A Variational Perspective over an Extremal Entropy Inequality

Sangwoo Park
Dept. of Electrical & Computer Eng.
Texas A&M University
College Station, Texas, USA
Email: swpark78@neo.tamu.edu

Erchin Serpedin
Dept. of Electrical & Computer Eng.
Texas A&M University
College Station, Texas, USA
Email: serpedin@ece.tamu.edu

Marwa Qaraqe
Qatar Biomedical Research Institute
CBMR, Doha, Qatar &
Texas A&M University
College Station, Texas, USA
Email: marwa@neo.tamu.edu

***Abstract*—This paper proposes a novel variational approach for proving the extremal entropy inequality (EEI) [1]. Unlike previous proofs [1], [2], the proposed variational approach is simpler and it does not require neither the classical entropy power inequality (EPI) [1], [2] nor the channel enhancement technique [1]. The proposed approach is versatile and can be easily adapted to numerous other applications such as proving or extending other fundamental information theoretic inequalities such as the EPI, worst additive noise lemma, and Cramér-Rao inequality.**

I. Introduction

Extremal entropy inequality (EEI), motivated by multi-terminal information theoretic problems such as the vector Gaussian broadcast channel and the distributed source coding with a single quadratic distortion constraint was proposed by Liu and Viswanath [1]. EEI is an extension of the vector version of the classical entropy power inequality (EPI) and is obtained by enforcing a constraint on the covariance matrix of the random variable. So far, the EEI has found numerous applications such as assessing the capacity of the vector Gaussian broadcast channel [1], evaluating the data rates of distributed source coding applications with quadratic distortion constraints [1], and determining the capacity of the Gaussian wire-tap channel [3]. Also, EEI has attracted increased attention lately for its usefulness in information theory, and it was re-studied by several different authors [2], [4], [5], [6], [7]. Because of the covariance matrix constraint, EEI could not be proved directly by using the classical EPI. Therefore, new techniques (e.g., [2], [5], [8]) were adopted in the proofs reported in [1], [2], [5].

The first two proofs of EEI were proposed in [1], and were designed from an optimization viewpoint. Exploiting Karush-Kuhn-Tucker (KKT) conditions and the channel enhancement technique [8], four different optimization problems were sequentially constructed, and they were solved using additional techniques such as worst additive noise lemma, data processing inequality, and classical EPI. Unlike the proofs based on the channel enhancement technique and KKT conditions [1], Park et al. proposed a novel alternative proof without using KKT conditions [2]. The main technique adopted in [2] was the equality condition in the data processing inequality. The proposed proof was simpler, and it is a more direct approach compared to the proofs in [1].

The main purpose of this paper is to introduce a general functional analysis framework [9] for establishing the EEI in a simpler way. In the proposed functional approach, calculus of variations techniques are exploited and showed to present great use in establishing the EEI as well as numerous other results [5]. Unlike the previous proofs, the proposed variational approach does not exploit neither the classical EPI nor the channel enhancement technique. In addition, the proposed functional approach represents not only a novel alternative method for establishing the EEI but also a powerful approach for establishing numerous other fundamental information theoretic results such as the EPI, the worst additive noise lemma, Cramér-Rao inequality, as illustrated in the extended journal paper version of this work [5]. In this paper only a brief sketch of the major steps undertaken to prove the EEI via calculus of variations is provided. For more detailed explanations and additional applications of the proposed techniques, the reader is directed to reference [5]. The rest of this paper is organized as follows. First, the EEI and the outline of its proof are described in Section II. The more mathematical details of the proof are presented in Section III. Several conclusions outlining the contributions of this work are stated in Section IV.

II. Extremal Entropy Inequality

The EEI is introduced by means of the following theorem:

Theorem 1: Assume that $\mu \geq 1$ is an arbitrary but fixed constant, and $\mathbf{\Sigma}$ is a positive semi-definite matrix. Independent Gaussian random vectors W_G with covariance matrix $\mathbf{\Sigma}_W$ and V_G with covariance matrix $\mathbf{\Sigma}_V$ are supposed to be independent of an arbitrary random vector X with covariance matrix $\mathbf{\Sigma}_X \preceq \mathbf{\Sigma}$. Covariance matrices $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_V$ are both assumed to be positive definite. Then, there exists a Gaussian random vector X_G^* with covariance matrix $\mathbf{\Sigma}_{X^*}$ which satisfies the following inequality:

$$h(X+W_G) - \mu h(X+V_G) \leq h(X_G^*+W_G) - \mu h(X_G^*+V_G), \tag{1}$$

where $\mathbf{\Sigma}_{X^*} \preceq \mathbf{\Sigma}$.

