An Information Inequality Motivated by the Gaussian Z-Interference Channel

Amin Gohari, Chandra Nair, David Ng

Abstract

We establish an information inequality that is motivated by the capacity region computation for the Gaussian Z-interference channel. This yields an improved slope for the capacity region at Costa's corner point. We believe the inequality may also be of independent interest as it provides a non-trivial upper bound on the entropy of sums of independent random variables.

1 Introduction

A scalar Gaussian Z-interference channel is modeled as

$$Y_1 = X_1 + Z_1, \tag{1}$$
$$Y_2 = X_2 + aX_1 + Z_2 \tag{2}$$

where X_1, X_2 correspond to the two transmitted symbols, $Z_1, Z_2 \sim \mathcal{N}(0,1)$ are standard Gaussians, and Y_1, Y_2 correspond to the two received symbols. We impose average power constraints P_1, P_2 on X_1, X_2 respectively. The definition of the achievable rate region and the capacity region is standard in literature and can be found in Chapter 6 of [1]. The capacity region is known when $a \geq 1$ [2], which is also known as the strong interference regime. Determining the capacity region of the scalar Gaussian Z-interference channel for $a < 1$ has been an unresolved central open problem in the area of multi-user information theory.

It is known that capacity region of the Gaussian Z-interference channel has two extreme (*corner*) points:

- *Sato's corner point* [2]:

$$(R_1, R_2) = \left(\frac{1}{2}\log(1+P_1), \frac{1}{2}\log\left(1 + \frac{P_2}{1+a^2P_1}\right)\right),$$

- *Costa's corner point* [3], [4]:

$$(R_1, R_2) = \left(\frac{1}{2}\log\left(1 + \frac{a^2P_1}{1+P_2}\right), \frac{1}{2}\log(1+P_2)\right).$$

The latter extreme point, which is also the point of interest in this paper, has had a rich history in this field. Costa [5] developed the celebrated *concavity of entropy power* result as an intermediate step towards establishing the extremality of the latter point. However Sason [6] observed a gap in one of Costa's results (Lemma 1 of [3]) in the finishing part of the proof. In [4] the authors replaced the use of Pinsker's inequality by Talagrands HWI inequality [7] along with certain continuity arguments to fix the proof of Lemma 1 in [3] and thus establish the extremality of *Costa's corner point*. The argument in [4] also provided an outer bound to the capacity region of the Gaussian Z-interference channel that had the property that: if $R_2 \geq C_2 - \epsilon$, then $R_1 \leq \frac{1}{2}\log\left(1 + \frac{a^2P_1}{1+P_2}\right) - O(\sqrt{\epsilon})$. This non-linear trade-off between R_1 and R_2 is unavoidable (see Remark 2 [4]) if one follows the proof idea to establish the extremality of the corner point, pioneered by Costa [3]. This implies that for no finite value of λ will the supporting hyperplane (to the outer bound) of the form $R_1 + \lambda R_2$ pass through Costa's corner point.

On the other hand, for the Gaussian Z-interference channel, the optimality or sub-optimality of the Han-Kobayashi's rate region (with Gaussian signaling) [8] is not yet determined. Further it has been established in [9] that for every $\lambda \geq \lambda_0$, the supporting hyperplane (to the Han-Kobayashi achievable rate region with Gaussian signaling) of the form $R_1 + \lambda R_2$ passes through Costa's corner point, where

$$\lambda_0 = 1 + \max\left\{\frac{-\log a^2 - \frac{1-a^2}{(1+a^2P_1+P_2)}}{\log(1+P_2) - \frac{P_2}{1+P_2}}, \frac{(1-a^2)(1+P_2)}{a^2P_2}\right\}.$$

Motivated by this inconsistent behaviour of the outer and inner bounds near the corner point, [10] (specifically near the corner point) and [11] (for the optimality of the entire capacity region) studied the multi-letter extension of Han-Kobayashi region (whose limit is the capacity region) for the Gaussian Z-Interference channel. In particular the following conjecture was made in [11].

Conjecture 1. *For $\mu \geq 0$, $N_2 \geq 0$ and $\Sigma_1, A_2 \succeq 0$, the maximum*

$$\max_{\substack{p(\mathbf{x}_1)p(\mathbf{x}_2) \\ \mathbb{E}[\mathbf{X}_2\mathbf{X}_2^T] \preceq A_2}} \Big[\mu h(\mathbf{X}_2 + \mathbf{X}_1 + \mathbf{Z}_1 + \mathbf{Z}_2) + h(\mathbf{X}_1 + \mathbf{Z}_1) - (\mu+1)h(\mathbf{X}_1 + \mathbf{Z}_1 + \mathbf{Z}_2) - \mathrm{tr}(\Sigma_1 \mathbb{E}[\mathbf{X}_1\mathbf{X}_1^T])\Big]$$

where $\mathbf{Z}_1 \sim \mathcal{N}(0, I)$, $\mathbf{Z}_2 \sim \mathcal{N}(0, N_2 I)$ and $\mathbf{X}_i, \mathbf{Z}_i$ $(i = 1, 2)$ are random variables in $\mathbb{R}^k$ $(k \geq 1)$, is attained by Gaussian $\mathbf{X}_1$ and $\mathbf{X}_2$.

Further it was shown that if this conjecture was true, then Han-Kobayashi achievable region with Gaussian signaling would equal the capacity region. In this paper, we establish the above conjecture for large enough μ (see Theorem 1), $A_2 = P_2 I$, and any choice of Σ_1.

In a related development [12] established the extremality of Costa's corner point by traditional arguments of deriving outer bounds in information theory. Expanding on that [13] showed that the supporting hyperplane (to a new outer bound developed in [13]) of the form $R_1 + \lambda R_2$ passes through Costa's corner point, whenever $\lambda \geq \lambda_1$, where

$$\lambda_1 = 1 + \begin{cases} \frac{(1+P_2)(1-a^2)}{a^2 P_2} \frac{\left(1+\sqrt{1+4a^2(1-a^2)P_2}\right)^2}{4a^2(1-a^2)P_2} & a^2 < \frac{1}{2} \\ \frac{(1+P_2)(1-a^2)}{a^2 P_2} \frac{\left(1+\sqrt{1+P_2}\right)^2}{P_2} & a^2 \geq \frac{1}{2} \end{cases}.$$

Thus we have a linear-trade off between R_1 and R_2 around the corner point, giving rise to a behavior similar to that predicted by the Han-Kobayashi inner bound. However the outer-bound developed in [13] does not directly imply Conjecture 1 for any choice of parameters. The result in this paper is inspired and related to the arguments used in [13] for the interference channel.

