

Article

A Monotone Path Proof of an Extremal Result for Long Markov Chains

Jia Wang ^{1,*} and Jun Chen ^{2,*} ¹ Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China² Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada

* Correspondence: jiaawang@sjtu.edu.cn (J.W.); junchen@ece.mcmaster.ca (J.C.)

Received: 31 January 2019; Accepted: 11 March 2019; Published: 13 March 2019



Abstract: We prove an extremal result for long Markov chains based on the monotone path argument, generalizing an earlier work by Courtade and Jiao.

Keywords: entropy power inequality; Karush–Kuhn–Tucker; Markov chain; mean squared error; semidefinite programming

1. Introduction

Shannon’s entropy power inequality (EPI) and its variants assert the optimality of the Gaussian solution to certain extremal problems. They play important roles in characterizing the information-theoretic limits of network information theory problems that involve Gaussian sources and/or channels (see, e.g., [1–15]). Many different approaches have been developed for proving such extremal results. Two notable ones are the doubling trick [16,17] and the monotone path argument [18–21]. Roughly speaking, the former reaches the desired conclusion by establishing the subadditivity of the relevant functional while the latter accomplishes its goal by constructing a monotone path with one end associated with an arbitrary given point in the feasible region and the other associated with the optimal solution to the Gaussian version of the problem. Though the doubling trick typically yields simpler proofs, the monotone path argument tends to be more informative. Indeed, it shows not only the existence of the Gaussian optimizer, but also the fact that every Karush–Kuhn–Tucker (KKT) point (also known as the stationary point) of the Gaussian version of the problem is in fact globally optimal. Such information is highly useful for numerical optimization.

Several years ago, inspired by the Gaussian two-terminal source coding problem, Courtade [22] conjectured the following EPI-type extremal result for long Markov chains.

Conjecture 1. Let \mathbf{X} and \mathbf{Z} be two independent n -dimensional zero-mean Gaussian random (column) vectors with covariance matrices $\Sigma_{\mathbf{X}} \succ \mathbf{0}$ and $\Sigma_{\mathbf{Z}} \succ \mathbf{0}$ respectively, and define $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Then, for any $\mu \geq 0$,

$$\inf_{U, V: U \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow V} I(\mathbf{X}; U) - \mu I(\mathbf{Y}; U) + I(\mathbf{Y}; V|U) - \mu I(\mathbf{X}; V|U) \quad (1)$$

has a Gaussian minimizer (i.e., the infimum is achieved by some (U, V) jointly Gaussian with (\mathbf{X}, \mathbf{Y})). Here, $U \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow V$ means $U, \mathbf{X}, \mathbf{Y}$, and V form a Markov chain in that order.

Later, Courtade and Jiao [23] proved this conjecture using the doubling trick. It is natural to ask whether this conjecture can also be proved via the monotone path argument. We shall show in this paper that it is indeed possible. Our work also sheds some light on the connection between these two proof strategies.

In fact, we shall prove a strengthened version of Conjecture 1 with some additional constraints imposed on (U, V) . For any random (column) vector \mathbf{S} and random object ω , let $\mathbf{D}_{\mathbf{S}|\omega}$ denote the distortion covariance matrix incurred by the minimum mean squared error (MMSE) estimator of \mathbf{S} from ω (i.e., $\mathbb{E}[(\mathbf{S} - \mathbb{E}[\mathbf{S}|\omega])(\mathbf{S} - \mathbb{E}[\mathbf{S}|\omega])^t]$), where $(\cdot)^t$ is the transpose operator. The main result of this paper is as follows.

Theorem 1. For any $\mu \geq 0$,

$$\inf_{\substack{U, V: U \leftrightarrow X \leftrightarrow Y \leftrightarrow V \\ (\mathbf{D}_{X|Y, U}, \mathbf{D}_{Y|X, V}) \in \mathcal{D}}} -\mu h(\mathbf{X}|U) + \mu h(\mathbf{Y}|U) + (\mu - 1)h(\mathbf{X}|U, V) - h(\mathbf{Y}|V) + h(\mathbf{X}|V) \tag{2}$$

has a Gaussian minimizer, where $\mathcal{D} = \{(\mathbf{D}_a, \mathbf{D}_b) : \mathbf{0} \prec \mathbf{D}_a \preceq \mathbf{D}_1, \mathbf{0} \prec \mathbf{D}_b \preceq \mathbf{D}_2\}$ with \mathbf{D}_1 and \mathbf{D}_2 satisfying $\mathbf{0} \prec \mathbf{D}_1 \preceq \mathbf{D}_{X|Y}$ and $\mathbf{0} \prec \mathbf{D}_2 \preceq \mathbf{D}_{Y|X}$.

Remark 1. The objective functions in (1) and (2) are equivalent. Indeed,

$$\begin{aligned} & I(\mathbf{X}; U) - \mu I(\mathbf{Y}; U) + I(\mathbf{Y}; V|U) - \mu I(\mathbf{X}; V|U) \\ & \approx -h(\mathbf{X}|U) + \mu h(\mathbf{Y}|U) + h(\mathbf{Y}|U) - h(\mathbf{Y}|U, V) - \mu h(\mathbf{X}|U) + \mu h(\mathbf{X}|U, V) \\ & = -(\mu + 1)h(\mathbf{X}|U) + (\mu + 1)h(\mathbf{Y}|U) - h(\mathbf{Y}|U, V) + \mu h(\mathbf{X}|U, V) + h(\mathbf{Y}|\mathbf{X}, U, V) - h(\mathbf{Y}|\mathbf{X}, V) \\ & = -(\mu + 1)h(\mathbf{X}|U) + (\mu + 1)h(\mathbf{Y}|U) - I(\mathbf{Y}; \mathbf{X}|U, V) + \mu h(\mathbf{X}|U, V) - h(\mathbf{Y}|\mathbf{X}, V) \\ & = -(\mu + 1)h(\mathbf{X}|U) + (\mu + 1)h(\mathbf{Y}|U) + (\mu - 1)h(\mathbf{X}|U, V) + h(\mathbf{X}|\mathbf{Y}, U) - h(\mathbf{Y}|\mathbf{X}, V) \\ & \approx -(\mu + 1)h(\mathbf{X}|U) + (\mu + 1)h(\mathbf{Y}|U) + (\mu - 1)h(\mathbf{X}|U, V) + h(\mathbf{X}|U) - h(\mathbf{Y}|U) - h(\mathbf{Y}|\mathbf{X}, V) \\ & \approx -\mu h(\mathbf{X}|U) + \mu h(\mathbf{Y}|U) + (\mu - 1)h(\mathbf{X}|U, V) - h(\mathbf{Y}|V) + h(\mathbf{X}|V), \end{aligned}$$

where “ \approx ” means that the two sides are equal up to an additive constant.

Remark 2. Conjecture 1 corresponds to the special case where $\mathbf{D}_1 = \mathbf{D}_{X|Y}$ and $\mathbf{D}_2 = \mathbf{D}_{Y|X}$.

The rest of the paper is organized as follows. Section 2 is devoted to the analysis of the Gaussian version of the optimization problem in (2). The key construction underlying our monotone path argument is introduced in Section 3. The proof of Theorem 1 is presented in Section 4. We conclude the paper in Section 5.

