

Uplink-Downlink Duality via Minimax Duality

Wei Yu

Abstract

The sum capacity of a Gaussian vector broadcast channel is the saddle point of a Gaussian mutual information game where the transmitter maximizes the mutual information by choosing the best transmit covariance matrix subject to a power constraint and the receiver minimizes the mutual information by choosing a least-favorable noise covariance matrix subject to a diagonal constraint. This result has been proved using two different approaches: decision-feedback equalization and uplink-downlink duality. 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 to be characterized in terms of the dual variables. It also reveals the structure of the least-favorable noise: the least-favorable noise is not unique and different least-favorable noise covariance matrices are related to each other via a linear estimation relation. Further, the new Lagrangian duality interpretation allows uplink-downlink duality to be generalized to Gaussian vector broadcast channels with arbitrary linear constraints. In particular, it is shown that the dual of a broadcast channel with individual per-antenna power constraint is a multiple-access channel with a diagonal uncertain noise. Finally, we observe that 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 duality representation.

Keywords

Broadcast channels, Multiple-access channel, Channel capacity, Duality, Gaussian channels, Minimax optimization, Multiuser channels, Multiple-antenna systems, Optimization methods

I. INTRODUCTION

There is a curious input-output reciprocity for Gaussian vector channels. Consider a Gaussian vector channel under a power constraint:

$$\mathbf{Y} = H\mathbf{X} + \mathbf{Z}, \quad (1)$$

where \mathbf{X} and \mathbf{Y} are vector-valued input and output respectively, H is the channel matrix, and \mathbf{Z} is the additive i.i.d. Gaussian vector noise with an identity covariance 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:

$$\mathbf{Y}' = H^T\mathbf{X}' + \mathbf{Z}', \quad (2)$$

This relation has been observed in [1]. Reciprocity holds because the computation of the Gaussian vector channel capacity under a power constraint involves the water-filling of total power over the set

Manuscript submitted to *IEEE Transactions on Information Theory* on April 15, 2004. The material in this paper was presented in part at *Canadian Workshop on Information Theory 2003*, in Waterloo, Canada, in part at *DIMACS Workshop on Network Information Theory* in Rutgers University, New Jersey, U.S.A. in 2003, and in part at *IEEE International Symposium on Information Theory* in Chicago, U.S.A. in 2004. Wei Yu is with the Electrical and Computer Engineering Department, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada. email: weiyu@comm.utoronto.ca. This work was supported by the Natural Science and Engineering Research Council (NSERC) under the Canada Research Chair program and under a discovery grant, and in part by Bell Canada University Laboratory and Communication and Information Technology Ontario (CITO).

of inverted 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 \mathbf{X}' and \mathbf{X} (also \mathbf{Z}' and \mathbf{Z}) do not have to have the same dimension.

Interestingly, input-output reciprocity generalizes to multiuser channels. Let $\mathbf{Y}^T = [\mathbf{Y}_1^T \cdots \mathbf{Y}_K^T]$. Consider a Gaussian vector broadcast channel

$$\mathbf{Y}_i = H_i \mathbf{X} + \mathbf{Z}_i, \quad i = 1, \dots, K, \quad (3)$$

where \mathbf{Y}_i 's do not cooperate and $H^T = [H_1^T \cdots H_K^T]$. Although the capacity region of the broadcast channel is still not completely known, several recent papers ([2] [3] and [4]) have independently solved for the sum capacity of the Gaussian vector broadcast channel. In particular, [3] and [4] showed that the sum capacity of a Gaussian vector broadcast channel under a power constraint is equal to the sum capacity 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. Let $\mathbf{X}'^T = [\mathbf{X}'_1{}^T \cdots \mathbf{X}'_K{}^T]$. The reciprocal multiple access channel is of the form:

$$\mathbf{Y}' = H_i^T \mathbf{X}'_i + \mathbf{Z}'_i, \quad i = 1, \dots, K, \quad (4)$$

and it has a sum capacity:

$$C = \max_{S_{x'x'}} I(\mathbf{X}'_1 \cdots \mathbf{X}'_K; \mathbf{Y}'), \quad (5)$$

where $S_{x'x'}$ is a block diagonal matrix consisting of covariance matrices of $\mathbf{X}'_1 \cdots \mathbf{X}'_K$, and the maximization is over all positive semi-definite matrices $S_{x'x'}$ subject to a sum power constraint $\text{trace}(S_{x'x'}) \leq P$. 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.

The objective of this paper is to provide a deeper understanding of uplink-downlink duality. To motivate the main results of the paper, a brief overview of Gaussian vector broadcast channel sum capacity is necessary. The solution to broadcast channel sum capacity problem is based on two key ideas. First, an achievability result can be obtained using a precoding technique, known as ‘‘writing-on-dirty-paper’’ [5] that effectively pre-subtracts multi-user interference at the transmitter [6]. Second, a converse theorem can be obtained using the idea that the broadcast channel capacity is bounded by a cooperative bound with a ‘‘least-favorable’’ noise.

The least-favorable-noise converse is based on the fact that the sum capacity of the broadcast channel is bounded by the capacity of the vector channel with cooperative receivers: $\max I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K)$. Further, as observed in [7], the sum capacity of a broadcast channel depends only on the marginal distribution of the noises and not on their correlation. This is because the receivers of a broadcast channel cannot cooperate, so they are ignorant of the actual noise correlation. Therefore, the sum capacity of a broadcast channel must be bounded by the minimum mutual information

$$C = \min_{S_{zz}} \max_{S_{xx}} I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K), \quad (6)$$

minimized over all possible joint distributions of the noises. For a Gaussian vector broadcast channel with a convex input constraint, restricting the above minimax problem to joint Gaussian input and

noise distributions is without loss of generality. Further, it can be shown that this upper bound is achievable. One way to establish the achievability is to design a decision-feedback equalizer as a joint receiver in the broadcast channel [2]. As shown in [2], a decision-feedback equalizer designed for a least-favorable noise would have a feedforward matrix that 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 at the receiver is needed whatsoever and $\min I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K)$ is achievable. This minimum can be further maximized over all possible input distributions. Since min-max is equal to max-min 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 uplink-downlink duality. This duality approach is used in both [3] and [4]. The proof in [3] is based on the observation that the noise covariance matrix for the broadcast channel corresponds to an input cost constraint in the reciprocal channel. Further, the cost constraint corresponding to the least-favorable noise covariance is precisely the one that de-couples 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. In addition, [3] also showed that uplink-downlink duality can be interpreted via a convex duality.

The approach in [4] is based on another different idea. The authors of [4] 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, [4] 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 [2] [3] and [4] nevertheless contain several subtle differences. First of all, the duality approaches of [3] and [4] appears to apply only to Gaussian vector broadcast channels with a power constraint, while the decision-feedback equalization approach [2] applies to broadcast channels with arbitrary convex constraints. Therefore, the result in [2] appears to be more general. Second, the minimax result [2] applies to only the sum capacity point, while the duality result [4] applies to the entire achievable rate region. (Recent progress toward proving a converse has been reported in [8] and [9]. A full converse has been reported in [10].) Third, numerically, the duality expression (5) is much easier to compute 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. Finally, none of [2] [3] or [4] gives an explicit and complete solution to the minimax problem, especially when the least-favorable noise covariance is singular.

