The Worst Additive Noise under a Covariance Constraint

Suhas N. Diggavi †and Thomas M. Cover‡
†180, Park Avenue, AT& T Shannon Laboratories,
Bldg 103, Florham Park, New Jersey, NJ 07932, USA.
Email: suhas@research.att.com, Fax: (973) 360-8178
‡254, Packard Bldg, Information Systems Laboratory,
Stanford University, Stanford, CA 94305, USA.
Email:{suhas,cover}@isl.stanford.edu, Fax: (973) 360-8178

Abstract

The maximum entropy noise under a lag p autocorrelation constraint is known by Burg's theorem to be the p^{th} order Gauss-Markov process satisfying these constraints. The question is, what is the worst additive noise for a communication channel given these constraints? Is it the maximum entropy noise?

The problem becomes one of extremizing the mutual information over all noise processes with covariances satisfying the correlation constraints $R_0, \ldots, R_p$. For high signal powers, the worst additive noise is Gauss-Markov of order p as expected. But for low powers the worst additive noise is Gaussian with a covariance matrix in a convex set which depends on the signal power.

Index terms: Burg's theorem, worst additive noise, mutual information game.

I. Introduction

This paper treats a simple problem: What is the noisiest noise under certain constraints? There are two possible contexts in which we might ask this question. One is, what is the noisiest random process satisfying, for example, a lag covariance constraint, $\mathbb{E}[Z_i Z_{i+k}] = R_k, k = 0, \ldots p$. Thus, we ask for the maximum entropy rate for such a process. It is well known from Burg's work [1], [2] that the maximum entropy noise process under p lag constraints is the pth order Gauss Markov process satisfying these constraints, *i.e.* it is the process that has minimal dependency on the past given the covariance constraints.

Another context in which we might ask this question is for an additive noise channel $Y = X + Z$, where the noise Z has covariance constraints $R_0, \ldots, R_p$ and the signal X has a power constraint P. What is the worst possible additive noise subject to these constraints? We expect the answer to be the maximum entropy noise, as for the first problem. Indeed, we find this is the case, but only when

This work was supported in part by NSF Grant NSF CCR-9973134 and ARMY (MURI) DAAD19-99-1-0252 This work was presented in part at the International Symposium on Information Theory (ISIT) 1997.

the signal power is high enough to fill the spectrum of the maximum entropy noise (yielding a white noise sum).

Consider the channel

$$Y_k = X_k + Z_k, \tag{1}$$

where X_k is the transmitted signal and Z_k is the additive noise. Transmission over additive Gaussian noise channels has been well studied over the past several decades [1]. The capacity is achieved by using Gaussian signaling and waterfilling over the noise spectrum [1]. The question of communication over partially known additive noise channels is addressed in [3], [4], [5], where the class of memoryless noise processes with average power constraint N_0 is considered. A game-theoretic problem [3], [4], [5] is formulated with a mutual information pay-off, where the sender maximizes mutual information, and the noise minimizes it, subject to average power constraints. It has been shown that an i.i.d. Gaussian signaling scheme and an i.i.d. Gaussian noise distribution are robust, in that any deviation of either the signal or noise distribution reduces or increases (respectively) the mutual information. Hence the solution to this game-theoretic problem yields a rate of $\frac{1}{2}\log(1+P/N_0)$, where P and N_0 are the signal and noise power constraints respectively.

An excellent survey for communication under channel uncertainties is given in [6]. In [7], [8], a game-theoretic problem with Gaussian inputs transmitted over a jamming channel (having an average power constraint) is studied under a mean-squared error pay-off function (for estimation/detection). The problem of worst power constrained noise when the inputs are limited to the binary alphabet is considered in [9].

The more general M-dimensional problem with average noise power constraint is considered in [10], where it is shown that even when the channel is not restricted to be memoryless, the white Gaussian codebook and white Gaussian noise constitute a unique saddlepoint. In [11], [12] (and references therein) it was shown that a Gaussian codebook and minimum Euclidean distance decoding achieves rate $\frac{1}{2}\log(1+P/N_0)$ under an average power constraint. Therefore, for average signal and noise power constraints the maximum entropy noise is the worst additive noise for communication. We ask whether this principle is true in more generality.

Suppose the noise is not memoryless and we have covariance constraints. If the signal is Gaussian with covariance K_x and the noise is Gaussian with covariance K_z, the mutual information $I(X;X+Z)$ is given by $I(X;X+Z) = \frac{1}{2}\log(\frac{|K_x+K_z|}{|K_z|})$. It is well known that the mutual information is maximized by choosing a signal covariance K_x that waterfills K_z [1]. The question we ask is about communication over partially known additive noise channels subject to covariance constraints. We first formulate the game-theoretic problem with mutual information as the pay-off. The signal maximizes the mutual information and the noise minimizes it by choosing distributions subject to covariance constraints. Note that the problem considered is similar in formulation to the compound channel problem [13], and therefore is more benign than the allowed noise in arbitrarily varying channels [6], [12]. In [14], [15] the problem where a memoryless interference which is statistically dependent on the input was considered. In this paper the additive noise is independent of the input but need not be memoryless.

We first show that Gaussian signaling and Gaussian noise constitute a saddlepoint to the mutual information game with covariance constraints. Therefore we can restrict our attention to the solution of a determinant game with payoff $\frac{1}{2}\log(\frac{|K_x+K_z|}{|K_z|})$. To solve this problem one chooses the signal covariance K_x and noise covariance K_z to maximize and minimize (respectively) the payoff $\frac{1}{2}\log(\frac{|K_x+K_z|}{|K_z|})$ subject to covariance constraints. Throughout this paper we impose an expected power constraint on the signal,

$$\mathbb{E}\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\mathrm{tr}\ (K_x) \leq P.$$

We will also assume that the noise covariance K_z lies in a given convex set $\mathcal{K}_z$, but the noise distribution is otherwise unspecified. For example, the set $\mathcal{K}_z$ of covariances K_z satisfying correlation constraints $R_0, \ldots, R_p$ is a convex set. Also, for some of the results in the paper, we assume $K_z > 0,$ for all $K_z \in \mathcal{K}_z$, *i.e.* the noise processes are not degenerate.

We study the properties of the saddlepoints to the pay-off function $\frac{1}{2}\log(\frac{|K_x+K_z|}{|K_z|})$. We show that the signaling covariance matrix K_x is unique and waterfills a set of worst noise covariance matrices. The set of worst noise covariance matrices is shown to be convex and hence the signaling scheme is protected against any mixture of noise covariances. Therefore choosing a Gaussian signaling scheme with covariance K_x^* which waterfills the class of worst covariance matrices will achieve the minimax mutual information. This establishes a single optimal strategy for the sender (Gaussian with a certain covariance matrix designed to waterfill the minimax noise) and a convex set of possible noise covariances, all of which look the same "below the water line".

Next, we re-examine the question of whether the maximum entropy noise is the worst additive noise when we have a banded matrix constraint specified up to a certain covariance lag on the noise covariance matrix. In this case we show that if we have sufficient input power, the maximum entropy noise is also the worst additive noise in the sense that it achieves the saddlepoint and minimizes the mutual information.

We put forth the game theoretic problem in Section II, establish the existence of a saddlepoint and also study its properties. We consider the banded noise covariance constraint in Section III. In Section IV we show this minimax rate is actually achievable using a random Gaussian codebook and Mahalanobis distance decoding.

II. Problem Formulation

The general problem is that of finding the maximum reliable communication rate over all noise distributions subject to covariance constraints. Throughout this section we assume that the constraint sets $\mathcal{K}_x$ and $\mathcal{K}_z$ are closed, bounded and convex. Note that we have implicitly associated with $\mathcal{K}_x$ and $\mathcal{K}_z$ the topology of $n \times n$ symmetric matrices, *i.e.,* that associated with $\mathbb{R}^M$, where $M = n(n+1)/2$. We need to show that there exists a codebook that is simultaneously good for all such noise distributions. We first guess that this problem can be solved by solving the minimax mutual information game. Later in Section IV we examine a random coding scheme and a decoding rule that achieves this rate. Hence the signal designer maximizes the mutual information and the noise (nature) minimizes it, and this is the minimax communication capacity.

Therefore we consider minimax problem

$$\inf_{p_Z \in \mathcal{Z}} \sup_{p_X \in \mathcal{X}} I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}^{(n)}), \tag{2}$$

where $\mathcal{Z} = Closure\{p_Z : \mathbb{E}[Z] = 0, K_z \in \mathcal{K}_z\}$, $\mathcal{X} = Closure\{p_X : \mathbb{E}[X] = 0, tr(K_x) \leq nP\}$, and p_X, p_Z are probability measures on defined on the Borel σ-algebra of $\mathbb{R}^n$. The closure is defined in terms of the weak topology of probability measures on $\mathbb{R}^n$ ([16], Section 2.2). We note that if the covariance constraint sets $\mathcal{K}_x, \mathcal{K}_z$ are closed, then the sets $\mathcal{X}, \mathcal{Z}$ can be proved to be closed (without the closure operation) if the random processes are assumed to have finite fourth moments. If there exist probability measures p_X^* and p_Z^* such that

$$I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}^{*(n)}) \leq I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}^{*(n)}) \leq I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}^{(n)}), \tag{3}$$

for all $p_X \in \mathcal{X}, p_Z \in \mathcal{Z}$, where $\mathbf{X}^{*(n)}$ and $\mathbf{Z}^{*(n)}$ are distributed according to measures p_X^* and p_Z^* respectively, then (p_X^*, p_Z^*) is defined as a saddlepoint for $I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}^{(n)})$, and $I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}^{*(n)})$ is called the value of the game. To show the existence of such a saddlepoint we examine some properties of the mutual information under input and noise constraints. We first show that there exist saddlepoints which have Gaussian probability measures p_X and p_Z.

Lemma II.1: ([1], Chapter 9) Let $\mathbf{Z}$ and $\mathbf{Z}_G$ be random vectors in $\mathbb{R}^n$ with the same covariance matrix K_z. If $\mathbf{Z}_G \sim \mathcal{N}(0, K_z)$ and $\mathbf{Z}$ has any other distribution, then

$$\mathbb{E}_{\mathbf{Z}_G}[\log(f_{\mathbf{Z}_G}(\mathbf{Z}))] = \mathbb{E}_{\mathbf{Z}}[\log(f_{\mathbf{Z}_G}(\mathbf{Z}))], \tag{4}$$

where $f_{\mathbf{Z}_G}(\cdot)$ denotes the probability density function of $\mathbf{Z}_G$, and $\mathbb{E}_{\mathbf{Z}_G}[\cdot]$ and $\mathbb{E}_{\mathbf{Z}}[\cdot]$ denote the expectations with respect to $\mathbf{Z}_G$ and $\mathbf{Z}$ respectively.