Proof: The proposed simplified proof of the EEI runs as follows. First, select a Gaussian random vector $\tilde{W}_G$ whose covariance matrix $\mathbf{\Sigma}_{\tilde{W}}$ satisfies $\mathbf{\Sigma}_{\tilde{W}} \preceq \mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_{\tilde{W}} \preceq \mathbf{\Sigma}_V$. Since the Gaussian random vectors V_G and W_G can be represented simultaneously as the sum of two independent random vectors $\tilde{W}_G$ and $\hat{V}_G$, and $\tilde{W}_G$ and $\hat{W}_G$, respectively, using Lemma 2 from the Appendix, the left-hand side of (1) is expressed as follows:

$$\begin{aligned} & h(X+W_G) - \mu h(X+V_G) \\ \leq\ & h(X+\tilde{W}_G) - \mu h(X+V_G) + h(W_G) - h(\tilde{W}_G) \\ =\ & h(X+\tilde{W}_G) - \mu h(X+\tilde{W}_G+\hat{V}_G) \\ & \qquad + h(\tilde{W}_G+\hat{W}_G) - h(\tilde{W}_G). \end{aligned} \tag{2}$$

Since the equation in (2) will be maximized over $f_X(\mathbf{x})$, where $f_X(\mathbf{x})$ is the probability density function of random vector X, the last two terms in (2) are ignored, and by defining the new random vectors Y and $\hat{X}$ as $X+\tilde{W}_G+\hat{V}_G$ and $X+\tilde{W}_G$, respectively, the inequality in (1) is equivalently expressed as the following variational problem:

$$\max_{f_{\hat{X}}, f_Y} \quad h(\hat{X}) - \mu h(Y) + \mu(\mu-1) h(\hat{V}_G) \tag{3}$$

such that

$$\begin{aligned} & \iint f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} - 1 = 0, \\ & \iint f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \mathbf{x}\mathbf{x}^T d\mathbf{x} d\mathbf{y} - \mathbf{\Sigma}_{\hat{X}} \preceq \mathbf{0}, \\ & \iint f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \mathbf{y}\mathbf{y}^T d\mathbf{x} d\mathbf{y} - \mathbf{\Sigma}_{Y^*} = \mathbf{0}, \\ & \iint f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \Big(\mathbf{y}\mathbf{y}^T - \mathbf{x}\mathbf{x}^T - (\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})^T \Big) d\mathbf{x} d\mathbf{y} = \mathbf{0}, \\ & -\iint f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \log f_{\hat{X}}(\mathbf{x}) d\mathbf{x} d\mathbf{y} = p_{\hat{X}}, \\ & f_Y(\mathbf{y}) = \int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x}, \end{aligned} \tag{4}$$

where $p_{\hat{X}}$ stands for a constant, $\hat{X} = X + \tilde{W}_G$, $Y = \hat{X} + \hat{V}_G$, $W_G = \tilde{W}_G + \hat{W}_G$, $V_G = \tilde{W}_G + \hat{V}_G$, $\mathbf{\Sigma}_{\hat{X}} = \mathbf{\Sigma} + \mathbf{\Sigma}_{\tilde{W}}$, $\mathbf{\Sigma}_{Y^*} = \mathbf{\Sigma}_{X^*} + \mathbf{\Sigma}_V$, and $\mathbf{\Sigma}_{X^*}$ is the covariance matrix of the optimal solution X^*.

We will solve this variational problem by exploiting Euler's equations. As it is shown in Section III, the cost function in (3) is maximized when both $\hat{X}$ and Y are Gaussian random vectors. Therefore, from (2) one can infer further that

$$\begin{aligned} & h(X+\tilde{W}_G) - \mu h(X+\tilde{W}_G+\hat{V}_G) \\ & \qquad + h(\tilde{W}_G+\hat{W}_G) - h(\tilde{W}_G) \\ \leq\ & h(X_G^*+\tilde{W}_G) - \mu h(X_G^*+\tilde{W}_G+\hat{V}_G) \\ & \qquad + h(\tilde{W}_G+\hat{W}_G) - h(\tilde{W}_G). \end{aligned} \tag{5}$$

The important thing to remark here is that solving this variational problem requires only standard calculus of variations techniques, i.e., the proposed approach does not require neither the classical EPI nor the worst additive noise lemma. Based on Lemmas 1 and 2 presented the Appendix, the right-hand side of the equation (5) is further simplified to

$$\begin{aligned} & h(X_G^*+\tilde{W}_G) - \mu h(X_G^*+\tilde{W}_G+\hat{V}_G) \\ & \qquad + h(\tilde{W}_G+\hat{W}_G) - h(\tilde{W}_G) \\ =\ & h(X_G^*+W_G) - \mu h(X_G^*+\tilde{W}_G+\hat{V}_G). \end{aligned} \tag{6}$$

