Fifty-Fifth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 3-6, 2017

On the Equality Condition for the I-MMSE Proof of the Entropy Power Inequality

Alex Dytso*, Ronit Bustin**, H. Vincent Poor*, and Shlomo Shamai (Shitz)**

Abstract—**The paper establishes the equality condition in the I-MMSE proof of the entropy power inequality (EPI). This is done by establishing an exact expression for the deficit between the two sides of the EPI. Interestingly, a necessary condition for the equality is established by making a connection to the famous Cauchy functional equation.**

I. Introduction

The classical entropy power inequality (EPI) formulated by Shannon in [1] states that for two independent continuous random vectors $\mathbf{V}$ and $\mathbf{W}$

$$e^{\frac{2}{n}h(\mathbf{V}+\mathbf{W})} \geq e^{\frac{2}{n}h(\mathbf{V})} + e^{\frac{2}{n}h(\mathbf{W})}, \quad (1)$$

where equality in (1) is attained if and only if $\mathbf{V}$ and $\mathbf{W}$ are Gaussian with proportional covariances (i.e., $\mathbf{K}_W = c\mathbf{K}_V$ for some scalar $c > 0$). Via the transformation

$$\mathbf{X}_1 = \frac{\mathbf{V}}{\sqrt{1-\alpha}}, \quad \mathbf{X}_2 = \frac{\mathbf{W}}{\sqrt{\alpha}},$$
$$\alpha = \frac{e^{\frac{2}{n}h(\mathbf{W})}}{e^{\frac{2}{n}h(\mathbf{V})} + e^{\frac{2}{n}h(\mathbf{W})}},$$

the EPI can be shown to be equivalent to Lieb's inequality [2]

$$h\left(\sqrt{1-\alpha}\mathbf{X}_1 + \sqrt{\alpha}\mathbf{X}_2\right) \geq (1-\alpha)h(\mathbf{X}_1) + \alpha h(\mathbf{X}_2), \quad \forall\, \alpha \in [0,1], \quad (2)$$

and where equality in (2) holds if and only if $\mathbf{K}_{\mathbf{X}_1} = \mathbf{K}_{\mathbf{X}_2}$.

There are several proofs of the EPI which follow three distinct methods: using integration over a path of a continuous Gaussian perturbation [3]–[7]; using the sharp version of Young's inequality and properties of Rényi entropy [2], [4], [5]; and mass transport proof using the Knöthe map [8], [9]. For a comprehensive list of references and a detailed history of the EPI, the reader is referred to [10] and [11], and references therein.

As was recently pointed out in [8] not all available proofs settle the *equality* case in (1) and (2). In particular, for the class of proofs via Gaussian perturbations, the case of equality has not yet been established in the proof given in [6], which relies on the so-called I-MMSE relationship [12].

The goal of this paper is to close this gap by establishing the equality case in the proof of the EPI via the I-MMSE relationship. Equality is established by determining an exact expression for the deficit in (2) and showing that the deficit is zero if and only if $\mathbf{X}_1$ and $\mathbf{X}_2$ are Gaussian with identical covariances.

Notation: Deterministic scalar/vector quantities are denoted by lowercase normal/bold letters, matrices by bold uppercase letters, random variables by uppercase letters, and random vectors by bold uppercase letters. For a random vector $\mathbf{V}$ we denote the covariance matrix by $\mathbf{K}_\mathbf{V}$, determinant by $\det(\mathbf{K}_\mathbf{V})$, transpose by $\mathbf{V}^T$, and trace by $\mathrm{Tr}\{\mathbf{V}\}$. The Euclidian norm of a vector $\mathbf{v}$ is denoted by $\|\mathbf{v}\|$. The gradient operator is denoted by ∇. The $\mathbb{E}[\cdot]$ denotes the expectation operator.

Assumptions: Throughout the paper, we assume that all random vectors treated in this work have covariance matrices with bounded entries and continuous, positive, and differentiable probability densities. Therefore, quantities such as entropies, expectations, and conditional expectations are well defined throughout the paper. The interested reader is referred to [8], [13] and [14] where it is shown that the set of aforementioned assumptions is sufficient to prove the EPI in (1).

II. Preliminary Results

In this section, we present necessary mathematical tools needed in this paper.

We define the minimum mean square error (MMSE) of estimating $\mathbf{X} \in \mathbb{R}^n$ from $\mathbf{Y} \in \mathbb{R}^k$ as

$$\mathrm{mmse}(\mathbf{X} \mid \mathbf{Y}) = \mathbb{E}\left[\|\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}]\|^2\right]. \quad (3)$$

The first result of this section establishes the penalty, incurred in the MMSE, for using a sub-optimal estimator.

Lemma 1. *Let $f : \mathbb{R}^n \to \mathbb{R}^k$ be measurable and such that $\mathbb{E}\left[\|f(\mathbf{Y})\|^2\right] \leq \infty$. Then,*

$$\mathrm{mmse}(\mathbf{X} \mid \mathbf{Y}) = \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^2\right] - \mathbb{E}\left[\|f(\mathbf{Y}) - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}]\|^2\right]. \quad (4)$$

*A. Dytso and H.V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA (email: adytso, poor@princeton.edu).

**R. Bustin and S. Shamai (Shitz) are with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: bustin@technion.ac.il, sshlomo@ee.technion.ac.il).

