Cross-Layer Optimization on Routing and Power Control of MIMO Ad Hoc Networks

Jia Liu[†] Tae Yoon Park[†] Y. Thomas Hou[†] Yi Shi[†] Hanif D. Sherali[‡]

[†] The Bradley Department of Electrical and Computer Engineering

[‡] The Grado Department of Industrial and Systems Engineering

Virginia Tech, Blacksburg, VA, USA

Abstract

MIMO-based systems have great potential to improve network capacity for wireless mesh networks (WMNs). Due to unique physical layer characteristics associated with MIMO systems, network performance is tightly coupled with mechanisms at physical layer and link layer. So far, research on MIMO-based WMNs is still in its infancy and little results are available in this important area. In this paper, we consider the problem of jointly optimizing power and bandwidth allocation at each node and multi-hop/multi-path routing in a MIMO-based WMN where links operate in orthogonal channels. To solve this problem, we develop a mathematical solution procedure, which combines Lagrangian dual decomposition, gradient projection, cutting-plane methods, and subgradient methods. We provide theoretical insights in deriving gradient projection and cutting plane methods. We also use simulations to verify the efficacy of our algorithm.

I Introduction

Since Winters's [1], Telatar's [2] and Foschini's [3] pioneering works predicting the potential of high spectral efficiency provided by multiple antenna systems, the last decade has witnessed the soar of research activity on Multiple-Input Multiple-Output (MIMO) technologies. Without costs of extra spectrum, MIMO technology, which exploits the rich scattering characteristic of wireless channels, is able to increase channel capacities substantially than conventional communication systems.

However, compared to the research on the capacity of single-user MIMO, for which many results are available (see [4] and [5] and references therein), the capacity issue of *multiuser* MIMO systems is much less studied and many fundamental problems remain unsolved [5]. With the emergence of wireless mesh networks (WMNs), which are multiuser and multihop in nature, the need to extend the MIMO communication concept from single-user systems to multiuser systems has become increasingly compelling. In a WMN, however, applying MIMO technique becomes far more complicated. Power control and power allocation at each link as well as multihop/multipath routing across the network all interact with one and another, and cross-layer optimization is not only desirable, but also necessary.

In this paper, we study the problem of cross-layer optimization on multi-hop/multi-path routing, power control, power allocation, and bandwidth allocation (CRPBA) for MIMO-based mesh network where links operate in orthogonal channels. Specifically, we consider how to support a set of user communication sessions by jointly optimizing power control, power allocation, bandwidth allocation, and flow routing such that some network utility (e.g., proportional fairness) is maximized. This problem, to the best of the authors' knowledge, has not been studied thus far.

A Main Contributions

The main contribution of this paper are the following:

1. We developed a mathematical solution procedure to solve CRPBA by combining Lagrangian decomposition, gradient projection, cutting-plane algorithms, and subgradient algorithms.

2. For the challenging link layer subproblem, we develop a rigorous gradient projection method as opposed to the heuristic one in [6].

3. Our proposed cutting-plane method can not only solve the Lagrangian dual, but also easily recover the optimal primal feasible solutions, thus circumventing a major difficulty of the popular subgradient-based approaches for solving Lagrangian dual problems. On the other hand, we also propose a primal optimal feasible solution recovery strategy for subgradient algorithm based only on dual information, and we show its convergence to primal feasibility and optimality.

B Paper Organization

The remainder of this paper is organized as follows. In Section II, we review related work. In Section III, we discuss the network model and problem formulation. Section IV introduces the the basic decomposition framework and the important techniques in solving the subproblems in each protocol. Section V and Section A introduce cutting-plane and subgradient algorithms for solving Lagrangian dual problem, respectively, and the recovering schemes for optimal primal feasible solutions. Numerical results are presented in Section VII. Section VIII concludes this paper.

II Related Work

Research on applying MIMO to WMN is still in its infancy and results remain extremely limited. In this section, we provide a synopsis of related work on the MIMO research evolution from single-hop ad hoc networks to mesh networks.

Within single-hop wireless networks, there is much research on MIMO-based systems for cellular networks [7–12], which are infrastructure-based (i.e., base station). In this setting, there are two types of channels: the MIMO multiple access channel (MIMO-MAC), which is associated to "uplink," and the MIMO broadcast channel (MIMO-BC), which is associated to "downlink". In [12], Yu et al. showed that the maximal sum rate of uplink MIMO-MAC can be solved via the so-called "Iterative Water-Filling" method by exploiting the nice convexity structure of the problem. Very recently, Weingarten et al. [13] showed that the general capacity region for MIMO-BC coincides with "dirty-paper coding" (DPC) rate region [14]. In particular, it was shown that there is a nice duality between MIMO-BC and MIMO-MAC [15–17]. By exploiting this duality, the maximal sum rate problem of MIMO-BC can be computed by solving an equivalent MIMO-MAC problem under a sum power constraint [8–10, 18, 19].

For multiuser single-hop MIMO-based communications without any infrastructure, the maximal sum rate problem becomes much more challenging due to its non-convex nature. In [20], Jorswieck and Boche analyzed the worst-case performance of a multiuser MIMO system with interference.

In [21, 22], Demirkol and Ingram introduced an iterative (trial-and-error) method based on stream control for some simple network topologies. In [23], Sundaresan et al. proposed a MAC control scheme along idea similar to those in [21, 22]. In [24], Chen and Gans analyzed the asymptotic behavior of network spectral efficiency with L simultaneous co-channel transmission pairs; they showed that, in the absence of channel state information (CSI) at the transmitters, the network asymptotic spectral efficiency is limited by n_r nats/s/Hz as $L \to \infty$, and at least $n_t + n_r + 2\sqrt{n_t \cdot n_r}$ nats/s/Hz when CSI is available at the transmitters. In contrast to asymptotic analysis, in [6], Ye and Blum studied a finite-sized network and proposed a gradient projection based method, which is local optimal.

For multi-hop MIMO-based ad hoc networks, research results remain limited. In [25], Hu and Zhang studied the joint problem of medium access control and routing, with a consideration of optimal hop distance to minimize end-to-end delay. In [26], Sundaresan and Sivakumar used simulations to study various characteristics and tradeoffs (multiplexing gain vs. diversity gain) of MIMO links that can be leveraged by routing layer protocols in rich multipath environments to improve performance. In [27], Lee et al. proposed a distributed algorithm for MIMO-based multi-hop ad hoc networks, in which diversity and multiplexing gains of each link are controlled to achieve the optimal rate-reliability tradeoff. The optimization problem assumes fixed SINRs and fixed routes between source and destination nodes. In these efforts, node power control, per-antenna power allocation and their impact on upper layers are not considered.

It is important to contrast the MIMO-enabled mesh network research outlined in this project and another active line of research under the name of “Cooperative Communications” (also called “Distributed MIMO”) [28, 29]. Typically, under cooperative communications, each node is only equipped with a *single* antenna. Cooperative communications rely on using relay channels at cooperating nodes to form a virtual multiple antenna array. The main objective is to improve transmission reliability, energy efficiency, and network connectivity, instead of increasing spectrum efficiency.

III Network Model

We first introduce notation for matrices, vectors, and complex scalars in this paper. We use boldface to denote matrices and vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^\dagger$ denotes the conjugate transpose. $\mathrm{Tr}\{\mathbf{A}\}$ denotes the trace of $\mathbf{A}$. $\mathrm{Diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ represents the block diagonal matrix with matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$ on its main diagonal. We let $\mathbf{I}$ denote the identity matrix with dimension determined from context. $\mathbf{A} \succeq 0$ represents that $\mathbf{A}$ is Hermitian and positive semidefinite (PSD). $\mathbf{1}$ and $\mathbf{0}$ denote vectors whose elements are all ones and zeros, respectively, and their dimensions are determined from context. $(\mathbf{v})_m$ represents the m^{th} entry of vector $\mathbf{v}$. For a real vector $\mathbf{v}$ and a real matrix $\mathbf{A}$, $\mathbf{v} \geq \mathbf{0}$ and $\mathbf{A} \geq \mathbf{0}$ mean that all entries in $\mathbf{v}$ and $\mathbf{A}$ are nonnegative, respectively. We let $\mathbf{e}_i$ be the unit column vector where the i^{th} entry is 1 and all other entries are 0. The dimension of $\mathbf{e}_i$ is determined from context as well. The operator "$\langle , \rangle$" represents vector or matrix inner product operation.

A Link Capacity Model

In this paper, it is assumed that the system has perfect channel knowledge, that is, the transmitters have perfect channel state information (CSI). Let the matrix $\mathbf{H}_l \in \mathbb{C}^{n_r \times n_t}$ represent the wireless channel gain matrix from the transmitting node to the receiving node of link l, where n_t and n_r are the numbers of transmitting and receiving antenna elements at each node, respectively. Although wireless channels in reality are time-varying, we consider a "constant" channel model in this paper, i.e., $\mathbf{H}_l$'s coherence time is larger than the transmission period we consider. This simplification is of much interest for the insight it provides and its application in finding the ergodic capacity for block-wise fading channels [5]. The received complex base-band signal vector for MIMO link l with n_t transmitting antennas and n_r receiving antennas in a Gaussian channel is given by

$$\mathbf{r}_l = \sqrt{\rho_l}\mathbf{H}_l\mathbf{t}_l + \mathbf{n}_l, \tag{1}$$

where $\mathbf{r}_l$ and $\mathbf{t}_l$ represent the received and transmitted signal vectors, $\mathbf{n}_l$ is the normalized additive white Gaussian noise vector, ρ_j captures path-loss effect. By adopting the path-loss model with path-loss exponent being equal to α, ρ_l can be computed as [30] $\rho_l = (\frac{G_t G_r \lambda^2}{(4\pi)^2})/(N_0 W D_l^\alpha)$, where

D_l denotes the length of link l, N_0 represents the power spectral density of white Gaussian noise, and W denotes the communication bandwidth, G_t and G_r are transmit and receive antenna gains, respectively, which are assumed to be 1 in this paper, λ is the wavelength of the transmitted signal.

Let matrix $\mathbf{Q}_l$ represent the covariance matrix of a zero-mean Gaussian input symbol vector $\mathbf{t}_l$ at link l, i.e., $\mathbf{Q}_l = \mathbb{E}\left\{\mathbf{t}_l \cdot \mathbf{t}_l^{\dagger}\right\}$. The definition of $\mathbf{Q}_l$ implies that it is Hermitian and $\mathbf{Q}_l \succeq 0$. Physically, $\mathbf{Q}_l$ represents the power allocation in different antenna elements in link l's transmitter and correlations between each of these elements. In this paper, we use the complex matrix $\mathbf{Q} \triangleq \left[\begin{array}{cccc} \mathbf{Q}_1 & \mathbf{Q}_2 & \ldots & \mathbf{Q}_L \end{array}\right] \in \mathbb{C}^{n_t \times (n_t \cdot L)}$ to denote the collection of all input covariance matrices. The link capacity of a MIMO link l in an AWGN channel can be written as

$$\Phi_l(W_l, \mathbf{Q}_l) \triangleq W_l \log_2 \det\left(\mathbf{I} + \rho_l \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^{\dagger}\right), \qquad (2)$$

where W_l represents the communication bandwidth of link l. It can be readily verified that $\Phi_l(\mathbf{Q}_l)$ is a monotone increasing convex function for $W_l > 0$ and $\mathbf{Q}_l \succeq 0$.