2 Main

The main result in this paper is the following:

Theorem 1. *Let X_1^n, X_2^n, Z_1^n, Z_2^n be mutually independent random variables in $\mathbb{R}^n$ $(n \geq 1)$ with $Z_1^n \sim \mathcal{N}(0, N_1 I)$ and $Z_2^n \sim \mathcal{N}(0, N_2 I)$, where $N_1, N_2 > 0$. Suppose*

$$\mathbb{E}[X_1^n] = \mathbb{E}[X_2^n] = 0, \qquad \mathbb{E}[\|X_1^n\|^2], \mathbb{E}[\|X_2^n\|^2] < \infty.$$

Then for any

$$\mu \geq \frac{N_2}{N_1} \frac{1}{\left(1 - \sqrt{\frac{N_1+N_2}{\frac{\mathbb{E}[\|X_2^n\|^2]}{n} + N_1 + N_2}}\right)^2} =: \mu_1$$

we have

$$\begin{aligned} &\mu h(X_2^n + X_1^n + Z_1^n + Z_2^n) + h(X_1^n + Z_1^n) - (\mu+1)h(X_1^n + Z_1^n + Z_2^n) \\ &\leq \frac{n}{2}\left[\mu \log\left(\frac{\mathbb{E}[\|X_2^n\|^2]}{n} + N_1 + N_2\right) + \log N_1 - (\mu+1)\log(N_1+N_2)\right]. \end{aligned}$$

Remark 1. *The following points are worth noting:*

i) The inequality is tight when $X_2^n \sim \mathcal{N}(0, P_2 I)$ and $X_1^n = 0$. Hence it establishes Conjecture 1 when $\mu \geq \mu_1$ and $A_2 = P_2 I$.

ii) The inequality does not hold for

$$\mu < \frac{N_2}{N_1} \frac{1}{\left(1 + \frac{N_1+N_2}{\frac{\mathbb{E}[\|X_2^n\|^2]}{n}}\right)}.$$

This can be seen by setting $X_2^n \sim \mathcal{N}(0, P_2 I)$ and $X_1^n = \epsilon I$, and taking the derivative of the left-hand-side of the inequality with respect to ϵ at $\epsilon = 0$.

Let $\mathcal{M}$ denote the set of distributions $p(x_1^n, x_2^n, \hat{v}, \hat{w})$ such that

$$X_1^n \to (\hat{V}, \hat{W}) \to X_2^n, \qquad \hat{W} \to \hat{V} \to X_2^n \tag{3}$$

form Markov chains and $X_1^n \perp (\hat{V}, X_2^n)$. The alphabet sets of $\hat{V}, \hat{W}$ are assumed to be arbitrary. Observe that by taking $\hat{V}, \hat{W}$ to be constants, any pair of independent $X_1^n, X_2^n \in \mathcal{M}$. Let $\mu \geq 0$ and $0 \leq \alpha \leq 1$. For independent X_1^n, X_2^n we upper bound the terms in Theorem 1 as follows:

$$\begin{aligned}
&\mu h(X_2^n + X_1^n + Z_1^n + Z_2^n) + h(X_1^n + Z_1^n) - (\mu+1) h(X_1^n + Z_1^n + Z_2^n) \\
&= \mu I(X_2^n; X_2^n + X_1^n + Z_1^n + Z_2^n) + h(X_1^n + Z_1^n) - h(X_1^n + Z_1^n + Z_2^n | X_2^n) \\
&\overset{(a)}{\leq} \mu\alpha I(X_2^n; X_2^n + Z_1^n + Z_2^n) + \mu(1-\alpha)\big((h(X_2^n + X_1^n + Z_1^n + Z_2^n) - h(X_1^n + Z_1^n)\big) \\
&\quad + (1 + \mu(1-\alpha))\big(h(X_1^n + Z_1^n | X_2^n) - h(X_1^n + Z_1^n + Z_2^n | X_2^n)\big) \\
&\leq \sup_{\substack{p(\hat{v},\hat{w}|x_1^n,x_2^n) \\ p(x_1^n,x_2^n,\hat{v},\hat{w}) \in \mathcal{M}}} \Big[\mu\alpha I(X_2^n; X_2^n + Z_1^n + Z_2^n | \hat{V}) + \mu(1-\alpha)\big((h(X_2^n + X_1^n + Z_1^n + Z_2^n | \hat{W}) - h(X_1^n + Z_1^n | \hat{W})\big) \\
&\quad + (1 + \mu(1-\alpha))\big(h(X_1^n + Z_1^n | X_2^n, \hat{W}) - h(X_2^n + X_1^n + Z_1^n + Z_2^n | X_2^n, \hat{W})\big)\Big]
\end{aligned} \tag{4}$$

where (a) follows from data-processing inequality. For any $p(x_1, x_2)$ define the functional (and its natural extension to vectors):

$$\Theta_{\mu,\alpha}(X_1, X_2) := \sup_{\substack{p(\hat{v},\hat{w}|x_1,x_2) \\ p(x_1,x_2,\hat{v},\hat{w}) \in \mathcal{M}}} \theta_{\mu,\alpha}(X_1, X_2 | \hat{V}, \hat{W}) \tag{5}$$

where

$$\begin{aligned}
\theta_{\mu,\alpha}(X_1, X_2 | \hat{V}, \hat{W}) := \mu\alpha \cdot I(X_2; X_2 + Z_1 + Z_2 | \hat{V}) \\
+ \mu(1-\alpha) \cdot \big[h(X_2 + X_1 + Z_1 + Z_2 | \hat{W}) - h(X_1 + Z_1 | \hat{W})\big] \\
+ (1 + \mu(1-\alpha)) \cdot \big[h(X_1 + Z_1 | X_2, \hat{W}) - h(X_1 + Z_1 + Z_2 | X_2, \hat{W})\big]
\end{aligned} \tag{6}$$

We will show that $\Theta_{\mu,\alpha}(X_1, X_2)$ is sub-additive, i.e.

$$\Theta_{\mu,\alpha}(X_1^n, X_2^n) \leq \sum_{i=1}^n \Theta_{\mu,\alpha}(X_{1i}, X_{2i}).$$

We will use this to show that subject to an upper bound on $\mathbb{E}[\|X_1^n\|^2]$, $\mathbb{E}[\|X_2^n\|^2]$ (even under more relaxed conditions), the functional $\Theta_{\mu,\alpha}(X_1, X_2)$ is maximized by Gaussians. Finally, we will optimize within the space of Gaussian distributions and arrive at a proof of Theorem 1 (see Section 3 for more details).

However, it is easier to establish the Gaussian optimality (Proposition 5) for a perturbed functional $\Theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n)$ and then take the limit of the perturbations (Lemma 2). This is for a technical reason: the perturbed functional has a *strict sub-additivity* property that allows one to conclude that Gaussians are the only optimizers, building on the ideas in [14]. We believe that some of these techniques, the tailoring of the perturbations for instance, developed here (and in [13]) can be of independent interest.

2.1 Applications of the Inequality

2.1.1 Slope at Costa's corner point

We now show how Theorem 1 yields a slope to the capacity region at Costa's corner point. The argument is essentially same as Lemma 4 in [10], which was one of the starting points of the study of the functional considered in this paper.

Proposition 1. *For any $\lambda \geq \lambda_2$, where*

$$\lambda_2 = 1 + \frac{(1+P_2)(1-a^2)}{a^2 P_2} \frac{\left(1 + \sqrt{1+P_2}\right)^2}{P_2},$$

the supporting hyperplane of the form $R_1 + \lambda R_2$ to the capacity region of the Gaussian Z-interference channel (defined by equations (1)-(2)) passes through Costa's corner point.

