The Effect of Maximal Rate Codes on the Interfering Message Rate

Ronit Bustin*, H. Vincent Poor* and S. Shamai (Shitz)†

* Dept. of Electrical Engineering, Princeton University, email: bustin.poor@princeton.edu
† Dept. of Electrical Engineering, Technion—Israel Institute of Technology,
   email: sshlomo@ee.technion.ac.il

Abstract

It is shown that the effect of a maximum rate transmission on an additional transmitted message, over the additive Gaussian noise channel, is, effectively, that of an additional additive Gaussian noise, meaning, that the behavior of the mutual information and minimum mean-square error (MMSE) are as if additional additive Gaussian noise were transmitted. This property provides corner points of the two-user Gaussian interference channel. The result holds under conditions requiring the point-wise convergence of the MMSE.

I. INTRODUCTION

In this work we examine the scenario of two transmitters interfering with each other over an additive Gaussian noise channel, where one of the two transmits at the maximum possible rate. The two-user Gaussian interference channel has been extensively investigated; however, its capacity region is, in general, still an open problem (see [1] and references therein for a recent overview of this problem).

Our focus, in this work, is on the following channel:

\[ Y_n = \sqrt{\text{snr}_1} X_n + \sqrt{a \text{snr}_2} Z_n + N_n \]  

where \( N_n \) represents a standard additive Gaussian noise vector with independent components, \( X_n \) carries the intended message and \( Z_n \) is the interfering signal. \( X_n \) and \( Z_n \) are independent of each other and independent of the additive Gaussian noise vector. The subscript \( n \) denotes that all vectors are length-\( n \) vectors. \( \text{snr}_1 \) and \( \text{snr}_2 \) are both non-negative scalar parameters and allow us to assume an average power constraint of 1 on both \( X_n \) and \( Z_n \) without loss of generality. Finally, the parameter \( a \) is also a non-negative scalar parameter, and is used here for consistency with the two-user Gaussian interference problem discussed in Section IV. We assume that there is a
sequence of point-to-point capacity achieving codebooks (i.e., that approach capacity, as $n \to \infty$, with vanishing probability of error). $X_n$ carries a message from the length-$n$ codebook. Thus, when $n \to \infty$,

$$R_x = \frac{1}{2} \log (1 + \text{snr}_1).$$

(2)

The above assumptions specify a corner point for the Gaussian interference channel, and the maximum possible rate of the transmission of $Z_n$ is an open question discussed thoroughly in [1] (where also the epsilon proximity to the corner point is studied). The conjecture, originated by Costa [2] (see also [1] and [3]), is that the rate of $Z_n$ is determined such that, as $n \to \infty$, it can be fully decoded at the receiver while considering $X_n$ as additive Gaussian noise. The achievability of such a corner point is straightforward: first reliably decode $Z_n$ while considering $X_n$ as additional additive Gaussian noise, remove it, and then reliably decode $X_n$. In this paper we provide a converse proof that shows that this is indeed the capacity region corner point. We do so, under several mild conditions: the first and second moment of both $X_n$ and $Z_n$ converge as $n \to \infty$, and the minimum mean-square error (MMSE) when estimating $Z_n$ from the channel output, and when estimating $\sqrt{\text{snr}_1} X_n + \sqrt{\text{snr}_2} Z_n$ from the channel output, converge as $n \to \infty$. Moreover, we show that the aforementioned achievability scheme is possible for any set of codebook sequences, complying with the above conditions, that attain this corner point. That is, that the reliable decoding of both messages is always possible.

The approach taken in this work is the exploration of the behavior of the derivatives of mutual information quantities with respect to the signal-to-noise ratio (SNR), rather than the mutual information quantities themselves. Using the fundamental relationship between information theory and estimation theory (I-MMSE) [4], the derivatives of some of these quantities are known, and given as the MMSE function in the estimation of one of the inputs from the output of an additive Gaussian channel. However, not all of the quantities follow this known relationship. Moreover, the mutual information quantities are related, meaning the behavior of one effects the others. Thus, these quantities are the center of our investigation. Basically, we use the fact that an optimal codebook over the additive Gaussian channel has a precise behavior for all SNR values [5], [6]. This fact enforces a specific behavior on the derivatives of related mutual information quantities.

The rest of this paper is constructed as follows: Section II contains some preliminary results, after which we formulate the main problem dealt with in this work. The main result and its proof is given in Section III. The Gaussian two-user interference channel is examined in Section IV. We conclude the paper in Section V.

II. PRELIMINARY RESULTS AND PROBLEM FORMULATION

A. Preliminary Results

Throughout the paper we mark any vector of finite length $n$ with a subscript $n$, and when $n \to \infty$ we remove the subscript. Moreover, assuming a length-$n$ input to a channel, denoted by $X_n$, and the corresponding output, denoted by $Y_n$, the per-component mutual information is

$$\frac{1}{n} I (X_n; Y_n).$$

(3)
Given a sequence of distributions over \( \{X_n\}_{n \geq 1} \), if the limit of (3) exists as \( n \to \infty \), it will be denoted as
\[
I(X; Y) = \lim_{n \to \infty} \frac{1}{n} I(X_n; Y_n). 
\] (4)

Similarly, the MMSE when estimating \( X_n \) from the channel output is defined as
\[
\text{MMSE}(X_n|Y_n) = \frac{1}{n} \text{Tr}(E_{X_n}(Y_n)) 
\] (5)
where \( E_{X_n}(Y_n) \) is the MMSE matrix when estimating \( X_n \) from \( Y_n \). If this sequence converges as \( n \to \infty \), then the limit is denoted by
\[
\text{MMSE}(X|Y) = \lim_{n \to \infty} \text{MMSE}(X_n|Y_n(\gamma)). 
\] (6)

Similarly, the limit of the MMSE matrix, if it exists, is denoted by
\[
E_X(Y) = \lim_{n \to \infty} E_{X_n}(Y_n). 
\] (7)

As we examine the behavior of the quantities in the limit, as \( n \to \infty \), we require some preliminary results. First note the following:

**Lemma 1:** The MMSE matrix, \( E_{X_n}(Y_n) \), converges point-wise if and only if the sequence of optimal estimators converges \( a.s. \) Moreover, the convergence of the MMSE matrix guarantees the point-wise convergence of \( \text{MMSE}(X_n|Y_n) \).

**Proof:** The first result is a direct consequence of the uniqueness, \( a.s. \), of the optimal estimator. The second claim is a simple consequence of the definition of \( \text{MMSE}(X_n|Y_n) \) as the normalized trace of \( E_{X_n}(Y_n) \).

Our second observation regards the additive white Gaussian noise (AWGN) channel.

**Lemma 2:** Given a sequence of distributions over \( \{X_n\}_{n \geq 1} \) with variance converging to 1 as \( n \to \infty \), if the following holds
\[
\lim_{n \to \infty} \text{MMSE}(X_n|Y_n) = \sqrt{\gamma \text{snr}} X_n + N_n = \text{MMSE}(X|Y_n) = \sqrt{\gamma \text{snr}} X + N = \frac{1}{1 + \gamma \text{snr}} 
\] (8)
where \( N_n \) is AWGN, then the MMSE matrix, \( E_{X_n}(Y_n) \), converges point-wise to the mean-square error (MSE) matrix of the bit-wise optimal linear estimator.

