Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints

Wei Yu, *Member, IEEE*, and Tian Lan, *Student Member, IEEE*

Abstract—This paper considers the transmitter optimization problem for a multiuser downlink channel with multiple transmit antennas at the base-station. In contrast to the conventional sum-power constraint on the transmit antennas, this paper adopts a more realistic per-antenna power constraint, because in practical implementations each antenna is equipped with its own power amplifier and is limited individually by the linearity of the amplifier. Assuming perfect channel knowledge at the transmitter, this paper investigates two different transmission schemes under the per-antenna power constraint: a minimum-power beamforming design for downlink channels with a single antenna at each remote user and a capacity-achieving transmitter design for downlink channels with multiple antennas at each remote user. It is shown that in both cases, the per-antenna downlink transmitter optimization problem may be transformed into a dual uplink problem with an uncertain noise. This generalizes previous uplink–downlink duality results and transforms the per-antenna transmitter optimization problem into an equivalent minimax optimization problem. Further, it is shown that various notions of uplink–downlink duality may be unified under a Lagrangian duality framework. This new interpretation of duality gives rise to efficient numerical optimization techniques for solving the downlink per-antenna transmitter optimization problem.

Index Terms—Beamforming, broadcast channel, capacity region, dirty-paper coding, Lagrangian duality.

I. INTRODUCTION

CONSIDER a wireless multi-antenna downlink transmission scenario with a base-station equipped with N transmit antennas sending independent information to K remote users, each equipped with M receive antennas. This downlink channel is often modeled as a vector Gaussian broadcast channel

$$\mathbf{y_i} = \mathbf{H_i}\mathbf{x} + \mathbf{z_i}, \qquad i = 1, \ldots, K \qquad (1)$$

where $\mathbf{x} = [x_1 \cdots x_N]^T$ is an $N \times 1$ complex vector representing the transmit signal, $\mathbf{H_i}$'s are $M \times N$ complex channel matrices which are assumed to be known to both the transmitter and the receivers, $\mathbf{y_i}$'s are $M \times 1$ complex vectors representing the received signal, and $\mathbf{z_i}$'s are the i.i.d. additive complex Gaussian noise vectors with variance $\sigma^2/2$ on each of its real and imaginary components. A great deal of recent progress has been made in the characterization of optimal transmission schemes for the downlink channel. For example, in the special case in which each remote user is equipped with a single antenna, i.e., the multiple-input, single-output (MISO) case, a *linear* beamforming strategy may be employed at the transmitter. In this scenario, K beamforming vectors may be used to create K data streams. The achievable data rate of each data stream depends on its signal-to-interference-and-noise ratio (SINR). The transmitter optimization problem can then be formulated as a problem of minimizing the total transmit power subject to a minimum SINR for each user. This problem has been solved in [1]–[8]. Alternatively, it is possible to achieve a higher throughput by employing *nonlinear* interference presubtraction methods at the transmitter. In this case, multiuser interference may be presubtracted using the method of "dirty-paper coding" [9], which has been shown to achieve the Shannon capacity of the downlink channel. The transmitter optimization problem under dirty-paper coding can be formulated as a weighted rate maximization problem subject to a total power constraint at the transmitter. The solution to this capacity region maximization problem for the general multiple-input multiple-output (MIMO) case has been provided in [10]–[14].

A key technique in the solutions of both the MISO minimum-power beamforming problem [1]–[8] and the MIMO capacity region maximization problem [10]–[14] is the idea of uplink–downlink duality. It can be shown that under a sum power constraint, the optimal beamforming SINR region for a downlink MISO channel is identical to that of a dual uplink channel in which the transmitter and the receivers are reversed and the channel matrices are transposed [1], [2]. Likewise, with interference presubtraction at the transmitter, a similar MIMO downlink capacity region duality can be established [12], [13], [15]. For both the beamforming and the capacity region problems, duality is useful, because the uplink problem has an analytical structure which is computationally easier to handle. Thus, the downlink problem may be more easily solved by solving its uplink dual.

However, all existing duality results in the literature depend crucially on the sum-power constraint across all transmit antennas, which is of the form

$$\sum_{i=1}^{N} \mathbb{E}|x_i|^2 \leq P. \qquad (2)$$

While analytically attractive, such a sum power constraint is often unrealistic in practice. In a physical implementation of a multi-antenna base-station, each antenna has its own power amplifier in its analog front-end, and is limited individually by the linearity of the power amplifier. Thus, a power constraint imposed on a per-antenna basis is more realistic:

$$\mathbb{E}|x_i|^2 \le P_i, \qquad i = 1, \ldots, N. \qquad (3)$$

The main objective of this paper is to provide a solution to the optimal beamforming and the capacity region problems for a multi-antenna downlink channel with per-antenna power constraints. The per-antenna power constraint is also motivated by future wireless systems in which base-stations can be connected via high-speed links and are capable of cooperatively transmitting information to and receiving information from mobile users. In this case, an individual power constraint must be applied to a group of antennas on each base-station. A similar situation is applicable for downlink wireline systems in which joint transmission may be done at the central office by coordinating multiple modems. In this case, an individual power constraint is again applicable to each modem.

This paper shows that the solution to the per-antenna problem arises naturally from a new interpretation of uplink–downlink duality. In the existing literature, the beamforming duality and the capacity region duality are derived using seemingly different methods. This paper shows that they can be unified under the framework of Lagrangian duality in convex optimization. This novel viewpoint leads to a solution to the per-antenna downlink problem as follows. We show that for both the beamforming problem and the capacity region problem, the dual of a multi-antenna downlink channel with per-antenna power constraints is an uplink channel with an uncertain and diagonally constrained noise. For the beamforming problem with one antenna at each remote user, the SINR regions of the two respective uplink and downlink channels are identical; for the capacity region problem with an arbitrary number of antennas at the remote users, the capacity regions of the two respective channels are also identical.

The formulation of the dual of the per-antenna constrained downlink problem as an uplink problem with an uncertain noise also gives rise to numerical methods for solving the per-antenna downlink problem. For the downlink beamforming problem, this paper proposes two efficient numerical algorithms for finding the optimal beamformers under the per-antenna power constraints: an iterative approach that updates the dual uplink transmit and noise covariance matrices in each iteration, and an interior-point based algorithm that updates the transmit and noise covariance matrices at the same time. For the capacity region problem, we apply a Newton's method for minimax optimization [16] that finds an efficient search direction for the maximization and the minimization parts of the problem simultaneously.

This paper makes the following assumptions in the problem formulation. First, perfect channel knowledge is assumed to be available at the transmitter. In a practical implementation, the channel realization has to be estimated at the mobile stations and fedback to the base-station. Second, only the single-cell case is considered in this paper; out-of-cell interference is assumed to be stationary; no multi-base-station joint beamforming is assumed to be possible.

A. Related Work

The beamforming duality between the uplink and downlink channels was discovered by Rashid-Farrokhi, Liu, and Tassiulas [1] and Visotsky and Madhow [2]. They showed that the optimal downlink beamforming problem under SINR constraints can be solved efficiently by an iterative uplink beamformer and power update algorithm. It is well known that the uplink beamforming problem is much easier to solve [3]. Thus, by transforming the nonconvex downlink beamforming problem into the uplink domain, the downlink problem may be solved efficiently as well. In particular, [2] offered an optimality proof for uplink–downlink duality based on an examination of the Karush–Kuhn–Tucker (KKT) condition for the optimization problems. The convergence proof for the iterative algorithm is based on the properties of fixed-point iteration. In a separate work, a different approach to the same problem was provided by Schubert and Boche [4], where an iterative algorithm was proposed to solve not only the minimal transmit power problem for a fixed SINR, but also the maximum minimal-SINR problem under a sum-power constraint. In [5], the same authors also extended their solution to include "dirty-paper coding." In another different approach, Bengtsson and Ottersten ([6], [7]) introduced a semidefinite programming (SDP) framework for the downlink beamforming problem. They proved that despite the apparent nonconvexity of the problem, its SDP relaxation achieves the global optimum of the original problem. In a recent work, Wiesel, Eldar and Shamai [8] formulated the problem as a second-order cone programming (SOCP) problem, and showed that the downlink beamforming problem can be directly turned into a convex problem. They further proposed a simple and fast fixed-point iteration algorithm for its solution. All of these above approaches essentially solve the same problem: the downlink beamforming problem with a sum-power constraint. The main point of this paper is that uplink–downlink duality may be extended to downlink problems with a per-antenna power constraints via a Lagrangian duality approach.

Recently, a great deal of progress has also been made in characterizing the information-theoretical capacity region of the downlink channel. Under a sum power constraint, Caire and Shamai [10] showed that the so-called dirty-paper precoding strategy [9] is optimal for the sum capacity of a downlink channel with two transmit antennas. This result has since been generalized for sum capacity but with an arbitrary number of users and an arbitrary number of transmit and receive antennas in [11]–[13], and for the entire capacity region by Weingarten, Steinberg, and Shamai [14]. The sum capacity result in [11] and that in [12] and [13], although equivalent, also have subtle differences. The approach of [11] is based on a generalized decision-feedback equalizer and is applicable to a downlink channel with arbitrary input constraints, while the alternative approaches in [12] and [13] establish the capacity result via uplink–downlink duality, but only for the sum-power constrained case. Uplink–downlink duality here refers to the fact that the dirty-paper region for the downlink channel and the capacity region of the reciprocal uplink channel are the same under a sum power constraint. An objective of this paper is to generalize this capacity region duality to downlink channels
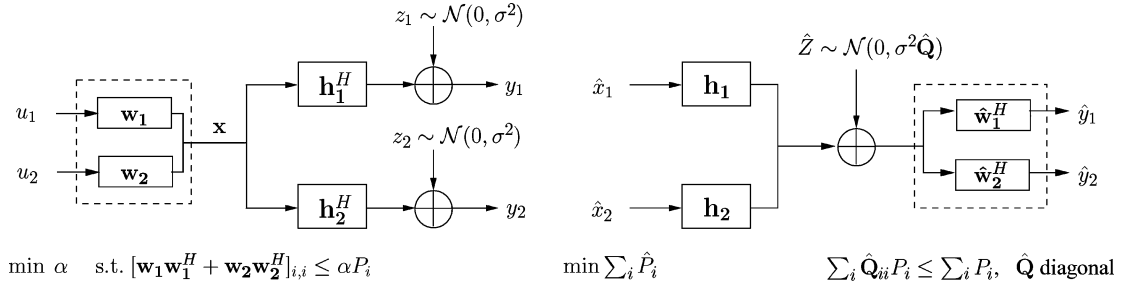
Fig. 1. Uplink–downlink beamforming duality with per-antenna power constraints.

with per-antenna power constraints. The generalization of uplink–downlink duality to a per-antenna power constrained case has been undertaken in part, for sum capacity, in our previous work [17]. This paper treats the entire capacity region.

B. Organization

The remainder of this paper is organized as follows. In Section II, we consider the beamforming problem and establish the duality between the downlink channel with per-antenna power constraints and the uplink channel with an uncertain noise. In Section III, we establish the same duality relation for the capacity region problem with per-antenna power constraints. Several extensions and generalizations are treated in Section IV. Numerical algorithms are proposed in Section V. Simulation results are presented in Section VI. Conclusions are drawn in Section VII.

Throughout this paper, we use upper-case bold face letters, such as $\mathbf{H}$, for matrices and lower-case bold face letters, such as $\mathbf{w}$, for vectors. All vectors are column vectors unless otherwise stated. Further, $\succeq$ is used to denote matrix inequality defined on the cone of nonnegative definite matrices.

II. BEAMFORMING DUALITY

A. Channel Model

In this section, we investigate the multi-antenna transmitter beamforming problem for a downlink channel with per-antenna power constraints at the transmitter and where each remote user is equipped with a single antenna. In this case, the channel model (1) simplifies to the following: the received vector $\mathbf{y_i}$ becomes a complex scalar, denoted by y_i, and the channel matrix $\mathbf{H_i}$ becomes an $1 \times N$ complex vector, denoted as $\mathbf{h_i}^H$. This channel model is also applicable to the multi-receive-antenna case with fixed receiver beamforming vectors.