Therefore, taking into account all the inequalities (2), (5), and (6), the following EEI results:

$$\begin{aligned} & h(X+W_G) - \mu h(X+V_G) \\ \leq\ & h(X_G^*+W_G) - \mu h(X_G^*+V_G), \end{aligned}$$

and the proof is completed. ■

III. Details of the Proof based on the Calculus of Variations

The problem in (3) is more appropriately re-formulated as follows:

$$\begin{aligned} \max_{f_{\hat{X}}, f_Y} \quad & \int\int f_X(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \Big(\mu \log f_Y(\mathbf{y}) - \log f_{\hat{X}}(\mathbf{x}) \\ & -\mu(\mu-1) \log f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \Big) d\mathbf{x} d\mathbf{y} \end{aligned} \tag{7}$$

such that

$$\begin{aligned} & \iint f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = 1, \\ & \iint \Big(y_i y_j - x_i x_j - (y-x)_i (y-x)_j \Big) f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = 0, \\ & \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\iint x_i x_j \xi_i \xi_j f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} \right) \leq \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}^2 \xi_i \xi_j, \\ & \iint y_i y_j f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y} = \sigma^2_{Y^*_{ij}}, \\ & -\iint f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \log f_{\hat{X}}(\mathbf{x}) d\mathbf{x} d\mathbf{y} = p_{\hat{X}}, \\ & f_Y(\mathbf{y}) = \int\int f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x} d\mathbf{y}, \end{aligned} \tag{8}$$

where the arbitrary deterministic non-zero vector $\boldsymbol{\xi}$ is defined as $[\xi_1, \ldots, \xi_n]^T$, $\sigma^2_{Y^*_{ij}}$ denotes the i^{th} row and j^{th} column element of $\mathbf{\Sigma}_{Y^*}$, $i, j = 1, \ldots, n$.

Using Lagrange multipliers, the functional problem and its constraints in (7) are expressed as

$$\max_{f_{\hat{X}}, f_Y} \quad \int \left(\int K(\mathbf{x}, \mathbf{y}, f_{\hat{X}}, f_Y) d\mathbf{x} \right) + \tilde{K}(\mathbf{y}, f_Y) d\mathbf{y}, \tag{9}$$

where

$$
\begin{aligned}
& K(\mathbf{x},\mathbf{y},f_{\hat{X}},f_Y) \\
={}& f_{\hat{X}}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \Bigg(\mu \log f_Y(\mathbf{y}) - \log f_{\hat{X}}(\mathbf{x}) \\
& + \sum_{i=1}^{n}\sum_{j=1}^{n} \Big(\gamma_{ij} y_i y_j - \gamma_{ij} x_i x_j - \gamma_{ij} (y-x)_i (y-x)_j \\
& \qquad + \theta x_i x_j \xi_i \xi_j + \phi_{ij} y_i y_j \Big) \\
& - \mu(\mu-1) \log f_{\hat{V}}(\mathbf{y}-\mathbf{x}) + \alpha_0 - \alpha_1 \log f_{\hat{X}}(\mathbf{x}) - \lambda(\mathbf{y}) \Bigg), \\
& \tilde{K}(\mathbf{y}, f_Y) = \lambda(\mathbf{y}) f_Y(\mathbf{y}),
\end{aligned}
\tag{10}
$$

where α_0, α_1, γ_{ij}, θ, ϕ_{ij}, and $\lambda(\mathbf{y})$ denote the Lagrange multipliers.

The first-order variation condition takes the form:

$$
\begin{aligned}
& K'_{f_{\hat{X}}} \Big|_{f_{\hat{X}} = f_{\hat{X}^*}, f_Y = f_{Y^*}} \\
={}& f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \Bigg(\mu \log f_{Y^*}(\mathbf{y}) - \log f_{\hat{X}^*}(\mathbf{x}) - 1 - \alpha_1 + \alpha_0 \\
& + \sum_{i=1}^{n}\sum_{j=1}^{n} \Big(\gamma_{ij} y_i y_j - \gamma_{ij} x_i x_j - \gamma_{ij} (y-x)_i (y-x)_j \\
& \qquad + \theta x_i x_j \xi_i \xi_j + \phi_{ij} y_i y_j \Big) \\
& - \mu(\mu-1) \log f_{\hat{V}}(\mathbf{y}-\mathbf{x}) - \alpha_1 \log f_{\hat{X}^*}(\mathbf{x}) - \lambda(\mathbf{y}) \Bigg), \\
={}& 0
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
& \left(\int K d\mathbf{x} + \tilde{K} \right)'_{f_Y} \Bigg|_{f_Y = f_{\hat{X}^*}, f_Y = f_{Y^*}} \\
={}& \mu \frac{\int f_{\hat{X}^*}(\mathbf{x}) f_{\hat{V}}(\mathbf{y}-\mathbf{x}) d\mathbf{x}}{f_{Y^*}(\mathbf{y})} + \lambda(\mathbf{y}) \\
={}& 0,
\end{aligned}
\tag{12}
$$

where $K'_{f_{\hat{X}}}$ and $\tilde{K}'_{f_Y}$ are the first-order partial derivatives with respect to $f_{\hat{X}}$ and f_Y, respectively.[1]