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 a complete solution to the Gaussian vector broadcast channel sum capacity problem. Toward this end, uplink-downlink duality is re-interpreted in the framework of a Lagrangian duality in minimax Gaussian mutual information optimization. This general minimax duality theory not only unifies the two previous results in uplink-downlink duality but also allows an explicit solution to the minimax problem based

on the dual problem. In addition, this solution reveals the structure of the least-favorable noise. It is shown that the least-favorable noise is often not unique, and different least-favorable noises are related to each other via a linear estimation relation. Further, the minimax framework allows uplink-downlink duality to be generalized to broadcast channels under arbitrary linear covariance constraints. In particular, when multiple linear constraints are present, the dual of a broadcast channel becomes a multiple-access channel with a uncertain noise. Finally, it is illustrated 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 duality expression.

The organization for the rest of the paper is as follows. In Section II, a general theory of minimax duality for the Gaussian mutual information is established. This minimax duality result is based on the KKT condition for the minimax optimization problem, and it is intimately connected to the Lagrangian theory in convex optimization. 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 illustrates the connection between uplink-downlink duality and Lagrangian duality. In Section IV, the duality result is used to characterize the saddle-point of the Gaussian minimax mutual information game. The expressions for the optimal transmit covariance matrix and the least-favorable noise are found in terms of the dual variables. Finally in Section V, 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 constraint is found. Section VI contains concluding remarks.

II. DUALITY THEORY FOR GAUSSIAN CHANNELS

In this section, we present a general theory of duality for the minimax Gaussian mutual information expression. 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:

$$\text{tr}(S_{xx}Q_x) \leq 1 \quad (8)$$

$$\text{tr}(S_{zz}Q_z) \leq 1, \quad (9)$$

where Q_x and Q_z are parameters of the linear constraints, and they are assumed to be symmetric positive semi-definite matrices. Implicitly, S_{xx} and S_{zz} are constrained to be positive semi-definite 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 it 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 the dual of the above minimax problem. The relation is most transparent when H is invertible and the saddle point solution S_{xx} and S_{zz}

are strictly positive definite. For the moment, we explicitly assume the invertibility and positive semi-definiteness conditions. These are technical conditions that will be removed later.

The first step in developing the 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 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:

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

Thus, 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, substitute (13) into (10) and solve for S_{xx} :

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

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

$$\Sigma_{xx} = \lambda_z Q_z \quad , \quad \Sigma_{zz} = \lambda_x Q_x \quad (15)$$

$$S_{xx} = \nu_z \Psi_z \quad , \quad S_{zz} = \nu_x \Psi_x \quad (16)$$

Equations (13) (14) can be re-written 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:

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

then,

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

Therefore, associated with the original minimax problem (7), there is a ‘‘dual’’ minimax problem:

$$C(H^T, \Psi_x, \Psi_z) = \max_{\Sigma_{xx}} \min_{\Sigma_{zz}} \frac{1}{2} \log \frac{|H^T \Sigma_{xx} H + \Sigma_{zz}|}{|\Sigma_{zz}|}, \quad (21)$$

with linear covariance constraints:

$$\text{tr}(\Sigma_{xx}\Psi_x) \leq 1, \quad (22)$$

$$\text{tr}(\Sigma_{zz}\Psi_z) \leq 1. \quad (23)$$

Further, it can be verified using (10) (11) and (15) that at the saddle-point:

$$\begin{aligned} \log \frac{|H^T \Sigma_{xx} H + \Sigma_{zz}|}{|\Sigma_{zz}|} &= \log \frac{|H^T \lambda_z Q_z H + \lambda_x Q_x|}{|\lambda_x Q_x|} \\ &= \log \frac{|H^T (S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1})H + H^T (HS_{xx}H^T + S_{zz})^{-1}H|}{|H^T (HS_{xx}H^T + S_{zz})^{-1}H|} \\ &= \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|}. \end{aligned} \quad (24)$$

Thus, the minimax problems (7) and (21) 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 (21).
- The optimal dual variable λ_z in the minimization part of (7) is the optimal dual variable ν_x in the maximization part of (21).
- The optimizing variables of (7) are related to the constraints of (21) by $(S_{xx}, S_{zz}) = (\nu_z \Psi_z, \nu_x \Psi_x)$.
- The optimizing variables of (21) 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-dual solution 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-dual solution of the dual minimax problem $C(H^T, \Psi_x, \Psi_z)$ is precisely $\Sigma_{xx} = \lambda_z Q_z$, $\Sigma_{zz} = \lambda_x Q_x$, $\nu_x = \lambda_z$ and $\nu_z = \lambda_x$. Further, $C(H^T, \Psi_x, \Psi_z) = C(H, Q_x, Q_z)$.

Proof: The derivation leading to the theorem provides a proof 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 I. \square

To summarize, there is a 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. This minimax duality is established based on the KKT conditions of the optimization problem, and it is a particular feature of the Gaussian mutual information optimization problem.

Minimax duality is related to Lagrangian duality in optimization. This relation is not yet apparent from the derivation above. In fact, the Lagrangian optimization dual of (7), although also a minimax problem, involves only a scalar maximizing variable and a scalar minimizing variable. However, it turns out that minimax duality, when applied to the Gaussian vector broadcast channel, is equivalent to Lagrangian duality. This point is elaborated upon in the next section.

Minimax Problem:	$\max_{S_{xx}} \min_{S_{zz}} \frac{1}{2} \log \frac{ HS_{xx}H^T + S_{zz} }{ S_{zz} }$	$\max_{\Sigma_{xx}} \min_{\Sigma_{zz}} \frac{1}{2} \log \frac{ H^T\Sigma_{xx}H + \Sigma_{zz} }{ \Sigma_{zz} }$
Constraints:	$\text{tr}(S_{xx}Q_x) \leq 1$ $\text{tr}(S_{zz}Q_z) \leq 1$	$\text{tr}(\Sigma_{xx}\Psi_x) \leq 1$ $\text{tr}(\Sigma_{zz}\Psi_z) \leq 1$
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:	$\lambda_x = \nu_z$ $\lambda_z = \nu_x$	$\nu_x = \lambda_z$ $\nu_z = \lambda_x$

TABLE I
MINIMAX DUALITY

III. UPLINK-DOWNLINK DUALITY

The main motivation for studying minimax duality 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 [2], the broadcast channel sum capacity is shown to be the solution of a minimax mutual information problem, while in [11] [4] and [3], 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.