2. The Gaussian Version

In this section, we consider the Gaussian version of the optimization problem in (2) defined by imposing the restriction that (U, V) and (\mathbf{X}, \mathbf{Y}) are jointly Gaussian. Specifically, let $U^g = \mathbf{X} + \mathbf{N}_1$ and $V^g = \mathbf{Y} + \mathbf{N}_2$, where \mathbf{N}_1 and \mathbf{N}_2 are two independent n -dimensional zero-mean Gaussian random (column) vectors with covariance matrices $\mathbf{\Sigma}_1 \succ \mathbf{0}$ and $\mathbf{\Sigma}_2 \succ \mathbf{0}$, respectively. It is assumed that $(\mathbf{N}_1, \mathbf{N}_2)$ is independent of (\mathbf{X}, \mathbf{Y}) ; as a consequence, the Markov chain constraint $U^g \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow V^g$ is satisfied. Clearly,

$$\begin{aligned} & -\mu h(\mathbf{X}|U^g) + \mu h(\mathbf{Y}|U^g) + (\mu - 1)h(\mathbf{X}|U^g, V^g) - h(\mathbf{Y}|V^g) + h(\mathbf{X}|V^g) \\ & \approx -\mu h(\mathbf{X}|\mathbf{Y}, U^g) + (\mu - 1)h(\mathbf{X}|U^g, V^g) - h(\mathbf{Y}|\mathbf{X}, V^g) \\ & \approx -\frac{\mu}{2} \log |\mathbf{D}_{X|Y, U^g}| + \frac{\mu - 1}{2} \log |\mathbf{D}_{X|U^g, V^g}| - \frac{1}{2} \log |\mathbf{D}_{Y|X, V^g}|. \end{aligned}$$

Moreover,

$$D_{X|Y,U^g} = (\Sigma_X^{-1} + \Sigma_Z^{-1} + \Sigma_1^{-1})^{-1}, \tag{3}$$

$$D_{Y|X,V^g} = (\Sigma_Z^{-1} + \Sigma_2^{-1})^{-1}, \tag{4}$$

$$D_{X|U^g,V^g} = (\Sigma_X^{-1} + \Sigma_1^{-1} + (\Sigma_Z + \Sigma_2)^{-1})^{-1}. \tag{5}$$

Now, it is straightforward to write down the Gaussian version of the optimization problem in (2) with Σ_1 and Σ_2 as variables. However, as shown in the sequel, through a judicious change of variables, one can obtain an equivalent version that is more amenable to analysis.

Given $\lambda \in (0, 1)$, we introduce two random (column) vectors M_X and M_Y , independent of (X, Y, U^g, V^g) , such that the joint distribution of (M_X, M_Y) is the same as that of

$$\left(\sqrt{\frac{1-\lambda}{\lambda}}(X - \mathbb{E}[X|U^g, V^g]), -\sqrt{\frac{\lambda}{1-\lambda}}(Y - \mathbb{E}[Y|U^g, V^g]) \right).$$

Denote the covariance matrix of $(M_X^t, M_Y^t)^t$ by $\Sigma_{(M_X^t, M_Y^t)^t}$. We have

$$\begin{aligned} \Sigma_{(M_X^t, M_Y^t)^t}^{-1} &= \begin{pmatrix} \sqrt{\frac{\lambda}{1-\lambda}}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\sqrt{\frac{1-\lambda}{\lambda}}\mathbf{I} \end{pmatrix} (\Sigma_{(X^t, Y^t)^t}^{-1} + \Sigma_{(N_1^t, N_2^t)^t}^{-1}) \begin{pmatrix} \sqrt{\frac{\lambda}{1-\lambda}}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\sqrt{\frac{1-\lambda}{\lambda}}\mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\lambda}{1-\lambda}(\Sigma_X^{-1} + \Sigma_Z^{-1} + \Sigma_1^{-1}) & \Sigma_Z^{-1} \\ \Sigma_Z^{-1} & \frac{1-\lambda}{\lambda}(\Sigma_Z^{-1} + \Sigma_2^{-1}) \end{pmatrix}. \end{aligned}$$

Define $W_X = X + M_X$ and $W_Y = Y + M_Y$. Since

$$\begin{aligned} D_{(X^t, Y^t)^t|W_X, W_Y} &= (\Sigma_{(X^t, Y^t)^t}^{-1} + \Sigma_{M_X, M_Y}^{-1})^{-1} \\ &= \begin{pmatrix} \Sigma_X^{-1} + \Sigma_Z^{-1} + \frac{\lambda}{1-\lambda}(\Sigma_X^{-1} + \Sigma_Z^{-1} + \Sigma_1^{-1}) & \mathbf{0} \\ \mathbf{0} & \Sigma_Z^{-1} + \frac{1-\lambda}{\lambda}(\Sigma_Z^{-1} + \Sigma_2^{-1}) \end{pmatrix}^{-1}, \tag{6} \end{aligned}$$

X and Y must be conditionally independent given (W_X, W_Y) . It can be verified that

$$D_X^g := D_{X|W_X, W_Y, U^g} = (D_{X|W_X, W_Y}^{-1} + \Sigma_1^{-1})^{-1} = (1-\lambda)(\Sigma_X^{-1} + \Sigma_Z^{-1} + \Sigma_1^{-1})^{-1},$$

$$D_Y^g := D_{Y|W_X, W_Y, V^g} = (D_{Y|W_X, W_Y}^{-1} + \Sigma_2^{-1})^{-1} = \lambda(\Sigma_Z^{-1} + \Sigma_2^{-1})^{-1},$$

which implies

$$\Sigma_1 = ((1-\lambda)D_X^g - \Sigma_X^{-1} - \Sigma_Z^{-1})^{-1}, \tag{7}$$

$$\Sigma_2 = (\lambda(D_Y^g)^{-1} - \Sigma_Z^{-1})^{-1}. \tag{8}$$

Substituting (7) and (8) into (3–5) gives

$$D_{X|Y,U^g} = \frac{1}{1-\lambda} D_X^g, \tag{9}$$

$$D_{Y|X,V^g} = \frac{1}{\lambda} D_Y^g, \tag{10}$$

$$D_{X|U^g,V^g} = \frac{1}{1-\lambda} \left((D_X^g)^{-1} - \frac{1}{\lambda(1-\lambda)} \Sigma_Z^{-1} D_Y^g \Sigma_Z^{-1} \right)^{-1}. \tag{11}$$

Therefore, the Gaussian version of the optimization problem in (2) can be written as

$$\inf_{(\mathbf{D}_X^g, \mathbf{D}_Y^g) \in \mathcal{D}'} -\frac{\mu}{2} \log |\mathbf{D}_X^g| - \frac{\mu-1}{2} \log \left| (\mathbf{D}_X^g)^{-1} - \frac{1}{\lambda(1-\lambda)} \Sigma_Z^{-1} \mathbf{D}_Y^g \Sigma_Z^{-1} \right| - \frac{1}{2} \log |\mathbf{D}_Y^g|, \tag{12}$$

where $\mathcal{D}' = \{(\mathbf{D}_a, \mathbf{D}_b) : \mathbf{0} \prec \mathbf{D}_a \preceq (1-\lambda)\mathbf{D}_1, \mathbf{0} \prec \mathbf{D}_b \preceq \lambda\mathbf{D}_2\}$. It is clear that the infimum in (12) is achievable by some $(\mathbf{D}_X^*, \mathbf{D}_Y^*) \in \mathcal{D}'$. Moreover, such $(\mathbf{D}_X^*, \mathbf{D}_Y^*)$ must satisfy the following KKT conditions:

$$-\mu(\mathbf{D}_X^*)^{-1} + (\mu-1)(\mathbf{D}_X^*)^{-1} \left((\mathbf{D}_X^*)^{-1} - \frac{1}{\lambda(1-\lambda)} \Sigma_Z^{-1} \mathbf{D}_Y^* \Sigma_Z^{-1} \right)^{-1} (\mathbf{D}_X^*)^{-1} + \Pi_1 = \mathbf{0}, \tag{13}$$

$$\frac{(\mu-1)}{\lambda(1-\lambda)} \Sigma_Z^{-1} \left((\mathbf{D}_X^*)^{-1} - \frac{1}{\lambda(1-\lambda)} \Sigma_Z^{-1} \mathbf{D}_Y^* \Sigma_Z^{-1} \right)^{-1} \Sigma_Z^{-1} - (\mathbf{D}_Y^*)^{-1} + \Pi_2 = \mathbf{0}, \tag{14}$$

$$\Pi_1(\mathbf{D}_X^* - (1-\lambda)\mathbf{D}_1) = \mathbf{0}, \tag{15}$$

$$\Pi_2(\mathbf{D}_Y^* - \lambda\mathbf{D}_2) = \mathbf{0}, \tag{16}$$

where $\Pi_1, \Pi_2 \succeq \mathbf{0}$.

