Fifty-second Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 3, 2014

An Extremal Inequality for Long Markov Chains

Thomas A. Courtade
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Email: courtade@eecs.berkeley.edu

Jiantao Jiao
Department of Electrical Engineering
Stanford University
Email: jiantao@stanford.edu

***Abstract*—Let $\mathbf{X}, \mathbf{Y}$ be jointly Gaussian vectors, and consider random variables U, V that satisfy the Markov constraint $U - \mathbf{X} - \mathbf{Y} - V$. We prove an extremal inequality relating the mutual informations between all $\binom{4}{2}$ pairs of random variables from the set $(U, \mathbf{X}, \mathbf{Y}, V)$. As a first application, we show that the rate region for the two-encoder quadratic Gaussian source coding problem follows as an immediate corollary of the the extremal inequality. In a second application, we establish the rate region for a vector-Gaussian source coding problem where Löwner-John ellipsoids are approximated based on rate-constrained descriptions of the data.**

I. Introduction

In this paper, we prove the following extremal result, which can be viewed as an entropy power inequality for long Markov chains:

Theorem 1. *For $n \times n$ positive definite matrices Σ_X, Σ_Z, let $\mathbf{X} \sim N(\mu_X, \Sigma_X)$ and $\mathbf{Z} \sim N(\mu_Z, \Sigma_Z)$ be independent n-dimensional Gaussian vectors, and define $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. For any U, V such that $U - \mathbf{X} - \mathbf{Y} - V$ form a Markov chain, the following inequality holds:*

$$2^{-\frac{2}{n}(I(\mathbf{Y};U)+I(\mathbf{X};V))} \geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}}\, 2^{-\frac{2}{n}(I(\mathbf{X};U)+I(\mathbf{Y};V))} + 2^{-\frac{2}{n}(I(\mathbf{X};\mathbf{Y})+I(U;V))}. \quad (1)$$

This result is motivated by multiterminal settings where such Markov chains often arise, but appropriate extremal inequalities for handling them do not exist. Indeed, we will argue shortly that (1) can lead to straightforward solutions of multiterminal Gaussian source coding problems, as the classical entropy power inequality does in point-to-point settings.

In the simplest case, where $\mathbf{Y} = \rho\mathbf{X} + \mathbf{Z}$, $\Sigma_X = I_n$ and $\Sigma_Z = (1-\rho^2)I_n$, Theorem 1 implies

$$2^{-\frac{2}{n}(I(\mathbf{Y};U)+I(\mathbf{X};V))} \geq \rho^2 2^{-\frac{2}{n}(I(\mathbf{X};U)+I(\mathbf{Y};V))} + (1-\rho^2)2^{-\frac{2}{n}I(V;U)}. \quad (2)$$

If V is degenerate, (2) further simplifies to an inequality shown by Oohama in [1], which proved to be instrumental in establishing the rate-distortion region for the one-helper quadratic Gaussian source coding problem. Together with Oohama's work, the sum-rate constraint established by Wagner *et al.* in their *tour de force* [2] completely characterized the rate-distortion region for the two-encoder quadratic Gaussian source coding problem. It turns out that the sum-rate constraint of Wagner *et al.* can be recovered as an immediate corollary to (2), thus unifying the works of Oohama and Wagner *et al.* under a common inequality. The entire argument is given as follows.

First Application: Recovery of the scalar-Gaussian sum-rate constraint

Using the Markov relationship $U - \mathbf{X} - \mathbf{Y} - V$, we can rearrange the exponents in (2) to obtain the equivalent inequality

$$2^{-\frac{2}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V))} \geq 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)}\left(1-\rho^2+\rho^2 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)}\right). \quad (3)$$

The left- and right-hand sides of (3) are monotone decreasing in $\frac{1}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V))$ and $\frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V)$, respectively. Therefore, if

$$\frac{1}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V)) \geq \frac{1}{2}\log\frac{1}{D} \quad \text{and} \quad (4)$$

$$\frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V) \leq R \quad (5)$$

for some pair (R, D), then we have $D \geq 2^{-2R}\left(1-\rho^2+\rho^2 2^{-2R}\right)$, which is a quadratic inequality with respect to the term 2^{-2R}. This is easily solved using the quadratic formula to obtain:

$$2^{-2R} \leq \frac{2D}{(1-\rho^2)\beta(D)} \;\Rightarrow\; R \geq \frac{1}{2}\log\frac{(1-\rho^2)\beta(D)}{2D}, \quad (6)$$

where $\beta(D) \triangleq 1+\sqrt{1+\frac{4\rho^2 D}{(1-\rho^2)^2}}$. Note that Jensen's inequality and the maximum-entropy property of Gaussians imply

$$\frac{1}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V)) \geq \frac{1}{2}\log\frac{1}{\mathsf{mmse}(\mathbf{X}|U,V)\mathsf{mmse}(\mathbf{Y}|U,V)}, \quad (7)$$

where $\mathsf{mmse}(\mathbf{X}|U,V) \triangleq \frac{1}{n}\|\mathbf{X}-\mathbb{E}[\mathbf{X}|U,V]\|^2$, and $\mathsf{mmse}(\mathbf{Y}|U,V)$ is defined similarly. Put $U = f_x(\mathbf{X})$ and $V = f_y(\mathbf{Y})$, where $f_x : \mathbb{R}^n \to [1:2^{nR_x}]$ and $f_y : \mathbb{R}^n \to [1:2^{nR_y}]$. Supposing $\mathsf{mmse}(\mathbf{X}|U,V) \leq d_x$ and $\mathsf{mmse}(\mathbf{Y}|U,V) \leq d_y$, inequalities (4)-(7) together imply

$$R_x + R_y \geq \frac{1}{2}\log\frac{(1-\rho^2)\beta(d_x d_y)}{2d_x d_y}, \quad (8)$$

which is precisely the sum-rate constraint for the two-encoder quadratic Gaussian source coding problem.

This work was supported by the NSF Center for Science of Information under grant agreement CCF-0939370.

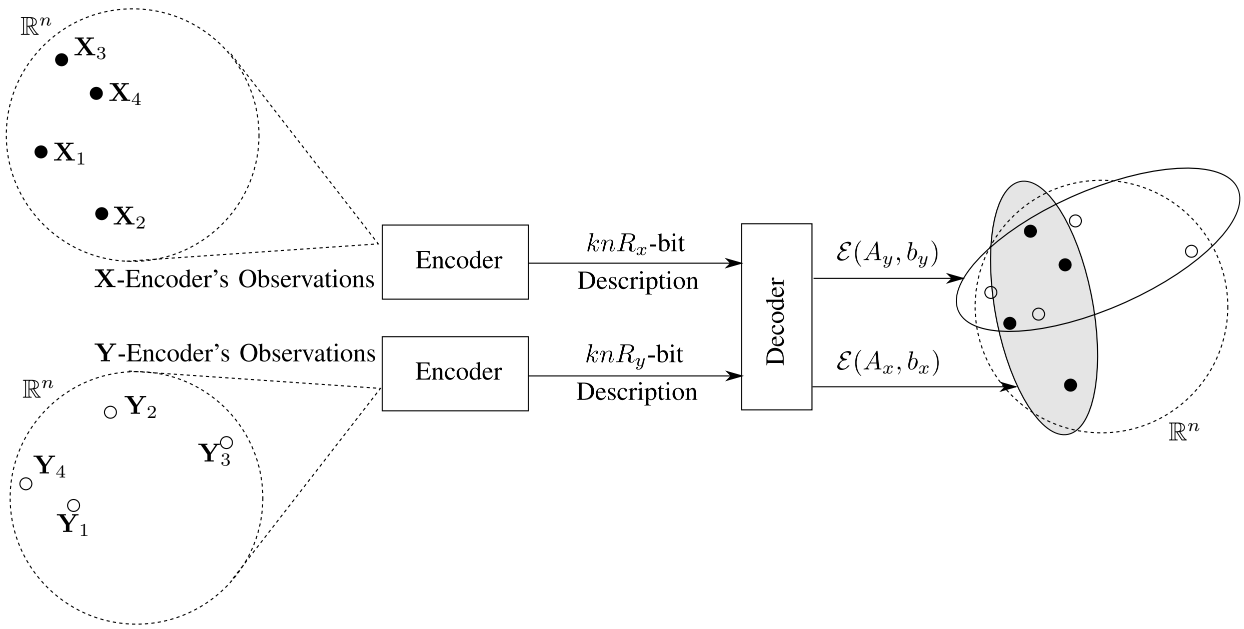

Fig. 1: Computation of covering-ellipsoids from compressed descriptions of the observed data points ($k = 4$). Note the decoder only computes the ellipsoids $\mathcal{E}(A_x, b_x), \mathcal{E}(A_y, b_y)$. The data points at the output of the decoder are only shown for reference.

