\mathbf{\Phi} + \lambda^\star\eta^2\mathbf{I}_{\text{sum}})^\dagger\mathbf{b} = P_r, \tag{36}$$

through which $\lambda^\star$ is an implicit function of η. Substituting the optimal solution in (33) into the MSE expression in (32) and using (36) to replace η^{-2}, we have

$$\text{MSE}_{\min}(\eta) = \text{tr}(\mathbf{R}_s) - \mathbf{b}^H(\mathbf{\Phi} + \lambda^\star\eta^2\mathbf{I}_{\text{sum}})^\dagger$$
$$\times(\mathbf{\Phi} + (2\lambda^\star\eta^2 - \theta)\mathbf{I}_{\text{sum}})(\mathbf{\Phi} + \lambda^\star\eta^2\mathbf{I}_{\text{sum}})^\dagger\mathbf{b}. \tag{37}$$

Up to now, the minimum MSE, which does not depend on ϕ, has been expressed as a function of η. From (35), $\text{MSE}_{\min}(\eta)$ is a monotonically decreasing function of η in the interval $[0, \eta_c]$ and therefore can only be minimized when $\eta \geq \eta_c$. Although $\lambda^\star$ does not have an explicit formula, the minimum MSE in (37) depends only on the product $\lambda^\star\eta^2 \triangleq \gamma$, which can take all nonnegative values. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be a set of orthonormal basis vectors for $\mathcal{N}(\mathbf{\Phi})$ and define $\mathbf{\Psi}_e \triangleq \mathbf{\Phi} + \gamma\mathbf{I}_{\text{sum}} + \sum_{k=1}^p \mathbf{u}_k\mathbf{u}_k^H$.

Using the same tactics as in the proof of Proposition 6, we get the derivative of $\text{MSE}_{\min}(\gamma)$ as

$$\frac{\text{dMSE}_{\min}(\gamma)}{\text{d}\gamma} = 2(\gamma - \theta)\mathbf{b}^H\mathbf{\Psi}_e^{-1}I_{\text{sum}}\mathbf{\Psi}_e^{-1}I_{\text{sum}}\mathbf{\Psi}_e^{-1}\mathbf{b}. \tag{38}$$

Obviously, $\text{MSE}_{\min}(\gamma)$ is monotonically decreasing if $0 \leq \gamma < \theta$, and monotonically increasing if $\gamma > \theta$. Therefore, $\gamma_o = \theta = \text{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)/P_r$ is the unique solution to minimize (37) and it is straightforward to derive (29) and (31). ∎

We may visualize η^2 as a target signal power level at the destination and η^{-1} as an automatic gain control factor. For $\eta \leq \eta_c$, the power budget at the relays is sufficient to support the unconstrained optimal solution ($\lambda^\star = 0$). As seen in (35), the first part, $\text{tr}(\mathbf{R}_s) - \mathbf{b}^H\mathbf{\Phi}^\dagger\mathbf{b}$, does not depend on η. The second part, $\eta^{-2}\text{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)$, decreases monotonically as a function of $\eta \leq \eta_c$, indicating weaker effects of the noise term $\mathbf{n}$ in (3) on the decoding process. Once η exceeds the threshold η_c, the power budget becomes insufficient and therefore the power regularization term $\lambda^\star\eta^2\mathbf{I}_{\text{sum}}$ is introduced. This slightly increases the first part of (32) (the first four terms), but the overall MSE still decreases because the second part $\eta^{-2}\text{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)$ is reduced by more. Nonetheless, there is a critical and therefore optimal η_o above which the latter cannot completely compensate for the former any more.

An alternative formulation is to introduce η as early as in the definition of the objective function in (6), which was used before for other relaying systems [12], [29], [30]. In this case, the objective function would be a convex function of $\mathbf{f}$, but not of both η and $\mathbf{f}$. Therefore, it does not formally guarantee optimality to set to zero the partial derivatives with respect to both η and $\mathbf{f}$. We also note that if $M = 1$ (a single relay), the optimal relaying matrix in (29) would be in agreement with the result in [30].

IV. PER-RELAY POWER CONSTRAINTS

Due to practical reasons such as the dynamic range of power amplifiers, it may sometimes be more appropriate to consider the per-relay power constraints. In this section, we study the optimality conditions and propose a power balancing algorithm to compute the optimal solution. Our analysis provides some insights into the power usage at the relays.

A. KKT Conditions and the Optimal Solution

The Lagrangian function for the relay optimization problem with the per-relay power constraints in (16) is given by

$$L(\mathbf{f}, \boldsymbol{\lambda}) = \text{MSE}(\mathbf{f}) + \sum_{k=1}^M \lambda_k \left(\mathbf{f}^H\mathbf{I}_{(k)}\mathbf{f} - P_k\right), \tag{39}$$

where $\boldsymbol{\lambda} \triangleq \text{col}(\lambda_1, \ldots, \lambda_M)$. By comparing with (18), we note that many results for the weighted sum power constraint extend to the per-relay power constraints, simply by replacing λw_k with λ_k. Subsequently, we skip the details to focus on presenting the main results. Redefine $\mathbf{\Psi} \triangleq \mathbf{\Phi} + \sum_{k=1}^M \lambda_k^\star\mathbf{I}_{(k)}$. The optimal $\mathbf{f}^\star$ and its dual-optimal variables $\lambda_k^\star, 1 \leq k \leq M$, satisfy the following KKT conditions:

$$\nabla_{\bar{\mathbf{f}}}L(\mathbf{f}, \boldsymbol{\lambda}^\star)|_{\mathbf{f}=\mathbf{f}^\star} = \mathbf{\Psi}\mathbf{f}^\star - \mathbf{b} = \mathbf{0}, \tag{40a}$$

$$\mathbf{f}^{\star H}\mathbf{I}_{(k)}\mathbf{f}^\star \leq P_k, \tag{40b}$$

$$\lambda_k^\star \geq 0, \tag{40c}$$

$$\lambda_k^\star \left(\mathbf{f}^{\star H} \mathbf{I}_{(k)} \mathbf{f}^\star - P_k \right) = 0, \tag{40d}$$

for all $1 \leq k \leq M$. Akin to Theorem 4 and Corollary 5, the minimum-norm solution to (40a) is $\mathbf{f}^\star = \boldsymbol{\Psi}^\dagger \mathbf{b}$ with an alternative form $\mathbf{f}_k^\star = (\mathbf{S}_k + \lambda_k^\star \mathbf{I}_{N_R^2})^\dagger \mathbf{T}_k \boldsymbol{\Sigma}^{-1} \text{vec}(\mathbf{I}_{N_S})$, where $\boldsymbol{\Sigma} \triangleq \mathbf{R_s}^{-T} \otimes \mathbf{I}_{N_S} + \sum_{l=1}^M \mathbf{T}_l^H (\mathbf{S}_l + \lambda_l^\star \mathbf{I}_{N_R^2})^\dagger \mathbf{T}_l$.

The only difference from the weighted sum power case is the existence of *multiple* dual variables and complementary slackness conditions. This requires algorithms that are more sophisticated than bisection or Newton's method, e.g., interior-point methods such as the path-following and the primal-dual methods [20], [28]. Software packages for these algorithms, e.g., Gurobi, CPLEX and SeDuMi, are available and can be used in Matlab via YALMIP [31] or CVX [32].