A. Gaussian Vector Broadcast Channel Sum Capacity

Consider a Gaussian vector broadcast channel

$$\mathbf{Y}_i = H_i\mathbf{X} + \mathbf{Z}_i, \quad i = 1, \dots, K, \quad (25)$$

where \mathbf{X} is the transmit signal, \mathbf{Y}_i 's are non-coordinated receivers, $\mathbf{Z}^T = [\mathbf{Z}_1^T \mathbf{Z}_2^T]$ is a Gaussian noise vector with an identity covariance matrix, and $H^T = [H_1^T \dots H_K^T]$. A sum power constraint $\mathbf{E}[\mathbf{X}^T\mathbf{X}] \leq P$ 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”, Costa [5] showed that if a Gaussian channel is corrupted by an interference signal S that is known *non-causally* to the transmitter but not to the receiver, i.e.

$$\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}, \quad (26)$$

the capacity of the channel is the same as if \mathbf{S} does not exist. Thus, in a broadcast channel, if $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ where \mathbf{X}_1 and \mathbf{X}_2 are Gaussian vectors, \mathbf{X}_1 can transmit information to \mathbf{Y}_1 as if \mathbf{X}_2 does not exist, and \mathbf{X}_2 can still transmit to \mathbf{Y}_2 with \mathbf{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

2-user 2-antenna case by Caire and Shamai [6], and has since been generalized by several authors [2] [4] [3].

The approach in [2] is based on the observation that interference pre-subtraction at the transmitter is identical to a decision-feedback equalizer 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 $\mathbf{Y}_1 \cdots \mathbf{Y}_K$. Clearly, such coordination is not possible in a broadcast channel. But, precisely because $\mathbf{Y}_1 \cdots \mathbf{Y}_K$ cannot coordinate, they are also ignorant of the noise correlation between $\mathbf{Z}_1 \cdots \mathbf{Z}_K$. Thus, the sum capacity of the broadcast channel must be bounded by the cooperative capacity with the *least favorable* noise correlation:

$$C \leq \min_{S_{zz}} I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K), \quad (27)$$

where S_{zz} is the covariance matrix for $\mathbf{Z}^T = [\mathbf{Z}_1^T \cdots \mathbf{Z}_K^T]$, and the minimization is over all S_{zz} whose block diagonal terms are the covariance matrices of $\mathbf{Z}_1, \cdots, \mathbf{Z}_K$. This outer bound is due to Sato [7].

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

$$S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1} = \begin{bmatrix} \Phi_1 & & 0 \\ & \ddots & \\ 0 & & \Phi_K \end{bmatrix} = \Phi, \quad (28)$$

where Φ_i is the dual variable corresponding to the i th 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 de-couples 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 S_{xx} , the capacity of the broadcast channel is precisely $\min_{S_{zz}} I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K)$.

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:

$$\text{tr}(S_{xx}) \leq P. \quad (29)$$

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

$$C \leq \min_{S_{zz}} \max_{S_{xx}} I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K), \quad (30)$$

(as $\max_{S_{xx}} I(\mathbf{X}; \mathbf{Y})$ is the capacity of the cooperative channel, and Sato’s bound minimizes over all cooperative capacity.) Further, $\min_{S_{zz}} I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K)$ is achievable with any Gaussian inputs, thus

it can be maximized over all S_{xx} :

$$C \geq \max_{S_{xx}} \min_{S_{zz}} I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K). \quad (31)$$

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:

$$\begin{aligned} \max_{S_{xx}} \min_{S_{zz}} \quad & \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|} \\ \text{s.t.} \quad & S_{zz}(i, i) = I, \quad i = 1, \dots, K. \\ & \text{tr}(S_{xx}) \leq P, \\ & S_{xx}, S_{zz} \geq 0, \end{aligned} \quad (32)$$

where $S_{zz}(i, i)$ refers to the i th block diagonal of S_{zz} . Interestingly, because of the minimax duality, the above minimax optimization problem is considerably easier to solve than the minimization problem (27).

B. Uplink-Downlink Duality via Minimax Duality

Uplink-downlink duality refers to the duality between a multiple access channel and a broadcast channel [11] [3] [4]. In [11], it was observed that under an input power constraint, the achievable rate region of a broadcast channel using the precoding technique is identical to the capacity region of a dual multiple access channel with the channel matrix transposed and a sum power constraint applied to all inputs. In [3], it was observed that the uplink-downlink duality is also closely related to convex Lagrangian duality. Based on convex duality, [4] and [3] showed that the sum capacity of the broadcast channel is the same as the sum capacity of the dual multiple access channel under a sum power constraint. The main objective of this section is to give a new derivation of uplink-downlink duality via minimax duality. The new derivation provides new insights into the structure of the optimization problem.

To simplify matters, let's again assume that H is square and invertible, and also that the maximizing S_{xx} and the least-favorable S_{zz} are both full rank. Further, assume that each receiver in the broadcast channel is equipped with a single-antenna. The starting point of the new derivation is the KKT condition for the minimax problem (32). Because the objective function in (32) is concave in S_{xx} and convex in S_{zz} , the KKT condition completely characterizes the saddle point. The KKT condition is:

$$H^T(HS_{xx}H^T + S_{zz})^{-1}H = \lambda I \quad (33)$$

$$S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1} = \Phi \quad (34)$$

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 (34) by H^T on the left and H on the right and substituting in (33):

$$H^T S_{zz}^{-1} H = H^T \Phi H + \lambda I. \quad (35)$$

The above is equivalent to

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

Observe that since Ψ is a diagonal matrix, (36) is precisely the KKT condition for a multiple access channel with a transmit covariance matrix Φ and an input constraint of the form $\text{tr}(\Phi S_{zz}) \leq 1$. The above equation also gives an explicit solution for S_{zz} . Further, substituting (36) into (33), it is not difficult to see that

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

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

Equations (36) and (37) define a set of KKT conditions for a dual minimax problem. From (36) and (37), it is clear that the minimax dual of (32) is the following problem:

$$\begin{aligned} \max_{\Phi} \min_{\lambda} \quad & \frac{1}{2} \log \frac{|H^T\Phi H + \lambda I|}{|\lambda I|} \\ \text{s.t.} \quad & \text{tr}(\Phi S_{zz}) \leq 1 \\ & \text{tr}(\lambda S_{xx}) \leq 1 \\ & \Phi, \lambda \geq 0, \end{aligned} \quad (38)$$

As the original minimax problem (32) and the dual problem (38) have equivalent KKT conditions, (38) is in fact the Lagrangian optimization dual of (32), where Φ is the Lagrangian dual variable associated with the diagonals of the noise covariance matrix and λ is the Lagrangian dual variable associated with the transmit power constraint. 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 Lagrangian duality between (32) and (38) is equivalent to the uplink-downlink duality as developed in [11] [3] and [4]. This is so 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 (36) and to recognize following two features. First, the noise of the dual problem is an identity matrix scaled by λ . Second, the constraint of the dual problem, S_{zz} , has identity matrices on the diagonal. Thus, the input constraint on $\text{tr}(\Phi S_{xx})$ is equivalent to a sum power constraint on $\text{tr}(\Phi)$:

$$\text{tr}(\Phi S_{zz}) = \sum_{i=1}^K \text{tr}(\Phi_i). \quad (39)$$

As shown in the next theorem, after a proper scaling of the power constraint

$$\Sigma = \frac{\Phi}{\lambda}, \quad (40)$$

it is possible to prove that (36) is precisely the KKT condition for a single maximization problem

$$\begin{aligned} \max_{\Sigma} \quad & \frac{1}{2} \log |H^T \Sigma H + I| \\ \text{s.t.} \quad & \Sigma \text{ is block diagonal} \\ & \text{tr}(\Sigma) \leq P, \end{aligned} \quad (41)$$