Second Application: Distributed compression of minimal-volume ellipsoids

Above, recovery of the quadratic Gaussian sum-rate constraint (8) demonstrated the utility of Theorem 1 in proving nontrivial results. Now, we consider a new problem which, the the authors' knowledge, is not a consequence of known results in the literature. In particular, we study the problem of compressing ellipsoids that cover a set of points which, subject to rate constraints, have approximately minimal volume. Such ellipsoids are similar to Löwner-John ellipsoids, which are defined as the (unique) ellipsoid of minimal volume that covers a finite set of points [3]. These minimum-volume ellipsoids and their approximations play a prominent role in the fields of optimization, data analysis, and computational geometry (e.g., [4]).

To begin, we recall that an n-dimensional ellipsoid $\mathcal{E}$ can be parameterized by a positive semidefinite matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$ as follows:

$$\mathcal{E} = \mathcal{E}(A, b) = \{x \in \mathbb{R}^n : \|Ax - b\| \le 1\}.$$

The volume of $\mathcal{E}(A, b)$ is related to the determinant of A by

$$\mathrm{vol}\left(\mathcal{E}(A, b)\right) = \frac{c_n}{|A|},$$

where $c_n \sim \frac{1}{\sqrt{n\pi}}\left(\frac{2\pi e}{n}\right)^{n/2}$ is the volume of the n-dimensional unit ball.

Fix $\rho \in (0, 1)$, and let $\{\Sigma_n : n \ge 1\}$ be a sequence of positive definite $n\times n$ matrices. Suppose $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_k, \mathbf{Y}_k)$ are k independent pairs of jointly Gaussian vectors, each equal in distribution to $(\mathbf{X}, \mathbf{Y})$, where $\mathbb{E}[\mathbf{X}\mathbf{X}^T] = \mathbb{E}[\mathbf{Y}\mathbf{Y}^T] = \Sigma_n$, and $\mathbb{E}[\mathbf{X}\mathbf{Y}^T] = \rho\Sigma_n$.

A $(n, R_x, R_y, \nu_x, \nu_y, k, \Sigma_n, \epsilon)$-code consists of encoding functions $f_x : \mathbb{R}^{kn} \to \{1, 2, \dots, 2^{knR_x}\}$ and $f_y : \mathbb{R}^{kn} \to \{1, 2, \dots, 2^{knR_y}\}$, and a decoding function

$$\psi : (f_x(\mathbf{X}_1, \dots, \mathbf{X}_k), f_y(\mathbf{Y}_1, \dots, \mathbf{Y}_k)) \mapsto (A_x, A_y, b_x, b_y)$$

which satisfies the error-probability constraints

$$\max_{1\le i\le k} \Pr\{\mathbf{X}_i \notin \mathcal{E}(A_x, b_x)\} < \epsilon$$
$$\max_{1\le i\le k} \Pr\{\mathbf{Y}_i \notin \mathcal{E}(A_y, b_y)\} < \epsilon,$$

and the volume constraints

$$\Big(\mathrm{vol}(\mathcal{E}(A_x, b_x))\Big)^{1/n} \le (1+\epsilon)c_n^{1/n}\sqrt{n\nu_x|\Sigma_n|^{1/n}}$$
$$\Big(\mathrm{vol}(\mathcal{E}(A_y, b_y))\Big)^{1/n} \le (1+\epsilon)c_n^{1/n}\sqrt{n\nu_y|\Sigma_n|^{1/n}}.$$

We remark that $\sqrt{n}c_n^{1/n} \to \sqrt{2\pi e}$ as $n \to \infty$ by Stirling's approximation, which explains the normalization factor of $\sqrt{n}$ in the volume constraint.

Definition 1. *For a sequence $\{\Sigma_n : n \ge 1\}$ of positive definite $n \times n$ matrices, a tuple $(R_x, R_y, \nu_x, \nu_y, k)$ is $\{\Sigma_n : n \ge 1\}$-achievable if there exists a sequence of $(n, R_x, R_y, \nu_x, \nu_y, k, \Sigma_n, \epsilon_n)$ codes satisfying $\epsilon_n \to 0$ as $n \to \infty$.*

If $(R_x, R_y, \nu_x, \nu_y, k)$ is a Pareto-optimal $\{\Sigma_n : n \ge 1\}$-achievable point, the corresponding ellipsoids $\mathcal{E}(A_x, b_x), \mathcal{E}(A_y, b_y)$ can be viewed as the best approximations to Löwner-John ellipsoids subject to rate-constrained descriptions of the data. That is, the two ellipsoids cover the k points observed at their respective encoders, and are (essentially) the minimum-volume such ellipsoids that can be computed from rate-constrained descriptions of the data. The general problem setup is illustrated in Figure 1.

Theorem 2. *For any sequence $\{\Sigma_n : n \geq 1\}$ of positive definite $n \times n$ matrices, a tuple $(R_x, R_y, \nu_x, \nu_y, k)$ is $\{\Sigma_n : n \geq 1\}$-achievable if and only if*

$$R_x \geq \frac{1}{2} \log \left[\frac{1}{\nu_x} \left(1 - \rho^2 + \rho^2 2^{-2R_y}\right) \right]$$
$$R_y \geq \frac{1}{2} \log \left[\frac{1}{\nu_y} \left(1 - \rho^2 + \rho^2 2^{-2R_x}\right) \right]$$
$$R_x + R_y \geq \frac{1}{2} \log \frac{(1-\rho^2)\beta(\nu_x \nu_y)}{2\nu_x \nu_y},$$

where $\beta(z) \triangleq 1 + \sqrt{1 + \frac{4\rho^2 z}{(1-\rho^2)^2}}$.

The direct part of Theorem 2 follows from an application of the achievability scheme for the two-encoder quadratic Gaussian source coding problem. However, the converse result does not appear to be a similar consequence since the matrices A_x, A_y describing the principal axes of the ellipsoids are allowed to depend on the source realizations. Nonetheless, with Theorem 1 at our disposal, the proof of the converse is fairly routine. Since the primary goal of this paper is to give a treatment of the extremal inequality (1), the proof of Theorem 2 has been omitted due to space constraints and can be found in the full paper [5].

The remainder of this paper is organized as follows: a treatment of the scalar version of (1) is given in Section II, and the vector generalization is considered in Section III. Closing remarks are provided in Section IV.

II. Scalar Setting

We begin the journey toward our main result by studying the scalar version of Theorem 1. Most of our effort will carry over to the vector setting, but the notation in the scalar case is less cumbersome. Therefore, for the remainder of this section, we will assume that X, Y are jointly Gaussian, each with unit variance and correlation ρ. Our main result in this section is the following rearrangement of (1).

