

The Structure of Least-Favorable Noise in Gaussian Vector Broadcast Channels

Wei Yu

ABSTRACT. The sum capacity of the Gaussian vector broadcast channel is the saddle point of a Gaussian mutual information game in which 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 established using a decision-feedback equalization approach under the assumption that the least-favorable noise covariance matrix is non-singular. This paper generalizes the above result to the case where the least-favorable noise is singular. In particular, it is shown that the least-favorable noise is not unique, and different least-favorable noise covariance matrices are related to each other by a linear estimation relation.

1. Introduction

A communication model with a single transmitter sending independent information to multiple receivers at the same time is referred to as a broadcast channel. Although the capacity region of a general broadcast channel is still an unsolved problem, recent progress has been made in the special case of the sum capacity of Gaussian vector broadcast channels [1] [2] [3] [4]. A Gaussian vector broadcast channel with K receivers can be modeled as follows:

$$(1.1) \quad \mathbf{Y}_i = H_i \mathbf{X} + \mathbf{Z}_i, \quad i = 1, \dots, K$$

where \mathbf{X} is a vector-valued transmit signal, \mathbf{Y}_i is the vector-valued receive signal, H_i is the channel matrix, and \mathbf{Z}_i is the additive Gaussian vector noise at receiver i . The solution to the Gaussian vector broadcast channel sum capacity problem is based on two key ideas. First, an achievability result can be obtained using a precoding technique that effectively pre-subtracts multi-user interference at the transmitter. The precoding technique is known as “writing-on-dirty-paper” [5]. 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.

1991 *Mathematics Subject Classification.* Primary 94A15, 94A24, 94A05; Secondary 93E24.

Key words and phrases. multi-user information theory, broadcast channel, least-favorable noise, decision-feedback equalization.

This work was supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chairs program.

The least-favorable-noise converse is based on the following: the sum capacity of the broadcast channel is clearly bounded by the capacity of the vector channel with cooperative receivers: $\max I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K)$. Further, Sato [6] observed that the sum capacity of a broadcast channel depends only on the marginal distribution of the noises and not on their correlation. This is because 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

$$(1.2) \quad \min_{S_{zz}} \max_{S_{xx}} I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K),$$

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 tight.

The achievability of the minimax mutual information expression can be established using several different methods. In [2], it is shown that if a decision-feedback equalizer is used as a joint receiver in the broadcast channel, the least-favorable noise would induce 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 in a decision-feedback equalizer is necessary and $\min I(\mathbf{X}; \mathbf{Y}_1 \cdots \mathbf{Y}_K)$ is achievable. The above 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 (1.2) is indeed the sum capacity of a Gaussian vector broadcast channel. However, the proof in [2] appears to apply only to least-favorable noise covariances that are non-singular.

In [3] and [4], a completely different approach is taken. Focusing on sum power constrained channels, [4] showed that the precoding region for a Gaussian vector broadcast channel is precisely a “dual” multiple access channel with the transmitter and receivers interchanged. Based on this duality, [4] showed that the minimization for the least-favorable noise and the maximization for the dual multiple access channel sum capacity have the same solution. In addition, [4] gives an explicit solution for the least-favorable noise. The derivation in [4] is most transparent if the least-favorable noise is non-singular. To circumvent the singular noise issue, [4] used a sequence of non-singular noise covariance to approximate the singular noise, thus proving the result in full generality. Strictly speaking, however, this solution is not constructive when the least-favorable noise is singular.

The singular noise issue is circumvented in [3] using a different technique. Although also based on duality, the treatment in [3] involves a transformation of the noise covariance matrix for the broadcast channel to an input cost constraint for the multiple access channel. This approach resolves the singularity issue, but it does not characterize the entire class of least-favorable noise.

It should be noted that the duality approach [3] [4] applies only to Gaussian vector broadcast channels with a power constraint. The decision-feedback equalization approach [2], on the other hand, applies to Gaussian vector broadcast channels with arbitrary convex constraints, and is therefore more general. However, neither approach has explicitly dealt with the issue of singular least-favorable noise so far.

