On Performance Optimization for Multi-Carrier MIMO Ad Hoc Networks

Jia Liu† Y. Thomas Hou† Yi Shi† Hanif D. Sherali*
†Department of Electrical and Computer Engineering
*Department of Industrial and Systems Engineering
Virginia Polytechnic Institute and State University, Blacksburg, VA, U.S.A.
{kevinlau, thou, yshi, hanifs}@vt.edu

ABSTRACT

Broadband multi-carrier MIMO (MC-MIMO) is a promising technology that could provide significant capacity gain for wireless ad hoc networks. For MC-MIMO networks, since the capacity is affected by potential mutual interference on subcarriers, scheduling for subcarriers and algorithms for power control/allocation become key problems to harness their potential. However, due to non-convexity and large size of the underlying problem, there are few results on this important problem. In this paper, we first show that the non-convex problem for MC-MIMO networks satisfies the so-called concave perturbation condition, which gives a zero duality gap for the problem. This important result allows us to tackle the problem in the dual domain. The dual approach has the highly desirable benefit of reducing the complexity of the underlying problem, which allows us to design a near-optimal off-line algorithm. In addition to the off-line algorithm, we also devise an online adaptive algorithm (OAA) without the need of channel distribution information (CDI). We show that OAA is able to achieve the same result as the off-line algorithm.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless communications*; G.1.6 [**Numerical Analysis**]: Optimization—*Non-linear Programming, Stochastic Programming*

General Terms

Algorithms, Performance, Theory

Keywords

Ad Hoc Networks, MIMO, Multi-Carrier Systems

1. INTRODUCTION

A critical factor affecting the success of wireless ad hoc networks for wide-scale deployment is *network capacity*: the end user wishes to have their communication experience over ad hoc networks comparable to that for wireline networks. For a wireless network, there are two complementary approaches to increase capacity. For a given channel bandwidth, one can increase the data rate per channel use (measured in bits/s/Hz). This can be done by employing advanced coding and signal processing techniques in space, time, and frequency domains (e.g., using multiple-input multiple-output systems (MIMO) [6, 26]). In fact, researchers have shown that, by employing multi-antenna arrays on both sides of a wireless channel, the capacity of a MIMO link scales approximately as $\min\{n_t, n_r\}\log_2(SNR)$, where n_t and n_r are the numbers of transmit and receive antennas, respectively. Such a capacity gain is achieved without extra cost of spectrum bandwidth. The capacity increase with the number of antennas is known as the *spatial multiplexing gain* [7, 33] and is largely responsible for MIMO's success.

Another approach is to increase the channel bandwidth, which can be done by either using broadband channels (i.e., increasing the bandwidth of the channel) or by exploiting idle channels using cognitive radio. For a broadband channel, a serious problem affecting communication reliability is the so-called *frequency-selective fading* effect [21], which must be addressed carefully. To understand this frequency-selective fading, consider a broadband channel for which the channel gain response is shown in Figure 1(a). Here, frequency-selective fading refers to the phenomenon that the channel gain response varies with frequency. This problem creates difficulty in using the entire bandwidth to transmit data because the transmitted signal's bandwidth is greater than the frequency range over which the channel response is flat. Under frequency-selective fading, the channel in time domain is associated with time dispersion and induces *inter-symbol interference* [21]. As a result, a transmitted symbol is corrupted by other subsequent transmitted symbols. To address this problem, multi-carrier (MC) techniques such as OFDM have been successfully developed to simplify channel equalization, and the low symbol rate in each subcarrier makes it much easier to handle time dispersion and eliminate ISI. As shown in Figure 1(b), a multi-carrier system partitions the entire broadband spectrum into a large number of narrowband subcarriers (typically, 128 – 4096 [21]). Within each narrowband subcarrier, the fading can be considered frequency-flat and can be handled more easily by error correction coding, simple equalization, or adaptive bit loading.

The complementary approaches of MIMO and broadband

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MobiHoc'09, May 18–21, 2009, New Orleans, Louisiana, USA.
Copyright 2009 ACM 978-1-60558-531-4/09/05 ...$5.00.

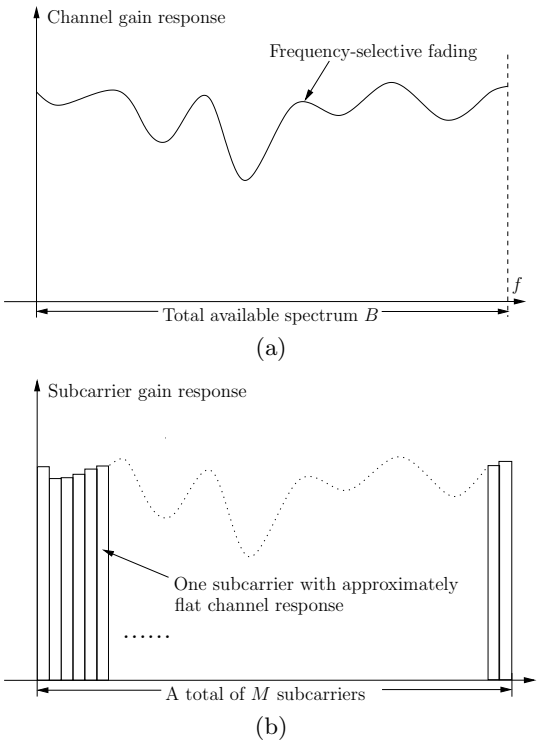


Figure 1: An illustration of a broadband wireless channel.

MC techniques offer great potential for increasing network capacity for wireless ad hoc networks [12]. The combination of these two approaches has since been termed multi-carrier MIMO (MC-MIMO) in the literature, and it has attracted much attention and many research activities, but mostly at the physical layer (see, e.g., [18, 19, 23, 31]). Under MC-MIMO, a broadband channel is partitioned into a large number of flat-fading narrowband subcarriers. Then, on each narrowband subcarrier, MIMO is used to increase the capacity for the subcarrier.

Although MC-MIMO has great potential to increase network capacity, a number of challenging problems remain to be addressed to optimize its performance. To illustrate, consider an MC-MIMO based single-hop ad hoc network example shown in Figure 2. Suppose that there are a total of K mutually-interfered MC-MIMO links transmitting concurrently, with each link representing a single-hop source-destination pair. The broadband in use is partitioned into M subcarriers. Due to potential mutual interference on subcarriers, appropriate algorithms for scheduling subcarrier usage and power allocation are necessary. By subcarrier scheduling, we mean that for each link in the network, we need to determine what subset of subcarriers need to be employed for transmission. The goal of subcarrier scheduling at each link is to avoid transmitting over those subcarriers that are highly interfered by other links or in deep fade modes. On the other hand, for a subcarrier on a link that is scheduled to transmit, we need to determine how much power should be allocated onto that subcarrier. Further, in each subcarrier, power allocation decision at each antenna element of a transmitter also needs to be made. Due to the complicated interference relationship among the links, optimal subcarrier scheduling and power control/allocation is a non-trivial problem.

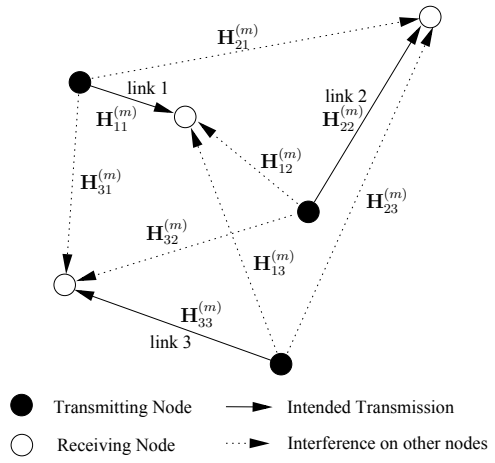


Figure 2: Network model of a MC-MIMO ad hoc network.

The difficulties associated with subcarrier scheduling and power control/allocation can be summarized as follows:

- **Large Number of Subcarriers:** In an MC system, the number of subcarriers needs to be large to ensure near flat fading on each subcarrier. As discussed, the number of subcarriers is at least 128 and could be as large as 4096. This results in a large optimization space for subcarrier scheduling and power control/allocation, which is not encountered in other types of wireless ad hoc networks.

- **Mutual Interference on Subcarrier:** For simultaneous transmissions on the same subcarrier at different links, there will be mutual interference. The level of mutual interferences is determined by the power control/allocation decision on these links. As expected, interference will affect the capacity on each interfering link on the same subcarrier, thus affecting the achieved network capacity.

- **Non-convexity:** As we shall soon show, the complicated interference relationship renders a non-convex structure for a performance optimization problem. Further, due to the large design space caused by the number of subcarriers, finding a global optimal solution is intractable: the state-of-the-art global optimization techniques (such as BB/RLT [22]) have an exponential complexity in MK matrix variables, where M and K are the numbers of subcarriers and links, respectively.

- **Channel Distribution Information:** In MC-MIMO ad hoc networks, the channel response of each subcarrier is not only frequency-variant but also time-variant. As a result, evaluating the ergodic capacity for each subcarrier requires channel distribution information (CDI) in time domain. However, acquiring such CDI knowledge needs an off-line training and measurement process, which is cumbersome and hinders the practical implementation of an MC-MIMO ad hoc network.

The goal of this paper is on performance optimization for MC-MIMO ad hoc networks through subcarrier scheduling and power control/allocation. We focus on capacity of an ad hoc network consisting of K links. The specific objective function we use is to maximize the weighted sum of rates of all links (simply called the MWSR problem in this paper). The MWSR problem can be viewed as a generalization of capacity maximization and is a pillar problem inherent in many important applications (see, e.g., [1, 4, 17, 29]).