Theorem 3. *Suppose X, Y are jointly Gaussian, each with unit variance and correlation ρ. Then, for any U, V satisfying $U - X - Y - V$, the following inequality holds:*

$$2^{-2(I(Y;U)+I(X;V|U))} \geq (1-\rho^2) + \rho^2 2^{-2(I(X;U)+I(Y;V|U))}. \tag{9}$$

A. Proof of Theorem 3

Instead of working directly with inequality (9), it will be convenient to consider a dual form. For $\lambda \geq 0$, define

$$F(\lambda) \triangleq \inf_{U,V:U-X-Y-V} \Big\{ I(X;U) - \lambda I(Y;U) + I(Y;V|U) - \lambda I(X;V|U) \Big\}. \tag{10}$$

The remainder of this section is devoted to characterizing the function $F(\lambda)$, which we split into a series of lemmas to highlight the main steps. We remark that that the infimum in (10) is attained for any λ. The proof of this is routine, and can be found in the full paper [5]. The bulk of the work ahead is devoted to establishing the existence of valid minimizers U, V for which $X|\{U = u\}$ is normal for almost every u.

To accomplish this, we now describe a simple construction that will be used throughout much of the sequel. This construction was first introduced for proving extremal inequalities in [6]. Suppose U, X, Y, V satisfy the Markov relationship $U - X - Y - V$, and consider two independent copies of U, X, Y, V, which will be denoted by the same variables with subscripts 1 and 2. Define

$$\hat{X}_1 = \frac{X_1 + X_2}{\sqrt{2}} \qquad \hat{X}_2 = \frac{X_1 - X_2}{\sqrt{2}}. \tag{11}$$

In a similar manner, define $\hat{Y}_1, \hat{Y}_2$. Note that $(\hat{X}_1, \hat{X}_2, \hat{Y}_1, \hat{Y}_2)$ and (X_1, X_2, Y_1, Y_2) are equal in distribution. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a one-to-one measurable transformation[1]. Define $\hat{U} = g(U_1, U_2)$ and $\hat{V} = g(V_1, V_2)$.

Lemma 1. *If U, X, Y, V minimize the functional* (10)*, and $\hat{X}_1, \hat{X}_2, \hat{Y}_1, \hat{Y}_2, \hat{U}, \hat{V}$ are constructed as above, then*

1) *For almost every y, $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$ conditioned on $\{\hat{Y}_2 = y\}$ is a valid minimizer of* (10)*.*
2) *For almost every y, $\hat{U}, \hat{X}_2, \hat{Y}_2, \hat{V}$ conditioned on $\{\hat{Y}_1 = y\}$ is a valid minimizer of* (10)*.*

Proof. We can assume $\lambda > 1$, else the data processing inequality implies that $F(\lambda) \geq 0$, which is easily attained.

Let ϕ_1 be such that $\hat{X}_1, \hat{Y}_1$ are independent of ϕ_1 and $\hat{U} - \hat{X}_1 - \hat{Y}_1 - \hat{V}$ conditioned on ϕ_1. Valid assignments of ϕ_1 include any nonempty subset of $\hat{X}_2, \hat{Y}_2$. Let ϕ_2 be defined similarly. Now, observe that we can write:

$$\begin{aligned}
2I(X;U) &= I(\hat{X}_1, \hat{X}_2; \hat{U}) = I(\hat{X}_1; \hat{U}) + I(\hat{X}_2; \hat{U}|\hat{X}_1) \\
&= I(\hat{X}_1; \hat{U}) + I(\hat{X}_2; \hat{U}) + I(\hat{X}_2; \hat{X}_1|\hat{U}) \\
&= I(\hat{X}_1; \hat{U}|\phi_1) + I(\hat{X}_2; \hat{U}|\phi_2) - I(\hat{X}_1; \phi_1|\hat{U}) \\
&\quad - I(\hat{X}_2; \phi_2|\hat{U}) + I(\hat{X}_1; \hat{X}_2|\hat{U}).
\end{aligned}$$

Similarly,

$$\begin{aligned}
2I(Y;U) &= I(\hat{Y}_1; \hat{U}|\phi_1) + I(\hat{Y}_2; \hat{U}|\phi_2) \\
&\quad - I(\hat{Y}_1; \phi_1|\hat{U}) - I(\hat{Y}_2; \phi_2|\hat{U}) + I(\hat{Y}_1; \hat{Y}_2|\hat{U})
\end{aligned}$$

Also, we have

$$\begin{aligned}
&2I(Y;V|U) \\
&= I(\hat{Y}_1, \hat{Y}_2; \hat{V}|\hat{U}) = I(\hat{Y}_1; \hat{V}|\hat{U}) + I(\hat{Y}_2; \hat{V}|\hat{U}) \\
&\quad - I(\hat{Y}_2; \hat{Y}_1|\hat{U}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\
&= I(\hat{Y}_1; \phi_1, \hat{V}|\hat{U}) + I(\hat{Y}_2; \phi_2, \hat{V}|\hat{U}) \\
&\quad - I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) \\
&\quad - I(\hat{Y}_2; \hat{Y}_1|\hat{U}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\
&= I(\hat{Y}_1; \hat{V}|\hat{U}, \phi_1) + I(\hat{Y}_2; \hat{V}|\hat{U}, \phi_2) \\
&\quad - I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\
&\quad + I(\hat{Y}_1; \phi_1|\hat{U}) + I(\hat{Y}_2; \phi_2|\hat{U}) - I(\hat{Y}_1; \hat{Y}_2|\hat{U}).
\end{aligned}$$

[1] Every uncountable Polish space is Borel isomorphic to $\mathbb{R}$ [7].

And, similarly,

$$
\begin{aligned}
&2I(X;V|U)\\
&= I(\hat{X}_1;\hat{V}|\hat{U},\phi_1) + I(\hat{X}_2;\hat{V}|\hat{U},\phi_2)\\
&\quad - I(\hat{X}_1;\phi_1|\hat{U},\hat{V}) - I(\hat{X}_2;\phi_2|\hat{U},\hat{V}) + I(\hat{X}_2;\hat{X}_1|\hat{U},\hat{V})\\
&\quad + I(\hat{X}_1;\phi_1|\hat{U}) + I(\hat{X}_2;\phi_2|\hat{U}) - I(\hat{X}_1;\hat{X}_2|\hat{U}).
\end{aligned}
$$

Assume U, V minimize the functional (10) subject to the Markov constraint $U - X - Y - V$, the existence of such U, V is established in [5]. Then, combining above, we have