Since the equalities in (11) and (12) must be satisfied for any $\mathbf{x}$ and $\mathbf{y}$, the following Gaussian density functions $f_{\hat{X}^*}$ and f_{Y^*} are obtained as solutions:

$$
\begin{aligned}
& f_{Y^*}(\mathbf{y}) \\
={}& (2\pi)^{-\frac{n}{2}} \left| -\frac{\mu}{2} (\mathbf{\Gamma} + \mathbf{\Phi})^{-1} \right|^{-\frac{1}{2}} \\
& \times \exp\left\{ -\frac{1}{2} \mathbf{y}^T \left(-\frac{\mu}{2} (\mathbf{\Gamma} + \mathbf{\Phi})^{-1} \right)^{-1} \mathbf{y} \right\} \\
& \times (2\pi)^{\frac{n}{2}} \left| -\frac{\mu}{2} (\mathbf{\Gamma} + \mathbf{\Phi})^{-1} \right|^{\frac{1}{2}} \exp\left\{ \frac{c_Y}{\mu} \right\} \\
={}& (2\pi)^{-\frac{n}{2}} \left| \mathbf{\Sigma}_{Y^*} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{y}^T \mathbf{\Sigma}_{Y^*}^{-1} \mathbf{y} \right\},
\end{aligned}
\tag{13}
$$

[1]Throughout the paper, to simplify the notations, the arguments of functionals or functions are omitted unless the arguments are ambiguous or confusing.

$$
\begin{aligned}
& f_{\hat{V}}(\mathbf{y}-\mathbf{x}) \\
={}& (2\pi)^{-\frac{n}{2}} \left| -\frac{\mu(\mu-1)}{2} \mathbf{\Gamma}^{-1} \right|^{-\frac{1}{2}} \\
& \times \exp\left\{ -\frac{1}{2} (\mathbf{y}-\mathbf{x})^T \left(-\frac{\mu(\mu-1)}{2} \mathbf{\Gamma}^{-1} \right)^{-1} (\mathbf{y}-\mathbf{x}) \right\} \\
& \times (2\pi)^{\frac{n}{2}} \left| -\frac{\mu(\mu-1)}{2} \mathbf{\Gamma}^{-1} \right|^{\frac{1}{2}} \exp\left\{ -\frac{c_{\hat{V}}}{\mu(\mu-1)} \right\} \\
={}& (2\pi)^{-\frac{n}{2}} \left| \mathbf{\Sigma}_{\hat{V}} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{y}-\mathbf{x})^T \mathbf{\Sigma}_{\hat{V}}^{-1} (\mathbf{y}-\mathbf{x}) \right\},
\end{aligned}
\tag{14}
$$

$$
\begin{aligned}
& f_{\hat{X}^*}(\mathbf{x}) \\
={}& (2\pi)^{-\frac{n}{2}} \left| -\frac{1-\alpha_1}{2} (\mathbf{\Gamma} - \theta \mathbf{\Xi})^{-1} \right|^{-\frac{1}{2}} \\
& \times \exp\left\{ -\frac{1}{2} \mathbf{x}^T \left(-\frac{1-\alpha_1}{2} (\mathbf{\Gamma} - \theta \mathbf{\Xi})^{-1} \right)^{-1} \mathbf{x} \right\} \\
& \times (2\pi)^{\frac{n}{2}} \left| -\frac{1-\alpha_1}{2} (\mathbf{\Gamma} - \theta \mathbf{\Xi})^{-1} \right|^{\frac{1}{2}} \\
& \times \exp\left\{ \frac{-\alpha_0 + \mu - 1 + \alpha_1 + c_{\hat{V}} + c_Y}{1-\alpha_1} \right\} \\
={}& (2\pi)^{-\frac{n}{2}} \left| \mathbf{\Sigma}_{\hat{X}^*} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{\Sigma}_{\hat{X}^*}^{-1} \mathbf{x} \right\},
\end{aligned}
\tag{15}
$$

where

$$
\begin{aligned}
& \mathbf{\Phi} = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1n} \\ \vdots & \ddots & \vdots \\ \phi_{n1} & \cdots & \phi_{nn} \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} \end{bmatrix}, \\
& \mathbf{\Xi} = \begin{bmatrix} \xi_1\xi_1 & \cdots & \xi_1\xi_n \\ \vdots & \ddots & \vdots \\ \xi_n\xi_1 & \cdots & \xi_n\xi_n \end{bmatrix}, \\
& \mathbf{x} = [x_1, \cdots, x_n]^T, \\
& \mathbf{y} = [y_1, \cdots, y_n]^T, \\
& \theta \geq 0.
\end{aligned}
\tag{16}
$$

