Uplink–Downlink Duality Via Minimax Duality

Wei Yu, *Member, IEEE*

Abstract—**The sum capacity of a Gaussian vector broadcast channel is the saddle point of a minimax Gaussian mutual information expression where the maximization is over the set of transmit covariance matrices subject to a power constraint and the minimization is over the set of noise covariance matrices subject to a diagonal constraint. This sum capacity result has been proved using two different methods, one based on decision-feedback equalization and the other based on a duality between uplink and downlink channels. This paper illustrates the connection between the two approaches by establishing that uplink–downlink duality is equivalent to Lagrangian duality in minimax optimization. This minimax Lagrangian duality relation allows the optimal transmit covariance and the least-favorable-noise covariance matrices in a Gaussian vector broadcast channel to be characterized in terms of the dual variables. In particular, it reveals that the least favorable noise is not unique. Further, the new Lagrangian interpretation of uplink–downlink duality allows the duality relation to be generalized to Gaussian vector broadcast channels with arbitrary linear constraints. However, duality depends critically on the linearity of input constraints. Duality breaks down when the input constraint is an arbitrary convex constraint. This shows that the minimax representation of the broadcast channel sum capacity is more general than the uplink–downlink duality representation.**

Index Terms—**Broadcast channel, Lagrangian duality, minimax optimization, multiple-input multiple-output (MIMO), multiple-access channel, multiple-antenna.**

I. INTRODUCTION

THERE is an interesting input–output reciprocity for Gaussian vector channels. Consider a Gaussian vector channel under a power constraint:

$$Y = HX + Z \qquad (1)$$

where X and Y are vector-valued input and output, respectively, H is the channel matrix, and Z is the additive independent and identically distributed (i.i.d.) Gaussian vector noise with a covariance matrix equal to an identity matrix. The capacity of the channel remains the same if the input and the output are interchanged, the channel matrix is transposed, and the same power constraint is applied to the reciprocal channel

$$\hat{Y} = H^T \hat{X} + \hat{Z}. \qquad (2)$$

This relation has been observed in [1]. Reciprocity holds because the computation of the Gaussian vector channel capacity under a power constraint involves a water-filling of the total power over the set of singular values of the channel matrix and the singular values of H and H^T are identical. This reciprocity relation is true even when the matrix H is not a square matrix. In this case, $\hat{X}$ and X (also $\hat{Z}$ and Z) do not have to have the same dimension.

Interestingly, the input–output reciprocity generalizes to multiuser channels. Let $Y^T = [Y_1^T \cdots Y_K^T]$. Consider a Gaussian vector broadcast channel

$$Y_i = H_i X + Z_i, \qquad i = 1, \ldots, K \qquad (3)$$

where Y_i's do not cooperate and $H^T = [H_1^T \cdots H_K^T]$. The capacity region of the Gaussian vector broadcast channel under a sum power constraint is exactly the same as the capacity region of a reciprocal multiple-access channel, where the roles of inputs and outputs are reversed, the channel matrix is transposed, and a sum power constraint is applied to all input terminals [2]–[4]. Let $\hat{X}^T = [\hat{X}_1^T \cdots \hat{X}_K^T]$. The reciprocal multiple-access channel has the form

$$\hat{Y} = H_i^T \hat{X}_i + \hat{Z}_i, \qquad i = 1, \ldots, K \qquad (4)$$

and it has a sum capacity

$$C = \max_{S_{\hat{x}\hat{x}}} I(\hat{X}_1 \cdots \hat{X}_K; \hat{Y}) \qquad (5)$$

where $S_{\hat{x}\hat{x}}$ is a block-diagonal matrix with covariance matrices $\hat{X}_1 \cdots \hat{X}_K$ on the diagonal, and the maximization is over all positive semidefinite matrices $S_{\hat{x}\hat{x}}$ subject to a sum power constraint $\mathrm{trace}(S_{\hat{x}\hat{x}}) \leq P$. Reciprocity implies that the capacity expression (5) is also the sum capacity of the Gaussian vector broadcast channel. This reciprocity relation is called "uplink–downlink duality," because the multiple-access channel corresponds to the uplink transmission and the broadcast channel corresponds to the downlink transmission in a wireless system.

This paper focuses on the sum capacity of the Gaussian broadcast channel and aims to provide a new understanding of uplink–downlink duality. Uplink–downlink duality is originally established in [2] and [3] by identifying a signal-to-interference-and-noise ratio (SINR) transformation between the multiple-access channel and the broadcast channel. Such a transformation, however, is not very intuitive. One of the main contributions of this paper is that we show uplink–downlink duality is in fact equivalent to Lagrangian duality in convex optimization. This new interpretation not only unifies several existing seemingly unrelated proofs for the sum capacity, but also permits a new characterization of the optimal input and noise covariance matrices in a Gaussian vector broadcast

channel. As the dual problem often lies in a lower dimensional space, the computation of sum capacity and the optimal transmission strategy for a broadcast channel can be done more efficiently via its dual.

Further, this new interpretation allows a generalization of uplink–downlink duality to broadcast channels with general linear input constraints. Generalized input constraints arise, for example, in a multiple-antenna broadcast channel with per-antenna power constraints. Interestingly, the dual of such a channel happens to be a multiple-access channel with an uncertain noise.

A. Literature Review

To put the new results in this paper into perspective, this subsection is devoted to an overview of the Gaussian vector broadcast channel. The capacity of Gaussian vector broadcast channel is first investigated in [5] and the sum capacity is proved for the case where the transmitter is equipped with two antennas. This sum capacity result is then independently generalized in [6], [2], and [3] for a broadcast channel with an arbitrary number of antennas. Finally, [4] fully established the entire capacity region for the Gaussian vector broadcast channel.

The achievability result for the Gaussian vector broadcast channel is obtained using a precoding technique known as "writing on dirty paper" [7]. The main idea is as follows [5]. To transmit independent information to multiple receivers at the same time, the transmitter may treat the signals intended for other users as known interference. Using dirty-paper precoding, such known interference can effectively be subtracted at the transmitter.

The converse result is established using the notion of "least favorable" noise. The idea is based on the observation that the capacity of a broadcast channel does not depend on the joint distribution of the noise but only on the marginals [8]. It then follows that the sum capacity of the vector broadcast channel is bounded by the minimum capacity of a cooperative vector channel

$$C = \min_{p(z)} \max_{p(x)} I(\boldsymbol{X}; \boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K) \qquad (6)$$

minimized over all possible joint distributions of the noises. For a Gaussian vector broadcast channel with a convex input constraint, the minimax solution is necessarily jointly Gaussian.

The main part of the sum capacity result is in proving that the minimax upper bound is achievable using dirty-paper precoding. Several proofs are available. One way to establish the achievability is to design a decision-feedback equalizer as a joint receiver in the broadcast channel [6]. As shown in [6], under a least-favorable-noise correlation, a decision-feedback equalizer can be designed so that the feedforward matrix is diagonal. In addition, the feedback section can be moved to the transmitter as a precoder. Thus, with a least favorable noise, no receiver cooperation is needed and $\min I(\boldsymbol{X}; \boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K)$ is achievable. This minimum can be further maximized over all possible input distributions. Since min-max and max-min are equal in a convex–concave function, this implies that (6) is indeed the sum capacity of a Gaussian vector broadcast channel.

The minimax capacity result for the broadcast channel can be obtained using a completely different approach, called up-

link–downlink duality. This duality approach is used in [2] and [3] and is based on an observation that the achievable SINRs of the uplink channel and the downlink channel are exactly the same. The proof in [2] further uses the fact that the noise covariance matrix for the broadcast channel corresponds to an input cost constraint in the reciprocal channel. It is then established that the input cost constraint corresponding to the least-favorable-noise covariance matrix is precisely the one that decouples the inputs of the reciprocal channel. This establishes the duality between the broadcast channel and the multiple-access channel and shows that (5) is the broadcast channel sum capacity.

The approach in [3] is based on a different idea. The authors of [3] observed that the precoding region for a Gaussian vector broadcast channel is exactly the capacity region of the reciprocal multiple-access channel. The proof of this fact involves a clever choice of transmit covariance matrix for the broadcast channel for each achievable point in the multiple-access channel and *vice versa.* Based on this duality, [3] showed that the minimax problem for the broadcast channel (6) and the maximization problem for the dual multiple-access channel (5) have precisely the same solution.

Although giving essentially the same result, the results in [6], [2], and [3] nevertheless contain several subtle differences. First, the duality approaches of [2] and [3] appear to apply only to Gaussian vector broadcast channels with a sum power constraint, while the decision-feedback equalization approach of [6] applies to broadcast channels with arbitrary convex constraints and is therefore more general. Second, the minimax result [6] applies only to the sum capacity point, while the duality result [3] applies to the entire achievable rate region. Third, the duality expression (5) is numerically much easier to evaluate as compared to the minimax expression (6). This is especially true when the least favorable noise is singular. Singular least favorable noise is not necessarily an unlikely event. In fact, it almost always happens when the number of receive dimensions is larger than the number of transmit dimensions in the broadcast channel.

B. Overview of Main Results

The purpose of this paper is to establish a connection between the duality approach and the minimax approach, to reconcile some of their differences, and to provide an explicit solution to the Gaussian vector broadcast channel sum capacity problem. Toward this end, uplink–downlink duality is re-interpreted in the framework of Lagrangian duality in minimax optimization. This general approach not only unifies the previous results on Gaussian broadcast channel sum capacity, but also allows an explicit solution to the minimax problem based on its dual. In fact, the new Lagrangian duality relation reveals that the least favorable noise is often not unique, and different least favorable noises are related to each other via a linear estimation relation. Finally, the minimax framework allows uplink–downlink duality to be generalized to broadcast channels under arbitrary linear covariance constraints. However, it is also shown that duality depends critically on the linearity of the input constraint. For a broadcast channel with a general convex input constraint, the minimax expression for the sum capacity is more general than the characterization based on uplink–downlink duality.