B Multicommodity Network Flows (Network Layer)

In this paper, the topology of a wireless ad hoc networks with mutually interfered MIMO links is represented by a directed graph, denoted by $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N}$ and $\mathcal{L}$ are the set of nodes and all possible MIMO-based links, respectively. We assume that $\mathcal{G}$ is always connected. Suppose that the cardinalities of the sets $\mathcal{N}$ and $\mathcal{L}$ are $|\mathcal{N}| = N$ and $|\mathcal{L}| = L$, respectively. In our model, a link (the line segment defined by a pair of nodes) is possible to exist if the length of the link is less than or equal to the maximum transmission range R_T, i.e., $\mathcal{L} = \{(i,j) : D_{ij} \leq R_T, i,j \in \mathcal{N}, i \neq j\}$, where D_{ij} represents the distance between node i and node j. R_T can be determined by a node's maximum transmission power.

The network topology of $\mathcal{G}$ can be represented by a *node-arc incidence matrix* (NAIM) [31]

$\mathbf{A} \in \mathbb{R}^{N \times L}$, whose entry a_{nl} associating with node n and arc l is defined as

$$a_{nl} = \begin{cases} 1 & \text{if } n \text{ is the transmitting node of arc } l \\ -1 & \text{if } n \text{ is the receiving node of arc } l \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

We define $\mathcal{O}(n)$ and $\mathcal{I}(n)$ as the sets of links that are outgoing from and incoming to node n, respectively. We use a multicommodity flow model for the routing of data packets across the wireless ad hoc network. In this model, several nodes send different data to their corresponding destinations, possibly through multipath and multihop routing. We assume that the flow conservation law at each node is satisfied, i.e., the network is a flow-balanced system.

Suppose that there are F flows in total in the network, representing F different commodities. The source and destination nodes of flow f, $1 \leq f \leq F$, are denoted as src(f) and dst(f), respectively. For each flow, we define a *source-sink vector* $\mathbf{s}_f \in \mathbb{R}^N$, whose entries, other than at the positions of src(f) and dst(f), are all zeros. In addition, from the flow conservation law, we must have $(\mathbf{s}_f)_{\mathrm{src}(f)} = -(\mathbf{s}_f)_{\mathrm{dst}(f)}$. Without loss of generality, we let$(\mathbf{s}_f)_{\mathrm{src}(f)} \geq 0$ and simply denote it as a scalar s_f. Therefore, we can further write the source-sink vector of flow f as

$$\mathbf{s}_f = s_f \begin{bmatrix} \cdots & 1 & \cdots & -1 & \cdots \end{bmatrix}^T, \tag{4}$$

where the dots represent zeros, and 1 and -1 are in the positions of src(f) and dst(f), respectively. Note that for the source-sink vector of a flow f, 1 does not necessarily appear before -1 as in (4), which is only for illustrative purpose. Using the notation "$=_{x,y}$" to represent the component-wise equality of a vector except at the x^{th} and the y^{th} entries, we have $\mathbf{s}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0}$. In addition, using the matrix $\mathbf{S} \triangleq \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_F \end{bmatrix} \in \mathbb{R}^{N \times F}$ to denote the collection of all source-sink vectors $\mathbf{s}_f$, we further have

$$\mathbf{S}\mathbf{e}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0}, \quad 1 \leq f \leq F, \tag{5}$$

$$\langle \mathbf{1}, \mathbf{S}\mathbf{e}_f \rangle = 0, \qquad 1 \leq f \leq F, \tag{6}$$

$$(\mathbf{S}\mathbf{e}_f)_{\mathrm{src}(f)} = s_f, \quad 1 \leq f \leq F, \tag{7}$$

where $\mathbf{e}_f$ is the f^{th} unit column vector.

On each link l, we let $x_l^{(f)} \geq 0$ be the amount of flow of commodity f in link l. We define $\mathbf{x}^{(f)} \in \mathbb{R}^L$ as the *flow vector* for commodity f. At each node n, components of the flow vector and source-sink vector for the same commodity satisfy the flow conservation law as follows:

$$\sum_{l \in \mathcal{O}(n)} x_l^{(f)} - \sum_{l \in \mathcal{I}(n)} x_l^{(f)} = (\mathbf{s}_f)_n, \quad 1 \leq n \leq N, \ 1 \leq f \leq F. \tag{8}$$

With NAIM, the flow conservation law across the whole network can be compactly written as $\mathbf{A}\mathbf{x}^{(f)} = \mathbf{s}_f, \quad 1 \leq f \leq F$. We use matrix $\mathbf{X} \triangleq \left[\begin{array}{cccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \ldots & \mathbf{x}^{(F)} \end{array} \right] \in \mathbb{R}^{L \times F}$ to denote the collection of flow vectors $\mathbf{x}^{(f)}$. With $\mathbf{X}$ and $\mathbf{S}$, the flow conservation law can be further compactly written as

$$\mathbf{A}\mathbf{X} = \mathbf{S}. \tag{9}$$

We impose a capacity constraint on each individual arc as follows. Let c_l be the capacity of arc l. Then we must have

$$\sum_{f=1}^{F} x_l^{(f)} \leq c_l, \quad 1 \leq l \leq L, \tag{10}$$

which means that the total amount of flow on each arc l cannot exceed its capacity limit. Eq. (10) can be further compactly written using matrix-vector notations as

$$\langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \leq c_l, \quad 1 \leq l \leq L. \tag{11}$$

In wireless ad hoc network, the link capacity c_l is not fixed and can be varied as the power allocation among each node changes. We will discuss the formula for computing MIMO link capacity in the next section. In summary, the multicommodity network flow model imposes the following group of constraints on the network flow variables $\mathbf{X}$, $\mathbf{S}$:

$$\left\{ \begin{array}{c} \mathbf{A}\mathbf{X} = \mathbf{S}, \\ \mathbf{X} \geq \mathbf{0}, \\ \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \leq c_l, \quad 1 \leq l \leq L, \end{array} \right. \tag{12}$$

where $\mathbf{S}$ satisfies $\mathbf{S}\mathbf{e}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0}$, $\langle \mathbf{1}, \mathbf{S}\mathbf{e}_f \rangle = 0$, and $(\mathbf{S}\mathbf{e}_f)_{\mathrm{src}(f)} = s_f$, for $f = 1, 2, \ldots, F$.

C Problem Formulation

The goal of this paper is to design an algorithm that performs the cross-layer optimization on multihop/multipath routing, power control, power allocation, and bandwidth allocation (CRPBA) for a MIMO-based WMN. We consider an FDMA MIMO-based WMN, where each node has been assigned non-overlapping (possibly reused) frequency bands for its incoming and outgoing links so that each node can simultaneously transmit and receive, and cause no interference to other nodes. How to perform channel assignments is a huge research topic by itself, and there are a vast amount of literature that discuss channel assignment problems. Thus in this paper, we focus on how to jointly optimize routing in the network layer and power control/allocation as well as bandwidth allocation in the link layer.

We adopt the well-known proportional fairness utility function, i.e., $\ln(s_f)$ for flow f [32]. We wish to perform the cross-layer optimization such that the sum of all utilities of flows is maximized. Since the network flows in a link cannot exceed the link's capacity limit, we have $\sum_{f=1}^{F} x_l^{(f)} \leq \Phi_l(W_l, \mathbf{Q}_l)$, for $1 \leq l \leq L$. Using matrix-vector notations, it can be further compactly written as

$$\langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \leq \Phi_l(W_l, \mathbf{Q}_l), \quad 1 \leq l \leq L. \tag{13}$$

In the link layer, since the total transmit power of each node is subject to a maximum power constraint, we have $\sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)}$, $1 \leq n \leq N$, where $P_{\max}^{(n)}$ represents the maximum transmit power of node n. Also, the sum of bandwidths of all the outgoing links for a node n cannot exceed the assigned bandwidth for node n, i.e., $\sum_{l \in \mathcal{O}(n)} W_l \leq B_n$, $1 \leq n \leq N$, where B_n is the assigned transmission band for node n. We use the matrix $\mathbf{W} = \begin{bmatrix} W_1 & W_2 & \ldots & W_L \end{bmatrix}^T \in \mathbb{R}^{L \times 1}$ to denote the collection of the bands for links from 1 to L. Coupling the MIMO link capacity model in Section A and the network flow model in Section B, we have the problem formulation for CRPBA as in (14).

$$\begin{array}{rlr} \textbf{CRPBA}\text{ :Maximize} & \sum_{f=1}^{F} \ln(s_f) & \\ \text{subject to} & \mathbf{AX} = \mathbf{S} & \\ & \mathbf{X} \geq \mathbf{0} & \\ & \mathbf{S}\mathbf{e}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0} & \forall\, f \\ & \langle \mathbf{1}, \mathbf{S}\mathbf{e}_f \rangle = 0 & \forall\, f \\ & (\mathbf{S}\mathbf{e}_f)_{\mathrm{src}(f)} = s_f & \forall\, f \\ & \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \leq \Phi_l(W_l, \mathbf{Q}_l) & \forall\, l \\ & \sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)} & \forall\, n \\ & \sum_{l \in \mathcal{O}(n)} W_l \leq B_n & \forall\, n \\ & \mathbf{Q}_l \succeq 0,\ W_l \geq 0 & \forall\, l \\ \text{Variables:} & \mathbf{S},\, \mathbf{X},\, \mathbf{Q}, \mathbf{W} & \end{array} \qquad (14)$$

IV Solution Procedure

It can be observed that the CRPBA possesses a special structure: The network layer variables and the link layer variables are coupled through the link capacity constraints $\langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \leq \Phi_l(W_l, \mathbf{Q}_l)$. Thus, we can exploit this special structure using Lagrangian dual decomposition to solve CRPBA efficiently. In [33], the authors used a similar decomposition technique to solve simultaneous routing and resource allocation problems. However, their routing setting was very different to our work and their link layer was not MIMO-based. Due to the spatial dimension resulted from MIMO, the link layer subproblem in this paper is completely different and substantially more challenging. Generally, given a nonlinear programming problem, several different Lagrangian dual problems can be constructed depending on which constraints are associated with Lagrangian dual variables [34]. For CRPBA, we associate Lagrangian multipliers u_i to the link capacity coupling constraints $\langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \leq \Phi_l(W_l, \mathbf{Q}_l)$. Hence, the Lagrangian can be written as [34]

$$\Theta(\mathbf{u}) = \sup_{\mathbf{S},\mathbf{X},\mathbf{Q},\mathbf{W}} \left\{ L(\mathbf{S}, \mathbf{X}, \mathbf{Q}, \mathbf{W}, \mathbf{u}) | (\mathbf{S}, \mathbf{X}, \mathbf{Q}, \mathbf{W}) \in \Gamma \right\},$$

where

$$L(\mathbf{S}, \mathbf{X}, \mathbf{Q}, \mathbf{W}, \mathbf{u}) \;=\; \sum_{f} \ln(s_f) + \sum_{l} u_l \left(\Phi_l(W_l, \mathbf{Q}_l) - \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \right) \qquad (15)$$

and Γ is defined as

$$\Gamma \triangleq \left\{ (\mathbf{S}, \mathbf{X}, \mathbf{Q}, \mathbf{W}) \;\middle|\; \begin{array}{ll} \mathbf{AX} = \mathbf{S} & \\ \mathbf{X} \geq \mathbf{0} & \\ \mathbf{Se}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0} & \forall\, f \\ \langle \mathbf{1}, \mathbf{Se}_f \rangle = 0 & \forall\, f \\ (\mathbf{Se}_f)_{\mathrm{src}(f)} = s_f & \forall\, f \\ \sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)} & \forall\, n \\ \mathbf{Q}_l \succeq 0 & \forall\, l \\ \sum_{l \in \mathcal{O}(n)} W_l \leq B_n & \forall\, n \end{array} \right\}$$