By solving the equations in (13-15), the following expressions for the Lagrange multipliers are obtained:

$$
\begin{aligned}
\alpha_0 &= \mu - (1-\alpha_1) + c_{\hat{V}} + c_Y + \frac{1-\alpha_1}{2} \log (2\pi)^n \left| \mathbf{\Sigma}_{\hat{X}^*} \right| \\
&= \mu - (1-\alpha_1) + \frac{\mu(\mu-1)}{2} \log (2\pi)^n \left| \mathbf{\Sigma}_{\hat{V}} \right| \\
&\quad - \frac{\mu}{2} \log (2\pi)^n \left| \mathbf{\Sigma}_{Y^*} \right| + \frac{1-\alpha_1}{2} \log (2\pi)^n \left| \mathbf{\Sigma}_{\hat{X}^*} \right|, \\
\mathbf{\Gamma} &= -\frac{\mu(\mu-1)}{2} \mathbf{\Sigma}_{\hat{V}}^{-1}, \\
\mathbf{\Phi} &= -\mathbf{\Gamma} - \frac{\mu}{2} \mathbf{\Sigma}_{Y^*}^{-1} \\
&= \frac{\mu(\mu-1)}{2} \mathbf{\Sigma}_{\hat{V}}^{-1} - \frac{\mu}{2} \left(\mathbf{\Sigma}_{\hat{X}^*} + \mathbf{\Sigma}_{\hat{V}} \right)^{-1},
\end{aligned}
$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{\hat{X}^*} &= -\frac{1-\alpha_1}{2}\left(\boldsymbol{\Gamma}-\theta\boldsymbol{\Xi}\right)^{-1} \\
&= \frac{1-\alpha_1}{2}\left(\frac{\mu\left(\mu-1\right)}{2}\boldsymbol{\Sigma}_{\hat{V}}^{-1}+\theta\boldsymbol{\Xi}\right)^{-1} && (17)\\
&\succeq \mathbf{0}, && (18)\\
\theta &\geq 0,\\
\alpha_1 &\leq 1-\mu, && (19)\\
c_{\hat{V}} &= \frac{\mu\left(\mu-1\right)}{2}\log\left(2\pi\right)^n\left|\boldsymbol{\Sigma}_{\hat{V}}\right|,\\
c_Y &= -\frac{\mu}{2}\log\left(2\pi\right)^n\left|\boldsymbol{\Sigma}_{Y^*}\right|,\\
\left|\boldsymbol{\Sigma}_{X^*}\right| &= \left(\frac{1}{2\pi e}\exp\left\{\frac{2}{n}p_X\right\}\right)^n \leq \left|\boldsymbol{\Sigma}\right|. && (20)
\end{aligned}$$

To make the second variation negative, the negative-definiteness of the following matrix is required:

$$\left[\begin{array}{cc} K''_{f_{\hat{X}^*}f_{\hat{X}^*}} & K''_{f_{\hat{X}^*}f_{Y^*}} \\ K''_{f_{Y^*}f_{\hat{X}^*}} & K''_{f_{Y^*}f_{Y^*}} \end{array}\right], \tag{21}$$

where $K''_{f_{\hat{X}^*}f_{\hat{X}^*}}$ and $K''_{f_{Y^*}f_{Y^*}}$ stand for the second-order partial derivatives with respect to $f_{\hat{X}^*}$ and f_{Y^*}, respectively, and $K''_{f_{\hat{X}^*}f_{Y^*}}$ denotes the second-order partial derivative with respect to $f_{\hat{X}^*}$ and f_{Y^*}. Thus, the following condition is required to hold:

$$\begin{aligned}
&\left[\begin{array}{cc} h_{\hat{X}} & h_Y \end{array}\right]\left[\begin{array}{cc} K''_{f_{\hat{X}^*}f_{\hat{X}^*}} & K''_{f_{\hat{X}^*}f_{Y^*}} \\ K''_{f_{Y^*}f_{\hat{X}^*}} & K''_{f_{Y^*}f_{Y^*}} \end{array}\right]\left[\begin{array}{c} h_{\hat{X}} \\ h_Y \end{array}\right]\\
&= K''_{f_{\hat{X}^*}f_{\hat{X}^*}}h_{\hat{X}}^2 + K''_{f_{Y^*}f_{Y^*}}h_Y^2 + (K''_{f_{\hat{X}^*}f_{Y^*}} + K''_{f_{Y^*}f_{\hat{X}^*}})h_Y h_{\hat{X}}\\
&\leq 0, && (22)
\end{aligned}$$

where $h_{\hat{X}}$ and h_Y are arbitrary admissible functions.