The organization for the rest of the paper is as follows. In Section II, a minimax Lagrangian duality relation for the Gaussian mutual information is established. In Section III, the minimax duality theory is applied to the Gaussian vector broadcast channel and a new derivation of uplink–downlink duality is given. This new derivation not only illustrates the connection between uplink–downlink duality and Lagrangian duality, but also gives a characterization of the saddle point of the Gaussian minimax mutual information expression in terms of the dual variables. In Section IV, the uplink–downlink duality is generalized to broadcast channels with arbitrary linear covariance constraints, and the sum capacity for a Gaussian vector broadcast channel with individual per-antenna power constraints is found. Section V contains concluding remarks.

II. DUALITY FOR GAUSSIAN MUTUAL INFORMATION

In this section, we illustrate a Lagrangian duality for a minimax Gaussian mutual information expression. Lagrangian duality refers to the fact that a convex minimization problem can be solved as a maximization problem over a set of dual variables where each dual variable corresponds to a constraint in the original primal problem [9], [10]. For a minimax problem, its dual is also a minimax problem. The Lagrangian duality, when applied to minimax problems, is termed minimax duality in this paper.

Consider a Gaussian channel in which the transmitter chooses a transmit covariance matrix S_{xx} to maximize the mutual information and the receiver chooses a noise covariance matrix S_{zz} to minimize the mutual information

$$C(H, Q_x, Q_z) = \max_{S_{xx}} \min_{S_{zz}} \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|} \quad (7)$$

subject to linear covariance constraints of the form

$$\text{tr}(S_{xx}Q_x) \le 1 \quad (8)$$
$$\text{tr}(S_{zz}Q_z) \le 1 \quad (9)$$

where Q_x and Q_z are parameters of the linear constraints, and they are assumed to be symmetric positive semidefinite matrices. Implicitly, S_{xx} and S_{zz} are constrained to be positive semidefinite matrices. For example, the usual power constraint corresponds to an identity matrix Q_x. The minimax capacity is a function of H, Q_x, and Q_z and is denoted as $C(H, Q_x, Q_z)$. This minimax problem may correspond to the capacity of a compound channel in which the transmitter must construct a codebook to achieve a vanishing probability of error for all possible realizations of the noise or a broadcast channel in which the noise correlation among the receivers may vary arbitrarily.

The main result of this section is a characterization of a Lagrangian dual of the above minimax problem. The duality formulation presented in this section is different from the conventional Lagrangian duality as in [9] and [10]. It relies instead on a transformation of the optimality conditions of the minimax problem and pertains to the particular form of the Gaussian mutual information expression only. Note that the Lagrangian dual of an optimization problem is in general nonunique. The duality between the two minimax problems presented in

this section is one possible way in which Lagrangian duality relation may be formulated.

The dual relation presented in this section is most transparent when H is invertible and the saddle-point solutions S_{xx} and S_{zz} are strictly positive definite. For the moment, we explicitly assume the invertibility and positive semidefiniteness conditions. These are technical conditions that will be removed later.

The first step in the development of minimax duality is a characterization of the saddle point of the minimax problem via its Karush–Kuhn–Tucker (KKT) condition. The KKT condition consists of the usual water-filling condition with respect to the maximization over S_{xx}

$$H^T(HS_{xx}H^T + S_{zz})^{-1}H = \lambda_x Q_x \quad (10)$$

and the least-favorable-noise condition with respect to the minimization over S_{zz}

$$S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1} = \lambda_z Q_z \quad (11)$$

where λ_x and λ_z are the appropriate Lagrangian dual variables. (The coefficient $\frac{1}{2}$ is omitted for simplicity.) The KKT condition is necessary and sufficient for optimality. Now, pre- and post-multiplying (11) by H^T and H, respectively, substituting (10) into (11), and rearranging the terms, it is easy to see that

$$H^T S_{zz}^{-1} H = H^T \lambda_z Q_z H + \lambda_x Q_x. \quad (12)$$

If H is invertible, then

$$H(H^T \lambda_z Q_z H + \lambda_x Q_x)^{-1}H^T = S_{zz}. \quad (13)$$

This is precisely the water-filling KKT condition for a Gaussian vector channel with H^T as the channel matrix, $\lambda_z Q_z$ as the transmit covariance matrix, and $\lambda_x Q_x$ as the noise covariance. The above is also an explicit solution for the minimizing S_{zz}. Further, substituting (13) into (10) and solving for S_{xx}, we obtain

$$(\lambda_x Q_x)^{-1} - (H^T \lambda_z Q_z H + \lambda_x Q_x)^{-1} = S_{xx}. \quad (14)$$

This is precisely the least-favorable-noise KKT condition with H^T as the channel matrix, $\lambda_x Q_x$ as the least favorable noise, and $\lambda_z Q_z$ as the transmit covariance matrix. Define

$$\Sigma_{xx} = \lambda_z Q_z, \quad \Sigma_{zz} = \lambda_x Q_x \quad (15)$$
$$S_{xx} = \nu_z \Psi_z, \quad S_{zz} = \nu_x \Psi_x. \quad (16)$$

Equations (13) and (14) can be rewritten as

$$H(H^T \Sigma_{xx} H + \Sigma_{zz})^{-1}H^T = \nu_x \Psi_x \quad (17)$$
$$\Sigma_{zz}^{-1} - (H^T \Sigma_{xx} H + \Sigma_{zz})^{-1} = \nu_z \Psi_z. \quad (18)$$

Now, if ν_x and ν_z are set to

$$\lambda_x = \nu_z \ , \ \lambda_z = \nu_x \quad (19)$$

then

$$\text{tr}(\Sigma_{xx}\Psi_x) = \text{tr}(S_{zz}Q_z) \quad (20)$$
$$\text{tr}(\Sigma_{zz}\Psi_z) = \text{tr}(S_{xx}Q_x). \quad (21)$$

TABLE I
MINIMAX DUALITY

Minimax Problem:	$\max\limits_{S_{xx}} \min\limits_{S_{zz}} \dfrac{1}{2} \log \dfrac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|}$	$\max\limits_{\Sigma_{xx}} \min\limits_{\Sigma_{zz}} \dfrac{1}{2} \log \dfrac{|H^T\Sigma_{xx}H + \Sigma_{zz}|}{|\Sigma_{zz}|}$
Constraints:	$\mathrm{tr}(S_{xx}Q_x) \leq 1$ $\mathrm{tr}(S_{zz}Q_z) \leq 1$	$\mathrm{tr}(\Sigma_{xx}\Psi_x) \leq 1$ $\mathrm{tr}(\Sigma_{zz}\Psi_z) \leq 1$
Primal Variables:	$S_{xx} = \nu_z\Psi_z$ $S_{zz} = \nu_x\Psi_x$	$\Sigma_{xx} = \lambda_z Q_z$ $\Sigma_{zz} = \lambda_x Q_x$
Dual Variables:	$\lambda_x = \nu_z$ $\lambda_z = \nu_x$	$\nu_x = \lambda_z$ $\nu_z = \lambda_x$

Therefore, associated with the original minimax problem (7), there is a "dual" minimax problem

$$C(H^T, \Psi_x, \Psi_z) = \max_{\Sigma_{xx}} \min_{\Sigma_{zz}} \frac{1}{2} \log \frac{|H^T\Sigma_{xx}H + \Sigma_{zz}|}{|\Sigma_{zz}|} \quad (22)$$

with linear covariance constraints

$$\mathrm{tr}(\Sigma_{xx}\Psi_x) \leq 1 \quad (23)$$
$$\mathrm{tr}(\Sigma_{zz}\Psi_z) \leq 1. \quad (24)$$

Further, it can be verified using (10), (11) and (15) that at the saddle point, we get (25) at the bottom of the page. Thus, the minimax problems (7) and (22) are duals of each other in the following sense.

- The optimal dual variable λ_x in the maximization part of (7) is the optimal dual variable ν_z in the minimization part of (22).
- The optimal dual variable λ_z in the minimization part of (7) is the optimal dual variable ν_x in the maximization part of (22).
- The optimizing variables of (7) are related to the constraints of (22) by $(S_{xx}, S_{zz}) = (\nu_z\Psi_z, \nu_x\Psi_x)$.
- The optimizing variables of (22) are related to the constraints of (7) by $(\Sigma_{xx}, \Sigma_{zz}) = (\lambda_z Q_z, \lambda_x Q_x)$.
- $C(H, Q_x, Q_z) = C(H^T, \Psi_x, \Psi_z)$.

The duality relation is summarized in Table I. This minimax duality result is formally stated as follows.

Theorem 1: Let $(S_{xx}, S_{zz}, \lambda_x, \lambda_z)$ be the primal and dual optimal solutions of the Gaussian minimax mutual information optimization problem $C(H, Q_x, Q_z)$. Define $\Psi_z = S_{xx}/\lambda_x$ and $\Psi_x = S_{zz}/\lambda_z$. Then, the primal and dual optimal solutions of the dual minimax problem $C(H^T, \Psi_x, \Psi_z)$ are precisely $\Sigma_{xx} = \lambda_z Q_z$, $\Sigma_{zz} = \lambda_x Q_x$, $\nu_x = \lambda_z$, and $\nu_z = \lambda_x$. Further, $C(H^T, \Psi_x, \Psi_z) = C(H, Q_x, Q_z)$.

The derivation leading to the theorem provides a proof of Theorem 1 for the case where H is square and invertible and where the optimal S_{xx} and S_{zz} are both full rank. The proof for the general case is given in Appendix A.

To summarize, there is an input–output duality for the Gaussian mutual information minimax problem. By transposing the channel matrix and interchanging the input and the output, the constraints of the original problem become the optimal solution of the dual problem (and *vice versa.*) Solving one minimax problem is equivalent to solving the other.

It should be emphasized that the Lagrangian dual of an optimization problem is in general not unique. The minimax duality relation formulated in this section is a particular form of Lagrangian dual formulation that allows the Gaussian vector broadcast channel to be related to a dual multiple-access channel, as shown in the next section. This duality relation is established based on the KKT conditions of the optimization problem, and it is a particular feature of the Gaussian mutual information expression.