The following result (Lemma II.2) has been proved by Ihara [17] based on a result by Pinsker [18]. The alternative proof given below shows the condition for which equality holds. In the proof we assume the noise has a probability density function.

Lemma II.2: Let $\mathbf{X}_G \sim \mathcal{N}(0, K_x)$, and let $\mathbf{Z}$ and $\mathbf{Z}_G$ be random vectors in $\mathbb{R}^n$ (independent of $\mathbf{X}_G$) with the same covariance matrix K_z. If $\mathbf{Z}_G \sim \mathcal{N}(0, K_z)$ and $\mathbf{Z}$ has any other distribution with covariance K_z, then

$$I(\mathbf{X}_G; \mathbf{X}_G + \mathbf{Z}) \geq I(\mathbf{X}_G; \mathbf{X}_G + \mathbf{Z}_G). \tag{5}$$

If $K_x > 0$, then equality is achieved iff $\mathbf{Z} \sim \mathcal{N}(0, K_z)$.

Proof: Let $\mathbf{Y} = \mathbf{X}_G + \mathbf{Z}$ and $\mathbf{Y}_G = \mathbf{X}_G + \mathbf{Z}_G$. Then $\mathbf{Y}_G \sim \mathcal{N}(0, K_x + K_z)$ and $\mathbf{Y}, \mathbf{Y}_G$ have the same covariance matrix $K_x + K_z$. We assume the existence of probability density functions for $\mathbf{Y}$ and $\mathbf{Z}$ denoted it by $f_{\mathbf{Y}}(\cdot)$ and $f_{\mathbf{Z}}(\cdot)$ respectively. The Gaussian density function for $\mathbf{Y}_G$ and $\mathbf{Z}_G$ are denoted by $f_{\mathbf{Y}_G}(\cdot)$ and $f_{\mathbf{Z}_G}(\cdot)$ respectively.

We have

$$
\begin{aligned}
I(\mathbf{X}_G; \mathbf{X}_G + \mathbf{Z}_G) - I(\mathbf{X}_G; \mathbf{X}_G + \mathbf{Z}) &= h(\mathbf{Y}_G) - h(\mathbf{Z}_G) - h(\mathbf{Y}) + h(\mathbf{Z}) \\
&= -\int \log(f_{\mathbf{Y}_G}(\mathbf{Y})) f_{\mathbf{Y}_G}(\mathbf{y}) d\mathbf{y} + \int \log(f_{\mathbf{Z}_G}(\mathbf{z})) f_{\mathbf{Z}_G}(\mathbf{z}) d\mathbf{z} \\
&\quad + \int \log(f_{\mathbf{Y}}(\mathbf{y})) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} - \int \log(f_{\mathbf{Z}}(\mathbf{z})) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \\
&\overset{(a)}{=} \int \log(\frac{f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}_G}(\mathbf{y})}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} + \int \log(\frac{f_{\mathbf{Z}_G}(\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \\
&= D(\mathbf{Y}||\mathbf{Y}_G) - D(\mathbf{Z}||\mathbf{Z}_G) \\
&= \int \log(\frac{f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Z}_G}(\mathbf{z})}{f_{\mathbf{Y}_G}(\mathbf{y}) f_{\mathbf{Z}}(\mathbf{z})}) f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y}, \mathbf{z}) d\mathbf{y} d\mathbf{z} \\
&\overset{(b)}{\leq} \log(\int \frac{f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Z}_G}(\mathbf{z})}{f_{\mathbf{Y}_G}(\mathbf{y}) f_{\mathbf{Z}}(\mathbf{z})} f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y}, \mathbf{z}) d\mathbf{y} d\mathbf{z}) \\
&\overset{(c)}{=} \log(\int [\frac{1}{f_{\mathbf{Y}_G}(\mathbf{y})} \int [f_{\mathbf{X}_G}(\mathbf{y} - \mathbf{z})] f_{\mathbf{Z}_G}(\mathbf{z}) d\mathbf{z}] f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}) \\
&\overset{(d)}{=} \log(\int \frac{f_{\mathbf{Y}_G}(\mathbf{y})}{f_{\mathbf{Y}_G}(\mathbf{y})} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}) \\
&= 0,
\end{aligned}
$$

where (a) follows from Lemma II.1, (b) follows from Jensen's inequality, (c) follows from $f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}) = \frac{f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y},\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} = f_{\mathbf{X}_G}(\mathbf{Y} - \mathbf{Z})$, and (d) follows from $f_{\mathbf{Y}_G}(\mathbf{y}) = \int f_{\mathbf{X}_G}(\mathbf{y} - \mathbf{z}) f_{\mathbf{Z}_G}(\mathbf{z}) d\mathbf{z}$. The equality in (b) (Jensen's inequality) is achieved iff

$$
\frac{f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Z}_G}(\mathbf{z})}{f_{\mathbf{Y}_G}(\mathbf{y}) f_{\mathbf{Z}}(\mathbf{z})} = 1, \text{ for } \quad \mathbf{y}, \mathbf{z} \text{ such that } \; f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y}, \mathbf{z}) = f_{\mathbf{X}_G}(\mathbf{y} - \mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) > 0. \tag{6}
$$

If $K_x > 0$, then the support set of $\mathbf{X}_G, \mathbf{Y}$ and $\mathbf{Y}_G$ is $\mathbb{R}^n$, and thus (6) is true for all $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{z}$ in the support set of $\mathbf{Z}$. Therefore we can write for some $\mathbf{z}$ in the support set of $\mathbf{Z}$,

$$
f_{\mathbf{Z}_G}(\mathbf{z}) \int f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = f_{\mathbf{Z}}(\mathbf{z}) \int f_{\mathbf{Y}_G}(\mathbf{y}) d\mathbf{y}, \tag{7}
$$

and so $f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{Z}_G}(\mathbf{z})$ for all $\mathbf{z}$ in the support set of $\mathbf{Z}$ as $\int f_{\mathbf{Y}_G}(\mathbf{y}) d\mathbf{y} = \int f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = 1$. Therefore, to achieve equality in (b) we need $\mathbf{Z} \sim \mathcal{N}(0, K_z)$ and therefore $\mathbf{Y} \sim \mathcal{N}(0, K_x + K_z)$. ∎

Using Lemma II.2 we examine the properties of the original minimax problem.

Theorem II.1: Consider the channel, $Y_i = X_i + Z_i$ for $i = 1, \dots, n$, and impose the constraints $p_X \in \mathcal{X}$ and $p_Z \in \mathcal{Z}$. Then there exists a pair (p_X^*, p_Z^*) (probability measures on $\mathbb{R}^n$) which is a saddlepoint for the pay-off function $I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}^{(n)})$. Moreover, the pair $(p_{X_G}^*, p_{Z_G}^*)$ is also a saddlepoint, where $p_{X_G}^*, p_{Z_G}^*$ are Gaussian distributions with the same covariances as p_X^*, p_Z^*, respectively. All saddlepoints have the same pay-off value $V \overset{def}{=} \min_{p_Z} \max_{p_X} I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}^{(n)})$. If $K_z > 0$, $\forall K_z \in \mathcal{K}_z$, then all saddlepoints are of the form $(p_{X_G}^*, p_Z^*)$, where the saddlepoint distribution $p_{X_G}^*$ is Gaussian and is unique.

Proof: We first argue that the set $\mathcal{X}$ of all probability measures having covariance matrices in $\mathcal{K}_x$ is convex. If $p_X^{(1)}, p_X^{(2)}$ are two probability measures with covariances $K_x^{(1)}, K_x^{(2)} \in \mathcal{K}_x$, then the covariance of $\lambda p_X^{(1)} + (1-\lambda) p_X^{(2)}$, $\lambda \in [0,1]$, is also in $\mathcal{K}_x$, by the convexity of $\mathcal{K}_x$. Thus $\mathcal{X}$ is convex. The same argument is true for the noise probability measure.

The mutual information $I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}^{(n)})$ is concave in p_X and convex in p_Z [1], and the constraint sets on the probability measures are closed, convex and bounded. Hence using the fundamental theorem of game theory [19], we know that there exists a saddlepoint (p_X^*, p_Z^*). Let $\mathbf{X}_G^{(n)}, \mathbf{Z}_G^{(n)}$ be Gaussian random vectors in $\mathbb{R}^n$ having the same covariances as p_X^*, p_Z^* respectively. Furthermore, let $\mathbf{X}^{*(n)}, \mathbf{Z}^{*(n)}$ be random vectors in $\mathbb{R}^n$ having probability measures p_X^*, p_Z^* respectively. Then

$$\begin{aligned} I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}_G^{(n)}) &= h(\mathbf{X}^{*(n)} + \mathbf{Z}_G^{(n)}) - h(\mathbf{X}^{*(n)} + \mathbf{Z}_G^{(n)} | \mathbf{X}^{*(n)}) \\ &= h(\mathbf{X}^{*(n)} + \mathbf{Z}_G^{(n)}) - h(\mathbf{Z}_G^{(n)}) \\ &\leq h(\mathbf{X}_G^{(n)} + \mathbf{Z}_G^{(n)}) - h(\mathbf{Z}_G^{(n)}) = I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}_G^{(n)}), \end{aligned} \tag{8}$$

where the inequality follows from the fact that the Gaussian distribution maximizes the entropy for a given covariance (see [1], Chapter 9). Similarly from Lemma II.2 we have

$$I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}_G^{(n)}) \leq I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}^{*(n)}), \tag{9}$$

for any distribution on $\mathbf{Z}^{*(n)}$, where $\mathbf{Z}_G^{(n)} \sim \mathcal{N}(0, K_z)$, and K_z is the covariance matrix of $\mathbf{Z}^{*(n)}$. Hence we have shown that

$$I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}_G^{(n)}) \leq I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}_G^{(n)}) \leq I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}^{*(n)}). \tag{10}$$

But, as we know that (p_X^*, p_Z^*) is a saddlepoint, we have the following double inequality,

$$I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}^{*(n)}) \leq I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}^{*(n)}) \leq I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}_G^{(n)}), \tag{11}$$

and hence we have

$$V \overset{def}{=} \min_{p_Z} \max_{p_X} I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}^{(n)}) = I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}_G^{(n)}) = I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}^{*(n)}). \tag{12}$$

This also shows an interchangability property, *i.e.*, if $(p_X^{(1)}, p_Z^{(1)})$ and $(p_X^{(2)}, p_Z^{(2)})$ are saddlepoints then $(p_X^{(1)}, p_Z^{(2)})$ and $(p_X^{(2)}, p_Z^{(1)})$ are also saddlepoints.