In a beamforming design, the transmit signal is of the form

$$\mathbf{x} = \sum_{i=1}^{K} u_i \mathbf{w_i} \qquad (4)$$

where u_i is a complex scalar denoting the information signal for user i, and $\mathbf{w_i}$ is a $N \times 1$ beamforming vector for user i. Without loss of generality, let $\mathbb{E}|u_i|^2 = 1$. The received signal is

$$y_i = \mathbf{h_i}^H \left(\sum_{j=1}^{K} u_j \mathbf{w_j} \right) + z_i, \qquad i = 1, \ldots, K \qquad (5)$$

where z_i is the additive white Gaussian complex noise with variance $\sigma^2/2$ on each of its real and imaginary components. It is easy to see that the SINR for each user may be expressed as

$$\mathrm{SINR}_i = \frac{\left| \mathbf{h_i}^H \mathbf{w_i} \right|^2}{\sum_{j \neq i} \left| \mathbf{h_i}^H \mathbf{w_j} \right|^2 + \sigma^2}. \qquad (6)$$

The SINR is directly related to system performance metrics such as the bit error rate (BER) and data rate. For example, under a fixed BER and assuming quadrature-amplitude modulation (QAM), a practical achievable rate can be computed as

$$R_i = \log \left(1 + \frac{\mathrm{SINR}_i}{\Gamma} \right), \qquad i = 1, \ldots, K \qquad (7)$$

where Γ is the SNR gap to capacity. The SNR gap is always greater than 1 (0 dB), and it gives an approximate relation between the SINR and the rate [18]. The downlink beamforming problem is illustrated in Fig. 1.

B. Duality With Per-Antenna Power Constraints

The focus of this paper is on the design of beamforming vectors $\mathbf{w_i}$ to minimize the per-antenna power on each transmit antenna while enforcing a set of SINR constraints on each user. One way to formulate such a per-antenna optimization problem is to uniformly minimize the margin of $\mathbb{E}|\mathbf{x}_i|^2/P_i$ over all antennas, i.e.,

$$\begin{align} \text{minimize} \quad & \alpha \qquad &(8) \\ \text{subject to} \quad & \mathbb{E}|\mathbf{x_i}|^2 \leq \alpha P_i, \qquad \forall i \qquad &(9) \\ & \mathrm{SINR}_i \geq \gamma_i, \qquad \forall i \qquad &(10) \end{align}$$

where $\gamma_1, \ldots, \gamma_K$ is a set of given SINR targets, and $P_1, \ldots, P_N$ is a set of given per-antenna power targets. This formulation provides a single measure that reflects the individual transmit power on each antenna.[1]

If the design objective were to minimize the sum transmit power, the downlink beamforming problem could have been easily solved via a dual uplink channel with the same SINR constraints [1]–[8]. In this paper, we revisit the notion of duality and provide a derivation of uplink–downlink beamforming duality based on Lagrangian duality in convex optimization. This viewpoint not only illustrates the duality results of [1]–[8] in a new perspective, but also allows the downlink beamforming problem with per-antenna power constraints to be solved.

---

[1]Alternatively, one might formulate a problem to minimize the total power subject to some individual per-antenna power constraints. Such a formulation would lead to a different dual problem.

We begin the development by restating the downlink beamforming problem as follows:

$$\text{minimize} \quad \alpha \sum_{i=1}^{N} P_i \tag{11}$$

$$\text{subject to} \quad \left[\sum_{j=1}^{K} \mathbf{w_j} \mathbf{w_j}^H \right]_{i,i} \leq \alpha P_i, \quad \forall i. \tag{12}$$

$$\frac{\left| \mathbf{h_i}^H \mathbf{w_i} \right|^2}{\sum_{j \neq i} \left| \mathbf{h_i}^H \mathbf{w_j} \right|^2 + \sigma^2} \geq \gamma_i, \quad \forall i. \tag{13}$$

where $[\cdot]_{i,i}$ denotes the (i,i)-entry of a matrix, and (12) is the per-antenna power constraint. The optimization variables are α and $\mathbf{w_i}$; γ_i, $\mathbf{h_i}$, and P_i are fixed. For the rest of this paper, we make the assumption that the SINR constraints in (8) are such that there exists at least one feasible solution. A rigorous discussion on feasibility can be found in [8]. Note that a constant factor $\sum_{i=1}^{N} P_i$ is included in the objective function [11] so that the objective of the minimization is the total power.

The optimization problem as stated in (11) is not convex. However, it has the following fundamental property.

Proposition 1: Strong duality holds for the optimization problem (11). In other words, the Lagrangian dual of (11) achieves the same optimal value as the original problem (11) itself. The optimal duality gap is zero.

Strong duality is the fundamental reason that the apparently nonconvex optimization problem (11) may be solved globally and efficiently via its uplink dual problem. Strong duality has been observed in [6] and [8] for the sum-power constrained problem. The main idea of the proof is that the apparently nonconvex problem may be transformed into a convex problem. The proof of strong duality for the per-antenna problem is similar and is presented in Section A of the Appendix.

Because of strong duality, the optimal value of the downlink problem can be found by solving for its dual. The main result of this section is that this Lagrangian duality also leads to an uplink–downlink duality. The dual channel in this case is still constrained by the same SINR requirements, but it has an unusual noise whose covariance matrix is uncertain.

The uplink beamforming problem is illustrated in Fig. 1. The dual uplink system is derived from the downlink system by interchanging the input and the output vectors and by transposing the channel matrices. In this case, each remote transmitter is equipped with a single antenna and transmits with a power

$$\hat{P}_i = \mathbb{E} |\hat{x}_i|^2. \tag{14}$$

In an uplink system with multiple antennas at the receiver, the design objective of the receiver beamforming problem is to jointly optimize the power allocation $\hat{P}_i$, and the receiver beamforming vectors $\hat{\mathbf{w}}_\mathbf{i}^H$ to satisfy a set of SINR constraints γ_i. Let $\sigma^2 \hat{\mathbf{Q}}$ be the uplink receiver noise covariance matrix. The uplink optimization problem may be posed as follows:

$$\text{minimize} \quad \sum_{i=1}^{K} \hat{P}_i$$

$$\text{subject to} \quad \frac{\hat{P}_i \left| \hat{\mathbf{w}}_\mathbf{i}^H \mathbf{h_i} \right|^2}{\sum_{j \neq i} \hat{P}_j \left| \hat{\mathbf{w}}_\mathbf{i}^H \mathbf{h_j} \right|^2 + \hat{\mathbf{w}}_\mathbf{i}^H \sigma^2 \hat{\mathbf{Q}} \hat{\mathbf{w}}_\mathbf{i}} \geq \gamma_i. \tag{15}$$

We now state the main result of this section.

Theorem 1: The optimal downlink beamforming problem (11) with per-antenna power constraints $(P_1, \ldots, P_N)$ can be solved via a dual uplink channel in which the SINR constraints remain the same and the noise is uncertain. More precisely, the Lagrangian dual of the downlink beamforming problem (11) is the following minimax problem:

$$\max_{\mathbf{Q}} \min_{\lambda_i, \hat{\mathbf{w}}_\mathbf{i}} \quad \sum_{i=1}^{K} \lambda_i \sigma^2$$

$$\text{subject to} \quad \frac{\lambda_i \sigma^2 \left| \hat{\mathbf{w}}_\mathbf{i}^H \mathbf{h_i} \right|^2}{\sum_{j \neq i} \lambda_j \sigma^2 \left| \hat{\mathbf{w}}_\mathbf{i}^H \mathbf{h_j} \right|^2 + \hat{\mathbf{w}}_\mathbf{i}^H \sigma^2 \mathbf{Q} \hat{\mathbf{w}}_\mathbf{i}} \geq \gamma_i$$

$$\text{tr}(\mathbf{Q}\boldsymbol{\Phi}) \leq \text{tr}(\boldsymbol{\Phi}), \quad \mathbf{Q} \text{ diagonal}, \quad \mathbf{Q} \succeq 0 \tag{16}$$

where $\boldsymbol{\Phi} = \text{diag}(P_1, \ldots P_N)$ is a diagonal matrix of per-antenna powers, λ_i is the dual variable associated with the SINR constraint, and $\mathbf{Q} = \text{diag}(q_1, \ldots, q_N)$ is a diagonal matrix of dual variables associated with the per-antenna power constraints in the downlink problem. This dual problem can be interpreted as an uplink problem (15) in which $\hat{P}_i = \lambda_i \sigma^2$ is the dual uplink power, and the SINR constraints have to be satisfied for all diagonal dual uplink noise covariance matrices $\sigma^2 \hat{\mathbf{Q}} = \sigma^2 \mathbf{Q}$.

Proof: We first derive the dual of the downlink beamforming problem (11). With a simple manipulation of the SINR constraints (13), the Lagrangian for the downlink optimization problem (11) is given by

$$L(\alpha, \mathbf{w}_i, q_i, \lambda_i) = \alpha \sum_{i=1}^{N} P_i + \sum_{i=1}^{N} q_i \left\{ \left[\sum_{j=1}^{K} \mathbf{w_j} \mathbf{w_j}^H \right]_{i,i} - \alpha P_i \right\}$$

$$- \sum_{i=1}^{K} \lambda_i \left\{ \frac{1}{\gamma_i} \left| \mathbf{h_i}^H \mathbf{w_i} \right|^2 - \sum_{j \neq i} \left| \mathbf{h_i}^H \mathbf{w_j} \right|^2 - \sigma^2 \right\} \tag{17}$$

where q_i's are the Lagrange multipliers corresponding to per-antenna power constraints (12), and λ_i's are the Lagrange multipliers corresponding to SINR constraints (13).

Let $\mathbf{Q} = \text{diag}(q_1, \ldots, q_N)$ and $\boldsymbol{\Phi} = \text{diag}(P_1, \ldots, P_N)$. Rearranging the terms of (17), we obtain

$$L(\alpha, \mathbf{w}_i, \mathbf{Q}, \lambda) = \sum_{i=1}^{K} \lambda_i \sigma^2 - \alpha \left\{ \text{tr}(\mathbf{Q}\boldsymbol{\Phi}) - \text{tr}(\boldsymbol{\Phi}) \right\}$$

$$+ \sum_{i=1}^{K} \mathbf{w_i}^H \left\{ \mathbf{Q} + \sum_{j \neq i} \lambda_j \mathbf{h_j} \mathbf{h_j}^H - \frac{\lambda_i}{\gamma_i} \mathbf{h_i} \mathbf{h_i}^H \right\} \mathbf{w_i} \tag{18}$$

The dual objective is therefore

$$g(\mathbf{Q}, \lambda_i) = \min_{\mathbf{w_i}} \min_{\alpha} L(\alpha, \mathbf{w}_i, \mathbf{Q}, \lambda_i). \tag{19}$$

Since α must be positive and there are no constraints on the beamformer $\mathbf{w_i}$, it is easy to see that $g(\mathbf{Q}, \lambda_i) = -\infty$ if $\text{tr}(\mathbf{Q}\boldsymbol{\Phi}) \geq \text{tr}(\boldsymbol{\Phi})$ or $\mathbf{Q} + \sum_{j \neq i} \lambda_j \mathbf{h_j} \mathbf{h_j}^H - (\lambda_i / \gamma_i) \mathbf{h_i} \mathbf{h_i}^H$ is not positive semidefinite. As $\mathbf{Q}$ and λ_i should be chosen such that the Lagrangian dual $g(\mathbf{Q}, \lambda_i)$ is finite, the above two

inequalities impose constraints on the dual objective function. Formally, the Lagrangian dual problem can be stated as follows:

$$\max_{\mathbf{Q}} \max_{\lambda_i} \quad \sum_{i=1}^{K} \lambda_i \sigma^2$$
$$\text{subject to} \quad \sum_{j=1}^{K} \lambda_j \mathbf{h_j} \mathbf{h_j}^H + \mathbf{Q} \succeq \left(1 + \frac{1}{\gamma_i}\right) \lambda_i \mathbf{h_i} \mathbf{h_i}^H$$
$$\text{tr}(\mathbf{Q}\boldsymbol{\Phi}) \leq \text{tr}(\boldsymbol{\Phi}), \quad \mathbf{Q} \text{ diagonal}, \quad \mathbf{Q} \succeq 0 \quad (20)$$

By strong duality, the optimal value of the original downlink problem is the same as that of the dual problem.