III. UPLINK–DOWNLINK DUALITY

The main motivation for studying the minimax duality relation is that it arises naturally in the characterization of the sum capacity of Gaussian vector broadcast channels. As mentioned earlier, the sum capacity for the Gaussian vector broadcast channel has been solved independently using two seemingly different approaches. In [6], the broadcast channel sum capacity is shown to be the solution of a minimax mutual information problem, while in [11], [3], and [2], the broadcast channel sum capacity is shown to be the capacity of a dual multiple-access channel with a sum power constraint. The objective of this section is to unify the two approaches and to show that the duality between the broadcast channel and the multiple-access channel is a special case of minimax duality.

$$\log \frac{|H^T\Sigma_{xx}H + \Sigma_{zz}|}{|\Sigma_{zz}|} = \log \frac{|H^T\lambda_z Q_z H + \lambda_x Q_x|}{|\lambda_x Q_x|}$$
$$= \log \frac{|H^T(S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1})H + H^T(HS_{xx}H^T + S_{zz})^{-1}H|}{|H^T(HS_{xx}H^T + S_{zz})^{-1}H|}$$
$$= \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|}. \quad (25)$$

A. Gaussian Vector Broadcast Channel Sum Capacity

Consider a Gaussian vector broadcast channel

$$\boldsymbol{Y}_i = H_i\boldsymbol{X} + \boldsymbol{Z}_i, \qquad i = 1,\ldots,K \tag{26}$$

where $\boldsymbol{X}$ is the transmit signal, $\boldsymbol{Y}_i$ is the received signal for user i. $\boldsymbol{Z}^T = [\boldsymbol{Z}_1^T \cdots \boldsymbol{Z}_K^T]$ is the set of Gaussian noise vectors with covariance matrices equal to identity matrices, and $H^T = [H_1^T \cdots H_K^T]$ is the set of channel matrices. A sum power constraint $\mathbb{E}[\boldsymbol{X}^T\boldsymbol{X}] \leq P$ is imposed on the input. A key ingredient in the characterization of the capacity is a connection between the broadcast channel and channels with side information. In a classic result known as "writing on dirty paper," [7] Costa showed that if a Gaussian channel is corrupted by an interference signal $\boldsymbol{S}$ that is known noncausally to the transmitter but not to the receiver, i.e.,

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{S} + \boldsymbol{Z} \tag{27}$$

the capacity of the channel is the same as if $\boldsymbol{S}$ does not exist. Thus, in a broadcast channel, if $\boldsymbol{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2$ where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are Gaussian vectors, $\boldsymbol{X}_1$ can transmit information to $\boldsymbol{Y}_1$ as if $\boldsymbol{X}_2$ does not exist, and $\boldsymbol{X}_2$ can transmit to $\boldsymbol{Y}_2$ with $\boldsymbol{X}_1$ regarded as noise. This precoding strategy turns out to be optimal for sum capacity in a Gaussian broadcast channel. This is proved for the two-user two-antenna case by Caire and Shamai [5] and has since been generalized by several authors [6], [3], [2].

The approach in [6] is based on the observation that interference presubtraction at the transmitter is identical to decision-feedback equalization with feedback "moved" to the transmitter. However, while the decision-feedback structure is capacity achieving for the Gaussian vector channel, it also requires coordination at the receivers because it has a feedforward matrix that operates on the entire set of $\boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K$. Clearly, such coordination is not possible in a broadcast channel. But, precisely because $\boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K$ cannot coordinate, they are also ignorant of the noise correlation between $\boldsymbol{Z}_1 \cdots \boldsymbol{Z}_K$. Thus, the sum capacity of the broadcast channel must be bounded by the cooperative capacity with the *least favorable* noise correlation

$$C \leq \min_{S_{zz}} I(\boldsymbol{X}; \boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K) \tag{28}$$

where S_{zz} is the covariance matrix for $\boldsymbol{Z}^T = [\boldsymbol{Z}_1^T \cdots \boldsymbol{Z}_K^T]$, and the minimization is over all S_{zz} whose block-diagonal terms are the covariance matrices of $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_K$. This outer bound is due to Sato [8].

Now, assume that the transmit signal of the broadcast channel is a Gaussian signal with a fixed covariance matrix S_{xx}. Then, the KKT condition associated with the minimization problem is

$$S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1} = \begin{bmatrix} \Phi_1 & & 0 \\ & \ddots & \\ 0 & & \Phi_K \end{bmatrix} = \Phi \tag{29}$$

where Φ_i is the dual variable corresponding to the ith diagonal constraint. Interestingly, $S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1}$ also corresponds to the feedforward matrix of the decision-feedback equalizer. So, if the noise covariance is least favorable, the feedforward matrix of the decision-feedback equalizer would be diagonal. Thus, after moving the feedback operation to the transmitter, the entire equalizer decouples into independent receivers for each user, and no coordination is needed whatsoever. Consequently, the Sato outer bound is achievable. Thus, assuming Gaussian signaling with a fixed transmit covariance matrix S_{xx}, the capacity of the broadcast channel is precisely $\min_{S_{zz}} I(\boldsymbol{X}; \boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K)$.

This mutual information minimization problem does not appear to have a closed-form solution, and it is not necessarily easy to solve numerically either. However, there is a special case for which the minimization is easy. This happens when S_{xx} is being maximized at the same time.

Consider an example in which the input signal is subject to a power constraint

$$\text{tr}(S_{xx}) \leq P. \tag{30}$$

In this case, it is not difficult to see that Sato's bound becomes

$$C \leq \min_{S_{zz}} \max_{S_{xx}} I(\boldsymbol{X}; \boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K) \tag{31}$$

Further, $\min_{S_{zz}} I(\boldsymbol{X}; \boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K)$ is achievable with any Gaussian inputs, thus, it can be maximized over all S_{xx}

$$C \geq \max_{S_{xx}} \min_{S_{zz}} I(\boldsymbol{X}; \boldsymbol{Y}_1 \cdots \boldsymbol{Y}_K). \tag{32}$$

Now, the Gaussian mutual information expression is concave in S_{xx} and convex in S_{zz}, so min-max is equal to max-min. In addition, the saddle point is Gaussian. Thus, the sum capacity is precisely the solution to the following problem:

$$\begin{aligned}
\max_{S_{xx}} \min_{S_{zz}} \quad & \frac{1}{2}\log\frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|} \\
\text{subject to} \quad & S_{zz}(i,i) = I, \qquad i = 1,\ldots,K. \\
& \text{tr}(S_{xx}) \leq P \\
& S_{xx}, S_{zz} \geq 0
\end{aligned} \tag{33}$$

where $S_{zz}(i,i)$ refers to the ith block-diagonal term of S_{zz}.

Interestingly, the preceding minimax optimization problem is considerably easier to solve than the minimization problem (28) alone. This is because the minimax problem (33) has a dual expression which allows it to be transformed into a lower dimensional space, as is explained in the next section.

B. Uplink–Downlink Duality Via Minimax Duality

Uplink–downlink duality refers to the observation that the sum capacity of a multiple-access channel and a broadcast channel are the same under the same input sum power constraint [11], [2], [3]. This uplink–downlink duality relation has since been generalized to the entire capacity region [4]. The original proofs of uplink–downlink duality are based on the equivalence of the achievable SINRs in the uplink and downlink channels. For example, [3] proposed an elaborate covariance transformation that maps the SINRs of the uplink channel to the SINRs of the downlink channel, and *vice versa*. The main objective of this section is to give a new derivation of uplink–downlink duality via minimax Lagrangian duality. The new derivation provides new insights into the structure of the optimization problem.

The derivation is again simplest with the assumptions that H is square and invertible and that the maximizing S_{xx} and the

least favorable S_{zz} are both full rank. Further, let us assume that each receiver in the broadcast channel is equipped with a single antenna. The following simplified derivation captures the essence of the problem. The starting point is the KKT condition for the minimax problem (33). The KKT condition is a sufficient and necessary characterization of the saddle point, because the objective function in (33) is concave in S_{xx} and convex in S_{zz}. The KKT condition is

$$H^T(HS_{xx}H^T + S_{zz})^{-1}H = \lambda I \tag{34}$$

$$S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1} = \Phi \tag{35}$$

where λ is the dual variable associated with the power constraint and Φ is a diagonal matrix of dual variables associated with the diagonal constraint on the noise covariance matrix. Multiplying (35) by H^T on the left and H on the right and substituting in (34), we obtain

$$H^T S_{zz}^{-1} H = H^T \Phi H + \lambda I. \tag{36}$$

The above is equivalent to

$$H(H^T \Phi H + \lambda I)^{-1} H^T = S_{zz}. \tag{37}$$

Observe that since Φ is a diagonal matrix, (37) is precisely the KKT condition for a multiple-access channel with a transmit covariance matrix Φ and an input constraint of the form $\mathrm{tr}(\Phi S_{zz}) \leq 1$. The above equation also gives an explicit solution for S_{zz}. Further, substituting (37) into (34), it is not difficult to see that

$$(\lambda I)^{-1} - (H^T \Phi H + \lambda I)^{-1} = S_{xx}. \tag{38}$$

This is an explicit solution for S_{xx}. It is in fact also a least-favorable-noise condition for the dual problem.

Equations (37) and (38) define a set of KKT conditions for a dual minimax problem. From (37) and (38), it is clear that the following problem is a dual of (33):

$$\max_{\Phi} \min_{\lambda} \quad \frac{1}{2} \log \frac{|H^T \Phi H + \lambda I|}{|\lambda I|}$$
$$\text{subject to} \quad \mathrm{tr}(\Phi S_{zz}) \leq 1$$
$$\mathrm{tr}(\lambda S_{xx}) \leq 1$$
$$\Phi, \lambda \geq 0. \tag{39}$$

As the original minimax problem (33) and the dual problem (39) have equivalent KKT conditions, (39) is in fact a Lagrangian optimization dual of (33). Note that the development so far assumes that S_{xx} and S_{zz} are full rank. These are technical conditions which may be removed using the same technique as in the proof of Theorem 1.

The main point of this section is to show that the Lagrangian duality relation between (33) and (39) can be used to establish uplink–downlink duality as developed in [11], [2], and [3]. This is because the dual minimax problem may be simplified into a single maximization problem corresponding to a multiple-access channel. Such a simplification enables significant computational saving in solving the original minimax problem.