$$
\begin{aligned}
&2F(\lambda)\\
=&I(\hat{X}_1;\hat{U}|\phi_1) + I(\hat{X}_2;\hat{U}|\phi_2) - I(\hat{X}_1;\phi_1|\hat{U})\\
&- I(\hat{X}_2;\phi_2|\hat{U}) + I(\hat{X}_1;\hat{X}_2|\hat{U})\\
&- \lambda\Big(I(\hat{Y}_1;\hat{U}|\phi_1) + I(\hat{Y}_2;\hat{U}|\phi_2) - I(\hat{Y}_1;\phi_1|\hat{U})\\
&\quad - I(\hat{Y}_2;\phi_2|\hat{U}) + I(\hat{Y}_1;\hat{Y}_2|\hat{U})\Big)\\
&+ I(\hat{Y}_1;\hat{V}|\hat{U},\phi_1) + I(\hat{Y}_2;\hat{V}|\hat{U},\phi_2)\\
&\quad - I(\hat{Y}_1;\phi_1|\hat{U},\hat{V}) - I(\hat{Y}_2;\phi_2|\hat{U},\hat{V}) + I(\hat{Y}_2;\hat{Y}_1|\hat{U},\hat{V})\\
&\quad + I(\hat{Y}_1;\phi_1|\hat{U}) + I(\hat{Y}_2;\phi_2|\hat{U}) - I(\hat{Y}_1;\hat{Y}_2|\hat{U})\\
&- \lambda\Big(I(\hat{X}_1;\hat{V}|\hat{U},\phi_1) + I(\hat{X}_2;\hat{V}|\hat{U},\phi_2)\\
&\quad - I(\hat{X}_1;\phi_1|\hat{U},\hat{V}) - I(\hat{X}_2;\phi_2|\hat{U},\hat{V}) + I(\hat{X}_2;\hat{X}_1|\hat{U},\hat{V})\\
&\quad + I(\hat{X}_1;\phi_1|\hat{U}) + I(\hat{X}_2;\phi_2|\hat{U}) - I(\hat{X}_1;\hat{X}_2|\hat{U})\Big)\\
=&I(\hat{X}_1;\hat{U}|\phi_1) - \lambda I(\hat{Y}_1;\hat{U}|\phi_1)\\
&\quad + I(\hat{Y}_1;\hat{V}|\hat{U},\phi_1) - \lambda I(\hat{X}_1;\hat{V}|\hat{U},\phi_1)\\
&+ I(\hat{X}_2;\hat{U}|\phi_2) - \lambda I(\hat{Y}_2;\hat{U}|\phi_2)\\
&\quad + I(\hat{Y}_2;\hat{V}|\hat{U},\phi_2) - \lambda I(\hat{X}_2;\hat{V}|\hat{U},\phi_2)\\
&- (\lambda+1)\Big(I(\hat{X}_1;\phi_1|\hat{U}) - I(\hat{X}_2;\phi_2|\hat{U}) + I(\hat{X}_1;\hat{X}_2|\hat{U})\Big)\\
&+ (\lambda+1)\Big(I(\hat{Y}_1;\phi_1|\hat{U}) + I(\hat{Y}_2;\phi_2|\hat{U}) - I(\hat{Y}_1;\hat{Y}_2|\hat{U})\Big)\\
&- I(\hat{Y}_1;\phi_1|\hat{U},\hat{V}) - I(\hat{Y}_2;\phi_2|\hat{U},\hat{V}) + I(\hat{Y}_2;\hat{Y}_1|\hat{U},\hat{V})\\
&+ \lambda\Big(I(\hat{X}_1;\phi_1|\hat{U},\hat{V}) + I(\hat{X}_2;\phi_2|\hat{U},\hat{V}) - I(\hat{X}_2;\hat{X}_1|\hat{U},\hat{V})\Big)\\
\geq&2F(\lambda)\\
&- (\lambda+1)\Big(I(\hat{X}_1;\phi_1|\hat{U}) - I(\hat{X}_2;\phi_2|\hat{U}) + I(\hat{X}_1;\hat{X}_2|\hat{U})\Big)\\
&+ (\lambda+1)\Big(I(\hat{Y}_1;\phi_1|\hat{U}) + I(\hat{Y}_2;\phi_2|\hat{U}) - I(\hat{Y}_1;\hat{Y}_2|\hat{U})\Big)\\
&- I(\hat{Y}_1;\phi_1|\hat{U},\hat{V}) - I(\hat{Y}_2;\phi_2|\hat{U},\hat{V}) + I(\hat{Y}_2;\hat{Y}_1|\hat{U},\hat{V})\\
&+ \lambda\Big(I(\hat{X}_1;\phi_1|\hat{U},\hat{V}) + I(\hat{X}_2;\phi_2|\hat{U},\hat{V}) - I(\hat{X}_2;\hat{X}_1|\hat{U},\hat{V})\Big).
\end{aligned}
$$

The last inequality follows since $\hat{U} - \hat{X}_1 - \hat{Y}_1 - \hat{V}$ conditioned on ϕ_1 is a candidate minimizer of the functional, and same for $\hat{U} - \hat{X}_2 - \hat{Y}_2 - \hat{V}$ conditioned on ϕ_2. Hence, we can conclude that the following must hold

$$
\begin{aligned}
&(\lambda+1)\Big(I(\hat{X}_1;\phi_1|\hat{U}) + I(\hat{X}_2;\phi_2|\hat{U}) - I(\hat{X}_1;\hat{X}_2|\hat{U})\Big)\\
&\quad + I(\hat{Y}_1;\phi_1|\hat{U},\hat{V}) + I(\hat{Y}_2;\phi_2|\hat{U},\hat{V}) - I(\hat{Y}_2;\hat{Y}_1|\hat{U},\hat{V})\\
&\geq (\lambda+1)\Big(I(\hat{Y}_1;\phi_1|\hat{U}) + I(\hat{Y}_2;\phi_2|\hat{U}) - I(\hat{Y}_1;\hat{Y}_2|\hat{U})\Big) \qquad (12)\\
&\quad + \lambda\Big(I(\hat{X}_1;\phi_1|\hat{U},\hat{V}) + I(\hat{X}_2;\phi_2|\hat{U},\hat{V}) - I(\hat{X}_2;\hat{X}_1|\hat{U},\hat{V})\Big).
\end{aligned}
$$

Now, set $\phi_1 = \hat{X}_2, \phi_2 = \hat{Y}_1$. The LHS of (12) is given by

$$
\begin{aligned}
&(\lambda+1)\Big(I(\hat{X}_1;\phi_1|\hat{U}) + I(\hat{X}_2;\phi_2|\hat{U}) - I(\hat{X}_1;\hat{X}_2|\hat{U})\Big)\\
&\quad + I(\hat{Y}_1;\phi_1|\hat{U},\hat{V}) + I(\hat{Y}_2;\phi_2|\hat{U},\hat{V}) - I(\hat{Y}_2;\hat{Y}_1|\hat{U},\hat{V})\\
=&(\lambda+1)\Big(I(\hat{X}_1;\hat{X}_2|\hat{U}) + I(\hat{X}_2;\hat{Y}_1|\hat{U}) - I(\hat{X}_1;\hat{X}_2|\hat{U})\Big)\\
&\quad + I(\hat{Y}_1;\hat{X}_2|\hat{U},\hat{V}) + I(\hat{Y}_2;\hat{Y}_1|\hat{U},\hat{V}) - I(\hat{Y}_2;\hat{Y}_1|\hat{U},\hat{V})\\
=&(\lambda+1)I(\hat{Y}_1;\hat{X}_2|\hat{U}) + I(\hat{Y}_1;\hat{X}_2|\hat{U},\hat{V}).
\end{aligned}
$$

Also, the RHS of (12) can be expressed as

$$
\begin{aligned}
&(\lambda+1)\Big(I(\hat{Y}_1;\phi_1|\hat{U}) + I(\hat{Y}_2;\phi_2|\hat{U}) - I(\hat{Y}_1;\hat{Y}_2|\hat{U})\Big)\\
&+ \lambda\Big(I(\hat{X}_1;\phi_1|\hat{U},\hat{V}) + I(\hat{X}_2;\phi_2|\hat{U},\hat{V}) - I(\hat{X}_2;\hat{X}_1|\hat{U},\hat{V})\Big)\\
&= (\lambda+1)\Big(I(\hat{Y}_1;\hat{X}_2|\hat{U}) + I(\hat{Y}_2;\hat{Y}_1|\hat{U}) - I(\hat{Y}_1;\hat{Y}_2|\hat{U})\Big)\\
&+ \lambda\Big(I(\hat{X}_1;\hat{X}_2|\hat{U},\hat{V}) + I(\hat{X}_2;\hat{Y}_1|\hat{U},\hat{V}) - I(\hat{X}_2;\hat{X}_1|\hat{U},\hat{V})\Big)\\
&= (\lambda+1)I(\hat{Y}_1;\hat{X}_2|\hat{U}) + \lambda I(\hat{Y}_1;\hat{X}_2|\hat{U},\hat{V}).
\end{aligned}
$$

Substituting into (12), we find that $(\lambda-1)I(\hat{Y}_1;\hat{X}_2|\hat{U},\hat{V}) \leq 0 \Rightarrow I(\hat{Y}_1;\hat{X}_2|\hat{U},\hat{V}) = 0$. Therefore, (12) is met with equality, and it follows that:

$$
\begin{aligned}
F(\lambda) &= I(\hat{X}_1;\hat{U}|\hat{X}_2) - \lambda I(\hat{Y}_1;\hat{U}|\hat{X}_2)\\
&\quad + I(\hat{Y}_1;\hat{V}|\hat{U},\hat{X}_2) - \lambda I(\hat{X}_1;\hat{V}|\hat{U},\hat{X}_2)\\
&= I(\hat{X}_2;\hat{U}|\hat{Y}_1) - \lambda I(\hat{Y}_2;\hat{U}|\hat{Y}_1)\\
&\quad + I(\hat{Y}_2;\hat{V}|\hat{U},\hat{Y}_1) - \lambda I(\hat{X}_2;\hat{V}|\hat{U},\hat{Y}_1).
\end{aligned}
$$