Proof. Costa's corner point is in the Han-Kobyashi achievable region. Hence it suffices to establish that for any achievable rate pair we must have

$$R_1 + \lambda R_2 \leq \frac{1}{2}\log\left(1 + \frac{a^2 P_1}{1+P_2}\right) + \frac{\lambda}{2}\log(1+P_2).$$

Consider an equivalent formulation, [3], of the Z-interference channel

$$\begin{aligned} Y_1 &= X_1 + Z_1, \\ Y_2 &= X_2 + X_1 + Z_1 + Z_2 \end{aligned}$$

where $Z_1 \sim \mathcal{N}(0,1)$, $Z_2 \sim \mathcal{N}(0, N_2)$ (where $N_2 := \frac{1}{a^2} - 1$ and $0 < a < 1$) and X_i, Y_i, Z_i $(i = 1, 2)$ are random variables in $\mathbb{R}$, under the average power constraints P_1 on X_1, and $\frac{1}{a^2}P_2$ on X_2 where $P_1, P_2 \geq 0$. From Fano's inequality, any sequence of codebooks of rate (R_1, R_2), whose probability of error goes to zero, must satisfy

$$\begin{aligned} & R_1 + \lambda R_2 - \epsilon_n \\ &\leq \frac{1}{n}\Big[I(X_1^n;Y_1^n) + \lambda I(X_2^n;Y_2^n)\Big] \\ &= \frac{1}{n}\Big[h(Y_2^n) - h(Y_1^n|X_1^n) + (\lambda-1)h(Y_2^n) + h(Y_1^n) - \lambda h(Y_2^n|X_2^n)\Big] \\ &= \frac{1}{n}\Big[h(X_2^n + X_1^n + Z_1^n + Z_2^n) - h(Z_1^n) + (\lambda-1)h(X_2^n + X_1^n + Z_1^n + Z_2^n) + h(X_1^n + Z_1^n) - \lambda h(X_1^n + Z_1^n + Z_2^n)\Big] \\ &\overset{(a)}{\leq} \frac{1}{2}\Big[\log\left(\frac{P_2}{a^2} + P_1 + \frac{1}{a^2}\right) + (\lambda-1)\log\left(\frac{P_2}{a^2} + \frac{1}{a^2}\right) - \lambda\log\left(\frac{1}{a^2}\right)\Big] \\ &= \frac{1}{2}\Big[\log(1 + a^2P_1 + P_2) + (\lambda-1)\log(1+P_2)\Big]. \end{aligned}$$

Here (a) follows from upper bounding $h(X_2^n + X_1^n + Z_1^n + Z_2^n)$ with the value of a Gaussian with the same power, and the latter terms using Theorem 1, where we use $\lambda \geq \lambda_2$. □

Remark 2. *Since Costa's corner point is achievable, there is a sequence of (random) codebooks whose value attains equality in the limit. This implies that inequality marked (a) in the proof above, must be an equality for the codebooks. To get the equality for the term $h(X_2^n + X_1^n + Z_1^n + Z_2^n)$, we need $X_2^n \sim \mathcal{N}\left(0, \frac{P_2}{a^2}I\right)$ and $X_1^n \sim \mathcal{N}(0, P_1 I)$. However equality in the latter term occurs when $X_2^n \sim \mathcal{N}\left(0, \frac{P_2}{a^2}I\right)$ and $X_1^n = 0$. Together this implies that distributions of codebooks come arbitrarily close to achieve the equality in Theorem 1 (when the expressions are normalized by n) asymptotically as n tends to infinity.*

Remark 3. *This result improves the bound given in [13] for the slope at the corner for the regime $a^2 < \frac{1}{2}$.*

2.1.2 Upper bound on entropy of sum

Another application of the inequality is to apply it to independent variables.

Proposition 2. *Let X_1^n and X_2^n be independent variables in $\mathbb{R}^n$ satisfying $\max\left\{\mathbb{E}[\|X_1^n\|^2], \mathbb{E}[\|X_2^n\|^2]\right\} \leq nP$. Then, for all*

$$\mu \geq \frac{N_2}{N_1}\frac{\sqrt{P+N_2+N_1}}{(\sqrt{P+N_1+N_2} - \sqrt{N_1+N_2})^2}$$

we obtain the following inequality:

$$\begin{aligned} \mu h(X_2^n + X_1^n + Z_1^n + Z_2^n) \leq{}& \frac{\mu+1}{2}\big(h(X_1^n + Z_1^n + Z_2^n) + h(X_2^n + Z_1^n + Z_2^n)\big) - \frac{1}{2}\big(h(X_1^n + Z_1^n) + h(X_2^n + Z_1^n)\big) \\ &+ \frac{n\mu}{2}\log\left(1 + \frac{P}{N_1+N_2}\right) - \frac{n}{2}\log\left(1 + \frac{N_2}{N_1}\right). \end{aligned}$$

Proof. The proof immediately follows from two applications of Theorem 1 by interchanging X_1^n and X_2^n. □

Remark 4. *The celebrated EPI yields a lower bound to the entropy of the sum, while the above inequality yields an upper bound to the entropy of sum of two independent random variables.*

3 Proof Outlines

For $i = 1, .., n$ define

$$Y_{1i} := X_{1i} + Z_{1i},$$
$$T_{2i} := X_{2i} + Z_{1i} + Z_{2i},$$
$$Y_{2i} := X_{2i} + X_{1i} + Z_{1i} + Z_{2i} = T_{2i} + X_{1i},$$

Proof of Theorem 1. For any $0 \le \alpha \le 1$ we have

$$\mu h(Y_2^n) + h(Y_1^n) - (\mu+1)h(Y_2^n|X_2^n) \overset{(a)}{\le} \Theta_{\mu,\alpha}(X_1^n, X_2^n) \le G^{(n)}\left(\frac{\mathbb{E}[\|X_1^n\|^2]}{n}, \frac{\mathbb{E}[\|X_2^n\|^2]}{n}\right)$$

where (a) follows from (4) and (5), and $G^{(n)}(P_1, P_2)$ in the last step is defined as in Proposition 3 below. Since α is arbitrary, we can take infimum over α and we get the desired result by Proposition 3 (iv). □

Proposition 3. *For $P_1, P_2 \ge 0$ denote*

$$G^{(n)}(P_1, P_2) := \sup_{\substack{p_1(x_1^n)p_2(x_2^n): \\ \mathbb{E}[X_1^n]=\mathbb{E}[X_2^n]=0 \\ \mathbb{E}[\|X_1^n\|^2]\le nP_1,\, \mathbb{E}[\|X_2^n\|^2]\le nP_2}} \Theta_{\mu,\alpha}(X_1^n, X_2^n),$$

where $\Theta_{\mu,\alpha}(X_1^n, X_2^n)$ is defined in (5). Let

$$\tilde{g}_{\mu,\alpha}(A_2, B_1, B_2, \Sigma) := \frac{1}{2}\Big[\mu\alpha \cdot \big[\log(A_2 + N_1 + N_2) - \log(N_1 + N_2)\big] + \mu(1-\alpha)\cdot\log(B_1 + B_2 + 2\Sigma + A_2 + N_1 + N_2) + \log(B_1 + N_1) - (1 + \mu(1-\alpha))\cdot\log(B_1 + N_1 + N_2)\Big].$$

Then we have the following:

(i) $G^{(n)}(P_1, P_2) = n \cdot G^{(1)}(P_1, P_2)$.

(ii) For $P_1, P_2 > 0$,

$$G^{(n)}(P_1, P_2) \le n \cdot \sup_{\substack{B_1, B_2 \ge 0 \\ \frac{B_1}{P_1} + \frac{B_2}{P_2} \le 1}} \tilde{g}_{\mu,\alpha}(P_2 - \frac{P_1 B_2}{P_1 - B_1}, B_1, B_2, \sqrt{B_1 B_2})$$

where $\frac{0}{0}$ is understood to be 0.