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:

Theorem 2: Consider the Gaussian mutual information minimax problem for a broadcast channel (32) and the maximization problem for a multiple access channel (41). Assume that H is square and invertible, the solution to the minimax problem (32), (S_{xx}, S_{zz}) , is non-singular, and the solution to the maximization problem (41), Σ is non-singular. Then, the primal-dual solution $(S_{xx}, S_{zz}, \lambda, \Phi)$ of (32) may be obtained from the primal-dual solution (Σ, ν) of the multiple-access channel (41) as follows:

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

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

$$\Phi = \nu\Sigma, \quad (44)$$

$$\lambda = \nu. \quad (45)$$

Conversely, the primal-dual 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 capacity.

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

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

Now, re-write the multiple access channel as follows:

$$\begin{aligned} \max_{\Sigma} \quad & \frac{1}{2} \log \left| \sum_{i=1}^K H_i^T \Sigma_i H_i + I \right| \\ \text{s.t.} \quad & \sum_{i=1}^K \text{tr}(\Sigma_i) \leq P, \end{aligned} \quad (47)$$

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

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

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

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

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

To establish the duality relation between the multiple access channel and the broadcast channel, it remains to 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 $\text{tr}(\Sigma) \leq P$, and the constraint associated with (49) is of the form $\text{tr}(\Phi\Gamma) = \text{tr}(\Phi) \leq P'$. 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 sensitivity 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 matrices S_{zz} . Then,

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

and

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

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

$$\text{tr}(S_{xx}) \leq P(1 + \delta) \quad (52)$$

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

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

where δ is a small positive real number. Observed that since the proportional increases in signal and noise powers are the same, from the structure of the expression (32), the saddle-point (S_{xx}, S_{zz}) would be scaled by exactly the same proportion also. 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

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

Therefore,

$$\text{tr}(\Sigma) = \sum_{i=1}^K \text{tr}(\Phi_i)/\lambda = P. \quad (55)$$

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 capacity. Using the dual variable relations (44) – (45), it can be seen that:

$$\begin{aligned} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|} &= \log \frac{|H((\lambda I)^{-1} - (H^T(\lambda \Sigma)H + \lambda I)^{-1})H^T + H(H^T(\lambda \Sigma)H + \lambda I)^{-1}H^T|}{|H(H^T\Phi H + \lambda I)^{-1}H^T|} \\ &= \log \frac{|H^T\Phi H + \lambda I|}{|\lambda I|} \\ &= \log |H\Sigma H^T + I| \end{aligned} \quad (56)$$

Optimization Problem:	$\max_{S_{xx}} \min_{S_{zz}} \frac{1}{2} \log \frac{ HS_{xx}H^T + S_{zz} }{ S_{zz} }$	$\max_{\Sigma} \frac{1}{2} \log H^T \Sigma H + I $
Constraints:	$\text{tr}(S_{xx}) \leq P$ $S_{zz}(i, i) = I$	$\text{tr}(\Sigma) \leq P$
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:	$\Phi = \nu \Sigma$ $\lambda = \nu$	$\nu = \lambda$

TABLE II
UPLINK-DOWNLINK DUALITY

This proves the theorem. \square

The duality relation between the multiple access channel and the broadcast channel is summarized in Table II. Duality is a statement of the fact that the solution to the minimax problem (32) can be obtained from the solution to a maximization problem (41) and vice versa. Since (41) is much easier to solve numerically, uplink-downlink duality gives an efficient way to solve the minimax problem (32).

C. Singular Least-Favorable Noise

The duality result 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 the channel matrix H is square and invertible. Also, the least-favorable noise, the maximizing input covariance matrix for the broadcast channel and the maximizing input covariance matrices 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 diagonal blocks of maximizing 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:

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

where Υ is a positive semi-definite matrix satisfying the complementary slackness condition

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

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 identities. However, if Σ is low rank, then $H(H^T(\nu \Sigma)H + \nu I)^{-1}H^T$ do 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 minimax duality is established with inequality constraints of the type $\text{tr}(S_{zz}Q_z) \leq 1$. However, the constraints associated with the broadcast channel is of the form:

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

rather than

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

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

The purpose of this section is to show that a least-favorable noise covariance can be obtained by adding a positive semi-definite 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. Such a positive semi-definite matrix is not unique, so the least-favorable noise covariance matrix is not unique. In fact, in some cases, the class of least favorable noises can be further enlarged by taking an additional step as will be shown later.

Consider the candidate least-favorable noise, re-labeled as $S_{zz}^{(0)}$:

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

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

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

Now, S_{xx} may be rank deficient. In this case, the dimension of H must be reduced, and the least-favorable noise must be re-computed using the reduced channel. The re-computed noise covariance could be of lower rank, thus enlarging the class of positive semi-definite matrices that can be added to its diagonal terms. For the rest of the section, it is assumed that this channel reduction step has already taken place.

The expression for the least favorable noise (61) has appeared in [3] as the cost constraint of a dual multiple access channel, although the singular noise issue was not specifically dealt with in [3]. The idea of forcing the diagonal terms to be identity matrices has appeared in [4]. To show that the singular noise is a solution to the minimax problem, the proof in [4] constructed a sequence of non-singular noises, and show that the result is obtained in the limit. Further, the result in [4] does not characterize the entire set of least favorable noises. In the following, the class of least favorable noises for the minimax problem is characterized explicitly.

Theorem 3: Consider the Gaussian mutual information minimax problem for a broadcast channel (32) and the maximization problem for a multiple access channel (41). The solution (S_{xx}, S_{zz}) of the broadcast channel problem (32) may be obtained from the primal-dual solution (Σ, ν) of the multiple-access channel (41) as follows:

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

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

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

Proof: The proof is fairly lengthy. As a first step, let's assume that the broadcast channel has only a single antenna at each receiver. This simplifies the algebra considerably and it shows all the essential structure of the least favorable noise. The result can be easily generalized to the multi-receive-antenna case. The generalization is presented at the end of the proof. Also, in the rest of the proof, $\nu\Sigma$ is denoted as Φ for notational convenience.