Next, we show that this dual problem (20) is equivalent to (16). We start by noting that (16) can be interpreted as an uplink problem (15) if we identify $\lambda_i \sigma^2 = \hat{P}_i$ as the dual uplink power and $\sigma^2 \mathbf{Q} = \sigma^2 \hat{\mathbf{Q}}$ as the dual uplink noise covariance. Next, the minimization over $\hat{\mathbf{w}_i}$ in (16) can be solved explicitly for fixed power $\lambda_i \sigma^2$ and fixed noise covariance $\sigma^2 \mathbf{Q}$. Clearly, the optimal receiver beamforming vector $\hat{\mathbf{w}_i}$ that maximizes the SINR is the minimum-mean-squared-error (MMSE) filter.[2] In order to take into account the possibility that the noise covariance matrix $\mathbf{Q}$ may not be strictly positive definite, we use a general form of the MMSE filter as

$$\hat{\mathbf{w}_i} = \left(\sum_{j=1}^{K} \lambda_j \sigma^2 \mathbf{h_j} \mathbf{h_j}^H + \sigma^2 \mathbf{Q}\right)^{\dagger} \mathbf{h_i} \quad (21)$$

where $(.)^{\dagger}$ denotes the matrix pseudoinverse [19]. In fact, when the noise covariance is low rank, the MMSE filter is not unique, but a minimum-norm MMSE filter can always be found by the above expression. When this minimum-norm MMSE filter is used, the SINR constraints are always met with equality.

Now, by substituting the MMSE $\hat{\mathbf{w}_i}$ above into the SINR constraint in (16), we obtain the following constraint:

$$\left(1 + \frac{1}{\gamma_i}\right) \lambda_i \mathbf{h_i}^H \left(\sum_{j=1}^{K} \lambda_j \mathbf{h_j} \mathbf{h_j}^H + \mathbf{Q}\right)^{\dagger} \mathbf{h_i} \geq 1. \quad (22)$$

Next, we claim that the above SINR constraint may be reversed, and the minimization over λ_i in the optimization problem (16) may also be reversed as a maximization, so that (16) may be rewritten as

$$\max_{\mathbf{Q}} \max_{\lambda_i} \quad \sum_{i=1}^{K} \lambda_i \sigma^2$$
$$\text{subject to} \quad \left(1 + \frac{1}{\gamma_i}\right) \lambda_i \mathbf{h_i}^H \left(\sum_{j=1}^{K} \lambda_j \mathbf{h_j} \mathbf{h_j}^H + \mathbf{Q}\right)^{\dagger} \mathbf{h_i} \leq 1$$
$$\text{tr}(\mathbf{Q}\boldsymbol{\Phi}) \leq \text{tr}(\boldsymbol{\Phi}), \quad \mathbf{Q} \text{ diagonal}, \quad \mathbf{Q} \succeq 0. \quad (23)$$

The reversal of the SINR constraints and the reversal of the minimization as a maximization do not affect the solution to the optimization problem (16), because, in an uplink problem, for

each fixed noise covariance $\sigma^2 \mathbf{Q}$, the minimization of power $\lambda_i \sigma^2$ under a set of minimum SINR constraints and the maximization of power under a set of maximum SINR constraints both require the SINR constraints to be met with equality, i.e.,

$$\left(1 + \frac{1}{\gamma_i}\right) \lambda_i \mathbf{h_i}^H \left(\sum_{j=1}^{K} \lambda_j \mathbf{h_j} \mathbf{h_j}^H + \mathbf{Q}\right)^{\dagger} \mathbf{h_i} = 1. \quad (24)$$

This implies that the optimal λ_i in both cases is just the unique fixed point of the above equations [8]. Thus, (16) and (23) have identical solutions due to the uniqueness of the fixed point.

Now, using Lemma 1, which is stated and proved in Section B of the Appendix, the SINR constraint in (23) can be shown to be equivalent to

$$\sum_{j=1}^{K} \lambda_j \mathbf{h_j} \mathbf{h_j}^H + \mathbf{Q} \succeq \left(1 + \frac{1}{\gamma_i}\right) \lambda_i \mathbf{h_i} \mathbf{h_i}^H. \quad (25)$$

Note that this is exactly the SINR constraint in (20). Thus, once interpreted as an uplink problem, (16) may be rewritten as (20). Therefore, the Lagrangian dual of the downlink beamforming problem (11) is equivalent to (16), an uplink beamforming problem with an uncertain noise. □

Corollary 1: At the optimum, the uplink problem (16) and the downlink problem (11) have the same objective value

$$\alpha \sum_{i=1}^{K} P_i = \sum_{i=1}^{K} \lambda_i \sigma^2. \quad (26)$$

Proof: The corollary follows directly from Proposition 1 and Theorem 1. As the uplink problem (16) and the downlink problem (20) are Lagrangian duals of each other, and since strong duality holds, the optimal values of the two problems must be the same. □

Note that in certain degenerate cases, the per-antenna power constraints αP_i in (16) are not met with equality. An example of this is the following 2-user channel with $\mathbf{h}_1 = [1,0]$, $\mathbf{h}_2 = [0,1]$, $P_1 = 1$, $P_2 = 10^6$, $\sigma = 1$, and $\gamma_1 = \gamma_2 = 1$. In this case, the power constraint P_2 is not tight, and strictly speaking, the optimal downlink beamformers $\mathbf{w}_i$'s are also not unique, as mentioned earlier (but the minimum-norm MMSE $\mathbf{w}_i$ is always unique.) In the dual uplink channel, the nonuniqueness is reflected in the fact that the dual variable $\mathbf{Q}$ is not strictly positive definite, and that the MMSE filter $\hat{\mathbf{w}}_i$ is not unique. However, αP_i is always an upper bound on the minimum transmit power, i.e., $|\mathbf{w_i}|^2 \leq \alpha P_i$.

Corollary 2: The optimal beamforming vectors for the uplink problem (16) and for the downlink problem (11) are the same up to a scaling factor. Let the optimal $\hat{\mathbf{w}_i}$ be as in (21). Then, the optimal $\mathbf{w_i} = \sqrt{\delta_i} \hat{\mathbf{w}_i}$ where δ_i may be found by a matrix inversion

$$[\delta_1 \cdots \delta_K]^T = \mathbf{G}^{-1} \mathbf{1} \sigma^2 \quad (27)$$

where the matrix $\mathbf{G}$ is defined as follows: $\mathbf{G}_{ii} = (1/\gamma_i)|\hat{\mathbf{w}}_i^H \mathbf{h_i}|^2$ and $\mathbf{G}_{ij} = -|\hat{\mathbf{w}}_j^H \mathbf{h_i}|^2$ for $i \neq j$, and $\mathbf{1}$ is an all-one vector.

---

[2]The $\hat{\mathbf{w}}_i$ as expressed in (21) is optimal up to an arbitrary scaling factor.

TABLE I
PRIMAL AND DUAL VARIABLES IN BEAMFORMING DUALITY

Downlink		Uplink
$\mathbf{h_i}$	$\Longleftrightarrow$	$\mathbf{h_i}^H$
$\mathbf{w_i}$	$\Longleftrightarrow$	$\hat{\mathbf{w}}_\mathbf{i}^H$
λ_i	$\Longleftrightarrow$	$\hat{P}_i$
q_i	$\Longleftrightarrow$	$\hat{\mathbf{Q}}$

Proof: To find the optimal $\mathbf{w_i}$, we take the gradient of the Lagrangian for the downlink problem (18) with respect to $\mathbf{w_i}$ and set it to zero

$$\frac{\partial L}{\partial \mathbf{w_i}} = \left(\mathbf{Q} + \sum_{j \neq i} \lambda_j \mathbf{h_j}\mathbf{h_j}^H - \frac{\lambda_i}{\gamma_i}\mathbf{h_i}\mathbf{h_i}^H \right) \mathbf{w_i} = 0. \quad (28)$$

To obtain an expression for $\mathbf{w_i}$, add $(1 + (1/\gamma_i))\lambda_i\mathbf{h_i}\mathbf{h_i}^H\mathbf{w_i}$ to both sides of the equation and solve for $\mathbf{w_i}$

$$\mathbf{w_i} = \left(\sum_{j=1}^K \lambda_j \mathbf{h_j}\mathbf{h_j}^H + \mathbf{Q} \right)^\dagger \left(1 + \frac{1}{\gamma_i} \right) \lambda_i \mathbf{h_i}\mathbf{h_i}^H\mathbf{w_i}. \quad (29)$$

Note that $\mathbf{h_i}^H\mathbf{w_i}$ is a scalar. Comparing this expression for $\mathbf{w_i}$ with the expression for $\hat{\mathbf{w}}_\mathbf{i}$ as in (21), we see that $\mathbf{w_i} = \sqrt{\delta_i}\hat{\mathbf{w}}_\mathbf{i}$ where $\sqrt{\delta_i} = \sigma^2(1 + (1/\gamma_i))\lambda_i\mathbf{h_i}^H\mathbf{w_i}$. However, this expression for δ_i depends on $\mathbf{w_i}$ itself. To find δ_i's in terms of $\hat{\mathbf{w}}_\mathbf{i}$'s, which are available from the uplink channel [i.e., (21)], we note that the SINR constraints (13) must be all active at the global optimum point. So

$$\frac{1}{\gamma_i} \left| \mathbf{w_i}^H\mathbf{h_i} \right|^2 = \sum_{j \neq i} \left| \mathbf{w_j}^H\mathbf{h_i} \right|^2 + \sigma^2. \quad (30)$$

Substituting $\mathbf{w_i} = \sqrt{\delta_i}\hat{\mathbf{w}}_\mathbf{i}$ into the above, we obtain a set of K linear equations with K unknowns $\delta_1, \ldots, \delta_K$, which can be solved by an inversion of the matrix $\mathbf{G}$ defined earlier. The matrix inverse exists because the original optimization problem is assumed to be feasible. This solution for δ_i, combined with (21), gives an explicit solution of the downlink beamforming problem via its uplink dual. $\square$

It is interesting to compare the structures of the uplink optimization problem (16) and the downlink problem (11). It is clear that the Lagrange multipliers λ_i corresponding to the SINR constraints in the downlink problem play the role of the noise covariance in the uplink problem. The Lagrange multipliers q_i corresponding to the per-antenna power constraints in the downlink problem play the role of noise covariance matrix in the uplink. This correspondence between the primal and dual variables enhances and generalizes the previous sum-power duality as in [1]–[8]. The duality relation is summarized in Table I.

C. Beamforming With Dirty-Paper Coding

In the formulation of the downlink beamforming problem so far, multiuser interference is regarded as noise. However, from an information theoretical point of view, it is possible for a downlink transmitter to presubtract interference using a technique called "dirty-paper coding." The theoretical basis for dirty-paper coding is due to Costa [9], who proved that the capacity of a Gaussian channel with interference known noncausally at the transmitter is exactly the same as if the interference does not exist. Dirty-paper may be implemented in practice using Tomlinson–Harashima precoding-like techniques [20], [21]. Dirty-paper precoding at the transmitter for the downlink is akin to interference cancellation at the receiver for the uplink. In fact, a duality result can be formally established.

To implement dirty-paper coding, some prior presubtraction order must be fixed. Without loss of generality, let us suppose that the encoding order is $\{1, 2, \ldots, K\}$, i.e., user 1, as the first user to be encoded at the transmitter, sees all other users as interference; user 2 is the second to be encoded with user 1's interference subtracted, etc.; user K has all other users' interference subtracted. In this case, the SINR constraint becomes

$$\frac{\left| \mathbf{h_i}^H\mathbf{w_i} \right|^2}{\sum_{j>i} \left| \mathbf{h_i}^H\mathbf{w_j} \right|^2 + \sigma^2} \geq \gamma_i. \quad (31)$$

With a fixed ordering, [5], [13], and [15] showed that uplink–downlink duality continues to hold with a sum power constraint. The generalization of this result to the per-antenna power constrained case is straightforward.

Corollary 3: Under a fixed subtraction order, the dual of the optimal downlink beamforming problem with dirty-paper coding and with per-antenna power constraints at the transmitter is exactly the optimal uplink beamforming problem with successive interference cancellation and with an uncertain noise at the receiver. More precisely, the dual of the downlink problem (11) with an SINR expression replaced by (31) is the uplink problem (16) with the SINR expression replaced by

$$\frac{\lambda_i\sigma^2 \left| \hat{\mathbf{w}}_\mathbf{i}^H\mathbf{h_i} \right|^2}{\sum_{j>i} \lambda_j\sigma^2 \left| \hat{\mathbf{w}}_\mathbf{i}^H\mathbf{h_j} \right|^2 + \hat{\mathbf{w}}_\mathbf{i}^H\sigma^2\mathbf{Q}\hat{\mathbf{w}}_\mathbf{i}} \geq \gamma_i. \quad (32)$$

Further, the uplink and downlink beamforming vectors are the same up to a scaling factor.