1.1 Summary of Main Results

We investigate the MWSR problem by first assuming complete knowledge of CDI, and accordingly design an off-line solution. Then, we design a practical online adaptive solution that does not require complete CDI knowledge. The main results of this paper are summarized as follows:

- For the off-line algorithm, we show that the MWSR problem is non-convex but satisfies the so-called concave perturbation condition. This concavity property of the perturbation function guarantees a zero duality gap, meaning that we can equivalently solve the MWSR problem in its dual domain.
- The key benefit of the dual approach is that the dual problem can be decomposed into M subproblems, avoiding the exponential complexity in M that is inherent in the primal approach. Since each subproblem remains a non-convex problem that has an exponential complexity in K matrix variables, we develop an efficient near-optimal algorithm for solving each subproblem.
- In addition to the off-line algorithm, we also develop an online adaptive algorithm (OAA) that does not require full CDI. We show that OAA can adapt to any *unknown* underlying fading distribution, and converges with probability one to the same optimal solution obtained by the off-line algorithm.

1.2 Paper Organization

The remainder of this paper is organized as follows. Section 2 introduces the network model and problem formulation. In Section 3, we design an off-line algorithm for solving the MWSR problem in the dual domain. In Section 4, we present a practical online adaptive algorithm and establish its convergence property. Section 5 reviews some related work and Section 6 concludes this paper.

2. NETWORK MODEL AND PROBLEM FORMULATION

Since the mathematical model involves characterizing mutual interference among subcarriers, we first introduce the notation style used in the paper. We use boldface to denote matrices and vectors. For a matrix $\mathbf{A}$, we let $\mathbf{A}^\dagger$ denote the conjugate transpose; $\mathrm{Tr}\{\mathbf{A}\}$ denotes the trace of $\mathbf{A}$; and $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$. We use $\mathbf{A} \succeq 0$ to represent that $\mathbf{A}$ is Hermitian and positive semidefinite (PSD). We denote $\mathbf{I}$ as the identity matrix with its dimension determined from the underlying context. $\mathbf{1}$ and $\mathbf{0}$ denote vectors whose elements are all ones and zeros, respectively, again with their dimensions determined from the context. Finally, we use the operator "$\langle\cdot,\cdot\rangle$" to represent the inner product operation for vectors.

In this paper, we consider an MC-MIMO ad hoc network as illustrated in Figure 2. There are a total of K concurrent transmission pairs (or links) in the network and M subcarriers in the given communication band, with a total bandwidth B. As a result, the bandwidth of each subcarrier is $\frac{B}{M}$. As discussed earlier, for each subcarrier to be considered flat fading, M needs to be sufficiently large (e.g., from 128 to 4096 in practice).

Suppose that each node in the network is equipped with N antennas. We use $\mathbf{H}_{ij}^{(m)} \in \mathbb{C}^{N\times N}$ to represent the channel gain matrix from the transmitting node of link i to the receiving node of link j over subcarrier m. We let $\mathbf{H} = \{\mathbf{H}_{ij}^{(m)} : i,j = 1,\ldots,K,\ m = 1,\ldots,M\}$ denote the collection of all channel gain matrices (i.e., a total of K^2M matrices). The entries in each channel gain matrix are assumed to be i.i.d. complex Gaussian.

The normalized received base-band signal vector at the receiving node of link k over subcarrier m can be computed as

$$\mathbf{r}_k^{(m)} = \sqrt{\rho_{kk}}\mathbf{H}_{kk}^{(m)}\mathbf{t}_k^{(m)} + \sum_{i=1,i\neq k}^{K}\sqrt{\rho_{ik}}\mathbf{H}_{ik}^{(m)}\mathbf{t}_i^{(m)} + \mathbf{n},$$

where $\mathbf{t}_k^{(m)} \in \mathbb{C}^N$ and $\mathbf{r}_k^{(m)} \in \mathbb{C}^N$ represent the transmitted signal vector (with unit power) and the received signal vector of link k over subcarrier m, respectively; and $\mathbf{n}$ represents the complex additive white Gaussian noise vector with zero mean and unit variance. In the above expression, ρ_{ik} denotes the signal-to-noise ratio (SNR) of link k if $i = k$, or the interference-to-noise ratio (INR) from link i to link k if $i \neq k$.

Denote $\mathbf{Q}_k^{(m)}$ as the covariance matrix for the input symbol vector $\mathbf{t}_k^{(m)}$, i.e.,

$$\mathbf{Q}_k^{(m)} = \mathbb{E}\{\mathbf{t}_k^{(m)} \cdot \mathbf{t}_k^{(m)\dagger}\}.$$

Physically, $\mathbf{Q}_k^{(m)}$ represents the power allocation among the antennas of the transmitting node of link k over subcarrier m, and the trace of $\mathbf{Q}_k^{(m)}$ (i.e., the sum of all its diagonal entries) is the transmit power on link k over subcarrier m. By the definition of $\mathbf{Q}_k^{(m)}$, we have $\mathbf{Q}_k^{(m)} \succeq 0$ and $\mathrm{Tr}(\mathbf{Q}_k^{(m)}) \leq 1$, $\forall k, m$. Since the sum of power over all subcarriers at a transmitting node cannot exceed the node's maximum transmit power, we have

$$\mathbb{E}_{\mathbf{H}}\left[\sum_{m=1}^{M}\mathrm{Tr}(\mathbf{Q}_k^{(m)})\right] \leq 1, \quad \text{for } k = 1,2,\ldots,K.$$

where 1 represents normalized maximum power.

Denote $\mathbf{Q} = \{\mathbf{Q}_k^{(m)} : k = 1,\ldots,K,\ m = 1,\ldots,M\}$ as the collection of all power allocation matrices (a total of KM such matrices for the network). The ergodic capacity of link k over subcarrier m (in b/s/Hz) can be computed by [10]

$$C_k^{(m)} = \mathbb{E}_{\mathbf{H}}\left[\log_2\left|\mathbf{I} + \rho_{kk}\mathbf{H}_{kk}^{(m)}\mathbf{Q}_k^{(m)}\mathbf{H}_{kk}^{(m)\dagger}(\mathbf{R}_k^{(m)})^{-1}\right|\right],$$

where the expectation is taken over the distribution of $\mathbf{H}$, and $\mathbf{R}_k^{(m)}$ represents the aggregate interference and noise at the receiving node of link k on subcarrier m, and can be computed as

$$\mathbf{R}_k^{(m)} = \mathbf{I} + \sum_{i=1,i\neq k}^{K}\rho_{ik}\mathbf{H}_{ik}^{(m)}\mathbf{Q}_i^{(m)}\mathbf{H}_{ik}^{(m)\dagger}. \tag{1}$$

With the above expression, the data rate R_k for link k can

be computed as

$$R_k = \sum_{m=1}^{M} \frac{B}{M} C_k^{(m)}.$$

That is, R_k is the aggregated data rate of link k over all M subcarriers.

Putting together these constraints, the MWSR problem can be formulated as follows:

$$\begin{array}{rll} \textbf{MWSR}\text{: Maximize} & \sum_{k=1}^{K} w_k R_k & \\ \text{subject to} & \mathbb{E}_{\mathbf{H}}[\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{Q}_k^{(m)})] - 1 \leq 0, \forall k & (2) \\ & \mathbf{Q}_k^{(m)} \succeq 0, \forall k, \ \forall m, & \end{array}$$

where w_i, $i = 1, \ldots, K$, are pre-assigned weights and the $\mathbf{Q}_k^{(m)}$-matrices are optimization variables.

Note that the $\mathbf{Q}_k^{(m)}$-matrices in this formulation capture both subcarrier scheduling and power control/allocation decisions. If $\mathbf{Q}_k^{(m)}$ is a zero matrix, it means that subcarrier m will not be scheduled by link k for transmission. If $\mathbf{Q}_k^{(m)}$ is non-zero, it means that link k will schedule subcarrier m for transmission. Furthermore, power control/allocation has two stages. First, the transmitting node of each link will determine how much power will be allocated on each subcarrier that is scheduled for transmission. The allocated power for that particular subcarrier is represented by the trace of $\mathbf{Q}_k^{(m)}$. Second, for each subcarrier, the entries in $\mathbf{Q}_k^{(m)}$ determine the power allocation among each of the antenna elements.

Since the objective function of (2) is not concave in $\mathbf{Q}_k^{(m)}$, MWSR is a non-convex optimization problem. As a result, solving MWSR is a difficult task. As indicated earlier, to ensure flat fading on each subcarrier, the value of M is large, whereas state-of-the-art global optimization techniques (e.g., BB/RLT [22]) have an exponential complexity in the total number of matrix variables KM. Thus, a primal approach to solving (2) is intractable.

3. AN OFF-LINE SOLUTION TO MWSR

Although the MWSR problem is non-convex, we will show in Sections 3.1 and 3.2 that this problem in fact has a zero duality gap. This finding is based on examining the concavity of the perturbation function of the MWSR problem. This zero-duality gap result enables us to tackle the MWSR problem in the dual domain. In Section 3.3, we show that studying the MWSR problem in its dual domain has the highly desirable benefit of yielding a much lower complexity that is linear in M, due to the fact that the dual problem can be suitably decomposed. This complexity reduction is significant because, as we mentioned earlier, M is large in MC-MIMO networks. In Section 3.4, we propose an off-line algorithm to solve the MWSR problem in its dual domain, and Section 3.5 presents some numerical results for the off-line algorithm.

3.1 Concavity of Perturbation Function and Zero Duality Gap

It is well-known that non-convexity usually results in a non-zero duality gap in optimization. However, under some special conditions, some non-convex optimization problems could turn out to have a zero duality gap. If an optimization problem has a zero duality gap, it can then be solved by equivalently analyzing its dual problem. This approach is particularly relevant when the primal problem is hard to solve. In this section, we establish such a zero duality gap condition. In the next section, we will show that the challenging MWSR problem satisfies this condition, meaning that the MWSR problem has a zero duality gap.