**Proof:** The proof is given in Appendix A.

Another simple, but important, consequence of the convergence of the MMSE matrix is the following:

**Lemma 3:** If both \( E_{X_n}(Y_n) \) and the second moment of \( \{X_n\}_{n \geq 1} \) converge point-wise then the sequence of the second moments of the optimal estimator converges. That is, for any component \( i \) of the length-\( n \) input vector, denoting \( S_n = E \{X_i|Y_n\} \) we have that
\[
\lim_{n \to \infty} E \{S_n^2\} = E \{S^2\}. 
\] (9)

**Proof:** The proof is a direct consequence of the fact that
\[
E_{Y_n} \{S_n^2\} = E_{X_n} \{X_i^2\} - (E_{X_n}(Y_n))_{ii} 
\]
and the right-hand-side converges by the assumptions.
Using the above lemma we conclude two important convergence results for the sequence of optimal estimators. 

**Lemma 4:** For a sequence of random variables \( \{ S_n \}_{n \geq 1} \) that converges a.s., if their second moments also converge then,

\[
\lim_{n \to \infty} \mathbb{E} \{|S_n|\} = \mathbb{E} \{|S|\} \quad \text{and} \\
\lim_{n \to \infty} \mathbb{E} \{|SS_n|\} = \mathbb{E} \{S^2\}.
\]

**Proof:** The proof is given in Appendix B.

The above result allow us to conclude the following:

**Corollary 1:** For a given sequence \( \{X_n\}_{n \geq 1} \) with second moments converging point-wise as \( n \to \infty \), if the sequence of optimal estimators converges a.s., then it also converges in \( L^1 \).

**Proof:** This is a result of Scheffe [7, Theorem 6.2] which states that an a.s. converging sequence which also has the property in (10) converges in \( L^1 \). As shown in Lemma 4 (10) holds when the second moments of the sequence converge. This property holds for the sequence of optimal estimators when the second moments of the sequence \( \{X_n\}_{n \geq 1} \) converge point-wise, as \( n \to \infty \).

The importance of the above corollary is in applying the next result from martingale theory to the sequence of estimators:

**Proposition 1 (Proposition 11 [8]):** Let \( \{S_n\}_{n \geq 1} \) be a sequence of real random variables such that \( \lim S_n = S \) exists a.s.. Then \( S_n \to S \) in \( L^1 \) if and only if the sequence \( \{S_n\}_{n \geq 1} \) is uniformly integrable. Furthermore, if the collection \( \{S_n\}_{n \geq 1} \) is uniformly integrable, then for every \( \sigma \)-algebra \( \mathcal{G} \),

\[
\lim_{n \to \infty} \mathbb{E} \{S_n|\mathcal{G}\} = \mathbb{E} \{S|\mathcal{G}\} \text{ in } L^1.
\]

**B. Problem Formulation**

As said, the focus of this work is on the model presented in (1). To simplify our notation we define the following:

\[
W_n = \sqrt{\text{snr}_1} X_n + \sqrt{\text{asnr}_2} Z_n
\]

which is a signal of average power limited to \( \text{snr}_1 + \text{asnr}_2 \). Define also

\[
Y_n(\gamma) = \sqrt{\gamma} W_n + N_n, \text{ for } \gamma \geq 0.
\]

Although we are assuming the transmission of length-\( n \) codewords over this channel, we examine the limit as \( n \to \infty \). For that purpose we formulate the problem on the set of distribution sequences over \( \{X_n, Z_n\}_{n=1}^{\infty} \), where the distribution sequence over \( \{Z_n\}_{n=1}^{\infty} \) is independent of the distribution sequence over \( \{X_n\}_{n=1}^{\infty} \).

We further define, for any arbitrary component in the length-\( n \) vector, e.g., \( X \) (a component of \( X_n \)) and \( Z \) (a component of \( Z_n \)), the following two sequences of random variables:

\[
G_n = \mathbb{E} \{X|Y_n, Z_n\} \\
H_n = \mathbb{E} \{Z|Y_n\}
\]

June 17, 2014 DRAFT
for every $n$. Note: we chose not to denote the components using a subscript $i$, e.g. $X_i$, so as to reduce the complexity of the notation. The observations are identical for every component of the length-$n$ vectors. Note that the sequence is well-defined since, for every $n$, the optimal estimator is known a.s. Furthermore, note that each estimator is an estimator from the output of the channel given an entirely different input sequence (and not some concatenation of the $n-1$ length input to the channel). Our underlying assumption is on the sequence of distributions over $\{X_n\}_{n=1}^\infty$ that constructs a code sequence of maximum rate. Assuming finite second moments, e.g. $\mathbb{E}\{||X_n||^2\} < \infty$ for all distributions in the sequence, and due to the convergence of the $k^{th}$ empirical distribution of such a sequence to the $k^{th}$ capacity achieving distribution shown by Shamai and Verdú [9], we can conclude that the first and second moments converge point-wise. This is due to the fact that weak convergence and uniform integrability guarantee the convergence of the moments [10, Chapter 6]. Second, given this assumption over $\{X_n\}_{n=1}^\infty$, we have that

$$I(X; Y(1)) = \frac{1}{2}\log(1 + \text{snr}_1),$$

and we can conclude that

$$I(X; \sqrt{\text{snr}_1}X + N) = \frac{1}{2}\log(1 + \text{snr}_1),$$

where $N$ is a standard additive Gaussian noise vector with independent components. Using the result in [5] and [6] we can conclude that

$$I(X; \sqrt{\gamma\text{snr}_1}X + N) = \frac{1}{2}\log(1 + \gamma\text{snr}_1), \quad \forall \gamma \in [0, 1].$$

In words, for any code of maximum possible rate over the Gaussian interference channel, its mutual-information behavior over a clean AWGN channel is known for every $\gamma$. The importance of this observation is twofold. First, through the I-MMSE relationship [4] also the MMSE behavior of such a sequence over the clean AWGN channel is known exactly for every SNR, when $n \to \infty$. Thus, our first conclusion from (17) is that the MMSE of such a code sequence converges as $n \to \infty$, that is

$$\lim_{n \to \infty} \text{MMSE}(X_n|\sqrt{\gamma\text{snr}_1}X_n + N_n) = \text{MMSE}(X|\sqrt{\gamma\text{snr}_1}X + N) = \frac{1}{1 + \gamma\text{snr}_1}, \quad \forall \gamma \in [0, 1)$$

and zero for all $\gamma \geq 1$, due to the reliable decoding of the message. Using Lemma 2 we can conclude that the MMSE matrix of this sequence converges point-wise, and according to Lemma 1 the optimal estimators converge a.s. Second, for all $\gamma \in [0, 1],$

$$I(W; Y(\gamma)) = h(Y(\gamma)) - h(Y(\gamma)|W)$$

$$= h(Y(\gamma)) - h(Y(\gamma)|X, Z) = I(X, Z; Y(\gamma))$$

$$= I(Z; Y(\gamma)) + I(X; Y(\gamma)|Z)$$

$$= I(Z; Y(\gamma)) + I(X; \sqrt{\gamma\text{snr}_1}X + N)$$

$$= I(Z; Y(\gamma)) + \frac{1}{2}\log(1 + \gamma\text{snr}_1)$$

$$= I(Z; Y(\gamma)) + \frac{1}{2}\log(1 + \gamma\text{snr}_1)$$

June 17, 2014
where $h(\cdot)$ is the differential entropy function. Taking the derivative with respect to $\gamma$, and using the I-MMSE relationship [4], we have for $\gamma \in [0, 1)$

$$\operatorname{MMSE}(W|Y(\gamma)) = 2 \frac{d}{d\gamma} I(Z; Y(\gamma)) + \operatorname{MMSE}(X|\sqrt{\gamma}\gamma_{\text{snr}} X + N) \cdot \gamma_{\text{snr}}$$

(20)

where the last transition is due to (18). Due to reliable decoding at $\gamma = 1$ we have for $\gamma \geq 1$

$$\operatorname{MMSE}(W|Y(\gamma)) = 2 \frac{d}{d\gamma} I(Z; Y(\gamma)).$$

(21)

Note that for the above observation we require the existence of $\operatorname{MMSE}(W|Y(\gamma))$ (or $I(W; Y(\gamma))$), this follows from Lemma 1 and the following assumption:

$$\lim_{n \to \infty} E(W_n(Y_n)) = E(W(Y)).$$

(22)

Furthermore, note that $I(Z; Y(\gamma))$ is not the transmission of a codeword over an AWGN channel, and thus we do not know much about its derivative. Our main result targets exactly the behavior of this quantity, showing that under our assumptions it behaves as if $Z$ were transmitted over an AWGN channel.