III. ACHIEVABLE RATE REGION DUALITY

In the beamforming problem, we fix the SINR constraints and minimize the transmit power. In many cases, it is advantageous to reverse the problem and consider instead the maximization of achievable rates subject to a fixed transmit power constraint. Traditionally, the beamforming problem and the rate-region problem have been treated independently in the literature. In this section, we show that these two are intimately related and uplink–downlink duality extends naturally to the achievable rate region. In fact, duality is also applicable to the case where remote users are each equipped with multiple receive antennas.

A. Single-Receive-Antenna Case

Consider first the scenario in which each remote user is equipped with a single receive antenna. In this case, the beamforming results stated in the previous section can be directly translated to achievable rate region results, because the rate is directly related to the SINR by $R = \log(1 + \text{SINR}/\Gamma)$. Define the beamforming achievable rate region of a downlink

channel under a power constraint as the set of $(R_1, \ldots, R_K)$ for which there exists a set of beamformers that satisfy the power constraint and have their corresponding SINR's achieving the rates. Theorem 1 implies that the beamforming achievable rate regions of the uplink and downlink are identical, and Corollary 3 implies that the same is true with dirty-paper coding for each fixed encoding order. Further, since the achievable rate region duality holds for every order, the uplink and downlink rate regions under all possible orders, which are the unions of respective rate regions over all $K!$ possible orders must also be the same. Since it is known that with the use of capacity-achieving codes dirty-paper coding and successive interference subtraction achieve the capacity regions of the downlink and uplink channels respectively, as an immediate consequence of Corollary 3, we also have the capacity region duality. Capacity region duality has been previously proved under the sum power constraint in [12], [13], and [15] using a very different proof technique. From the preceding development, it is clear that capacity region duality may be derived naturally from beamforming duality. Further, it is also possible to extend the duality to the per-antenna power constrained case.

Proposition 2: The beamforming achievable rate region of a downlink channel with one antenna at each remote user and with per-antenna power constraints $\{P_1, \ldots, P_N\}$ on the transmit antennas is exactly the same as the beamforming achievable rate region of a dual uplink channel with a sum power constraint $\sum_{i=1}^{N} P_i$ across all the users and with an uncertain noise whose covariance matrix $\sigma^2 \hat{\mathbf{Q}}$ is diagonal and satisfies $\sum_{i=1}^{N} \hat{Q}_{ii} P_i \leq \sum_i P_i$. This is true under any SNR gap Γ. In particular, the capacity regions of the two channels are also the same.

B. Multiple-Receive-Antenna Case

The main motivation for considering the rate region duality rather than the SINR duality is that the concept of rate region allows a generalization of duality to uplink and downlink channels with multiple antennas at the remote users. Toward this end, we first define a beamforming achievable rate region for the multiple receive-antenna channel, then extend the result to the capacity region.

1) Achievable Rate Region Duality: Consider the general multi-antenna downlink channel modeled in (1)

$$\mathbf{y_i} = H_i \mathbf{x} + \mathbf{z_i}, \qquad i = 1, \ldots, K. \qquad (33)$$

We define the *beamforming achievable rate region* of the downlink as follows. In a beamforming design, each receiver employs a set of M beamformers $\mathbf{b_{i,m}}$ to create M independent data streams

$$y_{i,m} = \mathbf{b_{i,m}^H y_i}, \qquad m = 1, \ldots, M \qquad (34)$$

where the beamformers $\mathbf{b_{i,m}}$'s are $M \times 1$ unit-norm vectors. The transmitter employs MK beamformers, denoted as $\mathbf{w_{i,m}}$

$$\mathbf{x} = \sum_{i=1}^{K} \sum_{m=1}^{M} u_{i,m} \mathbf{w_{i,m}} \qquad (35)$$

where $u_{i,m}$ is a scalar denoting the information signal for the ith user and mth data stream. Let $\mathbb{E}[u_{i,m}^2] = 1$.

Without interference subtraction, the SINR for the ith user and mth data stream is

$$\mathrm{SINR}_{i,m} = \frac{\left| \mathbf{b_{i,m}^H H_i w_{i,m}} \right|^2}{\sum_{(j,k) \neq (i,m)} \left| \mathbf{b_{i,m}^H H_i w_{j,k}} \right|^2 + \sigma^2}. \qquad (36)$$

The achievable rate for user i is then

$$R_i = \sum_{m=1}^{M} \log \left(1 + \frac{\mathrm{SINR}_{i,m}}{\Gamma} \right). \qquad (37)$$

The *beamforming achievable rate region* for the downlink is the set of $(R_1, \ldots, R_K)$ satisfying the power constraint.

A similar beamforming achievable rate region may be defined if dirty-paper coding may be implemented. In this case, by fixing a total ordering over the doubly indexed (i, m), the SINR may be defined as

$$\mathrm{SINR}_{i,m} = \frac{\left| \mathbf{b_{i,m}^H H_i w_{i,m}} \right|^2}{\sum_{(j,k) \succ (i,m)} \left| \mathbf{b_{i,m}^H H_i w_{j,k}} \right|^2 + \sigma^2} \qquad (38)$$

where $(j, k) \succ (i, m)$ denotes the condition that either $j > i$, or $j = i$ and $k > m$. Clearly, the achievable rate region with dirty-paper coding is larger than that with beamforming alone. Also, both achievable rate regions depend on Γ. As will be seen later, the dirty-paper region becomes the capacity region when $\Gamma = 0$ dB. We now state the main theorem of this section.

Theorem 2: The beamforming achievable rate region of a downlink channel under a fixed set of per-antenna power constraints $\{P_1, \ldots, P_N\}$ is identical to the achievable rate region of a dual uplink channel with a sum power constraint $\sum_{i=1}^{N} P_i$ across all the users and with an uncertain noise whose covariance matrix $\sigma^2 \hat{\mathbf{Q}}$ is diagonal and satisfies $\sum_{i=1}^{N} \hat{Q}_{ii} P_i \leq \sum_{i=1}^{N} P_i$. This uplink–downlink duality holds either with or without dirty-paper coding and successive interference subtraction. This uplink–downlink duality holds for channels with an arbitrary number of transmit and receive antennas, and for any SNR gap Γ.

Proof: Trivially, if the set of receiver beamformers $\mathbf{b_{i,m}}$ are fixed, then the downlink channel reduces to the single-receive antenna case discussed in the previous section for which each different data stream becomes a virtual user and uplink–downlink duality holds. In other words, under a fixed transmit power constraint and a fixed set of $\mathbf{b_{i,m}}$'s, the sets of achievable $\mathrm{SINR}_{i,m}$'s for the uplink and downlink are the same. Since the SINR is related to rate by (37), the achievable rate regions for both the uplink and downlink channels are also the same. Note that in the dual uplink channel, $\mathbf{b_{i,m}}$ becomes a transmit beamforming vector.

Now, the true achievable rate regions for both the uplink and downlink are the unions of achievable rate regions over all possible $\mathbf{b_{i,m}}$'s (and all possible $(MK)!$ ordering if dirty-paper coding is used.) Therefore, the true beamforming achievable rate regions for the uplink and downlink channels must also be the same. This is true for any arbitrary number of transmit and receive antennas, for any arbitrary gap Γ, and with or without dirty-paper coding. $\square$

$\mathbb{E}|\mathbf{x_i}|^2 \leq P_i, \forall i$ $\qquad$ $\mathbb{E}[\hat{\mathbf{x}}_1^H \hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2^H \hat{\mathbf{x}}_2] \leq \sum_i P_i$ $\qquad$ $\sum_i \hat{\mathbf{Q}}_{ii} P_i \leq \sum_i P_i,$ $\hat{\mathbf{Q}}$ diagonal
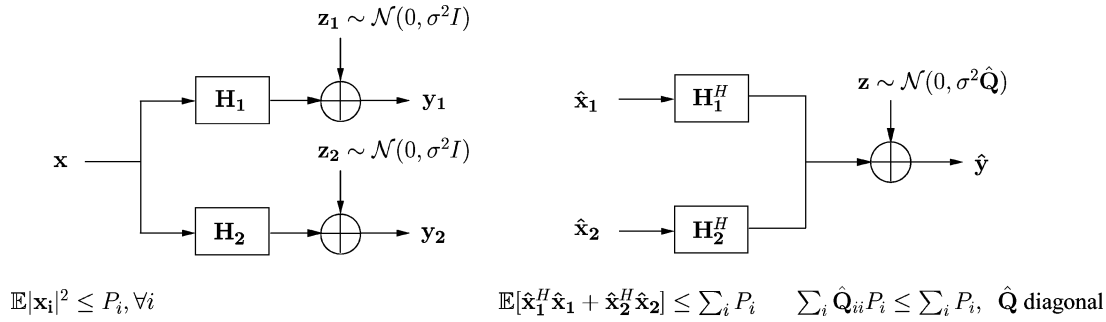
Fig. 2. Achievable rate region duality with per-antenna power constraint.

2) From Achievable Rate Region Duality to Capacity Region Duality: We are now ready to relate the beamforming achievable rate region to the information-theoretical capacity region for a multi-antenna channel. One would expect that with dirty-paper coding and successive decoding and with $\Gamma = 0$ dB, the beamforming achievable rate region would be the capacity region. While such a result appears intuitive (and indeed is true), several steps are needed to prove it rigorously.

Consider first the single-user case $\mathbf{y} = H\mathbf{x} + \mathbf{n}$. The information-theoretical capacity of the vector Gaussian channel is achieved with a water-filling covariance matrix for $\mathbf{x}$. The optimal transmitter is just the set of left-singular vectors of H. The optimal receiver is just the set of right-singular vectors of H. The singular vectors are exactly beamformers with which the vector channel is diagonalized into a set of parallel channels. Clearly, when $\Gamma = 0$ dB, the beamforming strategy is capacity achieving.

However, for a multiuser channel, boundary points of the capacity region are not necessarily achieved with water-filling covariance matrices. In this case, the information theoretically optimal transmission strategy is *not* a diagonalization of the effective channel by each user. Nevertheless, transmit and receive beamforming with interference cancellation and zero gap is still capacity achieving. This is true because of the following. First, any arbitrary transmit covariance matrix can be synthesized using the eigenvectors of the covariance matrix as the transmit beamformers, i.e., $\mathbf{x} = \mathbf{M}\mathbf{u}$, where $\mathbf{M}$ is a matrix of beamforming vectors, and $\mathbf{u}$ contains the information bits. Second, information in $\mathbf{u}$ can be recovered at the receiver via a set of MMSE receive beamformers with interference subtraction. This is because $I(\mathbf{x}; \mathbf{y}) = I(\mathbf{u}; \mathbf{y}) = I(u_1; \mathbf{y}) + I(u_2; \mathbf{y}|u_1) + \cdots + I(u_N; \mathbf{y}|u_1, \ldots, u_{N-1})$, and each of $I(u_k; \mathbf{y}|u_1, \ldots, u_{k-1})$ is achievable with MMSE receiver beamforming and interference subtraction [22, 8.3.4]. Therefore, beamforming with interference subtraction and with scalar zero-gap error correcting codes is capacity achieving. A similar argument can be made for dirty-paper coding. This fact provides a link between information theoretical capacity region and the beamforming region for the downlink channel.

Corollary 4: The capacity region of a downlink channel under a fixed set of per-antenna power constraints $\{P_1, \ldots, P_N\}$ is identical to the capacity region of a dual uplink channel with a sum power constraint $\sum_{i=1}^N P_i$ across all the users and with an uncertain noise whose covariance matrix $\sigma^2 \hat{\mathbf{Q}}$ is diagonal and satisfies $\sum_{i=1}^N \hat{\mathbf{Q}}_{ii} P_i \leq \sum_{i=1}^N P_i$. This

capacity region duality holds for channels with an arbitrary number of transmit and receive antennas.