Let us consider the general form of (2), which can be rewritten (after interchanging summations and expectations) as

$$\begin{array}{rl} \text{Maximize} & \sum_{m=1}^{M} f_m(\mathbf{x}_m) \\ \text{subject to} & \sum_{m=1}^{M} \mathbf{g}_m(\mathbf{x}_m) \leq \mathbf{0}, \end{array} \quad (3)$$

where $\mathbf{x}_m \in \mathbb{S} \subseteq \mathbb{C}^K$, $\forall m$, are vectors of the optimization variables; $f_m(\cdot) : \mathbb{C}^K \to \mathbb{R}$, $\forall m$, are functions that may or may not be concave, and $\mathbf{g}_m(\cdot) : \mathbb{C}^K \to \mathbb{R}^L$, $\forall m$, are vector-valued functions that may or may not be convex.

Now, consider problem (3) in the dual domain. By associating a dual vector variable $\mathbf{u} \in \mathbb{R}^L$ with the constraints, we can write the Lagrangian of (3) as

$$L(\mathbf{x}_m, \mathbf{u}) = \sum_{m=1}^{M} f_m(\mathbf{x}_m) - \mathbf{u}^T \sum_{m=1}^{M} \mathbf{g}_m(\mathbf{x}_m). \quad (4)$$

The Lagrangian dual function is the unconstrained maximization with respect to $\mathbf{x}_m$: $\Theta(\mathbf{u}) \triangleq \max_{\mathbf{x}_m} L(\mathbf{x}_m, \mathbf{u})$. The Lagrangian dual optimization problem is then given by:

$$\begin{array}{rl} \text{Minimize} & \Theta(\mathbf{u}) \\ \text{subject to} & \mathbf{u} \geq \mathbf{0}. \end{array} \quad (5)$$

In general, the weak duality theorem [2] says that the optimal objective value of the dual problem provides an upper bound on that for the primal problem, where the difference between this upper bound and the primal optimal value is called the duality gap.

It is worth pointing out that convexity of problem (3) (under suitable constraint qualification [2]) is only a *sufficient* condition for obtaining a zero duality gap, but not necessary [8]. In what follows, we will show that a non-convex problem having the structure of (3) yields a zero duality gap if its *perturbation function* is concave. First, let us introduce the concept of the perturbation function.

Definition 1. The perturbation function of (3), denoted as $\nu(\mathbf{y}) : \mathbb{R}^L \to \mathbb{R}$, is defined as the optimal value function of the following problem:

$$\nu(\mathbf{y}) = \max \left\{ \sum_{m=1}^{M} f_m(\mathbf{x}_m) : \sum_{m=1}^{M} \mathbf{g}_m(\mathbf{x}_m) \leq \mathbf{y} \right\}, \ \mathbf{y} \in \mathbb{R}^L.$$

From the definition, it is clear that $\nu(\mathbf{0})$ yields the same optimal objective value as that of the primal problem (3). The idea of perturbation function was first proposed in [9], where Geoffrion used this concept to study the stability of the nonlinear programs and its application in duality theory (later also called "value function" in [5] or "primal function" in [3]). In this paper, we use the perturbation function to analyze the duality gap of MWSR since this approach affords deep geometrical insights and does not require differential calculus, general minimax theorems, or conjugate function theory employed by most other studies in duality theory. Now, we state the concave perturbation condition in the following theorem.

Theorem 1 (Concave Perturbation Condition). *If the perturbation function $\nu(\mathbf{y})$ is concave, then the non-convex optimization problem in (3) and its dual problem in*

(5) has the same optimal objective value, i.e., the duality gap is zero.

Theorem 1 can be proved from a geometric perspective. If the condition stated in Theorem 1 holds, i.e., the perturbation function is concave, then there exists a supporting hyperplane to the hypograph of $\nu(\mathbf{y})$ at $\mathbf{y} = \mathbf{0}$. This supporting hyperplane condition is in turn a necessary and sufficient condition for the existence of a saddle point [2, Theorem 6.2.7], i.e., no duality gap.[1]

3.2 Perturbation Function for MWSR Problem

In the previous section, we have discussed that the concavity of the perturbation function leads to a zero duality gap for non-convex optimization problems. In this section, we show that the perturbation function of the MWSR problem in (2) is indeed concave. This result allows us to tackle the MWSR problem in its dual domain. Let $\nu(\mathbf{y})$ be the perturbation function of (2), i.e.,

$$\nu(\mathbf{y}) \triangleq \max \left\{ \sum_{k=1}^{K} w_k \sum_{m=1}^{M} \frac{B}{M} C_k^{(m)} \Bigg| \sum_{m=1}^{M} \mathbb{E}_{\mathbf{H}} \left[\mathrm{Tr}(\mathbf{Q}_k(f)) \right] - 1 \leq y_k, \forall k \right\}, \quad (6)$$

where $\mathbf{y} \in \mathbb{R}^K$ and $\mathbf{Q}_k(f) \succeq 0 \ \forall k, f$. We use a constructive approach to show that the perturbation function of (2) is concave. We first consider the problem (6) over one subcarrier, where the channel gain responses for all links are frequency-flat (as shown in Figure 3) due to the large value of M. Then, we will extend the concavity property of (6) to the case where channel response functions are frequency-varying. For each subcarrier, we have the following result. We note that the proof of time-sharing property in [30] is somewhat related to the idea in here. But our approach in showing the concave perturbation condition is much cleaner and straightforward.

LEMMA 1. *For each subcarrier where the channel gain responses $\mathbf{H}_{ik}$ are frequency-flat for all $i, k \in \{1, \ldots, K\}$, problem (2) satisfies the concave perturbation condition.*

PROOF. Suppose that $\mathbf{Q}^*_{\mathbf{y}_1}(f)$ and $\mathbf{Q}^*_{\mathbf{y}_2}(f)$ solve the problems to evaluate $\nu(\cdot)$ over subcarrier m when $\mathbf{y} = \mathbf{y}_1$ and $\mathbf{y} = \mathbf{y}_2$, respectively, where $f \in$ subcarrier m. Let $0 \leq \mu \leq 1$. To prove that the perturbation function $\nu(\mathbf{y})$ is concave over subcarrier m, we need to show that $\nu(\mu\mathbf{y}_1 + (1-\mu)\mathbf{y}_2) \geq \mu\nu(\mathbf{y}_1) + (1-\mu)\nu(\mathbf{y}_2)$ for all $0 \leq \mu \leq 1$. The basic idea of the proof is as follows. We construct a solution $\mathbf{Q}(f)$ that is feasible to the problem for evaluating $\nu(\mu\mathbf{y}_1 + (1-\mu)\mathbf{y}_2)$. Denote the objective value of (6) under $\mathbf{Q}(f)$ as $\hat{C}$. Clearly, we have $\nu(\mu\mathbf{y}_1 + (1-\mu)\mathbf{y}_2) \geq \hat{C}$. Note that if we can construct $\mathbf{Q}(f)$ such that $\hat{C} \geq \mu\nu(\mathbf{y}_1) + (1-\mu)\nu(\mathbf{y}_2)$ for all $0 \leq \mu \leq 1$, then we will have established the concavity of $\nu(\cdot)$.

[1] Incidentally, Hande *et al.* [11] and Yu *et al.* [30] also provided sufficient conditions for a zero duality gap. We note that the terms "optimized total utility function" [11] and "time-sharing property" [30] are, in essence, the same as our concave perturbation condition. However, the proof in [30] is mostly by visualization. In [11], the claim of the zero duality gap result following from "min common max crossing duality" in [3] is not accurate because the duality result in [3] is proved using stronger convexity assumptions. In fact, the result in [11] can be proved directly from [2, Theorem 6.2.7], as we have presented here.

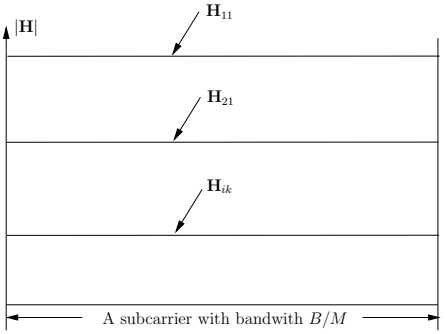


Figure 3: Constant channel responses in each subcarrier.

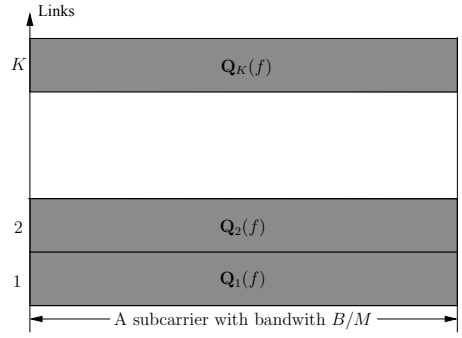


Figure 4: The frequency-flat solution structure in a subcarrier.

Without loss of generality, consider $\mathbf{Q}^*_{\mathbf{y}_1}(f)$, which solves the perturbation function $\nu(\mathbf{y}_1)$. Note that since all channel response functions $\mathbf{H}_{ik}$ are constant, $\mathbf{Q}^*_{\mathbf{y}_1}(f)$ must also be flat for all $f \in$ subcarrier m, as shown in Figure 4. Likewise, we can conclude that $\mathbf{Q}^*_{\mathbf{y}_2}(f)$ is flat over subcarrier m.