The work of A. Dytso and H.V. Poor was supported by the National Science Foundation under Grants CCF-1420575 and CNS-1456793. The work of S. Shamai and R.Bustin was supported by the Unions Horizon 2020 Research and Innovation Programme Grant 694630. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

Proof:

$$\begin{aligned}
&\mathbb{E}\left[\|f(\mathbf{Y})-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]\|^2\right]\\
&=\mathbb{E}\left[\|f(\mathbf{Y})-\mathbf{X}+\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]\|^2\right]\\
&=\mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^2\right]+\mathbb{E}\left[\|\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]\|^2\right]\\
&+2\mathbb{E}\left[\mathrm{Tr}\{(f(\mathbf{Y})-\mathbf{X})(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}])^T\}\right]\\
&=\mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^2\right]+\mathbb{E}\left[(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}])^2\right]\\
&-2\mathbb{E}\left[\mathrm{Tr}\{\mathbf{X}(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}])^T\}\right] \quad (5a)\\
&=\mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^2\right]+\mathbb{E}\left[\|\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]\|^2\right]\\
&-2\mathbb{E}\left[\|\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]\|^2\right] \quad (5b)\\
&=\mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^2\right]-\mathbb{E}\left[\|\mathbf{X}-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]\|^2\right],
\end{aligned}$$

where (5a) and (5b) are due to the orthogonality principle. This concludes the proof. ■

The necessary condition for the equality in (2) will be shown to be a consequence of a remarkably simple, yet powerful, Cauchy functional equation.

Lemma 2. *(*Cauchy Functional Equation.*) Over a space of measurable*[1] *functions from* $\mathbb{R}^n$ *to* $\mathbb{R}^n$ *the equation*

$$f(\mathbf{x}+\mathbf{y})=f(\mathbf{x})+f(\mathbf{y}), \quad (6)$$

is satisfied if and only if $f(\mathbf{x})=\mathbf{A}\,\mathbf{x}$ *(i.e., is linear) for some matrix* $\mathbf{A}\in\mathbb{R}^{n\times n}$.

Proof: See [15, Chapter 2]. ■

Cauchy functional equation has a very rich history, and the interested reader is referred to [15] for a comprehensive summary. Cauchy functional equation is used next to establish the following property of the conditional expectation.

Lemma 3. *Let* $\mathbf{V},\mathbf{W}\in\mathbb{R}^n$ *be independent random vectors with full support and* $f_1, f_2:\mathbb{R}^n\to\mathbb{R}^n$ *be measurable functions such that* $\mathbb{E}\left[\|f_1(\mathbf{V})\|^2\right], \mathbb{E}\left[\|f_2(\mathbf{W})\|^2\right]<\infty$. *Then, for any* $a_1,a_2\in\mathbb{R}$

$$\begin{aligned}
&\mathbb{E}\left[a_1 f_1(\mathbf{V})+a_2 f_2(\mathbf{W})\mid a_1\mathbf{V}+a_2\mathbf{W}\right]\\
&=a_1 f_1(\mathbf{V})+a_2 f_2(\mathbf{W}) \ \textit{a.s.}, \quad (7a)
\end{aligned}$$

if and only if $f_1(\mathbf{v})$ *and* $f_2(\mathbf{w})$ *are affine functions with the same slope, that is*

$$f_1(\mathbf{v})=\mathbf{A}\,\mathbf{v}+\mathbf{b},\ f_2(\mathbf{w})=\mathbf{A}\,\mathbf{w}+\mathbf{c}, \textit{a.s.}, \quad (7b)$$

for some $\mathbf{A}\in\mathbb{R}^{n\times n}$ *and* $\mathbf{b},\mathbf{c}\in\mathbb{R}^n$.

Proof: The proof of the sufficient condition follows trivially. To show the necessary condition observe that (7a) is equivalent to identifying a set of functions $\{h(\cdot)\}$ for which

$$h(a_1\mathbf{W}+a_2\mathbf{V})=a_1 f_1(\mathbf{V})+a_2 f_2(\mathbf{W}). \quad (8)$$

Since $\mathbf{V}$ and $\mathbf{W}$ are fully supported, we have that

$$h(a_1\mathbf{v}+a_2\mathbf{w})=a_1 f_1(\mathbf{v})+a_2 f_2(\mathbf{w}), \text{a.s. } (\mathbf{v},\mathbf{w}). \quad (9)$$

[1] In this paper, measurable is meant with respect to the Lebesgue measure.

In particular, since (9) holds a.s., there exist some fix $\mathbf{y}_1,\mathbf{y}_2$ such that

$$h(a_1\mathbf{v}+a_1\mathbf{y}_1+a_2\mathbf{y}_2)=a_1 f_1(\mathbf{v}+\mathbf{y}_1)+a_2 f_2(\mathbf{y}_2), \text{a.s. } \mathbf{v}, \quad (10a)$$

$$h(a_1\mathbf{y}_1+a_2\mathbf{w}+a_2\mathbf{y}_2)=a_1 f_1(\mathbf{y}_1)+a_2 f_2(\mathbf{w}+\mathbf{y}_2), \text{a.s. } \mathbf{w}. \quad (10b)$$

