Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2009, Article ID 370970, 8 pages
doi:10.1155/2009/370970

Research Article

An MMSE Approach to the Secrecy Capacity of the MIMO Gaussian Wiretap Channel

Ronit Bustin,[1] Ruoheng Liu,[2] H. Vincent Poor,[2] and Shlomo Shamai (Shitz)[1]

[1] *Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel*
[2] *Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA*

Correspondence should be addressed to Ronit Bustin, bustin@tx.technion.ac.il

Received 26 November 2008; Revised 15 March 2009; Accepted 21 June 2009

Recommended by Mérouane Debbah

This paper provides a closed-form expression for the secrecy capacity of the multiple-input multiple output (MIMO) Gaussian wiretap channel, under a power-covariance constraint. Furthermore, the paper specifies the input covariance matrix required in order to attain the capacity. The proof uses the fundamental relationship between information theory and estimation theory in the Gaussian channel, relating the derivative of the mutual information to the minimum mean-square error (MMSE). The proof provides the missing intuition regarding the existence and construction of an enhanced *degraded* channel that does not increase the secrecy capacity. The concept of enhancement has been used in a previous proof of the problem. Furthermore, the proof presents methods that can be used in proving other MIMO problems, using this fundamental relationship.

1. Introduction

The information theoretic characterization of secrecy in communication systems has attracted considerable attention in recent years. (See [1] for an exposition of progress in this area.) In this paper, we consider the general multiple-input multiple-output (MIMO) wiretap channel, presented in [2], with t transmit antennas and r and e receive antennas at the legitimate recipient and the eavesdropper, respectively:

$$\begin{aligned} \mathbf{Y}_r[m] &= \mathbf{H}_r\mathbf{X}[m] + \mathbf{W}_r[m], \\ \mathbf{Y}_e[m] &= \mathbf{H}_e\mathbf{X}[m] + \mathbf{W}_e[m], \end{aligned} \tag{1}$$

where $\mathbf{H}_r \in \mathbb{R}^{r\times t}$ and $\mathbf{H}_e \in \mathbb{R}^{e\times t}$ are assumed to be fixed during the entire transmission and are known to all three terminals. The additive noise terms $\mathbf{W}_r[m]$ and $\mathbf{W}_e[m]$ are zero-mean Gaussian vector processes independent across the time index m. The channel input satisfies a total power constraint:

$$\frac{1}{n}\sum_{m=1}^{n} \|\mathbf{X}[m]\|^2 \le P. \tag{2}$$

The secrecy capacity of a wiretap channel, defined by Wyner [3], as "perfect secrecy" capacity is the maximal rate such that the information can be decoded arbitrarily reliably by the legitimate recipient, while insuring that it cannot be deduced at any positive rate by the eavesdropper.

For a discrete memoryless wiretap channel with transition probability $P(\mathbf{Y}_r, \mathbf{Y}_e \mid \mathbf{X})$, a single-letter expression for the secrecy capacity was obtained by Csiszár and Körner [4]:

$$C_s = \max_{P(U,\mathbf{X})} \{I(U;\mathbf{Y}_r) - I(U;\mathbf{Y}_e)\}, \tag{3}$$

where U is an auxiliary random variable over a certain alphabet that satisfies the Markov relationship $U - \mathbf{X} - (\mathbf{Y}_r, \mathbf{Y}_e)$. This result extends to continuous alphabet cases with power constraint (2). Thus, in order to evaluate the secrecy capacity of the MIMO Gaussian wiretap channel we need to evaluate (3) under the power constraint (2). For the *degraded* case Wyner's single-letter expression of the secrecy capacity results from setting $U \equiv \mathbf{X}$ [3]:

$$C_s = \max_{P(\mathbf{X})} \{I(\mathbf{X};\mathbf{Y}_r) - I(\mathbf{X};\mathbf{Y}_e)\}. \tag{4}$$

The problem of characterizing the secrecy capacity of the MIMO Gaussian wiretap channel remained open until the work of Khisti and Wornell [5] and Oggier and Hassibi [6]. In their respective work, Khisti and Wornell [5] and

Oggier and Hassibi [6] followed an indirect approach using a Sato-like argument and matrix analysis tools. In [2] Liu and Shamai propose a more information-theoretic approach using the enhancement concept, originally presented by Weingarten et al. [7], as a tool for the characterization of the MIMO Gaussian broadcast channel capacity. Liu and Shamai have shown that an enhanced *degraded* version attains the same secrecy capacity as does the Gaussian input distribution. From the mathematical solution in [2] it is evident that such an enhanced channel exists; however it is not intuitive why, or how to construct such a channel.

A fundamental relationship between estimation theory and information theory for Gaussian channels was presented in [8]; in particular, it was shown that for the MIMO standard Gaussian channel,

$$\mathbf{Y} = \sqrt{\mathrm{snr}}\,\mathbf{H}\mathbf{X} + \mathbf{N} \tag{5}$$

and regardless of the input distribution, the mutual information and the minimum mean-square error (MMSE) are related (assuming real-valued inputs/outputs) by

$$\begin{aligned}\frac{d}{d\mathrm{snr}} I\left(\mathbf{X}; \sqrt{\mathrm{snr}}\,\mathbf{H}\mathbf{X} + \mathbf{N}\right) \\ = \frac{1}{2}\mathbb{E}\left\{\left|\left|\mathbf{H}\mathbf{X} - \mathbf{H}\mathbb{E}\left\{\mathbf{X} \mid \sqrt{\mathrm{snr}}\,\mathbf{H}\mathbf{X} + \mathbf{N}\right\}\right|\right|^2\right\},\end{aligned} \tag{6}$$

where $\mathbb{E}\{X \mid Y\}$ stands for the conditional mean of X given Y. This fundamental relationship and its generalizations [8, 9], referred to as the I-MMSE relations, have already been shown to be useful in several aspects of information theory: providing insightful proofs for entropy power inequalities [10], revealing the mercury/waterfilling optimal power allocation over a set of parallel Gaussian channels [11], tackling the weighted sum-MSE maximization in MIMO broadcast channels [12], illuminating extrinsic information of good codes [13], and enabling a simple proof of the monotonicity of the non-Gaussianness of independent random variables [14]. Furthermore, in [15] it has been shown that using this relationship one can provide insightful and simple proofs for multiuser single antenna problems such as the broadcast channel and the secrecy capacity problem. Similar techniques were later used in [16] to provide the capacity region for the Gaussian multireceiver wiretap channel.