Here, it is reasonable to consider the same complex-scaled equalizer $\eta^{-1} e^{-j\phi} \mathbf{Q}$ as in Section III-C, which results in the following theorem:

Theorem 8: Any equalizer $\mathbf{Q}$ can be replaced by a complex-scaled version $\eta_o^{-1} e^{-j\phi} \mathbf{Q}$ so that:

1) The optimal solution is

$$\mathbf{f}^\star = \eta_o \left(\boldsymbol{\Phi} + \sum_{k=1}^M \gamma_k^o \mathbf{I}_{(k)} \right)^\dagger \mathbf{b}, \tag{41}$$

where $\phi \in [0, 2\pi)$, $\gamma_k^o \triangleq \lambda_k^\star \eta_o^2$ ($1 \leq k \leq M$) satisfies

$$\sum_{k=1}^M \gamma_k^o P_k = \text{tr}(\mathbf{Q} \mathbf{R}_n \mathbf{Q}^H) \tag{42}$$

and $\eta_o > 0$ is the unique positive number such that $\gamma_k^o (\mathbf{f}^{\star H} \mathbf{I}_{(k)} \mathbf{f}^\star - P_k) = 0$ for $1 \leq k \leq M$.

2) The minimum MSE with the equalizer $\eta_o^{-1} e^{-j\phi} \mathbf{Q}$ is

$$\text{MSE}_{\min}(\eta_o) = \text{tr}(\mathbf{R}_s) - \mathbf{b}^H \left(\boldsymbol{\Phi} + \sum_{k=1}^M \gamma_k^o \mathbf{I}_{(k)} \right)^\dagger \mathbf{b}. \tag{43}$$

Proof: For any $\eta > 0$ and $0 \leq \phi < 2\pi$, the optimal solution is $\mathbf{f}^\star = \eta e^{j\phi} \boldsymbol{\Psi}^\dagger \mathbf{b}$, where $\boldsymbol{\Psi} \triangleq \boldsymbol{\Phi} + \sum_{k=1}^M \lambda_k^\star \eta^2 \mathbf{I}_{(k)}$, and the dual-optimal variables $\lambda_1^\star, \ldots, \lambda_M^\star$ are implicit functions of η through (40b)–(40d). The minimum MSE is also a function of $\eta > 0$, *viz.*,

$$\text{MSE}_{\min}(\eta) = \text{tr}(\mathbf{R}_s) - \mathbf{b}^H \boldsymbol{\Psi}^\dagger \mathbf{b}$$
$$+ \eta^{-2} \text{tr}(\mathbf{Q} \mathbf{R}_n \mathbf{Q}^H) - \sum_{k=1}^M \lambda_k^\star P_k. \tag{44}$$

Next, we minimize (44) over $\eta > 0$. With similar argument to that in Section III-C, there exists an η_c such that for all $\eta \leq \eta_c$, $\lambda_1^\star = \cdots = \lambda_M^\star = 0$. As a result, $\text{MSE}_{\min}(\eta)$ is a monotonically decreasing function in $(0, \eta_c]$ according to (44). This means that (44) is minimized only when $\eta > \eta_c$. In this situation, at least one of the dual variables is nonzero and hence $\sum_{k=1}^M \lambda_k^\star P_k \neq 0$. For convenience, rewrite the complementary slackness from (40d) as

$$\lambda_k^\star P_k = \lambda_k^\star \eta^2 \mathbf{b}^H \boldsymbol{\Psi}^{-1} \mathbf{I}_{(k)} \boldsymbol{\Psi}^{-1} \mathbf{I}. \tag{45}$$

Adding from $k = 1$ to M, we have

$$\eta^{-2} = \frac{\sum_{k=1}^M \lambda_k^\star \mathbf{b}^H \boldsymbol{\Psi}^{-1} \mathbf{I}_{(k)} \boldsymbol{\Psi}^{-1} \mathbf{b}}{\sum_{k=1}^M \lambda_k^\star P_k}. \tag{46}$$

Substituting (45) and (46) into (44) and using similar techniques to those in the proof of Theorem 7, we can prove that the optimal η_o, the corresponding $\lambda_k^\star$ and $\gamma_k^o = \lambda_k^\star \eta_o^2$ satisfy (42), and the minimum MSE takes the form of (44). $\blacksquare$

Equation (42) provides an elegant relationship between different γ_k^o. These parameters serve as regularization terms that control the transmit power of the relays. When the power budget P_k is higher, the required regularization tends to be lower and the value of γ_k^o tends to be smaller. Although complex scaling does not lead to closed-form expressions for the *individual* parameters $\{\gamma_k^o\}$, the results in (42), (41) and (43) lay the foundation for the power balancing algorithm proposed in the next subsection.

B. Power Balancing

The optimal solutions, (29) under the weighted sum power constraint and (41) under the per-relay power constraints, share some common structure. In particular, they are identical if the weights $\{w_k\}$ and P_r are chosen to satisfy

$$w_k \theta = w_k \frac{\text{tr}(\mathbf{Q} \mathbf{R}_n \mathbf{Q}^H)}{P_r} = \gamma_k^o \quad \text{and} \quad P_r = \sum_{k=1}^M w_k P_k. \tag{47}$$

The minimum MSE, (31) and (43), would also be equal. In other words, if we know this *equivalent* weighted sum power constraint, the optimal solution is immediately available from Theorem 7. Of course, scaling all w_k simultaneously by a common positive factor does not alter the optimal solution.

With this in mind, we propose a power balancing algorithm which finds these weights iteratively. The initial weights are all set to 1. In each iteration, the algorithm computes the optimal relaying matrices with the previous weights, compares the actual power of the relays with the per-relay power constraints, and adjusts the weights accordingly. If the actual power of the kth relay is higher than P_k, the weight w_k is increased and vice versa. The algorithm stops when all the constraints are satisfied. These steps are summarized in Algorithm 1.

Algorithm 1: Power Balancing

Initiate the counter $m = -1$;

Initiate the weights: $w_1(0) = \cdots = w_M(0) = 1$;

repeat

    Add the counter $m \leftarrow m + 1$;

    Compute the weighted sum power:

$$P_r(m) = \sum_{k=1}^M w_k(m) P_k; \tag{48}$$

    Compute the optimal $\mathbf{f}^\star(m)$ from (29) or (25);

    Compute $P_k(m) = \mathbf{f}^\star(m)^H \mathbf{I}_{(k)} \mathbf{f}^\star(m)$, $1 \leq k \leq M$;

    Update the weights for $1 \leq k \leq M$:

$$w_k(m+1) \leftarrow w_k(m) \frac{P_k(m)}{P_k}; \tag{49}$$

until $\max(P_1(m)/P_1, \ldots, P_M(m)/P_M) \leq 1$.

C. Remarks on Power Usage

With the per-relay power constraints, the optimal relaying strategy may not use the maximum power at some relays. We show this for a simplified single-antenna case but our analysis also captures the essence of multi-antenna systems. The matrices/vectors in the signal model in Section II become scalars, represented by the corresponding lowercase italic letters. For convenience, we assume g_k, f_k and h_k ($1 \leq k \leq M$) are all real positive numbers and extension to the complex scenarios is straightforward.

Define $\alpha \triangleq \sum_{l=1,l \neq k}^{M} g_l f_l h_l$ and $w \triangleq \sum_{l=1,l \neq k}^{M} g_l f_l w_l$ with $\operatorname{var}(w) = \sigma_w^2$. Without the kth relay, the signal received by the destination would be

$$r = \alpha s + w + n \tag{50}$$

and the SNR at the destination would be $\mathrm{SNR}_d \triangleq \alpha^2 \sigma_s^2 / (\sigma_w^2 + \sigma_n^2)$. With the kth relay, the signal would be

$$r' = (\alpha + g_k f_k h_k)s + g_k f_k w_k + w + n \tag{51}$$

and the SNR would be a function of f_k

$$\mathrm{SNR}_d' = \mathrm{SNR}_d \frac{(1 + g_k f_k h_k / \alpha)^2}{1 + g_k^2 f_k^2 \sigma_k^2 / (\sigma_w^2 + \sigma_n^2)}, \tag{52}$$

where the second operand ($\triangleq \kappa(f_k)$) is the *gain or penalty* due to the kth relay, depending on whether it is larger than or smaller than 1.