Now, suppose that $C^*_{\mathbf{y}_1}$ and $C^*_{\mathbf{y}_2}$ are the optimal objective values of $\nu(\mathbf{y}_1)$ and $\nu(\mathbf{y}_2)$, respectively. A construction of $\mathbf{Q}(f)$ is illustrated in Figure 5. We will show that this constructed $\mathbf{Q}(f)$ is feasible to the problem for $\nu(\mu\mathbf{y}_1 + (1-\mu)\mathbf{y}_2)$ and achieves at least $\mu C^*_{\mathbf{y}_1} + (1-\mu)C^*_{\mathbf{y}_2}$ for all $0 \leq \mu \leq 1$. Now, we divide the whole subcarrier into two subbands as shown in Figure 5. We assign a μ portion of the subcarrier as one subband, where we adopt $\mathbf{Q}^*_1(f)$ as the scheduling and power allocation scheme. We assign the remaining $(1-\mu)$ portion of the subcarrier as the second subband, where we adopt $\mathbf{Q}^*_2(f)$ as the scheduling and power allocation scheme. Evidently, the resultant $\mathbf{Q}(f)$ is feasible for $\nu(\mu\mathbf{y}_1 + (1-\mu)\mathbf{y}_2)$ because from the linearity of the power constraints in (6), we have $\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{Q}(f)) - 1 = \sum_{m=1}^{M} (\mu \mathrm{Tr}(\mathbf{Q}^*_1(f)) + (1-\mu)\mathrm{Tr}(\mathbf{Q}^*_2(f))) - 1 \leq \mu\mathbf{y}_1 + (1-\mu)\mathbf{y}_2$. Also, it is apparent that $\mathbf{Q}(f)$ achieves an objective value of $\mu C^*_1 + (1-\mu)C^*_2$ over subcarrier m. Hence, the concave perturbation condition holds for subcarrier m. □

Now, for problem (2), we have the following result:

THEOREM 2. *When the number of carriers M is sufficiently large, the MWSR problem satisfies the concave perturbation condition.*

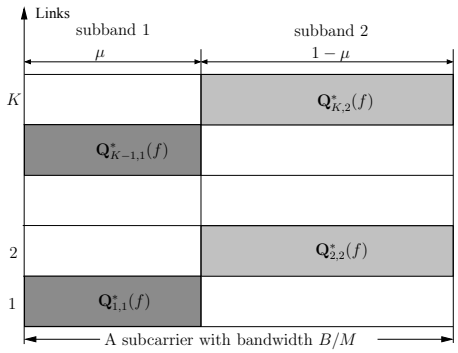


Figure 5: Solution construction for $\mathbf{Q}(f)$.

PROOF. As discussed earlier, for MC-MIMO, the number of subcarriers is sufficiently large (e.g., $128 \leq M \leq 4096$) so that the fading on each subcarrier can be considered flat. That is, for each $i, j \in \{1, \ldots, K\}$ and for each $m \in \{1, \ldots M\}$, $\mathbf{H}_{ij}^{(m)}$ is a constant. Based on Lemma 1, we conclude that the concave perturbation condition is satisfied for each subcarrier. Since we can construct a solution to show the concavity of the perturbation for each subcarrier, we can then construct a solution to show the concavity of the perturbation function for the entire band by combining the solutions for all subcarriers. Thus, the concave perturbation condition holds for the entire band. □

3.3 Problem Decomposition and Complexity Reduction

In previous sections, we have shown that the MWSR problem has a zero duality gap and thus can be solved in its dual domain. Note that when we solve the MWSR problem in the dual domain, the dual problem is of higher dimensionality since we introduce K dual variables. However, the dual domain approach has the desirable benefit of reducing the complexity of the primal problem since the dual problem can be decomposed into M subproblems, thus removing the exponential complexity in M (a large value) inherent in the primal problem. This complexity reduction far outweighs the dimensionality increase of the dual problem.

To see this, let us first associate a dual variable u_i with each constraint in (2) and denote $\mathbf{u} \triangleq [u_1\, u_2\, \ldots\, u_K]^T$ the collection of such dual variables. Then, the Lagrangian $L(\mathbf{Q}, \mathbf{u})$ can be computed as

$$L(\mathbf{Q}, \mathbf{u}) = \sum_{k=1}^{K} w_k \mathbb{E}_{\mathbf{H}}\Bigg[\sum_{m=1}^{M} \log_2 \Big| \mathbf{I} + \rho_{kk} \mathbf{H}_{kk}^{(m)} \mathbf{Q}_k^{(m)} \mathbf{H}_{kk}^{(m)\dagger} \cdot (\mathbf{R}_k^{(m)})^{-1} \Big| \Bigg] + \sum_{k=1}^{K} u_k \Bigg[1 - \mathbb{E}_{\mathbf{H}}\Big[\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{Q}_k^{(m)}) \Big] \Bigg].$$

The Lagrangian dual function is defined as

$$\Theta(\mathbf{u}) \triangleq \max_{\mathbf{Q}} \Big\{ L(\mathbf{Q}, \mathbf{u}) : \mathbf{Q}_k^{(m)} \succeq 0,\ \forall k,\ \forall m \Big\}. \quad (7)$$

Then, the dual problem can be written as follows:

$$\textbf{D-MWSR}\text{: Minimize} \quad \Theta(\mathbf{u}) \quad \text{subject to} \quad \mathbf{u} \geq \mathbf{0}. \quad (8)$$

After interchanging expectation and summation, we can rewrite (7) as

$$\Theta(\mathbf{u}) = \sum_{m=1}^{M} \mathbb{E}_{\mathbf{H}}\Big[\max_{\substack{\mathbf{Q}_k^{(m)} \succeq 0 \\ \forall k}} \Big\{ \sum_{k=1}^{K} F(\mathbf{Q}_k^{(m)}) \Big\} \Big] + \langle \mathbf{u}, \mathbf{1} \rangle, \quad (9)$$

where

$$F(\mathbf{Q}_k^{(m)}) \triangleq w_k \log |\mathbf{I} + \rho_{kk} \mathbf{H}_{kk}^{(m)} \mathbf{Q}_k^{(m)} \mathbf{H}_{kk}^{(m)\dagger} (\mathbf{R}_k^{(m)})^{-1}| - u_k \mathrm{Tr}(\mathbf{Q}_k^{(m)}).$$

Now, we can see that the Lagrangian dual function $\Theta(\mathbf{u})$ is decomposed into M subproblems, thus avoiding the exponential complexity in M that is inherent in the primal problem (2).

Algorithm 1 Solution to D-MWSR

1. Choose an initial starting solution $\mathbf{u}(0)$. Let $n = 0$.
2. Compute $\mathbf{Q}^*(n)$ by solving M subproblems, each of which corresponds a subcarrier m.
3. Compute the subgradient $d_{\mathbf{u}_k}(n)$ using (10) with $\mathbf{Q}^*(n)$, and choose an appropriate step size s_n.
4. Update the dual variables $\mathbf{u}(n)$ with using (11).
5. If $\|\mathbf{u}_k(n+1) - \mathbf{u}_k(n)\| < \epsilon$, then return $\mathbf{Q}^*(n)$ as the final optimal solution and stop. Otherwise, let $n \leftarrow n + 1$ and goto Step 2.

3.4 Solving the Dual Problem

In this section, we develop an algorithm to solve the dual problem in (8). Since $\Theta(\mathbf{u})$ is a point-wise minimum of a set of linear functions (in $\mathbf{u}$), it is convex regardless of the convexity property of the primal problem [2]. Thus, D-MWSR is a convex minimization problem and we can solve it by a subgradient algorithm [2]. The subgradient components at the nth iteration can be computed as

$$d_{u_k}(n) = 1 - \mathbb{E}_{\mathbf{H}}\Big[\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{Q}_k^{(m)*}(n)) \Big], \quad \forall k. \quad (10)$$

where $\mathbf{Q}_k^{(m)*}(n)$, $\forall k, m$, represents the optimal solution in (9) with $\mathbf{u}$ being replaced by $\mathbf{u}(n)$. The updates of dual variables can be computed as

$$u_k(n+1) = [u_k(n) - s_n d_{u_k}(n)]_+^U, \quad \forall k, \quad (11)$$

where $[\cdot]_+^U \triangleq \min\{\max\{\cdot, 0\}, U\}$ with U being some upper bound on u_k (for numerical stability), and where s_n is the step size during the nth iteration. Based on the subgradient algorithm's property [2], we have that if the step size selection satisfies

$$s_n \to 0, \quad \sum_{n=1}^{\infty} s_n \to \infty, \text{ and } \sum_{n=1}^{\infty} s_n^2 < \infty, \quad (12)$$

then the iterates generated by the subgradient method converge to the optimal solution $\mathbf{u}^*$ [2]. The subgradient algorithm is summarized in Algorithm 1.

A nice property for Algorithm 1 is that it can be implemented in a distributed fashion. This is because the subgradient computation in (10) and the dual variable updates in (11) only involve local power variables at each node and do not require global information.