An additional assumption that we require is on the distribution sequence over $\{Z_n\}_{n=1}^{\infty}$. For this sequence we require the following:

$$\lim_{n \to \infty} E\{Z_n\} = \mu_z, \lim_{n \to \infty} E\{|Z_n|^2\} = \rho_z^2 \quad \forall i$$

$$\lim_{n \to \infty} E_{Z_n}(Y_n) = E_Z(Y).$$

(23)

In other words, we also require the point-wise convergence of the first and second moments and the MMSE matrix of this sequence, as $n \to \infty$. Note that we do not require for this additive signal to be reliably decoded, or to even be a codeword.

To conclude this section our assumptions on the input sequences result in the point-wise convergence of the first and second moments and the MMSE matrices, the $a.s.$ convergence of the estimators, and their uniform integrability. This allows us to use Proposition 1 on both sequences of optimal estimators.

III. MAIN RESULT

Our main result is the following:

Theorem 1: For any distribution sequence over $\{X_n, Z_n\}_{n=1}^{\infty}$ for which

$$\lim_{n \to \infty} E\{X_n\} = \mu_x, \lim_{n \to \infty} E\{|X_n|^2\} = \rho_x^2 \quad \text{and} \quad \lim_{n \to \infty} E_{X_n}(Y_n) = E_X(Y)$$

and

$$\lim_{n \to \infty} E_{Z_n}(Y_n) = E_Z(Y)$$

(24)

$$\lim_{n \to \infty} E\{Z_n\} = \mu_z, \lim_{n \to \infty} E\{|Z_n|^2\} = \rho_z^2 \quad \text{and} \quad \lim_{n \to \infty} E_{Z_n}(Y_n) = E_Z(Y)$$

(25)

$$\lim_{n \to \infty} E_{W_n}(Y_n) = E_W(Y),$$

(26)
we have that

$$\frac{d}{d\gamma} I (Z; Y (\gamma)) = \begin{cases} \text{MMSE} \left( Z \sqrt{\frac{1 + \gamma \text{snr}_1}{1 + \gamma \text{snr}_1}} Z + N \right) \cdot \frac{\text{asnr}_2}{(1 + \gamma \text{snr}_1)^2}, & \gamma \in [0, 1) \\ \text{MMSE} (W|Y (\gamma)), & \gamma \geq 1 \end{cases} .$$

(27)

First note that (24) does not fully characterize $X_n$ when $n \to \infty$, and is obtained by “good” codes (capacity achieving codes). This result can be viewed as an extension of the result in [5] and [6] and it shows that an optimal point-to-point codebook not only behaves as an i.i.d. Gaussian random vector when we examine its input-output interfering input, mutual information and MMSE over the AWGN channel, but also has an i.i.d. Gaussian effect on the additional additive i.i.d. Gaussian noise, when the mutual information and MMSE of $Z_n$ and the output are considered.

In order to prove the above claim we need two assisting results:

**Theorem 2:** The following behavior holds for any distribution sequence over $\{X_n, Z_n\}_{n=1}^{\infty}$ for which (24)-(26) hold:

$$\frac{d}{d\gamma} I (Z; Y (\gamma)) = \text{MMSE} (Z|Y (\gamma)) \cdot \frac{\text{asnr}_2}{1 + \gamma \text{snr}_1}.$$  

(28)

**Proof:** First note that from (26) and Lemma 1 we can conclude the point-wise convergence of $\text{MMSE}(W_n|Y_n(\gamma))$.

Recall from (20) that the following relationship holds for all $\gamma \in [0, 1)$:

$$\text{MMSE}(W|Y (\gamma)) = \frac{2}{d\gamma} I (Z; Y (\gamma)) + \frac{\text{snr}_1}{1 + \gamma \text{snr}_1}.$$  

(28)

Now, we examine $\text{MMSE}(W|Y (\gamma))$. First let us examine the optimal finite MMSE-wise estimator:

$$E \{ W_n | Y_n (\gamma) \} = E \{ \sqrt{\text{snr}_1} X_n + \sqrt{\text{asnr}_2} Z_n | Y_n (\gamma) \}$$

$$= E \{ E \{ \sqrt{\text{snr}_1} X_n + \sqrt{\text{asnr}_2} Z_n | Y_n (\gamma), Z_n \} | Y_n (\gamma) \}$$

$$= E \{ E \{ \sqrt{\text{snr}_1} X_n | Y_n (\gamma), Z_n \} + \sqrt{\text{asnr}_2} Z_n | Y_n (\gamma) \} .$$

Now, we consider $E \{ \sqrt{\text{snr}_1} X_n | Y_n (\gamma), Z_n \}$:

$$E \{ \sqrt{\text{snr}_1} X_n | Y_n (\gamma), Z_n \} = E \{ \sqrt{\text{snr}_1} X_n | Y_n (\gamma) - \sqrt{\gamma \text{asnr}_2} Z_n, Z_n \}$$

$$= E \{ \sqrt{\text{snr}_1} X_n | Y_n (\gamma) - \sqrt{\gamma \text{asnr}_2} Z_n \} .$$  

(29)

where the second equality is due to the fact that $Y_n (\gamma) - \sqrt{\gamma \text{asnr}_2} Z_n = \sqrt{\gamma \text{snr}_1} X_n + N_n$ is independent of $Z_n$. The estimator $E \{ \sqrt{\text{snr}_1} X_n | Y_n (\gamma), Z_n \}$ is simply the estimator of $X_n$ from a clean AWGN channel. Our assumption on $X_n$ is that, for all $\gamma \in [0, 1)$,

$$\text{MMSE}(X|\sqrt{\gamma \text{snr}_1} X + N) = \lim_{n \to \infty} \frac{1}{n} \text{Tr}(E_{X_n} (\sqrt{\gamma \text{snr}_1} X_n + N_n)) = \frac{1}{1 + \gamma \text{snr}_1} .$$  

(30)

According to Lemma 2 and the assumption on the convergence of the first and second moments, we conclude that the MMSE matrix converges. Thus, using Lemma 1 we conclude that the sequence of estimators converges a.s. to

June 17, 2014 DRAFT
the bit-wise optimal linear estimator:

\[
\lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}1}} X_n | Y_n(\gamma), Z_n \right\} = \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z) \\
\lim_{n \to \infty} G_n = \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z).
\]

(31)

Note that the above is a well-defined sequence of random variables since the optimal MMSE estimator is a.s. unique for every \(n\). Moreover, for every finite \(n\) the optimal estimator is markedly different from the bit-wise linear estimator and only as \(n \to \infty\) does it converge to it (due to the given performance requirement and the uniqueness property). Due to the convergence of the second moment, and as the sequence of optimal estimators converges a.s., we can apply Corollary 1 and Proposition 1 and conclude that the sequence of estimator is uniformly integrable.