Motivated by these successes, this paper provides an alternative proof for the secrecy capacity of the MIMO Gaussian wiretap channel using the fundamental relationship presented in [8, 9], which results in a closed-form expression for the secrecy capacity, that is, an expression that does not include optimization over the input covariance matrix, a difficult problem on its own due to the nonconvexity of the expression [5]. Thus, another important contribution of this paper is the explicit characterization of the optimal input covariance matrix that attains the secrecy capacity. The proof presented here provides the intuition regarding the existence and construction of the enhanced *degraded* channel which is central in the approach of [2]. Furthermore, the methods presented here could be used to tackle other MIMO problems, using the fundamental relationships shown in [8, 9].

2. Definitions and Preliminaries

Consider a canonical version of the MIMO Gaussian wiretap channel, as presented in [2]:

$$\begin{aligned}\mathbf{Y}_r[m] &= \mathbf{X}[m] + \mathbf{W}_r[m], \\ \mathbf{Y}_e[m] &= \mathbf{X}[m] + \mathbf{W}_e[m],\end{aligned} \tag{7}$$

where $\mathbf{X}[m]$ is a real input vector of length t, and $\mathbf{W}_r[m]$ and $\mathbf{W}_e[m]$ are additive Gaussian noise vectors with zero means and covariance matrices $\mathbf{K}_r$ and $\mathbf{K}_e$, respectively, and are independent across the time index m. The noise covariance matrices $\mathbf{K}_r$ and $\mathbf{K}_e$ are assumed to be positive definite. The channel input satisfies a power-covariance constraint:

$$\frac{1}{n}\sum_{m=1}^{n}\left(\mathbf{X}[m]\mathbf{X}[m]^T\right) \preccurlyeq \mathbf{S}, \tag{8}$$

where $\mathbf{S}$ is a positive semidefinite matrix of size $t \times t$, and "$\preccurlyeq$" denotes "less or equal to" in the positive semidefinite partial ordering between real symmetric matrices. Note that (8) is a rather general constraint that subsumes constraints that can be described by a compact set of input covariance matrices [7]. For example, assuming $C_s(\mathbf{S})$ is the secrecy capacity under a covariance constraint (8) we have according to [7] the following:

$$\begin{aligned}C_s(P) &= \max_{\mathrm{tr}(\mathbf{S}) \le P} C_s(\mathbf{S}), \\ C_s(P_1, P_2, \ldots, P_t) &= \max_{\mathbf{S}_{ii} \le P_i, i=1,2,\ldots,t} C_s(\mathbf{S}),\end{aligned} \tag{9}$$

where $C_s(P)$ is the secrecy capacity under a total power constraint (2), and $C_s(P_1, P_2, \ldots, P_t)$ is the secrecy capacity under a per antenna power constraint. As shown in [2, 7], characterizing the secrecy capacity of the general MIMO Gaussian wiretap channel (1) can be reduced to characterizing the secrecy capacity of the canonical version (7). For full details the reader is referred to [7], and [17, Theorem 3].

We first give a few central definitions and relationships that will be used in the sequel. We begin with the following definition:

$$\mathbf{E} = \mathbb{E}\left\{(\mathbf{X} - \mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X} \mid \mathbf{Y}\})^T\right\}, \tag{10}$$

that is, $\mathbf{E}$ is the covariance matrix of the estimation error vector, known as the MMSE matrix. For the specific case in which the input to the channel is Gaussian with covariance matrix $\mathbf{K}_x$, we define

$$\mathbf{E}_G = \mathbf{K}_x - \mathbf{K}_x(\mathbf{K}_x + \mathbf{K})^{-1}\mathbf{K}_x, \tag{11}$$

where $\mathbf{K}$ is the covariance matrix of the additive Gaussian noise, $\mathbf{N}$. That is, $\mathbf{E}_G$ is the error covariance matrix of the joint Gaussian estimator.

The fundamental relationship between information theory and estimation theory in the Gaussian channel gave rise to a variety of other relationships [8, 9]. In our proof, we will use the following relationship, given by Palomar and Verdú in [9]:

$$\nabla_{\mathbf{K}} I(\mathbf{X}; \mathbf{X} + \mathbf{N}) = -\mathbf{K}^{-1}\mathbf{E}\mathbf{K}^{-1}, \tag{12}$$

where $\mathbf{K}$ is the covariance matrix of the additive Gaussian noise, $\mathbf{N}$.

Our first observation regarding the relationship given in (12) is detailed in the following lemma.

Lemma 1. *For any two symmetric positive semidefinite matrices* $\mathbf{K}_1$ *and* $\mathbf{K}_2$, *such that* $0 \preccurlyeq \mathbf{K}_1 \preccurlyeq \mathbf{K}_2$ *and positive semidefinite matrix* $\mathbf{A}$, *the integral* $\int_{\mathbf{K}_1 \rightsquigarrow \mathbf{K}_2} \mathbf{K}^{-1}\mathbf{A}(\mathbf{K})\mathbf{K}^{-1}\, d\mathbf{K}$ *is nonnegative (where* $\mathbf{K}_1 \rightsquigarrow \mathbf{K}_2$ *is any path from* $\mathbf{K}_1$ *to* $\mathbf{K}_2$*).*

The proof of the lemma is given in Appendix A.

3. The Degraded MIMO Gaussian Wiretap Channel

We first consider the *degraded* MIMO Gaussian wiretap channel, that is, $\mathbf{K}_r \preccurlyeq \mathbf{K}_e$.

Theorem 1. *The secrecy capacity of the degraded MIMO Gaussian wiretap channel* (7), $\mathbf{K}_r \preccurlyeq \mathbf{K}_e$, *under the power-covariance constraint* (8) *is*

$$C_s = \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{S}\mathbf{K}_r^{-1}\right) - \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{S}\mathbf{K}_e^{-1}\right). \tag{13}$$

Proof. Using (12) the difference to be maximized, according to Wyner's single-letter expression (4), can be written as

$$I(\mathbf{X};\mathbf{Y}_r) - I(\mathbf{X};\mathbf{Y}_e) = \int_{\mathbf{K}_r \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\mathbf{E}\mathbf{K}^{-1}\, d\mathbf{K}. \tag{14}$$

This is due to the independence of the line integral (A.3) on the path in any open connected set in which the gradient is continuous [18].

The error covariance matrix of any optimal estimator is upper bounded (in the positive semidefinite partial ordering between real symmetric matrices) by the error covariance matrix of the joint Gaussian estimator, $\mathbf{E}_G$, defined in (11), for the same input covariance. Formally, $\mathbf{E} \preccurlyeq \mathbf{E}_G$, and thus one can express $\mathbf{E}$ as follows: $\mathbf{E} = \mathbf{E}_G - \mathbf{E}_0$, where $\mathbf{E}_0$ is some positive semidefinite matrix.