Since $K''_{f_{\hat{X}^*}f_{\hat{X}^*}}$, $K''_{f_{\hat{X}^*}f_{Y^*}}$, $K''_{f_{Y^*}f_{\hat{X}^*}}$, and $K''_{f_{Y^*}f_{Y^*}}$ are defined as

$$\begin{aligned}
K''_{f_{\hat{X}^*}f_{\hat{X}^*}} &= -\frac{(1-\alpha_1)f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{\hat{X}^*}(\mathbf{x})},\\
K''_{f_{\hat{X}^*}f_{Y^*}} &= \frac{\mu f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})},\\
K''_{f_{Y^*}f_{\hat{X}^*}} &= \frac{\mu f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})},\\
K''_{f_{Y^*}f_{Y^*}} &= -\frac{\mu f_{\hat{X}^*}(\mathbf{x})f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})^2}, && (23)
\end{aligned}$$

the condition in (22) requires

$$\begin{aligned}
&-\frac{(1-\alpha_1)f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{\hat{X}^*}(\mathbf{x})}h_{\hat{X}}(\mathbf{x})^2+2\frac{\mu f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})}h_{\hat{X}}(\mathbf{x})h_Y(\mathbf{y})\\
&-\frac{\mu f_{\hat{X}^*}(\mathbf{x})f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{Y^*}(\mathbf{y})^2}h_Y(\mathbf{y})^2\\
&\leq -\frac{\mu f_{\hat{V}}(\mathbf{y}-\mathbf{x})}{f_{\hat{X}^*}(\mathbf{x})}\left(h_{\hat{X}}(\mathbf{x})-\frac{f_{\hat{X}^*}(\mathbf{x})}{f_{Y^*}(\mathbf{y})}h_Y(\mathbf{y})\right)^2\\
&\leq 0, && (24)
\end{aligned}$$

which it is satisfied since $\alpha_1 \geq 1-\mu$.

Since all the Lagrange multipliers exist, the necessarily optimal solutions $f_{\hat{X}^*}$ and f_{Y^*} maximize the functional problem in (3), and the proof is completed.

IV. Conclusion

In this paper, we proposed a novel functional approach for establishing the EEI based on calculus of variations. The proposed novel technique provides a simpler alternative method for proving the EEI without using the EPI or the channel enhancement technique. In addition, this method can be easily adopted to prove and extend other fundamental information theoretic results such as maximization of differential entropy, minimization of Fisher information (Cramér-Rao inequality), and worst additive noise lemma. Additional results in this regard are provided in the extended journal paper version [5] of this work.

Appendix

Lemma 1: There always exists a positive semi-definite matrix $\mathbf{K}$ which satisfies

$$\begin{aligned}
\boldsymbol{\Sigma}_{\tilde{W}} &\preceq \left(\mu-1\right)^{-1}\boldsymbol{\Sigma}_{\tilde{V}}, && (25)\\
\mathbf{K}\boldsymbol{\Sigma}_{X^*} &= \boldsymbol{\Sigma}_{X^*}\mathbf{K}=\mathbf{0}, && (26)
\end{aligned}$$

where $\boldsymbol{\Sigma}_{X^*} = (\mu-1)^{-1}\boldsymbol{\Sigma}_{\tilde{V}} - \boldsymbol{\Sigma}_{\tilde{W}}$, and $\boldsymbol{\Sigma}_{\tilde{W}} = \left(\boldsymbol{\Sigma}_W^{-1}+\mathbf{K}\right)^{-1}$.

Proof: Proving $\boldsymbol{\Sigma}_{\tilde{W}} \preceq \left(\mu-1\right)^{-1}\boldsymbol{\Sigma}_{\tilde{V}}$ is equivalent to showing the following relation:

$$\begin{aligned}
&\boldsymbol{\Sigma}_{\tilde{W}} \preceq \left(\mu-1\right)^{-1}\boldsymbol{\Sigma}_{\tilde{V}} && (27)\\
\iff\ &\boldsymbol{\Sigma}_W^{-1}+\mathbf{K} \succeq \left(\mu-1\right)\boldsymbol{\Sigma}_{\tilde{V}}^{-1}. && (28)
\end{aligned}$$

Since there always exists a non-singular matrix which simultaneously diagonalizes two positive semi-definite matrices [10], there exists a non-singular matrix $\mathbf{Q}$ which simultaneously diagonalizes both $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_{\tilde{V}}$:

$$\begin{aligned}
\mathbf{Q}^T\boldsymbol{\Sigma}_W\mathbf{Q} &= \mathbf{D}_W, && (29)\\
\mathbf{Q}^T\boldsymbol{\Sigma}_{\tilde{V}}\mathbf{Q} &= \mathbf{I}, && (30)
\end{aligned}$$

where $\mathbf{I}$ stands for the identity matrix, and $\mathbf{D}_W$ is a diagonal matrix. Since $\mathbf{Q}$ is a non-singular matrix, the inverse of $\mathbf{Q}$ always exists, and $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_{\tilde{V}}$ can be expressed as:

$$\begin{aligned}
\boldsymbol{\Sigma}_W &= \mathbf{Q}^{-T}\mathbf{D}_W\mathbf{Q}^{-1}, && (31)\\
\boldsymbol{\Sigma}_{\tilde{V}} &= \mathbf{Q}^{-T}\mathbf{Q}^{-1}. && (32)
\end{aligned}$$