By taking the first-order derivative, we know that $\kappa(f_k)$ is strictly monotonically increasing when

$$0 \leq f_k < f_o \triangleq \frac{h_k (\sigma_w^2 + \sigma_n^2)}{\alpha g_k \sigma_k^2} \tag{53}$$

and strictly monotonically decreasing when $f_k > f_o$. As $f_k \to \infty$, the limit would be

$$\kappa(\infty) \triangleq \lim_{f_k \to \infty} \kappa(f_k) = \frac{h_k^2 \sigma_s^2 / \sigma_k^2}{\alpha^2 \sigma_s^2 / (\sigma_w^2 + \sigma_n^2)} = \frac{\mathrm{SNR}_r}{\mathrm{SNR}_d}, \tag{54}$$

where $\mathrm{SNR}_r \triangleq h_k^2 \sigma_s^2 / \sigma_k^2$ is the SNR at the kth relay. Since the power transmitted by the relay is proportional to f_k^2, the above properties of $\kappa(f_k)$ lead to several interesting conclusions:

1) If $\mathrm{SNR}_r \geq \mathrm{SNR}_d$, $\kappa(\infty) \geq 1$ according to (54). Since $\kappa(0) = 1$, and $\kappa(f_k)$ increases in $[0, f_o]$ and decreases in $[f_o, \infty]$, $\kappa(f_k) > 1$ always holds in $(0, \infty)$, implying that the system always benefits from the use of the kth relay, no matter how much power the relay transmits. However, it is *not necessarily* better to use more power. Any $f_k > f_o$ would not be as good as f_o.

2) If $\mathrm{SNR}_r < \mathrm{SNR}_d$, $\kappa(\infty) < 1$ and there exists an f_c such that $\kappa(f_c) = 1$. Hence, $\kappa(f_k) > 1$ in the interval $(0, f_c)$, which means that the kth relay can still contribute to the SNR at the destination as long as it reduces its transmit power to a level low enough. The reason for this is that the signal components are added coherently, whereas the noise components are not.

3) If the value of f_k corresponding to the power constraint P_k falls in the interval $[0, f_o]$, the relay would use the maximum amount of power; otherwise, it would use only a portion ($f_k = f_o$). It is not justified to turn off a relay completely.

In practice, it is a waste of resources if a relay transmits only a small amount of power. Thanks to the randomness of the channels and users, this problem with the ideal narrowband configuration is probably not as important in practice. Firstly, most modern communication systems are based on a multicarrier scheme such as OFDM. A relay station may transmit less power on one subcarrier but more on another, so that the variation in the total power is smaller. Secondly, the multiple relays are simultaneously serving several randomly located users (with different subcarriers or time intervals). This will further reduce the disparity between the transmit power of different relays. Lastly, if the expected transmit power of a particular relay is abnormally small, the problem likely comes from inappropriate network layout and the relays should be relocated instead.

V. THE OPTIMAL EQUALIZER

In this section, we consider the joint design of the MIMO equalizer $\mathbf{Q}$ and the relaying matrices (or equivalently $\mathbf{f}$), under the weighted sum power constraint. For any $\mathbf{Q}$, the optimal $\mathbf{f}^\star$ is in the form of (29); for any $\mathbf{f}$ (or $\mathbf{F}$), the optimal equalizer $\mathbf{Q}^\star$ is the MMSE equalizer

$$\mathbf{Q}^\star = \mathbf{R}_{rs}^H \mathbf{R}_r^{-1}, \tag{55}$$

where $\mathbf{R}_{rs} \triangleq \mathrm{E}\{\mathbf{r}\mathbf{s}^H\} = \mathbf{GFHR}_s$ and $\mathbf{R}_r \triangleq \mathrm{E}\{\mathbf{r}\mathbf{r}^H\}$.

This observation suggests a block coordinate descent method. The algorithm starts from an initial $\mathbf{Q}_0$ and repeats the following steps: it first updates $\mathbf{f}$ using (29) while fixing $\mathbf{Q}$, and then calculates $\mathbf{Q}$ as in (55) while holding $\mathbf{f}$ constant. Thanks to the optimality in each step, the (bounded) sequence of MSE values is monotonically non-increasing, which must converge. As a result, the block coordinate descent algorithm is guaranteed to converge to a local optima. This idea is widely used in the literature such as [18], [33]. We also note that the design of a precoder is also possible through this framework.

The other approach is to consider the joint design as a two-step process. The first step is to design $\mathbf{f}$ as a function of $\mathbf{Q}$, which is what we have done in Section III. After this, the second step is to optimize $\mathbf{Q}$ to further minimize the MSE in (31). This approach handles the constraints in the convex problem (the first step), so that the remaining problem, though still non-convex, is an unconstrained one.[2] The line search algorithms are readily applicable to find a local minima. Beginning with $\mathbf{Q}_0$, these methods generate a sequence of iterates $\{\mathbf{Q}_n\}_{n=0}^{\infty}$ until a solution has been approximated with sufficient accuracy. Specifically, these algorithms choose a direction $\Delta\mathbf{Q}_n$ and search along this direction from the current $\mathbf{Q}_n$ for a new iterate $\mathbf{Q}_{n+1}$ with a lower MSE value. The distance to move along $\mathbf{Q}_n$ should satisfy criteria such as Wolfe's conditions [20]. In particular, the steepest descent method uses the opposite direction of the gradient (see Appendix C for derivation), *viz.*,

$$\begin{aligned}
\Delta\mathbf{Q}_n &= -\nabla_{\bar{\mathbf{Q}}}\mathrm{MSE}_{\min}|_{\mathbf{Q}=\mathbf{Q}_n} \\
&= -\eta_n^{-2}\mathbf{Q}_n\left(\mathbf{GF}_{(n)}\mathbf{R}_w\mathbf{F}_{(n)}^H\mathbf{G}^H + \mathbf{R}_n\right) \\
&\quad + \eta_n^{-1}\mathrm{unvec}\left(\boldsymbol{\Sigma}_n^{-1}\mathrm{vec}(\mathbf{I})\right)\mathbf{H}^H\mathbf{F}_{(n)}^H\mathbf{G}^H, \tag{56}
\end{aligned}$$

---

[2]An alternative, usually more popular, approach is to first set the equalizer $\mathbf{Q}$ as the MMSE equalizer. After substituting this optimal $\mathbf{Q}$ into (6), the MSE becomes a function of the matrices $\{\mathbf{F}_k\}$. The resulting problem is, however, not only non-convex but also with constraints.

where $\mathbf{F}_{(n)}$ is the optimal (block-diagonal) relaying matrix under $\mathbf{Q}_n$, and $\boldsymbol{\Sigma}_n$ and η_n are the corresponding intermediate matrices/variables when computing $\mathbf{F}_{(n)}$, *cf.* (5), (9) and (29). The steepest descent method is summarized in Algorithm 2.