Due to this representation of $\mathbf{E}$ we can express the mutual information difference, given in (14), in the following manner:

$$\begin{aligned} &I(\mathbf{X};\mathbf{Y}_r) - I(\mathbf{X};\mathbf{Y}_e) \\ &= \int_{\mathbf{K}_r \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\mathbf{E}\mathbf{K}^{-1}\, d\mathbf{K} \\ &= \int_{\mathbf{K}_r \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}(\mathbf{E}_G - \mathbf{E}_0)\mathbf{K}^{-1}\, d\mathbf{K} \\ &= \int_{\mathbf{K}_r \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\mathbf{E}_G\mathbf{K}^{-1}\, d\mathbf{K} - \int_{\mathbf{K}_r \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\mathbf{E}_0\mathbf{K}^{-1}\, d\mathbf{K} \\ &\le \int_{\mathbf{K}_r \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\mathbf{E}_G\mathbf{K}^{-1}\, d\mathbf{K}, \end{aligned} \tag{15}$$

where the last inequality is due to Lemma 1 and the fact that $\mathbf{K}_r \preccurlyeq \mathbf{K}_e$. Equality in (15) is attained when $\mathbf{X}$ is Gaussian. Thus, we obtain the following expression:

$$\begin{aligned} C_s &= \max_{0 \preccurlyeq \mathbf{K}_x \preccurlyeq \mathbf{S}} \left\{ \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{K}_x\mathbf{K}_r^{-1}\right) - \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{K}_x\mathbf{K}_e^{-1}\right) \right\} \\ &= \max_{0 \preccurlyeq \mathbf{K}_x \preccurlyeq \mathbf{S}} \left\{ \frac{1}{2}\log\det(\mathbf{K}_r + \mathbf{K}_x) - \frac{1}{2}\log\det(\mathbf{K}_e + \mathbf{K}_x) \right\} \\ &\quad + \frac{1}{2}\log\frac{\det\mathbf{K}_e}{\det\mathbf{K}_r} \\ &= \max_{0 \preccurlyeq \mathbf{K}_x \preccurlyeq \mathbf{S}} \left\{ -\frac{1}{2}\log\frac{\det((\mathbf{K}_r + \mathbf{K}_x) + (\mathbf{K}_e - \mathbf{K}_r))}{\det(\mathbf{K}_r + \mathbf{K}_x)} \right\} \\ &\quad + \frac{1}{2}\log\frac{\det\mathbf{K}_e}{\det\mathbf{K}_r} \\ &= \max_{0 \preccurlyeq \mathbf{K}_x \preccurlyeq \mathbf{S}} \left\{ -\frac{1}{2}\log\det\left(\mathbf{I} + (\mathbf{K}_r + \mathbf{K}_x)^{-1}(\mathbf{K}_e - \mathbf{K}_r)\right) \right\} \\ &\quad + \frac{1}{2}\log\frac{\det\mathbf{K}_e}{\det\mathbf{K}_r} \\ &= -\frac{1}{2}\log\det\left(\mathbf{I} + (\mathbf{K}_r + \mathbf{S})^{-1}(\mathbf{K}_e - \mathbf{K}_r)\right) \\ &\quad + \frac{1}{2}\log\frac{\det\mathbf{K}_e}{\det\mathbf{K}_r} \\ &= \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{S}\mathbf{K}_r^{-1}\right) - \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{S}\mathbf{K}_e^{-1}\right). \end{aligned} \tag{16}$$

□

4. The General MIMO Gaussian Wiretap Channel

In considering the general case, we first note that one can apply the generalized eigenvalue decomposition [19] to the following two symmetric positive definite matrices:

$$\mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_r^{-1}\mathbf{S}^{1/2}, \qquad \mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_e^{-1}\mathbf{S}^{1/2}. \tag{17}$$

That is, there exists an invertible general eigenvector matrix, $\mathbf{C}$, such that

$$\begin{aligned} \mathbf{C}^T\left[\mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_e^{-1}\mathbf{S}^{1/2}\right]\mathbf{C} &= \mathbf{I}, \\ \mathbf{C}^T\left[\mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_r^{-1}\mathbf{S}^{1/2}\right]\mathbf{C} &= \mathbf{\Lambda}_r, \end{aligned} \tag{18}$$

where $\mathbf{\Lambda}_r = \operatorname{diag}\{\lambda_{1,r}, \lambda_{2,r}, \ldots, \lambda_{t,r}\}$ is a positive definite diagonal matrix. Without loss of generality, we assume that there are b $(0 \le b \le t)$ elements of $\mathbf{\Lambda}_r$ larger than 1:

$$\lambda_{1,r} \ge \ldots \ge \lambda_{b,r} > 1 \ge \lambda_{b+1,r} \ge \cdots \lambda_{t,r}. \tag{19}$$

Hence, we can write $\mathbf{\Lambda}_r$ as

$$\mathbf{\Lambda}_r = \begin{pmatrix} \mathbf{\Lambda}_1 & 0 \\ 0 & \mathbf{\Lambda}_2 \end{pmatrix}, \tag{20}$$

where $\mathbf{\Lambda}_1 = \operatorname{diag}\{\lambda_{1,r},\ldots,\lambda_{b,r}\}$, and $\mathbf{\Lambda}_2 = \operatorname{diag}\{\lambda_{b+1,r},\ldots,\lambda_{t,r}\}$. Since the matrix $\mathbf{I}+\mathbf{S}^{1/2}\mathbf{K}_e^{-1}\mathbf{S}^{1/2}$ is positive definite, the problem of calculating the generalized eigenvalues and the matrix $\mathbf{C}$ is reduced to a standard eigenvalue problem [19]. Choosing the eigenvectors of the standard eigenvalue problem to be orthonormal, and the requirement on the order of the eigenvalues, leads to an invertible matrix $\mathbf{C}$, which is $\mathbf{I}+\mathbf{S}^{1/2}\mathbf{K}_e^{-1}\mathbf{S}^{1/2}$-orthonormal. Using these definitions we turn to the main theorem of this paper.