Proof: Based on results in information theory, it is known that the capacity region of a Gaussian multi-antenna multiple-access channel can be achieved with successive interference cancellation and that the capacity region of a Gaussian multi-antenna broadcast channel can be achieved using dirty-paper coding [14]. Each successive cancellation or dirty-paper coding step involves a Gaussian vector channel with colored noise. Now, as mentioned earlier, the vector channel capacity can be achieved with beamforming and scalar successive-decoding or dirty-paper-coding with zero-gap codes. Thus, the capacity region coincides with the beamforming achievable region with zero gap. Since duality holds for the beamforming achievable region by Theorem 2, it must also hold for the capacity region. □

Fig. 2 illustrates the achievable rate region duality between the uplink and downlink channels. Corollary 4 states that the capacity region for the broadcast channel can be computed by solving the following optimization problem corresponding to the dual uplink channel. Let $\mathbf{S}_i$ be the transmit covariance matrix for user i in the dual uplink; let $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_K \geq 0$ be weights characterizing different boundary points of the capacity region. The dual problem is

$$\min_{\hat{\mathbf{Q}}} \max_{\mathbf{S}_i} \quad \sum_{k=1}^{K} \mu_k \log \frac{\left| \sum_{i=1}^{k} \mathbf{H}_i^H \mathbf{S}_i \mathbf{H}_i + \sigma^2 \hat{\mathbf{Q}} \right|}{\left| \sum_{i=1}^{k-1} \mathbf{H}_i^H \mathbf{S}_i \mathbf{H}_i + \sigma^2 \hat{\mathbf{Q}} \right|} \tag{39}$$

$$\text{subject to} \quad \sum_{i=1}^{K} \operatorname{tr}(\mathbf{S}_i) \leq \operatorname{tr}(\boldsymbol{\Phi}) \tag{40}$$

$$\operatorname{tr}(\hat{\mathbf{Q}}\boldsymbol{\Phi}) \leq \operatorname{tr}(\boldsymbol{\Phi}) \tag{41}$$

$$\mathbf{S}_i \succeq 0, \quad \hat{\mathbf{Q}} \succeq 0, \quad \hat{\mathbf{Q}} \text{ diagonal} \tag{42}$$

where $\boldsymbol{\Phi} = \operatorname{diag}(P_1, \ldots, P_N)$. This dual problem is concave in $\mathbf{S}_i$'s and convex in $\hat{\mathbf{Q}}$, a fact which significantly simplifies the original downlink capacity region computation problem.

The capacity region duality under a sum-power constraint has been previously stated in [12], [13], [15]. The proof technique in the previous work (i.e., [13]) relies on a so-called "channel-flipping formula," which, for each set of uplink transmit covariance matrices, finds a set of corresponding downlink covariance matrices, and vice versa. Theorem 2 of this paper shows that duality is more general, as it applies to the $\Gamma \neq 0$ dB case and to the per-antenna power constraints as well.

When specialized to the $\Gamma = 0$ dB case, Theorem 2 and Corollary 4 also give a new derivation of the channel-flipping formula. Each set of transmit covariance matrices (in either uplink or downlink) defines a set of transmit and receive beamforming vectors. The correspondence between uplink and downlink beamformers immediately gives the corresponding covariance-matrix transformation between the uplink and the downlink.

Theorem 2 also holds if we restrict the transmission for each user to include only a fixed number of data streams. For example, each receiver in the downstream may be restricted to use only a single beamformer. In this case, duality still holds. Duality in this special case has been observed for the beamforming case in [23] and for the dirty-paper coding case in [24], and it leads to iterative methods of updating the transmit and receive downlink beamformers. Although such an iterative algorithm may not necessarily converge to the global optimum, it provides an efficient way of obtaining a set of locally optimal transmit and receive beamformers.

IV. GENERALIZED PER-ANTENNA POWER CONSTRAINT

A. Power Constraint Per Groups of Antennas

The results presented so far focus on a downlink channel with a power constraint on each antenna. As an extension, this section considers the case in which transmit antennas are partitioned into groups and a power constraint is imposed on a per-group basis. A duality for this more general case can be similarly derived.

Corollary 5: Let $\mathcal{N}_1, \ldots, \mathcal{N}_L$ be a set partition of $\{1, \ldots, N\}$. Suppose that a power constraint P_l is applied to each group of antennas $\mathcal{N}_l$, i.e., for each $l = 1, \ldots, L$

$$\sum_{i \in \mathcal{N}_l} \left[\sum_{j=1}^{K} \sum_{m=1}^{M} \mathbf{w}_{\mathbf{j},\mathbf{m}} \mathbf{w}_{\mathbf{j},\mathbf{m}}^H \right]_{i,i} \leq P_l. \tag{43}$$

Then, the dual of this downlink problem is an uplink channel with an uncertain noise. In this case, the noise covariance is a diagonal matrix of the form $\hat{\mathbf{Q}} = \mathrm{diag}(q_1, \ldots, q_1, q_2, \ldots, q_2, \ldots, q_L, \ldots, q_L)$, in which each q_l is repeated exactly $|\mathcal{N}_l|$ times, and the noise constraint is $\sum_{l=1}^{L} q_l P_l \leq \sum_{l=1}^{L} P_l$. This duality holds for both the beamforming case under the same set of SINR constraints (with or without dirty-paper coding) and for the achievable rate region case in which a sum power constraint $\sum_{l=1}^{L} P_l$ is applied to the dual uplink channel.

The proof of this result is a straightforward generalization of previous proofs. When $L = N$, we recover the per-antenna power constrained case; when $L = 1$, we recover the sum-power constrained duality.

B. Duality for Frequency-Selective Channels

We now extend the duality result to frequency-selective downlink channels with per-antenna power constraints. Assume a multi-input, multi-output orthogonal frequency-division multiplex (MIMO-OFDM) system which diagonalizes the channel by dividing the frequency band into F parallel independent MIMO subchannels. Let $\mathbf{H_i}(f)$ denote the ith user's

channel in frequency band f, where $f = 1, \ldots, F$. A separate beamformer is assigned for each frequency subchannel, for each user and for each data stream at the transmitter, i.e.,

$$\mathbf{x}(f) = \sum_{i=1}^{K} \sum_{m=1}^{M} u_{i,m}(f) \mathbf{w}_{\mathbf{i},\mathbf{m}}(f) \tag{44}$$

The receiver also uses a separate beamformer for each data stream and for each frequency subchannel. The power constraint is applied to each transmit antenna over the entire frequency band, i.e.,

$$\sum_{f=1}^{F} \left[\mathbf{x}(f) \mathbf{x}(f)^H \right]_{i,i} \leq P_i \qquad i = 1, \ldots, N. \tag{45}$$

The extension of the duality theorem to this case is again relatively straightforward. We state the following result without a proof.

Corollary 6: The dual of a downlink frequency-selective channel with a per-antenna power constraint (45) is an uplink frequency-selective channel with a sum-power constraint $\sum_{i=1}^{N} P_i$, in which the noise covariance $\sigma^2 \hat{\mathbf{Q}}$ on each frequency subchannel is identical and is constrained to be in the convex set: $\sum_{i=1}^{N} \hat{\mathbf{Q}}_{ii} P_i \leq \sum_{i=1}^{N} P_i$. This duality holds for both the beamforming case under the same set of SINR constraints (with or without dirty-paper coding) and for the achievable rate region case.

It is interesting to note that the structure of this per-antenna duality relation is markedly different from previous studies on the capacity region duality of broadcast and multiple-access fading channels under the sum power constraint [12].

V. NUMERICAL ALGORITHMS

The main motivation for establishing the duality between the uplink and the downlink is that the uplink input optimization problem is more amenable to numerical computation in many instances. In this section, we provide numerical algorithms to solve the downlink problem for two such cases.

First, we consider the downlink beamforming problem with a single antenna at each remote user. Under a sum power constraint, the dual uplink beamforming problem in this case can be solved using algorithms that update the power allocation and beamforming vectors iteratively [1]–[4]. The efficiency of such updates has been further improved in [8]. The optimal beamforming problem is even simpler when dirty-paper coding is used [5]. Although, in theory, the downlink beamforming problem can also be transformed into a semidefinite programming problem, direct iterative updates are in general more efficient.

Second, we consider the computation of the downlink capacity region for the multi-receive-antenna case. Although the achievable rate region for both the uplink and downlink problems is in general difficult to solve when interference subtraction is not used or when the SNR gap $\Gamma \neq 0$ dB, when the SNR gap $\Gamma = 0$ dB, the uplink capacity region problem is tractable. The weighted sum rate expression in this case is a convex function of transmit covariance matrices. This leads to an efficient solution to the downlink capacity region problem as well.

A. Iterative Algorithm for Downlink Beamforming

We first focus on the optimal per-antenna power constrained transmit beamforming problem for a downlink channel with one antenna at each remote user. In this case, Theorem 1 allows us to solve this downlink problem (11) by solving a dual uplink problem with an uncertain noise (16). This dual problem may be solved by iteratively computing the inner minimization on $(\lambda_i, \hat{\mathbf{w}}_\mathbf{i})$ and the outer maximization on $\mathbf{Q}$. Consider the following subproblem of of (16) with a fixed $\mathbf{Q}$

$$f(\mathbf{Q}) = \min_{\lambda_i, \hat{\mathbf{w}}_\mathbf{i}} \quad \sum_{i=1}^{K} \lambda_i \sigma^2$$
$$\text{s.t.} \quad \frac{\lambda_i \sigma^2 |\hat{\mathbf{w}}_\mathbf{i}^H \mathbf{h_i}|^2}{\sum_{j \neq i} \lambda_j \sigma^2 |\hat{\mathbf{w}}_\mathbf{i}^H \mathbf{h_j}|^2 + \hat{\mathbf{w}}_\mathbf{i}^H \sigma^2 \mathbf{Q} \hat{\mathbf{w}}_\mathbf{i}} \geq \gamma_i. \quad (46)$$

The main idea is that $f(\mathbf{Q})$ can be easily computed via a fixed-point algorithm previously proposed for uplink beamforming with a fixed noise covariance (e.g., using the method of [8] in Steps 1 and 2 shown below.) Then, the outer maximization of $f(\mathbf{Q})$ may be solved via a subgradient projection algorithm, where the subgradient may be found using the downlink beamformers obtained from the optimal uplink beamforming vectors.

Proposition 3: The function $f(\mathbf{Q})$ is concave in $\mathbf{Q}$. Further, $\text{diag}\{\sum_{i=1}^{K} \mathbf{w_i} \mathbf{w_i}^H\}$ is a subgradient of $f(\mathbf{Q})$, where $\mathbf{w_i}$ is the optimal downlink beamforming vector, which can be found using the algorithm in Corollary 2.

The proof of the proposition is provided in Appendix C. The use of $\text{diag}\{\sum_{i=1}^{K} \mathbf{w_i} \mathbf{w_i}^H\}$ as a subgradient for $\mathbf{Q}$ is intuitive, because $\mathbf{Q}$ is the Lagrangian multiplier for the per-antenna power constraints in the downlink. We propose the following algorithm based on the Euclidean projection $\mathcal{P}_{\mathcal{S}_\mathbf{Q}}$ of the subgradient of $f(\mathbf{Q})$ on the constraint set $\mathcal{S}_\mathbf{Q} = \{\mathbf{Q} : \text{tr}(\mathbf{\Phi} \mathbf{Q}) \leq \text{tr}(\mathbf{\Phi}), \mathbf{Q} \succeq 0\}$. This subgradient projection method is guaranteed to converge to the global optimum of $f(\mathbf{Q})$, since $f(\mathbf{Q})$ is concave [25]. The proposed algorithm is summarized as follows.

   0) Initialize $\mathbf{Q}^{(0)}$. Set $n = 0$,
   1) Fix $\mathbf{Q}^{(n)}$. Solve the subproblem $f(\mathbf{Q}^{(n)})$ by first finding the fixed-point λ_i^* of the following equation by iterative function evaluation [8]:

$$\lambda_i^* = \frac{1}{\left(1 + \frac{1}{\gamma_i}\right) \mathbf{h_i}^H \left(\sum_{j=1}^{K} \lambda_j^* \mathbf{h_j} \mathbf{h_j}^H + \mathbf{Q}^{(n)}\right)^\dagger \mathbf{h_i}}.$$

   2) Find the optimal uplink beamformers based on the optimal uplink power allocation λ_i^*

$$\hat{\mathbf{w}}_\mathbf{i} = \left(\sum_{j=1}^{K} \lambda_j^* \mathbf{h_j} \mathbf{h_j}^H + \mathbf{Q}^{(n)}\right)^\dagger \mathbf{h_i}.$$

   3) Update the downlink beamformers by $\mathbf{w_i} = \sqrt{\delta_i} \hat{\mathbf{w}}_\mathbf{i}$ where $\delta = \mathbf{G}^\dagger \mathbf{1} \sigma^2$ as defined in Corollary 2.
   4) Update $\mathbf{Q}^{(n)}$ using the subgradient projection method with step size t_n

$$\mathbf{Q}^{(n+1)} = \mathcal{P}_{\mathcal{S}_\mathbf{Q}} \left\{\mathbf{Q}^{(n)} + t_n \text{diag}\left\{\sum_{i=1}^{K} \mathbf{w_i} \mathbf{w_i}^H\right\}\right\}.$$

   5) Set $n = n + 1$ and return to Step 1) until convergence.