Now denote by $\mathbf{D}_K$ a diagonal matrix whose i^{th} diagonal element is represented as d_{K_i}, and which it is defined as

$$d_{K_i} = \left\{\begin{array}{ll} 0 & \text{if } d_{W_i} \leq \left(\mu-1\right)^{-1} \\ \mu-1-\frac{1}{d_{W_i}} & \text{if } d_{W_i} > \left(\mu-1\right)^{-1} \end{array}\right. \tag{33}$$

where d_{W_i} represents the i^{th} diagonal element of $\mathbf{D}_W$. Define also the matrix $\mathbf{K}$ via

$$\mathbf{K} = \mathbf{Q}\mathbf{D}_K\mathbf{Q}^T. \tag{34}$$

Then the inequality (28) is equivalent to

$$\begin{aligned}
&\boldsymbol{\Sigma}_W^{-1}+\mathbf{K} \succeq \left(\mu-1\right)\boldsymbol{\Sigma}_{\tilde{V}}^{-1} && (35)\\
\Leftrightarrow &\left(\mathbf{Q}^{-T}\mathbf{D}_W\mathbf{Q}^{-1}\right)^{-1}+\mathbf{Q}\mathbf{D}_K\mathbf{Q}^T \succeq (\mu-1)\left(\mathbf{Q}^{-T}\mathbf{Q}^{-1}\right)^{-1} && (36)\\
\Leftrightarrow &\mathbf{D}_W^{-1}+\mathbf{D}_K \succeq \left(\mu-1\right)\mathbf{I}. && (37)
\end{aligned}$$

The inequality (37) always holds since $\mathbf{D}_K$ is defined as in (33). Therefore, the inequality (27) also holds.

We know that $\boldsymbol{\Sigma}_{X^*}$ is equal to $(\mu-1)^{-1}\boldsymbol{\Sigma}_{\tilde{V}} - \boldsymbol{\Sigma}_{\tilde{W}}$. Therefore,

$$\boldsymbol{\Sigma}_{X^*}\mathbf{K} \quad = \quad (\mu-1)^{-1}\left(\boldsymbol{\Sigma}_{\tilde{V}} - (\mu-1)\,\boldsymbol{\Sigma}_{\tilde{W}}\right)\mathbf{K}, \quad (38)$$

and the equation (38) can be re-expressed as

$$\begin{aligned}
&\boldsymbol{\Sigma}_{X^*}\mathbf{K}\\
&=(\mu-1)^{-1}\left(\boldsymbol{\Sigma}_{\tilde{V}} - (\mu-1)\,\boldsymbol{\Sigma}_{\tilde{W}}\right)\mathbf{K} && (39)\\
&=(\mu-1)^{-1}\Big(\mathbf{Q}^{-T}\mathbf{Q}^{-1} - (\mu-1)\Big(\left(\mathbf{Q}^{-T}\mathbf{D}_W\mathbf{Q}^{-1}\right)^{-1}\\
&\quad +\mathbf{Q}\mathbf{D}_K\mathbf{Q}^T\Big)^{-1}\Big)\mathbf{Q}\mathbf{D}_K\mathbf{Q}^T && (40)\\
&=(\mu-1)^{-1}\mathbf{Q}^{-T}\Big(\mathbf{I} - (\mu-1)\left(\mathbf{D}_W^{-1}+\mathbf{D}_K\right)^{-1}\Big)\mathbf{D}_K\mathbf{Q}^T && (41)\\
&=\mathbf{0}. && (42)
\end{aligned}$$

The equality (40) is due to relations (31), (32), and (34), and the equality (42) is due to (33). Similarly,

$$\begin{aligned}
&\mathbf{K}\boldsymbol{\Sigma}_{X^*}\\
&=(\mu-1)^{-1}\mathbf{K}\left(\boldsymbol{\Sigma}_{\tilde{V}} - (\mu-1)\,\boldsymbol{\Sigma}_{\tilde{W}}\right)\\
&=(\mu-1)^{-1}\mathbf{Q}\mathbf{D}_K\mathbf{Q}^T\Big(\mathbf{Q}^{-T}\mathbf{Q}^{-1}\\
&\qquad -(\mu-1)\Big(\left(\mathbf{Q}^{-T}\mathbf{D}_W\mathbf{Q}^{-1}\right)^{-1}+\mathbf{Q}\mathbf{D}_K\mathbf{Q}^T\Big)^{-1}\Big)\\
&=(\mu-1)^{-1}\mathbf{Q}\mathbf{D}_K\Big(\mathbf{I} - (\mu-1)\left(\mathbf{D}_W^{-1}+\mathbf{D}_K\right)^{-1}\Big)\mathbf{Q}^{-1}\\
&=\mathbf{0}.
\end{aligned}$$

Therefore, by defining $\boldsymbol{\Sigma}_{\tilde{W}} = (\boldsymbol{\Sigma}_W^{-1}+\mathbf{K})^{-1}$, we can make $\boldsymbol{\Sigma}_{\tilde{W}}$ satisfy

$$\boldsymbol{\Sigma}_{\tilde{W}} \preceq (\mu-1)^{-1}\boldsymbol{\Sigma}_{\tilde{V}}, \qquad \boldsymbol{\Sigma}_{X^*}\mathbf{K} = \mathbf{K}\boldsymbol{\Sigma}_{X^*} = \mathbf{0}, \qquad (43)$$

and the proof is completed. ∎

Lemma 2 (Data Processing Inequality [11]): When three random vectors Y_1, Y_2, and Y_3 represent a Markov chain $Y_1 \to Y_2 \to Y_3$, the following inequality is satisfied:

$$I(Y_1;Y_3) \le I(Y_1;Y_2). \qquad (44)$$

The equality holds if and only if $I(Y_1;Y_2|Y_3)=0$.