(iii) If $\mu \ge \mu_0$, where

$$\mu_0 := \frac{N_2}{N_1}\frac{1}{(1 - \sqrt{\frac{N_1+N_2}{P_2+N_1+N_2}})^2}$$

then there exists $0 \le \alpha \le 1$ such that the maximization on right hand side of (ii) is attained by $B_1 = B_2 = 0$.

(iv) If $\mu \ge \mu_0$ where μ_0 is defined as in (iii), then

$$\inf_{0\le\alpha\le1} G^{(n)}(P_1, P_2) \le \frac{n}{2}\big[\mu\log(P_2 + N_1 + N_2) + \log N_1 - (\mu+1)\log(N_1 + N_2)\big].$$

Proof. The rest of the paper yields an outline and preliminary results that are needed to prove the proposition. See Appendix F to see how the results mentioned later combine to yield a proof. □

3.1 Sub-additive functionals

Let $\mu \ge 0$ and $0 \le \alpha \le 1$. We shall now establish a sub-additivity result with relatively minimal assumptions on the random variables so that we can apply it for different settings later on. We also establish it for random vectors so that we may be able to utilize the results in future attempts to establish Conjecture 1.

Proposition 4. *Let $\epsilon \in \mathbb{R}$. Let $\mu \geq 0$ and $0 \leq \alpha \leq 1$. Let $\mathbf{X}_{2i}, \mathbf{T}_{2i}, \mathbf{Y}_{1i}, \mathbf{Y}_{2i}$ $(i = 1, \dots, n)$ be arbitrary real-vector-valued random variables and $\hat{V}, \hat{W}$ be any random variables. Define the functional*

$$\begin{aligned}\varphi_{\mu,\alpha}^{(\epsilon)}(\mathbf{X}_2^n, \mathbf{T}_2^n, \mathbf{Y}_1^n, \mathbf{Y}_2^n|\hat{V}, \hat{W}) := \mu\alpha \cdot I(\mathbf{X}_2^n; \mathbf{T}_2^n|\hat{V}) + \mu(1-\alpha) \cdot \big[h(\mathbf{Y}_2^n|\hat{W}) - h(\mathbf{Y}_1^n|\hat{W})\big] \\ + (1 + \mu(1-\alpha)) \cdot \big[h(\mathbf{Y}_1^n|\mathbf{X}_2^n, \hat{W}) - h(\mathbf{Y}_2^n|\mathbf{X}_2^n, \hat{W})\big] \\ - \epsilon\big[I(\mathbf{X}_2^n; \mathbf{T}_2^n + \mathbf{G}^n|\hat{V}) + h(\mathbf{Y}_1^n|\hat{W}) + h(\mathbf{Y}_1^n|\hat{V}, \hat{W})\big]\end{aligned} \tag{7}$$

where $\mathbf{G}_i$'s are independent Gaussian random variables of suitable dimension with identity covariance matrix. Suppose

$$(\mathbf{X}_2^{n\setminus i}, \mathbf{T}_2^{i-1}) \to (\mathbf{X}_{2i}, \hat{V}) \to \mathbf{T}_{2i},$$
$$(\mathbf{X}_2^{n\setminus i}, \mathbf{Y}_{1(i+1)}^n, \mathbf{Y}_2^{i-1}) \to (\mathbf{Y}_{1i}, \mathbf{X}_{2i}, \hat{W}) \to \mathbf{Y}_{2i}$$

form Markov chains. Then we have

$$\begin{aligned}&\varphi_{\mu,\alpha}^{(\epsilon)}(\mathbf{X}_2^n, \mathbf{T}_2^n, \mathbf{Y}_1^n, \mathbf{Y}_2^n|\hat{V}, \hat{W}) \\ &= \sum_{i=1}^n \varphi_{\mu,\alpha}^{(\epsilon)}(\mathbf{X}_{2i}, \mathbf{T}_{2i}, \mathbf{Y}_{1i}, \mathbf{Y}_{2i}|V_i, W_i) - \sum_{i=1}^n \Big[(1 + \mu(1-\alpha)) I(\mathbf{Y}_{1i}; \mathbf{X}_2^{n\setminus i}|\mathbf{X}_{2i}, \mathbf{Y}_{2i}, W_i) \\ &\quad + \epsilon I(\mathbf{T}_2^{i-1}; \mathbf{T}_{2i} + \mathbf{G}_i|\mathbf{T}_2^{i-1} + \mathbf{G}^{i-1}, \hat{V}) + \epsilon I(\mathbf{Y}_2^{i-1}; \mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{W}) + \epsilon I(\mathbf{T}_2^{i-1}, \mathbf{Y}_2^{i-1}; \mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{V}, \hat{W})\Big]\end{aligned}$$

where $V_i := (\hat{V}, \mathbf{T}_2^{i-1})$ and $W_i := (\hat{W}, \mathbf{Y}_{1(i+1)}^n, \mathbf{Y}_2^{i-1})$. Consequently, from the non-negativity of mutual information, we have

$$\varphi_{\mu,\alpha}^{(\epsilon)}(\mathbf{X}_2^n, \mathbf{T}_2^n, \mathbf{Y}_1^n, \mathbf{Y}_2^n|\hat{V}, \hat{W}) \leq \sum_{i=1}^n \varphi_{\mu,\alpha}^{(\epsilon)}(\mathbf{X}_{2i}, \mathbf{T}_{2i}, \mathbf{Y}_{1i}, \mathbf{Y}_{2i}|V_i, W_i).$$

Proof. See Appendix A. □

In order to understand the definition of $\varphi_{\mu,\alpha}^{(\epsilon)}$, one can compare it with the definition of $\theta_{\mu,\alpha}$ in (6). Besides an extra ϵ term, $\varphi_{\mu,\alpha}^{(\epsilon)}$ is defined for arbitrary $\mathbf{X}_{2i}, \mathbf{T}_{2i}, \mathbf{Y}_{1i}, \mathbf{Y}_{2i}$ $(i = 1, \dots, n)$ while $\theta_{\mu,\alpha}$ assumes an additive structure on $\mathbf{T}_{2i}, \mathbf{Y}_{1i}, \mathbf{Y}_{2i}$. This structure is required for the subsequent results and given formally in the next subsection.

Notation

We shall use the following notations unless otherwise specified. Let $\mathbf{X}_{1i}$, $\mathbf{X}_{2i}$ $(i = 1, \dots, n)$ be random variables in $\mathbb{R}^d$ $(d \geq 1)$ and let $\hat{V}, \hat{W}$ be random variables on arbitrary alphabet sets. The joint distribution of $(\mathbf{X}_1^n, \mathbf{X}_2^n, \hat{V}, \hat{W})$ is also arbitrary. Let

$$\begin{aligned}\mathbf{Y}_{1i} &:= \mathbf{X}_{1i} + \mathbf{Z}_{1i}, \\ \mathbf{T}_{2i} &:= \mathbf{X}_{2i} + \mathbf{Z}_{1i} + \mathbf{Z}_{2i}, \\ \mathbf{Y}_{2i} &:= \mathbf{X}_{2i} + \mathbf{X}_{1i} + \mathbf{Z}_{1i} + \mathbf{Z}_{2i} = \mathbf{T}_{2i} + \mathbf{X}_{1i}.\end{aligned}$$

where

$$\mathbf{Z}_{1i} \sim \mathcal{N}(0, N_1 I), \qquad \mathbf{Z}_{2i} \sim \mathcal{N}(0, N_2 I)$$

are i.i.d. Gaussian random variables in $\mathbb{R}^d$ for $i = 1, \dots, n$, mutually independent of each other and of $(\mathbf{X}_1^n, \mathbf{X}_2^n, \hat{V}, \hat{W})$. We assume that $N_1, N_2 > 0$. Note that with this definition, $\varphi_{\mu,\alpha}^{(0)}$ as defined in (7) equals $\varphi_{\mu,\alpha}$ as defined in (6).