Algorithm 2: Steepest Descent

Choose the weights w_k, $1 \leq k \leq M$, P_r and $\epsilon > 0$;

Choose $\bar{\alpha} > 0$, $\rho, c \in (0, 1)$; {Line search parameters.}

Initiate the counter $n = -1$ and the equalizer $\mathbf{Q}_0$;

repeat

    Increment counter $n \leftarrow n + 1$;

    Compute $\Delta \mathbf{Q}_n$ from (56);

    Set $\alpha \leftarrow \bar{\alpha}$;

    repeat

        $\alpha \leftarrow \rho \alpha$; {Backtracking line search.}

    until $\text{MSE}_{\min}(\mathbf{Q}_n + \alpha \Delta \mathbf{Q}_n) \leq \text{MSE}_{\min}(\mathbf{Q}_n) - c\alpha \|\Delta \mathbf{Q}_n\|_{\text{F}}^2$;

    Update $\mathbf{Q}_{n+1} \leftarrow \mathbf{Q}_n + \alpha \Delta \mathbf{Q}_n$;

until $\|\Delta \mathbf{Q}_n\|_{\text{F}}^2 < \epsilon$.

As for the per-relay power constraints, Algorithm 1 is still applicable except that in each iteration, $\mathbf{Q}$ is updated along with $\mathbf{f}$ using the above methods.

VI. IMPLEMENTATION ISSUES AND COMPLEXITY

We discuss implementation issues for the proposed algorithms, including the requirements on communication and computing resources. Firstly, an important feature of the optimal methods is that they only require local CSI knowledge and a little additional shared information. As seen from (25), all the kth relay needs to know, in addition to its own backward channel $\mathbf{H}_k$ and forward channel $\mathbf{G}_k$, is the vector $\boldsymbol{\Sigma}^{-1}\text{vec}(\mathbf{I})$ (of size $N_S^2 \times 1$). Thanks to this attractive feature, these methods naturally lend themselves to distributed implementations. One possible implementation scheme is that a fusion center computes this vector and feeds it back to the relays via broadcasting. An alternative way is to compute the relaying matrices also at the fusion center. The fusion center can be the destination or one of the relays. Some dedicated resources are required but for a small number of relays (say 2 or 3), the overall complexity will be manageable.

Secondly, the existence of closed-form expressions such as (25) results in relatively low computational complexity. The major computing task consists of two parts: evaluating $\boldsymbol{\Sigma}^{-1}\text{vec}(\mathbf{I})$ and $\{\mathbf{f}_k\}$. The power balancing algorithm in Algorithm 1 for the per-relay constraints and the algorithms for the equalizer $\mathbf{Q}$ (*cf.* Section V) are primarily composed of repetitions of these operations. Although vectorization increases the dimensions of the matrices and vectors in the system model (*cf.* (11)), the complexity is not notably higher thanks to the properties of the SVDs and eigenvalue decompositions (EVDs) for Kronecker products. For example, we have

$$\left.\begin{array}{l} \mathbf{X} = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^H \\ \mathbf{Y} = \mathbf{U}_2 \mathbf{S}_2 \mathbf{V}_2^H \end{array}\right\}$$
$$\Rightarrow \mathbf{X} \otimes \mathbf{Y} = (\mathbf{U}_1 \otimes \mathbf{U}_2)(\mathbf{S}_1 \otimes \mathbf{S}_2)(\mathbf{V}_1 \otimes \mathbf{V}_2)^H,$$

which is essentially the SVD of $\mathbf{X} \otimes \mathbf{Y}$ except that the singular values are not sorted in descending order. A similar property holds for the EVD. Consequently, the EVD of the pseudo inverse $(\mathbf{S}_k + \lambda w_k \mathbf{I})^\dagger = \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^H$ can be obtained based on those of $(\mathbf{R}_{\mathbf{x}_k}^{-1/2} \mathbf{R}_{\mathbf{w}_k} \mathbf{R}_{\mathbf{x}_k}^{-1/2})^T$ and $\mathbf{G}_k^H \mathbf{Q}^H \mathbf{Q} \mathbf{G}_k$. Then, two matrix multiplications (not including those involving diagonal matrices) are needed to compute $\mathbf{T}_k^H (\mathbf{S}_k + \lambda w_k \mathbf{I})^\dagger \mathbf{T}_k = (\mathbf{T}_k^H \mathbf{U}_k) \boldsymbol{\Lambda}_k (\mathbf{T}_k^H \mathbf{U}_k)^H$. One additional matrix multiplication is necessary to compute $(\mathbf{S}_k + \lambda w_k \mathbf{I})^\dagger \mathbf{T}_k = \mathbf{U}_k \boldsymbol{\Lambda}_k (\mathbf{T}_k^H \mathbf{U}_k)^H$. Subsequently, we compute the sum matrix $\boldsymbol{\Sigma}$ and solve the linear equation $\boldsymbol{\Sigma}\mathbf{x} = \text{vec}(\mathbf{I})$ to get $\boldsymbol{\Sigma}^{-1}\text{vec}(\mathbf{I})$. In the end, $\mathbf{f}_k$ is obtained from (25), which requires only matrix-vector multiplications. In summary, the major operations include

- $2M$ EVDs of matrices with dimension $N_R \times N_R$;
- $3M$ matrix multiplications involving matrices of dimension $N_S^2 \times N_R^2$, $N_R^2 \times N_R^2$ or $N_R^2 \times N_S^2$.
- solving one linear equation of size $N_S^2 \times N_S^2$.

Thirdly, the proposed algorithms in Sections III and IV minimize the MSE over not only the relaying vector $\mathbf{f}$, but also the scaling factor $\eta e^{j\phi}$. In contrast, the interior-point methods can merely optimize $\mathbf{f}$ for a single η because the problem is not convex if $\mathbf{f}$ and η are simultaneously considered. To obtain the same result as our algorithms do, the interior-point methods have to run for different $\eta > 0$, which further increases their complexity.

In summary, the system complexity is well manageable for multi-antenna systems if the number of relays is small. The benefits brought by a three-relay configuration can be remarkable as shown by the simulation results in Section VII and previous publications such as [7], [11].

VII. NUMERICAL RESULTS

In this section, we first investigate the effects of channel gains on the power allocation among relays under the weighted sum power constraint. Next, we verify the convergence behaviors of the proposed iterative algorithms, including power balancing for the per-relay power constraints, block coordinate descent and steepest descent for the equalizer. In the end, we compare the BER results of the proposed designs and previous strategies.

The following assumptions are made. The variances of $\mathbf{s}$, $\mathbf{w}_k$ and $\mathbf{n}$ are respectively $\mathbf{R}_s = \sigma_s^2 \mathbf{I}$, $\mathbf{R}_{w_k} = \sigma_w^2 \mathbf{I}_{N_R}$ and $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_{N_D}$ and the covariance matrices of the noise terms have been normalized: $\sigma_w^2 = \sigma_n^2 = 1$. It is convenient to introduce two SNR parameters as follows. The first SNR is defined as $\rho_1 \triangleq \sigma_s^2/\sigma_w^2$, i.e., the ratio of transmitted signal power per antenna to the received noise power per antenna. The second SNR, defined in terms of the sum power $P_R \triangleq \sum_{k=1}^{M} P_k$ as $\rho_2 \triangleq P_R/(MN_R\sigma_n^2)$, gives the ratio of average transmitted power per relay antenna to the power of the noise induced at the individual destination antennas. In our simulations, the channel matrices have independent and identically distributed entries. Each entry is a zero mean circular symmetric complex Gaussian variable with unit variance.