Since $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}1$ conditioned on $\{\hat{Y}_1 = y\}$ is a candidate minimizer of (10), the second assertion of the claim follows. By a symmetric argument, if we set $\phi_1 = \hat{Y}_2, \phi_2 = \hat{X}_1$, the roles of the indices are reversed, and we find that

$$
\begin{aligned}
F(\lambda) &= I(\hat{X}_1;\hat{U}|\hat{Y}_2) - \lambda I(\hat{Y}_1;\hat{U}|\hat{Y}_2)\\
&\quad + I(\hat{Y}_1;\hat{V}|\hat{U},\hat{Y}_2) - \lambda I(\hat{X}_1;\hat{V}|\hat{U},\hat{Y}_2)\\
&= I(\hat{X}_2;\hat{U}|\hat{X}_1) - \lambda I(\hat{Y}_2;\hat{U}|\hat{X}_1)\\
&\quad + I(\hat{Y}_2;\hat{V}|\hat{U},\hat{X}_1) - \lambda I(\hat{X}_2;\hat{V}|\hat{U},\hat{X}_1).
\end{aligned}
$$

This establishes the first assertion of the claim and completes the proof. □

Lemma 2. *If U_λ, V_λ are valid minimizers of the functional* (10) *for parameter λ, then*

$$I(Y;U_\lambda) + I(X;V_\lambda|U_\lambda) = -F'(\lambda) \quad \text{for a.e. } \lambda. \qquad (13)$$

Proof. To begin, let $U_{\lambda+\Delta}, V_{\lambda+\Delta}$ be arbitrary, valid minimizers of the functional (10) for parameter $\lambda + \Delta$, and let

$U_{\lambda-\Delta}, V_{\lambda-\Delta}$ be arbitrary, valid minimizers of the functional (10) for parameter $\lambda-\Delta$. Next, note that $F(\lambda)$ is concave and (strictly) monotone decreasing in λ, and hence $F'(\lambda)$ exists for a.e. λ. Thus, for any $\Delta > 0$,

$$\begin{aligned}
&\frac{F(\lambda+\Delta)-F(\lambda)}{\Delta}\\
&=\frac{1}{\Delta}\Big(I(X;U_{\lambda+\Delta})+I(Y;V_{\lambda+\Delta}|U_{\lambda+\Delta})\\
&\qquad-(\lambda+\Delta)\left(I(Y;U_{\lambda+\Delta})+I(X;V_{\lambda+\Delta}|U_{\lambda+\Delta})\right)\Big)\\
&\quad-\frac{1}{\Delta}\Big(I(X;U_{\lambda})+I(Y;V_{\lambda}|U_{\lambda})\\
&\qquad-\lambda\left(I(Y;U_{\lambda})+I(X;V_{\lambda}|U_{\lambda})\right)\Big)\\
&=-\Big(I(Y;U_{\lambda+\Delta})+I(X;V_{\lambda+\Delta}|U_{\lambda+\Delta})\Big)\\
&\quad+\frac{1}{\Delta}\Big(I(X;U_{\lambda+\Delta})+I(Y;V_{\lambda+\Delta}|U_{\lambda+\Delta})\\
&\qquad-\lambda\left(I(Y;U_{\lambda+\Delta})+I(X;V_{\lambda+\Delta}|U_{\lambda+\Delta})\right)\Big)\\
&\quad-\frac{1}{\Delta}\Big(I(X;U_{\lambda})+I(Y;V_{\lambda}|U_{\lambda})\\
&\qquad-\lambda\left(I(Y;U_{\lambda})+I(X;V_{\lambda}|U_{\lambda})\right)\Big)\\
&\geq-\Big(I(Y;U_{\lambda+\Delta})+I(X;V_{\lambda+\Delta}|U_{\lambda+\Delta})\Big),
\end{aligned}$$

where the last inequality follows since $U_{\lambda+\Delta}, V_{\lambda+\Delta}$ is a candidate minimizer of (10) with parameter λ.

Similarly,

$$\begin{aligned}
&\frac{F(\lambda)-F(\lambda-\Delta)}{\Delta}\\
&=\frac{1}{\Delta}\Big(I(X;U_{\lambda})+I(Y;V_{\lambda}|U_{\lambda})\\
&\qquad-\lambda\left(I(Y;U_{\lambda})+I(X;V_{\lambda}|U_{\lambda})\right)\Big)\\
&\quad-\frac{1}{\Delta}\Big(I(X;U_{\lambda-\Delta})+I(Y;V_{\lambda-\Delta}|U_{\lambda-\Delta})\\
&\qquad-(\lambda-\Delta)\left(I(Y;U_{\lambda-\Delta})+I(X;V_{\lambda-\Delta}|U_{\lambda-\Delta})\right)\Big)\\
&=-\Big(I(Y;U_{\lambda-\Delta})-I(X;V_{\lambda+\Delta}|U_{\lambda+\Delta})\Big)\\
&\quad+\frac{1}{\Delta}\Big(I(X;U_{\lambda})+I(Y;V_{\lambda}|U_{\lambda})\\
&\qquad-\lambda\left(I(Y;U_{\lambda})+I(X;V_{\lambda}|U_{\lambda})\right)\Big)\\
&\quad-\frac{1}{\Delta}\Big(I(X;U_{\lambda-\Delta})+I(Y;V_{\lambda-\Delta}|U_{\lambda-\Delta})\\
&\qquad-\lambda\left(I(Y;U_{\lambda-\Delta})+I(X;V_{\lambda-\Delta}|U_{\lambda-\Delta})\right)\Big)\\
&\leq-\Big(I(Y;U_{\lambda-\Delta})+I(X;V_{\lambda-\Delta}|U_{\lambda-\Delta})\Big),
\end{aligned}$$

where the last inequality follows since $U_{\lambda-\Delta}, V_{\lambda-\Delta}$ is a candidate minimizer of (10) with parameter λ. Recalling concavity of $F(\lambda)$, we have shown

$$\begin{aligned}
&I(Y;U_{\lambda+\Delta})+I(X;V_{\lambda+\Delta}|U_{\lambda+\Delta})\\
&\geq -F'(\lambda)\\
&\geq I(Y;U_{\lambda-\Delta})+I(X;V_{\lambda-\Delta}|U_{\lambda-\Delta}).
\end{aligned}$$

As F' is monotone and well-defined up to a set of measure zero, we are justified in writing

$$-\lim_{z\to\lambda^+} F'(z) \geq I(Y;U_\lambda)+I(X;V_\lambda|U_\lambda) \geq -\lim_{z\to\lambda^-} F'(z).$$

Since F' is monotone, it is almost everywhere continuous, and so the LHS and RHS above coincide with $-F'(\lambda)$ for almost every λ. ☐

Since the derivative $F'(\lambda)$ is just a function of F itself, and not of a particular minimizer, we have the following

Corollary 1. *If U_λ, V_λ are valid minimizers of the functional* (10) *for parameter λ, then*

$$I(X;U_\lambda|Y)+I(Y;V_\lambda|X)=F(\lambda)-(\lambda-1)F'(\lambda) \quad (14)$$

for a.e. λ.