Singular least-favorable noise is not necessarily an unlikely event. In fact, the least-favorable noise is almost always singular in a broadcast channel where the

number of receive dimensions is larger than the number of transmit dimensions. The main purpose of this paper is to give a complete characterization of the least-favorable noise for an arbitrary channel. A combination of the decision-feedback equalization approach and the duality approach is taken. The main results are, first, the decision-feedback equalizer approach can be generalized to the case of singular least-favorable noise. Second, least-favorable noises are not unique, and they can be characterized via a duality between the broadcast channel and the multiple access channel.

2. Decision-Feedback Equalization with Singular Noise

Without loss of generality, consider a Gaussian vector broadcast channel with two receivers $\mathbf{Y} = H\mathbf{X} + Z$, where $H^T = [H_1^T H_2^T]$, $\mathbf{Y}^T = [\mathbf{Y}_1^T \mathbf{Y}_2^T]$ and $\mathbf{Z}^T = [\mathbf{Z}_1^T \mathbf{Z}_2^T]$. The objective of this section is to show that

$$(2.1) \quad \min_{S_{zz}} \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|}$$

is achievable even if the minimizing S_{zz} is singular. Recall that the minimization is over all S_{zz} with fixed block diagonal terms. The fixed block diagonals can be assumed to be identity matrices with no loss of generality. This is a generalization of the earlier result in [2] where the case for non-singular S_{zz} is solved. The earlier result showed that if S_{zz} satisfies the Karush-Kuhn-Tucker (KKT) condition for the minimization problem

$$(2.2) \quad S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1} = \begin{bmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{bmatrix},$$

then there exists a decision-feedback equalizer (DFE) with a diagonal feedforward matrix. When the channel matrix H is full rank, the minimizing noise covariance matrix is always non-singular. However, this is not the case when H has more rows than columns. This situation corresponds to the case where there are more receiver antennas than transmit antennas in the broadcast channel. To extend the achievability result to accommodate singular least-favorable noises, both the KKT condition and the design of the DFE need to be generalized.

The decision-feedback equalizer that is required to accommodate singular noise differs from the conventional structure in one crucial aspect. A conventional DFE never processes more output dimensions than input dimensions. For example, in a channel with two transmit antennas and three receive antennas, a conventional DFE always reduces the two-by-three channel to a two-by-two channel by receiver joint processing. However, in a broadcast channel, receivers cannot cooperate and joint processing is not possible. Thus, a non-trivial generalization of the conventional DFE structure is required.

Without loss of generality, S_{xx} is assumed to be fixed and full-rank. Let $S_{xx} = V\Sigma V^T$ be an eigenvalue decomposition. It is convenient to set $H' = HV\sqrt{\Sigma}M$ to be the effective channel, and let the input covariance matrix be simply an identity matrix. The choice of M will be made later.

Suppose that S_{zz} is a low-rank solution to the minimization problem (2.1). Decompose S_{zz} in the following form:

$$(2.3) \quad S_{zz} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_{\tilde{z}\tilde{z}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix},$$

where $S_{\tilde{z}\tilde{z}}$ is invertible and $[U_1 U_2]$ is an orthonormal matrix. The null space of S_{zz} must be a subspace of the null space of the channel, (because otherwise the capacity would be infinity). Thus, it must be possible to express the channel matrix H' as:

$$(2.4) \quad H' = U_1 \tilde{H}.$$

The minimization problem now becomes

$$(2.5) \quad \min_{S_{\tilde{z}\tilde{z}}} \frac{1}{2} \log \frac{|\tilde{H}\tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|}$$

s.t. $U_1 S_{\tilde{z}\tilde{z}} U_1^T$ has identities on the diagonal.

A necessary condition for the least-favorable noise is

$$(2.6) \quad S_{\tilde{z}\tilde{z}}^{-1} - (\tilde{H}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} = U_1^T \begin{bmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{bmatrix} U_1.$$

The objective of this section is to show that if the noise covariance matrix satisfies the above condition, then there exists a decision-feedback equalizer whose feedforward matrix is diagonal. The development of decision-feedback equalizer follows that of [2]. The difference between the new proof and the original treatment is highlighted here.