A. Weighted Sum Power Constraint: Power Allocation

To study the effects of channel gains on the power allocation among the relays, we consider a 1S-2R-1D system with $N_S = N_R = N_D = 2$, $\rho_1 = 20$ dB and $\mathbf{Q} = \mathbf{I}$. In particular, we multiply $\mathbf{H}_1$ by α which represents a relative channel gain,
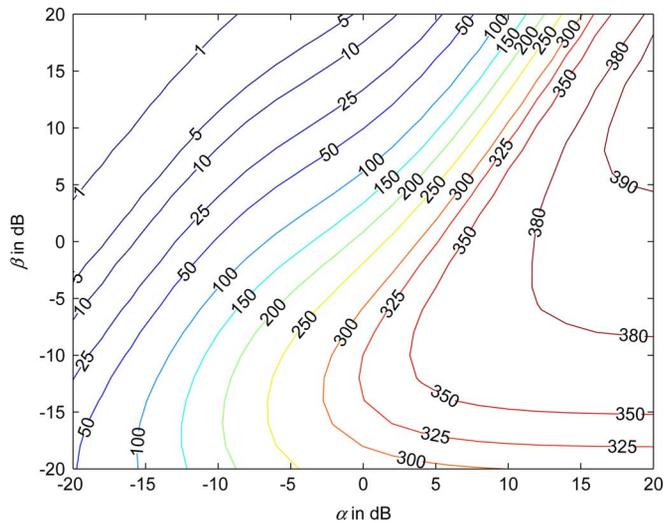
Fig. 2. Contours of the power of the first relay P_1' versus relative channel gains α and β.
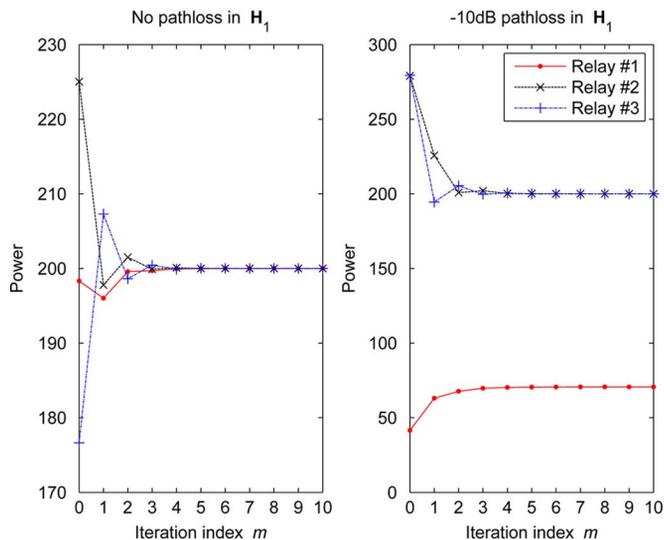
Fig. 3. Convergence behaviours of Algorithm 1: transmit power of the relays.

and multiply $\mathbf{G}_1$ by β which is independent from α. Then, the optimal relaying matrices and the transmit power of the relays (P_1', P_2') are all functions of these relative gains. A randomly generated set of channel matrices are used and the other parameters are chosen as $w_1 = w_2 = 1$ and $P_r = 400$. As shown by the contours for P_1' in Fig. 2, if β is fixed, the larger α becomes, the more power is allocated to the first relay; if α is held constant, the larger β comes, the less power to the first relay. In other words, the optimal relaying strategy tends to allocate more power to the relay with better backward channel and/or worse forward channel.

B. Convergence of Iterative Algorithms

Power Balancing for the Per-Relay Power Constraints: We study the convergence behaviors of Algorithm 1 base on a 1S-3R-1D system with $N_S = N_R = N_D = 4$, $\rho_1 = 20$ dB and the power constraints $P_1 = P_2 = P_3 = 200$. For a particular representative channel instance (randomly generated), Fig. 3 plots the power of the individual relays versus the iteration index m. Algorithm 1 usually converges within about 5 steps.

Our analysis in Section IV-C leads to the conclusion that the optimal relaying strategy does not necessarily use the maximum amount of power at some relays. This is verified by the right subfigure in Fig. 3, where the channel matrices used are the same as those in the left, except that $\mathbf{H}_1$ has a relative path loss of $\alpha = -10$ dB. We note that with optimal relaying strategy, the first relay is neither using the maximum power, 200, nor being turned off completely.

Equalizer Design: We study the convergence behaviors of the block coordinate descent method and the steepest descent method proposed in Section V, based on the same settings as above. The weights are $w_1 = w_2 = w_3 = 1$ and $P_r = 600$. The convergence behaviors are shown in Fig. 4 for a representative channel instance. The steepest descent method converges significantly faster than the block coordinate descent method. For the former, the MSE comes close to the locally optimal value after only 10 line-searches, whereas for the latter, it takes several thousand iterations. One interesting observation is that as far as we could verify, the optimal equalizer $\mathbf{Q}^\star$ does not depend on the algorithm used or the initial $\mathbf{Q}_0$ (except for a linear scaling factor). This suggests that the solution so obtained may be *globally* optimal, though it seems very difficult to prove and remains a conjecture for now because the expression of the Hessian is rather involved.

C. BER Performance

In this subsection, we compare the BER performance of the following relaying strategies:
1) Simplistic AF, $\mathbf{F}_k \propto \mathbf{I}$, [7];
2) MMSE-MMSE [7], [10];
3) CMMSE-MMSE [10];
4) Gradient-based MMSE [13];
5) Proposed method (sum power constraint, $\mathbf{Q} = \mathbf{I}$);
6) Proposed method (per-relay power constraints, $\mathbf{Q} = \mathbf{I}$);
7) Proposed joint design of the relaying matrices and the equalizer (sum power constraint);
8) Relay selection based on the JMMSE strategy [5].

For methods 1), 2), 3), 4) and 6), the total power is evenly split between different relays. For the selection-based strategy, the total power is allocated only to the single relay that would result in the minimum MSE based on the JMMSE relaying strategy [5]. We consider a 1S-3R-1D system with $N_S = N_R = N_D = 4$. In the simulations, each source antenna transmits independent uncoded 16-PSK symbol streams. The relay stations apply one of the above relaying schemes to their input signals and retransmit them. The destination applies a linear MMSE equalizer and then employs single-stream maximum likelihood decoding. The BER values are averaged over channel realizations.

First, we set ρ_2 to 20 dB and vary ρ_1 between 5 dB and 25 dB. Then, we set $\rho_1 = 20$ dB and vary ρ_2. The BER values are plotted in Figs. 5 and 6. As explained earlier, SAF cannot achieve distributed array gain and accordingly has the worst performance. The heuristic strategies including MMSE-MMSE and CMMSE-MMSE perform better. The gradient-based MMSE method provides further gain especially under low-to-mid SNR levels. The proposed MMSE-based strategies, 5), 6) and 7), outperform the above ones by large gaps. The choice depends on the compromise between performance and complexity: the joint design leads to lower BER but comes with higher computational complexity. It is worth
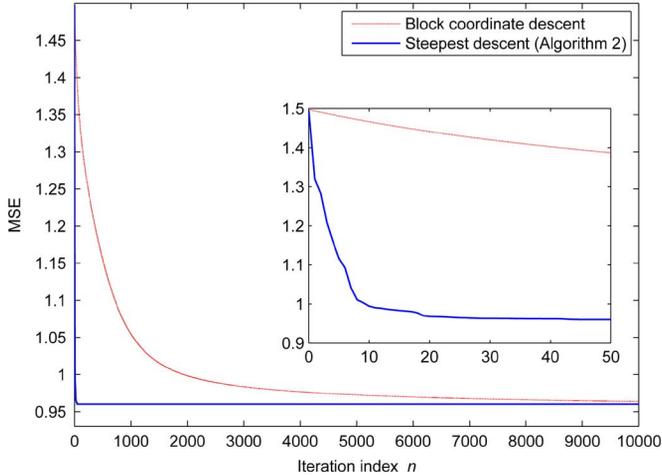
Fig. 4. Speed of convergence for the joint design: block coordinate descent and steepest descent (Algorithm 2).
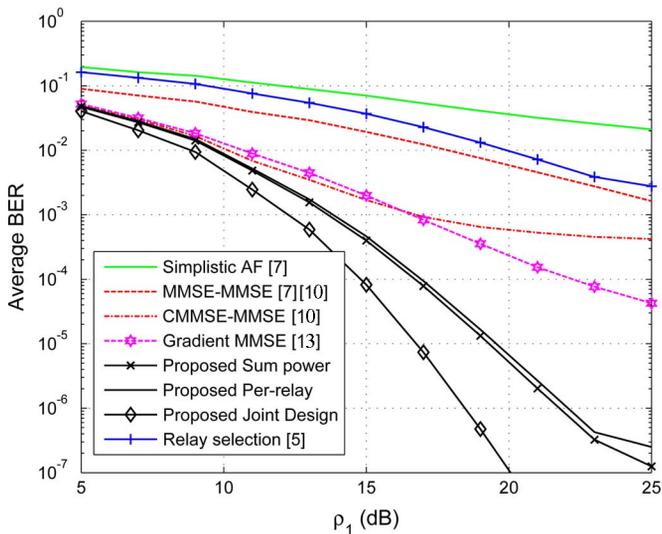
Fig. 5. Comparison of uncoded 16-PSK BER versus ρ_1 for different relaying strategies. $\rho_2 = 20$ dB.