3. The Key Construction

Let $(\mathbf{X}^*, \mathbf{Y}^*)$ be an identically distributed copy of (\mathbf{X}, \mathbf{Y}) . Moreover, let \mathbf{N}_1^* and \mathbf{N}_2^* be two n -dimensional zero-mean Gaussian random (column) vectors with covariance matrices Σ_1^* and Σ_2^* , respectively. It is assumed that $(\mathbf{X}^*, \mathbf{Y}^*)$, \mathbf{N}_1^* , and \mathbf{N}_2^* are mutually independent. Define $U^* = \mathbf{X}^* + \mathbf{N}_1^*$ and $V^* = \mathbf{Y}^* + \mathbf{N}_2^*$. We choose Σ_1^* and Σ_2^* such that (cf. (9)–(11))

$$\mathbf{D}_{X^*|Y^*, U^*} = \frac{1}{1-\lambda} \mathbf{D}_X^*, \tag{17}$$

$$\mathbf{D}_{Y^*|X^*, V^*} = \frac{1}{\lambda} \mathbf{D}_Y^*, \tag{18}$$

$$\mathbf{D}_{X^*|U^*, V^*} = \frac{1}{1-\lambda} \left((\mathbf{D}_X^*)^{-1} - \frac{1}{\lambda(1-\lambda)} \Sigma_Z^{-1} \mathbf{D}_Y^* \Sigma_Z^{-1} \right)^{-1} \tag{19}$$

for some $(\mathbf{D}_X^*, \mathbf{D}_Y^*) \in \mathcal{D}'$ satisfying the KKT conditions (13)–(16).

Let U and V be two arbitrary random objects jointly distributed with (\mathbf{X}, \mathbf{Y}) such that $U \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow V$ and $(\mathbf{D}_{X|Y,U}, \mathbf{D}_{Y|X,V}) \in \mathcal{D}$. It is assumed that $(\mathbf{X}, \mathbf{Y}, U, V)$ and $(\mathbf{X}^*, \mathbf{Y}^*, U^*, V^*)$ are mutually independent. We aim to show that the objective function in (2) does not increase when $(\mathbf{X}, \mathbf{Y}, U, V)$ is replaced with $(\mathbf{X}^*, \mathbf{Y}^*, U^*, V^*)$, from which the desired result follows immediately. To this end, we develop a monotone path argument based on the following construction.

For $\lambda \in [0, 1]$, define

$$\begin{pmatrix} \mathbf{X}_\lambda \\ \tilde{\mathbf{X}}_\lambda \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda}\mathbf{I} & \sqrt{1-\lambda}\mathbf{I} \\ \sqrt{1-\lambda}\mathbf{I} & -\sqrt{\lambda}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{X}^* \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{Y}_\lambda \\ \tilde{\mathbf{Y}}_\lambda \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda}\mathbf{I} & \sqrt{1-\lambda}\mathbf{I} \\ \sqrt{1-\lambda}\mathbf{I} & -\sqrt{\lambda}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}^* \end{pmatrix}.$$

It is easy to verify the following Markov structures (see Figure 1a):

$$\left\{ \begin{array}{l} (U, U^*) \leftrightarrow (\mathbf{X}, \mathbf{X}^*) \leftrightarrow (\mathbf{X}_\lambda, \tilde{\mathbf{Y}}_\lambda) \leftrightarrow (\mathbf{Y}, \mathbf{Y}^*) \leftrightarrow (V, V^*), \\ (U, U^*) \leftrightarrow (\mathbf{X}, \mathbf{X}^*) \leftrightarrow (\tilde{\mathbf{X}}_\lambda, \mathbf{Y}_\lambda) \leftrightarrow (\mathbf{Y}, \mathbf{Y}^*) \leftrightarrow (V, V^*), \end{array} \right. \quad \lambda \in [0, 1]. \tag{20}$$

Note that, as λ changes from 0 to 1, \mathbf{X}_λ (\mathbf{Y}_λ) moves from \mathbf{X}^* (\mathbf{Y}^*) to \mathbf{X} (\mathbf{Y}) while $\tilde{\mathbf{X}}_\lambda$ ($\tilde{\mathbf{Y}}_\lambda$) moves the other way around. This construction generalizes its counterpart in the doubling trick, which corresponds to the special case $\lambda = \frac{1}{2}$.

For $\lambda \in (0, 1)$, define $\mathbf{W}_X^* = \mathbf{X} + \mathbf{M}_X^*$ and $\mathbf{W}_Y^* = \mathbf{Y} + \mathbf{M}_Y^*$, where

$$\mathbf{M}_X^* = \sqrt{\frac{1-\lambda}{\lambda}}(\mathbf{X}^* - \mathbb{E}[\mathbf{X}^*|U^*, V^*]),$$

$$\mathbf{M}_Y^* = -\sqrt{\frac{\lambda}{1-\lambda}}(\mathbf{Y}^* - \mathbb{E}[\mathbf{Y}^*|U^*, V^*]).$$

In view of (6), we have the following Markov structure (see Figure 1b):

$$U \leftrightarrow \mathbf{X} \leftrightarrow (\mathbf{W}_X^*, \mathbf{W}_Y^*) \leftrightarrow \mathbf{Y} \leftrightarrow V. \tag{21}$$

Define $\mathbf{D}_X = \mathbf{D}_{X|X_\lambda, \bar{Y}_\lambda, U, U^*}$ and $\mathbf{D}_Y = \mathbf{D}_{Y|X_\lambda, \bar{Y}_\lambda, V, V^*}$. It can be verified that for $\lambda \in (0, 1)$,

$$\mathbf{D}_X = \mathbf{D}_{X|X_\lambda, \bar{Y}_\lambda, U, U^*, V, V^*} \tag{22}$$

$$= \mathbf{D}_{X|\mathbf{W}_X^*, \mathbf{W}_Y^*, U, U^*, V, V^*} \tag{23}$$

$$= \mathbf{D}_{X|\mathbf{W}_X^*, \mathbf{W}_Y^*, U, V} \tag{24}$$

$$= \mathbf{D}_{X|\mathbf{W}_X^*, \mathbf{W}_Y^*, U} \tag{25}$$

where (22) is due to (20), (23) is due to the existence of a bijection between $(X_\lambda, \bar{Y}_\lambda, U^*, V^*)$ and $(\mathbf{W}_X^*, \mathbf{W}_Y^*, U^*, V^*)$, (24) is due to the fact that (U^*, V^*) is independent of $(\mathbf{X}, \mathbf{W}_X, \mathbf{W}_Y, U, V)$, and (25) is due to (21). Similarly, we have $\mathbf{D}_Y = \mathbf{D}_{Y|\mathbf{W}_X^*, \mathbf{W}_Y^*, V}$ for $\lambda \in (0, 1)$.