Thus, applying Proposition 1 we have that for every \(\sigma\)-algebra \(\mathcal{G}\)

\[
\lim_{n \to \infty} E \left\{ G_n | \mathcal{G} \right\} = E \left\{ G | \mathcal{G} \right\} \text{ in } L^1
\]

\[
\lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}1}} X_n | Y_n(\gamma), Z_n \right\} | \mathcal{G} = E \left\{ \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z) | \mathcal{G} \right\} \text{ in } L^1.
\]

(32)

Also note that if we define the bit-wise optimal linear MMSE estimator, there is no meaning to the limit in the estimation of each component in the vector. Thus, we also have

\[
\lim_{n \to \infty} \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y_n(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z_n) = \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z)
\]

\[
\lim_{n \to \infty} E \left\{ \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y_n(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z_n) | \mathcal{G} \right\} = E \left\{ \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z) | \mathcal{G} \right\}.
\]

(33)

Thus, we can conclude that for any \(\sigma\)-algebra \(\mathcal{G}\)

\[
\lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}1}} X_n | Y_n(\gamma), Z_n \right\} | \mathcal{G} = \lim_{n \to \infty} E \left\{ \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y_n(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z_n) | \mathcal{G} \right\}
\]

(34)

and specifically,

\[
\lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}1}} X_n | Y_n(\gamma), Z_n \right\} | Y_n = \lim_{n \to \infty} E \left\{ \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y_n(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z_n) | Y_n \right\}
\]

(35)

in the \(L^1\) sense. From this we can proceed as follows:

\[
\lim_{n \to \infty} E \left\{ W_n | Y_n(\gamma) \right\} = \lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}1}} X_n | Y_n(\gamma), Z_n \right\} + \sqrt{\gamma_{\text{snr}2}} Z_n | Y_n(\gamma) \right\}
\]

\[
= \lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}1}} X_n | Y_n(\gamma), Z_n \right\} | Y_n(\gamma) \right\} + \lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}2}} Z_n | Y_n(\gamma) \right\}
\]

\[
= \lim_{n \to \infty} E \left\{ \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} (Y_n(\gamma) - \sqrt{\gamma_{\text{snr}2}} Z_n) | Y_n(\gamma) \right\} + \lim_{n \to \infty} E \left\{ \sqrt{\gamma_{\text{snr}2}} Z_n | Y_n(\gamma) \right\}
\]

\[
= \lim_{n \to \infty} \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} Y_n(\gamma) + \lim_{n \to \infty} \frac{\sqrt{\gamma_{\text{snr}2}}}{1 + \gamma_{\text{snr}1}} E \left\{ Z_n | Y_n(\gamma) \right\}
\]

\[
= \frac{\sqrt{\gamma_{\text{snr}1}}}{1 + \gamma_{\text{snr}1}} Y(\gamma) + \frac{\sqrt{\gamma_{\text{snr}2}}}{1 + \gamma_{\text{snr}1}} E \left\{ Z | Y(\gamma) \right\} \text{ in } L^1
\]

June 17, 2014 DRAFT
where transition $a$ is due to (35) and transition $b$ is due to Proposition 1 according to which the optimal estimators converge also in $L^1$. Now due to assumption (26) and Lemma 1 we can also conclude that the optimal estimators converge a.s., and thus the convergence shown above holds also in that sense.

Note that the above result also gives us the optimal (MMSE-wise) estimation scheme at $n \to \infty$, given our assumptions, and answers the following question: what is the estimation scheme that allows us to obtain this MMSE value at every $\gamma$?

Now let us consider the following proposed finite estimator:

$$
\hat{W}_n = \frac{\sqrt{\gamma_{\text{snr}_1}}}{1 + \gamma_{\text{snr}_1}} Y_n(\gamma) + \frac{\sqrt{\gamma_{\text{snr}_2}}}{1 + \gamma_{\text{snr}_1}} E\{Z_n|Y_n(\gamma)\}. \tag{36}
$$

According to Lemma 1 and our assumption in (25) we have that the optimal estimator $E\{Z_n|Y_n(\gamma)\}$ converges a.s. as $n \to \infty$. Thus, the sequence in (36) also converges a.s. as $n \to \infty$, that is,

$$
\lim_{n \to \infty} \hat{W}_n = \lim_{n \to \infty} E\{W_n|Y_n(\gamma)\} = \frac{\sqrt{\gamma_{\text{snr}_1}}}{1 + \gamma_{\text{snr}_1}} Y(\gamma) + \frac{\sqrt{\gamma_{\text{snr}_2}}}{1 + \gamma_{\text{snr}_1}} E\{Z|Y(\gamma)\}. \tag{37}
$$

Note that we have two sequences of estimators, one is the sequence of optimal estimators for every $n$, and the other sequence is $\hat{W}_n$, which is sub-optimal for any finite $n$. Both sequences converge a.s. to the same optimal MMSE-wise estimator as $n \to \infty$. Now, we claim that if the MSE of this sequence of estimators converges point-wise to some value, then it converges point-wise to $\text{MMSE}(W|Y(\gamma))$. This is due to the a.s. convergence of the estimators in (37) and Lemma 1.

We can calculate the MSE matrix of the proposed estimator, denoted by $E_W(W_n)$ (details are given in Appendix C), and after taking the limit of its normalized trace we obtain the following:

$$
\lim_{n \to \infty} \frac{1}{n} \text{Tr}\left( E_W(W_n) \right) = \frac{\text{MMSE}(Z|Y(\gamma))}{1 + \gamma_{\text{snr}_1}} + \frac{\text{MMSE}(Z|Y(\gamma))}{(1 + \gamma_{\text{snr}_1})^2} \tag{38}
$$

$$
+ \lim_{n \to \infty} \frac{1}{n} \text{Tr}\left( E\left( \left( \frac{\sqrt{\gamma_{\text{snr}_1}}}{1 + \gamma_{\text{snr}_1}} X_n - \frac{\sqrt{\gamma_{\text{snr}_1}}}{1 + \gamma_{\text{snr}_1}} Y_n(\gamma) - \sqrt{\gamma_{\text{snr}_2}} Z_n \right) \right)^T \left( \frac{1}{1 + \gamma_{\text{snr}_1}} \right) \right) \tag{39}
$$

where in the last transition we claim that the last two terms go to zero as $n \to \infty$. This is due to the orthogonality property of the optimal estimator. The details of this step are given in Appendix D. Using (39) in equation (20) we conclude our proof.

Note that Theorem 2 gives us a mutual information - MMSE (I-MMSE) like relationship (see e.g.[4]) for the quantity $I(Z;Y(\gamma))$, although the additive noise is not i.i.d. Gaussian, but rather $\sqrt{\gamma_{\text{snr}_1}} X + N$.

The second result required for the proof of Theorem 1 is the following:

**Theorem 3:** If for every distribution sequence over $\{Z_n\}_{n=1}^{\infty}$

$$
\frac{d}{d\gamma} I(Z;\sqrt{\gamma} Z + Q) = \frac{1}{2} \text{MMSE}(Z|\sqrt{\gamma} Z + Q), \tag{40}
$$
where $Q$ is some additive noise with a continuous distribution and variance one (at every index, but not required to be an independent random vector), independent of $Z$, then

$$\frac{d}{d\gamma} I(Z; \sqrt{\gamma}Z + Q) = \frac{1}{2} \text{MMSE}(Z|\sqrt{\gamma}Z + N)$$

(41)

where $N$ is a vector of independent standard normally distributed random variables.

**Proof:** Note that if $Q$ is i.i.d. Gaussian additive noise then (40) is the I-MMSE relationship [4] and it holds for any input distribution over $Z$. Our goal is to show that if (40) holds for any input distribution over $Z$ and some additive noise $Q$, then $Q$ must behave as i.i.d. additive Gaussian noise in terms of the MMSE function and the mutual information. Since our assumption is that (40) is valid for any input distribution over $Z$ we will examine, specifically, the i.i.d. Gaussian input distribution.