The derivation of the decision-feedback equalizer is based on minimum mean-square error (MMSE) estimation. The key observation is that MMSE estimator is not unique when both the channel and the noise are rank deficient. Consider the MMSE estimation of \mathbf{X} given $\mathbf{Y} = H\mathbf{X} + \mathbf{Z}$. The MMSE estimation is a matrix multiplication $\hat{\mathbf{X}} = W\mathbf{Y}$, where W satisfies a normal equation:

$$(2.7) \quad W S_{yy} = S_{xy}.$$

Because both S_{zz} and H are low rank, S_{yy} is also low rank. It is not difficult to verify that

$$(2.8) \quad W = \tilde{H}^T (\tilde{H}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} U_1^T + S U_2^T$$

satisfies the normal equation for any choice of S . A different DFE can be designed for each choice of S . The rest of the section is devoted to showing that there exists a choice of S (along with a choice of M) that makes the DFE feedforward matrix diagonal.

The first step in the design of a decision-feedback equalizer is noise whitening. Define

$$(2.9) \quad \hat{H} = S_{\tilde{z}\tilde{z}}^{-\frac{1}{2}} \tilde{H}.$$

The MMSE estimator matrix W can be re-written as

$$(2.10) \quad W = S_{xx} \hat{H}^T (\hat{H}\hat{H}^T + I)^{-1} S_{\tilde{z}\tilde{z}}^{-\frac{1}{2}} U_1^T + S U_2^T$$

The structure of the feedforward matrix involves a Cholesky factorization of $R_b = \mathbf{E}[(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^T]$. It turns out that R_b is independent of the choice of S . Using the matrix inversion lemma, it can be shown that

$$(2.11) \quad R_b = (\hat{H}^T \hat{H} + I)^{-1}.$$

Further, W can be re-written as follows:

$$(2.12) \quad W = (\hat{H}^T \hat{H} + I)^{-1} \hat{H}^T S_{\tilde{z}\tilde{z}}^{-\frac{1}{2}} U_1^T + S U_2^T.$$

In a decision-feedback equalizer, half of the Cholesky factorization $R_b = G^{-1}\Delta^{-1}G^{-T}$ is placed in the feedback section, and the remaining half is placed in the feedforward section. Thus, the feedforward matrix has the form:

$$(2.13) \quad F = \Delta^{-1}G^{-T}\hat{H}^T S_{\tilde{z}\tilde{z}}^{-\frac{1}{2}}U_1^T + S'U_2^T.$$

The goal of this section is to use the generalized least-favorable noise condition (2.6) to show that F can be made diagonal. Using the matrix inversion lemma,

$$(2.14) \quad U_1^T \begin{bmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{bmatrix} U_1 = S_{\tilde{z}\tilde{z}}^{-1} - (\tilde{H}\tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} \\ = \hat{H}(I + \hat{H}\hat{H}^T)^{-1}\hat{H}^T.$$

With some algebra (see [2] for details), it is possible to prove that, by an appropriate choice of M , the Cholesky factorization of R_b can be made to give:

$$(2.15) \quad \Delta^{-\frac{1}{2}}G^{-T}\hat{H}^T S_{\tilde{z}\tilde{z}}^{-\frac{1}{2}} = \begin{bmatrix} \sqrt{\Psi_1} & 0 \\ 0 & \sqrt{\Psi_2} \end{bmatrix}.$$

Therefore, by choosing

$$(2.16) \quad S' = \Delta^{-\frac{1}{2}} \begin{bmatrix} \sqrt{\Psi_1} & 0 \\ 0 & \sqrt{\Psi_2} \end{bmatrix} U_2,$$

the feedforward matrix becomes

$$(2.17) \quad F = \Delta^{-\frac{1}{2}} \begin{bmatrix} \sqrt{\Psi_1} & 0 \\ 0 & \sqrt{\Psi_2} \end{bmatrix} (U_1U_1^T + U_2U_2^T),$$

which is diagonal since

$$(2.18) \quad U_1U_1^T + U_2U_2^T = [U_1 \ U_2] \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = I.$$

To summarize, when the least-favorable noise is singular, it must satisfy a modified KKT condition (2.2). The decision-feedback equalizer structure with the singular noise is not unique. However, among the class of decision-feedback equalizers, there exists one whose feedforward matrix is diagonal. This is true for any fixed S_{xx} . Thus, in a Gaussian vector broadcast channel, the minimum mutual information $I(\mathbf{X}; \mathbf{Y}_1, \mathbf{Y}_2)$ is an achievable sum rate, even when the least favorable noise covariance matrix is singular.