mentioning that the proposed strategies are much superior to the selection-based one, which justifies the use of multiple relays.

VIII. CONCLUSION

In this paper, we have considered the MMSE-based joint design of the multiple relaying matrices. Under the weighted sum power constraint, we derived closed-form expressions for the optimal relaying matrices. The optimal strategy tends to allocate more power to those relays with better backward channels and/or worse forward channels, and to those with smaller weights. Under the per-relay power constraints, we proposed the power balancing algorithm (Algorithm 1) which is more efficient than general-purpose interior-point methods. The optimal strategy may not use the maximum amount of power at some relays, but does not turn off a relay either, no matter how low the SNR is at that relay. Additionally, under both types of constraints, a MIMO equalizer at the destination may be designed together with the relaying matrices. The steepest descent
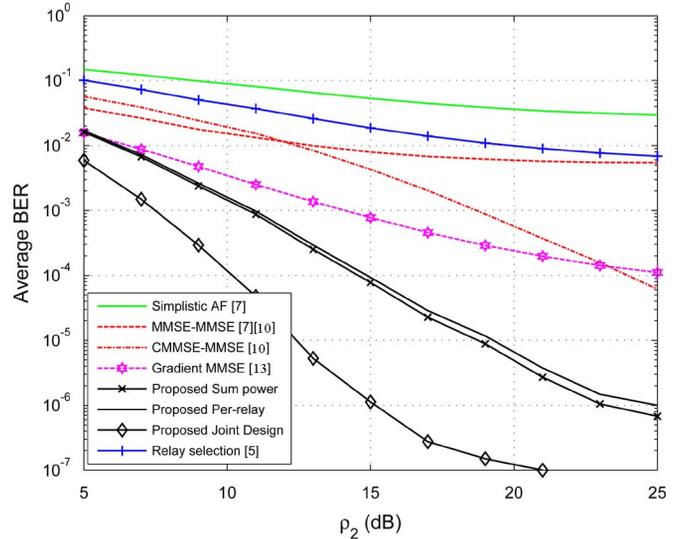
Fig. 6. Comparison of uncoded 16-PSK BER versus ρ_2 for different relaying strategies. $\rho_1 = 20$ dB.

method (Algorithm 2) converges much faster than the block coordinate descent method. The BER simulations show that all the proposed designs, under either type of constraints, with or without the equalizer, outperform previous ones by large margins. These simulations also illustrate significant performance advantage of multi-relay systems over single-relay ones.

APPENDIX A
PROOF OF COROLLARY 5

The Woodbury matrix identity does not hold for pseudo inverse in general and therefore we prove this corollary by substituting (25) into the left hand side (LHS) of (22). The kth sub-block of the column vector $\mathbf{\Psi f}^\star$ would be

$$
\mathbf{T}_k \left(\mathbf{R}_s^T \otimes \mathbf{I}\right) \sum_{l=1}^{M} \mathbf{T}_l^H \left(\mathbf{S}_l + \lambda^\star w_l \mathbf{I}_{N_R^2}\right)^\dagger \mathbf{T}_l \mathbf{\Sigma}^{-1}\mathrm{vec}(\mathbf{I})
$$

$$
+ \left(\mathbf{S}_k + \lambda^\star w_k \mathbf{I}_{N_R^2}\right)\left(\mathbf{S}_k + \lambda^\star w_k \mathbf{I}_{N_R^2}\right)^\dagger \mathbf{T}_k \mathbf{\Sigma}^{-1}\mathrm{vec}(\mathbf{I}).
$$
(57)

We always have $(\mathbf{S}_k + \lambda^\star w_k \mathbf{I}_{N_R^2})(\mathbf{S}_k + \lambda^\star w_k \mathbf{I}_{N_R^2})^\dagger \mathbf{T}_k = \mathbf{T}_k$: if $\lambda^\star w_k > 0$, the pseudo inverse operator is replaced by matrix inverse; if $\lambda^\star w_k = 0$,

$$
\mathbf{S}_k \mathbf{S}_k^\dagger \mathbf{T}_k = \left(\mathbf{H}_k^H \mathbf{R}_{\mathbf{x}_k}^{-1/2}\right)^T
$$
$$
\otimes \mathbf{G}_k^H \mathbf{Q}^H \mathbf{Q} \mathbf{G}_k \left(\mathbf{G}_k^H \mathbf{Q}^H \mathbf{Q} \mathbf{G}_k\right)^\dagger \mathbf{G}_k^H \mathbf{Q}^H
$$
$$
= \left(\mathbf{H}_k^H \mathbf{R}_{\mathbf{x}_k}^{-1/2}\right)^T \otimes \mathbf{G}_k^H \mathbf{Q}^H = \mathbf{T}_k
$$

because $\mathbf{Q}\mathbf{G}_k(\mathbf{G}_k^H \mathbf{Q}^H \mathbf{Q} \mathbf{G}_k)^\dagger \mathbf{G}_k^H \mathbf{Q}^H$ is a projection matrix so that $\mathbf{G}_k^H \mathbf{Q}^H$ is not changed. By inserting $(\mathbf{R}_s^T \otimes \mathbf{I})(\mathbf{R}_s^{-T} \otimes \mathbf{I})$ between $\mathbf{T}_k$ and $\mathbf{\Sigma}^{-1}$ in the second term of (57), the kth sub-block of $\mathbf{\Psi f}^\star$ is equal to $\mathbf{T}_k(\mathbf{R}_s^T \otimes \mathbf{I})\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathrm{vec}(\mathbf{I}) = \mathbf{T}_k(\mathbf{R}_s^T \otimes \mathbf{I})\mathrm{vec}(\mathbf{I})$. Therefore, $\mathbf{\Psi f}^\star = \mathbf{T}(\mathbf{R}_s^T \otimes \mathbf{I})\mathrm{vec}(\mathbf{I}) = \mathbf{b}$ and (25) is a solution of (22).

In addition, (25) is indeed the minimum-norm solution, that is, $\mathbf{f}^{\star H}\mathbf{f}^{\perp}_{(m_k)} = 0$ for m_k, $1 \le k \le M$. The following equality

$$\mathcal{N}(\mathbf{S}_k^{\dagger}) = \mathcal{N}(\mathbf{S}_k) = \mathcal{N}\left(\mathbf{I} \otimes \mathbf{G}_{m_k}^H \mathbf{Q}^H \mathbf{Q} \mathbf{G}_{m_k}\right)$$
$$= \mathcal{N}\left(\mathbf{I} \otimes \mathbf{Q} \mathbf{G}_{m_k}\right),$$

leads to $\mathbf{S}_k^{\dagger}\mathbf{f}^{\perp}_{m_k} = \mathbf{0}$. Since $\lambda^{\star} w_{m_k} = 0$, the inner product is

$$\mathbf{f}^{\star H}\mathbf{f}^{\perp}_{(m_k)} = \mathbf{f}^{\star H}_{m_k}\mathbf{f}^{\perp}_{m_k} = \text{vec}(\mathbf{I})^H \mathbf{\Sigma}^{-1} \mathbf{T}_{m_k}^H \mathbf{S}_k^{\dagger} \mathbf{f}^{\perp}_{m_k} = 0.$$

This completes the proof.