The key is to focus on the water-filling condition for the dual channel (37) and to recognize the following two features. First, the noise covariance matrix in the dual problem is an identity matrix scaled by λ. Second, S_{zz} corresponds to the constraint

in the dual problem. Since S_{zz} has identity matrices on the diagonal, the input constraint on $\mathrm{tr}(\Phi S_{xx})$ is equivalent to a sum power constraint on $\mathrm{tr}(\Phi)$

$$\mathrm{tr}(\Phi S_{zz}) = \sum_{i=1}^{K} \mathrm{tr}(\Phi_i). \tag{40}$$

As stated in the next lemma, after a proper scaling of Φ, it is possible to show that (37) is precisely the KKT condition for a single maximization problem

$$\text{maximize} \quad \frac{1}{2} \log |H^T \Sigma H + I|$$
$$\text{subject to} \quad \Sigma \text{ is block diagonal}$$
$$\mathrm{tr}(\Sigma) \leq P \tag{41}$$

which corresponds to the input optimization problem for a multiple-access channel. The duality between the broadcast channel and the multiple-access channel is formally stated as follows. The statement for the case where H is square and invertible is given first. The general theorem is stated in the next section.

Lemma 1: Consider the Gaussian mutual information minimax problem for a broadcast channel (33) and the maximization problem for a multiple-access channel (41). Assume that H is square and invertible, the solution of the minimax problem (33), (S_{xx}, S_{zz}), is nonsingular, and the solution of the maximization problem (41), Σ, is nonsingular. Then, the primal-dual optimal solution $(S_{xx}, S_{zz}, \lambda, \Phi)$ of (33) may be obtained from the primal-dual optimal solution (Σ, ν) of the multiple-access channel (41) as follows:

$$S_{xx} = (\nu I)^{-1} - (H^T(\nu \Sigma)H + \nu I)^{-1} \tag{42}$$

$$S_{zz} = H(H^T(\nu \Sigma)H + \nu I)^{-1} H^T \tag{43}$$

$$\Phi = \nu \Sigma \tag{44}$$

$$\lambda = \nu. \tag{45}$$

Conversely, the primal-dual optimal solution of the multiple-access channel (Σ, ν) may be obtained from the dual solution of the broadcast channel (Φ, λ) via (44) and (45). Further, the broadcast channel and the multiple-access channel have the same sum capacity.

Proof: The development leading to the proof shows that the primal-dual optimal solution to the minimax problem $(S_{xx}, S_{zz}, \Phi, \lambda)$ are as expressed in (42) and (43). In particular, they satisfy

$$H(H^T(\Phi/\lambda)H + I)^{-1} H^T = \lambda S_{zz}. \tag{46}$$

Now, rewrite the multiple-access channel problem as follows:

$$\text{maximize} \quad \frac{1}{2} \log \left| \sum_{i=1}^{K} H_i^T \Sigma_i H_i + I \right|$$
$$\text{subject to} \quad \sum_{i=1}^{K} \mathrm{tr}(\Sigma_i) \leq P \tag{47}$$

where $H^T = [H_1^T \cdots H_K^T]$, $\Sigma = \mathrm{diag}\{\Sigma_1, \ldots, \Sigma_K\}$. The KKT condition for the above problem is

$$H_i \left(\sum_{i=1}^{K} H_i^T \Sigma_i H_i + I \right)^{-1} H_i^T = \nu I, \tag{48}$$

TABLE II
UPLINK–DOWNLINK DUALITY

Optimization Problem:	$\max\limits_{S_{xx}} \min\limits_{S_{zz}} \frac{1}{2}\log\frac{\|HS_{xx}H^T + S_{zz}\|}{\|S_{zz}\|}$	$\max\limits_{\Sigma} \frac{1}{2}\log\|H^T\Sigma H + I\|$
Constraints:	$\operatorname{tr}(S_{xx}) \le P$ $S_{zz}(i,i) = I$	$\operatorname{tr}(\Sigma) \le P$
Primal Variables:	$S_{xx} = (\nu I)^{-1} - (H^T(\nu\Sigma)H + \nu I)^{-1}$ $S_{zz} = H(H^T(\nu\Sigma)H + \nu I)^{-1}H^T$	$\Sigma = \Phi/\lambda$
Dual Variables:	$\Phi = \nu\Sigma$ $\lambda = \nu$	$\nu = \lambda$

for all $i = 1, \ldots, K$. Writing it in matrix form, the above is equivalent to

$$H(H^T\Sigma H + I)^{-1}H^T = \nu\Gamma \qquad (49)$$

where Γ is a matrix whose diagonal entries are identity matrices.

To establish that the sum capacities of the multiple-access channel and the broadcast channel are the same, we check that (46) and (49) are identical. This is equivalent to checking that $S_{zz} = \Gamma$, $\nu = \lambda$, and $\Sigma = \Phi/\lambda$. First, both S_{zz} and Γ have identity matrices on their respective diagonal terms, thus, they can be made equal. Second, the dual variables ν and λ only need to be positive, so they can be made equal also. Third, the constraint on Σ is $\operatorname{tr}(\Sigma) \le P$, and the constraint associated with (46) is of the form

$$\operatorname{tr}((\Phi/\lambda)S_{zz}) = \sum_i \operatorname{tr}(\Phi_i)/\lambda \le P'.$$

For the two constraints to be the same, it remains to verify that $P' = P$.

In convex optimization, the dual variables λ and Φ have the interpretation of being the sensitivities of the saddle point with respect to the constraints. Let $C(P, S_{z_1 z_1} \cdots S_{z_K z_K})$ denote the sum capacity of the Gaussian vector broadcast channel with power constraint P and noise covariance matrix S_{zz}. Then

$$\lambda = \left.\frac{\partial C(P, S_{z_1 z_1} \cdots S_{z_K z_K})}{\partial P}\right|_{(S_{xx}, S_{zz})} \qquad (50)$$

and

$$\Phi_i = -\left.\frac{\partial C(P, S_{z_1 z_1} \cdots S_{z_K z_K})}{\partial S_{z_i z_i}}\right|_{(S_{xx}, S_{zz})}. \qquad (51)$$

Now, consider the following thought experiment. Suppose that the power constraint of the minimax problem is relaxed from P to $P(1 + \delta)$

$$\operatorname{tr}(S_{xx}) \le P(1 + \delta) \qquad (52)$$

and the noise covariance matrix in each terminal is also relaxed from I to $I(1 + \delta)$

$$S_{z_i z_i} = I(1 + \delta) \qquad (53)$$

where δ is a small positive real number. Observe that since the proportional increases in signal and noise powers are the same, from the structure of the expression (33), the saddle point (S_{xx}, S_{zz}) would also be scaled by exactly the same proportion. Consequently, the capacity would remain unchanged. Now, because λ and Φ_i are the sensitivities of the minimax expression with respect to the constraints, this implies that

$$(\delta P)\lambda = \sum_{i=1}^K \operatorname{tr}(\Phi_i \delta I). \qquad (54)$$

Therefore,

$$\operatorname{tr}(\Sigma) = \sum_{i=1}^K \operatorname{tr}(\Phi_i)/\lambda. \qquad (55)$$

This verifies the power constraint on the dual multiple-access channel.

Finally, it remains to verify that the broadcast channel and the dual multiple-access channel have the same sum capacity. Using the dual variable relations (44) and (45), the following relation (shown in (56) at the bottom of the page) can be seen. This proves the lemma. $\qquad\square$

The relation between the multiple-access channel and the broadcast channel is summarized in Table II. This relationship shows that the solution to the minimax problem (33) can be obtained from the solution to a maximization problem (41) and *vice versa*. Clearly, (41) is much easier to solve, as it only involves a single maximization. Further, the optimization variable is block diagonal, therefore, it lies in a lower dimensional space. Thus, uplink–downlink duality gives an efficient way to solve the minimax problem (33) numerically.

C. Singular Least Favorable Noise

The Lagrangian interpretation of uplink–downlink duality suggests that the least favorable noise in the minimax problem can be computed via a multiple-access channel. In particular, (43) gives an explicit formula for the covariance matrix of the least favorable noise. However, (43) is derived assuming that

$$\log\frac{\|HS_{xx}H^T + S_{zz}\|}{\|S_{zz}\|} = \log\frac{\|H((\lambda I)^{-1} - (H^T(\lambda\Sigma)H + \lambda I)^{-1})H^T + H(H^T(\lambda\Sigma)H + \lambda I)^{-1}H^T\|}{\|H(H^T\Phi H + \lambda I)^{-1}H^T\|}$$
$$= \log\frac{\|H^T\Phi H + \lambda I\|}{\|\lambda I\|}$$
$$= \log\|H\Sigma H^T + I\|. \qquad (56)$$

the channel matrix H is square and invertible. Also, the least favorable noise covariance matrix, the maximizing input covariance matrix for the broadcast channel, and the maximizing input covariance matrix for the dual multiple-access channel are all assumed to be full rank. This is not necessarily true in general. In practice, it often happens that the maximizing block diagonal Σ in the dual multiple-access channel does not have positive definite entries on its diagonal. This is the case, for example, when H has more rows than columns.

Formally, the KKT condition for the multiple-access channel should have been

$$H(H^T\Sigma H + I)^{-1}H^T = \nu\Gamma + \Upsilon \qquad (57)$$

where Υ is a positive semidefinite matrix satisfying the complementary slackness condition

$$\text{tr}(\Sigma\Upsilon) = 0. \qquad (58)$$

If Σ is full rank, then Υ is the zero matrix. This ensures that the diagonal terms of $H(H^T(\nu\Sigma)H + \nu I)^{-1}H^T$ are all identity matrices. However, if Σ is low rank, then $H(H^T(\nu\Sigma)H + \lambda I)^{-1}H^T$ does not necessarily have identity matrices on its diagonal. Thus, it cannot be a valid choice for a least favorable noise. Another way to interpret this phenomenon is to note that the minimax duality relation is established with inequality constraints of the type $\text{tr}(S_{zz}Q_z) \leq 1$. However, the constraint associated with the broadcast channel is of the form

$$S_{zz}(i,i) = I \qquad (59)$$

rather than

$$S_{zz}(i,i) \leq I. \qquad (60)$$

Thus, to establish duality, an additional step must be taken to ensure that the diagonal entries are equal to identity matrices.