3. The Structure of Least-Favorable Noise

3.1. Uplink-Downlink Duality. The previous section shows that whenever the noise covariance matrix satisfies the least-favorable noise condition, (2.2) or (2.6), the mutual information $I(\mathbf{X}; \mathbf{Y})$ is achievable in a broadcast channel even without receiver cooperation. However, it does not give an explicit method to compute the least-favorable noise. Neither (2.2) nor (2.6) appears to have a closed-form solution. Further, although $I(\mathbf{X}; \mathbf{Y})$ is a convex function of S_{zz} , the numerical computation of the least-favorable noise is not necessarily easy, especially when the minimizing noise covariance matrix is singular. Fortunately, there is one important special case for which the computation is easy. This is when the transmit covariance S_{xx} is being maximized at the same time.

Consider a Gaussian vector broadcast channel with a power constraint $\text{tr}(S_{xx}) \leq P$. Sato's outer bound states that the sum capacity is less than $\min \max I(\mathbf{X}; \mathbf{Y})$,

where the maximization is over the power constraint and the minimization is over all possible noise correlations. The achievability result in the previous section states that a sum rate of $\max \min I(\mathbf{X}; \mathbf{Y})$ is achievable. Since the mutual information is concave in S_{xx} and convex in S_{zz} , the minimization and maximization operations can be interchanged. Thus, the sum capacity of a Gaussian vector broadcast channel is precisely

$$(3.1) \quad C = \max_{S_{xx}} \min_{S_{zz}} \frac{1}{2} \log \frac{|HS_{xx}H^T + S_{zz}|}{|S_{zz}|},$$

where $\text{tr}(S_{xx}) \leq P$ and S_{zz} has fixed diagonal terms. Surprisingly, this minimax problem has a special structure that gives it a particularly simple solution. The new ingredient is the uplink-downlink duality¹.

Uplink-downlink duality refers to the observation that the achievable rate region of a broadcast channel using “writing-on-dirty-paper” precoding is the same as the capacity region of a “dual” multiple access channel with the roles of the input and output terminals interchanged and with a sum power constraint across all input terminals. This duality was established in [4] and [3]. In this section, a different derivation of duality is given. The main purpose is to illustrate that the minimax problem can be solved efficiently via duality.

To simplify matters, let’s first 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 (3.1). Because the objective function in (3.1) is concave in S_{xx} and convex in S_{zz} , the KKT condition completely characterizes the saddle point. The KKT condition is:

$$(3.2) \quad H^T(HS_{xx}H^T + S_{zz})^{-1}H = \lambda I$$

$$(3.3) \quad S_{zz}^{-1} - (HS_{xx}H^T + S_{zz})^{-1} = \Psi$$

where λ is the dual variable associated with the power constraint and Ψ is a diagonal matrix of dual variables associated with the diagonal constraint on the noise covariance matrix. Multiplying (3.3) by H^T on the left and H on the right and substituting in (3.2), we obtain

$$(3.4) \quad H^T S_{zz}^{-1} H = H^T \Psi H + \lambda I.$$

The above is equivalent to

$$(3.5) \quad H(H^T \Psi H + \lambda I)^{-1} H^T = S_{zz}.$$

Observe that (3.5) is precisely the KKT condition for a multiple access channel with diagonal matrix Ψ as the transmit covariance matrix and $\text{tr}(\Psi S_{zz}) \leq P$ as the input constraint. Since S_{zz} has 1’s on the diagonal, the constraint is equivalent to a sum power constraint on Ψ . After a proper scaling of the power constraint, Ψ , λ and S_{zz} , it can be shown that (3.5) is precisely the KKT condition for the following

¹Uplink-downlink duality can be generalized beyond the power constrained broadcast channels. Duality exists whenever the input constraint is a linear covariance constraint. See [7] for a detailed discussion.