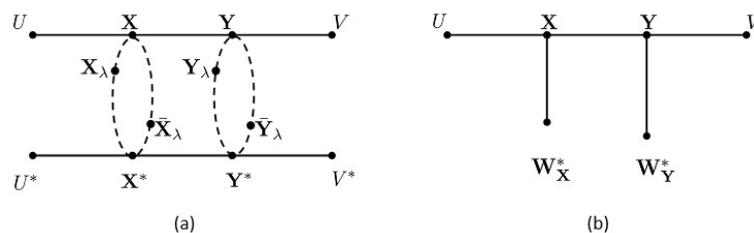


Figure 1. Illustrations of the Markov structures in (20) and (21).

4. Proof of Theorem 1

The following technical lemmas are needed for proving Theorem 1. Their proofs are relegated to the Appendices A–D.

Lemma 1. For $\lambda \in (0, 1)$,

$$\frac{d}{d\lambda} h(X_\lambda|U, U^*, V, V^*) = -\frac{n}{2(1-\lambda)} + \frac{1}{2(1-\lambda)^2} \text{tr} \left(\mathbf{D}_{X^*|U^*, V^*}^{-1} \mathbf{D}_{X|X_\lambda, U, U^*, V, V^*} \right), \tag{26}$$

$$\frac{d}{d\lambda} h(X_\lambda, \bar{Y}_\lambda|U, U^*) = -\frac{n}{2(1-\lambda)} + \frac{1}{2(1-\lambda)^2} \text{tr} \left(\mathbf{D}_{X^*|Y^*, U^*}^{-1} \mathbf{D}_X \right), \tag{27}$$

$$\frac{d}{d\lambda} h(X_\lambda, \bar{Y}_\lambda|V, V^*) = \frac{n}{2\lambda} - \frac{1}{2\lambda^2} \text{tr} \left(\mathbf{D}_{Y^*|X^*, V^*}^{-1} \mathbf{D}_Y \right). \tag{28}$$

Lemma 2. For $\lambda \in (0, 1)$,

$$\mathbf{D}_{X|X_\lambda, U, U^*, V, V^*} \succeq \Delta \succ \mathbf{0},$$

where

$$\Delta = \left(\mathbf{D}_X^{-1} - \frac{1}{(1-\lambda)^2} \Sigma_Z^{-1} \mathbf{D}_Y^* \left(\frac{1}{1-\lambda} \mathbf{D}_Y^* - \mathbf{D}_Y \right)^{-1} \mathbf{D}_Y^* \Sigma_Z^{-1} \right)^{-1}.$$

Lemma 3. $(\mathbf{B}^{-1} - \mathbf{A}^{-1})^{-1}$ is matrix convex in (\mathbf{A}, \mathbf{B}) for $\mathbf{A} \succ \mathbf{B} \succ \mathbf{0}$.

Lemma 4. Let \mathbf{X} be a Gaussian random vector and U be an arbitrary random object. Moreover, let \mathbf{N}_1 and \mathbf{N}_2 be two zero-mean Gaussian random vectors, independent of (\mathbf{X}, U) , with covariance matrices Σ_1 and Σ_2 respectively. If $\Sigma_2 \succ \Sigma_1 \succ \mathbf{0}$, then

$$\mathbf{D}_{\mathbf{X}|\mathbf{X}+\mathbf{N}_2,U} \succeq \left(\mathbf{D}_{\mathbf{X}|\mathbf{X}+\mathbf{N}_1,U} + \Sigma_2^{-1} - \Sigma_1^{-1} \right)^{-1}.$$

Now, we are in a position to prove Theorem 1. Define

$$h_\lambda = -\mu h(\bar{\mathbf{X}}_\lambda | \mathbf{X}_\lambda, U, U^*) + \mu h(\bar{\mathbf{Y}}_\lambda | \mathbf{X}_\lambda, U, U^*) + (\mu - 1)h(\bar{\mathbf{X}}_\lambda | \mathbf{X}_\lambda, U, U^*, V, V^*) - h(\bar{\mathbf{Y}}_\lambda | \mathbf{X}_\lambda, V, V^*) + h(\bar{\mathbf{X}}_\lambda | \mathbf{X}_\lambda, V, V^*).$$

Clearly,

$$h_\lambda|_{\lambda=0} = -\mu h(\mathbf{X}|U) + \mu h(\mathbf{Y}|U) + (\mu - 1)h(\mathbf{X}|U, V) - h(\mathbf{Y}|V) + h(\mathbf{X}|V),$$

$$h_\lambda|_{\lambda=1} = -\mu h(\mathbf{X}^*|U^*) + \mu h(\mathbf{Y}^*|U^*) + (\mu - 1)h(\mathbf{X}^*|U^*, V^*) - h(\mathbf{Y}^*|V^*) + h(\mathbf{X}^*|V^*).$$

Therefore, it suffices to show $\frac{dh_\lambda}{d\lambda} \leq 0$ for $\lambda \in (0, 1)$. Note that

$$h_\lambda = -\mu h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | U, U^*) + \mu h(\mathbf{X}_\lambda, \bar{\mathbf{Y}}_\lambda | U, U^*) + (\mu - 1)h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | U, U^*, V, V^*) - (\mu - 1)h(\mathbf{X}_\lambda | U, U^*, V, V^*) - h(\mathbf{X}_\lambda, \bar{\mathbf{Y}}_\lambda | V, V^*) + h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | V, V^*).$$

Since $h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | U, U^*) = h(\mathbf{X}, \mathbf{X}^* | U, U^*)$, $h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | U, U^*, V, V^*) = h(\mathbf{X}, \mathbf{X}^* | U, U^*, V, V^*)$, and $h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | V, V^*) = h(\mathbf{X}, \mathbf{X}^* | V, V^*)$, we have

$$\frac{d}{d\lambda} h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | U, U^*) = 0,$$

$$\frac{d}{d\lambda} h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | U, U^*, V, V^*) = 0,$$

$$\frac{d}{d\lambda} h(\mathbf{X}_\lambda, \bar{\mathbf{X}}_\lambda | V, V^*) = 0,$$

which, together with Lemma 1, implies

$$\begin{aligned} \frac{dh_\lambda}{d\lambda} &= -\frac{n}{2\lambda(1-\lambda)} + \frac{\mu}{2(1-\lambda)^2} \text{tr} \left(\mathbf{D}_{\mathbf{X}^*|\mathbf{Y}^*,U^*}^{-1} \mathbf{D}_X \right) + \frac{1}{2\lambda^2} \text{tr} \left(\mathbf{D}_{\mathbf{Y}^*|\mathbf{X}^*,V^*}^{-1} \mathbf{D}_Y \right) \\ &\quad - \frac{(\mu-1)}{2(1-\lambda)^2} \text{tr} \left(\mathbf{D}_{\mathbf{X}^*|U^*,V^*}^{-1} \mathbf{D}_{\mathbf{X}|\mathbf{X}_\lambda,U,U^*,V,V^*} \right). \end{aligned} \tag{29}$$

Combining Lemma 2 and (29) shows

$$\frac{dh_\lambda}{d\lambda} \leq -\frac{n}{2\lambda(1-\lambda)} - \frac{1}{2} f(\mathbf{D}_X, \mathbf{D}_Y), \tag{30}$$

where

$$f(\mathbf{D}_X, \mathbf{D}_Y) = -\frac{\mu}{(1-\lambda)^2} \text{tr} \left(\mathbf{D}_{X^*|Y^*,U^*}^{-1} \mathbf{D}_X \right) - \frac{1}{\lambda^2} \text{tr} \left(\mathbf{D}_{Y^*|X^*,V^*}^{-1} \mathbf{D}_Y \right) + \frac{(\mu-1)}{(1-\lambda)^2} \text{tr} \left(\mathbf{D}_{X^*|U^*,V^*}^{-1} \Delta \right).$$