It turns out that a least-favorable-noise covariance matrix can be obtained by adding a positive semidefinite matrix to $H(H^T\Phi H + \lambda I)^{-1}H^T$ so that the sum of the two matrices has identity matrices on the diagonal. Further, such a positive semidefinite matrix is not unique, so the least-favorable-noise covariance matrix is not unique.

Consider the candidate least favorable noise, relabeled as $S_{zz}^{(0)}$

$$S_{zz}^{(0)} = H(H^T(\nu\Sigma)H + \nu I)^{-1}H^T. \qquad (61)$$

The water-filling input covariance matrix S_{xx} with respect to the channel H and the noise $S_{zz}^{(0)}$ is

$$S_{xx} = (\nu I)^{-1} - (H^T(\nu\Sigma)H + \lambda I)^{-1}. \qquad (62)$$

Now, S_{xx} may be rank deficient. In this case, the dimension of H must be reduced, and the least favorable noise must be recomputed using the reduced channel. The recomputed noise covariance matrix could be lower rank, thus enlarging the class of positive semidefinite matrices that can be added to its diagonal terms. Without loss of generality, it is assumed here that the channel reduction step has already taken place.

The expression for the least favorable noise (61) has appeared in [2] as the cost constraint of a dual multiple-access channel, although the singular noise issue was not specifically dealt with in [2]. The idea of forcing the diagonal terms to be identity

matrices has appeared in [3]. In [3], in order to prove the sum capacity result when the least favorable noise is singular, a sequence of nonsingular noises is constructed and the result is obtained in the limit. However, the result in [3] does not characterize the entire set of least favorable noises.

The following theorem characterizes the nature of duality and the class of least favorable noises for the minimax problem explicitly. The proof is included in Appendix B.

Theorem 2: Consider the Gaussian mutual information minimax problem for a broadcast channel (33) and the maximization problem for a multiple-access channel (41). The solution (S_{xx}, S_{zz}) of the broadcast channel problem (33) may be obtained from the primal and dual optimal solutions (Σ, ν) of the multiple-access channel (41) as follows:

$$S_{xx} = (\nu I)^{-1} - (H^T(\nu\Sigma)H + \nu I)^{-1} \qquad (63)$$

$$S_{zz} = H(H^T(\nu\Sigma)H + \nu I)^{-1}H^T + S'_{zz} \qquad (64)$$

where S'_{zz} is any positive semidefinite matrix that makes the diagonals of S_{zz} identity matrices. Further, the broadcast channel and the multiple-access channel have the same sum capacity.

The least favorable noise in a Gaussian minimax mutual information expression is of the form $S_{zz}^{(0)} + S'_{zz}$. The addition of the term S'_{zz} does not change the mutual information $I(\boldsymbol{X}; \boldsymbol{Y})$. In some sense, $S_{zz}^{(0)}$ is the "smallest" possible noise within the class of least favorable noises. The reason is the following. The first step in the proof of Theorem 2 is a reduction of the channel matrix to the space spanned by a set of orthogonal vectors. Let U_1 be a matrix whose columns are these orthogonal vectors. After the channel reduction step, $HS_{xx}H^T$ is strictly positive everywhere in the span of U_1. Thus, for the capacity to be finite, the noise covariance matrix must also be strictly positive in the space spanned by U_1. Now, $S_{zz}^{(0)}$ is the only noise covariance matrix entirely in the U_1 space that satisfies the KKT condition. Thus, $S_{zz}^{(0)}$ must be the "smallest" noise covariance matrix that satisfies the KKT condition.

The class of least favorable noises $S_{zz}^{(0)} + S'_{zz}$ has a linear estimation interpretation. Let $\boldsymbol{Z}^{(0)}$ and $\boldsymbol{Z}'$ be independent Gaussian noise vectors with covariance matrices $S_{zz}^{(0)}$ and S'_{zz} respectively. Consider an orthogonal transformation

$$\boldsymbol{Z}_u^{(0)} = [U_1 U_2]\boldsymbol{Z}^{(0)} \quad \text{and} \quad \boldsymbol{Z}'_u = [U_1 U_2]\boldsymbol{Z}'$$

which takes the form

$$\boldsymbol{Z}_u^{(0)} = \begin{bmatrix} \boldsymbol{z}_1^0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{Z}'_u = \begin{bmatrix} \boldsymbol{z}_1' \\ \boldsymbol{z}_2' \end{bmatrix}. \qquad (65)$$

Consider two receivers, one has $\boldsymbol{Z}_u^{(0)}$ as noise and the other has $\boldsymbol{Z}_u^{(0)} + \boldsymbol{Z}'_u$ as noise. Clearly, the second receiver cannot do better than the first one. However, the second receiver does as well as the first receiver if the minimum-mean-squared-error estimation of $\boldsymbol{z}_1^0 + \boldsymbol{z}_1'$ given $\boldsymbol{z}_2'$ is exactly $\boldsymbol{z}_1'$. This relation characterizes the entire set of least favorable noises.[1] Mathematically, the above estimation relation is also related to the proof of Theorem 2. Since $\boldsymbol{z}_1^0$ and $\boldsymbol{z}_1'$ are independent, the estimation relation

[1]There is a subtlety in that, strictly speaking, $\boldsymbol{Z}'$ need not be Gaussian. However, this does not affect the sum capacity. See [12] for a discussion.
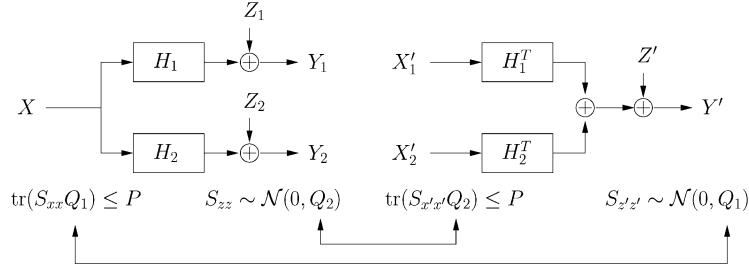
Fig. 1. Generalized uplink–downlink duality.

is equivalent to the condition $\mathbb{E}[\boldsymbol{z}_1'|\boldsymbol{z}_2'] = 0$. But, the covariance matrix of $\mathbb{E}[\boldsymbol{z}_1'|\boldsymbol{z}_2']$ is $S_{11} - S_{12}S_{22}^{-1}S_{21}$, where S_{11}, S_{22} and S_{12}, S_{21} are the entries of the covariance matrices S_{zz}'. So, $S_{11} - S_{12}S_{22}^{-1}S_{21} = 0$. This relation is crucial in the proof of Theorem 2.

IV. GENERALIZED DUALITY

A. Duality for Broadcast Channels With an Arbitrary Linear Constraint

An important benefit of the new Lagrangian interpretation of uplink–downlink duality is that it allows uplink–downlink duality to be generalized to Gaussian vector broadcast channels with arbitrary linear constraints. Such a generalization is useful from a computational perspective, because the multiple-access channel capacity (41) is considerably easier to compute than the minimax problem (33). This generalization of uplink–downlink duality to channels with arbitrary input constraints cannot be obtained directly from the previous proofs of uplink–downlink duality given in [11], [3] and [2]. This is because previous proofs of uplink–downlink duality depend critically on a transformation of SINRs which holds only for the sum power constraint. However, starting from the new Lagrangian interpretation of uplink–downlink duality the generalization is almost immediate.

Consider a broadcast channel with a linear covariance input constraint of the form

$$\text{tr}(S_{xx}Q) \leq P. \tag{66}$$

In this case, the noise covariance matrix I in the dual multiple-access channel (41) is replaced by the covariance matrix Q

$$C = \max_{\Sigma} \frac{1}{2} \log |H^T \Sigma H + Q|. \tag{67}$$

The same power constraint applies as before: $\text{tr}(\Sigma) \leq P$. Thus, the broadcast channel capacity with the new constraint may be computed by simply replacing the noise covariance matrix in the dual channel by the new input constraint matrix Q.

There is an interesting interplay between the cost constraint in the original channel and the noise covariance in the dual channel. In [2], it is shown that the least-favorable-noise covariance in the Gaussian broadcast channel may be transformed into a cost constraint for the dual multiple-access channel. The preceding derivation shows that the same relation holds between the cost constraint of the broadcast channel and the noise covariance of the multiple-access channel. Fig. 1 illustrates the transformation between the input cost constraint of a broadcast channel

and the noise covariance of the dual multiple-access channel. The relation is summarized in the following theorem.

Theorem 3: The dual of the Gaussian vector broadcast channel with an input constraint $\text{tr}(S_{xx}Q) \leq P$ is a multiple-access channel with a noise covariance matrix Q. The optimal solution (S_{xx}, S_{zz}) of the minimax problem corresponding to the broadcast channel may be obtained from the primal-dual optimal solution (Σ, ν) of the multiple-access channel as follows:

$$S_{xx} = (\nu I)^{-1} - (H^T(\nu\Sigma)H + \nu Q)^{-1} \tag{68}$$
$$S_{zz} = H(H^T(\nu\Sigma)H + \nu Q)^{-1}H^T + S_{zz}' \tag{69}$$

where S_{zz}' is any positive semidefinite matrix that makes the diagonals of S_{zz} identity matrices. Further, the broadcast channel and the multiple-access channel have the same sum capacity.

A key requirement for the duality between the broadcast channel and the multiple-access channel to hold is the linearity of the constraint. Without linearity, the dual of the minimax problem does not reduce to a single maximization problem. Consider a broadcast channel with an arbitrary convex constraint of the form

$$f(S_{xx}) \leq 0. \tag{70}$$

Its sum capacity can still be shown to be

$$\max_{S_{xx}} \min_{S_{zz}} \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|}. \tag{71}$$

(This is because the following two statements: min-max equals max-min and saddle point is Gaussian for any convex–concave function, hold under arbitrary convex input constraints.) The KKT condition of the minimax problem implies that

$$H(H^T\Phi H + \lambda Q)^{-1}H^T = S_{zz} \tag{72}$$

where Q is the gradient of $f(\cdot)$ at the saddle point S_{xx}. It is still possible to write $\Sigma = \Phi/\lambda$ and to formulate the dual multiple-access channel problem as

$$C = \max_{\Sigma} \frac{1}{2} \log |H^T \Sigma H + Q|. \tag{73}$$

However, without the linearity of $f(\cdot)$, the power constraint derivation (50)–(55) does not follow. Further, the dual noise covariance matrix $Q = f'(\cdot)$ now depends on the solution of the minimax problem, which is not known before the minimax problem (33) is explicitly solved. Therefore, although minimax duality still exists, it is not useful computationally. In this sense, the minimax expression (33) is a more fundamental charac-

terization of the Gaussian vector broadcast channel sum capacity. The minimax characterization applies to broadcast channels with any convex input covariance constraints, while uplink–downlink duality holds only when the input constraint is linear.