The Lagrangian dual problem of CRPBA can thus be written as [34]:

$$\begin{array}{rll} \mathbf{D}^{\mathrm{CRPBA}}: & \text{Minimize} & \Theta(\mathbf{u}) \\ & \text{subject to} & \mathbf{u} \geq \mathbf{0}. \end{array}$$

It is easy to recognize that, for any given Lagrangian multiplier $\mathbf{u}$, the Lagrangian in (15) can be separated into two terms:

$$\Theta(\mathbf{u}) = \Theta_{\mathrm{net}}(\mathbf{u}) + \Theta_{\mathrm{link}}(\mathbf{u}),$$

where Θ_{net} and Θ_{link} are two subproblems respectively corresponding to network layer and link layer:

$$\begin{array}{rll} \mathbf{D}_{\mathrm{net}}^{\mathrm{CRPBA}}: \Theta_{\mathrm{net}}(\mathbf{u}) \triangleq & \text{Maximize} \sum_f \ln(s_f) - \sum_l u_l \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle & \\ \text{subject to} & \mathbf{AX} = \mathbf{S} & \\ & \mathbf{X} \geq \mathbf{0} & \\ & \mathbf{Se}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0} & \forall\, f \\ & \langle \mathbf{1}, \mathbf{Se}_f \rangle = 0 & \forall\, f \\ & (\mathbf{Se}_f)_{\mathrm{src}(f)} = s_f & \forall\, f \\ \text{Variables:} & \mathbf{S},\, \mathbf{X} & \end{array}$$

$$\begin{array}{rll} \mathbf{D}_{\mathrm{link}}^{\mathrm{CRPBA}}: \Theta_{\mathrm{link}}(\mathbf{u}) \triangleq & \text{Maximize} \sum_l u_l \Phi_l(W_l, \mathbf{Q}_l) & \\ \text{subject to} & \sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)} & \forall\, n \\ & \sum_{l \in \mathcal{O}(n)} W_l \leq B_n & \forall\, n \\ & \mathbf{Q}_l \succeq 0 & \forall\, l \\ \text{Variables:} & \mathbf{Q},\, \mathbf{W} & \end{array}$$

The CRPBA Lagrangian dual problem can be thus transformed into the following *master* dual problem:

$$\begin{array}{rll} \mathbf{MD}^{\text{CRPBA}}: & \text{Minimize} & \Theta_{\text{net}}(\mathbf{u}) + \Theta_{\text{link}}(\mathbf{u}) \\ & \text{subject to} & \mathbf{u} \geq \mathbf{0} \end{array}$$

Now, the task of solving the decomposed Lagrangian dual problem boils down to how to evaluate the subproblems $\mathbf{D}_{\text{net}}^{\text{CRPBA}}$ and $\mathbf{D}_{\text{link}}^{\text{CRPBA}}$, and how to handle the master problem. Note that in the network layer subproblem $\mathbf{D}_{\text{net}}^{\text{CRPBA}}$, the objective function is concave and all constraints are affine. Therefore, $\mathbf{D}_{\text{net}}^{\text{CRPBA}}$ is readily solvable by using many polynomial time convex programming methods. However, solving $\mathbf{D}_{\text{link}}^{\text{CRPBA}}$ is not trivial because the objective function and constraints involve many complex matrices variables, even though it can be shown that $\mathbf{D}_{\text{link}}^{\text{CRPBA}}$ is a convex problem. In the following subsections, we will discuss each techniques we use to solve the link layer subproblem and the master problem in detail.

A Modified Gradient Projection Method (MGP)

In [35], we proposed using BB/RLT [36] to solve a similar problem for which, instead of having a weighted sum in the objective function, the maximum sum of link capacities needs to be determined. However, from the computational experience in [35], we know that BB/RLT is of exponential complexity, and it takes a very long time to converge even for networks of moderate size (e.g., 10 links) and 2×2 antenna arrays. In multihop network setting, the number of links is on the order of $O(N^2)$, which means that even a small network will contain too many links for BB/RLT to solve. Therefore, BB/RLT is not a practical choice in CRPBA.

In this paper, we propose a modified "gradient projection" (MGP) method to solve the link subproblem. Gradient projection, originally proposed by Rosen [37], is a classical nonlinear programming method aiming at solving constrained optimization problems. But its formal convergence proof has not been established until very recently [34]. The framework of MGP is shown in Algorithm 1.

Due to the complexity of the objective function, we cannot afford the luxury of performing an

Algorithm 1 Modified Gradient Projection Method

```
Initialization:
  Choose the initial conditions $\mathbf{W}^{(0)} = [W_1^{(0)}, W_2^{(0)}, \ldots, W_L^{(0)}]^T$, $\mathbf{Q}^{(0)} = [\mathbf{Q}_1^{(0)}, \mathbf{Q}_2^{(0)}, \ldots, \mathbf{Q}_L^{(0)}]^T$. Let $k = 0$.
Main Loop:
  1. Calculate the gradients $G_{W_l}^{(k)} = \nabla_{W_l}\Theta_{\text{link}}(\mathbf{u}, \mathbf{W}^{(k)}, \mathbf{Q}^{(k)})$ and $\mathbf{G}_{\mathbf{Q}_l}^{(k)} = \nabla_{\mathbf{Q}_l}\Theta_{\text{link}}(\mathbf{u}, \mathbf{W}^{(k)}, \mathbf{Q}^{(k)})$, for $l = 1, 2, \ldots, L$.
  2. Choose an appropriate step size $s_k$. Let $W_l^{(k)'} = W_l^{(k)} + s_k G_{W_l}^{(k)}$, $\mathbf{Q}_l^{(k)'} = \mathbf{Q}_l^{(k)} + s_k \mathbf{G}_{\mathbf{Q}_l}^{(k)}$, for $l = 1, 2, \ldots, L$.
  3. Let $[\bar{\mathbf{W}}_n^{(k)}, \bar{\mathbf{Q}}_n^{(k)}]^T$ be the projection of $[\mathbf{W}_n^{(k)'}, \mathbf{Q}_n^{(k)'}]^T$ onto $\Omega_+(n)$, where $\Omega_+(n) \triangleq \{(W_l, \mathbf{Q}_l) | l \in \mathcal{O}(n), W_l \geq 0, \mathbf{Q}_l \succeq 0, \sum_{l \in \mathcal{O}(n)} W_l \leq B_n, \sum_{l \in \mathcal{O}(n)} \text{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)}\}$.
  4. Choose appropriate step size $\alpha_k$. Let $W_l^{(k+1)} = W_l^{(k)} + \alpha_k(\bar{W}_l^{(k)} - W_l^{(k)})$, $\mathbf{Q}_l^{(k+1)} = \mathbf{Q}_l^{(k)} + \alpha_k(\bar{\mathbf{Q}}_l^{(k)} - \mathbf{Q}_l(k))$, $l = 1, 2, \ldots, L$.
  5. $k = k + 1$. If the maximum absolute value of the elements in $\mathbf{Q}_l^{(k)} - \mathbf{Q}_l^{(k-1)} < \epsilon$ and $W_l^{(k)} - W_l^{(k-1)} < \epsilon$, for $l = 1, 2, \ldots, L$, then stop; else go to step 1.
```

exact line search which requires the expense of excessive objective function evaluations. Therefore, we adopt the "Armijo rule" inexact line search method [34], which still enjoys provable convergence. The basic idea of Armijo rule is that at each step of the line search, we sacrifice accuracy for efficiency as long as we have sufficient improvement. According to Armijo rule, we choose $s_k = 1$ and $\alpha_k = \beta^{m_k}$ (the same as in [6]), where m_k is the first non-negative that satisfies

$$\Theta_{\text{link}}(\mathbf{Q}^{(k+1)}) - \Theta_{\text{link}}(\mathbf{Q}^{(k)}) \geq \sigma\beta^{m_k} \sum_{l=1}^{L} \text{Tr}\left[\nabla_{\mathbf{Q}_l}\Theta_{\text{link}}(\mathbf{Q}^{(k)})^\dagger \left(\bar{\mathbf{Q}}_l^{(k)} - \mathbf{Q}_l^{(k)}\right)\right],$$

where $0 < \beta < 1$ and $0 < \sigma < 1$ are fixed scalars. Note that the matrix inner product in (16) is defined as $\langle \mathbf{A}, \mathbf{B} \rangle = \text{Tr}\left(\mathbf{A}^\dagger \mathbf{B}\right)$.

In order to make the general MGP algorithm work, we will have to fill in two problem specific components, i.e., during the k^{th} duration, how to compute the gradient $\nabla\Theta_{\text{link}}(\mathbf{Q}(k))$ and how to project $\mathbf{Q}_l'(k)$ onto $\Omega_+(n) \triangleq \{\mathbf{Q}_l | l \in \mathcal{O}(n), \mathbf{Q}_l \succeq 0, \sum_l \text{Tr}\{\mathbf{Q}_l\} < P_{\max}\}$.

A.1 Compute the Gradients

It is evident that the gradient $G_{W_l} \triangleq \nabla_{W_l}\Theta_{\text{link}} = u_l \log_2 \det(\mathbf{I} + \rho_l \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^\dagger)$. The gradient $\mathbf{G}_{\mathbf{Q}_l} \triangleq \nabla_{\mathbf{Q}_l}\Theta_{\text{link}}(\mathbf{Q})$ depends on the partial derivatives of $\Theta_{\text{link}}(\mathbf{Q})$ with respect to $\mathbf{Q}_l$. Before compute the partial derivative of $\Theta_{\text{link}}(\mathbf{Q})$, we first introduce the following lemma from matrix differential calculus [38].

Lemma 1. *For matrices* $\mathbf{B} \in \mathbb{C}^{p\times m}$*,* $\mathbf{X} \in \mathbb{C}^{m\times n}$*, and* $\mathbf{C} \in \mathbb{C}^{n\times p}$*, if* $(\mathbf{BXC})$ *is invertible, then we have* $\frac{\partial \det(\mathbf{BXC})}{\partial \mathbf{X}} = \det(\mathbf{BXC})\left[\mathbf{C}(\mathbf{BXC})^{-1}\mathbf{B}\right]^T$.

With Lemma 1, we can have the following corollary.

Corollary 1. *For matrices* $\mathbf{A} \in \mathbb{C}^{p\times p}$*,* $\mathbf{B} \in \mathbb{C}^{p\times m}$*, rank*$(\mathbf{B}) = p$*,* $\mathbf{X} \in \mathbb{C}^{m\times n}$*,* $\mathbf{C} \in \mathbb{C}^{n\times p}$*, and rank*$(\mathbf{C}) = p$*, if* $(\mathbf{A}+\mathbf{BXC})$ *is invertible, then we have* $\frac{\partial}{\partial \mathbf{X}} \ln\det(\mathbf{A}+\mathbf{BXC}) = \left[\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}\right]^T$.