For $\delta \in \mathbb{R}$ let

$$\tilde{\mathbf{T}}_{2i}^{(\delta)} := \begin{pmatrix} 0 & I \\ 0 & \delta I \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1i} \\ \mathbf{X}_{2i} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_{1i} \\ \hat{\mathbf{Z}}_{1i} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_{2i} \\ \hat{\mathbf{Z}}_{2i} \end{pmatrix}, \qquad \tilde{\mathbf{Y}}_{2i}^{(\delta)} := \begin{pmatrix} I & I \\ 0 & \delta I \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1i} \\ \mathbf{X}_{2i} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_{1i} \\ \hat{\mathbf{Z}}_{1i} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_{2i} \\ \hat{\mathbf{Z}}_{2i} \end{pmatrix}$$

where

$$\hat{\mathbf{Z}}_{1i} \sim \mathcal{N}(0, \hat{N}_1 I), \qquad \hat{\mathbf{Z}}_{2i} \sim \mathcal{N}(0, \hat{N}_2 I)$$

are i.i.d. Gaussian random variables in $\mathbb{R}^d$ for $i = 1, \dots, n$, mutually independent of each other and of $(\mathbf{Z}_1^n, \mathbf{Z}_2^n, \mathbf{X}_1^n, \mathbf{X}_2^n, \hat{V}, \hat{W})$. We assume that $\hat{N}_1, \hat{N}_2 > 0$. Observe that $\tilde{\mathbf{T}}_{2i}^{(\delta)}$ and $\tilde{\mathbf{Y}}_{2i}^{(\delta)}$ are just perturbed versions of $\mathbf{T}_{2i}$ and $\mathbf{Y}_{2i}$. Define the *single-letterization* variables

$$V_i := (\hat{V}, (\mathbf{T}_2)_1^{i-1}), \qquad W_i := (\hat{W}, (\mathbf{Y}_1)_{i+1}^n, (\mathbf{Y}_2)_1^{i-1}),$$

and for $\delta \in \mathbb{R}$ their perturbed versions

$$\tilde{V}_i^{(\delta)} := (\hat{V}, (\tilde{\mathbf{T}}_2^{(\delta)})_1^{i-1}), \qquad \tilde{W}_i^{(\delta)} := (\hat{W}, (\mathbf{Y}_1)_{i+1}^n, (\tilde{\mathbf{Y}}_2^{(\delta)})_1^{i-1}).$$

Lemma 1. *(i) If*

$$(\mathbf{X}_1)_1^n \to (\hat{V}, \hat{W}) \to (\mathbf{X}_2)_1^n, \qquad \hat{W} \to \hat{V} \to (\mathbf{X}_2)_1^n$$

form Markov chains, then

$$\mathbf{X}_{1i} \to (\tilde{V}_i^{(\delta)}, \tilde{W}_i^{(\delta)}) \to \mathbf{X}_{2i}, \qquad \tilde{W}_i^{(\delta)} \to \tilde{V}_i^{(\delta)} \to \mathbf{X}_{2i}$$

form Markov chains.

(ii) If $(\mathbf{X}_1)_1^n \perp (\hat{V}, (\mathbf{X}_2)_1^n)$ *then* $\mathbf{X}_{1i} \perp (\tilde{V}_i^{(\delta)}, \mathbf{X}_{2i})$.

Proof. See Appendix B. □

With the notation as above, for any $\epsilon, \delta \in \mathbb{R}$, define the functional

$$\theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n | \hat{V}, \hat{W}) := \varphi_{\mu,\alpha}^{(\epsilon)}((\mathbf{X}_2)_1^n, (\tilde{\mathbf{T}}_2^{(\delta)})_1^n, (\mathbf{Y}_1)_1^n, (\tilde{\mathbf{Y}}_2^{(\delta)})_1^n | \hat{V}, \hat{W})$$

for any $p(\mathbf{x}_1^n, \mathbf{x}_2^n, \hat{v}, \hat{w})$, and the functional

$$\Theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n) := \sup_{\substack{p(\hat{v},\hat{w}|\mathbf{x}_1^n,\mathbf{x}_2^n): \\ p(\mathbf{x}_1^n,\mathbf{x}_2^n,\hat{v},\hat{w}) \in \mathcal{M}}} \theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n | \hat{V}, \hat{W})$$

for any $p(\mathbf{x}_1^n, \mathbf{x}_2^n)$, where $\mathcal{M}$ is as defined (analogously for vector versions) in (3). For $\delta = 0$ the quantities $\theta_{\mu,\alpha}^{(\epsilon,\delta)}$ and $\varphi_{\mu,\alpha}^{(\epsilon)}$ differ by only a constant

$$\theta_{\mu,\alpha}^{(\epsilon,0)}(\mathbf{X}_1^n, \mathbf{X}_2^n | \hat{V}, \hat{W}) = \varphi_{\mu,\alpha}^{(\epsilon)}(\mathbf{X}_2^n, \mathbf{T}_2^n, \mathbf{Y}_1^n, \mathbf{Y}_2^n | \hat{V}, \hat{W}) - h(\hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n).$$

Thus $\theta_{\mu,\alpha}^{(\epsilon,\delta)}$ can be viewed as a perturbation of $\varphi_{\mu,\alpha}^{(\epsilon)}$ via δ, and also as a perturbation of $\varphi_{\mu,\alpha}$ via both ϵ and δ.

Note that Proposition 4 together with Lemma 1 immediately establishes the following corollary, i.e. $\Theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1, \mathbf{X}_2)$ is sub-additive.

Corollary 1. *We have the following sub-additivity:*

$$\Theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n) \leq \sum_{i=1}^n \Theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_{1i}, \mathbf{X}_{2i}).$$

The next proposition establishes that some of the conditional distributions for any optimizer of the perturbed functional must be Gaussians.

Proposition 5. *Let* $\mathcal{K}_{1i}, \mathcal{K}_{2i}$ $(i = 1, \ldots, n)$ *be sets of* $d \times d$ *matrices. Denote*

$$v_{(\mathcal{K}_{11},\mathcal{K}_{21}),\ldots,(\mathcal{K}_{1n},\mathcal{K}_{2n})}^{(n)} := \sup_{\substack{p_1(\mathbf{x}_1^n)p_2(\mathbf{x}_2^n) \\ \mathbb{E}[\mathbf{X}_1^n]=\mathbb{E}[\mathbf{X}_2^n]=0 \\ \mathrm{Cov}(\mathbf{X}_{1i})\in\mathcal{K}_{1i},\, \mathrm{Cov}(\mathbf{X}_{2i})\in\mathcal{K}_{2i}}} \Theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n).$$

Then we have the following:

(i) Let $\epsilon, \delta \in \mathbb{R}$. *Then*

$$\sum_{i=1}^n v_{(\mathcal{K}_{1i},\mathcal{K}_{2i})}^{(1)} = v_{(\mathcal{K}_{11},\mathcal{K}_{21}),\ldots,(\mathcal{K}_{1n},\mathcal{K}_{2n})}^{(n)}.$$

(ii) Let $\epsilon > 0$ *and* $\delta \neq 0$, *or let* $\epsilon = \delta = 0$. *Suppose* $\mathcal{K}_1, \mathcal{K}_2$ *are compact convex sets of* $d \times d$ *matrices. Then there exists a maximizer* $p^*(\mathbf{x}_1, \mathbf{x}_2, v, w)$ *for* $v_{(\mathcal{K}_1,\mathcal{K}_2)}^{(1)}$ *such that* $p^*(\mathbf{x}_1|w)$, $p^*(\mathbf{x}_1|v, w)$ *and* $p^*(\mathbf{x}_2|v)$ *are Gaussians with covariance matrices independent of choice of* v *and* w.