B. Duality for Broadcast Channels With Per-Antenna Power Constraints

Finally, we discuss the situation in which multiple-input constraints are applied at the input. Consider a broadcast channel with L linear input covariance constraints

$$\mathrm{tr}(S_{xx}Q_1) \leq 1$$
$$\vdots$$
$$\mathrm{tr}(S_{xx}Q_L) \leq 1. \qquad (74)$$

Clearly, the dual noise is now $\sum_{l=1}^{L} \lambda_l Q_l$. Again, the dual noise is not directly available without solving the original minimax problem. However, in many practical situations, it may be easier to solve the dual minimax problem instead. Recall that the KKT condition associated with the dual problem is

$$H(H^T\Phi H + \lambda_1 Q_1 + \cdots + \lambda_L Q_L)^{-1}H^T = S_{zz} \qquad (75)$$

and

$$(\lambda_1 Q_1 + \cdots + \lambda_L Q_L)^{-1} - (H^T\Phi H + \lambda_1 Q_1 + \cdots + \lambda_L Q_L)^{-1}$$
$$= S_{xx}. \qquad (76)$$

The dual minimax problem is therefore

$$\max_{\Phi} \min_{\lambda_1 \cdots \lambda_L} \quad \frac{1}{2} \log \frac{\left| H^T\Phi H + \sum_{l=1}^{L} \lambda_l Q_l \right|}{\left| \sum_{l=1}^{L} \lambda_l Q_l \right|}$$

$$\text{subject to} \quad \mathrm{tr}(S_{zz}\Phi) \leq 1$$
$$\mathrm{tr}\left(S_{xx} \sum_{l=1}^{L} \lambda_l Q_l \right) \leq 1$$
$$\Phi, \lambda \geq 0. \qquad (77)$$

In the following, we describe one example of practical importance for which the above dual problem is easier to solve than the original minimax problem.

In many practical wireless and wireline downlink applications, individual per-antenna power constraints, rather than a sum power constraint, are imposed on the transmit antennas. The computation of capacity for such a channel has been considered in the past [13], but only suboptimal solutions are known. The duality relation illustrated in this paper is ideally suited to handle this case.

The per-antenna power constraints are essentially constraints on the diagonal terms of S_{xx}

$$S_{xx}(i, i) \leq P_i, \qquad i = 1 \cdots n. \qquad (78)$$

Thus, each Q_i in (74) is a matrix with $1/P_i$ on its ith diagonal term and zero elsewhere. Define

$$Q = \lambda_1 Q_1 + \cdots + \lambda_n Q_n = \mathrm{diag}(\lambda_1/P_1, \ldots, \lambda_n/P_n). \qquad (79)$$

The KKT conditions (75) and (76) imply that the dual minimax problem is

$$\min_{Q} \max_{\Phi} \quad \frac{1}{2} \log \frac{|H^T\Phi H + Q|}{|Q|}$$

$$\text{subject to} \quad \Phi, Q \text{ are block diagonal}$$
$$\mathrm{tr}(\Phi S_{zz}) \leq 1$$
$$\mathrm{tr}(Q S_{xx}) \leq 1. \qquad (80)$$

Recall that S_{zz} is a diagonal matrix. So, the constraint $\mathrm{tr}(\Phi S_{zz}) \leq 1$ reduces to a single trace constraint $\mathrm{tr}(\Phi) \leq 1$. Also, as Q is diagonal and S_{xx} has P_i on its diagonal terms, the constraint $\mathrm{tr}(Q S_{xx}) \leq 1$ reduces to $\sum_{i=1}^{n} \lambda_i \leq 1$, which is equivalent to $\sum_{i=1}^{n} Q_{ii}P_i \leq 1$ where Q_{ii} is the ith diagonal term of Q. The dual minimax problem is computationally much easier to solve than the original problem. This is because both Φ and Q are now diagonal matrices, thus they lie in a lower dimensional space.

In addition, (80) illustrates that the uplink dual problem in this case is itself a minimax Gaussian mutual information problem. This minimax expression can be interpreted as a multiple-access channel with an uncertain noise. The transmitter chooses the best diagonal input covariance to maximize the mutual information. Nature chooses a least favorable diagonal noise subject to a linear constraint to minimize the mutual information.

Theorem 4: The sum capacity of a Gaussian multiple-antenna broadcast channel with individual per-antenna transmit power constraints $P_1, P_2, \ldots, P_n$ is the same as the sum capacity of a dual multiple-access channel with a sum power constraint and with an uncertain noise. The sum power constraint of the dual channel is 1. The uncertain noise must have a diagonal covariance matrix Q with its diagonal values constrained by $\sum_i Q_{ii}P_i \leq 1$.

V. CONCLUSION

This paper illustrates a Lagrangian duality relation for a Gaussian mutual information minimax optimization problem. A central feature of the minimax duality relation is that the cost constraint in the original problem becomes the noise covariance in the dual problem and *vice versa*. This duality relation is particularly simple for the Gaussian vector broadcast channel for which a block-diagonal constraint applies to the noise covariance matrix and a single power constraint applies to the input covariance matrix. In this case, the dual minimax optimization problem reduces to a single maximization problem, which corresponds to a multiple-access channel. This approach shows that the uplink–downlink duality between a broadcast channel and a multiple-access channel can be derived directly from the Lagrangian theory, thus unifying several seemingly unrelated previous proofs of the sum capacity of the Gaussian vector broadcast channel. The new minimax Lagrangian duality approach also allows an explicit characterization of the saddle point in the minimax problem. It shows that the least favorable noise can be singular and is not unique.

Uplink–downlink duality may be generalized to broadcast channels with arbitrary linear constraints. In particular, it is shown that the dual of a Gaussian vector broadcast channel with individual per-antenna power constraint is a Gaussian multiple-access channel with an uncertain noise, where the noise covariance matrix is diagonally constrained. However, under general nonlinear convex constraints, duality breaks down. In this sense, the minimax derivation for the broadcast

channel sum capacity is more general than the derivation based on uplink–downlink duality.

APPENDIX A
PROOF OF THEOREM 1

The derivation leading to the statement of Theorem 1 gives a proof of the minimax duality relation for the case where H is square and invertible and (S_{xx}, S_{zz}) is full rank. The goal here is to show that minimax duality holds in general. Suppose that the optimal solution (S_{xx}, S_{zz}) for the minimax problem

$$\max_{S_{xx}} \min_{S_{zz}} \quad \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|} \tag{81}$$

$$\text{subject to} \quad \text{tr}(S_{xx}Q_x) \leq 1 \tag{82}$$

$$\text{tr}(S_{zz}Q_z) \leq 1 \tag{83}$$

is not full rank. Write

$$S_{xx} = U_1 S_{\tilde{x}\tilde{x}} U_1^T, \quad S_{zz} = U_2 S_{\tilde{z}\tilde{z}} U_2^T, \tag{84}$$

where $S_{\tilde{x}\tilde{x}}$ and $S_{\tilde{z}\tilde{z}}$ are full-rank matrices, and U_1 and U_2 are matrices consisting of orthonormal column vectors (so that $U_1^T U_1 = I$ and $U_2^T U_2 = I$). The idea is to apply minimax duality to $S_{\tilde{x}\tilde{x}}$ and $S_{\tilde{z}\tilde{z}}$. Write

$$Q_{\tilde{x}} = U_1^T Q_x U_1, \quad Q_{\tilde{z}} = U_2^T Q_z U_2. \tag{85}$$

Since S_{zz} is low rank, $HS_{xx}H^T$ must also be low rank, and in particular, must be contained in the space spanned by U_2. Thus, it is possible to write

$$H = U_2 \tilde{H} U_1. \tag{86}$$

Then, the minimax problem can be reformulated as follows:

$$\max_{S_{\tilde{x}\tilde{x}}} \min_{S_{\tilde{z}\tilde{z}}} \quad \frac{1}{2} \log \frac{|\tilde{H}S_{\tilde{x}\tilde{x}}\tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|} \tag{87}$$

$$\text{subject to} \quad \text{tr}(S_{\tilde{x}\tilde{x}}Q_{\tilde{x}}) \leq 1 \tag{88}$$

$$\text{tr}(S_{\tilde{z}\tilde{z}}Q_{\tilde{z}}) \leq 1. \tag{89}$$

First, we claim that $\tilde{H}S_{\tilde{x}\tilde{x}}\tilde{H}^T$ and $S_{\tilde{z}\tilde{z}}$ must have the same rank. The reason is as follows: $S_{\tilde{z}\tilde{z}}$ cannot have a lower rank than $\tilde{H}S_{\tilde{x}\tilde{x}}\tilde{H}^T$, as otherwise, the minimax expression becomes infinite. But, $\tilde{H}S_{\tilde{x}\tilde{x}}\tilde{H}^T$ cannot have a lower rank than $S_{\tilde{z}\tilde{z}}$, either. This is because if this happens, it must be possible to further reduce the rank of $S_{\tilde{z}\tilde{z}}$ while keeping the objective the same and reducing the constraint $\text{tr}(S_{\tilde{z}\tilde{z}}Q_{\tilde{z}})$ at the same time. (Note, $Q_{\tilde{z}}$ must be full rank, as otherwise, the minimizing $S_{\tilde{z}\tilde{z}}$ would have been unbounded.) Since the optimal $S_{\tilde{z}\tilde{z}}$ cannot have slack in the constraint, a higher rank $S_{\tilde{z}\tilde{z}}$ cannot be optimal.