APPENDIX B
PROOF OF PROPOSITION 6

We prove this proposition by showing that the derivative of $g(\lambda)$ is negative. The matrix $\mathbf{\Psi}$ can be singular, for example, when $w_k = 0$ for a particular k and $N_R > N_S$. Hence, the major difficulty is that the following property for matrix inverse

$$\frac{\mathrm{d}\mathbf{A}(\lambda)^{-1}}{\mathrm{d}\lambda} = -\mathbf{A}(\lambda)^{-1}\frac{\mathrm{d}\mathbf{A}(\lambda)}{\mathrm{d}\lambda}\mathbf{A}(\lambda)^{-1} \qquad (58)$$

does not hold for the pseudo inverse in general. Our tactics here is to make $\mathbf{\Psi}$ invertible by adding a matrix to it, but without changing the value of $g(\lambda)$. According to Proposition 3, $\mathcal{N}(\mathbf{\Psi})$ is the direct sum of all $\mathcal{F}_{m_k}$ satisfying $\lambda w_{m_k} = 0$, which does not depend on the specific value of λ as long as $\lambda > 0$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be a set of orthonormal basis vectors for $\mathcal{N}(\mathbf{\Psi})$ and define $\mathbf{\Psi}_e \triangleq \mathbf{\Psi} + \sum_{k=1}^{p} \mathbf{u}_k \mathbf{u}_k^H$ (a function of λ). The orthogonality relationship $\mathcal{R}(\mathbf{\Psi}) \perp \mathcal{N}(\mathbf{\Psi})$ leads to

$$\mathbf{\Psi}_e^{-1} = \left(\mathbf{\Psi} + \sum_{k=1}^{p} \mathbf{u}_k \mathbf{u}_k^H\right)^{-1} = \mathbf{\Psi}^{\dagger} + \sum_{k=1}^{p} \mathbf{u}_k \mathbf{u}_k^H. \qquad (59)$$

Since $\mathbf{b} \in \mathcal{R}(\mathbf{\Psi})$, we have $\mathbf{u}_k^H \mathbf{b} = 0$ and therefore

$$\mathbf{\Psi}_e^{-1}\mathbf{b} = \mathbf{\Psi}^{\dagger}\mathbf{b}. \qquad (60)$$

To this point, all the pseudo inverses ($\mathbf{\Psi}^{\dagger}$) can be replaced by matrix inverses ($\mathbf{\Psi}_e^{-1}$).

Based on the chain rule, the derivative of $g(\lambda)$ satisfies

$$g'(\lambda) = -2\mathbf{b}^H \mathbf{\Psi}_e^{-1} \mathbf{I}_{\text{sum}} \mathbf{\Psi}_e^{-1} \mathbf{I}_{\text{sum}} \mathbf{\Psi}_e^{-1} \mathbf{b} \le 0, \qquad (61)$$

for the matrix $\mathbf{\Psi}_e^{-1}$ is positive definite. Next we prove $g'(\lambda) \ne 0$ by contradiction. Assume there exists a $\lambda_0 > 0$ so that $g'(\lambda_0) = 0$. Since $\mathbf{\Psi}_e(\lambda_0)^{-1}$ is positive definite, $\mathbf{I}_{\text{sum}}\mathbf{\Psi}_e(\lambda_0)^{-1}\mathbf{b}$ must be a zero vector, which leads to

$$\mathbf{b} = (\mathbf{\Phi} + \lambda_0 \mathbf{I}_{\text{sum}})\mathbf{\Psi}(\lambda_0)^{\dagger}\mathbf{b} = (\mathbf{\Phi} + \lambda_0 \mathbf{I}_{\text{sum}})\mathbf{\Psi}_e(\lambda_0)^{-1}\mathbf{b}$$
$$= (\mathbf{\Phi} + \lambda \mathbf{I}_{\text{sum}})\mathbf{\Psi}_e(\lambda_0)^{-1}\mathbf{b}, \qquad \forall \lambda \ge 0.$$

This means that $\mathbf{f}^{\star} = \mathbf{\Psi}_e(\lambda_0)^{-1}\mathbf{b}$ is the solution of (22) for any $\lambda \ge 0$. As a result, the weighted sum power satisfies

$$0 \le g(\lambda) = \|\mathbf{I}_{\text{sum}}\mathbf{\Psi}(\lambda)^{\dagger}\mathbf{b}\|_2^2 = \|\mathbf{I}_{\text{sum}}\mathbf{\Psi}_e(\lambda_0)^{-1}\mathbf{b}\|_2^2 = 0.$$

This contradicts with $g(0) > 0$ and therefore, $g'(\lambda) < 0$ always holds. The limit of $g(\lambda)$ is

$$\lim_{\lambda \to \infty} g(\lambda) = \lim_{\lambda \to \infty} \mathbf{b}^H \mathbf{I}_{\text{sum}}^{\dagger} \mathbf{b}^H / \lambda^2 = 0.$$

APPENDIX C
GRADIENT OF THE MSE WITH RESPECT TO $\bar{\mathbf{Q}}$

From the MSE Expression in (31b), the partial derivative is

$$\frac{\partial \text{MSE}}{\partial \bar{q}_{ij}} = -\text{vec}(\mathbf{I})^H \mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \bar{q}_{ij}}\mathbf{\Sigma}^{-1}\text{vec}(\mathbf{I}). \qquad (62)$$

Using similar techniques to those in the proof of Proposition 6 and substituting (25), we can express (62) as

$$\eta_o^{-2}\sum_{k=1}^{M}\mathbf{f}_k^{\star H}\frac{\partial(\mathbf{S}_k + \theta w_k \mathbf{I})}{\partial \bar{q}_{ij}}\mathbf{f}_k^{\star} - \eta_o^{-1}\sum_{k=1}^{M}\mathbf{f}_k^{\star H}\frac{\partial \mathbf{T}_k}{\partial \bar{q}_{ij}}\mathbf{\Sigma}^{-1}\text{vec}(\mathbf{I}).$$

Define an indicator matrix $\mathbf{E}_{ji}$ whose (j,i)th entry is one and other entries are all zero, we have

$$\frac{\partial \mathbf{S}_k}{\partial \bar{q}_{ij}} = \left(\mathbf{R}_{x_k}^{-T/2}\mathbf{R}_{w_k}^T \mathbf{R}_{x_k}^{-T/2}\right) \otimes \left(\mathbf{G}_k^H \mathbf{E}_{ji}\mathbf{Q}\mathbf{G}_k\right),$$
$$\frac{\partial \theta}{\partial \bar{q}_{ij}} = \frac{\text{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{E}_{ji})}{P_R},$$
$$\frac{\partial \mathbf{T}_k}{\partial \bar{q}_{ij}} = \left(\mathbf{R}_{x_k}^{-T/2}\bar{\mathbf{H}}_k\right) \otimes \mathbf{G}_k^H \mathbf{E}_{ji}.$$