B. Interior-Point Algorithm for Downlink Beamforming

In this section, we develop an alternative interior-point algorithm for the optimal downlink beamforming problem (11). The idea is to solve its Lagrangian dual problem (20) directly using an interior-point method. Note that the first set of constraints in (20) is a linear matrix inequality constraint, which can be handled by introducing a logarithmic barrier [16]

$$\mathbf{\Psi}(\mathbf{Q}, \lambda_i) = \log |\mathbf{Q}| + \sum_{i=1}^{K} \log \lambda_i$$
$$+ \sum_{i=1}^{K} \log \left| \mathbf{Q} + \sum_{j=1}^{K} \lambda_j \mathbf{h_j} \mathbf{h_j}^H - \left(1 + \frac{1}{\gamma_i}\right) \lambda_i \mathbf{h_i} \mathbf{h_i}^H \right|.$$

The second set of constraints are equality constraints, which can be handled directly. Thus, the new optimization problem becomes

$$\max_{\mathbf{Q}, \lambda_i} \quad \sum_{i=1}^{K} \lambda_i \sigma^2 + \frac{1}{t} \mathbf{\Psi}(\mathbf{Q}, \lambda_i)$$
$$\text{s.t.} \quad \text{tr}(\mathbf{Q} \mathbf{\Phi}) \leq \text{tr}(\mathbf{\Phi}), \quad \mathbf{Q} \text{ diagonal}, \quad \mathbf{Q} \succeq 0. \quad (47)$$

where t is the magnitude of the barrier. It can be shown that the above optimization problem is concave in $\mathbf{Q}$ and λ_i, since $\log |\cdot|$ function is concave on the set of positive semidefinite matrices. The algorithm and the convergence analysis are standard.

C. Interior-Point Algorithm for Capacity Region Computation

In general, the achievable rate region maximization problem for a downlink channel with multiple receive antennas is a difficult problem. However, when the SNR gap $\Gamma = 0$ dB, finding the capacity region for the downlink is tractable. In particular, as shown in Corollary 4, the capacity region of a multi-antenna downlink with per-antenna power constraints $P_1, \ldots, P_N$ is the same as that of the dual uplink, which is a multiple-access channel with a capacity expression (39) that is concave in $\mathbf{S}_i$ and convex in $\hat{\mathbf{Q}}$.

One possible way of solving a minimax problem is to iterate between the minimization and the maximization. However, convergence is difficult to guarantee for such an algorithm. In this section, we propose the use of a novel infeasible-start Newton's method [16] that solves the maximization and the minimization at the same time. The key ingredient here is a novel stopping criterion that allows the iterative algorithm to approach the saddle point of the minimax problem directly.

As a first step, write $\mathbf{S}_i$'s and $\sigma^2 \hat{\mathbf{Q}}$ as vectors, and call them $\mathbf{s}_x^{(v)}$ and $\mathbf{s}_z^{(v)}$, respectively. This is automatic for $\mathbf{s}_z^{(v)}$, which is already diagonal. For an $M \times M$ matrix $\mathbf{S}_i$, we pack the upper triangular $M(M + 1)/2$ entries into the vector $\mathbf{s}_x^{(v)}$. Recognizing that the inequality constraints (40) are always satisfied with equality, i.e., $\sum_{i=1}^{K} \text{tr}(\mathbf{S}_i) = \text{tr}(\mathbf{\Phi})$ and $\text{tr}(\hat{\mathbf{Q}} \mathbf{\Phi}) = \text{tr}(\mathbf{\Phi})$, solving (39) is now equivalent to solving

$$\max_{\mathbf{s}_x^{(v)}} \min_{\mathbf{s}_z^{(v)}} \quad f\left(\mathbf{s}_x^{(v)}, \mathbf{s}_z^{(v)}\right)$$
$$\text{subject to} \quad \mathbf{A} \mathbf{s}_x^{(v)} = \mathbf{1}$$
$$\mathbf{B} \mathbf{s}_z^{(v)} = \mathbf{1} \quad (48)$$

where the positive semidefinite constraints are omitted for now, f denotes the objective function in (39) and matrices $\mathbf{A}$ and $\mathbf{B}$ represent the linear trace constraints (40) and (41).

Our algorithm is based on solving the KKT condition for the minimax optimization (39) as follows:

$$r_1 = \nabla_x f\left(\mathbf{s}_x^{(v)}, \mathbf{s}_z^{(v)}\right) + \mathbf{A}^T \nu = 0 \tag{49}$$

$$r_2 = \nabla_z f\left(\mathbf{s}_x^{(v)}, \mathbf{s}_z^{(v)}\right) + \mathbf{B}^T \mu = 0 \tag{50}$$

$$r_3 = \mathbf{A}\mathbf{s}_x^{(v)} - \mathbf{1} = 0 \tag{51}$$

$$r_4 = \mathbf{B}\mathbf{s}_z^{(v)} - \mathbf{1} = 0 \tag{52}$$

where ν and μ are Lagrangian multipliers associated with the equality constraints $\mathbf{A}\mathbf{s}_x^{(v)} = \mathbf{1}$ and $\mathbf{B}\mathbf{s}_z^{(v)} = \mathbf{1}$, respectively. For notational convenience, ∇_x is used to denote $\nabla_{\mathbf{s}_x^{(v)}}$, and likewise for ∇_z. Since the minimax optimization problem (39) is convex–concave, the KKT condition is sufficient and necessary for optimality. We define the residual $\mathbf{r} = \left[r_1^T, r_2^T, r_3^T, r_4^T\right]^T$ in a vector form. Thus, a solution $(\mathbf{s}_x^{(v)}, \mathbf{s}_z^{(v)}, \nu, \mu)$ is a global optimum if and only if the residual $\mathbf{r}(\mathbf{s}_x^{(v)}, \mathbf{s}_z^{(v)}, \nu, \mu) = 0$. This observation enables us to solve for the root of the residual directly. Toward this end, we approximate $\mathbf{r}(\mathbf{s}_x^{(v)}, \mathbf{s}_z^{(v)}, \nu, \mu)$ as a linear function using its gradient, and solve the equation $\mathbf{r} = 0$ as if it is linear. The update of $(\Delta\mathbf{s}_x^{(v)}, \Delta\mathbf{s}_z^{(v)}, \Delta\nu, \Delta\mu)$ can then be stated as

$$\begin{bmatrix} \Delta\mathbf{s}_x^{(v)} & \Delta\mathbf{s}_z^{(v)} & \Delta\nu & \Delta\mu \end{bmatrix}^T = -(\nabla\mathbf{r})^{-1}\mathbf{r}. \tag{53}$$

More explicitly, the above equation can be written as

$$\begin{bmatrix} \nabla_{xx}^2 f & \nabla_{xz}^2 f & \mathbf{A}^T & 0 \\ \nabla_{zx}^2 f & \nabla_{zz}^2 f & 0 & \mathbf{B}^T \\ \mathbf{A} & 0 & 0 & 0 \\ 0 & \mathbf{B} & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta\mathbf{s}_x^{(v)} \\ \Delta\mathbf{s}_z^{(v)} \\ \Delta\nu \\ \Delta\mu \end{bmatrix} = \begin{bmatrix} -\nabla_x f - \mathbf{A}^T\nu \\ -\nabla_z f - \mathbf{B}^T\mu \\ \mathbf{1} - \mathbf{A}\mathbf{s}_x^{(v)} \\ \mathbf{1} - \mathbf{B}\mathbf{s}_z^{(v)} \end{bmatrix} \tag{54}$$

where the square matrix above is known as the KKT matrix. The search direction $(\Delta\mathbf{s}_x^{(v)}, \Delta\mathbf{s}_z^{(v)}, \Delta\nu, \Delta\mu)$ is found via a matrix inversion.

The search direction derived above is actually a Newton's direction for both the minimization and the maximization at the same time. This observation is made in [16]. One way to interpret the search direction defined by (53) is that the minimax problem is being approximated by a quadratic minimax problem at each step, and the Newton's step represents a direction toward the saddle-point of the quadratic approximation. Note that unlike the conventional Newton's method, where the value of the objective function can be used to ensure that the behavior of the algorithm is monotonic, for minimax problems, a different metric is needed. The natural metric in our case is the norm of the residual, which can be used in the backtrack line search and as a stopping criterion.

The derivation so far assumes that the optimal $\mathbf{S}_i$'s and $\hat{\mathbf{Q}}$ are strictly positive definite. In general, the positivity constraints need to be taken into account in the interior-point method [16]

via a logarithmic barrier. More specifically, let $\Psi(\mathbf{S}_i, \hat{\mathbf{Q}}) = \sum_{i=1}^K \log|\mathbf{S}_i| - \log|\sigma^2\hat{\mathbf{Q}}|$ be the barrier function and define

$$f_t = f + \frac{1}{t}\Psi \tag{55}$$

then the earlier derivation follows with f replaced by f_t. The optimal solution for f_t will approach the optimal solution for f with a gap $(KM^2 + N)/t$.

We now summarize the interior-point algorithm for the minimax problem. The algorithm is guaranteed to converge to the global optimum. A detailed convergence analysis follows from the analysis in [16].

> 0) Initialize $\mathbf{S}_i^{(0)}$, $\hat{\mathbf{Q}}^{(0)}$, $\nu^{(0)}$, $\mu^{(0)}$, and $t > 0$.
> 1) Compute the search direction $(\Delta\mathbf{S}_i, \Delta\hat{\mathbf{Q}}, \Delta\nu, \Delta\mu)$ using (53) based on the current $\mathbf{S}_i^{(n)}$, $\hat{\mathbf{Q}}^{(n)}$, $\nu^{(n)}$, and $\mu^{(n)}$.
> 2) Set $(\mathbf{S}_i^{(n+1)}, \hat{\mathbf{Q}}^{(n+1)}, \nu^{(n+1)}, \mu^{(n+1)}) = (\mathbf{S}_i^{(n)} + s\Delta\mathbf{S}_i, \hat{\mathbf{Q}}^{(n)} + s\Delta\hat{\mathbf{Q}}, \nu^{(n)} + s\Delta\nu, \mu^{(n)} + s\Delta\mu)$. Use backtracking line search with parameter α to find the scalar s so that
> $$\left\|\mathbf{r}\left(\mathbf{S}_i^{(n+1)}, \hat{\mathbf{Q}}^{(n+1)}, \nu^{(n+1)}, \mu^{(n+1)}\right)\right\|_2$$
> $$< (1 - \alpha s)\left\|\mathbf{r}\left(\mathbf{S}_i^{(n)}, \hat{\mathbf{Q}}^{(n)}, \nu^{(n)}, \mu^{(n)}\right)\right\|_2.$$
> 3) If $\|\mathbf{r}\|_2 < \epsilon$, increase the magnitude of the barrier function by scaling t with a constant factor $\gamma > 1$.
> 4) Stop if the gap $(KM^2 + N)/t$ is less than the tolerance ϵ, or equivalently, if t is sufficiently large. Go to Step 1 otherwise.