Therefore, by adding the two equations in (10) and using (9), we arrive at

$$\begin{aligned}
&h(a_1\mathbf{v}+a_1\mathbf{y}_1+a_2\mathbf{y}_2)+h(a_1\mathbf{y}_1+a_2\mathbf{w}+a_2\mathbf{y}_2) \quad (11)\\
&=a_1 f_1(\mathbf{v}+\mathbf{y}_1)+a_2 f_2(\mathbf{y}_2)+a_1 f_1(\mathbf{y}_1)+a_2 f_2(\mathbf{w}+\mathbf{y}_2)\\
&=h(a_1\mathbf{v}+a_2\mathbf{w}+a_1\mathbf{y}_1+a_2\mathbf{y}_2)+a_2 f_2(\mathbf{y}_2)\\
&+a_1 f_1(\mathbf{y}_1), \ \text{a.s. } (\mathbf{v},\mathbf{w}). \quad (12)
\end{aligned}$$

Next, by letting $f(\mathbf{x})=h(\mathbf{x}+a_1\mathbf{y}_1+a_2\mathbf{y}_2)-a_2 f_2(\mathbf{y}_2)-a_1 f_1(\mathbf{y}_1)$, and using (12)

$$f(a_1\mathbf{v})+f(a_2\mathbf{w})=f(a_1\mathbf{v}+a_2\mathbf{w}) \text{ a.s. } (\mathbf{v},\mathbf{w}). \quad (13)$$

Now, (13) corresponds to Cauchy functional equation in Lemma 2 and $f(\cdot)$ is a linear function. As a result, we concluded that $h(\cdot)$ is an affine function

$$h(\mathbf{x})=\mathbf{A}\mathbf{x}+\mathbf{a}, \quad (14)$$

for some $\mathbf{A}\in\mathbb{R}^{n\times n}$ and $\mathbf{a}\in\mathbb{R}^n$. Finally, (14) and (10) imply that functions $f_1(\cdot)$ and $f_2(\cdot)$ are also affine with the same slope. This concludes the proof. ■

The following well-known property of the conditional expectation will be useful in manipulating some of our expressions.

Lemma 4. *(*Smoothing or Towering Property of the Conditional Expectation.*) Let sigma algebras* $\mathcal{G}_1,\mathcal{G}_2$ *be such that* $\mathcal{G}_1\subset\mathcal{G}_2$. *Then,*

$$\mathbb{E}\left[\mathbb{E}[\mathbf{X}\mid\mathcal{G}_2]\mid\mathcal{G}_1\right]=\mathbb{E}[\mathbf{X}\mid\mathcal{G}_1] \ \textit{a.s.} \quad (15a)$$

In particular, for $(\mathbf{X},\mathbf{W},\mathbf{V})$

$$\mathbb{E}\left[\mathbb{E}[\mathbf{X}\mid\mathbf{W},\mathbf{V}]\mid\mathbf{W}+\mathbf{V}\right]=\mathbb{E}[\mathbf{X}\mid\mathbf{W}+\mathbf{V}] \ \textit{a.s.} \quad (15b)$$

Proof: See [16, Chapter 10]. ■

The key step of the proof would be to establish that the equality holds if and only if a certain conditional expectation is a linear or an affine function. The following result shows that the conditional expectation is an affine function if and only if the input random variable is Gaussian.

Lemma 5. *Let* $\mathbf{Y}=\mathbf{X}+\mathbf{Z}$ *where* $\mathbf{X}$ *and* $\mathbf{Z}\sim\mathcal{N}(0,\mathbf{I})$ *are independent. Then,*

$$\mathbb{E}[\mathbf{X}\mid\mathbf{Y}]=\mathbf{A}\mathbf{Y}+\mathbf{v} \ \textit{a.s.}, \quad (16a)$$

if and only if $\mathbf{X}\sim\mathcal{N}(\mu_{\mathbf{X}},\mathbf{K}_{\mathbf{X}})$ *such that*

$$\mathbf{A}=\mathbf{K}_{\mathbf{XY}}\mathbf{K}_{\mathbf{Y}}^{-1}=\mathbf{K}_{\mathbf{X}}(\mathbf{I}+\mathbf{K}_{\mathbf{X}})^{-1},\ \mathbf{v}=(\mathbf{I}-\mathbf{A})\mu_{\mathbf{X}}. \quad (16b)$$

Proof: Lemma 5 is a well known result from estimation theory and the details of the proof can be found in [17].

Here, we only give a sketch of the proof of the necessary condition. To show the necessary condition, one must show that, in the MMSE sense, linear estimators are only optimal for Gaussian random vectors. For simplicity, we only look at the zero mean (i.e., $\mu_X = \mathbb{E}[X] = 0$ and $v = 0$) and the scalar case. Let $f(Y) = \mathrm{e}^{-itY}$, and let aY be an estimator that we claim to be optimal with $a = \frac{E[XY]}{E[Y^2]}$. Then, by the orthogonality principle we have that

$$\begin{aligned}
0 &= \mathbb{E}\left[(X - aY)\,\mathrm{e}^{-itY}\right] \\
&= \mathbb{E}\left[((1-a)X - aZ)\,\mathrm{e}^{-itY}\right] \\
&= (1-a)\mathbb{E}\left[X\mathrm{e}^{-itY}\right] - a\mathbb{E}\left[Z\mathrm{e}^{-itY}\right] \\
&= (1-a)\mathbb{E}\left[X\mathrm{e}^{-itX}\right]\mathbb{E}\left[\mathrm{e}^{-itZ}\right] - a\mathbb{E}\left[Z\mathrm{e}^{-itZ}\right]\mathbb{E}\left[\mathrm{e}^{-itX}\right] && \text{(17a)} \\
&= (1-a)i\phi_X^{'}(t)\phi_Z(t) - ia\phi_Z^{'}(t)\phi_X(t) && \text{(17b)} \\
&= (1-a)\,i\,\phi_X^{'}(t)\,\mathrm{e}^{-\frac{t^2}{2}} - ia\,t\,\mathrm{e}^{-\frac{t^2}{2}}\phi_X(t), && \text{(17c)}
\end{aligned}$$

where (17a) follows by the independence of X and Z, and (17b) follows by the derivative expression $\phi_X^{'}(t) = -i\,\mathbb{E}\left[X\mathrm{e}^{-itX}\right]$ (the derivative expression holds since by assumption $\mathbb{E}[X^2] < \infty$).