Let $\mathbf{Z}_G^{*(n)} \sim p_{Z_G}^*$ be one of the Gaussian noise saddlepoints. If $p_{\bar{X}} = \lambda p_X^* + \bar{\lambda} p_{X_G}^*$ then, by the concavity of the mutual information, we observe

$$V \geq I(\bar{\mathbf{X}}^{(n)}; \bar{\mathbf{X}}^{(n)} + \mathbf{Z}_G^{*(n)}) \geq \lambda I(\mathbf{X}^{*(n)}; \mathbf{X}^{*(n)} + \mathbf{Z}_G^{*(n)}) + \bar{\lambda} I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}_G^{*(n)}) = V, \tag{13}$$

where $\bar{\mathbf{X}}^{(n)} \sim p_{\bar{X}}, \mathbf{X}^{*(n)} \sim p_X^*, \mathbf{X}_G^{*(n)} \sim p_{X_G}^*$. Hence $h(\bar{\mathbf{Y}}^{(n)}) = h(\mathbf{Y}^{*(n)}) = h(\mathbf{Y}_G^{*(n)})$ where, $\bar{\mathbf{Y}}^{(n)} = \bar{\mathbf{X}}^{(n)} + \mathbf{Z}_G^{*(n)}, \mathbf{Y}^{*(n)} = \mathbf{X}^{*(n)} + \mathbf{Z}_G^{*(n)}, \mathbf{Y}_G^{*(n)} = \mathbf{X}_G^{*(n)} + \mathbf{Z}_G^{*(n)}$. If $K_z > 0$ then, $h(\mathbf{Y}_G^{*(n)}) < \infty$ and the entropy $h(\mathbf{Y}^{(n)})$ is *strictly* concave in p_Y and so we have $\mathbf{Y}^{*(n)} \sim \mathcal{N}(0, K_x^* + K_z^*)$. Therefore we have

$$\Psi_Y(\theta) = \Psi_{X^*}(\theta)\Psi_{Z_G^*}(\theta) = \Psi_{X_G^*}(\theta)\Psi_{Z_G^*}(\theta) = \Psi_{Y_G^*}(\theta), \tag{14}$$

where $\Psi_Y(\theta)$ is the characteristic function of $\mathbf{Y}^{(n)}$, and $\Psi_{Z_G^*}(\theta) = exp(\frac{1}{2}\theta^T K_z^* \theta)$. Hence, as $\Psi_{Z_G^*}(\theta)$ is non-zero for all θ we conclude that $p_X^* = p_{X_G}^*$, and that the $p_{X_G}^*$ is unique. ■

It is well known from convex analysis [20] that the set of minimizing arguments for a convex function is a convex set. In the next result we use this to show the set of worst noise distributions is a convex set.

Corollary II.1: Let $\mathbf{X}_G^* \sim p_{X_G}^*$ have the Gaussian input saddlepoint distribution, then the set of worst noise distributions

$$\mathcal{Z}^* = \{p_z^* \in \mathcal{Z} : p_z^* = \mathrm{argmin}_{p_z} I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}^{(n)})\}$$

is a convex set.

Proof: From Theorem II.1 we already know that the saddlepoints are of the form $(p_{X_G}^*, p_Z^*)$, where $p_{X_G}^*$ is unique. Let $(p_{X_G}^*, p_z^{(1)})$ and $(p_{X_G}^*, p_z^{(2)})$ be two saddlepoints, and $\gamma \in [0,1]$. Then,

$$I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \bar{\mathbf{Z}}^{(n)}) \leq \gamma I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}_1^{(n)}) + (1-\gamma) I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}_2^{(n)}) = V, \tag{15}$$

where $\mathbf{X}_G^{*(n)} \sim p_{X_G}^*, \mathbf{Z}_1^{(n)} \sim p_z^{(1)}, \mathbf{Z}_2^{(n)} \sim p_z^{(2)}, \bar{\mathbf{Z}}^{(n)} \sim \gamma p_z^{(1)} + (1-\gamma) p_z^{(2)}$ and V is the value of the game as defined in Theorem II.1. The above equation is due to the convexity of $I(\mathbf{X}^{(n)}, \mathbf{X}^{(n)} + \mathbf{Z}^{(n)})$ with p_Z [1]. Thus the inequality in (15) is satisfied with equality. Hence $(p_{X_G}^*, \gamma p_z^{(1)} + (1-\gamma) p_z^{(2)})$ is also a saddlepoint and therefore $\mathcal{Z}^*$ is a convex set. Moreover, this also implies that the set of worst covariance matrices

$$\mathcal{K}_z{}^* = \{K_z : K_z = \mathrm{argmin}_{K_z \in \mathcal{K}_z} \frac{1}{2} \log \frac{|K_x^* + K_z|}{|K_z|}\},$$

is a convex set. ■

We have shown that the saddlepoints are of the form $(p^*_{X_G}, p^*_Z)$, and that $(p^*_{X_G}, p^*_{Z_G})$ are also saddlepoints, where $p^*_{Z_G}$ is Gaussian with the same covariance as p^*_Z. We can make the following observation on the noise saddlepoint distributions p^*_Z.

Let the rank$(K_x) = \nu \leq n$, and the eigendecomposition of K_x be $K_x = U^T \Lambda_x U$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_\nu, 0, \ldots, 0)$. Hence we can write,

$$\begin{aligned} \tilde{\mathbf{Y}} &= U\mathbf{Y}^{(n)}, \;\; \tilde{\mathbf{X}} = U\mathbf{X}^{(n)}, \;\; \tilde{\mathbf{Z}} = U\mathbf{Z}^{(n)} \\ \tilde{\mathbf{X}} &= [\tilde{\mathbf{X}}_1^T, \tilde{\mathbf{X}}_2^T]^T, \;\; \tilde{\mathbf{Y}} = [\tilde{\mathbf{Y}}_1^T, \tilde{\mathbf{Y}}_2^T]^T, \;\; \tilde{\mathbf{Z}} = [\tilde{\mathbf{Z}}_1^T, \tilde{\mathbf{Z}}_2^T]^T, \end{aligned} \tag{16}$$

where $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2$ are of dimension $\nu, n-\nu$ respectively. The vectors $\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2$ are defined similarly. The following proposition has been contributed by Amos Lapidoth.

Proposition II.1: The noise saddlepoint distribution p^*_z is such that $\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2$ has a full rank Gaussian distribution where $C = \mathbb{E}[\tilde{\mathbf{Z}}_1 \tilde{\mathbf{Z}}_2^T] \left\{ \mathbb{E}[\tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_2^T] \right\}^{-1}$.

Note: This means that the estimation error of the best linear estimate of $\tilde{\mathbf{Z}}_1$ from $\tilde{\mathbf{Z}}_2$ is full rank Gaussian.

Proof: Let $\mathbf{Y}^{(n)} = \mathbf{X}_G^{(n)} + \mathbf{Z}^{(n)}$, where $\mathbf{X}_G^{(n)}$ has the Gaussian input saddlepoint distribution (see Theorem II.1). We define $\tilde{\mathbf{X}}^* = U\mathbf{X}_G^{(n)}$ and the notation from (16) is used. If K_x is not full rank *i.e.,* $\nu < n$, then $\tilde{\mathbf{X}}_2 = 0$ a.s, and $\tilde{\mathbf{Y}}_2 = \tilde{\mathbf{Z}}_2$ a.s. Let $C = [\tilde{\mathbf{Z}}_1 \tilde{\mathbf{Z}}_2^T] \left\{ \mathbb{E}[\tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_2^T] \right\}^{-1}$, then we have the following,

$$\begin{aligned} I(\mathbf{X}_G^{(n)}; \mathbf{Y}^{(n)}) &= I(\tilde{\mathbf{X}}^*; \tilde{\mathbf{Y}}) \\ &= I(\tilde{\mathbf{X}}_1^*; \tilde{\mathbf{Y}}) \\ &\geq I(\tilde{\mathbf{X}}_1^*; \tilde{\mathbf{Y}}_1 - C\tilde{\mathbf{Y}}_2) \\ &= I(\tilde{\mathbf{X}}_1^*; \tilde{\mathbf{Y}}_1 - C\tilde{\mathbf{Z}}_2) \\ &= I(\tilde{\mathbf{X}}_1^*; \tilde{\mathbf{X}}_1^* + \tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2) \\ &\overset{(a)}{\geq} \frac{1}{2} \log \frac{|K_{\tilde{\mathbf{X}}_1^*} + K_{\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2}|}{|K_{\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2}|} \\ &\overset{(b)}{=} \frac{1}{2} \log \frac{|K_x^* + K_z|}{|K_z|} \end{aligned} \tag{17}$$

where (a) is due to Lemma II.2. Moreover, as $K_{\tilde{\mathbf{X}}_1} \overset{def}{=} \mathbb{E}[\tilde{\mathbf{X}}_1^* \tilde{\mathbf{X}}_1^{*T}] > 0$, using Lemma II.2 we know that equality is achieved iff $\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2$ is Gaussian. Now, to show (b) we use the determinant relationship of block matrices using the Schur complement (defined as $A - BD^{-1}B^T$) [21]:

$$\begin{vmatrix} A & B \\ B^T & D \end{vmatrix} = |D||A - BD^{-1}B^T|. \tag{18}$$

Using (18) and

$$K_{\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2} = \mathbb{E}[\tilde{\mathbf{Z}}_1 \tilde{\mathbf{Z}}_1^T] - \mathbb{E}[\tilde{\mathbf{Z}}_1 \tilde{\mathbf{Z}}_2^T] \left\{ \mathbb{E}[\tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_2^T] \right\}^{-1} \mathbb{E}[\tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_1^T] \tag{19}$$

we obtain,

$$K_z = |\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_2^T]||\mathbb{E}[\tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_1^T] - \mathbb{E}[\tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_2^T]\left\{\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_2^T]\right\}^{-1}\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_1^T]| = |\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_2^T]||K_{\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2}| \tag{20}$$

$$K_x^* + K_z = |\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_2^T]||K_{\tilde{\mathbf{X}}_1^*} + \mathbb{E}[\tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_1^T] - \mathbb{E}[\tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_2^T]\left\{\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_2^T]\right\}^{-1}\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_1^T]| = |\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_2^T]||K_{\tilde{\mathbf{X}}_1^*} + K_{\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2}|$$

which completes the proof for (b) in (17). Therefore, equality is achieved in (17) iff $\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2$ has a full rank Gaussian distribution. ■

Now, this does not completely answer the question of whether *all* saddlepoints to this problem are Gaussian. The problem arises primarily because the mutual information is *not* necessarily a *strictly* convex function of p_Z and therefore the noise saddlpoint distribution p_Z^* need not be unique. However, using Theorem II.1, which shows the existence of Gaussian saddlepoints, and Proposition II.1 we believe that it is worthwhile to focus our attention on the the Gaussian mutual information game defined as follows.