VI. SIMULATIONS

In the first set of simulations, we justify the use of the per-antenna power constraint. Consider a downlink channel with $N = 3$ transmit antennas and $K = 10$ receivers each equipped with $M = 1$ antenna. Suppose that each transmit antenna is capable of transmitting a maximum power of 1/3 W. The channel matrix is drawn from an i.i.d. Gaussian distribution with mean 0 and variance 1. Using the uplink–downlink duality results derived in this paper, we can easily solve for the sum capacity in such a downlink channel under either the per-antenna power constraints $P_i \leq 1/3$ or a sum power constraint $\sum_i P_i \leq 1$. Fig. 3 illustrates the optimal transmitting power on one of the antennas for 100 different channel realizations under each type of constraints. Under a sum power constraint, the optimal transmit power fluctuates widely, and may be as high as 0.7 W. If one uses a naive strategy of transmitting independent symbols on each antenna with equal power, the resulting data rate would be significantly lower than the optimized strategy.[3]

Next, we present simulation results for the convergence behavior of the numerical algorithms proposed in Section V. First, we compare the performance of the iterative algorithm and the interior-point algorithm for the optimal downlink beamforming problem proposed in Sections V-A and V-B. The same downlink channel as in the previous simulation is used, but with $K = 10$

---

[3]It is possible to derive heuristics that work well. For example, it is possible to solve the sum-capacity optimization problem first, then scale the resulting transmit power and their correlations to obtain an approximate solution.
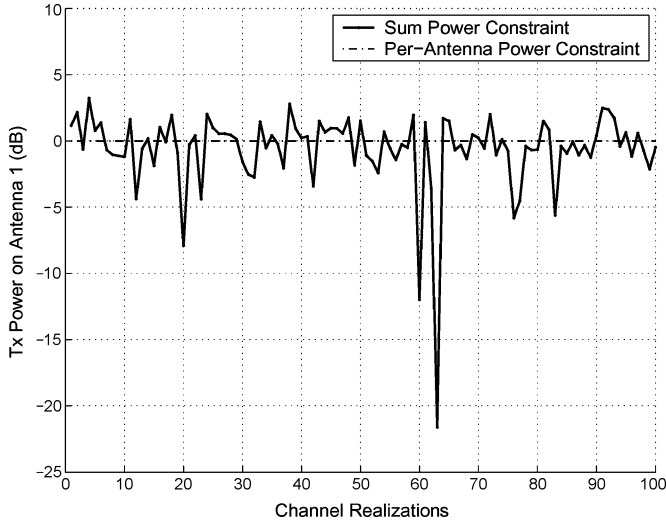
Fig. 3. Optimal transmit power on the first antenna for a downlink channel with three transmit antennas and with a sum power constraint across the three antennas over 100 channel realizations. The transmitting power is normalized so that 0 dB corresponds to 1/3.
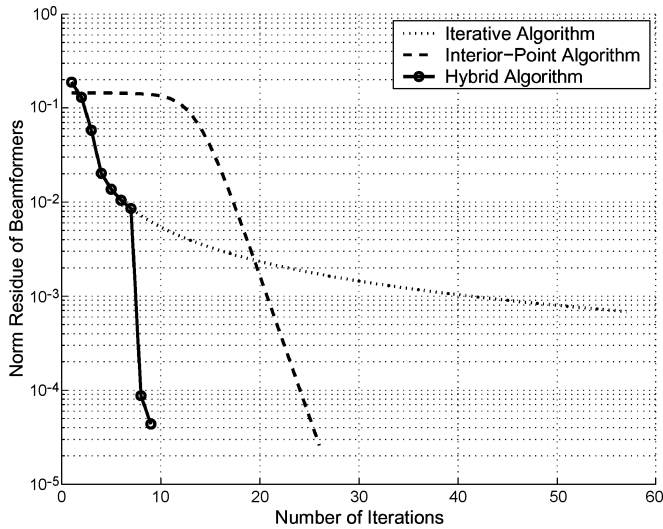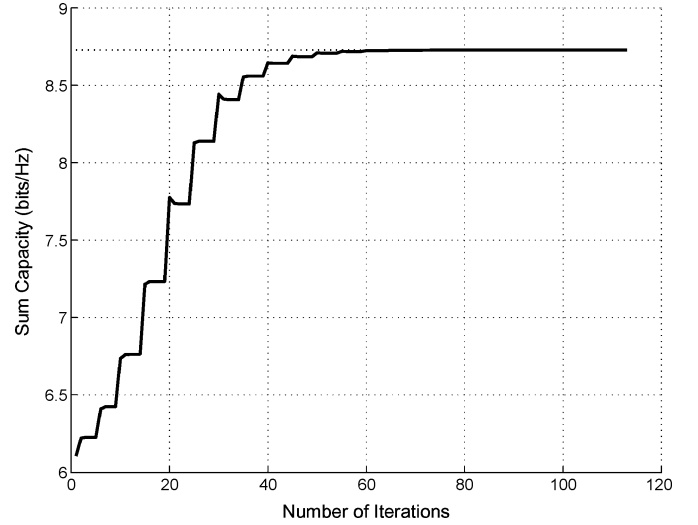

Fig. 5. Convergence behavior of the interior-point algorithm for capacity region computation for a downlink channel with $N = 5$ transmit antennas, $K = 20$ users, and $M = 2$ receive antennas for each user. The sum capacity versus the number of iterations is plotted.

Fig. 4. Convergence behavior of the iterative algorithm, the interior-point algorithm, and a hybrid algorithm for the optimal downlink beamforming problem with ten transmit antennas and ten users with one receive antenna each. Here, $\sum_i \|\mathbf{w}_i - \mathbf{w}_i^*\|_2$ versus the number of iterations is plotted.

then switches to the interior-point algorithm to obtain a high numerical accuracy. The performance of this hybrid method is also shown in Fig. 4. The hybrid algorithm is observed to be very efficient.

Finally, we illustrate the convergence of the interior-point algorithm presented in Section V-C for solving the downlink channel capacity region problem. In this scenario, a base station with $N = 5$ transmit antennas and $K = 20$ users with $M = 2$ antennas each is simulated. The tolerance is set to be $\epsilon = 10^{-6}$. The interior-point method parameter γ is set to 5, and backtracking line search parameters are set to be $\alpha = 1/3$, and $\beta = 2/3$. The sum capacity versus the number of iterations is plotted in Fig. 5. The algorithm is numerically well behaved.

VII. CONCLUSION

This paper illustrates an uplink–downlink duality for the multi-antenna downlink channel under the per-antenna power constraint. For the downlink beamforming problem under a fixed SINR constraint, we show that the dual of the downlink is an uplink channel with an uncertain noise. The same duality relation also holds between the achievable rate regions of the respective uplink and downlink channels. These duality results are based on an interpretation of uplink–downlink duality via Lagrangian duality. Uplink–downlink duality is useful because the downlink optimization problem can be solved more effectively in the dual uplink domain.

APPENDIX

A. Proof of Proposition 1

We prove in this appendix that strong duality holds for the downlink beamforming problem (11). The main idea is to use the technique in [8] to transform the problem into a second-order cone programing problem.

First, observe that an arbitrary phase rotation can be added to the beamforming vectors without affecting the SINR. Thus, if

users and $N = 10$ transmit antennas. The SINR targets are set to be $\gamma_i = 1$ for all receivers. The iterative algorithm uses a square summable step size $t_n = 1/n$ for the subgradient update. The interior-point algorithm uses a logarithmic barrier function. In Fig. 4, the Euclidean norm distance between the optimal beamformer and the beamformer at the nth iteration is plotted against the number of iterations. As illustrated in Fig. 4, the iterative algorithm is more efficient at the beginning, whereas the interior-point algorithm performs better as the power vector approaches the optimum.[4] We note that it is also possible to use a hybrid algorithm that first performs the iterative algorithm for a fixed number of iterations to obtain an approximate solution,

---

[4]The comparison is not entirely fair since the complexity per iteration for the two algorithms is different. However, the comparison between rates of convergence is still meaningful.

$\{\mathbf{w}_k\}$ is optimal, then so is $\{\mathbf{w}_k e^{j\phi_k}\}$. Without loss of generality, we may then choose $\{\phi_k\}$ so that $\mathbf{h}_k^H \mathbf{w}_k$ is real.

Next, let $\mathbf{W} = [\mathbf{w_1}, \ldots, \mathbf{w_K}]$ be a matrix of beamforming vectors. The SINR constraints can be written as

$$\left(1 + \frac{1}{\gamma_i}\right) |\mathbf{w_i}^H \mathbf{h_i}|^2 \geq \left\|\begin{matrix} \mathbf{h_i}^H \mathbf{W} \\ \sigma \end{matrix}\right\|^2, \qquad \forall i \tag{56}$$

where $\|\cdot\|$ denotes the $\mathcal{L}_2$ Euclidean vector norm. Because $\mathbf{h}_k^H \mathbf{w}_k$ can be assumed to be real, we may take square root of the above equation, the constraint becomes a second-order cone programming constraint [16], which is convex. Now, the downlink beamforming problem (11) may be rewritten as

$$\min_{\alpha, \mathbf{w_i}} \quad \alpha \sum_{i=1}^{N} P_i \tag{57}$$

$$\text{s.t.} \quad \left[\sum_{j=1}^{K} \mathbf{w_j} \mathbf{w_j}^H\right]_{i,i} \leq \alpha P_i, \qquad \forall i \tag{58}$$

$$\sqrt{1 + \frac{1}{\gamma_i}} \mathbf{w_i}^H \mathbf{h_i} \geq \left\|\begin{matrix} \mathbf{h_i}^H \mathbf{W} \\ \sigma \end{matrix}\right\|, \qquad \forall i. \tag{59}$$

Since both the optimization objective and constraints are convex, strong duality holds.

In addition, we also prove in this appendix that the Lagrangian dual of the convex form of the problem (57) is the same as the Lagrangian dual of the original form of the problem (11).[5]

The Lagrangian of (57) is given by

$$L(\alpha, \mathbf{w_i}, q_i, \mu_i) = \alpha \sum_{i=1}^{N} P_i + \sum_{i=1}^{N} q_i \left\{\left[\sum_{j=1}^{K} \mathbf{w_j} \mathbf{w_j}^H\right]_{i,i} - \alpha P_i\right\}$$
$$- \sum_{i=1}^{K} \mu_i \left\{\sqrt{1 + \frac{1}{\gamma_i}} \mathbf{w_i}^H \mathbf{h_i} - \left\|\begin{matrix} \mathbf{h_i}^H \mathbf{W} \\ \sigma \end{matrix}\right\|\right\}. \tag{60}$$

Let $\mathbf{Q} = \text{diag}(q_1, \ldots, q_N)$. The dual objective is therefore

$$g(\mathbf{Q}, \mu_i) = \min_{\mathbf{w_i}} \min_{\alpha} L(\alpha, \mathbf{w_i}, \mathbf{Q}, \mu_i). \tag{61}$$

Since strong duality holds, $g(\mathbf{Q}, \mu_i)$ maximized over $\mathbf{Q}$ and μ_i reaches a maximum at the optimal value of the primal problem (57). To compute $g(\mathbf{Q}, \mu_i)$, let

$$t_i = \sqrt{1 + \frac{1}{\gamma_i}} \mathbf{w_i}^H \mathbf{h_i} + \left\|\begin{matrix} \mathbf{h_i}^H \mathbf{W} \\ \sigma \end{matrix}\right\|. \tag{62}$$

[5]In general, an optimization problem may be written in several equivalent ways, each of which may lead to a different dual.

Then, the last term in (60) can be rewritten as

$$\mu_i \left\{\sqrt{1 + \frac{1}{\gamma_i}} \mathbf{w_i}^H \mathbf{h_i} - \left\|\begin{matrix} \mathbf{h_i}^H \mathbf{W} \\ \sigma \end{matrix}\right\|\right\}$$
$$= \frac{\mu_i}{t_i} \left\{\left(1 + \frac{1}{\gamma_i}\right)(\mathbf{w_i}^H \mathbf{h_i})^2 - \left\|\begin{matrix} \mathbf{h_i}^H \mathbf{W} \\ \sigma \end{matrix}\right\|^2\right\}$$
$$= \frac{\mu_i}{t_i} \left\{\left(1 + \frac{1}{\gamma_i}\right)(\mathbf{w_i}^H \mathbf{h_i})^2 - \left[\mathbf{h_i}^H \mathbf{W}\mathbf{W}^H \mathbf{h_i} + \sigma^2\right]\right\}$$
$$= \frac{\mu_i}{t_i} \left\{\left(1 + \frac{1}{\gamma_i}\right)(\mathbf{w_i}^H \mathbf{h_i})^2 - \mathbf{h_i}^H \left[\sum_{j=1}^{K} \mathbf{w_j}\mathbf{w_j}^H\right]\mathbf{h_i} - \sigma^2\right\}.$$

Substituting this into the Lagrangian, we obtain

$$L = \sum_{i=1}^{K} \frac{\mu_i}{t_i}\sigma^2 - \alpha \{\text{tr}(\mathbf{Q}\boldsymbol{\Phi}) - \text{tr}(\boldsymbol{\Phi})\}$$
$$+ \sum_{i=1}^{K} \mathbf{w_i}^H \left\{\mathbf{Q} + \sum_{j\neq i} \frac{\mu_j}{t_j}\mathbf{h_j}\mathbf{h_j}^H - \frac{\mu_i}{t_i\gamma_i}\mathbf{h_i}\mathbf{h_i}^H\right\}\mathbf{w_i} \tag{63}$$

where $\boldsymbol{\Phi} = \text{diag}(P_1, \ldots P_N)$. Note that t is lower bounded by σ and is strictly positive. Since the only constraint for maximization on μ_i is that $\mu_i\alpha \geq 0$, we can change the optimization variable to $\lambda_i = \mu_i/t_i$. Under this change of variable, the Lagrangian of the convex form of the optimization problem (57) is then exactly the same as the Lagrangian of the original form of the optimization problem (11) [as expressed later in (18)]. Therefore, the dual problems of the two equivalent forms [i.e., (57) and (11)] must also be the same. [A derivation of the Lagrangian (18) can be found in Section II-A. The dual problem is shown in (20).] $\quad\square$

B. Proof of Lemma 1

Lemma 1: Let $\mathbf{A}$ be an $n \times n$ symmetric positive semidefinite matrix and $\mathbf{b}$ be an $n \times 1$ vector. Then, $\mathbf{A} \succeq \mathbf{b}\mathbf{b}^H$ if and only if $\mathbf{b}^H \mathbf{A}^\dagger \mathbf{b} \leq 1$.