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

$$S_{zz}^{(0)} = U_1 S_{\tilde{z}\tilde{z}} U_1^T \quad (65)$$

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

$$H = U_1 \tilde{H}, \quad (66)$$

where \tilde{H} is invertible. The strategy is to solve the minimax problem over $S_{\tilde{z}\tilde{z}}$ and to show that the solution is $S_{zz}^{(0)} = H(H^T \Phi H + \lambda I)^{-1} H^T$. The reduced minimax problem is now:

$$\begin{aligned} \max_{S_{xx}} \min_{S_{\tilde{z}\tilde{z}}} & \frac{1}{2} \log \frac{|\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|} \\ \text{s.t.} & \quad \text{tr}(S_{xx}) \leq P \\ & \quad U_1 S_{\tilde{z}\tilde{z}} U_1^T \text{ has 1's on the diagonal} \\ & \quad S_{xx} \geq 0, \end{aligned} \quad (67)$$

The KKT condition of the minimax problem is:

$$\tilde{H}^T (\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} \tilde{H} = \lambda I \quad (68)$$

$$S_{\tilde{z}\tilde{z}}^{-1} - (\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} = U_1^T \Phi U_1 \quad (69)$$

Pre-multiplying the second equation above by \tilde{H}^T and post-multiplying by \tilde{H} , it is not difficult to verify that

$$H(H^T \Phi H + \lambda I)^{-1} H^T = U_1 S_{\tilde{z}\tilde{z}} U_1^T = S_{zz}^{(0)}. \quad (70)$$

Thus, the candidate $S_{zz}^{(0)}$ satisfies the least-favorable noise condition. Further, with $S_{zz}^{(0)}$ as the noise covariance, the optimal water-filling covariance S_{xx} is precisely

$$S_{xx} = (\nu I)^{-1} - (H^T (\nu \Sigma) H + \nu I)^{-1}. \quad (71)$$

The rest of the proof is devoted to showing that when a positive semi-definite 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 re-arranged so that the upper-left sub-matrix 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 non-zero entries only in the lower-right corner:

$$\Phi = \begin{bmatrix} \Phi' & 0 \\ 0 & 0 \end{bmatrix}, \quad S'_{zz} = \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix}. \quad (72)$$

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 $k \geq r$. The reason for this has to do with the channel reduction step mentioned before. Note that from (62) the optimal S_{xx} can be expressed as:

$$\begin{aligned} S_{xx} &= (\lambda I)^{-1} - (H^T \Phi H + \lambda I)^{-1} \\ &= (\lambda I)^{-1} H^T \sqrt{\Phi} (I + \sqrt{\Phi} H H^T \sqrt{\Phi})^{-1} \sqrt{\Phi} H (\lambda I)^{-1}. \end{aligned} \quad (73)$$

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 non-zero diagonal entries in at least r' positions, where r' is the rank of H . So, $k \geq r'$. But, after the channel reduction, the rank of H is the same as the rank of $S_{zz}^{(0)}$. This proves that $k \geq r$.

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:

$$S'_{zz} = \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} = [U_1 \quad U_2] \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}. \quad (74)$$

Recall that U_1 contains r column vectors and U_2 contains $n-k$ column vectors. So, $[U_1 \quad U_2]$ 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 sub-matrices:

$$U_1 = \begin{bmatrix} u_{11}^T \\ u_{12}^T \end{bmatrix} \quad U_2 = \begin{bmatrix} u_{21}^T \\ u_{22}^T \end{bmatrix}. \quad (75)$$

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

$$S_{11} - S_{12} S_{22}^{-1} S_{21} = 0. \quad (76)$$

This is because

$$\begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} [U_1 \quad U_2] = I, \quad (77)$$

so, from (74),

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} u_{11}^T & u_{12}^T \\ u_{21}^T & u_{22}^T \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix}. \quad (78)$$

This allows S_{ij} to be solved explicitly:

$$S_{11} = u_{12}^T S u_{12} \quad S_{12} = u_{12}^T S u_{22} \quad (79)$$

$$S_{21} = u_{22}^T S u_{12} \quad S_{22} = u_{22}^T S u_{22}. \quad (80)$$

Therefore,

$$S_{12} S_{22}^{-1} S_{21} = u_{12}^T S u_{22} (u_{22}^T S u_{22})^{-1} u_{22}^T S u_{12} = S_{11}, \quad (81)$$

thus proving (76).

The second useful fact is the following:

$$u_{21} = u_{11} u_{12}^T u_{22}^{-T}. \quad (82)$$

The proof is based on (74)

$$\begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} u_{11}^T & u_{12}^T \\ u_{21}^T & u_{22}^T \end{bmatrix}. \quad (83)$$

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

$$0 = u_{11} S_{11} u_{12}^T + u_{21} S_{21} u_{12}^T + u_{11} S_{12} u_{22}^T + u_{21} S_{22} u_{22}^T. \quad (84)$$

Substituting in (79)–(80), the above is equal to

$$0 = (u_{11} u_{12}^T + u_{21} u_{22}^T) S (u_{11} u_{11}^T + u_{22} u_{22}^T). \quad (85)$$

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

$$u_{11} u_{12}^T + u_{21} u_{22}^T = 0, \quad (86)$$

thus proving (82).

Finally, we are ready to show that $S_{zz}^{(0)} + S'_{zz}$ satisfies the least-favorable noise condition. Starting from (69)

$$S_{zz}^{-1} - (\tilde{H} S_{xx} \tilde{H}^T + S_{zz})^{-1} = U_1^T \Phi U_1, \quad (87)$$

the objective is to prove that

$$\begin{bmatrix} S_{zz} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1} - \begin{bmatrix} \tilde{H} S_{xx} \tilde{H}^T + S_{zz} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1} = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \Phi [U_1 \ U_2]. \quad (88)$$

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

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \Delta_D^{-1} & -\Delta_D^{-1} B D^{-1} \\ -D^{-1} C \Delta_D^{-1} & D^{-1} + D^{-1} C \Delta_D^{-1} B D^{-1} \end{bmatrix}, \quad (89)$$

where $\Delta_D = A - B D^{-1} C$. To evaluate the matrix inversions in (88), Δ_D^{-1} can be simplified using (76):

$$\Delta_D = S_{zz} + S_{11} - S_{12} S_{22}^{-1} S_{21} = S_{zz}. \quad (90)$$

Thus, the first matrix inversion in (88) is equal to

$$\begin{bmatrix} S_{\bar{z}\bar{z}}^{-1} & -S_{\bar{z}\bar{z}}^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}S_{\bar{z}\bar{z}}^{-1} & S_{22}^{-1} + S_{22}^{-1}S_{21}S_{\bar{z}\bar{z}}^{-1}S_{12}S_{22}^{-1} \end{bmatrix}, \quad (91)$$

and the second matrix inversion in (88) can be expanded similarly:

$$\begin{bmatrix} (\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1} & -(\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}(\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1} & S_{22}^{-1} + S_{22}^{-1}S_{21}(\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1}S_{12}S_{22}^{-1} \end{bmatrix}. \quad (92)$$