Proof. See Appendix C. □

The function $v_{(\mathcal{K}_1,\mathcal{K}_2)}^{(1)}$ does not have a simple form and it is difficult to establish its concavity directly. In the full version we use an indirect rotation and sub-additivity based argument to show its concavity in $\mathcal{K}_1$ and $\mathcal{K}_2$.

The next Lemma shows that the perturbed functional converges to the original functional.

Lemma 2. *Let $n \geq 1$. Let $\mathcal{D}$ be any set of distributions $p(\mathbf{x}_1^n, \mathbf{x}_2^n)$ such that $\mathbb{E}[\|\mathbf{X}_1^n\|^2]$ and $\mathbb{E}[\|\mathbf{X}_2^n\|^2]$ are bounded. Then*

$$\lim_{(\epsilon,\delta)\to(0,0)} \sup_{p(\mathbf{x}_1^n,\mathbf{x}_2^n)\in\mathcal{D}} \Theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n) = \sup_{p(\mathbf{x}_1^n,\mathbf{x}_2^n)\in\mathcal{D}} \Theta_{\mu,\alpha}^{(0,0)}(\mathbf{X}_1^n, \mathbf{X}_2^n).$$

Proof. See Appendix D. □

The next theorem finally establishes the Gaussian extremality.

Theorem 2. *Let $\mathcal{K}_1, \mathcal{K}_2$ be compact convex sets of $d \times d$ matrices. Let $\Theta_{\mu,\alpha}(\mathbf{X}_1, \mathbf{X}_2)$ be as defined (analogously for vectors) in (5). Then*

$$\sup_{\substack{p_1(\mathbf{x}_1)p_2(\mathbf{x}_2)\\ \mathbb{E}[\mathbf{X}_1]=\mathbb{E}[\mathbf{X}_2]=0\\ \mathrm{Cov}(\mathbf{X}_1)\in\mathcal{K}_1,\, \mathrm{Cov}(\mathbf{X}_2)\in\mathcal{K}_2}} \Theta_{\mu,\alpha}(\mathbf{X}_1, \mathbf{X}_2) = \sup_{\substack{A_2,B_1,B_2,C_1,C_2,\Sigma\in\mathbb{R}^{d\times d}\\ A_2,\begin{pmatrix} B_1 & \Sigma \\ \Sigma^T & B_2\end{pmatrix},\begin{pmatrix} C_1 & -\Sigma \\ -\Sigma^T & C_2\end{pmatrix}\succeq 0\\ B_1+C_1\in\mathcal{K}_1,\, A_2+B_2+C_2\in\mathcal{K}_2}} g_{\mu,\alpha}(A_2, B_1, B_2, \Sigma)$$

where

$$\begin{aligned} & g_{\mu,\alpha}(A_2, B_1, B_2, \Sigma) \\ := & \frac{1}{2}\Big[\mu\alpha \cdot \big[\log|A_2 + N_1 I + N_2 I| - \log|N_1 I + N_2 I|\big] \\ & + \mu(1-\alpha)\cdot\big[\log|B_1 + B_2 + \Sigma + \Sigma^T + A_2 + N_1 I + N_2 I| - \log|B_1 + N_1 I|\big] \\ & + (1+\mu(1-\alpha))\cdot\big[\log|B_1 - \Sigma(A_2+B_2)^{-1}\Sigma^T + N_1 I| - \log|B_1 - \Sigma(A_2+B_2)^{-1}\Sigma^T + N_1 I + N_2 I|\big]\Big]. \end{aligned}$$

Proof. See Appendix E . □

Remark 5. *The result in Theorem 2 is used to establish Proposition 3. Take $d = 1$ and $\mathcal{K}_1$ and $\mathcal{K}_2$ to be the set of non-negative numbers bounded by P_1 and P_2 respectively. Then, $g_{\mu,\alpha}(A_2, B_1, B_2, \Sigma)$ equals $G^{(1)}(P_1, P_2)$ as defined in Proposition 3. The expression $\tilde{g}_{\mu,\alpha}(A_2, B_1, B_2, \Sigma)$ used in Proposition 3 is just an easier-to-analyze upper bound of $g_{\mu,\alpha}(A_2, B_1, B_2, \Sigma)$. Therefore by a more careful analysis, it is possible to get slightly stronger results than that stated in Proposition 3.*

4 Conclusion

In this paper we establish an information inequality by proving Gaussian optimality for an information functional. This establishes, in some regimes, a conjecture that would imply the optimality of Han-Kobayashi region for the Gaussian Z-interference channel. The inequality provides an improved slope for the capacity region at the Costa's corner point for some parameter regimes. It also provides an upper bound on the entropy of sums of random variables. There are also some perturbation and other techniques which may be of independent interest.

References

[1] A. El Gamal and Y.-H. Kim, *Network Information Theory.* Cambridge University Press, 2012.

[2] H. Sato, "The capacity of the gaussian interference channel under strong interference (corresp.)," *IEEE Transactions on Information Theory*, vol. 27, no. 6, pp. 786–788, Nov 1981.

[3] M. H. M. Costa, "On the Gaussian interference channel," *Information Theory, IEEE Transactions on*, vol. 31, no. 5, pp. 607–615, Sep 1985.

[4] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *CoRR*, vol. abs/1504.04419, 2015. [Online]. Available: http://arxiv.org/abs/1504.04419

[5] M. Costa, "A new entropy power inequality," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 751–760, 1985.

[6] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. IT-50, pp. 1345–1356, Mar. 2004.

[7] M. Talagrand, "Transportation cost for gaussian and other product measures," *Geometric & Functional Analysis GAFA*, vol. 6, no. 3, pp. 587–600, May 1996. [Online]. Available: https://doi.org/10.1007/BF02249265

[8] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *Information Theory, IEEE Transactions on*, vol. 27, no. 1, pp. 49–60, jan 1981.

[9] M. H. M. Costa and C. Nair, "Gaussian z-interference channel: around the corner," *Information Theory and Applications Workshop (ITA)*, 2016.

[10] S. Beigi, S. Liu, C. Nair, and M. Yazdanpanah, "Some results on the scalar Gaussian interference channel," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2199–2203.

[11] M. Costa, C. Nair, D. Ng, and Y. N. Wang, "On the structure of certain non-convex functionals and the gaussian z-interference channel," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 1522–1527.

[12] A. Gohari and C. Nair, "New outer bounds for the two-receiver broadcast channel," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 1492–1497.

[13] A. Gohari and C. Nair, "Outer bounds for multiuser settings: The auxiliary receiver approach," *submitted to the IEEE Transactions on Information Theory*, 2020.