Proof. By chain rule, we have

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{X}} \ln\det(\mathbf{BXC}) = \frac{\partial \ln\det(\mathbf{BXC})}{\partial \det(\mathbf{BXC})} \cdot \frac{\partial \det(\mathbf{BXC})}{\partial \mathbf{X}} \\
&= \frac{1}{\det(\mathbf{BXC})} \cdot \det(\mathbf{BXC})\left[\mathbf{C}(\mathbf{BXC})^{-1}\mathbf{B}\right]^T \\
&= \left[\mathbf{C}(\mathbf{BXC})^{-1}\mathbf{B}\right]^T. \qquad (16)
\end{aligned}$$

Let $\mathbf{Y}$ such that $\mathbf{BYC} = \mathbf{A}+\mathbf{BXC}$, i.e., $\mathbf{Y} = \mathbf{X} + \mathbf{B}_R\mathbf{A}\mathbf{C}_L$, where $\mathbf{B}_R$ and $\mathbf{C}_L$ are the right and left inverses of $\mathbf{B}$ and $\mathbf{C}$, respectively. It then follows that $\partial\mathbf{Y}/\partial\mathbf{X} = \mathbf{I}$. Substitute it into (16), we have

$$\begin{aligned}
&\frac{\partial \ln\det(\mathbf{A}+\mathbf{BXC})}{\partial \mathbf{X}} = \frac{\partial \ln\det(\mathbf{BYC})}{\partial \mathbf{Y}} \cdot \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} = \\
&\left[\mathbf{C}(\mathbf{BYC})^{-1}\mathbf{B}\right]^T \cdot \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} = \left[\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}\right]^T \cdot \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \\
&= \left[\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}\right]^T.
\end{aligned}$$

□

We can now compute the partial derivative of $\Theta_{\text{link}}(\mathbf{Q})$ with respect to $\mathbf{Q}_l$, which is given by

$$\frac{\partial \Theta_{\text{link}}(\mathbf{Q})}{\partial \mathbf{Q}_l} \;=\; W\frac{\partial}{\partial \mathbf{Q}_l}\left[\log_2\det\left(\mathbf{I} + \rho_j\mathbf{H}_j\mathbf{Q}_j\mathbf{H}_j^{\dagger}\right)\right]. \qquad (17)$$

Assuming the channel gain matrices $\mathbf{H}$ are of full row rank, i.e., $rank(\mathbf{H}_l) = n_r$, $l = 1, 2, \ldots, L$. Applying Corollary 1 by letting $\mathbf{A} = \mathbf{I}$, $\mathbf{B} = \rho_l \mathbf{H}_l$, $\mathbf{X} = \mathbf{Q}_l$, and $\mathbf{C} = \mathbf{H}_l^{\dagger}$, we obtain

$$\frac{\partial \Theta_{\text{link}}(\mathbf{Q})}{\partial \mathbf{Q}_l} = \frac{2W_l u_l \rho_l}{\ln 2} \left[\mathbf{H}_l^{\dagger} \left(\mathbf{I} + \rho_l \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^{\dagger} \right)^{-1} \mathbf{H}_l \right]^T. \tag{18}$$

Notice that, in (18), we have used the fact $\mathbf{R}_l$ does not depend on $\mathbf{Q}_l$.

Further, for a function $f(z)$, where $z = x + jy$ is a complex variable, the derivative is defined as $\partial f(z)/\partial z = (1/2)(\partial f(z)/\partial x - j(\partial f(z)/\partial y))$, and the gradient is defined as $\nabla_z f(z) = \partial f(z)/\partial x + j(\partial f(z)/\partial y)$ [6, 39]. Hence, we have $\nabla_z f(z) = 2(\partial f(z)/\partial z)^*$. Therefore,

$$\mathbf{G}_{\mathbf{Q}_l} = \frac{2W_l u_l \rho_l}{\ln 2} \mathbf{H}_l^{\dagger} \left(\mathbf{I} + \rho_l \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^{\dagger} \right)^{-1} \mathbf{H}_l, \tag{19}$$

where we notice that $\left(\mathbf{I} + \rho_j \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^{\dagger} \right)^{-1}$ is a Hermitian matrix, hence $\mathbf{G}_{\mathbf{Q}_l}$ is also Hermitian.

A.2 Projection onto $\Omega_+(n)$

From (19), we notice that $\mathbf{G}_{\mathbf{Q}_l}$ is Hermitian. Thus, we have that $\mathbf{Q}_l'(k) = \mathbf{Q}_l(k) + s_k \mathbf{G}_{\mathbf{Q}_l}(k)$ is Hermitian. Therefore, we will be concerned with how to project Hermitian matrices onto the constraint set $\Omega_+(n)$. Since $\Omega_+(n)$ contains a constraint on sum power for each node n, we need to come up with a method projecting $|\mathcal{O}(n)|$ links onto $\Omega_+(n)$ simultaneously.

For a node n having $|\mathcal{O}(n)|$ outgoing links, we want to project the $|\mathcal{O}(n)|$ W-scalars and $|\mathcal{O}(n)|$ $\mathbf{Q}$-covariance matrices onto $\Omega_+(n)$ simultaneously, we construct a block diagonal matrix $\mathbf{D}_n$ as follows:

$$\mathbf{D}_n = \left[\begin{array}{c|c} \mathbf{W}_n & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{Q}_n \end{array} \right] \in \mathbb{C}^{|\mathcal{O}(n)|(n_t+1) \times |\mathcal{O}(n)|(n_t+1)}$$

where $\mathbf{W}_n$ is defined as $\mathbf{W}_n \triangleq \text{Diag}\{W_l : l \in \mathcal{O}(n)\} \in \mathbb{C}^{|\mathcal{O}(n)| \times |\mathcal{O}(n)|}$, and $\mathbf{Q}_n$ is defined as $\mathbf{Q}_n \triangleq \text{Diag}\{\mathbf{Q}_l : l \in \mathcal{O}(n)\} \in \mathbb{C}^{|\mathcal{O}(n)| n_t \times |\mathcal{O}(n)| n_t}$. Moreover, we introduce two more matrices $\mathbf{E}_1^{(n)}$ and $\mathbf{E}_2^{(n)}$

as follows:

$$\mathbf{E}_1^{(n)} = \left[\begin{array}{c|c} \mathbf{I}_{|\mathcal{O}(n)|} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \in \mathbb{C}^{|\mathcal{O}(n)|(n_t+1)\times|\mathcal{O}(n)|(n_t+1)},$$

$$\mathbf{E}_2^{(n)} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{|\mathcal{O}(n)|n_t} \end{array}\right] \in \mathbb{C}^{|\mathcal{O}(n)|(n_t+1)\times|\mathcal{O}(n)|(n_t+1)}.$$

It is easy to recognize that if $\mathbf{D}_n \in \Omega_+(n)$, we have $\mathrm{Tr}(\mathbf{E}_1^{(n)}\mathbf{D}_n) = \sum_{l\in\mathcal{O}(n)} W_l \leq B_n$ and $\mathrm{Tr}(\mathbf{E}_2^{(n)}\mathbf{D}_n) = \sum_{l\in\mathcal{O}(n)} \mathrm{Tr}\left(\mathbf{Q}_l\right) \leq P_{\max}^{(n)}$, and $\mathbf{D}_n \succeq 0$. In this paper, Frobenius norm, which is the counterpart of the Euclidean norm in the vector space, is used as the matrix distance criterion. By the definition of Frobenius norm, the square distance of two matrices $\mathbf{A}$ and $\mathbf{B}$ is $\|\mathbf{A}-\mathbf{B}\|_F = \left(\mathrm{Tr}\left[(\mathbf{A}-\mathbf{B})^\dagger(\mathbf{A}-\mathbf{B})\right]\right)^{\frac{1}{2}}$. Thus, given the block diagonal matrix $\mathbf{D}_n$, we wish to find a matrix $\tilde{\mathbf{D}}_n \in \Omega_+(n)$ such that $\tilde{\mathbf{D}}_n$ minimizes $\|\tilde{\mathbf{D}}_n - \mathbf{D}_n\|_F$, i.e.,:

$$\begin{array}{rl} \textsf{Minimize} & \|\tilde{\mathbf{D}}_n - \mathbf{D}_n\|_F \\ \textsf{subject to} & \mathrm{Tr}(\mathbf{E}_1^{(n)}\tilde{\mathbf{D}}_n) \leq B_n \\ & \mathrm{Tr}(\mathbf{E}_2^{(n)}\tilde{\mathbf{D}}_n) \leq P_{\max}^{(n)} \\ & \tilde{\mathbf{D}}_n \succeq 0 \end{array} \tag{20}$$

For more convenient algebraic manipulations, we instead study the following equivalent optimization problem:

$$\begin{array}{rl} \textsf{Minimize} & \frac{1}{2}\left\|\tilde{\mathbf{D}}_n - \mathbf{D}_n\right\|_F^2 \\ \textsf{subject to} & \mathrm{Tr}(\mathbf{E}_1^{(n)}\tilde{\mathbf{D}}_n) \leq B_n \\ & \mathrm{Tr}(\mathbf{E}_2^{(n)}\tilde{\mathbf{D}}_n) \leq P_{\max}^{(n)} \\ & \tilde{\mathbf{D}}_n \succeq 0 \end{array} \tag{21}$$

Notice that the objective function of this minimization problem is convex in $\tilde{\mathbf{D}}_n$, that the constraint $\tilde{\mathbf{D}}_n \succeq 0$ represents the convex cone of positive semidefinite matrices, and that the constraints $\mathrm{Tr}(\mathbf{E}_1^{(n)}\tilde{\mathbf{D}}_n) \leq B_n$ and $\mathrm{Tr}(\mathbf{E}_2^{(n)}\tilde{\mathbf{D}}_n) \leq P_{\max}^{(n)}$ are linear constraints, the problem is a convex optimization problem and we can solve this minimization problem by solving its Lagrangian dual.

Introducing Hermitian matrix $\mathbf{X}$ to the constraint $\tilde{\mathbf{D}}_n \succeq 0$, ν to the constraint $\mathrm{Tr}(\mathbf{E}_1^{(n)}\tilde{\mathbf{D}}_n) \leq B_n$, and μ to the constraint $\mathrm{Tr}(\mathbf{E}_2^{(n)}\tilde{\mathbf{D}}_n) \leq P_{\max}^{(n)}$, we can write the Lagrangian as

$$\begin{aligned} g(\mathbf{X},\nu,\mu) &= \min_{\tilde{\mathbf{D}}_n} \left\{ \frac{1}{2}\left\|\tilde{\mathbf{D}}_n - \mathbf{D}_n\right\|_F^2 - \mathrm{Tr}(\mathbf{X}^\dagger \tilde{\mathbf{D}}_n) \right. \\ &\quad \left. +\nu\left(\mathrm{Tr}[\mathbf{E}_1^{(n)}\tilde{\mathbf{D}}_n] - B_n\right) + \mu\left(\mathrm{Tr}[\mathbf{E}_2^{(n)}\tilde{\mathbf{D}}_n] - P_{\max}^{(n)}\right) \right\}. \end{aligned} \tag{22}$$

Since $g(\mathbf{X},\nu,\mu)$ is convex quadratic function in $\tilde{\mathbf{D}}_n$ and $\tilde{\mathbf{D}}_n$ becomes unconstrained after moving the positive semidefinite constraint to the objective function, we can compute the minimizer of (22) by simply setting the derivative of (22) to zero, i.e.,

$$(\tilde{\mathbf{D}}_n - \mathbf{D}_n) - \mathbf{X}^\dagger + \nu\mathbf{E}_1^{(n)} + \mu\mathbf{E}_2^{(n)} = 0.$$

Noting that $\mathbf{X}^\dagger = \mathbf{X}$, we have

$$\tilde{\mathbf{D}}_n = \mathbf{D}_n + \mathbf{X} - \nu\mathbf{E}_1^{(n)} - \mu\mathbf{E}_2^{(n)}. \tag{23}$$