We shall derive a lower bound for $f(\mathbf{D}_X, \mathbf{D}_Y)$. To this end, we first identify certain constraints on \mathbf{D}_X and \mathbf{D}_Y . Let $\mathbf{W} = \mathbf{X} + \tilde{\mathbf{Z}}$, where $\tilde{\mathbf{Z}}$ is a zero-mean Gaussian random vector, independent of (\mathbf{X}, U) , with covariance matrix $\Sigma_{\tilde{\mathbf{Z}}}$. We shall choose $\Sigma_{\tilde{\mathbf{Z}}}$ such that $\mathbf{D}_{X|W} = \mathbf{D}_{X|W_X^*, W_Y^*}$, which implies

$$\Sigma_{\tilde{\mathbf{Z}}} \prec \Sigma_Z, \tag{31}$$

$$\mathbf{D}_{X|W,U} = \mathbf{D}_{X|W_X^*, W_Y^*, U}. \tag{32}$$

It can be verified (cf. (6)) that

$$\begin{aligned} \Sigma_{\tilde{\mathbf{Z}}} &= \left(\Sigma_Z^{-1} + \frac{\lambda}{1-\lambda} (\Sigma_X^{-1} + \Sigma_Z^{-1} + (\Sigma_1^*)^{-1}) \right)^{-1} \\ &= \left(\Sigma_Z^{-1} + \lambda (\mathbf{D}_X^*)^{-1} \right)^{-1}. \end{aligned} \tag{33}$$

In view of (25), (31), (32), and Lemma 4,

$$\mathbf{D}_{X|Y,U} \succeq \left(\mathbf{D}_X^{-1} + \Sigma_Z^{-1} - \Sigma_{\tilde{\mathbf{Z}}}^{-1} \right)^{-1},$$

which, together with (33) and the constraint $\mathbf{0} \prec \mathbf{D}_{X|Y,U} \preceq \mathbf{D}_1$, implies

$$\mathbf{D}_X \preceq \left(\mathbf{D}_1^{-1} + \lambda (\mathbf{D}_X^*)^{-1} \right)^{-1}.$$

Similarly, we have

$$\mathbf{D}_Y \preceq \left(\mathbf{D}_2^{-1} + (1-\lambda) \mathbf{D}_Y^{*-1} \right)^{-1}. \tag{34}$$

Define $\bar{\mathbf{D}}_1 = \left(\mathbf{D}_1^{-1} + \lambda (\mathbf{D}_X^*)^{-1} \right)^{-1}$, $\bar{\mathbf{D}}_2 = \left(\mathbf{D}_2^{-1} + (1-\lambda) \mathbf{D}_Y^{*-1} \right)^{-1}$, and $\bar{\mathcal{D}} = \{(\mathbf{D}_a, \mathbf{D}_b) : \mathbf{0} \prec \mathbf{D}_a \preceq \bar{\mathbf{D}}_1, \mathbf{0} \prec \mathbf{D}_b \preceq \bar{\mathbf{D}}_2\}$. Consider

$$\min_{(\mathbf{D}_X, \mathbf{D}_Y) \in \bar{\mathcal{D}}} f(\mathbf{D}_X, \mathbf{D}_Y),$$

which is a convex semidefinite programming problem according to Lemma 3 (in view of (34) and Lemma 2, we have $\mathbf{D}_Y \prec \frac{1}{1-\lambda} \mathbf{D}_Y^*$ and $\Delta \succ \mathbf{0}$). Note that $(\mathbf{D}_X, \mathbf{D}_Y) \in \bar{\mathcal{D}}$ is an optimal solution to this convex programming problem if and only if it satisfies the following KKT conditions:

$$-\frac{\mu}{(1-\lambda)^2} \mathbf{D}_{X^*|Y^*,U^*}^{-1} + \frac{(\mu-1)}{(1-\lambda)^2} \left(\mathbf{D}_X^{-1} \Delta \mathbf{D}_{X^*|U^*,V^*}^{-1} \Delta \mathbf{D}_X^{-1} \right) + \Pi'_1 = \mathbf{0} \tag{35}$$

$$\begin{aligned} &-\frac{1}{\lambda^2} \mathbf{D}_{Y^*|X^*,V^*}^{-1} + \frac{(\mu-1)}{(1-\lambda)^2} \left((\mathbf{D}_Y^* - (1-\lambda) \mathbf{D}_Y)^{-1} \mathbf{D}_Y^* \Sigma_Z^{-1} \Delta \mathbf{D}_{X^*|U^*,V^*}^{-1} \Delta \Sigma_Z^{-1} \mathbf{D}_Y^* (\mathbf{D}_Y^* - (1-\lambda) \mathbf{D}_Y)^{-1} \right) \\ &+ \Pi'_2 = \mathbf{0}, \end{aligned} \tag{36}$$

$$\Pi'_1 (\mathbf{D}_X - \bar{\mathbf{D}}_1) = \mathbf{0}, \tag{37}$$

$$\Pi'_2 (\mathbf{D}_Y - \bar{\mathbf{D}}_2) = \mathbf{0}, \tag{38}$$

where $\Pi'_1, \Pi'_2 \succeq \mathbf{0}$. Let $\mathbf{D}_X = \mathbf{D}_X^*, \mathbf{D}_Y = \mathbf{D}_Y^*, \Pi'_1 = \frac{1}{1-\lambda}\Pi_1$, and $\Pi'_2 = \frac{1}{\lambda}\Pi_2$. One can readily verify by leveraging (17)–(19) that (35)–(38) with this particular choice of $(\mathbf{D}_X, \mathbf{D}_Y, \Pi'_1, \Pi'_2)$ can be deduced from (13)–(16); it is also easy to see that $(\mathbf{D}_X^*, \mathbf{D}_Y^*) \in \bar{\mathcal{D}}$. Therefore,

$$\min_{(\mathbf{D}_X, \mathbf{D}_Y) \in \bar{\mathcal{D}}} f(\mathbf{D}_X, \mathbf{D}_Y) = f(\mathbf{D}_X^*, \mathbf{D}_Y^*). \tag{39}$$

Combining (30), (39), and the fact that $f(\mathbf{D}_X^*, \mathbf{D}_Y^*) = -\frac{n}{\lambda(1-\lambda)}$ shows $\frac{dh_\lambda}{d\lambda} \leq 0$, which completes the proof.

5. Conclusions

We have generalized an extremal result by Courtade and Jiao. It is worth mentioning that recently Courtade [24] found a different generalization using the doubling trick. So far, we have not been able to prove this new result via the monotone path argument. A deeper understanding of the connection between these two methods is needed. It is conceivable that the convex-like property revealed by the monotone path argument and the subadditive property revealed by the doubling trick are manifestations of a common underlying mathematical structure yet to be uncovered.

Author Contributions: Methodology, J.W., J.C.; writing—original draft preparation, J.W.; writing—review and editing, J.C.