Now, using (87), the difference between the two can now be simplified:

$$\begin{bmatrix} U_1^T\Phi U_1 & -U_1^T\Phi U_1 S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}U_1^T\Phi U_1 & S_{22}^{-1}S_{21}U_1^T\Phi U_1 S_{12}S_{22}^{-1} \end{bmatrix}. \quad (93)$$

To prove (88), it remains to show that the above is equal to

$$\begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \Phi \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} U_1^T\Phi U_1 & U_1^T\Phi U_2 \\ U_2^T\Phi U_1 & U_2^T\Phi U_2 \end{bmatrix}. \quad (94)$$

Comparing (93) with (94), it is clear that the two are equal if the following holds:

$$\Phi U_2 = -\Phi U_1 S_{12}S_{22}^{-1}. \quad (95)$$

Recall that Φ has non-zero entries only in its upper-left diagonal. So, the left-hand side of (95) is

$$\Phi U_2 = \begin{bmatrix} \Phi' & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = \Phi' u_{21}. \quad (96)$$

The right-hand side of (95) is

$$-\Phi U_1 S_{12}S_{22}^{-1} = -\Phi' u_{11} u_{12}^T u_{22}^{-T}. \quad (97)$$

By (82), the left-hand side is equal to the right-hand side. This establishes the least-favorable noise condition (88).

Now, we verify that the broadcast channel and the multiple access channel have the same sum capacity. First, it is easy to see that:

$$\begin{aligned} \log \frac{|HS_{xx}H^T + S_{zz}^{(0)}|}{|S_{zz}^{(0)}|} &= \log \frac{|\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}}|}{|S_{\bar{z}\bar{z}}|} \\ &= \log \frac{|\tilde{H}((\lambda I)^{-1} - (H^T(\lambda\Sigma)H + \lambda I)^{-1})\tilde{H}^T + \tilde{H}(H^T(\lambda\Sigma)H + \lambda I)^{-1}\tilde{H}^T|}{|\tilde{H}(H^T\Phi H + \lambda I)^{-1}\tilde{H}^T|} \\ &= \log \frac{|H^T\Phi H + \lambda I|}{|\lambda I|} \\ &= \log |H\Sigma H^T + I|. \end{aligned} \quad (98)$$

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

$$C = \frac{1}{2} \log \frac{\begin{vmatrix} \tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}}{\begin{vmatrix} S_{\bar{z}\bar{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}}. \quad (99)$$

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

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D| \cdot |A - BD^{-1}C|, \quad (100)$$

it is not difficult to see that

$$\frac{|S_{22}| \cdot |\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}} + S_{11} - S_{12}S_{22}^{-1}S_{21}|}{|S_{22}| \cdot |S_{\bar{z}\bar{z}} + S_{11} - S_{12}S_{22}^{-1}S_{21}|} = \frac{|\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}}|}{|S_{\bar{z}\bar{z}}|}. \quad (101)$$

Therefore the capacity expressions (98) and (99) 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 $S_{zz}^{(0)} + S'_{zz}$. 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 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.

To generalize the result to the multi-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.. $(S_{xx}, S_{zz}^{(0)})$ satisfies the saddle-point condition. We need to show that the saddle-point KKT conditions remain to be satisfied if a positive semi-definite 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

$$S_{zz}^{(0)}(i, i) = V^T S_{\bar{z}\bar{z}}^{(0)}(i, i) V \quad (102)$$

be the eigenvalue decomposition of $S_{zz}^{(0)}(i, i)$, where $S_{\bar{z}\bar{z}}^{(0)}(i, i)$ 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:

$$\mathbf{Y}'_i = V H_i \mathbf{X} + V \mathbf{Z}_i. \quad (103)$$

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's and others being less than 1's. In such a new broadcast channel, receiver coordination is not needed within each \mathbf{Y}'_i . Thus, \mathbf{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 multi-receive antenna case. \square

D. Linear Estimation Interpretation

The least-favorable noise in a Gaussian minimax mutual information game is of the form $S_{zz}^{(0)} + S'_{zz}$. The addition of the term S'_{zz} does not change the mutual information $I(\mathbf{X}; \mathbf{Y})$. In some sense, $S_{zz}^{(0)}$ is the “smallest” possible noise within the class of least-favorable noises. This is because after the channel reduction step, $HS_{xx}H^T$ is strictly positive everywhere in the span of the U_1 space. Thus, for capacity not to become infinity, the noise covariance must also be strictly positive in the space spanned by U_1 . Now, $S_{zz}^{(0)}$ is the only noise covariance entirely in the U_1 space that satisfies the KKT condition. Thus, $S_{zz}^{(0)}$ must be the “smallest” noise covariance that satisfies the KKT condition.

The class of least-favorable noises $S_{zz}^{(0)} + S'_{zz}$ has a linear estimation interpretation. Let $\mathbf{Z}^{(0)}$ and \mathbf{Z}' be independent Gaussian noise vectors with covariance matrices $S_{zz}^{(0)}$ and S'_{zz} respectively. Consider an orthogonal transformation $\mathbf{Z}_u^{(0)} = [U_1 U_2] \mathbf{Z}^{(0)}$ and $\mathbf{Z}'_u = [U_1 U_2] \mathbf{Z}'$. These two noise vectors take the form

$$\mathbf{Z}_u^{(0)} = \begin{bmatrix} \mathbf{z}_1^0 \\ 0 \end{bmatrix}, \quad \mathbf{Z}'_u = \begin{bmatrix} \mathbf{z}'_1 \\ \mathbf{z}'_2 \end{bmatrix}. \quad (104)$$

Consider two receivers, one has $\mathbf{Z}_u^{(0)}$ as noise and the other has $\mathbf{Z}_u^{(0)} + \mathbf{Z}'_u$ 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 linear estimation of $\mathbf{z}_1^0 + \mathbf{z}'_1$ given \mathbf{z}'_2 is exactly \mathbf{z}'_1 . Since \mathbf{z}_1^0 and \mathbf{z}'_1 are independent, this condition is equivalent to $\mathbf{E}[\mathbf{z}'_1 | \mathbf{z}'_2] = 0$. But, the covariance matrix of $\mathbf{E}[\mathbf{z}'_1 | \mathbf{z}'_2]$ is $S_{11} - S_{12} S_{22}^{-1} S_{21}$. Thus, the covariance matrix of the additional noise S'_{zz} must satisfy the condition

$$S_{11} - S_{12} S_{22}^{-1} S_{21} = 0. \quad (105)$$

The above argument shows that the entire class of least-favorable noises are related to each other via a linear estimation relation¹. As the proof of Theorem 3 shows, this relation is crucial in the derivation of the least-favorable condition.

IV. GENERALIZED DUALITY

A. Duality with an Arbitrary Linear Constraint

The duality between the Gaussian broadcast channel and the Gaussian multiple access channel is important from a computational perspective. The multiple access channel capacity (41) is considerably easier to compute than the minimax problem (32). Although the duality result as established in [11] [4] and [3] applies only to a broadcast channel with a sum power constraint, it is clear from minimax duality that it may be generalized to broadcast channels with linear covariance input constraints of the form:

$$\text{tr}(S_x Q) \leq P. \quad (106)$$

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

$$C = \max_{\Sigma} \frac{1}{2} \log |H^T \Sigma H + Q|. \quad (107)$$

The same power constraint applies as before: $\text{tr}(\Sigma) \leq P$.