Therefore, from (17c) we have a differential equation of the form

$$(1-a)\,\phi_X^{'}(t) = at\,\phi_X(t). \tag{18}$$

The only nontrivial solution to the differential equation in (18) is given by the Gaussian distribution with the characteristic function given by $\phi_X(t) = \mathrm{e}^{-\frac{(a-1)}{a}\frac{t^2}{2}}$. This concludes the proof. ∎

We define the *score function* and the *Fisher information* of a continuous random vector $\mathbf{X}$ with the probability density function $f_{\mathbf{X}}(\mathbf{x})$ as

$$\rho_{\mathbf{X}}(\mathbf{x}) = \nabla_{\mathbf{x}} \log\left(f_{\mathbf{X}}(\mathbf{x})\right), \tag{19a}$$

$$J(\mathbf{X}) = \mathbb{E}\left[\rho_{\mathbf{X}}(\mathbf{X})^T \rho_{\mathbf{X}}(\mathbf{X})\right]. \tag{19b}$$

For the Gaussian noise channel, the score function of the output can be related to the conditional expectation.

Lemma 6. *Let $\mathbf{Y}_\gamma = \sqrt{\gamma}\,\mathbf{X} + \mathbf{Z}$ where $\mathbf{X}$ and $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I})$ are independent. Then,*

$$\rho_{\mathbf{Y}_\gamma}(\mathbf{Y}_\gamma) = \sqrt{\gamma}\,\mathbb{E}[\mathbf{X} \mid \mathbf{Y}_\gamma] - \mathbf{Y}_\gamma \text{ a.s.} \tag{20}$$

Proof: See [12, Eq.(56)]. ∎

We conclude this section by giving an expression for the differential entropy in terms of an integral of the MMSE which is a consequence of the I-MMSE relationship in [12].

Lemma 7. *For every continuous random vector $\mathbf{X} \in \mathbb{R}^n$,*

$$h(\mathbf{X}) = \frac{1}{2}\int_0^\infty \mathrm{mmse}(\mathbf{X}|\mathbf{Y}_\gamma) - \frac{n}{2\pi\mathrm{e} + \gamma} d\gamma, \tag{21a}$$

as long as

$$\lim_{t\to 0} h(\mathbf{X} + t\mathbf{Z}) = h(\mathbf{X}), \tag{21b}$$

where $\mathbf{Y}_\gamma = \sqrt{\gamma}\,\mathbf{X} + \mathbf{Z}$ and $\mathbf{X}$ is independent of $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I})$.

III. Main Results

The first main result of this section, which is a refinement of the bound in [6], establishes an exact expression for the deficit in (2).

Theorem 1. *For any independent continuous random vectors $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^n$ and any $\alpha \in [0, 1]$*

$$\begin{aligned}
h(\sqrt{1-\alpha}\mathbf{X}_1 + \sqrt{\alpha}\mathbf{X}_2) &= (1-\alpha)h(\mathbf{X}_1) + \alpha h(\mathbf{X}_2) \\
&\quad + \Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2), && \text{(22a)}
\end{aligned}$$

where

$$\begin{aligned}
&\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2) \\
&= \frac{1}{2}\int_0^\infty \mathbb{E}\left[\left\|\mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{\Sigma,\gamma}] - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{1,\gamma}, \mathbf{Y}_{2,\gamma}]\right\|^2\right] d\gamma, && \text{(22b)}
\end{aligned}$$

and where

$$\mathbf{X} = \sqrt{1-\alpha}\mathbf{X}_1 + \sqrt{\alpha}\mathbf{X}_2, \tag{22c}$$

$$\mathbf{Y}_{1,\gamma} = \sqrt{\gamma}\mathbf{X}_1 + \mathbf{Z}_1, \tag{22d}$$

$$\mathbf{Y}_{2,\gamma} = \sqrt{\gamma}\mathbf{X}_1 + \mathbf{Z}_2, \tag{22e}$$

$$\mathbf{Y}_{\Sigma,\gamma} = \sqrt{1-\alpha}\mathbf{Y}_{1,\gamma} + \sqrt{\alpha}\mathbf{Y}_{2,\gamma}, \tag{22f}$$

where $\mathbf{Z}_1 \sim \mathcal{N}(0, \mathbf{I})$, $\mathbf{Z}_2 \sim \mathcal{N}(0, \mathbf{I})$ and where $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Z}_1, \mathbf{Z}_2)$ are mutually independent.