The Gaussian mutual information game is defined with payoff

$$g(K_x, K_z) \stackrel{def}{=} I(\mathbf{X}_G^{(n)}; \mathbf{X}_G^{(n)} + \mathbf{Z}_G^{(n)}) = \frac{1}{2}\log\frac{|K_x + K_z|}{|K_z|}, \tag{21}$$

where we have constrained $\mathbf{X}^{(n)}$ and $\mathbf{Z}^{(n)}$ to be Gaussian with covariances $K_x \in \mathcal{K}_x$ and $K_z \in \mathcal{K}_z$. Note that all saddlepoints have the same value and hence the Gaussian saddlepoints yield the minimax rate. Later we will examine a sufficient condition under which the saddlepoint is indeed unique.

Note that as all saddlepoints covariances are characterized by (K_x^*, K_z), $K_z \in {\mathcal{K}_z}^*$. For example, if the input covariance constraint is an average power constraint, K_x^* must waterfill *all* the covariances in ${\mathcal{K}_z}^*$. From Corollary II.1, if the noise player chooses to use a mixture of covariances in ${\mathcal{K}_z}^*$ it does not gain, since the signal covariance K_x^* is already waterfilling any convex combination of $\{K_z\} \in {\mathcal{K}_z}^*$. Moreover, the noise cannot further reduce the mutual information by using any other distribution in $\mathcal{Z}^*$. In [22], [23] a problem with vector (parallel channels) inputs and outputs with power constraints on the signal *and* noise was considered. In our problem the transmitter does *not* know the noise covariance matrix and cannot use this information to form parallel channels. Moreover, the constraints on the processes are more general than power constraints (or trace constraints on the covariance matrix).

Next we examine the properties of the function $g(K_x, K_z)$. In particular we show that $\frac{1}{2}\log\frac{|K_x+K_z|}{|K_z|}$ is convex in K_z and concave in K_x.

Lemma II.3: The function $\log(\frac{|K_x+K_z|}{|K_z|})$ is convex in K_z, with strict convexity if $K_x > 0$.

Proof: Consider $\mathbf{Y} = \mathbf{X} + \mathbf{Z}_\theta$ and let $\mathbf{X} \sim \mathcal{N}(0, K_X)$, and let θ be independent of $\mathbf{X}$ and be distributed as

$$\theta = \begin{cases} 1, & \text{w.p.}\lambda \\ 2, & \text{w.p.}\bar{\lambda} \end{cases}, \tag{22}$$

where $\bar{\lambda} = 1-\lambda$. Let $\mathbf{Z}_1 \sim \mathcal{N}(0, K_{Z_1})$, $\mathbf{Z}_2 \sim \mathcal{N}(0, K_{Z_2})$ (mutually independent and independent of $\mathbf{X}$) and let us define

$$\mathbf{Z}_\theta = \begin{cases} \mathbf{Z}_1, & \text{if } \theta = 1 \\ \mathbf{Z}_2, & \text{if } \theta = 2 \end{cases} . \tag{23}$$

Consider the two expansions

$$\begin{aligned} I(\mathbf{X};\mathbf{Y},\theta) &= I(\mathbf{X};\theta) + I(\mathbf{X};\mathbf{Y}|\theta) \\ &= I(\mathbf{X};\mathbf{Y}) + I(\mathbf{X};\theta|\mathbf{Y}). \end{aligned} \tag{24}$$

Now, since $I(\mathbf{X};\theta) = 0$ and $I(\mathbf{X};\theta|\mathbf{Y}) \geq 0$, we have

$$I(\mathbf{X};\mathbf{Y}|\theta) \geq I(\mathbf{X};\mathbf{Y}). \tag{25}$$

However,

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}|\theta) = &\quad \lambda I(\mathbf{X};\mathbf{Y}|\theta = 0) + \bar{\lambda} I(\mathbf{X};\mathbf{Y}|\theta = 1) \\ = &\quad \lambda \tfrac{1}{2}\log(\tfrac{|K_x+K_{z_1}|}{|K_{z_1}|}) + \bar{\lambda}\tfrac{1}{2}\log(\tfrac{|K_x+K_{z_2}|}{|K_{z_2}|}). \end{aligned} \tag{26}$$

From Lemma II.2 we have,

$$I(\mathbf{X};\mathbf{X}+\mathbf{Z}) \geq I(\mathbf{X};\mathbf{X}+\mathbf{Z}_G) = \frac{1}{2}\log(\frac{|K_x+K_z|}{|K_z|}), \tag{27}$$

where $\mathbf{Z}_G \sim \mathcal{N}(0, K_z)$ and $K_z = \lambda K_{z_1} + \bar{\lambda} K_{z_2}$. Using (25 – 27) we have

$$\lambda \log(\frac{|K_x+K_{z_1}|}{|K_{z_1}|}) + \bar{\lambda}\log(\frac{|K_x+K_{z_2}|}{|K_{z_2}|}) \geq \log(\frac{|K_x+K_z|}{|K_z|}) \tag{28}$$

which gives the desired result. Note that if $K_x > 0$, the inequality in (27) is strict, by Lemma II.2, implying strict convexity. ■

The following lemma [24] has an information theoretic proof in [25].

Lemma II.4: If $K_z > 0$, the function $\log(\frac{|K_x+K_z|}{|K_z|})$ is strictly concave in K_x.

■

We now prove sufficient conditions under which the saddlepoint to the mutual information game is unique.

Lemma II.5: If there exists a saddlepoint (K_x^*, K_z^*) of $g(K_x, K_z)$, such that $K_x^* > 0$, then the saddlepoint (p_x^*, p_z^*) for the mutual information game is unique and Gaussian with covariances K_x^*, K_z^* respectively.

Proof: From Lemma II.2 $I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}^{(n)}) \geq I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}_G^{*(n)})$ and as $K_x^* > 0$, equality is achieved iff $\mathbf{Z}^{(n)} \sim \mathcal{N}(0, K_z^*)$. Now, let

$$\mathcal{K}_z{}^* = \{K_z : K_z = \text{argmin}_{K_z \in \mathcal{K}_z} \frac{1}{2}\log\frac{|K_x^* + K_z|}{|K_z|}\},$$

Now, since $g(K_x^*, K_z)$ is strictly convex for $K_x^* > 0$ (from Lemma II.3), we see that the above minimization has a unique minimum. Thus $p_z^* = \mathrm{argmin}_{p_z} I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}^{(n)})\}$ is unique and Gaussian. ■

This result also helps us make observations on the set of noise saddlepoint distributions for the case when K_x^* is not strictly positive definite. Here we use the notation of Proposition II.1 and (17). If $\mathrm{rank}(K_x) = \nu < n$, using the partition defined in (16), then we observe that $K_{\tilde{\mathbf{X}}_1^*} > 0$. Using Lemma II.5 on $I(\tilde{\mathbf{X}}_1^*; \tilde{\mathbf{X}}_1^* + \tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2)$ we see that for the noise saddlepoint distribution, $(\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2)$ has to be Gaussian with a *unique* covariance. Therefore, we can observe that the saddlepoint distributions are such that the Schur complement of the noise covariance matrix, projected onto the signal covariance eigendirections, is a constant. More precisely, the set of noise saddlepoint distributions is convex and such that the $\tilde{\mathbf{Z}}_1 - C\tilde{\mathbf{Z}}_2$ has a full rank Gaussian distribution with a covariance $\mathbb{E}[\tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_1^T] - \mathbb{E}[\tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_2^T]\left\{\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_2^T]\right\}^{-1}\mathbb{E}[\tilde{\mathbf{Z}}_2\tilde{\mathbf{Z}}_1^T]$ which is constant over the set.

We know [3] that for average signal and noise power, that the pair $(K_x = P\mathbf{I}, K_z = N_0\mathbf{I})$ is a saddlepoint. The result in Lemma II.5 shows that the saddlepoint is unique [10]. In the next section we find the worst additive noise for a banded covariance constraint.

III. Banded covariance constraint

In this section we constrain the noise distribution to have a a banded covariance matrix. Here we assume that we know the noise covariance lags up to the p^{th} lag as given by

$$\mathbb{E}[Z_i Z_{i+k}] = \alpha_k, \quad k = 0, \dots p, \quad \text{for all } i. \tag{29}$$

The noise is assumed to have zero mean. Now as the transmitter knows only partial information about the noise spectrum the question is what should be the input spectrum solving the mutual information game defined in (2). In this section we consider noise distributions $\mathcal{Z} = \{p(z) : \mathbb{E}[Z] = 0, K_z \in \mathcal{K}_z\}$ where $\mathcal{K}_z = \{K_z : (K_z)_{i,j} = \alpha(i-j), (i,j) \in \mathcal{S}\}$, and $\mathcal{S} = \{(i,j) : |i-j| \leq p, i,j = 1,2,\dots\}$ specifies the constraints on the correlation lags. Let the covariance matrix K_z^{**} be the maximum entropy covariance in $\mathcal{K}_z$ (specified by Burg's theorem [2]). The maximum entropy noise is a Gauss-Markov process with covariance lags satisfying the Yule-Walker equations ([1], pp 274–277). Clearly we can use a signal design which waterfills the eigenvalues of the maximum entropy extension K_z^{**}. Let us define this input covariance matrix to be K_x^{**}.