Substituting $\tilde{\mathbf{D}}_n$ back into (22), we have

$$\begin{aligned} g(\mathbf{X},\nu,\mu) &= \frac{1}{2}\left\|\mathbf{X} - \nu\mathbf{E}_1^{(n)} - \mu\mathbf{E}_2^{(n)}\right\|_F^2 - \nu B_n - \mu P_{\max}^{(n)} + \\ &\quad \mathrm{Tr}\left[\left(\nu\mathbf{E}_1^{(n)} + \mu\mathbf{E}_2^{(n)} - \mathbf{X}\right)\left(\mathbf{D}_n + \mathbf{X} - \nu\mathbf{E}_1^{(n)} - \mu\mathbf{E}_2^{(n)}\right)\right] \\ &= -\frac{1}{2}\left\|\mathbf{D}_n - \nu\mathbf{E}_1^{(n)} - \mu\mathbf{E}_2^{(n)} + \mathbf{X}\right\|_F^2 - \nu\mathbf{B}_n - \mu P_{\max}^{(n)} + \frac{1}{2}\|\mathbf{D}_n\|^2 \end{aligned} \tag{24}$$

Therefore, the Lagrangian dual problem can be written as

$$\begin{array}{ll} \textsf{Maximize} & -\frac{1}{2}\left\|\mathbf{D}_n - \nu\mathbf{E}_1^{(n)} - \mu\mathbf{E}_2^{(n)} + \mathbf{X}\right\|_F^2 - \nu\mathbf{B}_n - \mu P_{\max}^{(n)} + \frac{1}{2}\|\mathbf{D}_n\|^2 \\ \textsf{subject to} & \mathbf{X} \succeq 0, \nu \geq 0, \mu \geq 0 \end{array} \tag{25}$$

After solving (25) and by saddle-point optimality condition, we can have the optimal solution to the primal problem as:

$$\tilde{\mathbf{D}}_n^* = \mathbf{D}_n - \nu^*\mathbf{E}_1^{(n)} - \mu^*\mathbf{E}_2^{(n)} + \mathbf{X}^*, \tag{26}$$

where ν^*, μ^* and $\mathbf{X}^*$ are the optimal solutions to Lagrangian dual problem in (25).

Although the Lagrangian dual problem in (25) has a similar structure as that in the primal problem (having a positive semidefinitive matrix constraint), we find that this positive semidefinite matrix constraint can indeed be easily handled. To see this, we first introduce Moreau Decomposition Theorem from convex analysis [40].

Theorem 2. *(Moreau Decomposition) Let $\mathcal{K}$ be a closed convex cone. For $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^p$, the two properties below are equivalent:*

1. $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ *with* $\mathbf{x}_1 \in \mathcal{K}$, $\mathbf{x}_2 \in \mathcal{K}^o$ *and* $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = 0$,
2. $\mathbf{x}_1 = p_{\mathcal{K}}(\mathbf{x})$ *and* $\mathbf{x}_2 = p_{\mathcal{K}^o}(x)$,

where $\mathcal{K}^o \triangleq \{\mathbf{s} \in \mathbb{C}^p : \langle \mathbf{s}, \mathbf{y} \rangle \leq 0, \forall\, \mathbf{y} \in \mathcal{K}\}$ *is called the polar cone of cone* $\mathcal{K}$, $p_{\mathcal{K}}(\cdot)$ *represents the projection onto cone* $\mathcal{K}$.

In fact, we can see from Moreau decomposition that the properties of the projection onto a cone $\mathcal{K}$ is analogous to the projection onto a subspace, and the only difference is that the orthogonal subspace is now replaced by the polar cone.

From matrix theory, we know that the eigenvalues of a Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ are always real. Suppose that we sort these eigenvalues of $\mathbf{A}$, denoted by λ_i, $i = 1, \ldots, p$, in non-increasing order, i.e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, and perform eigenvalue decomposition on $\mathbf{A}$ yielding $\mathbf{A} = \mathbf{U}\mathrm{Diag}\{\lambda_i : i = 1, \ldots, p\}\mathbf{U}^\dagger$, where $\mathbf{U}$ is the unitary matrix formed by the eigenvectors corresponding to the non-increasing eigenvalues, then we have the positive semidefinite and negative semidefinite projections of $\mathbf{A}$ as follows:

$$\mathbf{A}_+ = \mathbf{U}\begin{bmatrix} \max\{\lambda_1, 0\} & & \\ & \ddots & \\ & & \max\{\lambda_p, 0\} \end{bmatrix}\mathbf{U}^\dagger, \quad \mathbf{A}_- = \mathbf{U}\begin{bmatrix} \min\{\lambda_1, 0\} & & \\ & \ddots & \\ & & \min\{\lambda_p, 0\} \end{bmatrix}\mathbf{U}^\dagger. \tag{27}$$

The proof of the results in (27) is a straightforward application of Theorem 2 by noting that $\mathbf{A}_+ \succeq 0$, $\mathbf{A}_- \preceq 0$, $\langle \mathbf{A}_+, \mathbf{A}_- \rangle = 0$, $\mathbf{A}_+ + \mathbf{A}_- = \mathbf{A}$, and the positive semidefinite cone and negative semidefinite cone are polar cones to each other.

Now we consider the term $\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} + \mathbf{X}$, which is the only term involving $\mathbf{X}$ in the dual objective function. We can rewrite it as $\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} - (-\mathbf{X})$, where we note that $-\mathbf{X} \preceq 0$. Finding a negative semidefinite matrix $-\mathbf{X}$ such that $\|\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} - (-\mathbf{X})\|_F$ is minimized is equivalent to finding the projection of $\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)}$ onto the negative semidefinite cone. From the previous discussions, we immediately have

$$-\mathbf{X} = \left(\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_- . \tag{28}$$

Substituting (28) back to the Lagrangian dual objective function, we have

$$\min_{\mathbf{X}} \left\| \mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} + \mathbf{X} \right\|_F = \left(\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_+ . \tag{29}$$

Thus, the matrix variable $\mathbf{X}$ in the Lagrangian dual problem can be analytically solved and the Lagrangian dual problem can be simplified to

$$\begin{array}{ll} \textsf{Maximize} & \psi(\nu, \mu) \triangleq -\frac{1}{2} \left\| \left(\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_+ \right\|_F^2 - \nu \mathbf{B}_n - \mu P_{\max}^{(n)} + \frac{1}{2} \|\mathbf{D}_n\|^2 \\ \textsf{subject to} & \nu \geq 0, \mu \geq 0 \end{array} \tag{30}$$

Suppose that after performing eigenvalue decomposition, we have $\mathbf{D}_n = \mathbf{U}_n \mathbf{\Lambda}_n \mathbf{U}_n^\dagger$, where $\mathbf{\Lambda}_n$ is the diagonal matrix formed by the eigenvalues of $\mathbf{D}_n$, $\mathbf{U}_n$ is the unitary matrix formed by the corresponding eigenvectors. From the fact that $\mathbf{E}_1^{(n)} = \mathbf{U}_n \mathbf{E}_1^{(n)} \mathbf{U}_n^\dagger$ and $\mathbf{E}_2^{(n)} = \mathbf{U}_n \mathbf{E}_2^{(n)} \mathbf{U}_n^\dagger$, we have

$$\left(\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_+ = \mathbf{U}_n \left(\mathbf{\Lambda}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_+ \mathbf{U}_n^\dagger . \tag{31}$$

Since $\mathbf{U}_n$ is unitary, we have

$$\left\| \left(\mathbf{D}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_+ \right\|_F^2 = \left\| \left(\mathbf{\Lambda}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_+ \right\|_F^2 . \tag{32}$$

In particular, we denote the eigenvalues in $\mathbf{\Lambda}_n$ corresponding to $\mathbf{W}_n$ and $\mathbf{Q}_n$ by $\lambda_i^{(\mathbf{W}_n)}$ and $\lambda_j^{(\mathbf{Q}_n)}$, respectively, and sort the eigenvalues in these two groups in non-increasing order as follows:

$$\mathbf{\Lambda}_n = \begin{bmatrix} \lambda_1^{(\mathbf{W}_n)} & & & & & \\ & \ddots & & & & \\ & & \lambda_{|\mathcal{O}(n)|}^{(\mathbf{W}_n)} & & & \\ & & & \lambda_1^{(\mathbf{Q}_n)} & & \\ & & & & \ddots & \\ & & & & & \lambda_{|\mathcal{O}(n)|\times n_t}^{(\mathbf{Q}_n)} \end{bmatrix}, \tag{33}$$

where $\lambda_1^{(\mathbf{W}_n)} \geq \ldots \geq \lambda_{|\mathcal{O}(n)|}^{(\mathbf{W}_n)}$ and $\lambda_1^{(\mathbf{Q}_n)} \geq \ldots \geq \lambda_{|\mathcal{O}(n)|\times n_t}^{(\mathbf{Q}_n)}$. It then follows that

$$\left\| \left(\mathbf{\Lambda}_n - \nu \mathbf{E}_1^{(n)} - \mu \mathbf{E}_2^{(n)} \right)_+ \right\|_F^2 = \sum_{i=1}^{|\mathcal{O}(n)|} \left(\max\left\{0, \lambda_i^{(\mathbf{W}_n)} - \nu \right\} \right)^2 + \sum_{j=1}^{|\mathcal{O}(n)|n_t} \left(\max\left\{0, \lambda_j^{(\mathbf{Q}_n)} - \mu \right\} \right)^2, \tag{34}$$

From (34), we have

$$\begin{aligned} \psi(\nu,\mu) &= -\frac{1}{2} \sum_{i=1}^{|\mathcal{O}(n)|} \left(\max\left\{0, \lambda_i^{(\mathbf{W}_n)} - \nu \right\} \right)^2 & (35) \\ &\quad -\nu B_n - \frac{1}{2} \sum_{j=1}^{|\mathcal{O}(n)|n_t} \left(\max\left\{0, \lambda_j^{(\mathbf{Q}_n)} - \mu \right\} \right)^2 - \mu P_{\max}^{(n)} + \frac{1}{2}\|\mathbf{D}_n\|^2 & \\ &= \psi(\nu) + \psi(\mu) + \frac{1}{2}\|\mathbf{D}_n\|^2, & (36) \end{aligned}$$

where $\psi(\nu) \triangleq -\frac{1}{2}\sum_{i=1}^{|\mathcal{O}(n)|}(\max\{0, \lambda_i^{(\mathbf{W}_n)} - \nu\})^2 - \nu B_n$ and $\psi(\mu) \triangleq -\frac{1}{2}\sum_{j=1}^{|\mathcal{O}(n)|n_t}(\max\{0, \lambda_j^{(\mathbf{Q}_n)} - \mu\})^2 - \mu P_{\max}^{(n)}$, i.e., we separate $\psi(\nu,\mu)$ into two parts. It is readily verifiable that $\psi(\nu,\mu)$ is continuous and piece-wise concave in ν and μ. Generally, for a piece-wise concave maximization problem, we can use subgradient method. In this problem, it is easy to derive the subgradients with respect to ν and μ as follows:

$$\frac{\partial \psi}{\partial \nu} = \sum_{i=1}^{|\mathcal{O}(n)|} \max\left\{0, \lambda_i^{(\mathbf{W}_n)} - \nu \right\} - B_n, \quad \frac{\partial \psi}{\partial \mu} = \sum_{j=1}^{|\mathcal{O}(n)|\times n_t} \max\left\{0, \lambda_i^{(\mathbf{Q}_n)} - \mu \right\} - P_{\max}^{(n)}. \tag{37}$$

However, due to the heuristic nature of its step size selection strategy, subgradient algorithm usually does not perform well. Noting that $\psi(\nu,\mu)$ is piece-wise quadratic and separable, we can consider ν and μ individually piece by piece.