Proof: According to Lemma 7, the entropy of a random vector $\mathbf{X} \in \mathbb{R}^n$, defined in (22c), is given by

$$\begin{aligned}
&h(\mathbf{X}) \\
&= \frac{1}{2}\int_0^\infty \mathbb{E}[\|\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{\Sigma,\gamma}]\|^2] - \frac{n}{2\pi\mathrm{e} + \gamma} d\gamma \\
&= \frac{1}{2}\int_0^\infty \mathrm{mmse}(\mathbf{X} \mid \mathbf{Y}_{1,\gamma}, \mathbf{Y}_{2,\gamma}) \\
&+ \mathbb{E}\left[\|\mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{\Sigma,\gamma}] - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{1,\gamma}, \mathbf{Y}_{2,\gamma}]\|^2\right] - \frac{n}{2\pi\mathrm{e} + \gamma} d\gamma, && \text{(23)}
\end{aligned}$$

where the last step follows by taking

$$\begin{aligned}
f(\mathbf{Y}_{1,\gamma}, \mathbf{Y}_{2,\gamma}) &= \mathbb{E}\left[\mathbf{X} \mid \sqrt{1-\alpha}\mathbf{Y}_{1,\gamma} + \sqrt{\alpha}\mathbf{Y}_{2,\gamma}\right] \\
&= \mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{\Sigma,\gamma}], && \text{(24)}
\end{aligned}$$

in Lemma 1.

Next, by the mutual independence of $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Z}_1, \mathbf{Z}_2)$, the first expectation in (23) reduces to

$$\begin{aligned}
&\mathrm{mmse}(\mathbf{X} \mid \mathbf{Y}_{1,\gamma}, \mathbf{Y}_{2,\gamma}) \\
&= (1-\alpha)\,\mathrm{mmse}(\mathbf{X}_1 \mid \mathbf{Y}_{1,\gamma}) + \alpha\,\mathrm{mmse}(\mathbf{X}_2 \mid \mathbf{Y}_{2,\gamma}). && \text{(25)}
\end{aligned}$$

Finally, by combining (23) and (25) we arrive at

$$\begin{aligned}
&h(\mathbf{X}) \\
&= \frac{1}{2}\int_0^\infty (1-\alpha)\,\mathrm{mmse}(\mathbf{X}_1 \mid \mathbf{Y}_{1,\gamma}) + \alpha\,\mathrm{mmse}(\mathbf{X}_2 \mid \mathbf{Y}_{2,\gamma}) \\
&+ \mathbb{E}\left[\|\mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{\Sigma,\gamma}] - \mathbb{E}[\mathbf{X} \mid \mathbf{Y}_{1,\gamma}, \mathbf{Y}_{2,\gamma}]\|^2\right] - \frac{n}{2\pi\mathrm{e} + \gamma} d\gamma \\
&= (1-\alpha)\,h(\mathbf{X}_1) + \alpha\,h(\mathbf{X}_2) + \Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2),
\end{aligned}$$

where $\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2)$ is defined in (22b). This concludes the proof. ∎

Clearly, $\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2)$ in (22b) is a non-negative quantity which leads to Lieb's inequality in (2).

We also would like to point out that another exact expression for the deficit (2) can be extracted from the proof in [8] and is given by

$$\begin{aligned}&\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2)\\&= I\left(\sqrt{1-\alpha}\,\Phi_1(\mathbf{X}_2^*)+\sqrt{\alpha}\,\Phi_1(\mathbf{X}_1^*);\sqrt{\alpha}\,\mathbf{X}_2^*-\sqrt{1-\alpha}\,\mathbf{X}_1^*\right)\\&+\mathbb{E}\left[\log\left(\frac{\det\left(\alpha\,\nabla\Phi_1(\mathbf{X}_1^*)+(1-\alpha)\,\nabla\Phi_2(\mathbf{X}_2^*)\,\right)}{\det\left(\nabla\Phi_1(\mathbf{X}_1^*)\right)^\alpha\det\left(\nabla\Phi_2(\mathbf{X}_2^*)\right)^{1-\alpha}}\right)\right],\end{aligned} \tag{26}$$

where $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ are i.i.d. Gaussian and $\Phi_i(\mathbf{X}_i^*)$ is a Knöthe map with a property that $\Phi_i(\mathbf{X}_i^*)$ has the same distribution as $\mathbf{X}_i$, $i\in\{1,2\}$. Knöthe map is a standard tool in the theory of optimal transport and the interested reader is referred to [18] and [19] for a detailed treatment of this subject.

A. On the Equality Condition

The following result establishes necessary and sufficient conditions for the equality in (2) and gives several equivalent statements for the equality.

Theorem 2. *The following statements are equivalent:*

$$\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2)=0, \tag{27a}$$

$$\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{\Sigma,\gamma}]=\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{1,\gamma},\mathbf{Y}_{2,\gamma}]\ \textit{a.s.}, \tag{27b}$$

$$\begin{aligned}&\mathbb{E}\left[\sqrt{1-\alpha}\,\mathbb{E}[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}]+\sqrt{\alpha}\,\mathbb{E}[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}]\mid\mathbf{Y}_{\Sigma,\gamma}\right]\\&=\sqrt{1-\alpha}\,\mathbb{E}[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}]+\sqrt{\alpha}\,\mathbb{E}[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}]\ \textit{a.s.},\end{aligned} \tag{27c}$$

$$\begin{aligned}&\mathbb{E}[\sqrt{1-\alpha}\rho_{\mathbf{Y}_{1,\gamma}}(\mathbf{Y}_{1,\gamma})+\sqrt{\alpha}\rho_{\mathbf{Y}_{2,\gamma}}(\mathbf{Y}_{2,\gamma})\mid\mathbf{Y}_{\Sigma,\gamma}]\\&=\sqrt{1-\alpha}\rho_{\mathbf{Y}_{1,\gamma}}(\mathbf{Y}_{1,\gamma})+\sqrt{\alpha}\rho_{\mathbf{Y}_{2,\gamma}}(\mathbf{Y}_{2,\gamma})\ \textit{a.s.},\end{aligned} \tag{27d}$$