optimization problem

$$(3.6) \quad \begin{aligned} \max_D \quad & \frac{1}{2} \log |H^T D H + I| \\ \text{s.t.} \quad & D \text{ is diagonal} \\ & \text{trace}(D) \leq P, \\ & D \geq 0, \end{aligned}$$

(where $D = \Psi/\lambda$.) Thus, there is a duality between a multiple access channel and a broadcast channel. The solution to the minimax problem (3.1) can be obtained from the solution to a maximization problem (3.6). Since (3.6) is much easier to solve numerically (see e.g. [8] [9]), uplink-downlink duality gives an efficient way to solve the minimax problem (3.1).

3.2. 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, (3.5) gives an explicit formula for the covariance matrix of the least-favorable noise. This formula has also appeared in both [3] and [4]. However, (3.5) is derived assuming that the channel matrix H is square and invertible and both the least-favorable noise and the maximizing input covariance matrices are full rank. In the general case, this is not necessarily true. In particular, the maximizing input covariance matrix D of the dual multiple access channel may not have non-zero entries everywhere on its diagonal. When the diagonals of D are all non-zero, the KKT condition for the maximization problem (3.6) ensures that the diagonal terms of $H(H^T \Psi H + \lambda I)^{-1} H^T$ are all 1's. Otherwise, slack variables need to be introduced and $H(H^T \Psi H + \lambda I)^{-1} H^T$ does not necessarily have 1's on its diagonal. Thus, it cannot be a valid choice of the least-favorable noise.

The purpose of the rest of the section is to show that a least-favorable noise covariance can be obtained by adding a positive semi-definite matrix to $H(H^T \Psi H + \lambda I)^{-1} H^T$ so that the sum of the two matrices has 1's 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. Consider the candidate least-favorable noise, re-labeled as $S_{zz}^{(0)}$:

$$(3.7) \quad S_{zz}^{(0)} = H(H^T \Psi H + \lambda I)^{-1} H^T.$$

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

$$(3.8) \quad S_{xx} = (\lambda I)^{-1} - (H^T \Psi H + \lambda I)^{-1}.$$

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 have fewer 1's on its diagonal, thus enlarging the class of positive semi-definite matrices that can be added to its diagonal terms. For the rest of the proof, it is assumed that this channel reduction step has already taken place.

The proof is fairly lengthy. To simplify the algebra, the broadcast channel considered here is assumed to have only a single antenna at each receiver. This special case exhibits the essential features of the least-favorable noise.

The first step of the proof is to verify that even though its diagonals may not be 1's, the candidate noise covariance matrix (3.7) satisfies the least-favorable noise condition. Assuming that channel has already been reduced, S_{zz} can be re-written as

$$(3.9) \quad S_{zz}^{(0)} = U_1 S_{\tilde{z}\tilde{z}} U_1^T$$

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

$$(3.10) \quad H = U_1 \tilde{H},$$

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 $S_{zz}^{(0)} = H(H^T \Psi H + \lambda I)^{-1} H^T$. The minimax problem is now

$$(3.11) \quad \begin{aligned} \max_{S_{xx}} \min_{S_{\tilde{z}\tilde{z}}} & \quad \frac{1}{2} \log \frac{|\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}}|}{|S_{\tilde{z}\tilde{z}}|} \\ \text{s.t.} & \quad \text{tr}(S_{xx}) \leq P \\ & \quad U_1 S_{\tilde{z}\tilde{z}} U_1^T \text{ has 1's on diagonal} \\ & \quad S_{xx} \geq 0, \end{aligned}$$

The KKT condition of the minimax problem is:

$$(3.12) \quad \tilde{H}^T (\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} \tilde{H} = \lambda I$$

$$(3.13) \quad S_{\tilde{z}\tilde{z}}^{-1} - (\tilde{H} S_{xx} \tilde{H}^T + S_{\tilde{z}\tilde{z}})^{-1} = U_1^T \Psi U_1$$

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

$$(3.14) \quad H(H^T \Psi H + \lambda I)^{-1} H^T = U_1 S_{\tilde{z}\tilde{z}} U_1^T = S_{zz}^{(0)}.$$

Thus, the candidate $S_{zz}^{(0)}$ satisfies the least-favorable noise condition.