We now show that the maximum entropy extension K_z^{**} is the worst noise when we have

$$\begin{aligned}\min_{K_z \in \mathcal{K}_z} \max_{K_x \in \mathcal{K}_x} \frac{1}{2}\log\left(\frac{|K_x + K_z|}{|K_z|}\right) \qquad (30)\\ = \min_{K_z \in \mathcal{K}_z} \frac{1}{2}\log\left(\frac{|\nu I|}{|K_z|}\right),\end{aligned}$$

for appropriate ν, which is true if the input power is high enough so that for all $K_z \in \mathcal{K}_z$, $K_x^o + K_z = \nu I$, where K_x^o waterfills K_z. Now $\nu = P + \sum_i \lambda_i / n$, where $\{\lambda_i\}$ are the eigenvalues of K_z. Thus the minimax problem becomes

$$\min_{K_z \in \mathcal{K}_z} [\frac{1}{2}\log|(P + \sum_i \lambda_i/n)I| - \frac{1}{2}\log|K_z|]. \tag{31}$$

But $\sum_i \lambda_i/n = \alpha_0$ is specified in (29), so the maximum in (31) is achieved by maximizing $\max_{K_z \in \mathcal{K}_z} \frac{1}{2}\log|K_z|$. However for this condition, we need the power P to be large. We examine the implication of this high power requirement. Notice that we need $\nu > \max_i \lambda_i$ for (30) to be true. Therefore we need $P > \max_i \lambda_i - \alpha_0$ for the naive high power requirement. This might require a power growing linearly with block size. In Theorem III.1 we show that this requirement is too pessimistic and that the worst additive noise is the maximum entropy noise for a *bounded* input power requirement. To show this, we recall two useful facts.

Fact III.1: $\frac{d \log|\mathbf{X}|}{d\mathbf{X}} = \mathbf{X}^{-1}$, for $\mathbf{X} = \mathbf{X}^T > 0$.

Fact III.2: For the maximum entropy completion of the noise specified in (29), the covariance matrix K_z^{**} satisfies $(K_z^{**-1})_{i,j} = 0$, for $(i,j) \notin \mathcal{S}$ as shown, for example in [1].

Now, using these facts we will show that the maximal entropy extension (K_z^{**}) of the noise and the corresponding signal waterfilling covariance matrix (K_x^{**}) do indeed form a saddlepoint for the game defined in (2) for sufficiently high input power.

Theorem III.1: Let $Y_i = X_i + Z_i$ for $i = 1, \dots, n$, and let $\{Z_i\}$ be a noise process satisfying the constraints given in (29). Let $\{X_i\}$ satisfy the expected power constraint P. If $K_x^{**} > 0$, we have

$$I(\mathbf{X}^{(n)}; \mathbf{X}^{(n)} + \mathbf{Z}_G^{*(n)}) \le I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}_G^{*(n)}) \le I(\mathbf{X}_G^{*(n)}; \mathbf{X}_G^{*(n)} + \mathbf{Z}^{(n)}), \tag{32}$$

for all $p_x \in \mathcal{X}, p_z \in \mathcal{Z}$ where $\mathbf{X}^{*(n)} \sim \mathcal{N}(0, K_x^{**})$, $\mathbf{Z}^{*(n)} \sim \mathcal{N}(0, K_z^{**})$, K_z^{**} is the maximum entropy extension of the noise and K_x^{**} is the corresponding waterfilling signal covariance matrix.

Proof: The first inequality is easy to show from the waterfilling argument. For the second inequality, we again use Lemma II.2 to reduce consideration to only Gaussian noise processes. Therefore, the problem reduces to

$$\min_{K_z} \frac{1}{2}\log(\frac{|K_x^{**} + K_z|}{|K_z|}) \tag{33}$$
$$\text{such that} \quad \mathbb{E}[Z_i Z_{i+k}] = \alpha_k, \;\; k = 0, \dots p, \text{for all} \;\; i.$$

This is again a convex minimization problem over a convex set and as $K_x^{**} > 0$, $\frac{1}{2}\log(\frac{|K_x^{**}+K_z|}{|K_z|})$ is a strictly convex functional (Lemma II.3) and hence it has a unique solution [26]. It remains to show that K_z^{**} satisfies the necessary and sufficient conditions for optimality [26]. Setting up the Lagrangian we have

$$\mathcal{L} = \frac{1}{2}\log(|K_x^{**} + K_z|) - \frac{1}{2}\log(|K_z|) + \sum_{(i,j)\in\mathcal{S}} \lambda_{i,j}(K_z)_{i,j}, \tag{34}$$

where $\mathcal{S} = \{(i,j) : j = i \pm k, k = 0, \ldots, p\}$ specifies the constraints on the correlation lags. Now differentiating with respect to K_z and using Fact III.1, we obtain

$$\frac{d\mathcal{L}}{dK_z} = (K_x^{**} + K_z)^{-1} - (K_z)^{-1} + \mathbf{A}, \tag{35}$$

where $\mathbf{A}$ is a banded matrix such that $(\mathbf{A})_{i,j} = 0$ for $(i,j) \notin \mathcal{S}$. Note that from Fact III.2 we have $(K_z^{**-1})_{i,j} = 0$ for $(i,j) \notin \mathcal{S}$. Hence it is clear that K_z^{**} satisfies the necessary and sufficient conditions for optimality, since $K_x^{**} + K_z^{**} = \nu\mathbf{I}$ for some constant ν. This is true as K_x^{**} is the waterfilling solution to K_z^{**}. Clearly from this it follows that K_z^{**} is the minimizing solution. Note that from Lemma II.5, as $K_x^{**} > 0$, this constitutes a unique saddlepoint to the problem. ■

To see what the power requirement is for $K_x^{**} > 0$ and Theorem III.1 to hold, we see that the power should be large enough so that we can "completely" waterfill the maximum entropy extension. The power needed for this is bounded, as we now argue. For the maximum entropy completion, the noise covariance matrix is Toepltiz ([1]) and therefore asymptotically the density of the eigenvalues on the real line tends to the power spectrum of the maximum entropy stochastic process [1]. Hence the condition for the power spectral density of the input process for "completely" waterfilling the maximum entropy process is that: $\nu - N_{ME}(f) > 0, \forall f \in [-1/2, 1/2)$, where $\nu = P - \int_{-1/2}^{1/2} N_{ME}(f)df$, the maximum entropy noise spectral density is denoted by $N_{ME}(f) = \frac{\sigma^2}{|1+\sum_{k=1}^{p} a_k exp(-j2\pi fk)|^2}$, where $a_1, \ldots, a_p, \sigma^2$, satisfy the Yule-Walker equations ([1], pp 274–277). If the maximum entropy process is stable (*i.e.* the noise spectral density does not have poles on the unit circle) then the input power needed for the above condition is finite, as $\sup_{f\in[-1/2,1/2)} N_{ME}(f) < \infty$. If the banded constraint is not degenerate then the Yule-Walker equations are not degenerate, *i.e.* we do not have a completely predictable process. Hence the maximum entropy completion (for the given banded constraint) cannot be unstable (or critically stable), completing the argument. Now, as we have chosen $K_x^{**} > 0$, we have a strictly convex minimization problem for K_z and we establish the result.

Example: This example shows how the maximum entropy noise and worst additive noise might differ. Let $\mathbb{E}[Z_i^2] = 1$ and $\mathbb{E}[Z_i Z_{i+1}] = 0.9$. Thus

$$\mathcal{K}_z = \left\{ K_z : K_z = \begin{bmatrix} 1 & 0.9 & ? \\ 0.9 & 1 & 0.9 \\ ? & 0.9 & 1 \end{bmatrix} \right\}, \tag{36}$$

and maximum entropy completion is

$$K_z^{**} = \begin{bmatrix} 1 & 0.9 & 0.81 \\ 0.9 & 1 & 0.9 \\ 0.81 & 0.9 & 1 \end{bmatrix} = \lambda_1 \psi_1 \psi_1^T + \lambda_2 \psi_2 \psi_2^T + \lambda_3 \psi_3 \psi_3^T, \tag{37}$$

where $\lambda_1 = 2.7406, \lambda_2 = 0.19, \lambda_3 = 0.0693$ are the eigenvalues of K_z^{**} and ψ_1, ψ_2, ψ_3 are the associated eigenvectors. If the power is large enough to waterfill K_z^{**} (*i.e.*, $tr(K_x) > 5.22$) then the conditions needed for Theorem III.1 are satisfied and the maximum entropy completion K_z^{**} is indeed the worst noise.

We now consider the power constraint, $tr(K_x) \leq 0.1$. Here the input power is insufficient to waterfill the maximum entropy completion. We find the saddlepoint (K_x^*, K_z^*) by numerically solving for $\max_{K_x \in \mathcal{K}_x} \min_{K_z \in \mathcal{K}_z} \frac{1}{2} \log \frac{|K_x + K_z|}{|K_z|}$. The worst noise additive K_z^* is then given by

$$K_z^* = \begin{bmatrix} 1 & 0.9 & 0.873 \\ 0.9 & 1 & 0.9 \\ 0.873 & 0.9 & 1 \end{bmatrix} = \lambda_1^* \eta_1 \eta_1^T + \lambda_2^* \eta_2 \eta_2^T + \lambda_3^* \eta_3 \eta_3^T, \tag{38}$$

where $\lambda_1^* = 0.091, \lambda_2^* = 0.127, \lambda_3^* = 2.782$ are the eigenvalues of K_z^* and η_1, η_2, η_3 are the associated eigenvectors. The optimal transmitter covariance matrix K_x^* is of rank 2, given by

$$K_x^* = \begin{bmatrix} 0.0275 & -0.0228 & -0.0045 \\ -0.0228 & 0.0450 & -0.0228 \\ -0.0045 & -0.0228 & 0.0275 \end{bmatrix}, \tag{39}$$

and

$$C = \max_{tr(K_x) \leq 0.1} \min_{K_z \in \mathcal{K}_z} \frac{1}{2} \log \frac{|K_x + K_z|}{|K_z|} = \frac{1}{2} \log \frac{|K_x^* + K_z^*|}{|K_z^*|} = 0.3915 \text{ nats.} \tag{40}$$

Thus for the this low signal power example, the worst additive noise is $\mathcal{N}(0, K_z^*)$, which differs from the $\mathcal{N}(0, K_z^{**})$ maximum entropy noise

Note that if the transmitter uses the minimax distribution $\mathcal{N}(0, K_x^*)$, but nature deviates from the noise distribution $\mathcal{N}(0, K_z^*)$ by using the maximum entropy noise $\mathcal{N}(0, K_z^{**})$, the transmission rate increases to $\frac{1}{2} \log \frac{|K_x^* + K_z^{**}|}{|K_z^{**}|} = 0.4196$ nats. Thus deviation by the noise player is strictly punished, and the maximum entropy noise is seen to be strictly suboptimal for low power.