$$\rho_{\mathbf{Y}_{\Sigma,\gamma}}(\mathbf{Y}_{\Sigma,\gamma})=\sqrt{1-\alpha}\rho_{\mathbf{Y}_{1,\gamma}}(\mathbf{Y}_{1,\gamma})+\sqrt{\alpha}\rho_{\mathbf{Y}_{2,\gamma}}(\mathbf{Y}_{2,\gamma})\ \textit{a.s.}, \tag{27e}$$

$$(1-\alpha)\,J\left(\mathbf{Y}_{1,\gamma}\right)+\alpha\,J(\mathbf{Y}_{2,\gamma})=J\left(\mathbf{Y}_{\Sigma,\gamma}\right). \tag{27f}$$

Moreover, equality in (27) *holds if and only if* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are Gaussian with identical covariances.*

Proof:

From (22b) it is immediate that $\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2)=0$ if and only if

$$\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{\Sigma,\gamma}]=\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{1,\gamma},\mathbf{Y}_{2,\gamma}]\ \text{a.s.} \tag{28}$$

This shows equivalence between (27a) and (27b).

The equivalence between (27b) and (27c) follows from the towering property in Lemma 4

$$\begin{aligned}&\mathbb{E}\left[\sqrt{1-\alpha}\,\mathbb{E}[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}]+\sqrt{\alpha}\,\mathbb{E}[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}]\mid\mathbf{Y}_{\Sigma,\gamma}\right]\\&=\mathbb{E}\left[\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{1,\gamma},\mathbf{Y}_{2,\gamma}]\mid\mathbf{Y}_{\Sigma,\gamma}\right]\\&=\mathbb{E}\left[\mathbf{X}\mid\mathbf{Y}_{\Sigma,\gamma}\right].\end{aligned} \tag{29}$$

Showing equivalence between (27d), (27e) and (27f) is deferred to Appendix A.

Next, we show that (27) is satisfied if and only if $\mathbf{X}_1$ and $\mathbf{X}_2$ are Gaussian random vectors with identical covariances. The sufficient condition follows by noting that if $\mathbf{X}_1\sim\mathcal{N}(0,\mathbf{K}_{\mathbf{X}_1})$ and $\mathbf{X}_2\sim\mathcal{N}(0,\mathbf{K}_{\mathbf{X}_2})$, then the estimators are linear and are given by

$$\begin{aligned}\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{\Sigma,\gamma}]&=\mathbf{K}_{\mathbf{X}\,Y_{\Sigma,\gamma}}\,\mathbf{K}_{\mathbf{Y}_{\Sigma,\gamma}}^{-1}\,\mathbf{Y}_{\Sigma,\gamma},\\\mathbb{E}[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}]&=\mathbf{K}_{\mathbf{X}_1\,\mathbf{Y}_{1,\gamma}}\,\mathbf{K}_{\mathbf{Y}_{1,\gamma}}^{-1}\,\mathbf{Y}_{1,\gamma},\\\mathbb{E}[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}]&=\mathbf{K}_{\mathbf{X}_2\,\mathbf{Y}_{2,\gamma}}\,\mathbf{K}_{\mathbf{Y}_{2,\gamma}}^{-1}\,\mathbf{Y}_{2,\gamma}.\end{aligned}$$

Therefore, the equality condition in (27b) holds only if

$$\mathbf{K}_{\mathbf{X}\,\mathbf{Y}_{\Sigma,\gamma}}\,\mathbf{K}_{\mathbf{Y}_{\Sigma,\gamma}}^{-1}=\mathbf{K}_{\mathbf{X}_1\,\mathbf{Y}_{1,\gamma}}\,\mathbf{K}_{\mathbf{Y}_{1,\gamma}}^{-1}, \tag{30a}$$

$$\mathbf{K}_{\mathbf{X}\,\mathbf{Y}_{\Sigma,\gamma}}\,\mathbf{K}_{\mathbf{Y}_{\Sigma,\gamma}}^{-1}=\mathbf{K}_{\mathbf{X}_2\,\mathbf{Y}_{2,\gamma}}\,\mathbf{K}_{\mathbf{Y}_{2,\gamma}}^{-1}. \tag{30b}$$

With a small amount of algebra it is not difficult to show that the equality in (30) holds only if $\mathbf{K}_{\mathbf{X}_1}=\mathbf{K}_{\mathbf{X}_2}$.

The necessary condition follows by letting $f_1(\mathbf{Y}_{1,\gamma})=\mathbb{E}[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}]$ and $f_2(\mathbf{Y}_{2,\gamma})=\mathbb{E}[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}]$, in which case the condition in (27c) reduces to

$$\begin{aligned}&\mathbb{E}[\sqrt{1-\alpha}\,f_1(\mathbf{Y}_{1,\gamma})+\sqrt{\alpha}\,f_2(\mathbf{Y}_{2,\gamma})\mid\mathbf{Y}_{\Sigma,\gamma}]\\&=\sqrt{1-\alpha}\,f_1(\mathbf{Y}_{1,\gamma})+\sqrt{\alpha}\,f_2(\mathbf{Y}_{2,\gamma}).\end{aligned} \tag{31}$$