Proof: We first prove the Lemma with the additional assumption that $\mathbf{A}$ is strictly positive definite. This assumption will be removed later. First, assume $\mathbf{A} \succeq \mathbf{b}\mathbf{b}^H$. Using the property of matrix determinant, we obtain

$$1 - \mathbf{b}^H \mathbf{A}^{-1}\mathbf{b} = 1 - (\mathbf{A}^{-1/2}\mathbf{b})^H(\mathbf{A}^{-1/2}\mathbf{b}) \tag{64}$$
$$= |I - \mathbf{A}^{-1/2}\mathbf{b}\mathbf{b}^H\mathbf{A}^{-1/2}| \tag{65}$$
$$= |\mathbf{A} - \mathbf{b}\mathbf{b}^H|/|\mathbf{A}| \geq 0. \tag{66}$$

This establishes the necessary condition $\mathbf{b}^H A^{-1}\mathbf{b} \leq 1$. To prove sufficiency, choose an $x > 0$; we can show that

$$0 \preceq \mathbf{A}\left(\mathbf{A} + \frac{1}{x}\mathbf{b}\mathbf{b}^H\right)^{-1}\mathbf{A} \tag{67}$$
$$= \mathbf{A}\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{b}\frac{1}{x + \mathbf{b}^H A^{-1}\mathbf{b}}\mathbf{b}^H A^{-1}\right)\mathbf{A} \tag{68}$$

$$= \mathbf{A} - \frac{1}{x + \mathbf{b}^H A^{-1} \mathbf{b}} \mathbf{b}\mathbf{b}^H \qquad (69)$$

$$\preceq \mathbf{A} - \frac{1}{x+1} \mathbf{b}\mathbf{b}^H. \qquad (70)$$

where (68) follows from the matrix inverse lemma, and the last step uses $\mathbf{b}^H A^{-1} \mathbf{b} \leq 1$. By letting $x \to 0$, we obtain $\mathbf{A} - \mathbf{b}\mathbf{b}^H \succeq 0$. This establishes the Lemma under the assumption $\mathbf{A} \succ 0$.

Next, if $\mathbf{A}$ is positive semidefinite with $\mathrm{rank}(\mathbf{A}) = n_1$, then there exists an unitary matrix $\mathbf{U}$, such that $\mathbf{U}\mathbf{A}\mathbf{U}^H = \mathrm{diag}\{\mathbf{U}_1 \mathbf{A}\mathbf{U}_1^H, 0\}$, where $\mathbf{U}_1$ represents the first n_1 rows of $\mathbf{U}$. Note that in the forward direction, if $\mathbf{A} \succeq \mathbf{b}\mathbf{b}^H$, then it is easy to see that $\mathbf{U}\mathbf{b}$ must be of the form $[\mathbf{U_1}\mathbf{b}^T \ 0^T]^T$. In the reverse direction, the second half of $\mathbf{U}\mathbf{b}$ may be set arbitrarily without affecting $\mathbf{b}^H \mathbf{A}^\dagger \mathbf{b}$. Thus, without loss of generality, we may regard $\mathbf{U}\mathbf{b} = [\mathbf{U_1}\mathbf{b}^T \ 0^T]^T$. Then

$$\mathbf{A} \succeq \mathbf{b}\mathbf{b}^H \Leftrightarrow \mathbf{U}(\mathbf{A} - \mathbf{b}\mathbf{b}^H)\mathbf{U}^H \succeq 0$$
$$\Leftrightarrow \mathbf{U}_1\mathbf{A}\mathbf{U}_1^H \succeq (\mathbf{U}_1\mathbf{b})(\mathbf{U}_1\mathbf{b})^H$$
$$\Leftrightarrow \mathbf{b}^H\mathbf{U}_1^H \left(\mathbf{U}_1\mathbf{A}\mathbf{U}_1^H\right)^{-1} \mathbf{U}_1\mathbf{b} \leq 1$$
$$\Leftrightarrow \mathbf{b}^H\mathbf{U}^H(\mathbf{U}\mathbf{A}\mathbf{U}^H)^{\dagger}\mathbf{U}\mathbf{b} \leq 1$$
$$\Leftrightarrow \mathbf{b}^H\mathbf{A}^{\dagger}\mathbf{b} \leq 1$$

where the third step follows from the positive definite case discussed earlier. This completes the proof of the Lemma. $\square$

C. Proof of Proposition 3

We prove in this appendix that the function $f(\mathbf{Q})$ is concave in $\mathbf{Q}$, and $\mathrm{diag}\{\sum_{i=1}^K \mathbf{w_i}\mathbf{w_i}^H\}$ is a subgradient of $f(\mathbf{Q})$. First, the concavity of $f(\mathbf{Q})$ follows directly from the fact that $f(\mathbf{Q})$ is the objective function of a dual problem, which is always concave. The following several steps verify this directly.

Transform $f(\mathbf{Q})$ into the downlink domain using the beamforming duality

$$f(\mathbf{Q}) = \min_{\mathbf{w_i}} \quad \sum_{i=1}^K \mathbf{w_i}^H \mathbf{Q} \mathbf{w_i}$$

$$\text{s.t.} \quad \frac{\left|\mathbf{w_i}^H \mathbf{h_i}\right|^2}{\sum_{j \neq i} \left|\mathbf{w_j}^H \mathbf{h_i}\right|^2 + \sigma^2} \geq \gamma_i, \qquad \forall i \qquad (71)$$

where $\mathbf{w_i}$ is the corresponding downlink beamformer.

Let $\mathbf{Q}_1$ and $\mathbf{Q}_2$ be two positive semidefinite matrices. We use $\mathbf{w}_{i,k}$ to denote the optimal beamformer for $f(\mathbf{Q}_k)$, $k = 1, 2$. To verify concavity, we have

$$f\left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2}\right) = \min_{\{\mathbf{w_i}:\mathrm{SINR}_i \geq \gamma_i\}} \sum_{i=1}^K \mathbf{w_i}^H \left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2}\right) \mathbf{w_i}$$

$$\geq \frac{1}{2}\sum_{i=1}^K \mathbf{w}_{i,1}^H \mathbf{Q}_1 \mathbf{w}_{i,1} + \frac{1}{2}\sum_{i=1}^K \mathbf{w}_{i,2}^H \mathbf{Q}_2 \mathbf{w}_{i,2}$$

$$= \frac{1}{2}f(\mathbf{Q}_1) + \frac{1}{2}f(\mathbf{Q}_2).$$

By definition, for a concave function $f(\mathbf{Q})$, $\mathbf{H}$ is a subgradient of f at $\mathbf{Q}_1$ if $f(\mathbf{Q}_2) \leq f(\mathbf{Q}_1) + \mathrm{tr}(\mathbf{H}(\mathbf{Q}_2 - \mathbf{Q}_1))$ for all $\mathbf{Q}_2$. Now
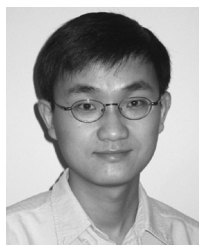
$$f(\mathbf{Q}_2) - f(\mathbf{Q}_1) = \sum_{i=1}^K \mathbf{w}_{i,2}^H \mathbf{Q}_2 \mathbf{w}_{i,2} - \sum_{i=1}^K \mathbf{w}_{i,1}^H \mathbf{Q}_1 \mathbf{w}_{i,1}$$

$$\leq \sum_{i=1}^K \mathbf{w}_{i,1}^H \mathbf{Q}_2 \mathbf{w}_{i,1} - \sum_{i=1}^K \mathbf{w}_{i,1}^H \mathbf{Q}_1 \mathbf{w}_{i,1}$$

$$= \mathrm{tr}\left(\left\{\sum_{i=1}^K \mathbf{w}_{i,1}\mathbf{w}_{i,1}^H\right\}(\mathbf{Q}_2 - \mathbf{Q}_1)\right)$$

$$= \mathrm{tr}\left(\mathrm{diag}\left\{\sum_{i=1}^K \mathbf{w}_{i,1}\mathbf{w}_{i,1}^H\right\}(\mathbf{Q}_2 - \mathbf{Q}_1)\right)$$

where the last step follows because $(\mathbf{Q}_2 - \mathbf{Q}_1)$ is diagonal. Thus, $\mathrm{diag}\{\sum_{i=1}^K \mathbf{w_i}\mathbf{w_i}^H\}$ is a subgradient of $f(\mathbf{Q})$. $\square$

REFERENCES

[1] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[2] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. IEEE. Veh. Tech. Conf.*, May 1999, vol. 1, pp. 851–856.

[3] F. Rashid-Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 46, no. 10, pp. 1313–1324, Oct. 1998.

[4] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[5] ——, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2324–2334, Jul. 2005.

[6] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. 37th Allerton Conf. Commun. Control Comput.*, Sep. 1999, pp. 987–996.

[7] ——, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications,* L. C. Godara, Ed. Boca Raton, FL: CRC, 2002.

[8] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

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

[10] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Veh. Technol.*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[11] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

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

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

[14] H. Weingarten, Y. Steinberg, and S. Shamai (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.

[15] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 68–78, May 2004.

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

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

[18] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communications*, 3rd ed. Norwell, MA: Kluwer, 2004.

[19] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[20] W. Yu, D. P. Varodayan, and J. M. Cioffi, "Trellis and convolutional precoding for transmitter-based interference presubtraction," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1220–1230, Jul. 2005.

[21] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.

[22] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, MA: Cambridge Univ. Press, 2005.

[23] J.-H. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint transmitter receiver diversity for efficient space-division multiaccess," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 16–27, Jan. 2002.

[24] R. Doostnejad, T. J. Lim, and E. Sousa, "Precoding for the MIMO broadcast channels with multiple antennas at each receiver," in *Proc. 39th Annu. Conf. Inf. Sci. Syst. (CISS)*, Baltimore, MD, Mar. 2005, CDROM.

[25] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, ser. Computational Mathematics. New York: Springer, 1985.

Tian Lan (S'03) received the B.Sc. degree in electronic information engineering from Tsinghua University, Beijing, China, in 2003 and the M.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2005. He is currently working towards the Ph.D. degree in electrical engineering at Princeton University, Princeton, NJ.

His research interests are in wireless communications, optimization, distributed algorithms, and network protocols.

Wei Yu (S'97–M'02) received the B.A.Sc. degree in computer engineering and mathematics from the University of Waterloo, Waterloo, ON, Canada, in 1997 and M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively.

Since 2002, he has been an Assistant Professor with the Electrical and Computer Engineering Department, University of Toronto, Toronto, ON, Canada, where he also holds a Canada Research Chair. His main research interests include multi-user information theory, optimization, wireless communications, and broadband access networks.

Prof. Wei Yu is an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS for a special issue on "Nonlinear Optimization of Communications Systems," and a Guest Editor of the *EURASIP Journal on Applied Signal Processing* for a special issue on "Advanced Signal Processing for Digital Subscriber Lines."