Next, we describe how to solve the M subproblems in Step 2 of Algorithm 1, where each subproblem corresponds to a subcarrier. From (9), we find that each subproblem is a maximization problem in K N-dimensional matrix variables

Algorithm 2 Solution to each subcarrier subproblem

1. Initialize $\mathbf{Q}_1^{(m)}(0)$, $\mathbf{Q}_2^{(m)}(0)$,...,$\mathbf{Q}_K^{(m)}(0)$. Let $n = 0$, $k = 1$.
2. Compute the gradients $\mathbf{G}_k^{(m)}(n)$.
3. Compute $\hat{\mathbf{Q}}_k^{(m)}(n+1) = \mathbf{Q}_k^{(m)}(n) + z_n \mathbf{G}_k^{(m)}(n)$ with step size z_n.
4. Let $\bar{\mathbf{Q}}_k^{(m)}(n+1)$ be the projection of $\hat{\mathbf{Q}}_k^{(m)}(n+1)$ onto the positive semidefinite cone.
5. Choose an appropriate step size α_n. Let $\mathbf{Q}_k^{(m)}(n+1) = \bar{\mathbf{Q}}_k^{(m)}(n+1) + \alpha_n(\bar{\mathbf{Q}}_k^{(m)}(n+1) - \mathbf{Q}_k^{(m)}(n))$.
6. If $\|\mathbf{Q}_k^{(m)}(n+1) - \mathbf{Q}_k^{(m)}(n)\| \leq \epsilon$, for $k = 1, 2, \ldots, K$, then stop; else let $n \leftarrow n+1$, $k \leftarrow (k+1) \bmod K$ and go to Step 2.

in the following form:

$$\begin{aligned} \text{Maximize} \quad & \sum_{k=1}^{K} \Big[w_k \log_2 \big| \mathbf{I} + \rho_{kk} \mathbf{H}_{kk}^{(m)} \mathbf{Q}_k^{(m)} \mathbf{H}_{kk}^{(m)\dagger} \cdot \\ & (\mathbf{R}_k^{(m)})^{-1} \big| - u_k \mathrm{Tr}(\mathbf{Q}_k^{(m)}) \Big] \\ \text{subject to} \quad & \mathbf{Q}_k^{(m)} \succeq 0, \forall k. \end{aligned} \tag{13}$$

In (13), $\mathbf{R}_k^{(m)}$ is defined in (1), which represents the total power of interferences plus noise that link k experiences on subcarrier m. It is this term that renders (13) a non-convex program. Such a non-convex optimization problem has exponential complexity in the K matrix variables. Therefore, solving (13) remains non-trivial. In this paper, we devise a distributed gradient projection (DGP) algorithm to solve (13), where the input covariance matrices for the links are updated in an *iterative* manner: In each iteration, we use the gradient projection method [2] to update the power matrix variable for a particular link to maximize (13) while all the other links' matrix variables are held fixed. Thus, except at the initiation, only link (n mod K) and link ($n+1$ mod K) (where a zero modulus value is replaced by K) need to exchange information (the information of K $\mathbf{Q}$ variables) in the nth iteration. We summarize our solution method for solving the subcarrier subproblem in Algorithm 2.

A key component in Algorithm 2 is the gradient computation in Step 2. By using the identity $\frac{\partial}{\partial \mathbf{X}} \ln\det(\mathbf{A}+\mathbf{BXC}) = \big[\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}\big]^T$ from matrix differential calculus [16], we are able to obtain the gradients as (for simplicity, we drop the iteration index n in all $\mathbf{Q}_k$ and u_k variables):

$$\begin{aligned} \mathbf{G}_k^{(m)} = & \frac{2w_k\rho_{kk}}{\ln 2} \mathbf{H}_{kk}^{(m)\dagger} \left(\mathbf{R}_k + \rho_k \mathbf{H}_{kk}^{(m)} \mathbf{Q}_k \mathbf{H}_{kk}^{(m)\dagger} \right)^{-1} \mathbf{H}_{kk}^{(m)} \\ & + \sum_{j=1, j\neq k}^{K} \frac{2w_j\rho_{jk}}{\ln 2} \mathbf{H}_{jk}^{(m)\dagger} \Big[(\mathbf{R}_j^{(m)} + \rho_{jj} \mathbf{H}_{jj}^{(m)} \mathbf{Q}_j^{(m)} \mathbf{H}_{jj}^{(m)\dagger})^{-1} \\ & - (\mathbf{R}_j^{(m)})^{-1} \Big] \mathbf{H}_{jk}^{(m)} - u_k \mathbf{I}. \end{aligned} \tag{14}$$

3.5 Numerical Results

We provide some numerical examples to illustrate the performance of the Algorithms 1 and 2. First, we examine the gap between the solution obtained by Algorithm 2 and the optimal solution obtained via exhaustive search. We first consider the ergodic rate region in the dual domain, i.e., the K-dimensional subspace for the maximum achievable link rates of problem (13). To visualize the capacity regions on 2-D graphs, we plot the ergodic rate region for a 2-link 2-antenna example with $\mathbf{u} = [0.8\ 0.8]^T$. By varying the non-negative weights w_1 and w_2 subject to $w_1 + w_2 = 1$, the entire rate region can be achieved. In this case, we are interested in the performance of Algorithm 2 under the strong interference (stressful situation) since each subproblem becomes less convex as INRs increase. The SNRs in this example are $\rho_{11} = \rho_{22} = 20$ dB. The INRs in this example are $\rho_{12} = 10$ dB and $\rho_{21} = 11$ dB. The ergodic rate region of this example is plotted in Figure 6(a). We find that even under strong interference, the gap between Algorithm 2 and exhaustive search remains close.

Next, we consider the optimal dual objective values of a number of 5-link 2-antenna examples under 5 different sets of dual variables. The first set of dual variables have identical values (randomly chosen to be $\mathbf{u} = 0.8 \cdot \mathbf{1}$ in this example). The second and the third sets of dual variables are randomly generated with mean 1 and 1.3, respectively. The fourth and the fifth sets use randomly generated weights and dual variables. For each set, we examine 16 different combinations of mean SNR and INR values. The optimal dual objective values computed by DGP and exhaustive search are plotted in Figures 6(b) to 6(f). Again, from all 80 cases, we observe that the performance of DGP and exhaustive search is very close. For example, under $\mathbf{u} = 0.8 \cdot \mathbf{1}$, when SNR for each link is 10 dB and INR for each link is 15 dB, i.e., the interference is much stronger than the intended signal, the dual objective values computed by DGP and exhaustive search are 11.17 and 12.13, respectively. In this extreme case, the gap is only 7.9%.

We use two 5-link network examples to see the convergence behavior of the subgradient algorithm. As shown in Figures 7(a) and 7(c), the five links are uniformly distributed in a 2000m×2000m square region. Each node in the network is equipped with 4 antennas. The path-loss index is 4. The solid lines represent intended transmissions and the dotted lines represent interferences. The total number of subcarriers is 128.

Figures 7(b) and 7(d) show the convergence process of the subgradient algorithm for minimizing the dual problem. It can be seen that after approximately 500 iterations, the subgradient algorithm converges and yields a dual objective value of 5.68 b/s/Hz and 7.57 b/s/Hz, respectively. For Example 1, we recover the corresponding primal feasible solution and find that the primal objective value is 5.65 b/s/Hz, i.e., the gap is less than 1%. Compared with the starting point before optimization (equal power allocation to all subcarriers and antennas at each node), which yields an objective value of 3.83 b/s/Hz, we can see the performance gain after optimization is significant (47.5%).

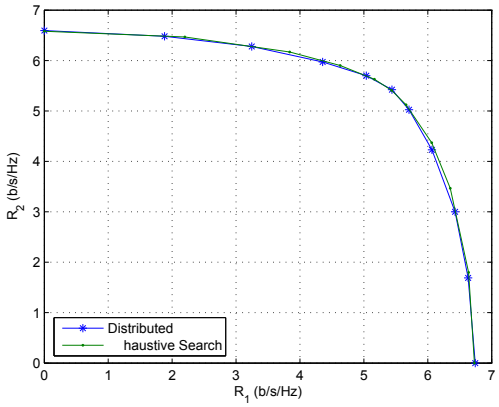
4. AN ONLINE ADAPTIVE ALGORITHM

Algorithm 1 is an off-line algorithm for the MWSR problem since it requires full CDI to compute the subgradients in (10). Such CDI usually requires a training process to acquire before communication takes place, thus making Algorithm 1 cumbersome to implement in practice.

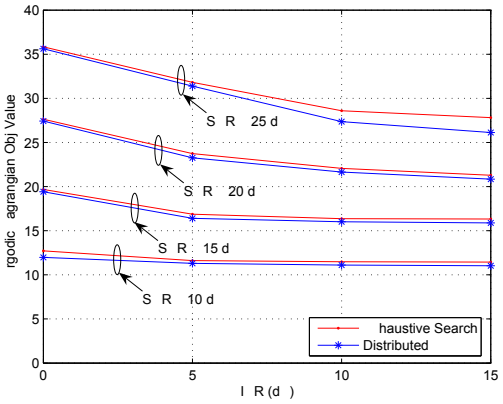
In this section, we present an *online adaptive algorithm* (OAA) for the MWSR problem. The main goal of OAA is to avoid the use of a training process and to adapt to unknown underlying fading distributions on the fly, thus rendering itself to an online algorithm desirable for practical implementation. In Section 4.1, we give a sketch of the main idea of the proposed online algorithm. Section 4.2 studies the convergence behavior and optimality characteristics of this algorithm.

4.1 Main Idea

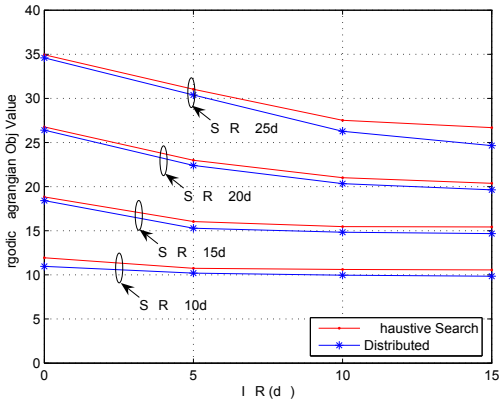
As discussed, the primary reason that Algorithm 1 is an



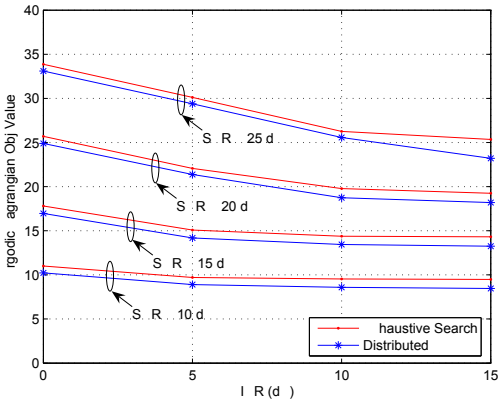
(a) Rate region of a 2-link example.