For example, we can start searching the optimal value of ν as follows. Let the pieces of $\psi(\nu)$ indexed by $\hat{I}$, $\hat{I}=0,1,\ldots,|\mathcal{O}\left(n\right)|$. Initially we set $\hat{I}=0$ and increase $\hat{I}$ subsequently. Also, we introduce $\lambda_0^{(\mathbf{W}_n)}=\infty$ and $\lambda_{|\mathcal{O}(n)|+1}^{(\mathbf{W}_n)}=-\infty$. If $\hat{I}=0$, we let endpoint objective value $\psi_{\hat{I}}\left(\lambda_0^{(\mathbf{W}_n)}\right)=0$, $\phi^*=\psi_{\hat{I}}\left(\lambda_0^{(\mathbf{W}_n)}\right)$, and let $\nu^*=\lambda_0^{(\mathbf{W}_n)}$. If $\hat{I}>|\mathcal{O}\left(n\right)|$, the search stops. For a particular index $\hat{I}$, suppose that $\nu\in\left[\lambda_{\hat{I}+1}^{(\mathbf{W}_n)},\lambda_{\hat{I}}^{(\mathbf{W}_n)}\right]\cap\mathbb{R}_+$, where $\mathbb{R}_+$ denotes the set of non-negative real numbers. Solve for $\nu_{\hat{I}}^*$ by setting

$$\frac{\partial}{\partial\nu}\psi_{\hat{I}}(\nu)\triangleq\frac{\partial}{\partial\nu}\left(-\frac{1}{2}\sum_{i=1}^{\hat{I}}\left(\lambda_i^{(\mathbf{W}_n)}-\nu\right)^2-\nu B_n\right)=0, \tag{38}$$

for which we have

$$\nu_{\hat{I}}^*=\frac{\sum_{i=1}^{\hat{I}}\lambda_i^{(\mathbf{W}_n)}-B_n}{\hat{I}}. \tag{39}$$

Now we consider the following two cases:

1. If $\nu_{\hat{I}}^*\in\left[\lambda_{\hat{I}+1}^{(\mathbf{W}_n)},\lambda_{\hat{I}}^{(\mathbf{W}_n)}\right]\cap\mathbb{R}_+$, then we find the optimal solution for ν because $\psi(\nu,\mu)$ is continuous concave quadratic in ν, and the point having zero-value first derivative, if exists, must be the unique global maximum solution. Thus, we can let $\nu^*=\nu_{\hat{I}}^*$ and the search is done.

2. If $\nu_{\hat{I}}^*\notin[\lambda_{\hat{I}+1}^{(\mathbf{W}_n)},\lambda_{\hat{I}}^{(\mathbf{W}_n)}]\cap\mathbb{R}_+$, we must have that the local maximum in the interval $\left[\lambda_{\hat{I}+1}^{(\mathbf{W}_n)},\lambda_{\hat{I}}^{(\mathbf{W}_n)}\right]\cap\mathbb{R}_+$ is achieved at one of the two end points. We note that the objective value $\psi_{\hat{I}}\left(\lambda_{\hat{I}}^{(\mathbf{W}_n)}\right)$ has been computed in the previous iteration. This is because from the continuity of the objective function, we have $\psi_{\hat{I}}\left(\lambda_{\hat{I}}^{(\mathbf{W}_n)}\right)=\psi_{\hat{I}-1}\left(\lambda_{\hat{I}}^{(\mathbf{W}_n)}\right)$. Thus, we only need to compute the other

endpoint objective value $\psi_{\hat{I}}\left(\lambda_{\hat{I}+1}^{(\mathbf{W}_n)}\right)$. If $\psi_{\hat{I}}\left(\lambda_{\hat{I}+1}^{(\mathbf{W}_n)}\right) < \psi_{\hat{I}}\left(\lambda_{\hat{I}}^{(\mathbf{W}_n)}\right) = \phi^*$, then we know ν^* is the optimal solution; else let $\nu^* = \lambda_{\hat{I}+1}^{(\mathbf{W}_n)}$, $\phi^* = \psi_{\hat{I}}\left(\lambda_{\hat{I}+1}^{(\mathbf{W}_n)}\right)$, let $\hat{I} = \hat{I} + 1$ and continue.

Since there are $|\mathcal{O}(n)| + 1$ intervals, the search process takes at most $|\mathcal{O}(n)| + 1$ steps to find the optimal solution ν^*. Likewise, the search process for μ can be done in a similar fashion.

After finding ν^* and μ^*, we have

$$\tilde{\mathbf{D}}_n^* = \left(\mathbf{D}_n - \nu^* \mathbf{E}_1^{(n)} - \mu^* \mathbf{E}_2^{(n)}\right)_+ = \mathbf{U}_n \left(\mathbf{\Lambda}_n - \nu^* \mathbf{E}_1^{(n)} - \mu^* \mathbf{E}_2^{(n)}\right)_+ \mathbf{U}_n^\dagger, \tag{40}$$

that is, the projection $\tilde{\mathbf{D}}_n$ can be computed by adjusting the eigenvalues of $\mathbf{D}_n$ using ν^* and μ^* and keeping the eigenvectors unchanged.

The projection of $\mathbf{D}_n$ onto $\Omega_+(n)$ is summarized in Algorithm 2 and Algorithm 3.

Algorithm 2 Projection onto $\Omega_+(n)$

```
1. Construct a block diagonal matrix $\mathbf{D}_n$. Perform eigenvalue decomposition $\mathbf{D}_n = \mathbf{U}_n \mathbf{\Lambda}_n \mathbf{U}_n^\dagger$, separate the eigenvalues in two groups corresponding to $\mathbf{W}_n$ and $\mathbf{Q}_n$, and sort them in non-increasing order within each group, respectively.
2. For each group of eigenvalues, call Algorithm 3 to find the optimal dual variable $\nu^*$ and $\mu^*$.
3. Compute $\tilde{\mathbf{D}}_n = \mathbf{U}_n(\mathbf{\Lambda}_n - \nu^* \mathbf{E}_1^{(n)} - \mu^* \mathbf{E}_2^{(n)})_+ \mathbf{U}^\dagger$.
```

Algorithm 3 Search the Optimal Dual Variable

```
Initiation:
  Introduce $\lambda_0 = \infty$ and $\lambda_K = -\infty$. Let $\hat{I} = 0$. Let endpoint objective $\psi_{\hat{I}}(\lambda_0) = 0$, $\phi^* = \psi_{\hat{I}}(\lambda_0)$, and $\mu^* = \lambda_0$.
Main Loop:
  1. If $\hat{I} > K$, return $\mu^*$; else let $\mu_{\hat{I}}^* = (\sum_{j=1}^{\hat{I}} \lambda_j - P)/\hat{I}$.
  2. If $\mu_{\hat{I}}^* \in [\lambda_{\hat{I}+1}, \lambda_{\hat{I}}] \cap \mathbb{R}_+$, then let $\mu^* = \mu_{\hat{I}}^*$ and return $\mu^*$.
  3. Compute $\psi_{\hat{I}}(\lambda_{\hat{I}+1})$. If $\psi_{\hat{I}}(\lambda_{\hat{I}+1}) < \phi^*$, then return $\mu^*$; else let $\mu^* = \lambda_{\hat{I}+1}$, $\phi^* = \psi_{\hat{I}}(\lambda_{\hat{I}+1})$, $\hat{I} = \hat{I} + 1$ and continue.
```

V Cutting-Plane Method for Solving $\mathrm{D}^{\mathrm{CRPA}}$

Compared to the popular subgradient-based approaches for solving Lagrangian dual problems, the attractive feature of the cutting-plane method is its speed of convergence and its simplicity in recovering optimal primal feasible solutions. As opposed to the cumbersomeness of subgradient method, in cutting-plane method, primal optimal feasible solutions can be exactly computed by averaging all the primal solutions (may or may not be primal feasible) using the dual variables as weights [34].

We briefly introduce the basic idea of cutting-plane method as follows. Letting $z = \Theta(\mathbf{u})$, the inequality $z \geq \sum_f \ln(s_f) + \sum_l u_l \left(\Phi_l(W_l, \mathbf{Q}_l) - \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle\right)$ must hold for all $(\mathbf{S}, \mathbf{X}, \mathbf{Q}, \mathbf{W}) \in \Gamma$. Thus, the dual problem is equivalent to

$$\begin{array}{rl} \text{Minimize} & z \\ \text{subject to} & z \geq \sum_f \ln(s_f) + \sum_l u_l \left(\Phi_l(W_l, \mathbf{Q}_l) - \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle\right) \\ & \mathbf{u} \geq 0, \end{array} \tag{41}$$

where $(\mathbf{S}, \mathbf{X}, \mathbf{Q}, \mathbf{W}) \in \Gamma$. Although (41) is a linear program with infinite constraints not known explicitly, we can consider the following approximating problem:

$$\begin{array}{rl} \text{Minimize} & z \\ \text{subject to} & z \geq \sum_f \ln\left(s_f^{(j)}\right) + \sum_l u_l \left(\Phi_l(W_l^{(j)}, \mathbf{Q}_l^{(j)}) - \langle \mathbf{1}, \mathbf{X}^{(j)T} \mathbf{e}_l \rangle\right) \\ & \mathbf{u} \geq 0, \end{array} \tag{42}$$

where the points $(\mathbf{S}^{(j)}, \mathbf{X}^{(j)}), \mathbf{Q}^{(j)}, \mathbf{W}^{(j)}) \in \Gamma$, $j = 1, \ldots, k-1$. The problem in (42) is a linear program with a finite number of constraints and can be solved efficiently. Let $(z^{(k)}, \mathbf{u}^{(k)})$ be an optimal solution to the approximating problem, which we refer to as the *master program.* If the solution is feasible to (41), then it is an optimal solution to the Lagrangian dual problem. To check the feasibility, we consider the following *subproblem*:

$$\begin{array}{rl} \text{Maximize} & \sum_f \ln(s_f) + \sum_l u_l^{(k)} \left(\Phi_l(W_l, \mathbf{Q}_l) - \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle\right) \\ \text{subject to} & (\mathbf{S}, \mathbf{X}, \mathbf{Q}, \mathbf{W}) \in \Gamma \end{array} \tag{43}$$

Suppose that $(\mathbf{S}^{(k)}, \mathbf{X}^{(k)}, \mathbf{Q}^{(k)}, \mathbf{W}^{(k)})$ is an optimal solution to the subproblem (43) and $\Theta^*(\mathbf{u}^{(k)})$ is the corresponding optimal objective value. If $z_k \geq \Theta^*(\mathbf{u}^{(k)})$, then $\mathbf{u}^{(k)}$ is an optimal solution to the Lagrangian dual problem. Otherwise, for $\mathbf{u} = \mathbf{u}^{(k)}$, the inequality constraint in (41) is not satisfied for $(\mathbf{S}^{(j)}, \mathbf{X}^{(j)}, \mathbf{Q}^{(j)}, \mathbf{W}^{(j)})$. Thus, we can add the constraint

$$z \geq \sum_f \ln\left(s_f^{(k)}\right) + \sum_l u_l \left(\Phi_l(W_l^{(k)}, \mathbf{Q}_l^{(k)}) - \langle \mathbf{1}, \mathbf{X}^{(k)T}\mathbf{e}_l \rangle\right) \tag{44}$$

to (42), and re-solve the master linear program. Obviously, $(z^{(k)}, \mathbf{u}^{(k)})$ violates (44) and will be cut off by (44). The cutting plane algorithm is summarized in Algorithm 4.