We begin by rewriting the derivative in (40) as follows:

$$\frac{d}{d\gamma} I(Z; \sqrt{\gamma}Z + Q) = \left( \frac{d}{da} I(Z; \sqrt{a}Q + Z) \right) \frac{da}{d\gamma}$$

(42)

where $a = \frac{1}{\gamma}$. The assumption in the above claim is for every input distribution over $Z$. The derivative on the right-hand side (RHS) in (42) was calculated for an i.i.d. Gaussian distribution over $Z$ in [11, Theorem 5]. Using their result (normalized by $n$) we have,

$$\frac{d}{d\gamma} I(Z; \sqrt{\gamma}Z + Q) = \frac{1}{2a} \text{MMSE}(Z|\sqrt{a}Q + Z) - 1 \cdot \frac{da}{d\gamma}$$

$$= \frac{\gamma}{2} \text{MMSE}(Z|\sqrt{\gamma}Z + Q) - 1 \cdot \left( -\frac{1}{\gamma^2} \right)$$

$$= -\frac{1}{2\gamma} (\text{MMSE}(Z|\sqrt{\gamma}Z + Q) - 1).$$

(43)

Using the assumption in (40), we have that

$$\frac{1}{2} \text{MMSE}(Z|\sqrt{\gamma}Z + Q) = -\frac{1}{2\gamma} (\text{MMSE}(Z|\sqrt{\gamma}Z + Q) - 1)$$

$$\text{MMSE}(Z|\sqrt{\gamma}Z + Q) = \frac{1}{1 + \gamma} = \text{LMMSE}(Z|\sqrt{\gamma}Z + Q).$$

(44)

As before, we see that the MMSE equals exactly the MSE obtained by linearly estimating every component of the vector, bit-by-bit. Thus, when $n \to \infty$ the additive noise is effectively normally distributed, making this bit-by-bit linear estimator the optimal MMSE-wise estimator. This concludes our proof.

We can now prove Theorem 1.

**Proof of Theorem 1:** When $\gamma \geq 1$ the result is a direct consequence of the assumption (24) and the chain rule of mutual information (19). As for $\gamma \in [0, 1)$, given Theorem 2 and the chain rule of differentiation, we have that

---

June 17, 2014 DRAFT
for every independent distribution sequence over \( \{ Z_n, X_n \}_{n=1}^{\infty} \) complying with (24)-(26),

\[
\text{MMSE} (Z|Y(\gamma)) = \frac{a\text{snr}^2}{(1 + \gamma\text{snr})^2} 
\]

\[
= 2 \frac{d}{d\gamma} I (Z; \sqrt{\gamma} a\text{snr} + \gamma\text{snr} X + N) 
= 2 \frac{d}{d\gamma} I \left( Z; \sqrt{\gamma} \frac{\text{snr}^1}{1 + \gamma\text{snr}^1} X + \sqrt{\frac{1}{1 + \gamma\text{snr}^1}} N \right) \cdot \frac{d\gamma}{d\gamma} 
\]

where we have defined \( \tilde{\gamma} = \frac{a\text{snr}^2}{1 + \gamma\text{snr}^1} \), and

\[
\frac{d\gamma}{d\gamma} = \frac{a\text{snr}^2 (1 + \gamma\text{snr}^1) - \gamma a\text{snr} \text{snr}^1}{(1 + \gamma\text{snr}^1)^2} = \frac{a\text{snr}^2}{(1 + \gamma\text{snr}^1)^2}. 
\]

We further define

\[
Q = \sqrt{\frac{\gamma\text{snr}^1}{1 + \gamma\text{snr}^1}} X + \sqrt{\frac{1}{1 + \gamma\text{snr}^1}} N 
\]

which is an additive noise of variance one (for every value of \( \gamma \)) at each time index, and of continuous distribution, since \( N \) is normally distributed. Returning to (45), we have that

\[
2 \frac{d}{d\gamma} I (Z; \sqrt{\gamma} Z + Q) = \text{MMSE} (Z|Y(\gamma)). 
\]

Since the above is valid for any distribution sequence over \( \{ Z_n \}_{n=1}^{\infty} \) (complying with (25) and (26), and specifically the i.i.d. Gaussian distribution) we can now apply the result of Theorem 3 to this expression and conclude that

\[
2 \frac{d}{d\gamma} I (Z; \sqrt{\gamma} Z + Q) = \text{MMSE} (Z|\sqrt{\gamma} Z + N) 
\]

and

\[
2 \frac{d}{d\gamma} I (Z; Y(\gamma)) = \text{MMSE} \left( Z|\sqrt{\gamma} a\text{snr}^2 + \frac{\gamma a\text{snr}^2}{1 + \gamma\text{snr}^1} Z + N \right) \cdot \frac{a\text{snr}^2}{(1 + \gamma\text{snr}^1)^2}, 
\]

meaning the additive interference has an i.i.d. Gaussian effect on \( Z \) in terms of the mutual information and MMSE.

\[ \text{IV. THE COSTA CONJECTURE} \]

Let us now consider the two-user Gaussian interference channel:

\[
Y_1 = \sqrt{\text{snr}^1} X + \sqrt{a\text{snr}^2} Z + N_1 \\
Y_2 = \sqrt{b\text{snr}^1} X + \sqrt{\text{snr}^2} Z + N_2, 
\]

where \( N_1 \) and \( N_2 \) represent standard additive Gaussian noise vectors with independent components (and can be assumed independent of each other). \( X \) and \( Z \) are assumed to be independent of each other and independent of the AWGN vectors. The average power constraint on both is 1; and \( 0 < a < 1 \) and \( 0 \leq b \). We assume maximum rate for the transmission of \( X \). For this setting we can apply Theorem 1 and obtain the next result:
Theorem 4: For any pair of codebooks for reliable transmission over the two-user Gaussian interference channel, (50), for which \(X\) is transmitted at maximum rate, \(E \{ ||X||^2 \} < \infty\), for all codebooks, and \(Z\) complying with the condition in (25) and (26), we have that

\[
\text{MMSE}(Z|\sqrt{\text{snr}}Z + N) = 0, \quad \forall \text{snr} \geq \frac{a_{\text{snr}2}}{1 + \text{snr}_1}.
\]

Proof: Any codebook of maximum rate, \(i.e., \frac{1}{2} \text{log}(1 + \text{snr}_1)\), and \(E \{ ||X||^2 \} < \infty\) complies with (24). Thus, the result of Theorem 1 applies to any such pair of reliable codebooks. Considering the limiting expression for the capacity of the two-user Gaussian interference channel [12], we have

\[
\bigcup_{P_Z, P_X} \left\{ R_x \leq I(Z; Y_1) = I(Z; Y(1)) \right\}.
\]

Let us assume that we have a pair of codebooks for the Gaussian interference channel. Thus, at \(Y(1)\) we can decode \(X\) reliably, meaning,

\[
I(Z; Y(1)) = I(Z; Y(1)|X) = I(Z; Y(1)|\sqrt{a_{\text{snr}2}Z + N}).
\]

On the other hand, in Theorem 1 (see also (48)), we have shown that

\[
I(Z; Y(1)) = I(Z; \sqrt{\frac{a_{\text{snr}2}}{1 + \text{snr}_1}}Z + N)
\]

meaning that, for any such pair of reliable codebooks, we have the following equality:

\[
I(Z; Y(1)) = I(Z; \sqrt{\frac{a_{\text{snr}2}}{1 + \text{snr}_1}}Z + N) = I(Z; \sqrt{a_{\text{snr}2}}Z + N)
\]

which by the I-MMSE relationship [4] means that the MMSE when estimating \(Z\) from the output of an AWGN channel is zero for any such codebooks for SNR in the region \(\frac{a_{\text{snr}2}}{1 + \text{snr}_1}, a_{\text{snr}_2}\), and hence any SNR \(\geq \frac{a_{\text{snr}2}}{1 + \text{snr}_1}\). This concludes our proof.