[14] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2087–2104, April 2014.

[15] T. Cover and J. Thomas, *Elements of Information Theory.* Wiley Interscience, 1991.

A Proof of Proposition 4

The proof follows from putting the following calculations together, with $(\star)$ denoting application of chain rule and $(\sharp)$ denoting application of Csiszár's sum identity.

(i) We have

$$\begin{aligned} I(\mathbf{X}_2^n;\mathbf{T}_2^n|\hat{V}) &\overset{(\star)}{=} \sum_{i=1}^n I(\mathbf{X}_2^n;\mathbf{T}_{2i}|\mathbf{T}_2^{i-1},\hat{V}) \\ &\overset{(a)}{=} \sum_{i=1}^n I(\mathbf{X}_{2i};\mathbf{T}_{2i}|\mathbf{T}_2^{i-1},\hat{V}) \\ &= \sum_{i=1}^n I(\mathbf{X}_{2i};\mathbf{T}_{2i}|V_i) \end{aligned}$$

where (a) holds since

$$\mathbf{X}_2^{n\setminus i} \to (\mathbf{X}_{2i},\mathbf{T}_2^{i-1},\hat{V}) \to \mathbf{T}_{2i}$$

forms a Markov chain.

(ii) We have

$$\begin{aligned} &h(\mathbf{Y}_2^n|\hat{W}) - h(\mathbf{Y}_1^n|\hat{W}) \\ &\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{2i}|\mathbf{Y}_2^{i-1},\hat{W}) - h(\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n,\hat{W})\Big] \\ &\overset{(\sharp)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{2i}|\mathbf{Y}_2^{i-1},\hat{W}) - h(\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n,\hat{W}) + I(\mathbf{Y}_2^{i-1};\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n,\hat{W}) - I(\mathbf{Y}_{1(i+1)}^n;\mathbf{Y}_{2i}|\mathbf{Y}_2^{i-1},\hat{W})\Big] \\ &\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{2i}|\mathbf{Y}_{1(i+1)}^n,\mathbf{Y}_2^{i-1},\hat{W}) - h(\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n,\mathbf{Y}_2^{i-1},\hat{W})\Big] \\ &= \sum_{i=1}^n \Big[h(\mathbf{Y}_{2i}|W_i) - h(\mathbf{Y}_{1i}|W_i)\Big]. \end{aligned}$$

(iii) We have

$$
\begin{aligned}
& h(\mathbf{Y}_1^n|\mathbf{X}_2^n,\hat{W}) - h(\mathbf{Y}_2^n|\mathbf{X}_2^n,\hat{W}) \\
&\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{X}_2^n,\mathbf{Y}_{1(i+1)}^n,\hat{W}) - h(\mathbf{Y}_{2i}|\mathbf{X}_2^n,\mathbf{Y}_2^{i-1},\hat{W})\Big] \\
&\overset{(\sharp)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{X}_2^n,\mathbf{Y}_{1(i+1)}^n,\hat{W}) - h(\mathbf{Y}_{2i}|\mathbf{X}_2^n,\mathbf{Y}_2^{i-1},\hat{W}) \\
&\qquad + I(\mathbf{Y}_{1(i+1)}^n;\mathbf{Y}_{2i}|\mathbf{X}_2^n,\mathbf{Y}_2^{i-1},\hat{W}) - I(\mathbf{Y}_2^{i-1};\mathbf{Y}_{1i}|\mathbf{X}_2^n,\mathbf{Y}_{1(i+1)}^n,\hat{W})\Big] \\
&\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{X}_2^n,\mathbf{Y}_{1(i+1)}^n,\mathbf{Y}_2^{i-1},\hat{W}) - h(\mathbf{Y}_{2i}|\mathbf{X}_2^n,\mathbf{Y}_{1(i+1)}^n,\mathbf{Y}_2^{i-1},\hat{W})\Big] \\
&= \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{X}_2^n,W_i) - h(\mathbf{Y}_{2i}|\mathbf{X}_2^n,W_i)\Big] \\
&\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{X}_{2i},W_i) - h(\mathbf{Y}_{2i}|\mathbf{X}_{2i},W_i) + I(\mathbf{Y}_{2i};\mathbf{X}_2^{n\backslash i}|\mathbf{X}_{2i},W_i) - I(\mathbf{Y}_{1i};\mathbf{X}_2^{n\backslash i}|\mathbf{X}_{2i},W_i)\Big] \\
&\overset{(a)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{X}_{2i},W_i) - h(\mathbf{Y}_{2i}|\mathbf{X}_{2i},W_i) + I(\mathbf{Y}_{2i};\mathbf{X}_2^{n\backslash i}|\mathbf{X}_{2i},W_i) - I(\mathbf{Y}_{1i},\mathbf{Y}_{2i};\mathbf{X}_2^{n\backslash i}|\mathbf{X}_{2i},W_i)\Big] \\
&\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{X}_{2i},W_i) - h(\mathbf{Y}_{2i}|\mathbf{X}_{2i},W_i)\Big] - \sum_{i=1}^n I(\mathbf{Y}_{1i};\mathbf{X}_2^{n\backslash i}|\mathbf{X}_{2i},\mathbf{Y}_{2i},W_i)
\end{aligned}
$$

where (a) holds since

$$\mathbf{Y}_{2i} \to (\mathbf{Y}_{1i},\mathbf{X}_{2i},W_i) \to \mathbf{X}_2^{n\backslash i}$$

forms a Markov chain.

(iv) Denoting $\mathbf{S}_{2i} := \mathbf{T}_{2i} + \mathbf{G}_i$ we have

$$
\begin{aligned}
I(\mathbf{X}_2^n;\mathbf{T}_2^n+\mathbf{G}^n|\hat{V}) &= I(\mathbf{X}_2^n;\mathbf{S}_2^n|\hat{V}) \\
&\overset{(\star)}{=} \sum_{i=1}^n I(\mathbf{X}_2^n;\mathbf{S}_{2i}|\mathbf{S}_2^{i-1},\hat{V}) \\
&\overset{(a)}{=} \sum_{i=1}^n I(\mathbf{X}_{2i};\mathbf{S}_{2i}|\mathbf{S}_2^{i-1},\hat{V}) \\
&\overset{(\star)}{=} \sum_{i=1}^n \Big[I(\mathbf{X}_{2i},\mathbf{S}_2^{i-1};\mathbf{S}_{2i}|\hat{V}) - I(\mathbf{S}_2^{i-1};\mathbf{S}_{2i}|\hat{V})\Big] \\
&\overset{(b)}{=} \sum_{i=1}^n \Big[I(\mathbf{X}_{2i},\mathbf{T}_2^{i-1};\mathbf{S}_{2i}|\hat{V}) - I(\mathbf{S}_2^{i-1};\mathbf{S}_{2i}|\hat{V})\Big] \\
&\overset{(\star)}{=} \sum_{i=1}^n \Big[I(\mathbf{X}_{2i};\mathbf{S}_{2i}|\mathbf{T}_2^{i-1},\hat{V}) + I(\mathbf{T}_2^{i-1};\mathbf{S}_{2i}|\hat{V}) - I(\mathbf{S}_2^{i-1};\mathbf{S}_{2i}|\hat{V})\Big] \\
&\overset{(c)}{=} \sum_{i=1}^n \Big[I(\mathbf{X}_{2i};\mathbf{S}_{2i}|\mathbf{T}_2^{i-1},\hat{V}) + I(\mathbf{T}_2^{i-1},\mathbf{S}_2^{i-1};\mathbf{S}_{2i}|\hat{V}) - I(\mathbf{S}_2^{i-1};\mathbf{S}_{2i}|\hat{V})\Big] \\
&\overset{(\star)}{=} \sum_{i=1}^n \Big[I(\mathbf{X}_{2i};\mathbf{S}_{2i}|\mathbf{T}_2^{i-1},\hat{V}) + I(\mathbf{T}_2^{i-1};\mathbf{S}_{2i}|\mathbf{S}_2^{i-1},\hat{V})\Big] \\
&= \sum_{i=1}^n I(\mathbf{X}_{2i};\mathbf{T}_{2i}+\mathbf{G}_i|V_i) + \sum_{i=1}^n I(\mathbf{T}_2^{i-1};\mathbf{T}_{2i}+\mathbf{G}_i|\mathbf{T}_2^{i-1}+\mathbf{G}^{i-1},\hat{V})
\end{aligned}
$$