According to Lemma 3 equality in (31) implies that $f_1(\mathbf{Y}_{1,\gamma})$ and $f_2(\mathbf{Y}_{2,\gamma})$ (or $\mathbb{E}[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}]$ and $\mathbb{E}[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}]$) are affine functions with the same slope. In other words, the conditional expectations are given by

$$\begin{aligned}\mathbb{E}[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}]&=\mathbf{A}\mathbf{Y}_{1,\gamma}+\mathbf{b},\\\mathbb{E}[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}]&=\mathbf{A}\mathbf{Y}_{2,\gamma}+\mathbf{c}.\end{aligned}$$

Moreover, by Lemma 5 the linearity of conditional expectations implies that $\mathbf{X}_1$ and $\mathbf{X}_2$ are Gaussian random vectors such that

$$\mathbf{A}=\sqrt{\gamma}\,\mathbf{K}_{\mathbf{X}_1}(\mathbf{I}+\gamma\,\mathbf{K}_{\mathbf{X}_1})^{-1}=\sqrt{\gamma}\,\mathbf{K}_{\mathbf{X}_2}(\mathbf{I}+\gamma\,\mathbf{K}_{\mathbf{X}_2})^{-1}. \tag{32}$$

From (32), it is evident that $\mathbf{X}_1$ and $\mathbf{X}_2$ have identical covariances. This concludes the proof. ∎

IV. Concluding Remark

In this work, we have established the equality condition for the I-MMSE proof of the EPI. Theorem 2 also establishes an equality condition for the following Fisher information inequality

$$(1-\alpha)J(\mathbf{X}_1)+\alpha J(\mathbf{X}_2)\geq J(\sqrt{1-\alpha}\mathbf{X}_1+\sqrt{\alpha}\mathbf{X}_2). \tag{33}$$

This should come as no surprise since the inequality in (33) is a key to establishing the proof of the EPI via *DeBruijn's identity* [3]–[5]. The equality condition in (33) was previously established in [13] by showing that a certain differential equation is satisfied only by Gaussian densities, and in [20] by checking the equality case of the *Variance Drop* inequality. In contrast, our proof relies on Cauchy functional equation and towering property of the conditional expectation.

It is also interesting to observe that the expression for the deficit

$$\begin{aligned}
&2\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2)\\
&=\int_0^\infty \mathbb{E}\left[\|\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{\Sigma,\gamma}]-\mathbb{E}[\mathbf{X}\mid\mathbf{Y}_{1,\gamma},\mathbf{Y}_{2,\gamma}]\|^2\right]d\gamma \quad (34)\\
&=\int_0^\infty \frac{1}{\gamma}\mathbb{E}\Big[\big\|\rho_{\mathbf{Y}_{\Sigma,\gamma}}(\mathbf{Y}_{\Sigma,\gamma})-\sqrt{1-\alpha}\rho_{\mathbf{Y}_{1,\gamma}}(\mathbf{Y}_{1,\gamma})\\
&-\sqrt{\alpha}\rho_{\mathbf{Y}_{2,\gamma}}(\mathbf{Y}_{2,\gamma})\big\|^2\Big]d\gamma, \quad (35)
\end{aligned}$$

is closely related to the mismatched representation of the *relative entropy* [21]

$$2D(P\|Q)=\int_0^\infty \mathbb{E}_P\left[\|\mathbb{E}_P[\mathbf{X}_1\mid\mathbf{Y}_\gamma]-\mathbb{E}_Q[\mathbf{X}_2\mid\mathbf{Y}_\gamma]\|^2\right]d\gamma \quad (36)$$

$$=\int_0^\infty \frac{1}{\gamma}I(P_{\mathbf{Y}_\gamma}\|Q_{\mathbf{Y}_\gamma})d\gamma, \quad (37)$$

where $\mathbf{X}_1\sim P$, $\mathbf{X}_2\sim Q$, and $\sqrt{\gamma}\,\mathbf{X}_1+\mathbf{Z}\sim P_{\mathbf{Y}_\gamma}$, $\sqrt{\gamma}\,\mathbf{X}_2+\mathbf{Z}\sim Q_{\mathbf{Y}_\gamma}$, and where

$$I(P_{\mathbf{Y}_\gamma}\|Q_{\mathbf{Y}_\gamma})=\mathbb{E}_P\left[\left\|\rho_{P_{\mathbf{Y}_\gamma}}(\mathbf{Y}_\gamma)-\rho_{Q_{\mathbf{Y}_\gamma}}(\mathbf{Y}_\gamma)\right\|^2\right], \quad (38)$$

is the *relative Fisher information distance*.

Moreover, in view of the exact characterization of $\Delta(\mathbf{X}_1\|\mathbf{X}_2)$ in (22b), it would be interesting to explore connections to the work in [22]. By using the expression for the deficit in (26), the authors of [22] provided lower bounds on $\Delta(\mathbf{X}_1\|\mathbf{X}_2)$, for uniformly log-concave densities in terms of *Wasserstein distance*:

$$\begin{aligned}
&\frac{2}{\alpha(1-\alpha)}\Delta_\alpha(\mathbf{X}_1\|\mathbf{X}_2)\\
&\geq \inf_{G_1,G_2}\left(W_2^2(P_1,G_1)+W_2^2(P_2,G_2)+W_2^2(G_1,G_2)\right), \quad (39)
\end{aligned}$$

where G_1,G_2 are Gaussian probability measures and where $\mathbf{X}_1\sim P_1$, $\mathbf{X}_2\sim P_2$. The expression in (39) gives a stability result in the sense that if deficit in the EPI is small then the distributions of random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ are close to Gaussian in the $W_2(\cdot,\cdot)$ distance.