The following corollary follows immediately:

Corollary 2: Under the assumption of Theorem 4 we have that

\[
2 \frac{d}{d\gamma} I(Z; Y(\gamma)) = \left\{ \begin{array}{ll}
\text{MMSE} \left( Z|\sqrt{\frac{\gamma a_{\text{snr}2}}{1 + \gamma \text{snr}_1}}Z + N \right) \cdot \frac{a_{\text{snr}2}}{(1 + \gamma \text{snr}_1)^2}, & \gamma \in [0, 1) \\
0, & \gamma \geq 1
\end{array} \right.
\]

The above result provides us with corner points of the capacity region of the two-user Gaussian interference channel, under the conditions stated in Theorem 4. Depending on the value of \(b\) we have three different types of interference channels: \(b = 0\) is a one-sided (or Z) interference channel, \(b \in (0, 1)\) is the weak interference channel and \(b \geq 1\) is the mixed interference channel. Note that for the reverse case, in which \(Z\) is transmitted at maximum rate, the corner points of the capacity region of both the one-sided case and the mixed case are known, and are the corners yielding sum-capacity [1], [3], [13].

June 17, 2014 DRAFT
Corollary 3: For the one-sided Gaussian interference channel ((50) with $b = 0$), when the input distributions comply with the conditions in Theorem 4, we have the following corner point:

$$\left(\frac{1}{2} \log (1 + \text{snr}_1), \frac{1}{2} \log \left(1 + \frac{a \text{snr}_2}{1 + \text{snr}_1}\right)\right).$$  

(55)

Proof: Using Theorem 4 the solution to the maximization yields, therefore,

$$\max I(Z; \sqrt{\text{snr}_2} Z + N) = \frac{1}{2} \log \left(1 + \frac{a \text{snr}_2}{1 + \text{snr}_1}\right).$$  

(56)

Thus, concluding our proof.  

Note, that the above optimization problem can also be written as follows:

$$\max I(Z; \sqrt{\text{snr}_2} Z + N)$$

s.t.  

$$\text{MMSE}(Z|Y(1)) = 0$$  

(57)

for which the solution is according to the I-MMSE trade-off shown in [14], with the assumption that the additive noise $N + \sqrt{\text{snr}_1} X$ is i.i.d. Gaussian. Moreover, we do not need to limit this optimization problem to an average power constraint on $Z$ and can associate it with other constraints on $Z$, such as a peak amplitude constraint.

Theorem 5: For the weak two-user Gaussian interference channel ((50) with $b \in (0, 1)$), when the input distributions comply with the conditions in Theorem 4, we have the following corner points:

$$\left(\frac{1}{2} \log (1 + \text{snr}_1), \frac{1}{2} \log \left(1 + \frac{b \text{snr}_1}{1 + \text{snr}_2}\right)\right)$$

and

$$\left(\frac{1}{2} \log \left(1 + \frac{b \text{snr}_1}{1 + \text{snr}_2}\right), \frac{1}{2} \log (1 + \text{snr}_2)\right).$$  

(58)

(59)

Proof: From (24) and the fact that $b \in (0, 1)$ we can conclude that,

$$\text{MMSE}(X|\sqrt{\gamma b \text{snr}_1} X + N) = \frac{1}{1 + \gamma b \text{snr}_1}. \quad \forall \gamma \in [0, 1].$$

Thus, using Theorem 1 we have that,

$$2 \frac{d}{d\gamma} I(Z; \sqrt{\gamma b \text{snr}_1} X + \sqrt{\text{snr}_2} Z + N_2) = \text{MMSE} \left(Z; \sqrt{\frac{\gamma \text{snr}_2}{1 + \gamma b \text{snr}_1}} Z + N\right) \cdot \frac{\text{snr}_2}{(1 + \gamma b \text{snr}_1)^2}.$$ 

meaning that,

$$I(Z; Y_2) = I \left(Z; \sqrt{\frac{\text{snr}_2}{1 + b \text{snr}_1}} Z + N\right).$$  

(60)

Thus, maximizing $I(Z; Y_2)$ results with

$$\min \left\{\frac{1}{2} \log \left(1 + \frac{a \text{snr}_2}{1 + \text{snr}_1}\right), \frac{1}{2} \log \left(1 + \frac{\text{snr}_2}{1 + b \text{snr}_1}\right)\right\} = \frac{1}{2} \log \left(1 + \frac{a \text{snr}_2}{1 + \text{snr}_1}\right)$$

(61)

due to Theorem 4 and the maximum entropy result. This provides us with (58), (59) is simply the symmetric result and can be derived in the same manner. This concludes our proof.  

Finally, we have the following result for the mixed case.
**Theorem 6:** For the mixed two-user Gaussian interference channel ((50) with \( b \geq 1 \)), when the input distributions comply with the conditions in Theorem 4, we have the following corner point:

\[
\left( \frac{1}{2} \log (1 + \text{snr}_1), \frac{1}{2} \log \left( 1 + \frac{a \text{snr}_2}{1 + \text{snr}_1} \right) \right).
\]

(62)

**Proof:** Since \( b \geq 1 \) and \( Z \) can be reliably decoded at \( Y_2 \), we can also reliably decode \( X \) at \( Y_2 \). Thus, a trivial outer bound on this corner point is the multiple access channel (MAC) corner point:

\[
\left( \frac{1}{2} \log (1 + \text{snr}_1), \frac{1}{2} \log \left( 1 + \frac{\alpha \text{snr}_2 + b \text{snr}_1}{1 + \text{snr}_1} \right) \right).
\]

(63)

However, we also have that

\[ I(Z; Y_2) \leq I(Z; \sqrt{\text{snr}_2} Z + N) \]

and the maximization of the RHS of this inequality using Theorem 4 gives us

\[ I(Z; Y_2) \leq I(Z; \sqrt{\text{snr}_2} Z + N) \leq \frac{1}{2} \log \left( 1 + \frac{\alpha \text{snr}_2}{1 + \text{snr}_1} \right). \]

This upper bound is tighter than the MAC outer bound. It can be shown that this corner point can be achieved using Gaussian point-to-point codes and joint decoding, as shown in [15]. This concludes our proof. \( \blacksquare \)

Thus, treating the interference as Gaussian noise when \( b \in [1, \frac{1 - a + \text{snr}_1}{\alpha \text{snr}_1}] \) is optimal, not only in terms of the generalized degrees of freedom, as recently shown in [16]. Moreover, note that non-unique decoding at \( Y_2 \) [17], which potentially could obtain higher achievable rates, does not improve the above corner point. Note further that for any pair of codebooks (complying with (25)) that achieves these corner points, both messages can be reliably decoded. Moreover, at \( Y_1 \) the decoding of both messages can always be done sequentially, due to Theorem 4, which guarantees MMSE of zero for the estimation of \( Z \).

**V. Discussion and Conclusions**

In this work we have examined the effect of maximum rate transmission on an additional transmitted message, over the additive Gaussian channel. We have shown that a maximum rate transmission creates, effectively, additional additive Gaussian noise, in terms of the mutual information and MMSE. This observation leads to the fact that, if one requires to reliably decode the maximum rate transmission, the MMSE of any additional transmitted signal must be zero. This condition is orthogonal to any additional condition on the rate of this additional transmission. Using these observations we obtain corner points of the two-user Gaussian interference channel. These results hold under several conditions of moments convergence and MMSE convergence.