Algorithm 4 Cutting Plane Algorithm for Solving $\mathbf{D}^{\text{CRPA}}$

```
Initialization:
  Find a point $(\mathbf{S}^{(0)}, \mathbf{X}^{(0)}, \mathbf{Q}^{(0)}, \mathbf{W}^{(0)}) \in \Gamma$. Let $k = 1$.
Main Loop:
  1. Solve the master program in (42). Let $(z^{(k)}, \mathbf{u}^{(k)})$ be an optimal solution.
  2. Solve the subproblem in (43). Let $(\mathbf{S}^{(k)}, \mathbf{X}^{(k)}, \mathbf{Q}^{(k)}, \mathbf{W}^{(k)})$ be an optimal point, and let $\Theta^*(\mathbf{u}^{(k)})$ be the
     corresponding optimal objective value.
  3. If $z^{(k)} \geq \Theta(\mathbf{u}^{(k)})$, then stop with $\mathbf{u}^{(k)}$ as the optimal dual solution. Otherwise, add the constraint (44) to the
     master program, replace $k$ by $k+1$, and go to step 1.
```

VI Subgradient Algorithm for Solving $\mathbf{D}^{\text{CRPA}}$and Distributed Implementation

Since the CRPA primal problem is not a concave problem, the dual objective function is piece-wise differentiable in general. For piece-wise differentiable convex minimization problems, subgradient method is one of the effective algorithms [34]. Let S be a nonempty convex set in $\mathbb{R}^n$ and let $f : S \to \mathbb{R}$ be a convex piece-wise differentiable function. Then $\mathbf{d}$ is called a subgradient of f at $\bar{\mathbf{x}} \in S$ if $f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \mathbf{d}^T(\mathbf{x} - \bar{\mathbf{x}})$, $\forall \mathbf{x} \in S$. The geometric interpretation of a subgradient corresponds to the slope of a possible supporting hyperplane of f at $\bar{\mathbf{x}}$. On the other hand, if f is concave, then the inequality is reversed.

Subgradient algorithm for minimization problems is a generalization of steepest descent algorithm in which the negative gradient direction is replaced by a suitable negative subgradient direction. For $\mathbf{D}^{\text{CRPA}}$, starting with an initial $\mathbf{u}^{(1)}$ and after evaluating subproblems $\mathbf{D}_{\text{net}}^{\text{CRPA}}$ and $\mathbf{D}_{\text{link}}^{\text{CRPA}}$ for $\mathbf{u}^{(k)}$ in the k^{th} iteration, we update the dual variables by

$$\mathbf{u}^{(k+1)} = \left[\mathbf{u}^k - \lambda_k \mathbf{d}^{(k)}\right]_+ . \tag{45}$$

In (45), the operator $[\cdot]_+$ projects a vector on to the nonnegative orthant, and λ_k denotes a positive scalar step size. $\mathbf{d}^{(k)}$ is a subgradient of the Lagrangian at point $\mathbf{u}^{(k)}$. It is proven in [34] that the subgradient algorithm converges if the step size λ_k satisfies $\lambda_k \to 0$ as $k \to \infty$ and $\sum_{k=0}^{\infty} \lambda_k = \infty$. However, one has to carefully select the step size to avoid stalling and accelerate the convergence. It is shown in [34] that the best choice of the step size λ_k is

$$\lambda_k = \frac{\beta_k[\Theta(\mathbf{u}^{(k)}) - \hat{\Theta}]}{\|\mathbf{d}_k\|}, \tag{46}$$

where $\beta_k > 0$ and $\hat{\Theta}$ is an estimate of the optimal value of Θ. However, the optimal step size selection strategy still requires global information. Another simple and useful step size selection strategies is the divergent harmonic series $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$. We will discuss this choice of step sizes in the next subsection in more detail.

From [34], we have that the subgradient for the Lagrangian dual problem is

$$\Phi_l(\mathbf{Q}^*(\mathbf{u})) - \langle \mathbf{1}, \mathbf{X}^*(\mathbf{u})^T \mathbf{e}_l \rangle, \quad l = 1, 2, \ldots, L. \tag{47}$$

In fact, the subgradient expression in (47) has a very interesting economics interpretation. Upon a closer look, one may immediately recognize that the expression represents the excess capacity on link l. That is, if d_l is positive, link l is not fully utilized, and more traffic can pass through; otherwise, the link is saturated or over utilized. Further, it is not difficult to see that the Lagrangian dual variable u_l plays the role of pricing on link l for each unit flow. The network layer acts like a customer and the link layer acts like a service provider. From the network layer objective function $\Theta_{\text{link}} = \max \left(\sum_f \ln(s_f) - \sum_l u_l \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \right)$, it is seen that the customer wants to maximize his net profit under the current prices $\mathbf{u}$. While from the link layer objective function $\Theta_{\text{link}} = \max \sum_l u_l \Phi_l(\mathbf{Q})$, it can be seen that the service provider wants to maximize his overall revenue by adjusting the links' capacities according to the current prices $\mathbf{u}$.

Recall that the dual variables update scheme is $\mathbf{u}^{(k+1)} = \left[\mathbf{u}^k - \lambda_k \mathbf{d}^{(k)}\right]_+$, where $\lambda_k > 0$ is the step size at iteration k. If a link l is under utilized, then $d_l^{(k)}$ is positive, which will in turn decrease the price $u_l^{(k+1)}$ in the $(k+1)^{th}$ iteration. This price reduction on link l will encourage the customer to use link l in the next iteration, and prompt the service provider to reduce the capacity on link l in the next iteration. On the other hand, if a link l is over utilized, then $d_l^{(k)}$ is negative, which will in turn increase the price $u_l^{(k+1)}$ in the $(k+1)^{th}$ iteration. This price increasing on link l will discourage the customer from using link l in the next iteration, and prompt the service provider to increase the capacity of link l because link l becomes more valuable. Eventually, when the subgradient algorithm converges, i.e., $\mathbf{d} \approx \mathbf{0}$, it means that every link in the network is almost fully utilized, and the prices' adjustments from iteration to iteration are so minor that it is no longer necessary for both the customer and the service provider to make changes to increase their profits. In summary, the subgradient algorithm for $\mathbf{MD}^{\text{CRPA}}$ can be regarded a pricing scheme that coordinates both the customer and the service provider simultaneously, driving them to a steady state where all links are fully utilized.

A Distributed Implementation of Subgradient Algorithm

As mentioned earlier, the Lagrangian dual problem is in the form of decomposed structure, which renders itself solvable by subgradient algorithm. Although the original motivation to apply the subgradient algorithm is due to the non-differentiability of Lagrangian dual objective function (piecewise concave/convex function in general), the fact that the subgradient method can be implemented in a distributed fashion is the main motivation for pursuing along this direction. Specifically, the subgradient method has the following properties when it comes to distributed implementation:

- Subgradient computation only requires local traffic information $\langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle$ and the available link capacity information $\Phi_l(W_l, \mathbf{Q}_l)$ at each link l: The subgradient can be computed as $\partial \Theta(\mathbf{u})/\partial u_l = \Phi_l(W_l, \mathbf{Q}_l) - \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle$. In this case, subgradient only involves local variables W_l, $\mathbf{Q}_l$, and $\mathbf{X}^T \mathbf{e}_l$ at each link and thus can be computed locally.

- The choice of step sizes can be chosen as $\lambda_k = \beta/k$, $k = 1, 2, \ldots$, where $0 < \beta \leq 1$ is a predefined constant. This step size choice obviously satisfies the convergence condition. It is seen that this choice of step size depends only upon the iteration index k, and does not require any other global knowledge. In conjunction with the first property, the dual variable, in the iterative form of $u_l^{(k+1)} = u_l^{(k)} + \lambda_k(\partial \Theta_l(\mathbf{u})/\partial u_l)$, may also be computed locally.

- The objective functions Θ_{link} and Θ_{net} can be decomposed such that each node in the network can perform the computation locally. Recall that the link layer subproblem is: Max $\Theta_{\text{link}} \triangleq \sum_{l=1}^{L} u_l C_l(\mathbf{Q}_l)$ s.t. $\sum_{l \in \mathcal{O}(n)} \text{Tr}\{\mathbf{Q}_l\} \le P_{\max}, \mathbf{Q}_l \succeq 0,\ \forall l$, where $\mathcal{O}(n)$ denotes the set of outgoing links from node n. For this subproblem, it is not difficult to observe that it has a special structure that can be decomposed into a set of new subproblems based on each node n as follows:

$$\text{Max } \Theta_{\text{link}}^{(n)} \triangleq u_l C_l(\mathbf{Q}_l) \quad \text{s.t.} \sum_{l \in \mathcal{O}(n)} \text{Tr}\{\mathbf{Q}_l\} \le P_{\max}, \mathbf{Q}_l \succeq 0, l \in \mathcal{O}(n). \tag{48}$$

The dual original link layer subproblem can then be simply transformed to $\Theta_{\text{link}} = \sum_{n=1}^{N} \Theta_{\text{link}}^{(n)}$. This suggests that the optimization of each problem $\Theta_{\text{link}}^{(n)}$ in (48) only requires local information of channel gains (all outgoing links from node n) and the locally-computed dual variable u_l. Thus, the link layer subproblem can be solved distributively. Likewise, for the network layer subproblem: Max $\Theta_{\text{net}} \triangleq \sum_{f=1}^{F} \ln(s_f) - \sum_l u_l \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle$, s.t. flow balance constraints for all flows, we may decompose it into a set of subproblems based on the source node of each flow f:

$$\text{Max } \Theta_{\text{net}}^{(f)} \triangleq \ln(s_f) - \sum_l u_l \langle \mathbf{1}, \mathbf{X}^T \mathbf{e}_l \rangle \text{ s.t. flow balance constraints for flow } f. \tag{49}$$

The original network layer subproblem can then be simply transformed to $\Theta_{\text{net}} = \sum_{f=1}^{F} \Theta_{\text{net}}^{(f)}$. Again, this suggests that the optimization of each problem $\Theta_{\text{net}}^{(f)}$ in (49) only requires the locally-computed dual variable u_l for the links. In each iteration, all links can send the locally-computed dual information back to the source node of each session. As a result, the network layer subproblem can be solved in a distributed fashion.

The key steps in this distributed algorithm include:

1. Initialize the iteration index $k = 0$, and choose initial values for $u_l^{(0)}$, for all l.
2. At each node n, solve the local link layer subproblem $\Theta_{\text{link}}^{(l)}$ based on $u_l^{(k)}$, where $l \in \mathcal{O}\left(n\right)$.
3. Each node sends the locally computed dual variable, or forwards the dual information upon receiving the dual information from its one-hop neighbors to each source node of each flow.
4. At the source node of each session, upon collecting all the required dual information, the source node computes the optimal flow $\mathbf{S}$ and $\mathbf{X}$, and then adjusts flow rates according to the optimal solution.
5. Based on current values of dual variables $u_l^{(k)}$ and the iteration number k, compute $u_l^{(k+1)}$. If $u_l^{(k+1)} - u_l^{(k)} < \epsilon$, or k has reached a predefined number of iterations, the algorithm stops. Otherwise, let $k = k + 1$ and continue.