Theorem 2. *The secrecy capacity of the MIMO Gaussian wiretap channel* (7)*, under the power-covariance constraint* (8)*, is*

$$\begin{aligned} C_s &= \frac{1}{2}\log\det(\mathbf{I}+\mathbf{S}\mathbf{K}_0^{-1}) - \frac{1}{2}\log\det(\mathbf{I}+\mathbf{S}\mathbf{K}_e^{-1}) \\ &= \frac{1}{2}\log\det(\mathbf{I}+\mathbf{K}_x^*\mathbf{K}_r^{-1}) - \frac{1}{2}\log\det(\mathbf{I}+\mathbf{K}_x^*\mathbf{K}_e^{-1}), \end{aligned} \tag{21}$$

where, using the invertible matrix $\mathbf{C}$ *defined in* (18) *one defines,*

$$\mathbf{K}_0 = \mathbf{S}^{1/2}\left[\mathbf{C}^{-T}\begin{pmatrix}\mathbf{\Lambda}_1 & 0\\ 0 & \mathbf{I}_{(t-b)\times(t-b)}\end{pmatrix}\mathbf{C}^{-1}-\mathbf{I}\right]^{-1}\mathbf{S}^{1/2}, \tag{22}$$

and letting $\mathbf{C}=[\mathbf{C}_1\mathbf{C}_2]$ *where* $\mathbf{C}_1$ *is the* $t\times b$ *submatrix and* $\mathbf{C}_2$ *is the* $t\times(t-b)$ *submatrix, one defines,*

$$\mathbf{K}_x^* = \mathbf{S}^{1/2}\mathbf{C}\begin{pmatrix}\left(\mathbf{C}_1^T\mathbf{C}_1\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{C}^T\mathbf{S}^{1/2}. \tag{23}$$

Proof. Following [7, Lemma 2], we may assume that $\mathbf{S}$ is (strictly) positive definite. We divide the proof into two parts: the converse part, that is, constructing an upper bound, and the achievability part-showing that the upper bound is attainable.

(a) Converse. Our goal is to evaluate the secrecy capacity expression (3). Due to the Markov relationship, $U-\mathbf{X}-(\mathbf{Y}_r,\mathbf{Y}_e)$, the difference to be maximized can be written as

$$\begin{aligned} &I(U;\mathbf{Y}_r)-I(U;\mathbf{Y}_e) \\ &\quad=\{I(\mathbf{X};\mathbf{Y}_r)-I(\mathbf{X};\mathbf{Y}_e)\}-\{I(\mathbf{X};\mathbf{Y}_r\mid U)-I(\mathbf{X};\mathbf{Y}_e\mid U)\}. \end{aligned} \tag{24}$$

We use the I-MMSE relationship (12) on each of the two differences in (24):

$$I(\mathbf{X};\mathbf{Y}_r)-I(\mathbf{X};\mathbf{Y}_e)=\int_{\mathbf{K}_r\rightsquigarrow\mathbf{K}_e}\mathbf{K}^{-1}\mathbf{E}\mathbf{K}^{-1}\,d\mathbf{K}, \tag{25}$$

where $\mathbf{E}=\mathbb{E}\{(\mathbf{X}-E[\mathbf{X}\mid\mathbf{Y}])(\mathbf{X}-E[\mathbf{X}\mid\mathbf{Y}])^T\}$, and

$$\begin{aligned} &I(\mathbf{X};\mathbf{Y}_r\mid U)-I(\mathbf{X};\mathbf{Y}_e\mid U) \\ &\quad=\mathbb{E}\{I(\mathbf{X};\mathbf{Y}_r\mid U=u)-I(\mathbf{X};\mathbf{Y}_e\mid U=u)\} \\ &\quad=\mathbb{E}\left\{\int_{\mathbf{K}_r\rightsquigarrow\mathbf{K}_e}\mathbf{K}^{-1}\mathbb{E}[(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U=u]) \right. \\ &\qquad\left.\times(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U=u])^T\mid U=u\right]\mathbf{K}^{-1}\,d\mathbf{K}\right\} \\ &\quad=\int_{\mathbf{K}_r\rightsquigarrow\mathbf{K}_e}\mathbf{K}^{-1}\mathbf{E}_u\mathbf{K}^{-1}\,d\mathbf{K}, \end{aligned} \tag{26}$$

where $\mathbf{E}_u=\mathbb{E}\{(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U])(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U])^T\}$. Thus, putting the two together, (24) becomes

$$I(U;\mathbf{Y}_r)-I(U;\mathbf{Y}_e)=\int_{\mathbf{K}_r\rightsquigarrow\mathbf{K}_e}\mathbf{K}^{-1}(\mathbf{E}-\mathbf{E}_u)\mathbf{K}^{-1}\,d\mathbf{K}. \tag{27}$$

We define, $\widetilde{\mathbf{E}}=\mathbf{E}-\mathbf{E}_u$, and obtain

$$\begin{aligned} \widetilde{\mathbf{E}} &= \mathbb{E}\left\{(\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U])(\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U])^T\right\} \\ &= \mathbb{E}\Big\{(\mathbb{E}[\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U]\mid\mathbf{Y}]-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U]) \\ &\qquad\times(\mathbb{E}[\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U]\mid\mathbf{Y}]-\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U])^T\Big\}. \end{aligned} \tag{28}$$

That is, $\widetilde{\mathbf{E}}$ is the error covariance of the optimal estimation of $\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U]$ from $\mathbf{Y}$, and as such it is positive semidefinite. It is easily verified that $\mathbf{K}_0$, defined in (22), satisfies both $\mathbf{K}_0\preccurlyeq\mathbf{K}_e$, and $\mathbf{K}_0\preccurlyeq\mathbf{K}_r$. The integral in (27) can be upper bounded using this fact and Lemma 1:

$$\begin{aligned} &I(U;\mathbf{Y}_r)-I(U;\mathbf{Y}_e) \\ &\quad=\int_{\mathbf{K}_0\rightsquigarrow\mathbf{K}_e}\mathbf{K}^{-1}\widetilde{\mathbf{E}}\mathbf{K}^{-1}\,d\mathbf{K}-\int_{\mathbf{K}_0\rightsquigarrow\mathbf{K}_r}\mathbf{K}^{-1}\widetilde{\mathbf{E}}\mathbf{K}^{-1}\,d\mathbf{K} \\ &\quad\le\int_{\mathbf{K}_0\rightsquigarrow\mathbf{K}_e}\mathbf{K}^{-1}\widetilde{\mathbf{E}}\mathbf{K}^{-1}\,d\mathbf{K}. \end{aligned} \tag{29}$$

Equality will be attained when the second integral equals zero. Using the upper bound in (29) we present two possible proofs that result with the upper bound given in (30). The more information-theoretic proof is given in the sequel, while the second, the more estimation-theoretic proof, is relegated to Appendix B.