Second, we claim that $S_{\tilde{x}\tilde{x}}$ and $S_{\tilde{z}\tilde{z}}$ must have the same rank, and $\tilde{H}$ must be square and invertible. The reason is as follows. The rank of $S_{\tilde{x}\tilde{x}}$ is at least as high as the rank of $\tilde{H}S_{\tilde{x}\tilde{x}}\tilde{H}^T$. However, if it were strictly higher, it would be possible to reduce the rank of $S_{\tilde{x}\tilde{x}}$ while keeping the objective the same and reducing the constraint $\text{tr}(S_{\tilde{x}\tilde{x}}Q_{\tilde{x}})$ at the same time. (Again, $Q_{\tilde{x}}$ must be full rank, as otherwise, the maximizing $S_{\tilde{x}\tilde{x}}$ would have been unbounded.) Again, since the optimal $S_{\tilde{x}\tilde{x}}$ cannot have slack in the constraint, a higher rank $S_{\tilde{x}\tilde{x}}$ cannot be optimal. Finally, the fact that the ranks of $S_{\tilde{x}\tilde{x}}$ and $\tilde{H}S_{\tilde{x}\tilde{x}}\tilde{H}^T$ are equal implies that $\tilde{H}$ is square and invertible.

The channel reduction step does not affect the dual variables λ_x and λ_z. Now, apply the minimax duality result to the reduced channel (87). Let $\Psi_{\tilde{z}} = S_{\tilde{x}\tilde{x}}/\lambda_x$ and $\Psi_{\tilde{x}} = S_{\tilde{z}\tilde{z}}/\lambda_z$. The dual minimax problem is of the form

$$\max_{\Sigma_{\tilde{x}\tilde{x}}} \min_{\Sigma_{\tilde{z}\tilde{z}}} \quad \frac{1}{2} \log \frac{|\tilde{H}^T \Sigma_{\tilde{x}\tilde{x}} \tilde{H} + \Sigma_{\tilde{z}\tilde{z}}|}{|\Sigma_{\tilde{z}\tilde{z}}|} \tag{90}$$

$$\text{subject to} \quad \text{tr}(\Sigma_{\tilde{x}\tilde{x}}\Psi_{\tilde{x}}) \leq 1 \tag{91}$$

$$\text{tr}(\Sigma_{\tilde{z}\tilde{z}}\Psi_{\tilde{z}}) \leq 1. \tag{92}$$

The solution of the dual problem is precisely

$$\Sigma_{\tilde{x}\tilde{x}} = \lambda_z Q_{\tilde{z}}, \quad \Sigma_{\tilde{z}\tilde{z}} = \lambda_x Q_{\tilde{x}}. \tag{93}$$

Now, define

$$\Sigma_{xx} = U_2^T \Sigma_{\tilde{x}\tilde{x}} U_2, \quad \Sigma_{zz} = U_1^T \Sigma_{\tilde{z}\tilde{z}} U_1 \tag{94}$$

and

$$\Psi_x = U_2^T \Psi_{\tilde{x}} U_2, \quad \Psi_z = U_1^T \Psi_{\tilde{z}} U_1. \tag{95}$$

Then, the reduced dual minimax problem may be rewritten as

$$\max_{\Sigma_{xx}} \min_{\Sigma_{zz}} \quad \frac{1}{2} \log \frac{|H^T \Sigma_{xx} H + \Sigma_{zz}|}{|\Sigma_{zz}|} \tag{96}$$

$$\text{subject to} \quad \text{tr}(\Sigma_{xx}\Psi_x) \leq 1 \tag{97}$$

$$\text{tr}(\Sigma_{zz}\Psi_z) \leq 1. \tag{98}$$

It is easy to verify that

$$C(H, Q_x, Q_z) = C(\tilde{H}, Q_{\tilde{x}}, Q_{\tilde{z}}) = C(\tilde{H}^T, \Psi_{\tilde{x}}, \Psi_{\tilde{z}})$$
$$= C(H^T, \Psi_x, \Psi_z) \tag{99}$$

and

$$\lambda_x = \nu_z, \quad \lambda_z = \nu_x \tag{100}$$

and

$$(S_{xx}, S_{zz}) = (\nu_z \Psi_z, \nu_x \Psi_x), \quad (\Sigma_{xx}, \Sigma_{zz}) = (\lambda_z Q_z, \lambda_x Q_x). \tag{101}$$

Therefore, minimax duality holds even if H is low rank and (S_{xx}, S_{zz}) are singular. $\square$

APPENDIX B
PROOF OF THEOREM 2

The first part of the proof is based on an earlier version of the proof [14] where the case in which each receiver has a single antenna is presented. The extension to the multiple-antenna case is presented at the end of this proof. In the rest of the proof, $\nu\Sigma$ is denoted as Φ for notational convenience.

Consider first the single-antenna case. The first step of the proof is to verify that even though the candidate noise covariance matrix (61) may not have diagonal terms that are identity matrices (or in the single-antenna case, 1's), it satisfies the least favorable-noise condition for the minimax problem. Assuming that the channel has already been reduced, the main idea is to represent $S_{zz}^{(0)}$ in its eigenvalue decomposition

$$S_{zz}^{(0)} = U_1 S_{\tilde{z}\tilde{z}} U_1^T \tag{102}$$

where $S_{\tilde{z}\tilde{z}}$ is invertible and columns of U_1 are orthonormal vectors. Further, H can be rewritten as

$$H = U_1 \tilde{H} \tag{103}$$

where $\tilde{H}$ is invertible. Then, the minimax problem is now reduced to

$$\max_{S_{xx}} \min_{S_{\tilde{z}\tilde{z}}} \quad \frac{1}{2} \log \frac{|\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|}$$
$$\text{subject to} \quad \text{tr}(S_{xx}) \le P$$
$$U_1 S_{\tilde{z}\tilde{z}} U_1^T \text{ has 1's on the diagonal}$$
$$S_{xx} \ge 0. \tag{104}$$

Since the KKT conditions for the reduced problem are

$$\tilde{H}^T (\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} \tilde{H} = \lambda I \tag{105}$$
$$S_{\tilde{z}\tilde{z}}^{-1} - (\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} = U_1^T \Phi U_1. \tag{106}$$

The candidate $S_{zz}^{(0)}$ is therefore the least favorable noise for the reduced problem

$$S_{zz}^{(0)} = U_1 S_{\tilde{z}\tilde{z}} U_1^T = H(H^T \Phi H + \lambda I)^{-1} H^T. \tag{107}$$

With $S_{zz}^{(0)}$ as the noise covariance matrix, the optimal water-filling covariance matrix S_{xx} is

$$S_{xx} = (\nu I)^{-1} - (H^T(\nu \Sigma) H + \nu I)^{-1}. \tag{108}$$

The rest of the proof is devoted to showing that when a positive semidefinite matrix, denoted as S'_{zz}, is added to $S_{zz}^{(0)}$ to makes its diagonal terms 1's, both the least favorable-noise condition and the water-filling condition remain to be satisfied. In fact, the entire class of saddle points is precisely $(S_{xx}, S_{zz}^{(0)} + S'_{zz})$.

The proof is divided into several parts. First, without loss of generality, the rows of H can be rearranged so that the upper left submatrix of $S_{zz}^{(0)}$ has 1's on its diagonal. This implies that Φ only has positive entries on its upper left diagonal, and S'_{zz} has nonzero entries only in the lower right corner

$$\Phi = \begin{bmatrix} \Phi' & 0 \\ 0 & 0 \end{bmatrix}, \qquad S'_{zz} = \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix}. \tag{109}$$

Note that S is a full-rank matrix. Let n be the number of receivers in the broadcast channel. So, Φ, S_{zz} and S'_{zz} are $n \times n$ matrices. Let the dimension of Φ' be $k \times k$. So, the dimension of S is $(n-k) \times (n-k)$. Also, let the rank of $S_{zz}^{(0)}$ be r. Now, observe that $k \ge r$. The reason for this has to do with the channel reduction step mentioned earlier. Note that from (62) the optimal S_{xx} can be expressed as

$$S_{xx} = (\lambda I)^{-1} - (H^T \Phi H + \lambda I)^{-1}$$
$$= (\lambda I)^{-1} H^T \sqrt{\Phi} (I + \sqrt{\Phi} H H^T \sqrt{\Phi})^{-1} \sqrt{\Phi} H (\lambda I)^{-1}. \tag{110}$$

Thus, the rank of S_{xx} is the same as the rank of $H^T \Phi H$. The channel reduction step guarantees that S_{xx} is full rank. This implies that $H^T \Phi H$ must be full rank. For this to be true, Φ must have nonzero diagonal entries in at least r' positions, where r'

is the rank of H. So, $k \ge r'$. But, after the channel reduction, the rank of H is the same as the rank of $S_{zz}^{(0)}$. This proves that $k \ge r$.

In general, k can be strictly larger than r. Physically, this implies that in a broadcast channel, the number of active receivers can be larger than the number of transmit dimensions. Also, note that since the rank of $S_{zz}^{(0)}$ is r and the rank of S'_{zz} is $(n-k)$, the rank of $S_{zz}^{(0)} + S'_{zz}$ is at most $(n-k+r)$, which is not full rank if k is strictly larger than r.

The next step of the proof involves the decomposition of $S_{zz}^{(0)} + S'_{zz}$ along the direction U_1. The strategy is the following. First, find an $n \times (n-k)$ matrix of orthonormal column vectors, denoted as U_2, extending the space spanned by the columns of U_1, such that S'_{zz} can be expressed as

$$S'_{zz} = \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}. \tag{111}$$

Recall that U_1 contains r column vectors and U_2 contains $n-k$ column vectors. So, $\begin{bmatrix} U_1 & U_2 \end{bmatrix}$ contains $n-k+r$ vectors, which do not necessarily span the whole space. Note also that S_{22} is a full-rank square matrix.