The rest of the proof is devoted to showing that the least-favorable noise condition remains to be satisfied if a positive semi-definite matrix, denoted as S'_{zz} , is added to $S_{zz}^{(0)}$ to makes its diagonal terms 1. 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 Ψ only has positive entries on its upper-left diagonal, and S'_{zz} has non-zero entries only in the lower-right corner:

$$(3.15) \quad \Psi = \begin{bmatrix} \Psi' & 0 \\ 0 & 0 \end{bmatrix}, \quad S'_{zz} = \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix}.$$

Let n be the number of receivers in the broadcast channel. So, Ψ , S_{zz} and S'_{zz} are $n \times n$ matrices. Let the dimension of Ψ' 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 (3.8) the optimal S_{xx} can be expressed as:

$$(3.16) \quad \begin{aligned} S_{xx} &= (\lambda I)^{-1} - (H^T \Psi H + \lambda I)^{-1} \\ &= (\lambda I)^{-1} H^T \sqrt{\Psi} (I + \sqrt{\Psi} H H^T \sqrt{\Psi})^{-1} \sqrt{\Psi} H (\lambda I)^{-1}. \end{aligned}$$

Thus, the rank of S_{xx} is the same as the rank of $H^T \Psi H$. The channel reduction step guarantees that S_{xx} is full rank. This implies that $H^T \Psi H$ must be full rank.

For this to be true, Ψ 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 in 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:

$$(3.17) \quad S'_{zz} = \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} = [U_1 \quad U_2] \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}.$$

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 that this U_2 may be different from the U_2 in section 2.)

Partition U_1 and U_2 into sub-matrices:

$$(3.18) \quad U_1 = \begin{bmatrix} u_{11}^T \\ u_{12}^T \end{bmatrix} \quad U_2 = \begin{bmatrix} u_{21}^T \\ u_{22}^T \end{bmatrix}.$$

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

$$(3.19) \quad S_{11} - S_{12}S_{22}^{-1}S_{21} = 0.$$

This is because

$$(3.20) \quad \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} [U_1 \quad U_2] = I,$$

so, from (3.17),

$$(3.21) \quad \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} u_{11}^T & u_{12}^T \\ u_{21}^T & u_{22}^T \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix}.$$

This allows S_{ij} to be solved explicitly:

$$(3.22) \quad S_{11} = u_{12}^T S u_{12} \quad S_{12} = u_{12}^T S u_{22}$$

$$(3.23) \quad S_{21} = u_{22}^T S u_{12} \quad S_{22} = u_{22}^T S u_{22}.$$

Therefore,

$$(3.24) \quad S_{12}S_{22}^{-1}S_{21} = u_{12}^T S u_{22} (u_{22}^T S u_{22})^{-1} u_{22}^T S u_{12} = S_{11},$$

thus proving (3.19).

The second useful fact is the following:

$$(3.25) \quad u_{21} = u_{11} u_{12}^T u_{22}^{-T}.$$

The proof is based on (3.17)

$$(3.26) \quad \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} u_{11}^T & u_{12}^T \\ u_{21}^T & u_{22}^T \end{bmatrix}.$$

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

$$(3.27) \quad 0 = u_{11}S_{11}u_{12}^T + u_{21}S_{21}u_{12}^T + u_{11}S_{12}u_{22}^T + u_{21}S_{22}u_{22}^T.$$

Substituting in (3.22)–(3.23), the above is equal to

$$(3.28) \quad 0 = (u_{11}u_{12}^T + u_{21}u_{22}^T)S(u_{11}u_{11}^T + u_{22}u_{22}^T).$$

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

$$(3.29) \quad u_{11}u_{12}^T + u_{21}u_{22}^T = 0,$$

thus proving (3.25).