¹There is a subtlety in that, strictly speaking, \mathbf{Z}' needs not be Gaussian. However, this does not affect the sum capacity. See [12] for a discussion.

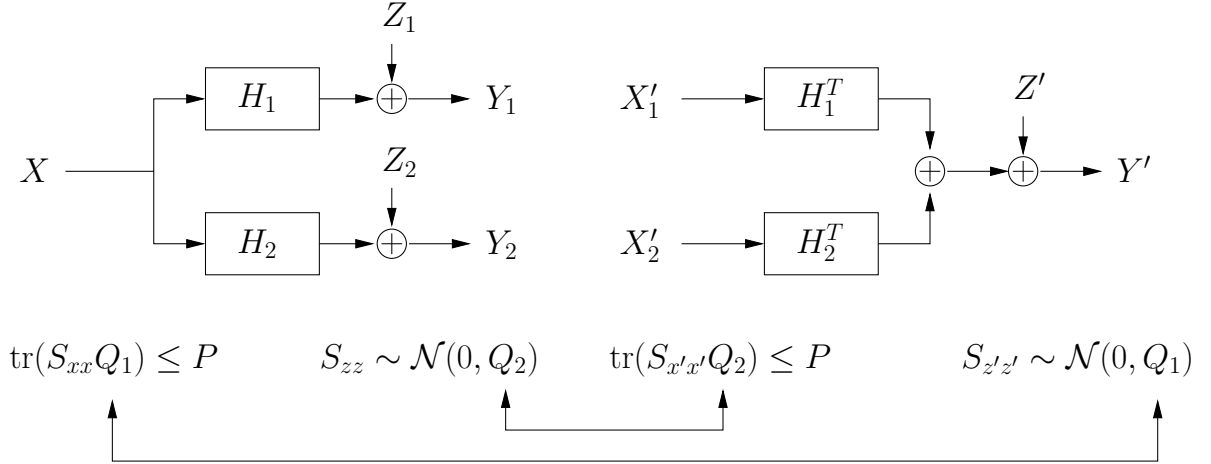


Fig. 1. Generalized Uplink-Downlink Duality

There is an interesting interplay between the cost constraint in the original channel and the noise covariance in the dual channel. In [3], the least-favorable noise covariance in the Gaussian broadcast channel may be transformed into a cost constraint for the dual multiple access channel. The above derivation shows that the same thing happens the other way around. The input cost constraint of a broadcast channel may be transformed into the noise covariance in the dual multiple access channel. Fig. 1 illustrates this transformation.

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

$$f(S_{xx}) \leq 0. \quad (108)$$

Its sum capacity can still be shown to be

$$\max_{S_{xx}} \min_{S_{zz}} \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|}. \quad (109)$$

(Recall that min-max equals max-min and saddle-point is Gaussian for any convex-concave function under arbitrary convex input constraints.) The KKT condition of the minimax problem implies that

$$H(H^T\Phi H + \lambda Q)^{-1}H^T = S_{zz}, \quad (110)$$

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

$$C = \max_{\Sigma} \frac{1}{2} \log |H^T\Sigma H + Q|. \quad (111)$$

However, without 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 (32) is explicitly solved. Therefore, although

minimax duality still exists, it is not useful computationally. In this sense, the minimax expression (32) is a more fundamental characterization of the Gaussian vector broadcast channel sum capacity. It applies to broadcast channels with any convex input covariance constraints. Duality holds only when the input constraint is linear.

B. Duality for Broadcast Channels with Per-Antenna Power Constraint

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:

$$\begin{aligned} \text{tr}(S_{xx}Q_1) &\leq 1 \\ &\vdots \\ \text{tr}(S_{xx}Q_L) &\leq 1. \end{aligned} \quad (112)$$

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:

$$H(H^T\Phi H + \lambda_1 Q_1 + \cdots + \lambda_L Q_L)^{-1}H^T = S_{zz}. \quad (113)$$

and

$$(\lambda_1 Q_1 + \cdots + \lambda_L Q_L)^{-1} - (H^T\Phi H + \lambda_1 Q_1 + \cdots + \lambda_L Q_L)^{-1} = S_{xx}. \quad (114)$$

The dual minimax problem is therefore:

$$\begin{aligned} \max_{\Phi} \min_{\lambda_1 \cdots \lambda_L} & \frac{1}{2} \log \frac{|H^T\Phi H + \sum_{l=1}^L \lambda_l Q_l|}{|\sum_{l=1}^L \lambda_l Q_l|} \\ \text{s.t.} & \text{tr}(S_{zz}\Phi) \leq 1 \\ & \text{tr}\left(S_{xx} \sum_{l=1}^L \lambda_l Q_l\right) \leq 1 \\ & \Phi, \lambda \geq 0, \end{aligned} \quad (115)$$

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, an individual per-antenna power constraint, rather than a sum power constraint, is imposed on transmit antenna of the broadcast channel. The computation of capacity for such a channel has been considered in the past [13], but only sub-optimal 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} :

$$S_{xx}(i, i) \leq P_i, \quad i = 1 \cdots n. \quad (116)$$

Thus, Q_i in (112) is just an all-zero matrix except in the i th diagonal term, which is $1/P_i$. Define

$$Q = \lambda_1 Q_1 + \cdots + \lambda_n Q_n = \text{diag}(\lambda_1/P_1, \cdots, \lambda_n/P_n). \quad (117)$$

The KKT conditions (113) - (114) imply that the dual minimax problem is:

$$\begin{aligned} \min_Q \max_{\Phi} \quad & \frac{1}{2} \log \frac{|H^T \Phi H + Q|}{|Q|} \\ \text{s.t.} \quad & \Phi, Q \text{ are block diagonal} \\ & \text{tr}(\Phi S_{zz}) \leq 1 \\ & \text{tr}(Q S_{xx}) \leq 1, \end{aligned} \tag{118}$$

Recall that S_{zz} is diagonal matrix. So the constraint $\text{tr}(\Phi S_{zz}) \leq 1$ reduces to a single trace constraint $\text{tr}(\Phi) \leq 1$. Also, as Q is diagonal and S_{xx} has P_i on its diagonal terms, the constraint $\text{tr}(Q S_{xx}) \leq 1$ 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 i th 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 much lower dimensional space.