Funding: This research was funded by the National Science Foundation of China grant number 61771305.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Lemma 1

Note that

$$\begin{aligned} \frac{d}{d\lambda} h(\mathbf{X}_\lambda | U, U^*, V, V^*) &= \frac{d}{d\lambda} \left(\frac{n}{2} \log \lambda + \frac{d}{d\lambda} h(\mathbf{X} + \mathbf{M}_X^* | U, V) \right) \\ &= \frac{n}{2\lambda} + \frac{d}{d\lambda} h(\mathbf{X} + \mathbf{M}_X^* | U, V) \\ &= \frac{n}{2\lambda} + \text{tr} \left(\nabla_{\Sigma_{\mathbf{M}_X^*}} h(\mathbf{X} + \mathbf{M}_X^* | U, V) \nabla_\lambda \Sigma_{\mathbf{M}_X^*} \right) \\ &= \frac{n}{2\lambda} - \frac{1}{\lambda^2} \text{tr} \left(\mathbf{D}_{\mathbf{X}^* | U^*, V^*} \nabla_{\Sigma_{\mathbf{M}_X^*}} h(\mathbf{X} + \mathbf{M}_X^* | U, V) \right), \end{aligned} \tag{A1}$$

where $\Sigma_{\mathbf{M}_X^*} = \frac{1-\lambda}{\lambda} \mathbf{D}_{\mathbf{X}^* | U^*, V^*}$. We have

$$\nabla_{\Sigma_{\mathbf{M}_X^*}} h(\mathbf{X} + \mathbf{M}_X^* | U, V) = \frac{1}{2} \mathbf{J}(\mathbf{X} + \mathbf{M}_X^* | U, V) \tag{A2}$$

$$= \frac{\lambda}{2} \mathbf{J}(\mathbf{X}_\lambda | U, U^*, V, V^*), \tag{A3}$$

where $\mathbf{J}(\cdot)$ denotes the Fisher information matrix, and (A2) is due to ([25], Theorem 4). Substituting (A3) into (A1) gives

$$\frac{d}{d\lambda} h(\mathbf{X}_\lambda | U, U^*, V, V^*) = \frac{n}{2\lambda} - \frac{1}{2\lambda} \text{tr} \left(\mathbf{D}_{\mathbf{X}^* | U^*, V^*} \mathbf{J}(\mathbf{X}_\lambda | U, U^*, V, V^*) \right), \tag{A4}$$

which, together with the fact ([26], Lemma 2) that

$$\mathbf{J}(\mathbf{X}_\lambda | U, U^*, V, V^*) = \frac{1}{1-\lambda} \mathbf{D}_{\mathbf{X}^* | U^*, V^*}^{-1} - \frac{\lambda}{(1-\lambda)^2} \mathbf{D}_{\mathbf{X}^* | U^*, V^*}^{-1} \mathbf{D}_{\mathbf{X} | \mathbf{X}_\lambda, U, U^*, V, V^*} \mathbf{D}_{\mathbf{X}^* | U^*, V^*}^{-1} \tag{A5}$$

proves (26).

It remains to prove (27) since (28) follows by symmetry. Note that

$$\begin{pmatrix} \mathbf{X}_\lambda \\ \tilde{\mathbf{Y}}_\lambda \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sqrt{1-\lambda}\mathbf{I} \end{pmatrix} \left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} + \begin{pmatrix} \mathbf{N}_X \\ \mathbf{N}_Y \end{pmatrix} \right),$$

where $\mathbf{N}_X = \sqrt{\frac{1-\lambda}{\lambda}}\mathbf{X}^*$ and $\mathbf{N}_Y = -\sqrt{\frac{\lambda}{1-\lambda}}\mathbf{Y}^*$. Define

$$\begin{aligned} \bar{\mathbf{N}}_X &= \mathbf{N}_X - \mathbb{E}[\mathbf{N}_X|U^*], \\ \bar{\mathbf{N}}_Y &= \mathbf{N}_Y - \mathbb{E}[\mathbf{N}_Y|U^*]. \end{aligned}$$

Denote the covariance matrix of $(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t$ by $\Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}$. It can be verified (cf. (A4) and (A5)) that

$$\frac{d}{d\lambda}h(\mathbf{X}_\lambda, \tilde{\mathbf{Y}}_\lambda|U, U^*) = \frac{(1-2\lambda)n}{2\lambda(1-\lambda)} + \frac{1}{2}\text{tr} \left(\nabla_\lambda \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t} \mathbf{J}((\mathbf{X}^t, \mathbf{Y}^t)^t + (\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t|U) \right), \tag{A6}$$

$$\mathbf{J}((\mathbf{X}^t, \mathbf{Y}^t)^t + (\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t|U) = \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} - \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} \mathbf{D}_{(\mathbf{X}^t, \mathbf{Y}^t)^t|X_\lambda, \tilde{Y}_\lambda, U, U^*} \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1}. \tag{A7}$$

Since

$$\Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t} = \begin{pmatrix} \frac{1-\lambda}{\lambda} \mathbf{D}_{X^*|U^*} & -\mathbf{D}_{X^*|U^*} \\ -\mathbf{D}_{X^*|U^*} & \frac{\lambda}{1-\lambda} (\mathbf{D}_{X^*|U^*} + \Sigma_Z) \end{pmatrix},$$

we have

$$\begin{aligned} \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} &= \begin{pmatrix} \frac{\lambda}{1-\lambda} (\mathbf{D}_{X^*|U^*}^{-1} + \Sigma_Z^{-1}) & \Sigma_Z^{-1} \\ \Sigma_Z^{-1} & \frac{1-\lambda}{\lambda} \Sigma_Z^{-1} \end{pmatrix}, \\ \nabla_\lambda \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t} &= \begin{pmatrix} -\frac{1}{\lambda^2} \mathbf{D}_{X^*|U^*} & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\lambda)^2} (\mathbf{D}_{X^*|U^*} + \Sigma_Z) \end{pmatrix}, \end{aligned}$$

which further implies

$$\frac{1}{2}\text{tr} \left(\nabla_\lambda \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t} \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} \right) = 0. \tag{A8}$$

Combining (A6)–(A8) gives

$$\frac{d}{d\lambda}h(\mathbf{X}_\lambda, \tilde{\mathbf{Y}}_\lambda|U, U^*) = \frac{(1-2\lambda)n}{2\lambda(1-\lambda)} - \frac{1}{2}\text{tr} \left(\Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} \nabla_\lambda \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t} \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} \mathbf{D}_{(\mathbf{X}^t, \mathbf{Y}^t)^t|X_\lambda, \tilde{Y}_\lambda, U, U^*} \right). \tag{A9}$$

In view of (20),

$$\mathbf{D}_{(\mathbf{X}^t, \mathbf{Y}^t)^t|X_\lambda, \tilde{Y}_\lambda, U, U^*} = \begin{pmatrix} \mathbf{D}_X & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{Y|X_\lambda, \tilde{Y}_\lambda} \end{pmatrix}. \tag{A10}$$

Moreover,

$$\begin{aligned} \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} \nabla_\lambda \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t} \Sigma_{(\bar{\mathbf{N}}_X^t, \bar{\mathbf{N}}_Y^t)^t}^{-1} &= \begin{pmatrix} \frac{1}{(1-\lambda)^2} (\mathbf{D}_{X^*|U^*}^{-1} + \Sigma_Z^{-1}) & \mathbf{0} \\ \mathbf{0} & \frac{1}{\lambda^2} \Sigma_Z^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{(1-\lambda)^2} \mathbf{D}_{X^*|Y^*, U^*}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\lambda^2} \Sigma_Z^{-1} \end{pmatrix}. \end{aligned} \tag{A11}$$

Substituting (A10) and (A11) into (A9) yields

$$\frac{d}{d\lambda} h(\mathbf{X}_\lambda, \tilde{\mathbf{Y}}_\lambda | U, U^*) = \frac{(1 - 2\lambda)n}{2\lambda(1 - \lambda)} + \frac{1}{2(1 - \lambda)^2} \text{tr} \left(\mathbf{D}_{\mathbf{X}^* | \mathbf{Y}^*, U^*} \mathbf{D}_X \right) - \frac{1}{2\lambda^2} \text{tr} \left(\boldsymbol{\Sigma}_Z^{-1} \mathbf{D}_Y |_{\mathbf{X}_\lambda, \tilde{\mathbf{Y}}_\lambda} \right). \quad (\text{A12})$$