Proof. Suppose U_λ, V_λ are valid minimizers. Then, we can write:

$$\begin{aligned}
&F(\lambda)\\
&=I(X;U_\lambda)-\lambda I(Y;U_\lambda)+I(Y;V_\lambda|U_\lambda)-\lambda I(X;V_\lambda|U_\lambda)\\
&=I(X;U_\lambda)-\lambda I(Y;U_\lambda)+I(Y;V_\lambda)\\
&\quad-\lambda I(X;V_\lambda)+(\lambda-1)I(U_\lambda;V_\lambda)\\
&=I(X;U_\lambda|Y)+I(Y;V_\lambda|X)\\
&\quad+(\lambda-1)\Big(I(U_\lambda;V_\lambda)-I(Y;U_\lambda)-I(X;V_\lambda)\Big)\\
&=I(X;U_\lambda|Y)+I(Y;V_\lambda|X)+(\lambda-1)F'(\lambda),
\end{aligned}$$

where the last line follows from Lemma 2. ☐

Lemma 3. *If U, V are valid minimizers of the functional* (10)*, and $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{V}$ are constructed as described above, then there exist valid minimizers $\widetilde{U}, \widetilde{V}$ such that*

$$I(X;U|Y) \geq I(X;\widetilde{U}|Y)+\frac{1}{2}I(\hat{X}_1;\hat{X}_2|\hat{U},\hat{Y}_1,\hat{Y}_2). \quad (15)$$

Proof. To begin, note that:

$$\begin{aligned}
&I(\hat{X}_1;\hat{X}_2|\hat{U},\hat{Y}_1,\hat{Y}_2)\\
&=I(\hat{X}_1;\hat{X}_2,\hat{U}|\hat{Y}_1,\hat{Y}_2)-I(\hat{X}_1;\hat{U}|\hat{Y}_1,\hat{Y}_2)\\
&=I(\hat{X}_1;\hat{U}|\hat{Y}_1,\hat{Y}_2,\hat{X}_2)-I(\hat{X}_1;\hat{U}|\hat{Y}_1,\hat{Y}_2)\\
&=I(\hat{X}_1,\hat{X}_2;\hat{U}|\hat{Y}_1,\hat{Y}_2)-I(\hat{X}_1;\hat{U}|\hat{Y}_1,\hat{Y}_2)-I(\hat{X}_2;\hat{U}|\hat{Y}_1,\hat{Y}_2)\\
&=2I(X;U|Y)-I(\hat{X}_1;\hat{U}|\hat{Y}_1,\hat{Y}_2)-I(\hat{X}_2;\hat{U}|\hat{Y}_1,\hat{Y}_2).
\end{aligned}$$

Thus, without loss of generality (relabeling indices 1 and 2 if necessary), we can assume

$$I(X;U|Y) \geq I(\hat{X}_1;\hat{U}|\hat{Y}_1,\hat{Y}_2)+\frac{1}{2}I(\hat{X}_1;\hat{X}_2|\hat{U},\hat{Y}_1,\hat{Y}_2).$$

Lemma 1 asserts that, for almost every y, the tuple $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$ conditioned on $\{\hat{Y}_2=y\}$ is a valid minimizer of (10). Hence, there must exist a y^* such that

$$I(X;U|Y) \geq I(\hat{X}_1;\hat{U}|\hat{Y}_1,\hat{Y}_2=y^*)+\frac{1}{2}I(\hat{X}_1;\hat{X}_2|\hat{U},\hat{Y}_1,\hat{Y}_2),$$

and $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$ conditioned on $\{\hat{Y}_2 = y^*\}$ is a valid minimizer of (10). Therefore, the claim follows by letting

$\widetilde{U}, X, Y, \widetilde{V}$ be equal in distribution to $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$ conditioned on $\{\hat{Y}_2 = y^*\}$. □

Corollary 2. *There exist U, V which are valid minimizers of the functional* (10)*, and satisfy*

$$I(\hat{X}_1; \hat{X}_2 | \hat{U}, \hat{Y}_1, \hat{Y}_2) = 0, \tag{16}$$

where $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{V}$ are constructed as described above.

Proof. Applying Lemma 3, we can inductively construct a sequence of valid minimizers $\{U^{(k)}, X, Y, V^{(k)}\}_{k\geq 1}$ which satisfy

$$I(X; U^{(k)}|Y) \geq I(X; U^{(k+1)}|Y) + \frac{1}{2} I(\hat{X}_1; \hat{X}_2 | \hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2)$$

for $k = 1, 2, \ldots$, where $\hat{U}^{(k)}, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{V}^{(k)}$ are constructed from two independent copies of $U^{(k)}, X, Y, V^{(k)}$. By Corollary 1, we must also have

$$I(X; U^{(k)}|Y) + I(Y; V^{(k)}|X) = F(\lambda) - (\lambda - 1)F'(\lambda)$$

for all $k = 1, 2, \ldots$. Therefore, for any n, we have:

$$\begin{aligned}
&F(\lambda) - (\lambda-1)F'(\lambda) \\
&= \frac{1}{n}\sum_{k=1}^{n} I(X; U^{(k)}|Y) + I(Y; V^{(k)}|X) \\
&\geq \frac{1}{n}\sum_{k=2}^{n} \Big(I(X; U^{(k)}|Y) + I(Y; V^{(k)}|X) \Big) \\
&\quad + \frac{1}{n}\Big(I(X; U^{(n+1)}|Y) + I(Y; V^{(1)}|X) \Big) \\
&\quad + \frac{1}{2n}\sum_{k=1}^{n} I(\hat{X}_1; \hat{X}_2 | \hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2) \\
&\geq \frac{n-1}{n}\Big(F(\lambda) - (\lambda-1)F'(\lambda) \Big) \\
&\quad + \frac{1}{2n}\sum_{k=1}^{n} I(\hat{X}_1; \hat{X}_2 | \hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2),
\end{aligned}$$

and thus

$$\sum_{k=1}^{n} I(\hat{X}_1; \hat{X}_2 | \hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2) \leq 2\Big(F(\lambda) - (\lambda-1)F'(\lambda) \Big).$$

Hence, the sum on the LHS above must converge as $n \to \infty$, implying

$$\lim_{k\to\infty} I(\hat{X}_1; \hat{X}_2 | \hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2) = 0.$$

Arguing carefully (see the full paper [5]), we can conclude that there exists an optimizer U, V for which $I(\hat{X}_1; \hat{X}_2 | \hat{U}, \hat{Y}_1, \hat{Y}_2)$ is exactly zero. □

Lemma 4. *[8] Let $\mathbf{A}_1$ and $\mathbf{A}_2$ be mutually independent n-dimensional random vectors. If $\mathbf{A}_1 + \mathbf{A}_2$ is independent of $\mathbf{A}_1 - \mathbf{A}_2$, then $\mathbf{A}_1$ and $\mathbf{A}_2$ are normally distributed.*

Corollary 3. *There exist optimizers U, V such that $X|\{U = u\}$ is Gaussian for a.e. u.*

Proof. By construction and Corollary 2, we can conclude that there exist optimizers U, V for which

$$I(X_1; X_2 | U_1, U_2, Y_1, Y_2) = I(\hat{X}_1; \hat{X}_2 | U_1, U_2, Y_1, Y_2) = 0.$$

Therefore, by Lemma 4, there exist optimizers U, V such that $X|\{U, Y = u, y\}$ is Gaussian for a.e. u, y.

Letting $P(x, y, u, v)$ denote the joint distribution of the above X, Y, U, V, we can use Markovity to write:

$$\begin{aligned}
P(x, y, u, v) &= P(u)P(y|u)P(x|u, y)P(v|y) &(17)\\
&= P(u)P(x|u)P(y|x)P(v|y). &(18)
\end{aligned}$$

Taking logarithms and rearranging, we have the identity

$$\log(P(x|u)) = \log(P(y|u)) + \log(P(x|u, y)) - \log(P(y|x)). \tag{19}$$

Since $X|\{U, Y = u, y\}$ is Gaussian for a.e. u, y, and X, Y are jointly Gaussian by assumption, the RHS of (19) is a quadratic function of x for a.e. u, y. Hence, $\log(P(x|u))$ is quadratic in x for a.e. u, and the claim follows. □

Lemma 5. *[1] For any U satisfying $U - X - Y$, the following inequality holds:*

$$2^{-2I(Y;U)} \geq 1 - \rho^2 + \rho^2 2^{-2I(X;U)}. \tag{20}$$

Proof. Consider any U satisfying $U - X - Y$. Let Y_u, X_u denote the random variables X, Y conditioned on $U = u$. By Markovity and definition of X, Y, we have that $Y_u = \rho X_u + Z$, where $Z \sim N(0, 1 - \rho^2)$ is independent of X_u. Hence, the conditional entropy power inequality implies that

$$\begin{aligned}
2^{2h(Y|U)} &\geq \rho^2 2^{2h(X|U)} + 2\pi e(1 - \rho^2) \\
&= 2\pi e \rho^2 2^{-2I(X;U)} + 2\pi e(1 - \rho^2).
\end{aligned}$$