The upper bound given in (29) can be viewed as the secrecy capacity of an MIMO Gaussian model, similar to the model given in (7), but with noise covariance matrices $\mathbf{K}_0$ and $\mathbf{K}_e$ and outputs $\mathbf{Y}_0[m]$ and $\mathbf{Y}_e[m]$, respectively. Furthermore, this is a *degraded* model, and it is well known that the general solution given by Csiszár and Körner [4],

reduces to the solution given by Wyner [3] by setting $U \equiv \mathbf{X}$. Thus, (29) becomes

$$\begin{aligned} & I(U;\mathbf{Y}_r) - I(U;\mathbf{Y}_e) \\ & \quad \le I(U;\mathbf{Y}_0) - I(U;\mathbf{Y}_e) \\ & \quad \le I(\mathbf{X};\mathbf{Y}_0) - I(\mathbf{X};\mathbf{Y}_e) \\ & \quad \le \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\mathbf{E}_G\mathbf{K}^{-1}\, d\mathbf{K} \\ & \quad \le \max_{0 \preccurlyeq \mathbf{K}_x \preccurlyeq \mathbf{S}} \left\{ \frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x\mathbf{K}_0^{-1}) \right. \\ & \qquad \left. -\frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x\mathbf{K}_e^{-1}) \right\} \\ & \quad = \frac{1}{2}\log\det(\mathbf{I} + \mathbf{S}\mathbf{K}_0^{-1}) \\ & \qquad - \frac{1}{2}\log\det(\mathbf{I} + \mathbf{S}\mathbf{K}_e^{-1}), \end{aligned} \tag{30}$$

where the third inequality is according to (15), and the last two transitions are due to Theorem 1, (16). This completes the converse part of the proof.

(b) Achievability. We now show that the upper bound given in (30) is attainable when $\mathbf{X}$ is Gaussian with covariance matrix $\mathbf{K}_x^*$, as defined in (23). The proof is constructed from the next three lemmas. We first prove that $\mathbf{K}_x^*$ is a legitimate covariance matrix, that is, it complies with the input covariance constraint (8).

Lemma 2. *The matrix $\mathbf{K}_x^*$ defined in* (23) *complies with the power-covariance constraint* (8)*, that is,*

$$0 \preccurlyeq \mathbf{K}_x^* \preccurlyeq \mathbf{S}. \tag{31}$$

The proof of Lemma 2 is given in Appendix C. In the next two Lemmas we show that $\mathbf{K}_x^*$ attains the upper bound given in (30).

Lemma 3. *The following equality holds:*

$$\frac{1}{2}\log\frac{\det(\mathbf{I} + \mathbf{S}\mathbf{K}_0^{-1})}{\det(\mathbf{I} + \mathbf{S}\mathbf{K}_e^{-1})} = \frac{1}{2}\log\frac{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_0^{-1})}{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_e^{-1})}. \tag{32}$$

Proof of Lemma 3. We first calculate the expression in the left hand side (assuming $\mathbf{S} \succ 0$), which is the upper bound in (30):

$$\begin{aligned} \frac{\det(\mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_0^{-1}\mathbf{S}^{1/2})}{\det(\mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_e^{-1}\mathbf{S}^{1/2})} &= \frac{\det\, \mathbf{C}^T(\mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_0^{-1}\mathbf{S}^{1/2})\mathbf{C}}{\det\, \mathbf{C}^T(\mathbf{I} + \mathbf{S}^{1/2}\mathbf{K}_e^{-1}\mathbf{S}^{1/2})\mathbf{C}} \\ &= \frac{\det\, \mathbf{\Lambda}_1}{\det\, \mathbf{I}} = \det\, \mathbf{\Lambda}_1, \end{aligned} \tag{33}$$

where we have used the generalized eigenvalue decomposition (18) and the definition of $\mathbf{K}_0$ (22). From (18) we note that,

$$\mathbf{K}_e^{-1} = \mathbf{S}^{-1/2}\left[\mathbf{C}^{-T}\begin{pmatrix}\mathbf{I} & 0\\ 0 & \mathbf{I}\end{pmatrix}\mathbf{C}^{-1} - \mathbf{I}\right]\mathbf{S}^{-1/2}. \tag{34}$$

Using (34) we can derive the following relationship (full details are given in Appendix D):

$$\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_0^{-1}) = \det\left(\left(\mathbf{C}_1^T\mathbf{C}_1\right)^{-1}\right)\det(\mathbf{\Lambda}_1). \tag{35}$$

And similarly we can derive

$$\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_e^{-1}) = \det\left(\left(\mathbf{C}_1^T\mathbf{C}_1\right)^{-1}\right). \tag{36}$$

Thus, we have

$$\frac{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_0^{-1})}{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_e^{-1})} = \det(\mathbf{\Lambda}_1), \tag{37}$$

which is the result attained in (33). This concludes the proof of Lemma 3. □

Lemma 4. *The following equality holds:*

$$\frac{1}{2}\log\frac{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_0^{-1})}{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_e^{-1})} = \frac{1}{2}\log\frac{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_r^{-1})}{\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_e^{-1})}. \tag{38}$$

Proof of Lemma 4. Due to the generalized eigenvalue decomposition (18) we have,

$$\mathbf{K}_r^{-1} = \mathbf{S}^{-1/2}\left[\mathbf{C}^{-T}\begin{pmatrix}\mathbf{\Lambda}_1 & 0\\ 0 & \mathbf{\Lambda}_2\end{pmatrix}\mathbf{C}^{-1} - \mathbf{I}\right]\mathbf{S}^{-1/2}. \tag{39}$$

Using similar steps as the ones used to obtain (35) we can show that,

$$\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_r^{-1}) = \det\left(\left(\mathbf{C}_1^T\mathbf{C}_1\right)^{-1}\right)\det(\mathbf{\Lambda}_1). \tag{40}$$

Thus, concluding the proof of Lemma 4. □

Putting all the above together we have that

$$\begin{aligned} & \frac{1}{2}\log\det(\mathbf{I} + \mathbf{S}\mathbf{K}_0^{-1}) - \frac{1}{2}\log\det(\mathbf{I} + \mathbf{S}\mathbf{K}_e^{-1}) \\ & \quad = \frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_0^{-1}) \\ & \qquad - \frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_e^{-1}) \\ & \quad = \frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_r^{-1}) \\ & \qquad - \frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x^*\mathbf{K}_e^{-1}), \end{aligned} \tag{41}$$

where the first equality is due to Lemma 3, and the second equality is due to Lemma 4. Thus, the upper bound given in (30) is attainable using the Gaussian distribution over $\mathbf{X}$, $U \equiv \mathbf{X}$, and $\mathbf{K}_\mathbf{x}^*$, defined in (23). This concludes the proof of Theorem 2. □

5. Discussion and Remarks

The alternative proof we have presented here uses the enhancement concept, also used in the proof of Liu and Shamai [2], in a more concrete manner. We have constructed a specific enhanced *degraded* model. The constructed model is the "tightest" enhancement possible in the sense that under the specified transformation, the matrix $\mathbf{C}^T[\mathbf{I}+\mathbf{S}^{1/2}\mathbf{K}_0^{-1}\mathbf{S}^{1/2}]\mathbf{C}$ is the "smallest" possible positive definite matrix, that is, both $\succcurlyeq \mathbf{\Lambda}_r$ and $\succcurlyeq \mathbf{I}$.