According to the property (14c), we have

$$\mathbf{f}_k^{\star H}\frac{\partial \mathbf{S}_k}{\partial \bar{q}_{ij}}\mathbf{f}_k^{\star} = \text{tr}\left(\mathbf{E}_{ji}\mathbf{Q}\mathbf{G}_k\mathbf{F}_k\mathbf{R}_{w_k}\mathbf{F}_k^H\mathbf{G}_k^H\right).$$

Because the power constraint is tightly satisfied, we have $\sum_{k=1}^{M} w_k \|\mathbf{f}_k^{\star}\|^2 = P_R$ and therefore

$$\sum_{k=1}^{M}\mathbf{f}_k^{\star H}\left(\frac{\partial \theta}{\partial \bar{q}_{ij}}w_k\mathbf{I}\right)\mathbf{f}_k^{\star} = P_R\frac{\partial \theta}{\partial \bar{q}_{ij}} = \text{tr}(\mathbf{E}_{ji}\mathbf{Q}\mathbf{R}_n).$$

The property (14c) also leads to

$$\mathbf{f}_k^{\star H}\frac{\partial \mathbf{T}_k}{\partial \bar{q}_{ij}}\mathbf{\Sigma}^{-1}\text{vec}(\mathbf{I})$$
$$= \text{tr}\left(\mathbf{E}_{ji}\,\text{unvec}(\mathbf{\Sigma}^{-1}\text{vec}(\mathbf{I}))\mathbf{H}_k^H\mathbf{F}_k^H\mathbf{G}_k^H\right).$$

Since $\text{tr}(\mathbf{E}_{ji}\mathbf{X}) = \mathbf{X}(i,j)$, the gradient can be expressed as in (56).

REFERENCES

[1] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 1433–1444, Jun. 2006.

[2] Y. J. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1774–1786, May 2007.

[3] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, pp. 2593–2604, Jun. 2007.

[4] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398–1407, Apr. 2007.

[5] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, pp. 517–519, Jul. 2008.

[6] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4837–4851, Dec. 2009.

[7] O. Oyman and A. J. Paulraj, "Design and analysis of linear distributed MIMO relaying algorithms," in *Proc. Inst. Electr. Eng.—Commun.*, Apr. 2006, vol. 153, pp. 565–572.

[8] H. Shi, T. Abe, T. Asai, and H. Yoshino, "A relaying scheme using QR decomposition with phase control for MIMO wireless networks," in *Proc. IEEE Int. Conf. Commun.*, Seoul, Korea, May 2005, vol. 4, pp. 2705–2711.

[9] Z. Ding, W. H. Chin, and K. Leung, "Distributed beamforming and power allocation for cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1817–1822, May 2008.

[10] C. Zhao and B. Champagne, "Non-regenerative MIMO relaying strategies—From single to multiple cooperative relays," in *Proc. 2nd Int. Conf. Wireless Commun. Signal Process.*, Suzhou, China, Oct. 2010, pp. 1–6.

[11] Y. Fu, L. Yang, and W.-P. Zhu, "A nearly optimal amplify-and-forward relaying scheme for two-hop MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 14, pp. 229–231, Mar. 2010.

[12] C. Zhao and B. Champagne, "MMSE-based non-regenerative parallel MIMO relaying with simplified receiver," in *Proc. IEEE Global Telecomm. Conf.*, Houston, TX, USA, Dec. 2011, pp. 1–5.

[13] K.-J. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3671–3681, Dec. 2010.

[14] Y. Rong, "Joint source and relay optimization for two-way MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 15, pp. 1329–1331, Dec. 2011.

[15] Y. Izi and A. Falahati, "Amplify-forward relaying for multiple-antenna multiple relay networks under individual power constraint at each relay," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 50, pp. 1–10, 2012.

[16] L. Sanguinetti, A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, pp. 1331–1346, Sep. 2012.

[17] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, pp. 5062–5073, Oct. 2008.

[18] A. Toding, M. R. A. Khandaker, and R. Yue, "Optimal joint source and relay beamforming for parallel MIMO relay networks," in *Proc. 6th Int. Conf. Wireless Commun. Netw. Mobile Comput.*, Chengdu, China, Sep. 2010, pp. 1–4.

[19] A. Toding, M. Khandaker, and Y. Rong, "Joint source and relay optimization for parallel MIMO relays using MMSE-DFE receiver," in *Proc. 16th Asia-Pacific Conf. Commun.*, Auckland, New Zealand, Oct.–Nov. 2010, pp. 12–16.

[20] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer Science+Business Media, 2006.

[21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[22] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *IEEE Trans. Commun.*, vol. 48, pp. 1681–1691, Oct. 2000.

[23] T. Kong and Y. Hua, "Optimal design of source and relay pilots for MIMO relay channel estimation," *IEEE Trans. Signal Process.*, vol. 59, pp. 4438–4446, Sep. 2011.

[24] Y. Rong, M. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay system via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2224–2233, Jun. 2012.

[25] 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, pp. 2198–2206, Dec. 2001.

[26] R. Gallager, *Principles of Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[27] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[28] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[29] G. Li, Y. Wang, T. Wu, and J. Huang, "Joint linear filter design in multi-user non-regenerative MIMO-relay systems," in *Proc. IEEE Int. Conf. Commun.*, Dresden, Germany, Jun. 2009, pp. 1–6.

[30] C. Song, K.-J. Lee, and I. Lee, "MMSE based transceiver designs in closed-loop non-regenerative MIMO relaying systems," *IEEE Trans. Wireless Commun.*, pp. 2310–2319, Jul. 2010.

[31] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004 [Online]. Available: http://users.isy.liu.se/johanl/yalmip

[32] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming. ver. 1.21, Apr. 2011 [Online]. Available: http://cvxr.com/cvx//cvx

[33] S. Ma, C. Xing, Y. Fan, Y.-C. Wu, T.-S. Ng, and H. Poor, "Iterative transceiver design for MIMO AF relay networks with multiple sources," in *Proc. Military Commun. Conf.*, Nov. 2010, pp. 369–374.
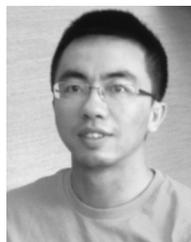
Chao Zhao (S'10) received the B.E. degree in electronics information engineering, and M.S. degrees in information and communication engineering from Harbin Institute of Technology, Harbin, Heilongjiang, China, in 2006 and 2008, respectively.

Since 2008, he has been a Ph.D. student at the Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, Canada. His research interests include array signal processing, MIMO relay communications, and computer architectures.

Benoit Champagne (SM'03) received the B.Ing. degree in engineering physics from the Ecole Polytechnique de Montréal in 1983, the M.Sc. degree in physics from the Université de Montréal in 1985, and the Ph.D. degree in electrical engineering from the University of Toronto in 1990.

From 1990 to 1999, he was an Assistant and then Associate Professor at INRS-Telecommunications, Université du Quebec, Montréal. In 1999, he joined McGill University, Montreal, where he is now a Full Professor within the Department of Electrical and Computer Engineering. He also served as Associate Chairman of Graduate Studies in the Department from 2004 to 2007. His research focuses on the development and performance analysis of advanced algorithms for the processing of information bearing signals by digital means. His interests span many areas of statistical signal processing, including detection and estimation, sensor array processing, adaptive filtering, and applications thereof to broadband communications and speech processing, where he has published extensively. His research has been funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada, the "Fonds de Recherche sur la Nature et les Technologies" from the Government of Quebec, Prompt Quebec, as well as some major industrial sponsors, including Nortel Networks, Bell Canada, InterDigital, and Microsemi.

Dr. Champagne has been an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the EURASIP JOURNAL ON APPLIED SIGNAL PROCESSING, and has served on the Technical Committees of several international conferences in the fields of communications and signal processing. His is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.