Note that when we have low signal power, the optimal K_x^* does not have full rank. In general (for a larger number of dimensions n) there could be a convex set of noise covariance matrices whose projections on the range space of K_x^* are identical but could be different in the null space of K_x^* (still satisfying the covariance constraints). Thus, the set of worst noise covariance matrices is convex and looks the same in the range space of K_x^* (or "below the waterline").

IV. Decoding scheme

It is difficult for the receiver to form a maximum likelihood detection scheme for all noise distributions. Therefore we propose using a simpler detection scheme based on a Gaussian metric and the second-order moments. However, as this is not the optimal metric, it falls into the category of mismatched decoding [11], and it is not obvious that the rate $\frac{1}{2}\log\frac{|K_x+K_z|}{|K_z|}$ is achievable using such a mismatched decoding scheme.

In this subsection we show that the rate $\frac{1}{2}\log\frac{|K_x+K_z|}{|K_z|}$ is achievable using a random Gaussian codebook and a Gaussian metric under some conditions on the noise process. In [11], [23], it was shown that $\frac{1}{2}\log(1+P/N_0)$ is achievable using a Gaussian codebook and a minimum Euclidean distance decoding metric. This result was extended to the vector single user channel where the transmitter knows the noise covariance matrix and hence can form parallel channels [11], [23]. In our case we do not assume that the transmitter knows the noise covariance but show that if the receiver knows K_z, then the rate $\frac{1}{2}\log\frac{|K_x+K_z|}{|K_z|}$ is achievable.

The coding game is played as follows. The transmitter knows the family $\mathcal{K}_z$ but not the specific covariance $K_z \in \mathcal{K}_z$ or the distribution. The transmitter chooses a distribution $p(x^{(n)})$ and 2^{nR_n} i.i.d codewords drawn according to $p(x^{(n)})$. The transmitter is also allowed to choose a random codebook, where the codebook is known to the receiver. The receiver is assumed to know K_z but not the noise distribution. The receiver chooses a given decoding rule based on the knowledge of the noise covariance and the transmitter codebook. The noise can choose any distribution $f(z^{(n)})$ satisfying the given covariance constraints $K_z \in \mathcal{K}_z$ and some regularity conditions (*C1* and *C2* below) on the noise process. We find the highest achievable rate for which the probability of error averaged over the random codebooks goes to zero.

Let us define $M(\mathbf{X}^{(n)};\mathbf{Y}^{(n)})$ as

$$\begin{aligned} M(\mathbf{X}^{(n)};\mathbf{Y}^{(n)}) = \frac{1}{2}\log\frac{|K_x+K_z|}{|K_z|} &+ \frac{1}{2}\mathbf{Y}^{(n)T}(K_x+K_z)^{-1}\mathbf{Y}^{(n)} \\ &-\frac{1}{2}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)}). \end{aligned} \tag{41}$$

Define $\mathbf{X}^{(n)}$ and $\mathbf{Y}^{(n)}$ to be jointly ϵ-typical if we have

$$\frac{1}{2n}\log\frac{|K_x+K_z|}{|K_z|} - \frac{1}{n}M(\mathbf{X}^{(n)};\mathbf{Y}^{(n)}) < \epsilon. \tag{42}$$

Our detection rule is that we declare $\mathbf{X}^{(n)}(i)$ to be decoded if it is the *only* codeword which is jointly ϵ-typical with the received $\mathbf{Y}^{(n)}$. Note that the detection rule is equivalent to a Gaussian decoding metric with a threshold detection scheme where an error is declared if there are more than one codewords below the threshold. This can be seen by rewriting (42) as

$$\frac{1}{2n}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)}) < \frac{1}{2n}\mathbf{Y}^{(n)T}(K_x+K_z)^{-1}\mathbf{Y}^{(n)} + \epsilon. \tag{43}$$

The conditions that we impose on the noise process are:
C1: $\lim_{n\to\infty} \Pr\left(|\frac{1}{n}\mathbf{z}^{(n)T}K_z^{-1}\mathbf{z}^{(n)} - \mathbb{E}[\frac{1}{n}\mathbf{z}^{(n)T}K_z^{-1}\mathbf{z}^{(n)}]| > \epsilon \right) = 0, \forall\epsilon > 0.$
C2: $\lim_{n\to\infty} \Pr\left(|\frac{1}{n}\mathbf{z}^{(n)T}(K_x(1+\gamma)+K_z)^{-1}\mathbf{z}^{(n)} - \mathbb{E}[\frac{1}{n}\mathbf{z}^{(n)T}(K_x(1+\gamma)+K_z)^{-1}\mathbf{z}^{(n)}]| > \epsilon \right) = 0, \forall\epsilon > 0, \gamma > 0.$

We begin by stating two lemmas which are proved in the appendix. Lemma IV.2 requires the use of conditions $C1$ and $C2$ on the noise process.

Lemma IV.1: If $\mathbf{X}^{(n)} \sim \mathcal{N}(0, K_x)$ and is independent of $\mathbf{Y}^{(n)}$, then

$$\begin{aligned} \mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{Y}^{(n)^T}(K_x+K_z)^{-1}\mathbf{Y}^{(n)} - \right.\right. & \left.\left. \tfrac{1}{2}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})\right)\right] \\ & = exp(-\tfrac{1}{2}log(|K_x+K_z|/|K_z|)) \end{aligned} \tag{44}$$

Lemma IV.2: If $\mathbf{X}^{(n)} \sim \mathcal{N}(0, K_x)$ and is independent of $\mathbf{Z}^{(n)}$, and $\mathbb{E}[\mathbf{Z}^{(n)}\mathbf{Z}^{(n)^T}] = K_z > 0$, and the noise satisfies $C1$ and $C2$, then we have

$$\begin{aligned} Pr[\frac{1}{2n}\mathbf{Z}^{(n)^T}K_z^{-1}\mathbf{Z}^{(n)} & > \tfrac{1}{2n}(\mathbf{Z}^{(n)}+\mathbf{X}^{(n)})^T(K_x+K_z)^{-1}(\mathbf{Z}^{(n)}+\mathbf{X}^{(n)})+\epsilon] \\ & \leq (1-\epsilon)exp(-n\tfrac{\epsilon^2}{8})+\epsilon. \end{aligned} \tag{45}$$

We define $P_e^{(n)}$ as the probability of error over a block of n samples averaged over transmitter codebooks, *i.e.*,

$$\begin{aligned} \lambda_i^{(n)}(\mathcal{C}) &= \Pr(\hat{i}(y^{(n)}) \neq i | x^{(n)}(i)) \\ P_e^{(n)}(\mathcal{C}) &= \frac{1}{2^{nR_n}}\sum_i \lambda_i^{(n)}(\mathcal{C}) \\ P_e^{(n)} &= \mathbb{E}_{\mathcal{C}} P_e^{(n)}(\mathcal{C}). \end{aligned}$$

We will show below that for rates R_n below $C_n = \frac{1}{2n}\log\frac{|K_x+K_z|}{|K_z|}$, there exist codes for which the probability of error goes to zero as $n \to \infty$.

Theorem IV.1: Let the channel $\mathbf{Y}^{(n)} = \mathbf{X}^{(n)} + \mathbf{Z}^{(n)}$, where $K_x \in \mathcal{K}_x$, $K_z \in \mathcal{K}_z$ and $\mathbf{Z}^{(n)}$ satisfies conditions *C1* and *C2*. Suppose the transmitter knows the family $\mathcal{K}_z$ but not the actual covariance $K_z \in \mathcal{K}_z$. Let the receiver know the covariance K_z of $\mathbf{Z}^{(n)}$, but not the distribution. Then there exists a sequence of $(2^{n(C_n-\epsilon)}, n)$ randomly drawn codes with decoding rule given in (42) such that the probability of error $P_e^{(n)} \to 0$.

Proof: Let $\mathbf{X}^{(n)}(i), i = 1, \ldots, 2^{nR_n}$, be independent codewords chosen from a Gaussian distribution with covariance K_x. Let us define the event $E_i = \{\mathbf{X}^{(n)}(i), \mathbf{Y}^{(n)}$ are jointly ϵ-typical $\}$, where typicality is defined in (42). As the index of the codewords is assumed to be chosen from a uniform distribution we can assume w.l.o.g. that $\mathbf{X}^{(n)}(W), W = 1$ was the transmitted codeword. Hence we can write the probability of error $P[\mathcal{E}|W=1]$ using the union bound as

$$P[\mathcal{E}|W=1] \leq Pr[E_1^c] + \sum_{i=2}^{2^{nR_n}} Pr[E_i]. \tag{46}$$

We can write $Pr[E_i]$ for $i \neq 1$ as

$$\begin{aligned}
\Pr[E_i] &= \Pr\left[\frac{1}{n}M(\mathbf{X}^{(n)}(i);\mathbf{Y}^{(n)}) > \frac{1}{2n}\log\frac{|K_x+K_z|}{|K_z|} - \epsilon\right] \\
&\overset{(a)}{\leq} \mathbb{E}[e^{M(\mathbf{X}^{(n)}(i);\mathbf{Y}^{(n)})-n\beta}] \\
&\overset{(b)}{=} e^{\frac{1}{2}\log\frac{|K_x+K_z|}{|K_z|}-n\beta}\mathbb{E}\left[exp(\frac{1}{2}\mathbf{Y}^{(n)T}(K_x+K_z)^{-1}\mathbf{Y}^{(n)}\right. \\
&\quad \left. -\frac{1}{2}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)}))\right] \\
&\overset{(c)}{=} e^{-n\beta} \\
&\overset{(d)}{=} e^{-n(C_n-\epsilon)},
\end{aligned} \tag{47}$$

where (a) follows from the Chernoff bound, using $\beta = C_n - \epsilon$ and $C_n = \frac{1}{2n}\log\frac{|K_x+K_z|}{|K_z|}$; (b) follows by expanding $M(\mathbf{X}^{(n)}(i);\mathbf{Y}^{(n)})$; (c) uses Lemma IV.1, and (d) uses the definition of β. Therefore using (46) and (47) we have,

$$\begin{aligned}
P[\mathcal{E}|W=1] &\leq Pr[E_1^c] + e^{-n(C_n-R_n-\epsilon)} \\
&\overset{(a)}{\leq} (1-\epsilon)exp(-n\frac{\epsilon^2}{8}) + \epsilon + e^{-n(C_n-R_n-\epsilon)}
\end{aligned} \tag{48}$$

where (a) follows from Lemma IV.2. Therefore, $\lim_{n\to\infty} P[\mathcal{E}|W=1]=0$, if $R_n \leq C_n - \epsilon$. ■