**Appendix**

**A. Proof of Lemma 2**

For any \( n \) the MMSE matrix can be upper bounded (in the positive semidefinite sense) by the MSE of the bit-wise optimal linear estimator, denoted by \( E_{X_n}^{\text{bitLinear}}(Y_n) \):

\[ E_{X_n}(Y_n) \preceq E_{X_n}^{\text{bitLinear}}(Y_n), \quad \forall n. \]

(65)
As \( n \to \infty \) the diagonal values of \( \mathcal{E}_{X_n}^{bitLinear}(Y_n) \) converge to \( \frac{1}{1 + \gamma_{snr}} \). As (65) holds for all \( n \), it holds also in the limit. Thus, we can conclude that the diagonal values of \( \mathcal{E}_{X_n}(Y_n) \) in the limit, as \( n \to \infty \), are all bounded by \( \frac{1}{1 + \gamma_{snr}} \). On the other hand, from (8) we know that the normalized trace converges to \( \frac{1}{1 + \gamma_{snr}} \), meaning that each and every diagonal value must converge exactly to \( \frac{1}{1 + \gamma_{snr}} \). Now, we claim that this leads to the convergence a.s. of the optimal estimator to the bit-wise optimal linear estimator. Assuming in contradiction that this is not the case, then we must have that

\[
\limsup_{n \to \infty} \left( \mathcal{E}_{X_n}^{bitLinear}(Y_n) - \mathcal{E}_{X_n}(Y_n) \right) > 0.
\]  

(66)

However, the diagonal values at the limit, as \( n \to \infty \) are all zero, and thus, due to the positive definite property we have that

\[
\lim_{n \to \infty} \left( \mathcal{E}_{X_n}^{bitLinear}(Y_n) - \mathcal{E}_{X_n}(Y_n) \right) = 0.
\]  

(67)

From this and the uniqueness a.s. we can conclude that the optimal estimators converge to the bit-wise linear estimator, a.s. Using Lemma 1 we conclude that also the MMSE matrix converges and from (67) we have that the convergence is to

\[
\lim_{n \to \infty} \mathcal{E}_{X_n}^{bitLinear}(Y_n).
\]  

(68)

This concludes the proof.

**B. Proof of Lemma 4**

Using Jensen’s inequality we have

\[
0 \leq (E \{|S_n| - |S|\})^2 \leq E \left\{ (|S_n| - |S|)^2 \right\}
\]

\[
= E \{ S_n^2 \} + E \{ S^2 \} - 2E \{ |S_n|S \}.
\]  

(69)

First, observe that

\[
2E \{ |S_n|S \} \leq E \{ S_n^2 \} + E \{ S^2 \}.
\]  

(70)

Taking the limit on both sides, as \( n \to \infty \), we have

\[
\liminf_{n \to \infty} 2E \{ |S_n|S \} \leq \limsup_{n \to \infty} E \{ |S_n|S \} \leq \lim_{n \to \infty} E \{ S_n^2 \} + E \{ S^2 \}
\]

\[
\liminf_{n \to \infty} E \{ |S_n|S \} \leq \limsup_{n \to \infty} E \{ |S_n|S \} \leq E \{ S^2 \}
\]

where we have used the assumption that the second moments converge, as \( n \to \infty \). Second, note that we have a positive sequence, \( \{|S_n|S\}_{n \geq 1} \), and thus according to Fatou’s Lemma [18, Lemma 4.8] we have that

\[
E \{ S^2 \} \leq \liminf_{n \to \infty} E \{ |S_n|S \}.
\]  

(71)

Putting the two together we have that

\[
\lim_{n \to \infty} E \{ |S_n|S \} = E \{ S^2 \}.
\]  

(72)
which proves (11). Now, going back to (69) and using (72) we obtain that
\[ 0 \leq \liminf_{n \to \infty} (E \{|S_n| - |S|\})^2 \leq \limsup_{n \to \infty} (E \{|S_n| - |S|\})^2 \leq 0 \] (73)
which proves (10). This concludes the proof.

C. Derivation of Equation 38

We can calculate the MSE matrix of the proposed estimator as follows:

\[
E_{W_n}(\hat{W}_n) = E\left\{ \left( W_n - \frac{\sqrt{\gamma_{\text{snr}}}}{1 + \gamma_{\text{snr}}} Y_n(\gamma) - \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} E\{Z_n|Y_n(\gamma)\} \right) \cdot \left( W_n - \frac{\sqrt{\gamma_{\text{snr}}}}{1 + \gamma_{\text{snr}}} Y_n(\gamma) - \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} E\{Z_n|Y_n(\gamma)\} \right)^T \right\}
\]

\[
= E\left\{ \left( \sqrt{\gamma_{\text{snr}}} X_n + \sqrt{a_{\text{snr}}} Z_n - \frac{\sqrt{\gamma_{\text{snr}}}}{1 + \gamma_{\text{snr}}} Y_n(\gamma) - \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} E\{Z_n|Y_n(\gamma)\} \right) \cdot \left( \sqrt{\gamma_{\text{snr}}} X_n + \sqrt{a_{\text{snr}}} Z_n - \frac{\sqrt{\gamma_{\text{snr}}}}{1 + \gamma_{\text{snr}}} Y_n(\gamma) - \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} E\{Z_n|Y_n(\gamma)\} \right)^T \right\}
\]

\[
= E\left\{ \left( \sqrt{\gamma_{\text{snr}}} X_n - Y_n(\gamma) + \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} Z_n - \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} E\{Z_n|Y_n(\gamma)\} \right) \cdot \left( \sqrt{\gamma_{\text{snr}}} X_n - Y_n(\gamma) + \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} Z_n - \frac{\sqrt{a_{\text{snr}}}}{1 + \gamma_{\text{snr}}} E\{Z_n|Y_n(\gamma)\} \right)^T \right\}
\]

June 17, 2014 DRAFT
\[
+E \left\{ \left( \frac{\sqrt{\text{snr}_2}}{1 + \gamma \text{snr}_1} (Z_n - E \{Z_n | Y_n(\gamma)\}) \right) \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Y_n(\gamma) - \sqrt{\text{snr}_2} Z_n) \right)^T \right\}.
\]

Taking the limit of the normalized trace of the above result we obtain (38).

D. Proof That Cross Terms Tend to Zero

The proof is identical for both terms so we will show it only for the first one.

\[
E \left\{ \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Y_n(\gamma) - \sqrt{\text{snr}_2} Z_n) \right) \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Z_n - E \{Z_n | Y_n(\gamma)\}) \right)^T \right\}
\]

\[
= E \left\{ \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Z_n - E \{Z_n | Y_n(\gamma)\}) \right) \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Y_n(\gamma) - \sqrt{\text{snr}_2} Z_n) \right) \right\}.
\]

(74)

where the last transition is due to the orthogonality property (the estimation error of the optimal estimator of \(X_n\) form \(Y_n\) and \(Z_n\) is orthogonal to any function of the measurements, for every \(n\)). Now we will take the limit as \(n \to \infty\) of the normalized trace of this matrix:

\[
\lim_{n \to \infty} \frac{1}{n} \text{Tr} \left( E \left\{ \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Y_n(\gamma) - \sqrt{\text{snr}_2} Z_n) \right) \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Z_n - E \{Z_n | Y_n(\gamma)\}) \right)^T \right\} \right)
\]

\[
\leq E \left\{ \lim_{n \to \infty} \frac{1}{n} \text{Tr} \left( E \left\{ \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Y_n(\gamma) - \sqrt{\text{snr}_2} Z_n) \right) \left( \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Z_n - E \{Z_n | Y_n(\gamma)\}) \right)^T \right\} \right) \right\}
\]

\[
= 0.
\]

(75)