From here, the lemma easily follows. □

Lemma 6.

$$\inf_{U:U-X-Y} \Big\{ I(X;U) - \lambda I(Y;U) \Big\} \tag{21}$$

$$= \begin{cases} \frac{1}{2}\left[\log\left(\frac{\rho^2(\lambda-1)}{1-\rho^2}\right) - \lambda \log\left(\frac{\lambda-1}{\lambda(1-\rho^2)}\right)\right] & \text{If } \lambda \geq 1/\rho^2 \\ 0 & \text{If } 0 \leq \lambda \leq 1/\rho^2. \end{cases}$$

Proof. The claim follows from Lemma 5 and elementary calculus. Details can be found in the full paper [5]. □

Lemma 7.

$$F(\lambda) = \inf_{U:U-X-Y} \Big\{ I(X;U) - \lambda I(Y;U) \Big\}.$$

Proof. We will assume $\lambda \geq 1/\rho^2$. The claim that $F(\lambda) = 0$ for $0 \leq \lambda < 1/\rho^2$ follows immediately by monotonicity of $F(\lambda)$. To this end, let U, V be optimizers such that $X|\{U = u\}$ is Gaussian for a.e. u. The existence of such U, V is guaranteed by Corollary 3. Let X_u, Y_u denote the random variables X, Y conditioned on $U = u$. By Markovity, X_u, Y_u are jointly Gaussian with

$$Y_u = \rho X_u + Z,$$

where $Z \sim N(0, 1-\rho^2)$ is independent of X_u. Letting σ_u^2 be the variance of X_u, the variance of Y_u is $\rho^2\sigma_u^2 + (1-\rho^2)$. Moreover, the squared linear correlation of X_u and Y_u is given by

$$\rho_u^2 \triangleq \frac{\rho^2\sigma_u^2}{\rho^2\sigma_u^2 + (1-\rho^2)}.$$

By Lemma 6,

$$\begin{aligned}&\inf_{V: V - Y_u - X_u} \left\{ I(Y_u; V) - \lambda I(X_u; V) \right\} \\ &= \frac{1}{2}\left[\log\left(\frac{\rho_u^2(\lambda-1)}{1-\rho_u^2}\right) - \lambda \log\left(\frac{\lambda-1}{\lambda(1-\rho_u^2)}\right)\right] \quad (22)\end{aligned}$$

whenever $\lambda \geq 1/\rho_u^2$, and the infimum is equal to zero otherwise.

By definition, we have

$$\begin{aligned} F(\lambda) &= I(X;U) - \lambda I(Y;U) + I(Y;V|U) - \lambda I(X;V|U) \\ &= \int \Big(h(X) - h(X|u) - \lambda(h(Y) - h(Y|U=u)) \\ &\qquad + I(Y;V|U=u) - \lambda I(X;V|U=u) \Big) dP_U(u) \\ &= \int \Big(-\frac{1}{2}\log\sigma_u^2 + \frac{\lambda}{2}\log(\rho^2\sigma_u^2 + (1-\rho^2)) \quad (23) \\ &\qquad + I(Y;V|U=u) - \lambda I(X;V|U=u) \Big) dP_U(u). \end{aligned}$$

If $\lambda \geq 1/\rho_u^2$, we can apply (22) to bound the integrand in (23) as follows

$$\begin{aligned} &-\frac{1}{2}\log\sigma_u^2 + \frac{\lambda}{2}\log(\rho^2\sigma_u^2 + (1-\rho^2)) \\ &\quad + I(Y;V|U=u) - \lambda I(X;V|U=u) \\ &\geq -\frac{1}{2}\log\sigma_u^2 + \frac{\lambda}{2}\log(\rho^2\sigma_u^2 + (1-\rho^2)) \\ &\quad + \frac{1}{2}\left[\log\left(\frac{\rho_u^2(\lambda-1)}{1-\rho_u^2}\right) - \lambda\log\left(\frac{\lambda-1}{\lambda(1-\rho_u^2)}\right)\right] \\ &= \frac{1}{2}\left[\log\left(\frac{\rho^2(\lambda-1)}{1-\rho^2}\right) - \lambda\log\left(\frac{\lambda-1}{\lambda(1-\rho^2)}\right)\right]. \end{aligned}$$

On the other hand, if $\lambda \leq 1/\rho_u^2$, then we can bound the integrand in (23) by

$$\begin{aligned} &-\frac{1}{2}\log\sigma_u^2 + \frac{\lambda}{2}\log(\rho^2\sigma_u^2 + (1-\rho^2)) \\ &\quad + I(Y;V|U=u) - \lambda I(X;V|U=u) \\ &\geq -\frac{1}{2}\log\sigma_u^2 + \frac{\lambda}{2}\log(\rho^2\sigma_u^2 + (1-\rho^2)) \\ &\geq \frac{1}{2}\left[\log\left(\frac{\rho^2(\lambda-1)}{1-\rho^2}\right) - \lambda\log\left(\frac{\lambda-1}{\lambda(1-\rho^2)}\right)\right], \end{aligned}$$

where the final inequality follows since $\lambda \leq 1/\rho_u^2 \Rightarrow \sigma_u^2 \leq \frac{1-\rho^2}{\rho^2(\lambda-1)}$, and $-\frac{1}{2}\log\sigma_u^2 + \frac{\lambda}{2}\log(\rho^2\sigma_u^2 + (1-\rho^2))$ is monotone decreasing in σ_u^2 for $\sigma_u^2 \leq \frac{1-\rho^2}{\rho^2(\lambda-1)}$. Therefore, we have established the inequality

$$F(\lambda) \geq \frac{1}{2}\left[\log\left(\frac{\rho^2(\lambda-1)}{1-\rho^2}\right) - \lambda\log\left(\frac{\lambda-1}{\lambda(1-\rho^2)}\right)\right].$$

The definition of $F(\lambda)$ together with Lemma 6 implies the reverse inequality, completing the proof. □

Since (21) is a dual characterization of the inequality (20), we have proved Theorem 3.

Remark 1. *Although Lemma 7 implies that the functional* (10) *is minimized when either U or V is degenerate, there are also minimizers for which this is not the case. For example, if $-1 \leq \rho_u, \rho_v \leq 1$ satisfy*

$$(1-\rho^2)(1-\rho^2\rho_u^2\rho_v^2) = \rho^2(\lambda-1)(1-\rho_u^2)(1-\rho_v^2),$$

then U, V defined according to

$$\begin{aligned} U &= \rho_u X + Z_u \\ V &= \rho_v Y + Z_v, \end{aligned}$$

where $Z_u \sim N(0, 1-\rho_u^2)$ and $Z_v \sim N(0, 1-\rho_v^2)$ are independent of everything else, also minimize (10).

III. Vector Setting

Now, we turn our attention to the vector case. Throughout the remainder of this section, let Σ_X, Σ_Z be positive definite $n \times n$ matrices. Suppose $\mathbf{X} \sim N(\mu_X, \Sigma_X)$ and $\mathbf{Z} \sim N(\mu_Z, \Sigma_Z)$ are independent n-dimensional Gaussian vectors, and define $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. We recall the statement of Theorem 1 here, along with conditions for equality:

Theorem 1. *For any U, V such that $U - \mathbf{X} - \mathbf{Y} - V$,*

$$\begin{aligned} &2^{-\frac{2}{n}(I(\mathbf{Y};U) + I(\mathbf{X};V|U))} \quad (24) \\ &\geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}}\, 2^{-\frac{2}{n}(I(\mathbf{X};U) + I(\mathbf{Y};V|U))} + 2^{-\frac{2}{n}I(\mathbf{X};\mathbf{Y})}. \end{aligned}$$

Moreover, equality holds iff $\mathbf{X}|\{U=u\} \sim N(\mu_u, \Sigma_{X|U})$ for all u, where $\mu_u \triangleq \mathbb{E}[\mathbf{X}|U=u]$, and $\Sigma_{X|U}$ is proportional to Σ_Z.