The specific enhancement results in a closed-form expression for the secrecy capacity, using $\mathbf{K}_0$. Furthermore, Theorem 2 shows that instead of $\mathbf{S}$ we can maximize the secrecy capacity by taking an input covariance matrix that "disregards" subchannels for which the eavesdropper has an advantage over the legitimate recipient (or is equivalent to the legitimate recipient). Mathematically, this allows us to switch back from $\mathbf{K}_0$ to $\mathbf{K}_r$, and thus to show that $\mathbf{K}_x^*$, explicitly defined, is the optimal input covariance matrix. Intuitively, $\mathbf{K}_x^*$ is the optimal input covariance for the legitimate receiver, since under the transformation, $\mathbf{C}$, it is $\mathbf{S}$ for the sub-channels for which the legitimate receiver has an advantage and zero otherwise.

The enhancement concept was used in addition to the I-MMSE approach in order to attain the upper bound in (30). The primary usage of these two concepts came together in (29), where we derived an initial upper bound. We have shown that the upper bound is attainable when $\mathbf{X}$ is Gaussian with covariance matrix $\mathbf{K}_x^*$. Thus, under these conditions the second integral in (29) should be zero, that is,

$$\begin{aligned}
&\int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_r} \mathbf{K}^{-1}\widetilde{\mathbf{E}}\mathbf{K}^{-1}\,d\mathbf{K} \\
&\quad = I(U;\mathbf{Y}_0) - I(U;\mathbf{Y}_r) \\
&\quad = I(\mathbf{X};\mathbf{Y}_0) - I(\mathbf{X};\mathbf{Y}_r) \\
&\quad = \frac{1}{2}\log\det\left(\mathbf{I}+\mathbf{K}_x^*\mathbf{K}_0^{-1}\right) \\
&\qquad - \frac{1}{2}\log\det\left(\mathbf{I}+\mathbf{K}_x^*\mathbf{K}_r^{-1}\right) \\
&\quad = 0,
\end{aligned} \tag{42}$$

where the second transition is due to the choice $U \equiv \mathbf{X}$, the third is due to the choice of a Gaussian distribution for $\mathbf{X}$ with covariance matrix $\mathbf{K}_x^*$, and the last equality is due to Lemma 4.

Appendices

A. Proof of Lemma 1

The inner product between matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A}\cdot\mathbf{B} = \operatorname{vec}\mathbf{A}^T\operatorname{vec}\mathbf{B}, \tag{A.1}$$

and the Schur product between matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$[\mathbf{A}\odot\mathbf{B}]_{ij} = [\mathbf{A}]_{ij}[\mathbf{B}]_{ij}. \tag{A.2}$$

For a function $\mathbf{G}$ with gradient $\nabla\mathbf{G}$ the line integral (type II) [18] is given by

$$\begin{aligned}
&\int_{\vec{r}_1 \rightsquigarrow \vec{r}_2} \nabla\mathbf{G}\,d\vec{r} \\
&\qquad = \int_{u=0}^{u=1} \nabla\mathbf{G}\left(\vec{r}_1 + u\left(\vec{r}_2 - \vec{r}_1\right)\right)\cdot\left(\vec{r}_2 - \vec{r}_1\right)du.
\end{aligned} \tag{A.3}$$

Thus in our case, where $\nabla\mathbf{G}$, $\vec{r}$ are $t\times t$ matrices, and $\nabla\mathbf{G} = \mathbf{K}^{-1}\mathbf{A}(\mathbf{K})\mathbf{K}^{-1}$ the integral over a path from $\mathbf{K}_1$ to $\mathbf{K}_2$ is equivalent to the following line integral:

$$\begin{aligned}
&\int_{u=0}^{1} \left(\mathbf{K}_1 + u(\mathbf{K}_2-\mathbf{K}_1)\right)^{-1}\mathbf{A}\left(\mathbf{K}_1 + u(\mathbf{K}_2-\mathbf{K}_1)\right) \\
&\qquad \times \left(\mathbf{K}_1 + u(\mathbf{K}_2-\mathbf{K}_1)\right)^{-1}\cdot(\mathbf{K}_2-\mathbf{K}_1)du \\
&= \int_{u=0}^{1} \underline{1}^T\left(\mathbf{K}_1 + u(\mathbf{K}_2-\mathbf{K}_1)\right)^{-1}\mathbf{A}\left(\mathbf{K}_1 + u(\mathbf{K}_2-\mathbf{K}_1)\right) \\
&\qquad \times \left(\mathbf{K}_1 + u(\mathbf{K}_2-\mathbf{K}_1)\right)^{-1}\odot(\mathbf{K}_2-\mathbf{K}_1)\underline{1}du.
\end{aligned} \tag{A.4}$$

Since the Schur product preserves the positive definite/semidefinite quality [20, 7.5.3], it is easy to see that when $0 \preccurlyeq \mathbf{K}_1 \preccurlyeq \mathbf{K}_2$, both are symmetric, and since $\mathbf{A}(\mathbf{K})$ is a positive semidefinite matrix for all $\mathbf{K}$, the integral is always nonnegative.

B. Second Proof of Theorem 2

The error covariance matrix of the optimal estimator $\widetilde{\mathbf{E}}$ can be written as $\widetilde{\mathbf{E}} = \widetilde{\mathbf{E}}_L - \mathbf{E}_0$, where both $\widetilde{\mathbf{E}}_L$ and $\mathbf{E}_0$ are positive semidefinite, and $\widetilde{\mathbf{E}}_L$ is the error covariance matrix of the optimal linear estimator of $\mathbb{E}[\mathbf{X}\mid\mathbf{Y},U]$ from $\mathbf{Y}$. Using this in (29), we have

$$\begin{aligned}
I(U;\mathbf{Y}_r) - I(U;\mathbf{Y}_e) &\le \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\widetilde{\mathbf{E}}K^{-1}\,d\mathbf{K} \\
&= \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\left(\widetilde{\mathbf{E}}_L - \mathbf{E}_0\right)\mathbf{K}^{-1}\,d\mathbf{K} \\
&= \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\widetilde{\mathbf{E}}_L\mathbf{K}^{-1}\,d\mathbf{K} \\
&\quad - \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\mathbf{E}_0\mathbf{K}^{-1}\,d\mathbf{K} \\
&\le \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\widetilde{\mathbf{E}}_L\mathbf{K}^{-1}\,d\mathbf{K},
\end{aligned} \tag{B.1}$$

where the last inequality is again due to Lemma 1. Equality will be attained when $\widetilde{\mathbf{E}}_L = \widetilde{\mathbf{E}}$, that is, when $\mathbf{E}_0 = 0$.