ACKNOWLEDGMENT

The authors wish to thank Prof. Thomas A. Courtade for valuable comments.

APPENDIX A
PROOF OF EQUIVALENCE BETWEEN (27d), (27e), AND (27f)

The equivalence between (27c) and (27d) follows from Lemma 6.

Next we show the equivalence between (27d) and (27e). Using Lemma 6 the score function can be written as

$$\rho_{\mathbf{Y}_{\Sigma,\gamma}}(\mathbf{Y}_{\Sigma,\gamma})=\sqrt{\gamma}\,\mathbb{E}\left[\mathbf{X}\mid\mathbf{Y}_{\Sigma,\gamma}\right]-(\mathbf{Y}_{\Sigma,\gamma}). \quad (40)$$

Next, by the towering property of the conditional expectation in Lemma 4

$$\begin{aligned}
&\sqrt{\gamma}\,\mathbb{E}\left[\mathbf{X}\mid\mathbf{Y}_{\Sigma,\gamma}\right]\\
&=\sqrt{\gamma}\,\mathbb{E}\left[\mathbb{E}\left[\mathbf{X}\mid\mathbf{Y}_{1,\gamma},\mathbf{Y}_{2,\gamma}\right]\mid\mathbf{Y}_{\Sigma,\gamma}\right]\\
&=\sqrt{\gamma}\,\mathbb{E}\left[\sqrt{1-\alpha}\mathbb{E}\left[\mathbf{X}_1\mid\mathbf{Y}_{1,\gamma}\right]+\sqrt{\alpha}\mathbb{E}\left[\mathbf{X}_2\mid\mathbf{Y}_{2,\gamma}\right]\mid\mathbf{Y}_{\Sigma,\gamma}\right]\\
&=\mathbb{E}\left[\sqrt{1-\alpha}\left(\rho_{\mathbf{Y}_{1,\gamma}}(\mathbf{Y}_{1,\gamma})+\mathbf{Y}_{1,\gamma}\right)\right.\\
&\left.+\sqrt{\alpha}\left(\rho_{\mathbf{Y}_{2,\gamma}}(\mathbf{Y}_{2,\gamma})+\mathbf{Y}_{2,\gamma}\right)\mid\mathbf{Y}_{\Sigma,\gamma}\right], \quad (41)
\end{aligned}$$

where the last step follows Lemma 6. Putting equation (40) and (41) together we arrive at

$$\begin{aligned}
&\rho_{\mathbf{Y}_{\Sigma,\gamma}}(\mathbf{Y}_{\Sigma,\gamma})\\
&=\mathbb{E}\left[\sqrt{1-\alpha}\rho_{\mathbf{Y}_{1,\gamma}}(\mathbf{Y}_{1,\gamma})+\sqrt{\alpha}\rho_{\mathbf{Y}_{2,\gamma}}(\mathbf{Y}_{2,\gamma})\mid\mathbf{Y}_{\Sigma,\gamma}\right], \quad (42)
\end{aligned}$$

which establishes equivalence between (27d) and (27e). The expression in (42) is sometimes called a convolution identity of the score function [20].

The equivalence between (27d) and (27f) follows from (42) and the definition of Fisher's information in (19). This concludes the proof.

REFERENCES

[1] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 379-423, 623-656, Jul., Oct. 1948.

[2] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," in *Inequalities*. Springer, 2002, pp. 359–365.

[3] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959.

[4] N. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, vol. 11, no. 2, pp. 267–271, 1965.

[5] A. Dembo, T. Cover, and J. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov 1991.

[6] S. Verdú and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2165–2166, 2006.

[7] D. Guo, S. Shamai, and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE Int. Symp. Inf. Theory*. IEEE, 2006, pp. 1011–1015.

[8] O. Rioul, "Yet another proof of the entropy power inequality," *arXiv preprint arXiv:1606.05969*, 2016.

[9] ——, "Optimal transport to the entropy-power inequality and a reverse inequality," *arXiv preprint arXiv:1701.08534*, 2017.

[10] ——, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, 2011.

[11] M. Madiman, J. Melbourne, and P. Xu, "Forward and reverse entropy power inequalities in convex geometry," in *Convexity and Concentration*. Springer, 2017, pp. 427–485.

[12] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[13] E. Carlen and A. Soffer, "Entropy production by block variable summation and central limit theorems," *Communications in mathematical physics*, vol. 140, no. 2, pp. 339–371, 1991.

[14] S. Bobkov and G. P. Chistyakov, "Entropy power inequality for the rényi entropy." *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 708–714, 2015.

[15] J. Aczél, *Lectures on functional equations and their applications*. Academic press, 1966, vol. 19.

[16] S. I. Resnick, *A probability path*. Springer Science & Business Media, 2013.

[17] H. V. Poor, *An introduction to signal detection and estimation*. Springer Science & Business Media, 2013.

[18] C. Villani, *Topics in optimal transportation*. American Mathematical Soc., 2003, no. 58.

[19] ——, *Optimal transport: old and new*. Springer Science & Business Media, 2008, vol. 338.
[20] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2317–2329, 2007.
[21] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3712–3720, Aug 2010.
[22] T. A. Courtade, M. Fathi, and A. Pananjady, "Wasserstein stability of the entropy power inequality for log-concave densities," *arXiv preprint arXiv:1610.07969*, 2016.