A. Proof of Theorem 1

Instead of working directly with inequality (24), it will be convenient to consider a dual form. As before, for $\lambda \geq 0$, define

$$\begin{aligned} \mathbf{F}(\lambda) \triangleq \inf_{U,V: U - \mathbf{X} - \mathbf{Y} - V} \Big\{ &I(\mathbf{X};U) - \lambda I(\mathbf{Y};U) \\ &+ I(\mathbf{Y};V|U) - \lambda I(\mathbf{X};V|U) \Big\} \quad (25) \end{aligned}$$

The remainder of this section is devoted to bounding the function $\mathbf{F}(\lambda)$.

To begin, we remark that the extension of the results up to Lemma 5 for the scalar setting immediately generalize to the present vector case by repeating the proofs verbatim. Namely, we have the key observation:

Corollary 4. *There exist U, V which minimize the functional* (25) *such that $\mathbf{X}|\{U=u\}$ is Gaussian for a.e. u.*

Therefore, we pick up at this point and sketch a proof of Theorem 1, beginning with a vector version of Lemma 5.

Lemma 8. *For any U such that $U - \mathbf{X} - \mathbf{Y}$,*

$$2^{-2I(\mathbf{Y};U)/n} \geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}}\, 2^{-2I(\mathbf{X};U)/n} + 2^{-2I(\mathbf{X};\mathbf{Y})/n}.$$

Moreover, equality holds iff $\mathbf{X}|\{U = u\} \sim N(\mu_u, \Sigma)$ for all u, where $\mu_u \triangleq \mathbb{E}[\mathbf{X}|U = u]$, and $\Sigma \propto \Sigma_Z$.

The proof is a straightforward generalization of the scalar version given for Lemma 5, and has been omitted due to space constraints. Details can be found in the full paper [5].

Lemma 9. *Let U be such that $U - \mathbf{X} - \mathbf{Y}$.*
1) *If $\lambda \geq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$, then*

$$I(\mathbf{X};U) - \lambda I(\mathbf{Y};U) \geq \frac{n}{2}\log\left(\frac{|\Sigma_X|^{1/n}(\lambda-1)}{|\Sigma_Z|^{1/n}}\right) - \frac{\lambda n}{2}\log\left(\frac{|\Sigma_X+\Sigma_Z|^{1/n}(\lambda-1)}{|\Sigma_Z|^{1/n}\lambda}\right). \tag{26}$$

2) *If $0 \leq \lambda \leq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$, then*

$$I(\mathbf{X};U) - \lambda I(\mathbf{Y};U) \geq -\frac{\lambda n}{2}\log\left(\frac{|\Sigma_X+\Sigma_Z|^{1/n}}{|\Sigma_X|^{1/n}+|\Sigma_Z|^{1/n}}\right). \tag{27}$$

Proof. The claim follows from the Minkowski determinant theorem [9], Lemma 8 and elementary calculus. Complete details can be found in the full paper [5]. □

Note that the lower bound (26) is achieved if U can be chosen such that $\mathbf{X}|\{U = u\} \sim N(\mu_u, \Sigma_{X|U})$ for each u, and $\Sigma_{X|U} = \alpha\Sigma_Z$. The lower bound (27) is only attainable if Σ_X and Σ_Z are proportional. In this case, the RHS of (27) is precisely zero.

Lemma 10.
1) *If $\lambda \geq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$, then*

$$\mathbf{F}(\lambda) \geq \frac{n}{2}\left[\log\left(\frac{|\Sigma_X|^{1/n}(\lambda-1)}{|\Sigma_Z|^{1/n}}\right) - \lambda\log\left(\frac{|\Sigma_X+\Sigma_Z|^{1/n}(\lambda-1)}{|\Sigma_Z|^{1/n}\lambda}\right)\right]. \tag{28}$$

2) *If $0 \leq \lambda \leq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$, then*

$$\mathbf{F}(\lambda) \geq -\frac{\lambda n}{2}\log\left(\frac{|\Sigma_X+\Sigma_Z|^{1/n}}{|\Sigma_X|^{1/n}+|\Sigma_Z|^{1/n}}\right). \tag{29}$$

The proof of Lemma 10 is similar to that of Lemma 7, except we require Lemma 9 in place of Lemma 6. A complete proof is omitted due to space constraints; full details can be found in [5].

To conclude, observe that Lemma 10 bounds the dual form (25) of the desired inequality (24). It is a straightforward exercise in calculus to show that this bound is sufficient to prove Theorem 1. Details can be found in [5, Appendix C].

IV. Closing Remarks

The focus of this paper was on the extremal result asserted by Theorem 1, and not on operational coding problems. However, since the entropy-power-like inequality of Theorem 1 leads to what is arguably the simplest solution for the two-encoder quadratic Gaussian source coding problem (an archetypical problem in network information theory), we have little doubt that it will find other interesting applications. We provided Theorem 2 as one such example. As another example, Theorem 3 can be applied to show that jointly Gaussian auxiliaries exhaust the rate region for multiterminal source coding under logarithmic loss [10] when the sources are Gaussian. This leads to yet another solution for the two-encoder quadratic Gaussian source coding problem, and unifies the two problems under the paradigm of compression under logarithmic loss.

Acknowledgement

The authors acknowledge several conversations with Tsachy Weissman, Kartik Venkat, and Vignesh Ganapathi-Subramanian which led to deeper intuition and insight. The first author also wishes to acknowledge an inspiring discussion with Chandra Nair that took place at the 2014 International Zurich Seminar on Communications following the presentation of [11]. This conversation and the following exchanges generated the spark which led to a successful proof of Theorem 3, a long-held conjecture of the first author that was first made public on the *Information Theory b-Log* in March, 2013 [12], [13].

References

[1] Y. Oohama, "Gaussian multiterminal source coding," *Information Theory, IEEE Transactions on*, vol. 43, pp. 1912–1923, Nov 1997.

[2] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *Information Theory, IEEE Transactions on*, vol. 54, pp. 1938–1961, May 2008.

[3] M. Henk, "Löwner–John ellipsoids," *Optimization Stories, Documenta Mathematica*, pp. 95–106, 2012.

[4] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[5] T. A. Courtade and J. Jiao, "An extremal inequality for long Markov chains," *CoRR*, vol. abs/1404.6984, 2014.

[6] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *Information Theory, IEEE Trans. on*, vol. 60, pp. 2087–2104, April 2014.

[7] S. M. Srivastava, *A course on Borel sets*, vol. 180. Springer, 1998.

[8] S. Ghurye, I. Olkin, *et al.*, "A characterization of the multivariate normal distribution," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 533–541, 1962.

[9] M. Marcus and H. Minc, *A survey of matrix theory and matrix inequalities*, vol. 14. Courier Dover Publications, 1992. p. 115.

[10] T. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *Information Theory, IEEE Transactions on*, vol. 60, pp. 740–761, Jan 2014.

[11] T. Courtade, J. Jiao, and T. Weissman, "On an extremal data processing inequality for long Markov chains," in *International Zurich Seminar on Communications (IZS)*, pp. 1 – 4, Feb. 2014.

[12] T. Courtade, "An extremal conjecture: Experimenting with online collaboration," *Information Theory b-Log*, March 5, 2013. http://blogs.princeton.edu/blogit/2013/03/05/.

[13] T. Javidi, Ed., "In the blogosphere...," *IEEE Information Theory Society Newsletter*, vol. 63, pp. 2087–2104, June 2013.

[14] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*, vol. 902. John Wiley & Sons, 2011.

[15] R. Durrett, *Probability: theory and examples*, vol. 3. Cambridge university press, 2010.