We denote $Z = \mathbb{E}[\mathbf{X}\mid\mathbf{Y},U]$. The optimal linear estimator has the following form:

$$\widetilde{\mathbf{E}}_L = \mathbf{C}_z - \mathbf{C}_{zy}\mathbf{C}_y{}^{-1}\mathbf{C}_{yz}, \tag{B.2}$$

where $\mathbf{C}_z$ is the covariance matrix of $\mathbf{Z}$, $\mathbf{C}_{zy}$ and $\mathbf{C}_{yz}$ are the cross-covariance matrices of $\mathbf{Z}$ and $\mathbf{Y}$, and $\mathbf{C}_y$ is the

covariance matrix of $\mathbf{Y}$. We can easily calculate $\mathbf{C}_{zy}$ and $\mathbf{C}_y$ (assuming zero mean):

$$\begin{aligned}\mathbf{C}_{zy} &= \mathbb{E}\Big\{\mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U]\mathbf{Y}^T\Big\} \\ &= \mathbb{E}\Big\{\mathbb{E}\Big[\mathbf{X}\mathbf{Y}^T \mid \mathbf{Y}, U\Big]\Big\} \\ &= \mathbb{E}\Big[\mathbf{X}\mathbf{Y}^T\Big] \\ &= \mathbf{C}_{xy} = \mathbf{K}_x \\ \mathbf{C}_y &= (\mathbf{K}_x + \mathbf{K}).\end{aligned} \tag{B.3}$$

Regarding $\mathbf{C}_z$ we can claim the following:

$$\begin{aligned}0 &\preccurlyeq \mathbb{E}\Big\{(\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U])(\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U])^T\Big\} \\ &= \mathbf{K}_x - \mathbb{E}\Big\{\mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U]\mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U]^T\Big\}\end{aligned} \tag{B.4}$$

thus,

$$\mathbb{E}\Big\{\mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U]\mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U]^T\Big\} = \mathbf{C}_z \preccurlyeq \mathbf{K}_x, \tag{B.5}$$

where equality, $\mathbf{C}_z = \mathbf{K}_x$, is attained when the estimation error is zero, that is, when $\mathbf{X} = \mathbb{E}[\mathbf{X} \mid \mathbf{Y}, U]$. Since $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ this can only be achieved when $U \equiv \mathbf{X}$ or $U \equiv \mathbf{N}$; however since the Markov property, $U - \mathbf{X} - (\mathbf{Y}_e, \mathbf{Y}_r)$, must be preserved, we conclude that $U \equiv \mathbf{X}$ in order to achieve equality.

We have $\mathbf{K}_x - \mathbf{C}_0 = \mathbf{C}_z$, where $\mathbf{C}_0$ is a positive semidefinite matrix, and the linear estimator is

$$\widetilde{\mathbf{E}}_L = \mathbf{K}_x - \mathbf{C}_0 - \mathbf{K}_x(\mathbf{K}_x + \mathbf{K})^{-1}\mathbf{K}_x. \tag{B.6}$$

Substituting this into the integral in (B.1) we have

$$\begin{aligned}&I(U;\mathbf{Y}_r) - I(U;\mathbf{Y}_e) \\ &\quad\le \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\widetilde{\mathbf{E}}_L\mathbf{K}^{-1}\, d\mathbf{K} \\ &\quad\le \int_{\mathbf{K}_0 \rightsquigarrow \mathbf{K}_e} \mathbf{K}^{-1}\Big(\mathbf{K}_x - \mathbf{K}_x(\mathbf{K}_x + \mathbf{K})^{-1}\mathbf{K}_x\Big)\mathbf{K}^{-1}\, d\mathbf{K} \\ &\quad= \frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x\mathbf{K}_0^{-1}) - \frac{1}{2}\log\det(\mathbf{I} + \mathbf{K}_x\mathbf{K}_e^{-1}) \\ &\quad\le \frac{1}{2}\log\det(\mathbf{I} + \mathbf{S}\mathbf{K}_0^{-1}) - \frac{1}{2}\log\det(\mathbf{I} + \mathbf{S}\mathbf{K}_e^{-1}),\end{aligned} \tag{B.7}$$

where the second inequality is due to Lemma 1, and the last inequality is due to Theorem 1, (16). The resulting upper bound equals the one given in (30). The rest of the proof follows via similar steps to those in the proof given in Section 4.

C. Proof of Lemma 2

Since the sub-matrix $\mathbf{C}_1^T\mathbf{C}_1$ is positive semidefinite it is evident that $0 \preccurlyeq \mathbf{K}_x^*$. Thus, it remains to show that $\mathbf{K}_x^* \preccurlyeq \mathbf{S}$.

Since $\mathbf{C}$ is invertible, in order to prove $\mathbf{K}_x^* \preccurlyeq \mathbf{S}$, it is enough to show that

$$\begin{pmatrix}\left(\mathbf{C}_1^T\mathbf{C}_1\right)^{-1} & 0 \\ 0 & 0\end{pmatrix} \preccurlyeq \mathbf{C}^{-1}\mathbf{C}^{-T} = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}. \tag{C.1}$$

We notice that,

$$\mathbf{C}^T\mathbf{C} = [\mathbf{C}_1\mathbf{C}_2]^T[\mathbf{C}_1\mathbf{C}_2] = \begin{pmatrix}\mathbf{C}_1^T\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{C}_2 \\ \mathbf{C}_2^T\mathbf{C}_1 & \mathbf{C}_2^T\mathbf{C}_2\end{pmatrix}. \tag{C.2}$$