This result needs to be interpreted with caution, as it is proved that the average error probability, averaged over randomly chosen codebooks, goes to zero. This does not show that a single codebook will suffice for all noise distributions in $\mathcal{K}_z$. Randomization may protect against noise distributions which are designed for specific codebooks. Given this caveat, we have shown that despite having a mismatched decoder (which treats the noise as Gaussian), we can transmit information reliably at rate $R_n = \frac{1}{2n}\log\frac{|K_x+K_z|}{|K_z|}$ using a codebook consisting of independently drawn Gaussian codewords.

Note that we have *not* used the "worst" covariance K_z^* for the decoding rule. It seems difficult to show whether the rate $R_n = \frac{1}{2n}\log\frac{|K_x+K_z^*|}{|K_z^*|}$ is achievable using the worst covariance for decoding rather than assuming that the noise covariance K_z is known at the decoder. It can be shown that the equivalent of Lemma IV.1 can be shown for K_z^* as well (and the proof is almost identical to that in the Appendix). However, to show the equivalent of Lemma IV.2 may be harder. An encouraging sign is an adaptation of the result in [13] Lemma 6.10 (pp 212–214, in the context of a convex class of compound channels), where it is shown that

$$\mathbb{E}_{\mathbf{Y},\mathbf{X}}[\log(\frac{f_{\mathbf{Y}^*|\mathbf{X}}(\mathbf{y}|\mathbf{z})}{f_{\mathbf{Y}^*}(\mathbf{y})})] \geq I(\mathbf{X};\mathbf{Y}^*),$$

where $\mathbf{Y}^*$ corresponds to the output of the channel that achieves the saddlepoint in the mutual information game. Using a similar set up, in our case this translates to:

$$\mathbb{E}_{\mathbf{X},\mathbf{Z}}[\mathbf{z}^T K_z^{*-1}\mathbf{z}] < \mathbb{E}_{\mathbf{X},\mathbf{Z}}[(\mathbf{z}+\mathbf{x})^T(K_x + K_z^*)^{-1}(\mathbf{z}+\mathbf{x})].$$

We can perhaps use this in order to prove a coding theorem using K_z^* for the decoding. However, this is just a conjecture, we have not proved such a result and it is not clear whether it is true.

V. Concluding Remarks

The existence of Gaussian saddlepoints in the mutual information game (under covariance constraints on signal and noise) implies the robustness of Gaussian codebooks. The problem of robust signal design reduces to waterfilling on the worst noise processes subject to covariance constraints. We show that for high signal power, the worst noise with a banded covariance constraint is the maximum entropy noise. However, the maximum entropy noise is not the worst noise for low signal powers. Hence robust signal design depends on the noise constraints as well as the available signal power.

ACKNOWLEDGEMENT

We gratefully acknowledge stimulating discussions with Erik Ordentlich and Bijit Halder. We wish to thank Amos Lapidoth for a very detailed and helpful review of the manuscript. He (along with his student Pascal Vontobel) also made the observation leading to Proposition II.1 and contributed its proof. We also wish to thank the referees for helpful reviews and we used their suggestion to use the characteristic function argument in (14).

Appendix

I.

Lemma A.1: If $\mathbf{X} \sim \mathcal{N}(0, K_x)$,

$$\mathbb{E}_{\mathbf{X}}[exp(-b(\mathbf{X}-\mathbf{a})^T\mathbf{A}^{-1}(\mathbf{X}-\mathbf{a})/2)] = \frac{|\mathbf{A}/b|^{1/2}}{|\mathbf{A}/b + K_x|^{1/2}} exp(-\mathbf{a}^T(K_x + \mathbf{A}/b)^{-1}\mathbf{a}/2) \tag{49}$$

Proof: We can always write $\mathbf{X} = \mathbf{C}\psi$, where $\psi \sim \mathcal{N}(0, I)$ and $\mathbf{C}$ is a nxm matrix. Here m denotes the rank of K_x. Therefore, we have

$$\begin{aligned}\mathbb{E}_{\mathbf{X}}[exp(-b(\mathbf{X}-\mathbf{a})^T\mathbf{A}^{-1}(\mathbf{X}-\mathbf{a})/2)] &= \mathbb{E}_{\psi}[exp(-b(\mathbf{C}\psi-\mathbf{a})^T\mathbf{A}^{-1}(\mathbf{C}\psi-\mathbf{a})/2)] \\ &\overset{(a)}{=} \frac{1}{|I + \mathbf{C}^T\mathbf{A}^{-1}b\mathbf{C}|^{1/2}} exp\left(-\mathbf{a}^T(\mathbf{A}^{-1}b\right. \\ &\quad \left. -\mathbf{A}^{-1}b\mathbf{C}(I + \mathbf{C}^T\mathbf{A}^{-1}b\mathbf{C})^{-1}\mathbf{C}^T\mathbf{A}^{-1}b)\mathbf{a}/2\right) \\ &\overset{(b)}{=} \frac{|\mathbf{A}/b|^{1/2}}{|\mathbf{A}/b + K_x|^{1/2}} exp(-\mathbf{a}^T(K_x + \mathbf{A}/b)^{-1}\mathbf{a}/2)\end{aligned} \tag{50}$$

where (a) follows from $\psi \sim \mathcal{N}(0, I)$ and (b) uses the matrix inversion lemma and the facts $K_x = \mathbf{C}\mathbf{C}^T$, $|I + UV| = |I + VU|$[27]. ■

Lemma IV.1: If $\mathbf{X}^{(n)} \sim \mathcal{N}(0, K_x)$ *and is independent of* $\mathbf{Y}^{(n)}$*, where* $\mathbf{Y}^{(n)}$ *has an arbitrary distribution, then we have*

$$\begin{aligned}\mathbb{E}[exp(\frac{1}{2}\mathbf{Y}^{(n)^T}(K_x + K_z)^{-1}\mathbf{Y}^{(n)} - & \tfrac{1}{2}(\mathbf{Y}^{(n)} - \mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)} - \mathbf{X}^{(n)}))] \\ & = exp(-\tfrac{1}{2}log(|K_x + K_z|/|K_z|)).\end{aligned} \tag{51}$$

Proof of Lemma IV.1:

$$\begin{aligned}&\mathbb{E}[exp(\frac{1}{2}\mathbf{Y}^{(n)^T}(K_x + K_z)^{-1}\mathbf{Y}^{(n)} - \frac{1}{2}(\mathbf{Y}^{(n)} - \mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)} - \mathbf{X}^{(n)}))] \\ &= \mathbb{E}_Y[e^{\frac{1}{2}\mathbf{Y}^{(n)^T}(K_x+K_z)^{-1}\mathbf{Y}^{(n)}}\mathbb{E}_{X|Y}[e^{-\frac{1}{2}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})}|\mathbf{Y}^{(n)}]] \\ &\stackrel{(a)}{=} \mathbb{E}_Y[e^{\frac{1}{2}\mathbf{Y}^{(n)^T}(K_x+K_z)^{-1}\mathbf{Y}^{(n)}}\mathbb{E}_X[e^{-\frac{1}{2}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})^T K_z^{-1}(\mathbf{Y}^{(n)}-\mathbf{X}^{(n)})}]] \\ &\stackrel{(b)}{=} \mathbb{E}_Y[\frac{|K_z|^{1/2}}{|K_x + K_z|^{1/2}}] \\ &= e^{-\frac{1}{2}\log\frac{|K_x+K_z|}{|K_z|}}\end{aligned} \tag{52}$$

where (a) follows from the fact that $\mathbf{X}^{(n)}$ and $\mathbf{Y}^{(n)}$ are independent, (b) follows from Lemma A.1. ■

Lemma IV.2: If $\mathbf{X}^{(n)} \sim \mathcal{N}(0, K_x)$ *and is independent of* $\mathbf{Z}^{(n)}$*, and* $\mathbb{E}[\mathbf{Z}^{(n)}\mathbf{Z}^{(n)^T}] = K_z > 0$*, then we have,*

$$\begin{aligned}Pr[\frac{1}{2n}\mathbf{Z}^{(n)^T}K_z^{-1}\mathbf{Z}^{(n)} & > \tfrac{1}{2n}(\mathbf{Z}^{(n)} + \mathbf{X}^{(n)})^T(K_x + K_z)^{-1}(\mathbf{Z}^{(n)} + \mathbf{X}^{(n)}) + \epsilon] \\ & \leq (1 - \epsilon)exp(-n\tfrac{\epsilon^2}{8}) + \epsilon\end{aligned} \tag{53}$$