In addition, (118) illustrates that the uplink dual problem in this case is itself a Gaussian mutual information game. This mutual information game can be interpreted as a multiple access channel with 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 a Gaussian multi-antenna broadcast channel with individual per-antenna transmit power constraints P_1, P_2, \dots, P_n is the same as a dual multiple-access channel with a sum power constraint and with uncertain noise. The sum power constraint 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. CONCLUSIONS

This paper illustrates a minimax duality for a Gaussian mutual information optimization problem. The central feature of minimax duality is that the cost constraint in the original problem becomes the noise covariance in the dual problem and vice versa. The uplink-downlink duality between a broadcast channel and a multiple access channel can be shown to be a special case of minimax duality. Uplink-downlink duality is particularly simple under a block diagonal constraint on the noise covariance matrix and a single power constraint on the input covariance. In this case, the dual minimax optimization problem reduces to a single maximization problem. Uplink-downlink duality allows an explicit characterization of the saddle-point in the minimax problem. It shows that the least-favorable noise can be singular, it is not unique, and all least-favorable noises are related to each other via a linear estimation relation.

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 power constraint is a Gaussian multiple access channel with uncertain noise, where the noise covariance matrix is diagonally constrained. However, under general nonlinear convex constraints, duality breaks down. In this sense, the minimax expression for the broadcast channel sum capacity is more general than uplink-downlink duality.

APPENDIX

Proof of Theorem 1: The derivation leading to the statement of Theorem 1 gives a proof of minimax duality 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

$$\max_{S_{xx}} \min_{S_{zz}} \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|} \quad (119)$$

$$\text{s.t.} \quad \text{tr}(S_{xx}Q_x) \leq 1 \quad (120)$$

$$\text{tr}(S_{zz}Q_z) \leq 1 \quad (121)$$

is not full rank. Write

$$S_{xx} = U_1 S_{\tilde{x}\tilde{x}} U_1^T, \quad S_{zz} = U_2 S_{\tilde{z}\tilde{z}} U_2^T, \quad (122)$$

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

$$Q_{\tilde{x}} = U_1^T Q_x U_1, \quad Q_{\tilde{z}} = U_2^T Q_z U_2. \quad (123)$$

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

$$H = U_2 \tilde{H} U_1. \quad (124)$$

Then, the minimax problem can be reformulated as follows:

$$\max_{S_{\tilde{x}\tilde{x}}} \min_{S_{\tilde{z}\tilde{z}}} \frac{1}{2} \log \frac{|\tilde{H} S_{\tilde{x}\tilde{x}} \tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|} \quad (125)$$

$$\text{s.t.} \quad \text{tr}(S_{\tilde{x}\tilde{x}} Q_{\tilde{x}}) \leq 1 \quad (126)$$

$$\text{tr}(S_{\tilde{z}\tilde{z}} Q_{\tilde{z}}) \leq 1 \quad (127)$$

The first claim is 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 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 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.

The second claim is 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 is strictly higher, then 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 does not change the dual variables λ_x and λ_z . Now, apply the minimax duality result for the reduced channel (125). 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:

$$\max_{\Sigma_{\tilde{x}\tilde{x}}} \min_{\Sigma_{\tilde{z}\tilde{z}}} \frac{1}{2} \log \frac{|\tilde{H}^T \Sigma_{\tilde{x}\tilde{x}} \tilde{H} + \Sigma_{\tilde{z}\tilde{z}}|}{|\Sigma_{\tilde{z}\tilde{z}}|} \quad (128)$$

$$\text{s.t.} \quad \text{tr}(\Sigma_{\tilde{x}\tilde{x}} \Psi_{\tilde{x}}) \leq 1 \quad (129)$$

$$\text{tr}(\Sigma_{\tilde{z}\tilde{z}} \Psi_{\tilde{z}}) \leq 1. \quad (130)$$

The solution of the dual problem is precisely:

$$\Sigma_{\tilde{x}\tilde{x}} = \lambda_z Q_{\tilde{z}} \quad , \quad \Sigma_{\tilde{z}\tilde{z}} = \lambda_x Q_{\tilde{x}} \quad (131)$$

Now, define:

$$\Sigma_{xx} = U_2^T \Sigma_{\tilde{x}\tilde{x}} U_2 \quad , \quad \Sigma_{zz} = U_1^T \Sigma_{\tilde{z}\tilde{z}} U_1, \quad (132)$$

and

$$\Psi_x = U_2^T \Psi_{\tilde{x}} U_2 \quad , \quad \Psi_z = U_1^T \Psi_{\tilde{z}} U_1. \quad (133)$$

Then, the reduced dual minimax problem may be rewritten as:

$$\max_{\Sigma_{xx}} \min_{\Sigma_{zz}} \frac{1}{2} \log \frac{|H^T \Sigma_{xx} H + \Sigma_{zz}|}{|\Sigma_{zz}|} \quad (134)$$

$$\text{s.t.} \quad \text{tr}(\Sigma_{xx} \Psi_x) \leq 1, \quad (135)$$

$$\text{tr}(\Sigma_{zz} \Psi_z) \leq 1. \quad (136)$$

It is easy to verify that

$$C(H, Q_x, Q_z) = C(\tilde{H}, Q_{\tilde{x}}, Q_{\tilde{z}}) = C(\tilde{H}^T, \Psi_{\tilde{x}}, \Psi_{\tilde{z}}) = C(H^T, \Psi_x, \Psi_z), \quad (137)$$

and

$$\lambda_x = \nu_z \quad , \quad \lambda_z = \nu_x \quad (138)$$

and

$$(S_{xx}, S_{zz}) = (\nu_z \Psi_z, \nu_x \Psi_x) \quad , \quad (\Sigma_{xx}, \Sigma_{zz}) = (\lambda_z Q_z, \lambda_x Q_x) \quad (139)$$

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

ACKNOWLEDGMENT

The author wishes to thank Tian Lan for many useful discussions.

REFERENCES

- [1] E. Telatar, "Capacity of the multiple antenna Gaussian channel," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov/Dec 1999.
- [2] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *submitted to IEEE Trans. Info. Theory*, Nov. 2001, available at <http://www.comm.utoronto.ca/~weiyu/publications>.
- [3] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Info. Theory*, July 2003.

- [4] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Info. Theory*, October 2003.
- [5] M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [6] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *submitted to IEEE Trans. Info. Theory*, July 2001.
- [7] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. Info. Theory*, vol. 24, no. 3, pp. 374–377, May 1978.
- [8] D. Tse and P. Viswanath, "On the capacity of the multiple antenna broadcast channel," in *DIMACS Workshop on Signal Processing for Wireless Transmission*, October 2002.
- [9] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, and A. Goldsmith, "Outer bounds for multi-antenna broadcast channels," in *DIMACS Workshop on Signal Processing for Wireless Transmission*, October 2002.
- [10] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Conference on Information Sciences and Systems (CISS)*, 2004.
- [11] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *submitted to IEEE Trans. Info. Theory*, August 2002.
- [12] S. N. Diggavi and T. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Info. Theory*, vol. 47, pp. 3072–3081, Nov 2001.
- [13] S. Jafar and A. Goldsmith, "Transmitter optimization for multiple antenna cellular systems," in *IEEE Inter. Symp. Info. Theory (ISIT)*, 2002.