Partition U_1 and U_2 into submatrices

$$U_1 = \begin{bmatrix} u_{11}^T \\ u_{12}^T \end{bmatrix}, \quad U_2 = \begin{bmatrix} u_{21}^T \\ u_{22}^T \end{bmatrix}. \tag{112}$$

Two useful facts about S_{ij} and u_{ij} are derived next. First

$$S_{11} - S_{12} S_{22}^{-1} S_{21} = 0. \tag{113}$$

This is because

$$\begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix} = I. \tag{114}$$

So, from (111)

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} u_{11}^T & u_{12}^T \\ u_{21}^T & u_{22}^T \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix}. \tag{115}$$

This allows S_{ij} to be solved explicitly

$$S_{11} = u_{12}^T S u_{12} \qquad S_{12} = u_{12}^T S u_{22} \tag{116}$$
$$S_{21} = u_{22}^T S u_{12} \qquad S_{22} = u_{22}^T S u_{22}. \tag{117}$$

Therefore,

$$S_{12} S_{22}^{-1} S_{21} = u_{12}^T S u_{22} (u_{22}^T S u_{22})^{-1} u_{22}^T S u_{12} = S_{11} \tag{118}$$

thus proving (113).

The second useful fact is the following:

$$u_{21} = -u_{11} u_{12}^T u_{22}^{-T}. \tag{119}$$

The proof is based on (111)

$$\begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} u_{11}^T & u_{12}^T \\ u_{21}^T & u_{22}^T \end{bmatrix}. \tag{120}$$

Multiply out the right-hand side. Using the fact that the upper right submatrix of the right-hand side is zero, we get

$$0 = u_{11} S_{11} u_{12}^T + u_{21} S_{21} u_{12}^T + u_{11} S_{12} u_{22}^T + u_{21} S_{22} u_{22}^T. \tag{121}$$

Substituting in (116) and (117), (121) is equal to

$$0 = (u_{11} u_{12}^T + u_{21} u_{22}^T) S (u_{11} u_{11}^T + u_{22} u_{22}^T). \tag{122}$$

Since S is full rank and $(u_{11}u_{11}^T + u_{22}u_{22}^T)$ is full rank, the preceding implies

$$u_{11}u_{12}^T + u_{21}u_{22}^T = 0 \qquad (123)$$

thus proving (119).

Finally, we are ready to show that $S_{zz}^{(0)} + S_{zz}'$ satisfies the least favorable-noise condition. Starting from (106)

$$S_{\tilde{z}\tilde{z}}^{-1} - (\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} = U_1^T \Phi U_1 \qquad (124)$$

the objective is to prove that

$$\begin{bmatrix} S_{\tilde{z}\tilde{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1} - \begin{bmatrix} \tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \Phi \begin{bmatrix} U_1 & U_2 \end{bmatrix}. \qquad (125)$$

The strategy is to simplify the above using Schur's complement formula

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \Delta_D^{-1} & -\Delta_D^{-1}BD^{-1} \\ -D^{-1}C\Delta_D^{-1} & D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \end{bmatrix} \qquad (126)$$

where $\Delta_D = A - BD^{-1}C$. To evaluate the matrix inversions in (125), Δ_D^{-1} can be simplified using (113)

$$\Delta_D = S_{\tilde{z}\tilde{z}} + S_{11} - S_{12}S_{22}^{-1}S_{21} = S_{\tilde{z}\tilde{z}}. \qquad (127)$$

Thus, the first matrix inversion in (125) is equal to

$$\begin{bmatrix} S_{\tilde{z}\tilde{z}}^{-1} & -S_{\tilde{z}\tilde{z}}^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}S_{\tilde{z}\tilde{z}}^{-1} & S_{22}^{-1} + S_{22}^{-1}S_{21}S_{\tilde{z}\tilde{z}}^{-1}S_{12}S_{22}^{-1} \end{bmatrix} \qquad (128)$$

and the second matrix inversion in (125) can be expanded similarly in (129) at the bottom of the page. Now, using (124), the difference between the two can be simplified

$$\begin{bmatrix} U_1^T \Phi U_1 & -U_1^T \Phi U_1 S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}U_1^T \Phi U_1 & S_{22}^{-1}S_{21}U_1^T \Phi U_1 S_{12}S_{22}^{-1} \end{bmatrix}. \qquad (130)$$

To prove (125), it remains to show that the above is equal to

$$\begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \Phi \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} U_1^T \Phi U_1 & U_1^T \Phi U_2 \\ U_2^T \Phi U_1 & U_2^T \Phi U_2 \end{bmatrix}. \qquad (131)$$

Comparing (130) with (131), it is clear that the two are equal if the following holds:

$$\Phi U_2 = -\Phi U_1 S_{12}S_{22}^{-1}. \qquad (132)$$

Recall that Φ has nonzero entries only in its upper left diagonal. So, the left-hand side of (132) is

$$\Phi U_2 = \begin{bmatrix} \Phi' & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = \Phi' u_{21}. \qquad (133)$$

The right-hand side of (132) is

$$-\Phi U_1 S_{12}S_{22}^{-1} = -\Phi' u_{11}u_{12}^T u_{22}^{-T}. \qquad (134)$$

By (119), the left-hand side is equal to the right-hand side. This establishes the least favorable-noise condition (125).

Now, we verify that the broadcast channel and the multiple-access channel have the same sum capacity. First, the relation (135) at the bottom of the page can be easily verified. It remains to prove that substituting $S_{zz}^{(0)} + S_{zz}'$ for $S_{zz}^{(0)}$ in the above does not change capacity. With $(S_{xx}, S_{zz}^{(0)} + S_{zz}')$, the capacity is

$$C = \frac{1}{2} \log \frac{\begin{vmatrix} \tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}}{\begin{vmatrix} S_{\tilde{z}\tilde{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}}. \qquad (136)$$

Using the relation $S_{11} - S_{12}S_{22}^{-1}S_{21} = 0$ and the Schur's complement formula for the determinant

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D| \cdot |A - BD^{-1}C| \qquad (137)$$

it is not difficult to see that

$$\frac{|S_{22}| \cdot |\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}} + S_{11} - S_{12}S_{22}^{-1}S_{21}|}{|S_{22}| \cdot |S_{\tilde{z}\tilde{z}} + S_{11} - S_{12}S_{22}^{-1}S_{21}|} = \frac{|\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|}. \qquad (138)$$

Therefore, the capacity expressions (135) and (136) have the same value. In fact, this also implies that S_{xx} is the water-filling covariance matrix for the entire class of noise covariances matrices $S_{zz}^{(0)} + S_{zz}'$. This is so because the addition of S_{zz}' can only reduce the minimax capacity. The fact that it does not shows that S_{xx} must be the maximizing covariance matrix for all $S_{zz}^{(0)} + S_{zz}'$ simultaneously. The earlier part of the proof shows that all $S_{zz}^{(0)} + S_{zz}'$ satisfy the least favorable-noise condition with respect to S_{xx} simultaneously. Combining the two results, we conclude that saddle points of the minimax problem are precisely the class of covariance matrices $(S_{xx}, S_{zz}^{(0)} + S_{zz}')$. This proves the theorem for the single receiver antenna case.

$$\begin{bmatrix} (\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} & -(\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}(\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} & S_{22}^{-1} + S_{22}^{-1}S_{21}(\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1}S_{12}S_{22}^{-1} \end{bmatrix}. \qquad (129)$$

$$\log \frac{|HS_{xx}H^T + S_{zz}^{(0)}|}{|S_{zz}^{(0)}|} = \log \frac{|\tilde{H}S_{xx}\tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|}$$
$$= \log \frac{|\tilde{H}((\lambda I)^{-1} - (H^T(\lambda\Sigma)H + \lambda I)^{-1})\tilde{H}^T + \tilde{H}(H^T(\lambda\Sigma)H + \lambda I)^{-1}\tilde{H}^T|}{|\tilde{H}(H^T\Phi H + \lambda I)^{-1}\tilde{H}^T|}$$
$$= \log \frac{|H^T\Phi H + \lambda I|}{|\lambda I|}$$
$$= \log |H\Sigma H^T + I|. \qquad (135)$$

To generalize the result to the multiple-antenna case, consider the candidate least favorable noise $S_{zz}^{(0)}$, with block-diagonal terms $S_{zz}^{(0)}(i,i)$, which are not necessarily identity matrices. Here, $(S_{xx}, S_{zz}^{(0)})$ satisfies the saddle-point condition. We need to show that the saddle-point KKT conditions remain to be satisfied if a positive semidefinite matrix is added to $S_{zz}^{(0)}$ to make the diagonal terms identity matrices. The strategy is to whiten the block-diagonal terms of $S_{zz}^{(0)}$ and to consider a new broadcast channel with whitening filters at each receiver. Let

$$S_{zz}^{(0)}(i,i) = V^T S_{\tilde{z}\tilde{z}}^{(0)}(i,i) V \tag{139}$$

be the eigenvalue decomposition of $S_{zz}^{(0)}(i,i)$, where $S_{\tilde{z}\tilde{z}}^{(0)}(i,i)$ is a diagonal matrix of eigenvalues. Since $S_{zz}^{(0)}(i,i) \leq I$, the eigenvalues are all less than or equal to 1. The new broadcast channel is of the form

$$\boldsymbol{Y}_i' = V H_i \boldsymbol{X} + V \boldsymbol{Z}_i. \tag{140}$$

Now, if we find the equivalent of $S_{zz}^{(0)}$ for the above new broadcast channel, it would have had a diagonal structure with some diagonal entries being equal to 1 and others being less than 1. In such a new broadcast channel, receiver coordination is not needed within each $\boldsymbol{Y}_i'$. Thus, $\boldsymbol{Y}_i'$ can be split into multiple receivers each equipped with a single antenna only to which the previous saddle-point characterization applies. An inverse transformation V^T on S_{zz} recovers the saddle point for the original problem. This concludes the proof for the general multireceive antenna case.   □

REFERENCES

[1] E. Telatar, "Capacity of the multiple antenna Gaussian channel," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[2] P. Viswanath and D. N. C. 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, Jul. 2003.

[3] 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.

[4] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, to be published.

[5] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiple-antenna Gaussain broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[6] 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.

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

[8] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 374–377, May 1978.

[9] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

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

[11] 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. 768–783, May 2004.

[12] S. N. Diggavi and T. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[13] S. Jafar and A. Goldsmith, "Transmitter optimization for multiple antenna cellular systems," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 50.

[14] W. Yu, "The structure of the least-favorable noise in Gaussian vector broadcast channels," in *Advances in Network Information Theory: DIMACS Workshop on Network Information Theory*, P. Gupta, G. Kramer, and A. J. van Wijngaarden, Eds. Providence, RI: Amer. Math. Soc., 2004.