Using blockwise inversion [20] we have

$$\begin{aligned}&\left(\mathbf{C}^T\mathbf{C}\right)^{-1} \\ &= \begin{pmatrix}\mathfrak{I} + \mathfrak{I}\mathbf{C}_1^T\mathbf{C}_2\mathbf{M}^{-1}\mathbf{C}_2^T\mathbf{C}_1\mathfrak{I} & -\mathfrak{I}\mathbf{C}_1^T\mathbf{C}_2\mathbf{M}^{-1} \\ -\mathbf{M}^{-1}\mathbf{C}_2^T\mathbf{C}_1\mathfrak{I} & \mathbf{M}^{-1}\end{pmatrix},\end{aligned} \tag{C.3}$$

where $\mathfrak{I}$ denotes $\left(\mathbf{C}_1^T\mathbf{C}_1\right)^{-1}$ and

$$\mathbf{M} = \mathbf{C}_2^T\mathbf{C}_2 - \mathbf{C}_2^T\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{C}_2 \succ 0 \tag{C.4}$$

due to the positive definite quality of $\mathbf{C}^T\mathbf{C}$ and the Schur Complement Lemma [20]. Hence,

$$\begin{aligned}&\left(\mathbf{C}^T\mathbf{C}\right)^{-1} - \begin{pmatrix}\mathfrak{I} & 0 \\ 0 & 0\end{pmatrix} \\ &\quad= \begin{pmatrix}\mathfrak{I}\mathbf{C}_1^T\mathbf{C}_2\mathbf{M}^{-1}\mathbf{C}_2^T\mathbf{C}_1\mathfrak{I} & -\mathfrak{I}\mathbf{C}_1^T\mathbf{C}_2\mathbf{M}^{-1} \\ -\mathbf{M}^{-1}\mathbf{C}_2^T\mathbf{C}_1\mathfrak{I} & \mathbf{M}^{-1}\end{pmatrix} \\ &\quad= \begin{pmatrix}I & -\mathfrak{I}\mathbf{C}_1^T\mathbf{C}_2 \\ 0 & \mathbf{I}\end{pmatrix}\begin{pmatrix}0 & 0 \\ 0 & \mathbf{M}^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{I} & 0 \\ -\mathbf{C}_2^T\mathbf{C}_1\mathfrak{I} & \mathbf{I}\end{pmatrix} \\ &\quad\succcurlyeq 0.\end{aligned} \tag{C.5}$$

D. Deriving Equation (35)

$$
\begin{aligned}
&\det\left(\mathbf{I}+\mathbf{K}_x^*\mathbf{K}_0^{-1}\right)\\
&\quad=\det\left(\mathbf{I}+\mathbf{S}^{1/2}\mathbf{C}\begin{pmatrix}\mathfrak{I} & 0\\ 0 & 0\end{pmatrix}\mathbf{C}^T\right.\\
&\qquad\left.\times\left[\mathbf{C}^{-T}\begin{pmatrix}\mathbf{\Lambda}_1 & 0\\ 0 & \mathbf{I}\end{pmatrix}\mathbf{C}^{-1}-\mathbf{I}\right]\mathbf{S}^{-1/2}\right)\\
&\quad=\det\left(\mathbf{I}+\begin{pmatrix}\mathfrak{I} & 0\\ 0 & 0\end{pmatrix}\mathbf{C}^T\left[\mathbf{C}^{-T}\begin{pmatrix}\mathbf{\Lambda}_1 & 0\\ 0 & \mathbf{I}\end{pmatrix}\mathbf{C}^{-1}-\mathbf{I}\right]\mathbf{C}\right)\\
&\quad=\det\left(\left(\mathbf{I}-\begin{pmatrix}\mathfrak{I} & 0\\ 0 & 0\end{pmatrix}\mathbf{C}^T\mathbf{C}+\begin{pmatrix}\mathfrak{I}\mathbf{\Lambda}_1 & 0\\ 0 & 0\end{pmatrix}\right)\right)\\
&\quad=\det\left(\mathbf{I}-\begin{pmatrix}\mathfrak{I} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathfrak{I}^{-1} & \mathbf{C}_1^T\mathbf{C}_2\\ \mathbf{C}_2^T\mathbf{C}_1 & \mathbf{C}_2^T\mathbf{C}_2\end{pmatrix}\right.\\
&\qquad\left.+\begin{pmatrix}\mathfrak{I}\mathbf{\Lambda}_1 & 0\\ 0 & 0\end{pmatrix}\right)\\
&\quad=\det\left(\mathbf{I}-\begin{pmatrix}\mathbf{I} & \mathfrak{I}\mathbf{C}_1^T\mathbf{C}_2\\ 0 & 0\end{pmatrix}+\begin{pmatrix}\mathfrak{I}\mathbf{\Lambda}_1 & 0\\ 0 & 0\end{pmatrix}\right)\\
&\quad=\det\begin{pmatrix}\mathfrak{I}\mathbf{\Lambda}_1 & -\mathfrak{I}\mathbf{C}_1^T\mathbf{C}_2\\ 0 & \mathbf{I}\end{pmatrix}\\
&\quad=\det\mathfrak{I}\,\det(\mathbf{\Lambda}_1).
\end{aligned}
\tag{D.1}
$$

Acknowledgments

This work has been supported by the Binational Science Foundation (BSF), the FP7 Network of Excellence in Wireless Communications NEWCOM++, and the U.S. National Science Foundation under Grants CNS-06-25637 and CCF-07-28208.

References

[1] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4-5, pp. 355–580, 2008.

[2] T. Liu and S. Shamai (Shitz), "A note on secrecy capacity of the multi-antenna wiretap channel," *IEEE Transaction on Information Theory*, vol. 55, no. 6, pp. 2547–2553, 2009.

[3] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[4] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[5] A. Khisti and G. Wornell, "The MIMOME channel," in *Proceedings of the 45th Annual Allerton Conference on Communication, Control and Computing*, Monticello, Ill, USA, September 2007.

[6] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '08)*, pp. 524–528, Toronto, Canada, July 2008.

[7] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[8] 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, 2005.

[9] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141–154, 2006.

[10] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '06)*, pp. 1011–1015, Seattle, Wash, USA, July 2006.

[11] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, 2006.

[12] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.

[13] M. Peleg, A. Sanderovich, and S. Shamai (Shitz), "On extrinsic information of good binary codes operating over Gaussian channels," *European Transactions on Telecommunications*, vol. 18, no. 2, pp. 133–139, 2007.

[14] A. M. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianness of the sum of independent random variables: a simple proof," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4295–4297, 2006.

[15] D. Guo, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: properties of the minimum mean-square error," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '08)*, Toronto, Canada, July 2008.

[16] E. Ekrem and S. Ulukus, "Secrecy capacity region of the Gaussian multi-receive wiretap channel," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '09)*, Seoul, Korea, June-July 2009.

[17] R. Liu, T. Liu, H. V. Poor, and S. Shamai (Shitz), "Multiple-input multiple-output Gaussian broadcast channels with coonfidential messages," submitted to *IEEE Transactions on Information Theory* and in Proceedings of *IEEE International Symposium on Information Theory (ISIT'09)*, Seoul, Korea, June-July 2009.

[18] T. M. Apostol, *Calculus, Multi-Variable Calculus and Linear Algebra, with Applications to Differential Equations and Probability*, Wiley, New York, NY, USA, 2nd edition, 1969.

[19] G. Strang, *Linear Algebra and Its Applications*, Wellesley-Cambridge Press, Wellesley, Mass, USA, 1998.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*, University Press, Cambridge, UK, 1985.