B Recover the Primal Feasible Solution

Generally, additional work is usually needed to find an optimal primal feasible solution from the solutions of dual problem. As the dual function Θ is evaluated at a given $\mathbf{u}$, we obtain a solution, denoted by $[\mathbf{S}_{\mathbf{u}}^* \, \mathbf{X}_{\mathbf{u}}^* \, \mathbf{W}_{\mathbf{u}}^*, \mathbf{Q}_{\mathbf{u}}^*]^T$, to a problem that is closely related to the original problem, in which the constraints are perturbed. Let $I = \{l : u_l > 0\}$ represent the set of link indices whose Lagrangian multipliers are strictly positive. For some $\epsilon > 0$, if $\left|\Phi_l(\mathbf{W}_{\mathbf{u}}^*, \mathbf{Q}_{\mathbf{u}}^*) - \langle \mathbf{1}, \mathbf{X}_{\mathbf{u}}^{*T} \mathbf{e}_l \rangle\right| \leq \epsilon$, $\forall l \in I$, we have a *near-feasible solution.* Furthermore, suppose that $\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_F$ solve CRPA, we must have [34]

$$\sum_f \ln\left((\mathbf{S}_{\mathbf{u}}^* \mathbf{e}_f)_{\text{src}(f)}\right) \leq \sum_f \ln\left(\bar{s}_f\right) + \epsilon \sum_{l \in I} u_l.$$

Thus, $[\mathbf{S}^* \, \mathbf{X}^* \, \mathbf{W}^*, \mathbf{Q}^*]^T$ is also called a *near-optimal solution.* Based on this, one may achieve a primal feasible solution from the near-optimal solution by enforcing the violated constraints to be feasible, e.g., fixing the $\mathbf{Q}^*$ and solving $\mathbf{S}^*$ and $\mathbf{X}^*$ by solving a convex programming problem.

However, there is no primal optimality guarantee for such achieved primal feasible solutions. In this paper, we adopt the following primal feasible optimal solution recovering scheme that only requires the solution and step size information from the process of solving the Lagrangian dual problem. Let $[\mathbf{S}^*_{\mathbf{u}_k}\,\mathbf{X}^*_{\mathbf{u}_k}\,\mathbf{Q}^*_{\mathbf{u}_k}]^T$ be the solvers for $\Theta_{\mathbf{u}_k}$, and let λ_k be the corresponding step size in the subgradient algorithm. The primal optimal feasible solution recovery is shown in Algorithm 5.

Algorithm 5 Recovering A Primal Feasible Solution

```
1. In the $k^{th}$ iteration of the subgradient algorithm, let weights $\omega_j^k = 1/k$, for $j = 1, \ldots, k$.
2. Let $[\overline{\mathbf{S}}_k\,\overline{\mathbf{X}}_k\,\overline{\mathbf{W}}_k, \overline{\mathbf{Q}}_k]^T = \sum_{j=1}^{k} \omega_j^k [\mathbf{S}^*_{\mathbf{u}_j}\,\mathbf{X}^*_{\mathbf{u}_j}\,\mathbf{W}^*_{\mathbf{u}_j}\,\mathbf{Q}^*_{\mathbf{u}_j}]^T$
```

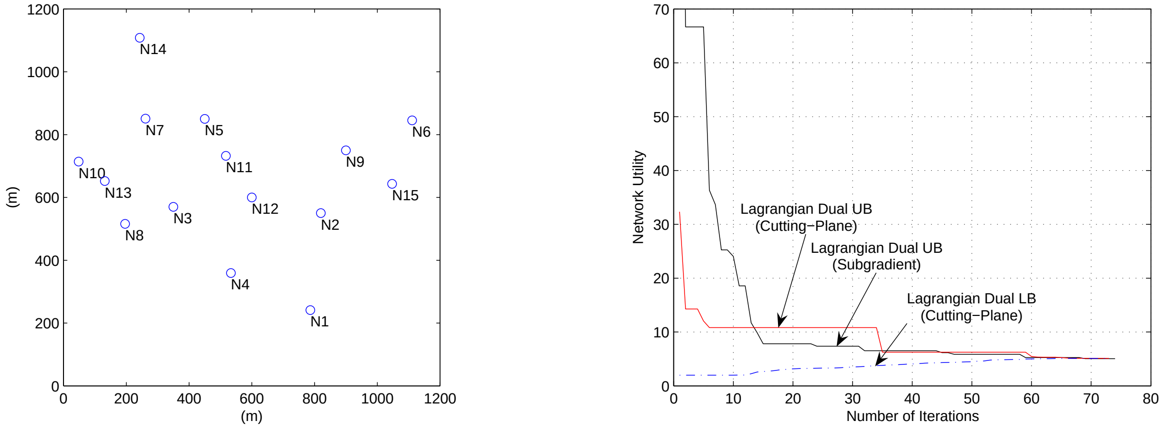
It is evident from our step size selection strategy that $\lambda_{k+1} \leq \lambda_k$ for all k and $\lim_{k\to\infty} k\lambda_k = \infty$. Thus, all conditions in Theorem 1 in [41] are satisfied, which means as $k \to \infty$, $[\overline{\mathbf{S}}_k\,\overline{\mathbf{X}}_k\,\overline{\mathbf{W}}_k, \overline{\mathbf{Q}}_k]^T$ will approach a KKT point.

VII Numerical Results

In this section, we present some pertinent numerical results through simulations to provide further insights on solving CRPA. We first describe our simulation settings. N nodes in a MIMO ad hoc network are uniformly distributed in a square region. Each node in the network is equipped with two transmitting antennas and two receiving antennas. The maximum transmit power for each node is set to $P_{\max} = 10$dBm. The network operates in 2.4GHz ISM band. The channel bandwidth is $W = 30$MHz. The path-loss index is chosen to be $\alpha = 2$.

We illustrate a 15-node network example, as shown in Fig. 1(a), to show the convergence process of the subgradient algorithm for solving $\mathbf{D}^{\text{CRPA}}$. All possible links (within the transmission range of using maximum power to transmit) are also illustrated in Fig. 1(a). In this example, there are three flows transmitting across the network: node 14 to node 1, node 6 to node 10, and node 5 to node 4, respectively. The convergence process is illustrated in Fig. 1(b).

The step size selection strategy for the distributed subgradient is $\lambda_k = \frac{0.1}{k}$. It can be seen that for this 15-node network example, both centralized cutting-plane algorithm and distributed



(a) Network Topology.

(b) Convergence behavior of cutting-plane (centralized) and subgradient (distributed) algorithms for the 15-node network example.

Figure 1: A 15-node MIMO-based mesh network example with $n_t = n_r = 2$.

subgradient algorithm converge in approximately 70 iterations, the optimal value of the network utility function is 5.10. The optimal routing paths for the flows are: 1) N13 → N5 → N11 → N4 → N1; 2) N6 → N9 → N11 → N7 → N10; 3) N5 → N11 → N12 → N4.

VIII Conclusion

In this paper, we investigated the problem of cross-layer optimization of routing, power control, power allocation, and bandwidth allocation for MIMO-based mesh networks. We developed a mathematical solution procedure, which combines Lagrangian decomposition, gradient projection, cutting-plane, and subgradient methods. We provided the theoretical insights of our proposed algorithms and conducted simulations to verify their efficacy. Our results show that the nice decoupled structure and the high efficiency of our proposed algorithm make it an attractive method for optimizing the performance of MIMO-based mesh networks.

References

[1] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 5, pp. 871–878, Jun. 1987.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–355, Mar. 1998.

[4] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[5] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 1, pp. 684–702, Jun. 2003.

[6] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[7] S. Cartreux, L. J. Greenstein, and P. F. Dressen, "Simulation results for an interference-limited multiple-input multiple-output cellular system," *IEEE Commun. Lett.*, vol. 4, no. 11, pp. 334–336, Nov. 2000.

[8] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[9] T. Lan and W. Yu, "Input optimization for multi-antenna broadcast channels and per-antenna power constraints," in *Proc. IEEE GLOBECOM*, Dallas, TX, U.S.A., Nov. 2004, pp. 420–424.

[10] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 802–811, Jun. 2003.

[11] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[12] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[13] H. Weingarten, Y. Steinberg, and S. S. (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[14] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[15] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[16] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

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

[18] ——, "A dual decomposition approach to the sum power Gaussian vector multiple-access channel sum capacity problem," in *Proc. Conf. Information Sciences and Systems (CISS)*, Baltimore, MD, U.S.A., 2003.

[19] J. Liu, Y. T. Hou, and H. D. Sherali, "Conjugate gradient projection method for multi-antenna Gaussian broadcast channels," *Technical Report, Deptment of ECE, Virginia Tech*, Nov. 2006. [Online]. Available: http://www.ece.vt.edu/thou/Publications/publications.html

[20] E. A. Jorswieck and H. Boche, "Performance analysis of capacity of MIMO systems under multiuser interference based on worst-case noise behavior," *EURASIP Journal on Wireless Communications and Networking*, no. 2, pp. 273–285, Feb. 2004.

[21] M. F. Demirkol and M. A. Ingram, "Power-controlled capacity for interfering MIMO links," in *Proc. IEEE Veh. Technol. Conf.*, Atlantic City, NJ, U.S.A., Oct. 2001, pp. 187–191.

[22] ——, "Stream control in network with interfereing MIMO links," in *Proc. IEEE WCNC*, New Orleans, LA, U.S.A., Mar. 2003, pp. 343–348.

[23] K. Sundaresan, R. Sivakumar, M. A. Ingram, and T.-Y. Chang, "Medium access control in ad hoc networks with MIMO links: Optimization considerations and algorithms," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 350–365, Oct. 2004.

[24] B. Chen and M. J. Gans, "MIMO communications in ad hoc networks," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2773–2783, Jul. 2006.

[25] M. Hu and J. Zhang, "MIMO ad hoc networks: Medium access control, saturation throughput, and optimal hop distance," *Special Issue on Mobile Ad Hoc Networks, Journal of Communications and Networks*, pp. 317–330, Dec. 2004.

[26] K. Sundaresan and R. Sivakumar, "Routing in ad hoc networks with MIMO links," in *Proc. IEEE ICNP*, Boston, MA, U.S.A., Nov. 2005, pp. 85–98.

[27] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Price-based distributed algorithms for rate-reliability tradeoff in network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 962–976, May 2006.

[28] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[29] A. Scaglione, D. L. Goeckel, and J. N. Laneman, "Cooperative communications in mobile ad hoc networks," *IEEE Signal Process. Mag.*, vol. 23, no. 5, pp. 18–29, Sep. 2006.

[30] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 2000.

[31] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows.* New York-Chichester-Brisbane-Toronto-Singapore: John Wiley & Sons Inc., 1990.

[32] F. P. Kelly, A. K. Malullo, and D. K. H. Tan, "Rate control in communications networks: Shadow prices, proporotional fairness and stablity," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[33] L. Xiao, M. Johansson, and S. P. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, Jul. 2004.

[34] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. New York, NY: John Wiley & Sons Inc., 2006.

[35] J. Liu, Y. T. Hou, Y. Shi, and H. Sherali, "Optimization of multiuser MIMO networks with interference," *IEEE Trans. Wireless Commun.*, submitted for publication.

[36] H. D. Sherali and C. H. Tuncbilek, "A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique," *Journal of Global Optimization*, no. 2, pp. 101–112, 1992.

[37] J. B. Rosen, "The gradient projection method for nonlinear programming, Part I, Linear constraints," *SIAM J. Applied Mathematics*, vol. 8, pp. 181–217, 1960.

[38] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Economics.* New York: Wiley, 1999.

[39] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1996.

[40] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis.* Berlin: Springer-Verlag, 2001.

[41] H. D. Sherali and G. Choi, “Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs,” *Operations Research Letters*, vol. 19, no. 3, pp. 105–113, Sep. 1996.