Finally, we are ready to show that $S_{zz}^{(0)} + S'_{zz}$ satisfies the least-favorable noise condition. Starting from (3.13)

$$(3.30) \quad S_{\bar{z}\bar{z}}^{-1} - (\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1} = U_1^T\Psi U_1,$$

the objective is to prove that

$$(3.31) \quad \begin{bmatrix} S_{\bar{z}\bar{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1} - \begin{bmatrix} \tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1} = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \Psi [U_1 \ U_2].$$

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

$$(3.32) \quad \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \Delta_D^{-1} & -\Delta_D^{-1}BD^{-1} \\ -D^{-1}C\Delta_D^{-1} & D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \end{bmatrix},$$

where $\Delta_D = A - BD^{-1}C$. To evaluate the matrix inversions in (3.31), Δ_D^{-1} can be simplified using (3.19):

$$(3.33) \quad \Delta_D = S_{\bar{z}\bar{z}} + S_{11} - S_{12}S_{22}^{-1}S_{21} = S_{\bar{z}\bar{z}}.$$

Thus, the first matrix inversion in (3.31) is equal to

$$(3.34) \quad \begin{bmatrix} S_{\bar{z}\bar{z}}^{-1} & -S_{\bar{z}\bar{z}}^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}S_{\bar{z}\bar{z}}^{-1} & S_{22}^{-1} + S_{22}^{-1}S_{21}S_{\bar{z}\bar{z}}^{-1}S_{12}S_{22}^{-1} \end{bmatrix},$$

and the second matrix inversion in (3.31) can be expanded similarly:

$$(3.35) \quad \begin{bmatrix} (\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1} & -(\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}(\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1} & S_{22}^{-1} + S_{22}^{-1}S_{21}(\tilde{H}S_{xx}\tilde{H}^T + S_{\bar{z}\bar{z}})^{-1}S_{12}S_{22}^{-1} \end{bmatrix}.$$

Now, using (3.30), the difference between the two can now be simplified:

$$(3.36) \quad \begin{bmatrix} U_1^T\Psi U_1 & -U_1^T\Psi U_1 S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}U_1^T\Psi U_1 & S_{22}^{-1}S_{21}U_1^T\Psi U_1 S_{12}S_{22}^{-1} \end{bmatrix}.$$

To prove (3.31), it remains to show that the above is equal to

$$(3.37) \quad \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \Psi [U_1 \ U_2] = \begin{bmatrix} U_1^T\Psi U_1 & U_1^T\Psi U_2 \\ U_2^T\Psi U_1 & U_2^T\Psi U_2 \end{bmatrix}.$$

Comparing (3.36) with (3.37), it is clear that the two are equal if the following holds:

$$(3.38) \quad \Psi U_2 = -\Psi U_1 S_{12}S_{22}^{-1}.$$

Recall that Ψ has non-zero entries only in its upper-left diagonal. So, the left-hand side of (3.38) is

$$(3.39) \quad \Psi U_2 = \begin{bmatrix} \Psi' & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = \Psi' u_{21}.$$

The right-hand side of (3.38) is

$$(3.40) \quad -\Psi U_1 S_{12} S_{22}^{-1} = -\Psi' u_{11} u_{12}^T u_{22}^{-T}.$$

By (3.25), the left-hand side is equal to the right-hand side. This establishes the least-favorable noise condition (3.31).

3.3. Linear Estimation Interpretation. The previous section shows that the least-favorable noise in a Gaussian minimax mutual information game is of the type $S_{zz}^{(0)} + S'_{zz}$. Note that adding the term S'_{zz} does not change the mutual information $I(\mathbf{X}; \mathbf{Y})$. This can be seen using the following argument. The value of the minimax expression with $(S_{xx}, S_{zz}^{(0)})$ is

$$(3.41) \quad C = \frac{1}{2} \log \frac{|\tilde{H} S_{xx} \tilde{H}^T + S_{\bar{z}\bar{z}}|}{|S_{\bar{z}\bar{z}}|}.$$

With $(S_{xx}, S_{zz}^{(0)} + S'_{zz})$, the capacity is

$$(3.42) \quad C = \frac{1}{2} \log \frac{\begin{vmatrix} \tilde{H} S_{xx} \tilde{H}^T + S_{\bar{z}\bar{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}}{\begin{vmatrix} S_{\bar{z}\bar{z}} + S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}}.$$

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

$$(3.43) \quad \begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D| \cdot |A - B D^{-1} C|,$$

it is not difficult to show that

$$(3.44) \quad \frac{|S_{22}| \cdot |\tilde{H} S_{xx} \tilde{H}^T + S_{\bar{z}\bar{z}} + S_{11} - S_{12} S_{22}^{-1} S_{21}|}{|S_{22}| \cdot |S_{\bar{z}\bar{z}} + S_{11} - S_{12} S_{22}^{-1} S_{21}|} = \frac{|\tilde{H} S_{xx} \tilde{H}^T + S_{\bar{z}\bar{z}}|}{|S_{\bar{z}\bar{z}}|}.$$

Therefore the capacity expressions (3.41) and (3.42) have the same value.