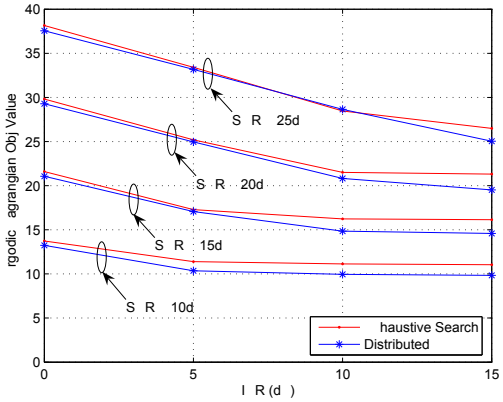
(b) Objective value comparison (equal weights, $\mathbf{u} = 0.8 \cdot \mathbf{1}$).



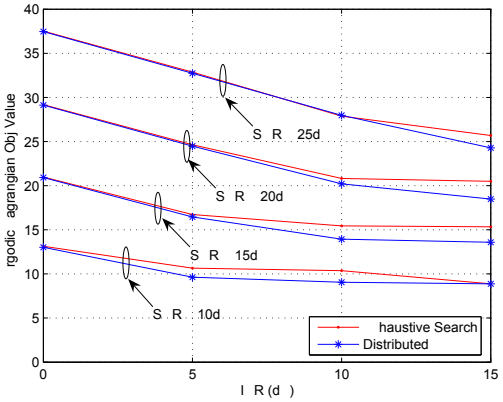
(c) Objective value comparison (equal weights, randomly generated dual variables, mean = 1).



(d) Objective value comparison (random weights, randomly generated dual variables, mean = 1.3).



(e) Objective value comparison (random weights, mean = 1, randomly generated dual variables, mean = 1).

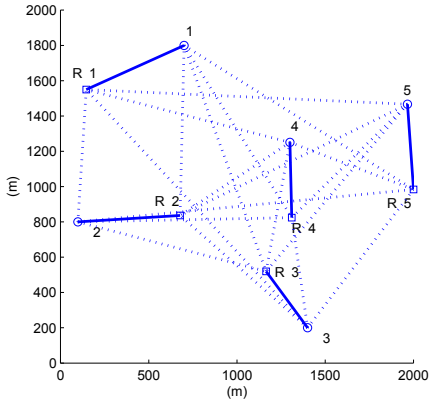


(f) Objective value comparison (random weights, mean = 1, randomly generated dual variables, mean = 1.2).

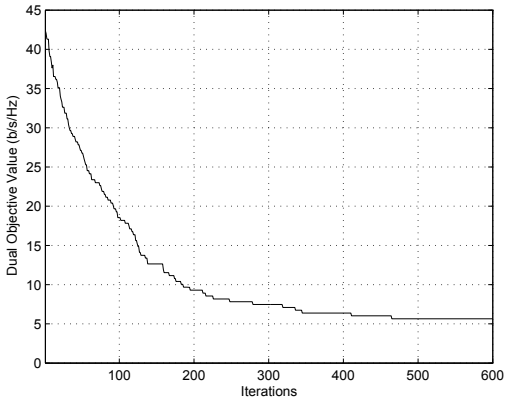
Figure 6: Performance comparison between Algorithm 2 and exhaustive search.

off-line algorithm is that the subgradients in (10) require computing the expectation $\mathbb{E}_{\mathbf{H}}(\cdot)$. For an online algorithm, we ask the following question: Can we remove the expectation computation while retaining the structure of Algorithm 1? If this can be done, then we will have a good chance of transforming Algorithm 1 to an online algorithm.

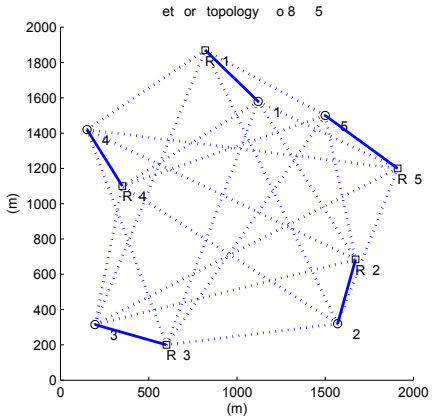
Upon a closer look at Algorithm 1, we realize that, to eliminate the expectation computation, the key step is to have



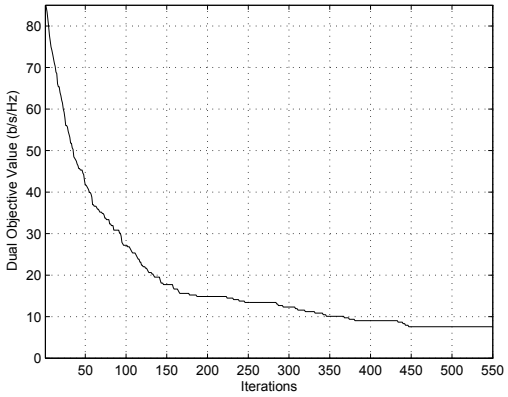
(a) Network topology of Example 1.



(b) Convergence process of Example 1.



(c) Network topology of Example 2.



(d) Convergence process of Example 2.

Figure 7: Network examples and convergence processes of Algorithm 1.

an adequate approximation for the subgradient expression in (10). More importantly, this approximation should not rely on full CDI. This motivates us to resort to stochastic approximation techniques [14]. An easy and practical choice for such an approximation is the following: Instead of using full CDI to accurately compute (10), we only use one observation of the current channel state to compute the subgradient at each iteration. More precisely, the subgradient during the nth iteration is computed using the following approximation:

$$\hat{d}_{u_k}(n) = 1 - \sum_{m=1}^{M} \mathrm{Tr}(\mathbf{Q}_k^{(m)*}), \quad \forall k. \tag{15}$$

Accordingly, in Algorithm 1, the dual variable update is modified to:

$$\hat{u}_k(n+1) = [\hat{u}_k(n) - s_n \hat{d}_{u_k}(n)]_+^U, \quad \forall k. \tag{16}$$

The reason to use (15) as an approximation is that (15) is an unbiased estimation of (10), and can be viewed as a variant of the recursive stochastic approximation algorithm [14]. Comparing the use of (15) in lieu of (10) in Algorithm 1, we find that the expectation computation has been removed and each subgradient computation is solely based on the current channel state. Also, the basic structure of OAA remains the same as that of the off-line algorithm, which means that OAA can also be implemented in a distributed fashion. Thus, the approach based on subgradient approximation in (15) and dual updates in (16) meets our goal.

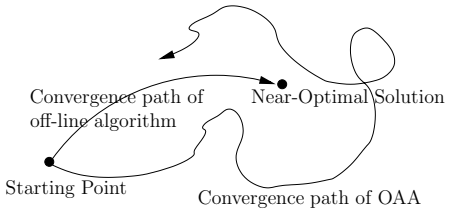


Figure 8: Illustration of OAA's convergence path with respect to that of the off-line algorithm.

However, we note that the such changes in (15) and (16) could have a profound impact on the algorithm's convergence behavior. This is because the iterates of OAA are no longer deterministic and the convergence path may deviate randomly from that of the off-line algorithm, as shown in Figure 8. As a result, the convergence behavior of OAA needs to be investigated. Further, even if OAA converges,

the quality of the achieved solution (how far it deviates from the near-optimal solution achieved by the off-line algorithm) remains unclear. These issues will be our focus in Section 4.2.

4.2 Convergence and Optimality

In this section, we show that OAA converges with probability one to the *same* near-optimal solution obtained by the off-line algorithm. We employ recent results in stochastic approximation theory [14] to establish the convergence and near-optimality of OAA. Our approach is similar to that by Zhang *et al.* in [32], which studied the impact of noisy message exchanges on the well-known distributed network utility maximization (NUM) framework [13]. The goal of [32] is to understand how an existing algorithm would fare under a new environment where some ideal assumptions may not hold. By contrast, our goal here is to design a new online algorithm to approximate an off-line algorithm so as to eliminate the requirement of CDI. Due to this difference, different algorithms are used in the dual domain: we use a dual decomposition algorithm with dual updates only, while [32] employs a primal-dual algorithm, where primal and dual variables are updated simultaneously.

In this section, we outline the main results on the convergence and near-optimality of OAA and the key steps of the proofs. First, we will show that the iterates $\{\hat{\mathbf{u}}(n) : n = 1, 2, \ldots\}$ generated by OAA are recurrent with probability one to an arbitrarily small neighborhood of the optimal solution $\mathbf{u}^*$.

Defining a Lyapunov function $V(\hat{\mathbf{u}}) \triangleq \|\hat{\mathbf{u}} - \mathbf{u}^*\|^2$ and denoting a ball of radius $\sqrt{\epsilon}$ centered at $\mathbf{u}^*$ as $\mathbb{B}_\epsilon \triangleq \{\hat{\mathbf{u}} : V(\hat{\mathbf{u}}) \leq \epsilon\}$ for some given $\epsilon > 0$, we have the following lemma.