In Lemma 2, Y_1, Y_2, and Y_3 are defined as X_G^*, $X_G^*+\tilde{W}_G$, and $X_G^*+\tilde{W}_G+\hat{W}_G$, respectively. Therefore, the equality condition, $I(Y_1;Y_2|Y_3)=0$ is expressed as

$$\begin{aligned}
&I(Y_1;Y_2|Y_3)\\
&= h(Y_1|Y_3) - h(Y_1|Y_2,Y_3)\\
&=\frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{Y_1|Y_3}\right| - \frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{Y_1|Y_2}\right|\\
&=\frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{Y_1} - \boldsymbol{\Sigma}_{Y_1}\boldsymbol{\Sigma}_{Y_3}^{-1}\boldsymbol{\Sigma}_{Y_1}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{Y_1} - \boldsymbol{\Sigma}_{Y_1}\boldsymbol{\Sigma}_{Y_2}^{-1}\boldsymbol{\Sigma}_{Y_1}\right|\\
&=\frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{X^*} - \boldsymbol{\Sigma}_{X^*}\left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}+\boldsymbol{\Sigma}_{\hat{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{X^*} - \boldsymbol{\Sigma}_{X^*}\left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&=\frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{X^*}\right|\left|I - \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}+\boldsymbol{\Sigma}_{\hat{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|\boldsymbol{\Sigma}_{X^*}\right|\left|I - \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&=\frac{1}{2}\log(2\pi e)^n\left|I - \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}+\boldsymbol{\Sigma}_{\hat{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|I - \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&=\frac{1}{2}\log(2\pi e)^n\left|I - \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{W}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&\quad -\frac{1}{2}\log(2\pi e)^n\left|I - \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}\right|\\
&=0. && (45)
\end{aligned}$$

If $\left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{W}\right)^{-1}\boldsymbol{\Sigma}_{X^*} = \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}$, the equality in (45) is satisfied, the equality condition in Lemma 2 holds, and therefore, the equality in (6) is proved. The validity of $\left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{W}\right)^{-1}\boldsymbol{\Sigma}_{X^*} = \left(\boldsymbol{\Sigma}_{X^*}+\boldsymbol{\Sigma}_{\tilde{W}}\right)^{-1}\boldsymbol{\Sigma}_{X^*}$ is proved by Lemma 1. Therefore, $I(Y_1;Y_2|Y_3)=0$

ACKNOWLEDGMENT

This paper was made possible by QNRF-NPRP grant 4-1293-2-513, and in part by NSF grant 0915444 and QF-QSLP.

REFERENCES

[1] T. Liu and P. Viswanath, "An Extremal Inequality Motivated by Multiterminal Information-Theoretic Problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839 - 1851, May 2007.

[2] S. Park, E. Serpedin, and K. Qaraqe, "An Alternative Proof of an Extremal Inequality," *IEEE Trans. Inf. Theory*, (submitted for publication) arXiv:1201.6681, Nov. 2012.

[3] T. Liu and S. Shamai, "A Note on the Secrecy Capacity of the Multiple-Antenna Wiretap Channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547 - 2553, Jun 2009.

[4] R. Liu, T. Liu, H. Poor, and S. Shamai, "A vector generalization of Costas entropy-power inequality with applications," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1865-1879, Apr. 2010.

[5] S. Park, E. Serpedin and K. Qaraqe, "A Unifying Veriational Perspective on Some Fundamental Information Theoretic Inequalities," *IEEE Trans. Inf. Theory*, (submitted for publication) arxiv.org/pdf/1211.4795.pdf, Sept. 2012.

[6] Y. Geng and C. Nair, "The capacity region of the two-receiver vector Gaussian broadcast channel with private and common messages," *IEEE Trans. Inf. Theory*, arxiv.org/pdf/1202.0097, February 2012.

[7] E. Ekrem and S. Ulukus, "Capacity Region of Gaussian MIMO Broadcast Channels With Common and Confidential Messages," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5669-5680, Sep. 2012.

[8] H. Weingarten, Y. Steinberg and S. Shamai, "The Capacity Region of the Gaussian Mutiple-Input Multiple-Output Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936-3964, Sep 2006.

[9] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, New York: Dover, 1991.

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

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley-Interscience, 1991.