The last transition is due to (31) and (33). Transition \(e\) requires more justification. We have the following sequence of elements (component-wise):

\[
E \{ \sqrt{\text{snr}_1} X | Y_n, Z_n \} - \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Y - \sqrt{\text{snr}_2} Z) \left( \frac{\sqrt{\text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - E \{Z | Y_n\}) \right)
\]

\[
= \left( G_n - \frac{\sqrt{\text{snr}_1}}{1 + \gamma \text{snr}_1} (Y - \sqrt{\text{snr}_2} Z) \right) \left( \frac{\sqrt{\text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H_n) \right).
\]

(76)
Due to the linearity of the expectation we need to show the following:

\[
\lim_{n \to \infty} E \left\{ \frac{\sqrt{\gamma \text{snr}}}{1 + \gamma \text{snr}_1} (Y - \sqrt{\gamma \text{snr}_2} Z) \left( \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H_n) \right) \right\} = E \left\{ \frac{\sqrt{\gamma \text{snr}}}{1 + \gamma \text{snr}_1} (Y - \sqrt{\gamma \text{snr}_2} Z) \left( \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H) \right) \right\}
\]

\[
\lim_{n \to \infty} E \left\{ G_n \left( \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H_n) \right) \right\} = E \left\{ G \left( \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H) \right) \right\}. 
\]  

(78)

The first claim can be obtained in a straightforward manner; using orthogonality,

\[
\lim_{n \to \infty} E \left\{ \frac{\sqrt{\gamma \text{snr}_1}}{1 + \gamma \text{snr}_1} (Y - \sqrt{\gamma \text{snr}_2} Z) \left( \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H_n) \right) \right\} = \lim_{n \to \infty} \frac{\gamma \text{snr}_1 \text{snr}_2}{(1 + \gamma \text{snr}_1)^2} (E \{ H_n^2 \} - \rho_{z_i}^2) 
\]

(79)

which converges due to the convergence of the second moment of \{Z_n\}_{n \geq 1} and the convergence of the second moment of the sequence \{H_n\}_{n \geq 1} (Lemma 3). The second claim is a bit harder to show. Note that

\[
E \left\{ (X - G_n) \left( \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H_n) \right) \right\} = 0, \quad \forall n
\]

(80)

due to the orthogonality of the estimation error. Thus, it also converges to zero as \( n \to \infty \). Moreover, we have that

\[
E \left\{ X \left( \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} (Z - H_n) \right) \right\} = \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} E \{ XZ \} - \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} E \{ XH_n \}
\]

\[
= \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} \mu_X \mu_Z - \frac{\sqrt{\gamma \text{snr}_2}}{1 + \gamma \text{snr}_1} E \{ XH_n \}. 
\]

(81)

Since we assume that the first moments converge, as \( n \to \infty \), it remains to be shown that

\[
\lim_{n \to \infty} E \{ XH_n \} = E_{X, Y} \{ XH \} 
\]

(82)

in order to prove that (81) converges, which leads to the required convergence in (78). According to Lemma 4 (11) we have that

\[
\lim_{n \to \infty} E \{ |HH_n| \} = E \{ H^2 \}. 
\]

(83)

Moreover, the sequence \{HH_n\}_{n \geq 1} converges a.s.; thus using Scheffe’s result [7, Theorem 6.2] we can conclude that the sequence \{HH_n\}_{n \geq 1} converges in \( L^1 \), and we can apply Proposition 1 on it to conclude also that

\[
\lim_{n \to \infty} E \{ HH_n \} = E \{ H^2 \}. 
\]

(84)

This leads to an \( L^2 \) convergence for this sequence,

\[
E \left\{ (H_n - H)^2 \right\} = E \{ H_n^2 \} + E \{ H^2 \} - 2E \{ HH_n \}
\]

\[
\lim_{n \to \infty} E \left\{ (H_n - H)^2 \right\} = 0. 
\]

(85)

Now, let us return to the sequence \{XH_n\}_{n \geq 1}. This sequence also converges a.s. Using the Cauchy-Schwartz inequality we have that

\[
0 \leq (E \{ X(H_n - H) \})^2 \leq E_{X,Y} \{ X^2 \} E \left\{ (H_n - H)^2 \right\}. 
\]

(86)

Taking the limit as \( n \to \infty \) on both sides and due to the finite second moment of \{X \}_{n \geq 1} we obtain that

\[
\lim_{n \to \infty} E \{ X(H_n - H) \} = 0
\]

and thus (82) holds. This concludes our proof.
REFERENCES

[1] I. Sason, “On the corner points of the capacity region of a two-user Gaussian interference channel,” submitted to the IEEE Transactions on Information Theory, August 2013, arXiv:1306.4934.

[2] M. H. M. Costa, “On the Gaussian interference channel,” IEEE Transactions on Information Theory, vol. 31, no. 5, pp. 607–615, September 1985.

[3] I. Sason, “On achievable rate regions for the Gaussian interference channel,” IEEE Transactions on Information Theory, vol. 50, no. 6, pp. 1345–1356, June 2004.

[4] D. Guo, S. Shamai (Shitz), and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” IEEE Transactions on Information Theory, vol. 51, no. 4, pp. 1261–1282, April 2005.

[5] M. Peleg, A. Sanderovich, and S. Shamai (Shitz), “Intrinsic information of good codes operating over Gaussian channels,” European Transactions on Telecommunications, vol. 18, no. 2, pp. 133–139, 2007.

[6] N. Merhav, D. Guo, and S. Shamai (Shitz), “Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions,” IEEE Transactions on Information Theory, vol. 50, no. 3, pp. 1400–1416, March 2010.

[7] D. Williams, Probability with Martingales. Cambridge University Press, 1991.

[8] S. Lalley, “Conditional expectation and martingales,” course notes: Statistics 313 Spring 2013/ Stochastic Processes II, 2013, http://galton.uchicago.edu/ lalley/Courses/313/index.html.

[9] S. Shamai (Shitz) and S. Verdú, “The empirical distribution of good codes,” IEEE Transactions on Information Theory, vol. 43, no. 3, pp. 836–846, May 1997.

[10] A. DasGupta, Asymptotic Theory of Statistics and Probability. Springer text in Statatics, 2008.

[11] D. Guo, S. Shamai (Shitz), and S. Verdú, “Proof of entropy power inequalities via MMSE,” in Proc. IEEE International Symposium on Information Theory (ISIT 2006), Seattle, WA, July 9-14 2006.

[12] R. Ahlswede, “Multi-way communication channels,” in Proc. IEEE International Symposium on Information Theory (ISIT 1971), pp. 23–52, Tsakhkador, Armenia, U.S.S.R, Sptember 1971.

[13] A. S. Motahari and A. K. Khandani, “Capacity bounds for the Gaussian interference channel,” IEEE Transactions on Information Theory, vol. 55, no. 2, pp. 620–643, February 2009.

[14] R. Bustin and S. Shamai (Shitz), “MMSE of “bad” codes,” IEEE Transactions on Information Theory, vol. 59, no. 2, pp. 733–743, February 2013.

[15] F. Baccelli, A. El Gamal, and D. Tse, “Interference networks with point-to-point codes,” in Proc. IEEE International Symposium on Information Theory (ISIT 2011), Saint Petersburg, Russia, July 31 - August 5 2011.

[16] A. Dytso, D. Tuninetti, and N. Devroye, “On discrete alphabets for Gaussian interference channels with a receiver lacking codebook knowledge,” Information Theory and Applications Workshop (ITA 2014), UCSD, San Diego, CA, USA, February 9 - 14, 2014, arXiv:1401.5536.

[17] S. S. Bidokhti, V. M. Prabhakaran, and S. N. Diggavi, “Is non-unique decoding necessary?” IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2594 – 2610, May 2014.

[18] R. Bartle, The Elements of Integration and Lebesgue Measure. Wiley Interscience, 1995.