LEMMA 2. *If the step sizes $\{s_n\}$ satisfy (12), then the iterates $\{\hat{\mathbf{u}}(n) : n = 1, 2, \ldots\}$ generated by OAA are recurrent with probability one to $\mathbb{B}_\epsilon$.*

The main idea of the proof is similar to that in [32], so we just give a sketch of the important steps here. The recurrence of $\{\hat{\mathbf{u}}(n) : n = 1, 2, \ldots\}$ to $\mathbb{B}_\epsilon$ can be proved by showing that $V(\hat{\mathbf{u}})$ has a negative drift [14]. To see this, we first examine the structure of the stochastic subgradients during the nth iteration. It can be readily verified that the stochastic subgradient $\hat{d}_{u_k}(n)$ can be decomposed as $\hat{d}_{u_k}(n) = d_{u_k}(n) + \xi_{u_k}(n)$, where $\xi_{u_k}(n)$ is a noise term and is defined as

$$\begin{aligned}\xi_{u_k}(n) &\triangleq \hat{d}_{u_k}(n) - \mathbb{E}_{\mathbf{H}}^{(n)}\left[\hat{d}_{u_k}(n)\right] \\ &= \mathbb{E}_{\mathbf{H}}^{(n)}\left[\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{Q}_k^{(m)*}(n))\right] - \sum_{m=1}^{M} \mathrm{Tr}(\mathbf{Q}_k^{(m)*}(n)). \quad (17)\end{aligned}$$

In (17), $\mathbb{E}_{\mathbf{H}}^{(n)}[\cdot]$ denotes the expectation over the distribution of $\mathbf{H}$ conditioned on the previous $(n-1)$ iterates. Then, the stochastic dual variable updates can be written as

$$\hat{u}_k(n+1) = \hat{u}_k(n) - s_n\left[d_{u_k}(n) + \xi_{u_k}(n)\right] + z_{u_k}(n),\ \forall k,$$

where $z_{u_k}(n)$ is a correction term that projects the stochastic subgradient back to the non-negative orthant. Noting that the correction term $\xi_{u_k}(n)$ is a non-expansive mapping [3, Prop. 2.2.1], which does not increase the norm of $\hat{u}_k(n) - s_n\left[d_{u_k}(n) + \xi_{u_k}(n)\right]$, we have

$$\begin{aligned}|\hat{u}_k(n+1) - u_k^*|^2 \leq |\hat{u}_k(n) - u_k^*|^2 - 2s_n(\hat{u}_k(n) - u_k^*) \times \\ \left[d_{u_k}(n) + \xi_{u_k}(n)\right] + s_n^2\left[d_{u_k}(n) + \xi_{u_k}(n)\right]^2.\end{aligned}$$

Suppose that the primal problem is feasible. Then, the dual objective function $\Theta(\mathbf{u})$ is bounded from below. Thus, its subgradient $d_{u_k}(n)$ is bounded. From OAA's dual updates in (16), we conclude that $\hat{u}_k(n)$ is bounded for all n. Thus, $\Theta(\hat{\mathbf{u}}(n))$ is bounded, which in turn implies that $\hat{d}_{u_k}(n)$ is also bounded. Since $\hat{d}_{u_k}(n)$ is bounded, we have that $|\xi_{u_k}(n)|$ is bounded as well. The boundedness of these terms implies that they can be driven to zero by the diminishing step size s_n as $n \to \infty$. Also, since $\mathbb{E}_{\mathbf{H}}^{(n)}[\xi_{u_k}(n)] = 0$, we have $s_n(\hat{u}_k(n) - u_k^*)\mathbb{E}_{\mathbf{H}}^{(n)}[\xi_{u_k}(n)] = 0$. Thus, by rearranging terms and taking expectation on both sides, it follows that (in vector form),

$$\begin{aligned}\mathbb{E}_{\mathbf{H}}^{(n)}\left[\|\hat{\mathbf{u}}(n+1) - \mathbf{u}^*\|^2\right] - \|\hat{\mathbf{u}}(n) - \mathbf{u}^*\|^2 \leq \\ -2s_n\langle\hat{\mathbf{u}}(n) - \mathbf{u}^*, \mathbf{d}_{\mathbf{u}}(n)\rangle + O(s_n^2). \quad (18)\end{aligned}$$

Since $s_n > 0$, we can see from (18) and the definition of $V(\hat{\mathbf{u}})$ that in order for $V(\hat{\mathbf{u}})$ to have a negative drift, it is equivalent to showing that $\langle\hat{\mathbf{u}}(n) - \mathbf{u}^*, \mathbf{d}_{\mathbf{u}}(n)\rangle \geq 0$. This holds because for $\hat{\mathbf{u}}(n)$, we have

$$\langle\hat{\mathbf{u}}(n) - \mathbf{u}^*, \mathbf{d}_{\mathbf{u}}(n)\rangle \geq \Theta(\hat{\mathbf{u}}(n)) - \Theta(\mathbf{u}^*) \geq 0. \quad (19)$$

In (19), the first inequality follows from the convexity of $\Theta(\mathbf{u})$ (the convexity of $\Theta(\mathbf{u})$ implies $\Theta(\hat{\mathbf{u}}(n)) \geq \Theta(\mathbf{u}^*) + \mathbf{d}_{\mathbf{u}}(n)^T(\hat{\mathbf{u}}(n) - \mathbf{u}^*)$). The second inequality follows from the fact that $\mathbf{u}^*$ is a global minimizer of $\Theta(\mathbf{u})$. This shows that, when $\hat{\mathbf{u}}(n) \notin \mathbb{B}_\epsilon$, there exists $\delta_\epsilon \geq 0$ such that $\langle\hat{\mathbf{u}}(n) - \mathbf{u}^*, \mathbf{d}_{\mathbf{u}}(n)\rangle \geq \delta_\epsilon$, which means that the drift in (18) is negative. It then follows that $\{\hat{\mathbf{u}}(n) : n = 1, 2, \ldots\}$ returns to $\mathbb{B}_\epsilon$ with probability one.

After showing the recurrence of $\{\hat{\mathbf{u}}(n) : n = 1, 2, \ldots\}$ in Lemma 2, we now show that, if the noise term $\xi_{\mathbf{u}}(n)$ drives $\hat{\mathbf{u}}(n)$ away from $\mathbb{B}_\epsilon$, the deviation of the iterates $\{\hat{\mathbf{u}}(n)\}$ is bounded. This result is stated in the following lemma.

LEMMA 3. *If the step sizes $\{s_n\}$ satisfy (12), then for any $\epsilon > 0$, there exists an n_0 such that for $n \geq n_0$, the trajectory of $\{\hat{\mathbf{u}}(n) : n > n_0\}$ is bounded by the contraction region $\mathbb{B}_{3\epsilon}$ almost surely.*

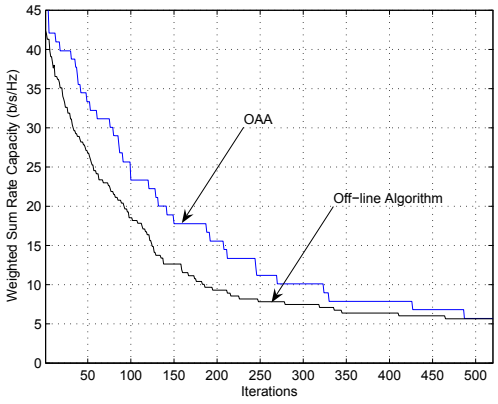
The proof of this lemma is similar to that in [32] and so we only give a brief sketch of the proof here. We first show that if n_0 is sufficiently large and $\hat{\mathbf{u}}(n_0+1)$ moves out of $\mathbb{B}_\epsilon$, then $\hat{\mathbf{u}}(n_0+1)$ resides in $\mathbb{B}_{2\epsilon}$ with probability one. This is true because by Chebyshev's inequality, we have

$$\Pr\left(s_{n_0}\|\xi_{\mathbf{u}}(n_0)\| > \epsilon\right) \leq \frac{s_{n_0}^2\mathbb{E}_{\mathbf{H}}^{(n)}[\|\xi_{\mathbf{u}}(n_0)\|^2]}{\epsilon^2}. \quad (20)$$

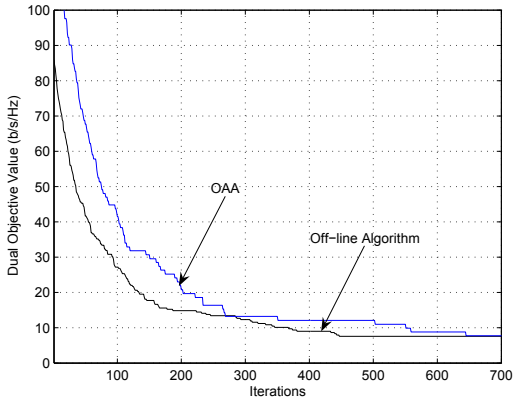
Since $s_n^2 \to 0$, then when n_0 is large enough, the change in one step is almost surely no greater than ϵ. Thus, $\hat{\mathbf{u}}(n_0+1)$ resides in $\mathbb{B}_{2\epsilon}$ with probability one.