Proof of Lemma IV.2:

$$\begin{aligned}&\Pr[\frac{1}{2n}\mathbf{z}^{(n)T}K_z^{-1}\mathbf{z}^{(n)} > \frac{1}{2n}(\mathbf{z}^{(n)} + \mathbf{x}^{(n)})^T(K_x + K_z)^{-1}(\mathbf{z}^{(n)} + \mathbf{x}^{(n)}) + \epsilon|\mathbf{z}^{(n)}] \\ &\stackrel{(a)}{\leq} \mathbb{E}[e^{\gamma(\frac{1}{2}\mathbf{Z}^{(n)T}K_z^{-1}\mathbf{Z}^{(n)} - \frac{1}{2}(\mathbf{Z}^{(n)}+\mathbf{X}^{(n)})^T(K_x+K_z)^{-1}(\mathbf{Z}^{(n)}+\mathbf{X}^{(n)}) - n\epsilon)}] \\ &\stackrel{(b)}{=} e^{-n\gamma\epsilon}e^{\frac{\gamma}{2}\mathbf{Z}^{(n)T}K_z^{-1}\mathbf{Z}^{(n)}}\mathbb{E}_X[e^{-\frac{\gamma}{2}(\mathbf{Z}^{(n)}+\mathbf{X}^{(n)})^T(K_x+K_z)^{-1}(\mathbf{Z}^{(n)}+\mathbf{X}^{(n)})}] \\ &\stackrel{(c)}{=} e^{-n\gamma\epsilon}e^{\frac{\gamma}{2}\mathbf{Z}^{(n)T}K_z^{-1}\mathbf{Z}^{(n)}}\frac{|(K_x + K_z)/\gamma|^{1/2}}{|K_x + (K_x + K_z)/\gamma|^{1/2}}e^{-\frac{1}{2}\mathbf{Z}^{(n)T}(K_x+(K_x+K_z)/\gamma)^{-1}\mathbf{Z}^{(n)}}\end{aligned} \tag{54}$$

where (a) follows from the Chernoff bound, γ is the Chernoff parameter, (b) follows from the independence of $\mathbf{X}^{(n)}$ and $\mathbf{z}^{(n)}$, and (c) follows from Lemma A.1. Let us define,

$$\begin{aligned}E(n, \gamma, \mathbf{z}^{(n)}) = \gamma\epsilon + \frac{1}{2n}\log(\frac{|K_x + (K_x + K_z)/\gamma|}{|(K_x + K_z)/\gamma|}) \\ -\frac{1}{2n}\mathbf{z}^{(n)T}(\gamma K_z^{-1} - (K_x + (K_x + K_z)/\gamma)^{-1})\mathbf{z}^{(n)}\end{aligned} \tag{55}$$

Hence we have the RHS of (54) given by $e^{-nE(n,\gamma,\mathbf{z}^{(n)})}$. We can rewrite $E(n,\gamma,\mathbf{z}^{(n)})$ as,

$$E(n,\gamma,\mathbf{z}^{(n)}) = \gamma\epsilon + \frac{1}{2n}\sum_{i=1}^{n}\log(1+\frac{\gamma}{1+\delta_i}) - \frac{1}{2n}\mathbf{z}^{(n)T}(\gamma K_z^{-1} - (K_x + (K_x+K_z)/\gamma)^{-1})\mathbf{z}^{(n)} \tag{56}$$

where $\delta_i = 1/\lambda_i(K_z^{-1/2}K_xK_z^{-*/2})$.

$$\begin{aligned} E(n,\gamma,\mathbf{z}^{(n)}) &\overset{(a)}{\geq} \gamma\epsilon + \frac{1}{2n}\sum_{i=1}^{n}\frac{\gamma}{1+\gamma+\delta_i} - \frac{1}{2n}\mathbf{z}^{(n)T}(\gamma K_z^{-1} - (K_x + (K_x+K_z)/\gamma)^{-1})\mathbf{z}^{(n)} \\ &\overset{(b)}{=} \gamma\epsilon + \frac{1}{\gamma+1}\frac{1}{2n}\mathrm{trace}([\gamma K_z^{-1} - (K_x + (K_x+K_z)/\gamma)^{-1}]K_z) \\ &\quad - \frac{1}{2n}\mathbf{z}^{(n)T}(\gamma K_z^{-1} - (K_x + (K_x+K_z)/\gamma)^{-1})\mathbf{z}^{(n)} \end{aligned} \tag{57}$$

where in (a) we have used $\log(1+x) \geq \frac{x}{1+x}$ for $x \geq 0$ and (b) is due to $\sum_{i=1}^{n}\frac{\gamma}{1+\gamma+\delta_i} = \frac{1}{\gamma+1}\mathrm{trace}([\gamma K_z^{-1} - (K_x + (K_x+K_z)/\gamma)^{-1}]K_z)$.

Let $\mathcal{C}_1 = \{\mathbf{z}^{(n)} : |\frac{1}{n}\mathbf{z}^{(n)T}K_z^{-1}\mathbf{z}^{(n)} - \mathbb{E}[\frac{1}{n}\mathbf{z}^{(n)T}K_z^{-1}\mathbf{z}^{(n)}]| < \epsilon/2\}$, $\mathcal{C}_2 = \{\mathbf{z}^{(n)} : |\frac{1}{n}\mathbf{z}^{(n)T}(K_x(1+\gamma)+K_z)^{-1}\mathbf{z}^{(n)} - \mathbb{E}[\frac{1}{n}\mathbf{z}^{(n)T}(K_x(1+\gamma)+K_z)^{-1}\mathbf{z}^{(n)}]| < \epsilon/2\}$. If $\mathcal{A} = \mathcal{C}_1 \cap \mathcal{C}_2$ then from *C1* and *C2* we have $Pr[\mathcal{A}] > 1-\epsilon$ for all $n \geq N(\epsilon)$. If we evaluate $E(n,\gamma,\mathbf{z}^{(n)})$ when $\mathbf{z}^{(n)} \in \mathcal{A}$ and denote it by $E(n,\gamma,\mathbf{z}^{(n)}|\mathcal{A})$ we have,

$$\begin{aligned} E(n,\gamma,\mathbf{z}^{(n)}|\mathcal{A}) &\overset{(a)}{\geq} \gamma\epsilon - \frac{\gamma}{2}[\epsilon-\gamma] \\ &= \frac{\gamma}{2}[\epsilon-\gamma] \\ &\overset{(b)}{\geq} \frac{\epsilon^2}{8} \end{aligned} \tag{58}$$

where (a) follows because $\mathbf{z}^{(n)} \in \mathcal{A}$ and (b) follows by choosing $\gamma \leq \epsilon$. The result follows by using (54), (58) $\gamma = \frac{\epsilon}{2}$, and $Pr[\mathcal{A}] > 1-\epsilon$. ■

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley and Sons, Inc., 1991.

[2] B. S. Choi and T. M. Cover, “An information theoretic proof of the Burg’s maximum entropy spectrum,” *Proeedings of the IEEE*, vol. 72, pp. 1094–1095, August 1984.

[3] N. M. Blachman, “Communication as a game,” in *Proceedings Wescon Conference*, pp. 61–66, August 1957.

[4] R. L. Dobrushin, “Optimum information transmission through a channel with unknown parameters,” *Radioteck. Elektron.*, vol. 4, pp. 1951–1956, December 1959.

[5] R. J. McEliece and W. E. Stark, “An information theoretic study of communication in the presence of jamming,” in *Proceedings Int. Conf. on Communications*, pp. 45.3.1–45.4.5, 1981.

[6] A. Lapidoth and P. Narayan, “Reliable Communication under channel uncertainty,” *IEEE Trans. on Information Theory, Commemorative Issue*, vol. 44, pp. 2148–2177, October 1998.

[7] T. Başar, “The Gaussian test channel with an intelligent jammer,” *IEEE Trans. on Information Theory*, vol. 29, pp. 152–157, January 1983.

[8] T. Başar and Y.-W. Wu, "A complete characterization of minimax and maximin encoder-decoder policies for communication channels with incomplete statistical description," *IEEE Trans. on Information Theory*, vol. 31, pp. 482–489, January 1985.

[9] S. Shamai and S. Verdu, "Worst case power constrained noise for binary input channels," *IEEE Transactions on Information Theory*, vol. 38, pp. 1494–1511, September 1992.

[10] C. R. Baker and I.-F. Chao, "Information capacity of channels with partially unknown noise. I: Finite dimensional channels," *SIAM Journal Applied Math.*, vol. 56, pp. 946–963, June 1996.

[11] A. Lapidoth, *Mismatched decoding of the multiple-access channel and some related issues in lossy source compression.* PhD thesis, Stanford University, Stanford, CA., 1995.

[12] I. Csiszar and P. Narayan, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Transactions on Information Theory*, vol. 37, pp. 18–26, Jan 1991.

[13] I. Csiszar and J. Korner, *Information Theory: Coding theorems for Discrete Memoryless Channels.* New York: Academic, 1982.

[14] N. M. Blachman, "On the capacity of a bandlimited channel perturbed by statistically dependent interference," *IRE Trans. on Information Theory*, vol. 8, pp. 48–55, January 1962.

[15] N. M. Blachman, "Effect of statistically dependent interference upon channel capacity," *IRE Trans. on Information Theory*, vol. 8, pp. 53–57, September 1962.

[16] R. Durrett, *Probability: theory and examples.* Duxbery Press, 2nd ed., 1995.

[17] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," *Information and Control*, vol. 37, pp. 34–39, 1978.

[18] M. S. Pinsker, "Calculation of the rate of information production by means of stationary random processes and the capacity of stationary channel," *Dokl. Akad. Nauk, USSR 111*, pp. 753–756, 1956. (in Russian).

[19] M. J. Osborne and A. Rubinstein, *A course in game theory.* Cambridge, Mass.: MIT Press, 1994.

[20] R. T. Rockafellar, *Convex Analysis.* Princeton, New Jersey: Princeton University Press, 1970.

[21] R. A. Horn and C. R. Johnson, *Matrix analysis.* United Kingdom: Cambridge University Press, 1990.

[22] B. Hughes and P. Narayan, "The capacity of a vector Gaussian arbitrarily varying channel," *IEEE Transactions on Information Theory*, vol. 34, pp. 995–1003, Sept. 1988.

[23] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Transactions on Information Theory*, vol. 42, pp. 1520–1529, Sept 1996.

[24] K. Fan, "On a theorem of Weyl concerning the eigenvalues of linear transformations II," *Proceedings National Academy of Sciences U.S.*, vol. 36, pp. 31–35, 1950.

[25] T. M. Cover and J. A. Thomas, "Determinant inequalities via information theory," *SIAM Journal of Matrix Analysis and it Applications*, vol. 9, pp. 384–392, 1988.

[26] D. G. Luenberger, *Optimization by Vector Space Methods.* New York: Wiley, 1969.

[27] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice Hall, 2nd ed., 1991.