It can be verified that

$$\begin{aligned} \mathbf{D}_{(\mathbf{X}^t, \mathbf{Y}^t)^t | \mathbf{X}_\lambda, \tilde{\mathbf{Y}}_\lambda} &= \mathbf{D}_{(\mathbf{X}^t, \mathbf{Y}^t)^t | \mathbf{X} + \mathbf{N}_X, \mathbf{Y} + \mathbf{N}_Y} \\ &= \left(\boldsymbol{\Sigma}_{(\mathbf{X}^t, \mathbf{Y}^t)^t}^{-1} + \boldsymbol{\Sigma}_{(\mathbf{N}_X^t, \mathbf{N}_Y^t)^t}^{-1} \right)^{-1} \\ &= \begin{pmatrix} (1 - \lambda)(\boldsymbol{\Sigma}_X^{-1} + \boldsymbol{\Sigma}_Z^{-1})^{-1} & \mathbf{0} \\ \mathbf{0} & \lambda \boldsymbol{\Sigma}_Z \end{pmatrix}, \end{aligned}$$

which implies

$$\mathbf{D}_{Y | \mathbf{X}_\lambda, \tilde{\mathbf{Y}}_\lambda} = \lambda \boldsymbol{\Sigma}_Z. \quad (\text{A13})$$

Substituting (A13) into (A12) proves (27).

Appendix B. Proof of Lemma 2

It can be verified that $\mathbb{E}[\mathbf{W}_Y^* | \mathbf{X}, \mathbf{Y}, \mathbf{W}_X^*] = \mathbf{A}\mathbf{X} + \mathbf{Y} - \mathbf{A}\mathbf{W}_X^*$, where $\mathbf{A} = \frac{1}{1-\lambda} \mathbf{D}_Y^* \boldsymbol{\Sigma}_Z^{-1}$; moreover, $\mathbf{Q} := \mathbf{W}_Y^* - \mathbb{E}[\mathbf{W}_Y^* | \mathbf{X}, \mathbf{Y}, \mathbf{W}_X^*]$ is a zero-mean Gaussian random vector with covariance matrix $\boldsymbol{\Sigma}_Q = \frac{1}{1-\lambda} \mathbf{D}_Y^*$, and is independent of $(\mathbf{X}, \mathbf{Y}, \mathbf{W}_X^*, U, V)$. Define $\tilde{\mathbf{Y}} = \mathbf{A}\mathbf{X} + \mathbf{Y}$. We have

$$\begin{aligned} \mathbf{D}_{(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t | \mathbf{W}_X^*, \mathbf{W}_Y^*, U, V} &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A} & \mathbf{I} \end{pmatrix} \mathbf{D}_{(\mathbf{X}^t, \mathbf{Y}^t)^t | \mathbf{W}_X^*, \mathbf{W}_Y^*, U, V} \begin{pmatrix} \mathbf{I} & \mathbf{A}^t \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{D}_X & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_Y \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}^t \\ \mathbf{0} & \mathbf{I} \end{pmatrix}. \end{aligned} \quad (\text{A14})$$

On the other hand,

$$\begin{aligned} \mathbf{D}_{(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t | \mathbf{W}_X^*, \mathbf{W}_Y^*, U, V} &= \mathbf{D}_{(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t | \mathbf{W}_X^*, U, V, \tilde{\mathbf{Y}} + \mathbf{Q}} \\ &\preceq \left(\mathbf{D}_{(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t | \mathbf{W}_X^*, U, V} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_Q^{-1} \end{pmatrix} \right)^{-1}, \end{aligned} \quad (\text{A15})$$

where (A15) is due to the fact that the distortion covariance matrix incurred by the MMSE estimator of $(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t$ from $(\mathbf{W}_X^*, U, V, \tilde{\mathbf{Y}} + \mathbf{Q})$ is no greater than (in the semidefinite sense) that incurred by the linear MMSE estimator of $(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t$ from $(\mathbb{E}[(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t | \mathbf{W}_X^*, U, V], \tilde{\mathbf{Y}} + \mathbf{Q})$. Combining (A14) and (A15) yields

$$\mathbf{D}_{(\mathbf{X}^t, \tilde{\mathbf{Y}}^t)^t | \mathbf{W}_X^*, U, V} \succeq \begin{pmatrix} \mathbf{D}_X^{-1} + \mathbf{A}^t \mathbf{D}_Y^{-1} \mathbf{A} & -\mathbf{A}^t \mathbf{D}_Y^{-1} \\ -\mathbf{D}_Y^{-1} \mathbf{A} & \mathbf{D}_Y^{-1} - \boldsymbol{\Sigma}_Q^{-1} \end{pmatrix}^{-1}.$$

Comparing the upper-left submatrices on the two sides of the above inequality gives $\mathbf{D}_{X | \mathbf{W}_X^*, U, V} \succeq \boldsymbol{\Delta}$, which, together with the fact that $\mathbf{D}_{X | \mathbf{X}_\lambda, U, U^*, V, V^*} = \mathbf{D}_{X | \mathbf{W}_X^*, U, V}$, proves $\mathbf{D}_{X | \mathbf{X}_\lambda, U, U^*, V, V^*} \succeq \boldsymbol{\Delta}$. Moreover, we have

$$\mathbf{0} \prec \mathbf{D}_{(\mathbf{X}^t, \mathbf{Y}^t)^t | (\mathbf{W}_X, U, V)} \preceq \begin{pmatrix} \mathbf{D}_X^{-1} + \mathbf{A}^t \mathbf{D}_Y^{-1} \mathbf{A} & -\mathbf{A}^t \mathbf{D}_Y^{-1} \\ -\mathbf{D}_Y^{-1} \mathbf{A} & \mathbf{D}_Y^{-1} - \boldsymbol{\Sigma}_Q^{-1} \end{pmatrix},$$

where the second inequality is due to (A14) and (A15). Therefore, Δ , which is equal to the inverse of the upper-left submatrix of

$$\begin{pmatrix} \mathbf{D}_X^{-1} + \mathbf{A}^t \mathbf{D}_Y^{-1} \mathbf{A} & -\mathbf{A}^t \mathbf{D}_Y^{-1} \\ -\mathbf{D}_Y^{-1} \mathbf{A} & \mathbf{D}_Y^{-1} - \Sigma_Q^{-1} \end{pmatrix}^{-1},$$

must be positive definite.

Appendix C. Proof of Lemma 3

It is known [27] that $\mathbf{B}\mathbf{S}^{-1}\mathbf{B}$ is matrix convex in (\mathbf{S}, \mathbf{B}) for symmetric matrix \mathbf{B} and positive definite matrix \mathbf{S} . The desired conclusion follows from the fact that $(\mathbf{B}^{-1} - \mathbf{A}^{-1})^{-1} = \mathbf{B} + \mathbf{B}(\mathbf{A} - \mathbf{B})^{-1}\mathbf{B}$ and that affine transformations preserve convexity.