In some sense, $S_{zz}^{(0)}$ is the “smallest” possible noise in the class of least-favorable noises. This is because after the channel reduction step, $H S_{xx} H^T$ 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.

$S_{zz}^{(0)}$ is not, of course, the only such noise. The previous section shows that $(S_{xx}, S_{zz}^{(0)} + S'_{zz})$ also satisfies the least-favorable noise condition. However, the previous section did not explicitly prove that S_{xx} is the water-filling covariance matrix for $S_{zz}^{(0)} + S'_{zz}$. But, 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 $S_{zz}^{(0)} + S'_{zz}$. Thus, the entire class of $(S_{xx}, S_{zz}^{(0)} + S'_{zz})$ are saddle-points of the minimax problem.

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

$$(3.45) \quad \mathbf{Z}_u^{(0)} = \begin{bmatrix} \mathbf{z}_1^0 \\ 0 \end{bmatrix}, \quad \mathbf{Z}'_u = \begin{bmatrix} \mathbf{z}'_1 \\ \mathbf{z}'_2 \end{bmatrix}.$$

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

$$(3.46) \quad S_{11} - S_{12} S_{22}^{-1} S_{21} = 0.$$

The above argument shows that the entire class of least-favorable noises are related to each other via a linear estimation relation². As the previous section shows, this relation is crucial in the derivation of the least-favorable condition.

4. Conclusions

This paper deals with the least-favorable noise in a Gaussian vector broadcast channel. The least-favorable noise is not necessarily non-singular. The achievability proof for the broadcast channel sum capacity is extended to the case of singular least-favorable noises. The proof is based on the fact that the decision-feedback equalizer is not unique when the noise is singular. Among all possible equalizers, there exists one with a diagonal feedforward matrix.

In the second part of the paper, a new derivation for the duality between the multiple access channel and the broadcast channel is given. The duality relation gives a natural characterization of the entire class of least-favorable noises. The least-favorable noise is not unique, and all least-favorable noises are related to each other by a linear estimation relation.

References

- [1] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *submitted to IEEE Trans. Info. Theory*, July 2001.
- [2] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *submitted to IEEE Trans. Info. Theory*, Nov. 2001.
- [3] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and downlink-uplink duality," *submitted to IEEE Trans. Info. Theory*, July 2002.
- [4] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *submitted to IEEE Trans. Info. Theory*, Aug. 2002.
- [5] M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [6] 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.
- [7] W. Yu, "Uplink-downlink duality via minimax duality," *Canadian Workshop on Information Theory*, May 2003.

²There is a slight subtlety in that, strictly speaking, \mathbf{Z}' needs not be Gaussian. However, this does not affect the sum capacity. See [10] for a discussion.

- [8] W. Yu, "A dual decomposition approach to the sum power Gaussian vector multiple access channel sum capacity," *Conference on Information Sciences and Systems (CISS)*, March 2003.
- [9] S. Vishwanath, R. Wonjong, N. Jindal, S. Jafar, and A. Goldsmith, "Sum power iterative waterfilling for Gaussian vector broadcast channels," *IEEE Inter. Symp. Info. Theory (ISIT)*, June 2003. Also in *Asilomar Conf.*, Nov. 2002.
- [10] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Info. Theory*, vol. 47, pp. 3072–3081, Nov. 2001.

ELECTRICAL AND COMPUTER ENGINEERING DEPARTMENT, UNIVERSITY OF TORONTO
Current address: 10 King's College Road, Toronto, Ontario M5S 3G4, Canada
E-mail address: weiyu@comm.utoronto.ca