Let $\bar{\mathbb{B}}_\epsilon$ denote the complement of $\mathbb{B}_\epsilon$. We now show that if $\hat{\mathbf{u}}(n_0+1) \in \mathbb{B}_{2\epsilon} \cap \bar{\mathbb{B}}_\epsilon$, then for all $n > n_0+1$, $\hat{\mathbf{u}}(n) \in \mathbb{B}_{3\epsilon}$ with probability one. That is, the trajectory of the series starting from $\hat{\mathbf{u}}(n_0+1)$ stays in $\mathbb{B}_{3\epsilon}$ almost surely. By using the inequality [14, Eq. (1.4)], we have that, for $m \geq n_0+1$,

$$\Pr\left\{\sup\left|\sum_{n=n_0+1}^{m} s_n\|\xi_{\mathbf{u}}^{(n)}\|\right| \geq \epsilon\right\} \leq \frac{\bar{\mathbb{E}}_{\mathbf{H}}^{(n)}[\|\xi_{\mathbf{u}}^{(n)}\|^2]}{\epsilon^2}\sum_{\substack{n=\\ n_0+1}}^{\infty} s_n^2,$$



(a) Example 1 in Figure 7(a)



(b) Example 2 in Figure 7(c)

Figure 9: Convergence of the OAA algorithm.

where $\bar{\mathbb{E}}_{\mathbf{H}}^{(n)}[\|\xi_{\mathbf{u}}(n)\|^2] \triangleq \limsup_n \mathbb{E}_{\mathbf{H}}^{(n)}[\|\xi_{\mathbf{u}}(n)\|^2]$. It then follows from the conditions on s_n in (12) and the boundedness of $\|\xi_{\mathbf{u}}(n)\|$ that

$$\lim_{m\to\infty} \Pr\left\{\sup\left|\sum_{n=n_0+1}^{m} s_n\|\xi_{\mathbf{u}}(n)\|\right| \geq \epsilon\right\} = 0, \ \forall \epsilon > 0.$$

In other words, the accumulated distance deviating from $\hat{\mathbf{u}}(n_0+1)$ is almost surely less than ϵ. It then follows that the trajectory resides in $\mathbb{B}_{3\epsilon}$ with probability one.

Combining Lemmas 2 and 3, and noting that ϵ can be made arbitrary small, we have the following theorem.

THEOREM 3. *If the step sizes $\{s_n\}$ satisfy (12), then the iterates $\{\hat{\mathbf{u}}(n), n = 1, 2, \ldots\}$ generated by OAA converge to the same near-optimal solution $\mathbf{u}^*$ of (8) with probability one.*

We use the previous two 5-link MC-MIMO ad hoc network examples to illustrate the convergence behavior of the OAA algorithm. For comparison, we plot the convergence processes of OAA and the off-line algorithm in Figures 9(a) and 9(b), respectively. It can be seen that after approximately 500 and 700 iterations, respectively, OAA converges to the same solutions obtained by the off-line algorithm in these two examples. We note that although the number of iterations of OAA is slightly larger than that for the off-line algorithm, the running time of OAA is shorter because OAA does not require computing expectations over the fading distribution.

5. RELATED WORK

There has been great interest on MC-MIMO at the physical layer in recent years, e.g., joint transceiver design [18,19], channel estimation [23, 31], multiple access strategies [27], limited feedback and antenna selection techniques [20]. An excellent overview of MC-MIMO physical layer techniques can be found in [12]. Another closely-related line of research is on joint subcarrier scheduling and power spectrum optimization for multi-carrier communications systems with single-antenna [1, 15, 24, 25, 28]. Andrews and Zhang [1] studied the multiple carrier scheduling problem for WiMAX downlink and designed several constant factor approximation algorithms to solve the problem. Wong *et al.* [28] showed that adaptive subcarrier loading and modulation can substantially increase the capacity of a multi-carrier communication system. Song and Li [24,25] proposed a framework for subcarrier assignment and power allocation in OFDM wireless networks and designed various optimization algorithms based on this framework. However, their framework did not consider the channel variations in time domain. In [15] , Liu *et al.* proposed a scheduling algorithm at the MAC layer for multiple connections with different QoS requirements, where each connection employs adaptive modulation and coding schemes at the physical layer over fading channels. The limitation of [15] is that the scheduling algorithm does not consider power control.

6. CONCLUSION

In this paper, we investigated the important problem of subcarrier scheduling and power control/allocation for broadband MC-MIMO ad hoc networks. We considered the so-called MWSR problem, which is a challenging non-convex optimization problem. With the use of a perturbation function, we showed that the MWSR problem has a zero duality gap, thus enabling us to study the problem in its dual domain. This approach is highly desirable as it is able to significantly reduce the problem complexity. We designed an off-line algorithm in the dual domain, which can achieve near-optimal performance. We also proposed an online adaptive algorithm (OAA) that eliminates the full CDI requirement in the off-line algorithm. We showed that this OAA converges with probability one to the same solution obtained by the off-line algorithm.

7. ACKNOWLEDGEMENTS

The authors thank the anonymous reviewers for their constructive comments. The work of Y.T. Hou, J. Liu, and Y. Shi has been supported in part by the National Science Foundation (NSF) under Grant CNS-0721421 and Office of Naval Research (ONR) under Grant N00014-08-1-0084. The work of H.D. Sherali has been supported in part by NSF Grant CMMI-0552676.

8. REFERENCES

[1] M. Andrews and L. Zhang. Scheduling algorithms for multi-carrier wireless data systems. In *Proc. ACM Mobicom*, pages 3–14, Montréal, Québec, Canada, Sept.9-14, 2007.

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

[3] D. P. Bertsekas, A. Ndeić, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 1 edition, 2003.

[4] H. Boche and M. Wiczanowski. Stability optimal transmission policy for the multiple antenna multiple access channel in the geometric view. *EURASIP Signal Process. J. (Special Issue on Advances in Signal Processing-Assisted Cross-Layer Designs)*, 86(8):1815–1833, Aug. 2006.

[5] J. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Exampels*. Springer, New York, NY, 2 edition, 2006.

[6] G. J. Foschini. Layered space-time architecture for wireless communication in a fading envorinment when using multi-element antennas. *Bell Labs Tech. J.*, 1(2):41–59, 1996.

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

[8] D. Y. Gao. *Duality Principles in Nonconvex Systems: Theory, Methods and Applications*. Kluwer Academic Publishers., Dordrecht/Boston/London, 2000.

[9] A. M. Geoffrion. Duality in nonlinear programming: A simplified applications-oriented development. *SIAM Review*, 13(1):1–37, Jan. 1971.

[10] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE J. Sel. Areas Commun.*, 21(1):684–702, June 2003.

[11] P. Hande, S. Zhang, and M. Chiang. Distributed rate allocation for inelastic flows. *IEEE/ACM Trans. Netw.*, 15(6):1240–1253, Dec. 2007.

[12] M. Jiang and L. Hanzo. Multiuser MIMO-OFDM for next generation wireless systems. *Proc. IEEE*, 95(7):1430–1469, July 2007.

[13] F. P. Kelly, A. K. Malullo, and D. K. H. Tan. Rate control in communications networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[14] H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York, 2003.

[15] Q. Liu, X. Wang, and G. B. Giannakis. A cross-layer scheduling algorithm with QoS support in wireless networks. *IEEE Trans. Veh. Technol.*, 55(3):839–847, May 2006.

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

[17] M. J. Neely, E. Modiano, and C. E. Rohrs. Power allocation and routing in multibeam satellites with time-varying channels. *IEEE/ACM Trans. Netw.*, 11(2):138–152, Feb. 2003.

[18] D. P. Paloma. Convex primal decomposition for multicarrier linear MIMO transceivers. *IEEE Trans. Signal Process.*, 53(12):4661–4674, Dec. 2005.

[19] D. P. Paloma, J. M. Cioffi, and M. A. Lagunas. Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization. *IEEE Trans. Signal Process.*, 51(9):2381–2401, Sept. 2003.

[20] T. Pande, D. J. Love, and J. V. Krogmeier. Reduced feedback MIMO-OFDM precoding and antenna selection. *IEEE Trans. Signal Process.*, 55(5):2284–2293, May 2007.

[21] T. S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, Upper Saddle River, NJ, 2002.

[22] H. D. Sherali and W. P. Adams. *A Reformulation-Linearization-Technique for Solving Discrete and Continuous Nonconvex Problems*. Kluwer Academic Publishing, Boston, MA, 1999.

[23] C. Shin, R. W. Heath, and E. J. Powers. Blind channel estimation for MIMO-OFDM systems. *IEEE Trans. Veh. Technol.*, 56(2):670–685, Mar. 2007.

[24] G. Song and Y. Li. Cross-layer optimization for OFDM wireless networks – Part I: Theoretical framework. *IEEE Trans. Wireless Commun.*, 4(2):614–624, Mar. 2005.

[25] G. Song and Y. Li. Cross-layer optimization for OFDM wireless networks – Part II: Algorithm development. *IEEE Trans. Wireless Commun.*, 4(2):625–634, Mar. 2005.

[26] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Trans. Telecomm.*, 10(6):585–596, Nov. 1999.

[27] S. Visuri and H. Bölcskei. Multiple-access strategies for frequency-selective MIMO channels. *IEEE Trans. Inf. Theory*, 52(9):3980–3993, Sept. 2006.

[28] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch. Multiuser OFDM with adaptive subcarrier, bit, and power allocation. *IEEE J. Sel. Areas Commun.*, 17(10):1747–1758, Oct. 1999.

[29] E. M. Yeh and A. S. Cohen. Information theory, queueing, and resource allocation in multi-user fading communications. In *Proc. Conf. Inf. Sci. Syst.*, Princton, NJ, Mar. 2004.

[30] W. Yu and R. Li. Dual methods for nonconvex spectrum optimization of multicarrier systems. *IEEE Trans. Commun.*, 54(7):1310–1322, July 2006.

[31] Y. Zeng, A. R. Leyman, and T.-S. Ng. Joint semiblind frequency offset and channel estimation for multiuser MIMO-OFDM uplink. *IEEE Trans. Commun.*, 55(12):2270–2278, Dec. 2006.

[32] J. Zhang, D. Zheng, and M. Chiang. The impact of stochastic noisy feedback on distributed network utility maximization. *IEEE Trans. Inf. Theory*, 54(2):645–665, Feb. 2008.

[33] L. Zheng and D. N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inf. Theory*, 49(5):1073–1096, May 2003.