Appendix D. Proof of Lemma 4

Without loss of generality, we assume that $\mathbf{N}_2 = \mathbf{N}_1 + \tilde{\mathbf{N}}$, where $\tilde{\mathbf{N}}$ is a zero-mean Gaussian random vector with covariance matrix $\tilde{\Sigma} = \Sigma_2 - \Sigma_1$ and is independent of $(\mathbf{X}, \mathbf{N}_1, U)$. As a consequence, $\mathbb{E}[\mathbf{N}_1|\mathbf{N}_2] = \mathbf{H}\mathbf{N}_2$, where $\mathbf{H} = (\Sigma_1^{-1} + \tilde{\Sigma}^{-1})^{-1}\tilde{\Sigma}^{-1}$; moreover, $\mathbf{M} := \mathbf{N}_1 - \mathbb{E}[\mathbf{N}_1|\mathbf{N}_2]$ is a zero-mean Gaussian random vector with covariance matrix $\Sigma_M = (\Sigma_1^{-1} + \tilde{\Sigma}^{-1})^{-1}$, and is independent of $(\mathbf{X}, \mathbf{N}_2, U)$. Note that

$$\begin{aligned} \mathbf{J}(\mathbf{X} + \mathbf{N}_1|\mathbf{X} + \mathbf{N}_2, U) &= \mathbf{J}((\mathbf{I} - \mathbf{H})\mathbf{X} + \mathbf{M}|\mathbf{X} + \mathbf{N}_2, U) \\ &= (\Sigma_1^{-1} + \tilde{\Sigma}^{-1}) - \Sigma_1^{-1} \mathbf{D}_{\mathbf{X}|\mathbf{X}+\mathbf{N}_1, U} \Sigma_1^{-1}, \end{aligned} \quad (\text{A16})$$

where the second equality follows by ([26], Lemma 2). On the other hand,

$$\mathbf{J}(\mathbf{X} + \mathbf{N}_1|\mathbf{X} + \mathbf{N}_2, U) \succeq \mathbf{D}_{\mathbf{X}+\mathbf{N}_1|\mathbf{X}+\mathbf{N}_2, U}^{-1} \quad (\text{A17})$$

$$\begin{aligned} &= \mathbf{D}_{(\mathbf{I}-\mathbf{H})\mathbf{X}+\mathbf{M}|\mathbf{X}+\mathbf{N}_2, U}^{-1} \\ &= \left((\mathbf{I} - \mathbf{H}) \mathbf{D}_{\mathbf{X}|\mathbf{X}+\mathbf{N}_2, U} (\mathbf{I} - \mathbf{H}^t) + \Sigma_M \right)^{-1}, \end{aligned} \quad (\text{A18})$$

where (A17) is due to the conditional version of the Cramer–Rao inequality. Combining (A16) and (A18) yields the desired result.

References

1. Bergmans, P. A simple converse for broadcast channels with additive white Gaussian noise (corresp.). *IEEE Trans. Inf. Theory* **1974**, *20*, 279–280. [CrossRef]
2. Weingarten, H.; Steinberg, Y.; Shamai, S. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inf. Theory* **2006**, *52*, 3936–3964. [CrossRef]
3. Prabhakaran, V.; Tse, D.; Ramchandran, K. Rate region of the quadratic Gaussian CEO problem. In Proceedings of the IEEE International Symposium on Information Theory, (ISIT), Chicago, IL, USA, 27 June–2 July 2004; p. 117.
4. Oohama, Y. Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder. *IEEE Trans. Inf. Theory* **2005**, *51*, 2577–2593. [CrossRef]
5. Wang, J.; Chen, J. Vector Gaussian two-terminal source coding. *IEEE Trans. Inf. Theory* **2013**, *59*, 3693–3708. [CrossRef]
6. Wang, J.; Chen, J. Vector Gaussian multiterminal source coding. *IEEE Trans. Inf. Theory* **2014**, *60*, 5533–5552. [CrossRef]
7. Liu, T.; Shamai, S. A note on the secrecy capacity of the multiple-antenna wiretap channel. *IEEE Trans. Inf. Theory* **2009**, *55*, 2547–2553. [CrossRef]
8. Chen, J. Rate region of Gaussian multiple description coding with individual and central distortion constraints. *IEEE Trans. Inf. Theory* **2009**, *55*, 3991–4005. [CrossRef]

9. Motahari, A.S.; Khandani, A.K. Capacity bounds for the Gaussian interference channel. *IEEE Trans. Inf. Theory* **2009**, *55*, 620–643. [[CrossRef](#)]
10. Shang, X.; Kramer, G.; Chen, B. A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels. *IEEE Trans. Inf. Theory* **2009**, *55*, 689–699. [[CrossRef](#)]
11. Annapureddy, V.S.; Veeravalli, V.V. Gaussian interference networks: Sum capacity in the low interference regime and new outer bounds on the capacity region. *IEEE Trans. Inf. Theory* **2009**, *55*, 3032–3050. [[CrossRef](#)]
12. Song, L.; Chen, J.; Wang, J.; Liu, T. Gaussian robust sequential and predictive coding. *IEEE Trans. Inf. Theory* **2013**, *59*, 3635–3652. [[CrossRef](#)]
13. Song, L.; Chen, J.; Tian, C. Broadcasting correlated vector Gaussians. *IEEE Trans. Inf. Theory* **2015**, *61*, 2465–2477. [[CrossRef](#)]
14. Khezeli, K.; Chen, J. A source-channel separation theorem with application to the source broadcast problem. *IEEE Trans. Inf. Theory* **2016**, *62*, 1764–1781. [[CrossRef](#)]
15. Tian, C.; Chen, J.; Diggavi, S.N.; Shamai S. Matched multiuser Gaussian source channel communications via uncoded schemes. *IEEE Trans. Inf. Theory* **2017**, *63*, 4155–4171. [[CrossRef](#)]
16. Geng, Y.; Nair, C. The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *IEEE Trans. Inf. Theory* **2014**, *60*, 2087–2104. [[CrossRef](#)]
17. Geng, Y.; Nair, C. Reflection and summary of the paper: The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *IEEE IT Soc. Newlett.* **2017**, *67*, 5–7.
18. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **1959**, *2*, 101–112. [[CrossRef](#)]
19. Blachman, N.M. The convolution inequality for entropy powers. *IEEE Trans. Inf. Theory* **1965**, *11*, 267–271. [[CrossRef](#)]
20. Dembo, A.; Cover, T.M.; Thomas, J.A. Information theoretic inequalities. *IEEE Trans. Inf. Theory* **1991**, *37*, 1501–1518. [[CrossRef](#)]
21. Liu, T.; Viswanath, P. An extremal inequality motivated by multiterminal information-theoretic problems. *IEEE Trans. Inf. Theory* **2007**, *53*, 1839–1851. [[CrossRef](#)]
22. Courtade, T.A. An Extremal Conjecture: Experimenting With Online Collaboration. *Information Theory b-Log*, 5 March 2013. Available online: <http://blogs.princeton.edu/blogit/2013/03/05/> (accessed on 15 March 2013).
23. Courtade, T.A.; Jiao, J. An extremal inequality for long Markov chains. *arXiv* **2014**, arXiv:1404.6984v1.
24. Courtade, T.A. A strong entropy power inequality. *IEEE Trans. Inf. Theory* **2018**, *64*, 2173–2192. [[CrossRef](#)]
25. Palomar, D.P.; Verdú, S. Gradient of mutual information in linear vector Gaussian channels. *IEEE Trans. Inf. Theory* **2006**, *52*, 141–154. [[CrossRef](#)]
26. Zhou, Y.; Xu, Y.; Yu, W.; Chen, J. On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks. *IEEE Trans. Inf. Theory* **2016**, *62*, 7402–7418. [[CrossRef](#)]
27. Marshall, A.W.; Olkin, I. *Inequalities: Theory of Majorization and Its Applications*; Academic Press: New York, NY, USA, 1979.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).