where (a) holds since

$$\mathbf{X}_2^{n\backslash i} \to (\mathbf{X}_{2i},\mathbf{S}_2^{i-1},\hat{V}) \to \mathbf{S}_{2i}$$

forms a Markov chain, (b) holds since

$$(\mathbf{T}_2^{i-1},\mathbf{S}_2^{i-1}) \to (\mathbf{X}_{2i},\hat{V}) \to \mathbf{S}_{2i}$$

forms a Markov chain, (c) holds since

$$\mathbf{S}_2^{i-1} \to (\mathbf{T}_2^{i-1}, \hat{V}) \to \mathbf{S}_{2i}$$

forms a Markov chain.

(v) We have

$$\begin{aligned} h(\mathbf{Y}_1^n|\hat{W}) &\overset{(\star)}{=} \sum_{i=1}^n h(\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{W}) \\ &\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \mathbf{Y}_2^{i-1}, \hat{W}) + I(\mathbf{Y}_2^{i-1}; \mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{W}) \Big] \\ &= \sum_{i=1}^n h(\mathbf{Y}_{1i}|W_i) + \sum_{i=1}^n I(\mathbf{Y}_2^{i-1}; \mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{W}). \end{aligned}$$

(vi) We have

$$\begin{aligned} h(\mathbf{Y}_1^n|\hat{V}, \hat{W}) &\overset{(\star)}{=} \sum_{i=1}^n h(\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{V}, \hat{W}) \\ &\overset{(\star)}{=} \sum_{i=1}^n \Big[h(\mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \mathbf{Y}_2^{i-1}, \mathbf{T}_2^{i-1}, \hat{V}, \hat{W}) + I(\mathbf{T}_2^{i-1}, \mathbf{Y}_2^{i-1}; \mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{V}, \hat{W}) \Big] \\ &= \sum_{i=1}^n h(\mathbf{Y}_{1i}|V_i, W_i) + \sum_{i=1}^n I(\mathbf{T}_2^{i-1}, \mathbf{Y}_2^{i-1}; \mathbf{Y}_{1i}|\mathbf{Y}_{1(i+1)}^n, \hat{V}, \hat{W}). \end{aligned}$$

B Proof of Lemma 1

(i) The Markov chain

$$(\mathbf{X}_1)_1^n \to (\hat{V}, \hat{W}) \to (\mathbf{X}_2)_1^n$$

implies

$$(\mathbf{X}_{1i}, (\mathbf{Y}_1)_{i+1}^n, (\mathbf{X}_1)_1^{i-1}) \to (\hat{V}, \hat{W}) \to (\mathbf{X}_{2i}, (\tilde{\mathbf{T}}_2^{(\delta)})_1^{i-1})$$

which in turn implies

$$\mathbf{X}_{1i} \to (\hat{V}, \hat{W}, (\mathbf{Y}_1)_{i+1}^n, (\mathbf{X}_1)_1^{i-1}, (\tilde{\mathbf{T}}_2^{(\delta)})_1^{i-1}) \to \mathbf{X}_{2i}$$

which, since $\tilde{\mathbf{Y}}_{2i}^{(\delta)} = \tilde{\mathbf{T}}_{2i}^{(\delta)} + \begin{pmatrix} \mathbf{X}_{1i} \\ 0 \end{pmatrix}$, is the same as

$$\mathbf{X}_{1i} \to (\hat{V}, \hat{W}, (\mathbf{Y}_1)_{i+1}^n, (\tilde{\mathbf{Y}}_2^{(\delta)})_1^{i-1}, (\tilde{\mathbf{T}}_2^{(\delta)})_1^{i-1}) \to \mathbf{X}_{2i}$$

or equivalently

$$\mathbf{X}_{1i} \to (\tilde{V}_i^{(\delta)}, \tilde{W}_i^{(\delta)}) \to \mathbf{X}_{2i}.$$

Moreover, from both of the Markov chains in the assumption we have a Markov chain

$$(\hat{W}, (\mathbf{X}_1)_1^n) \to \hat{V} \to (\mathbf{X}_2)_1^n$$

which implies

$$(\hat{W}, (\mathbf{Y}_1)_{i+1}^n, (\mathbf{X}_1)_1^{i-1}) \to \hat{V} \to (\mathbf{X}_{2i}, (\tilde{\mathbf{T}}_2^{(\delta)})_1^{i-1})$$

which in turn implies

$$(\hat{W}, (\mathbf{Y}_1)_{i+1}^n, (\mathbf{X}_1)_1^{i-1}) \to (\hat{V}, (\tilde{\mathbf{T}}_2^{(\delta)})_1^{i-1}) \to \mathbf{X}_{2i}$$

which, since $\tilde{\mathbf{Y}}_{2i}^{(\delta)} = \tilde{\mathbf{T}}_{2i}^{(\delta)} + \begin{pmatrix} \mathbf{X}_{1i} \\ 0 \end{pmatrix}$, is the same as

$$(\hat{W}, (\mathbf{Y}_1)_{i+1}^n, (\tilde{\mathbf{Y}}_2^{(\delta)})_1^{i-1}) \to (\hat{V}, (\tilde{\mathbf{T}}_2^{(\delta)})_1^{i-1}) \to \mathbf{X}_{2i}$$

or equivalently

$$\tilde{W}_i^{(\delta)} \to \tilde{V}_i^{(\delta)} \to \mathbf{X}_{2i}.$$

(ii) This is obvious from definition.

C Proof of Proposition 5

We will need three lemmas for proving (ii) of the Proposition.

Lemma 3 (Lemma 1 of [11]). *Let Q be a random variable and let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be real-vector-valued random variables such that for any q the conditional distribution $p(\mathbf{x}, \mathbf{y}, \mathbf{z}|q)$ has everywhere non-zero density. Suppose*

$$\mathbf{X} \to (\mathbf{Y}, Q) \to \mathbf{Z} \quad \textit{and} \quad \mathbf{Y} \to (\mathbf{X}, Q) \to \mathbf{Z}$$

form Markov chains. Then

$$(\mathbf{X}, \mathbf{Y}) \to Q \to \mathbf{Z}$$

forms a Markov chain.

Lemma 4 (Lemma 2 of [11]). *Let $\mathbf{X}_1, \mathbf{X}_2$ be real-vector-valued random variables and $\mathbf{Z}_1, \mathbf{Z}_2$ be Gaussian random variables such that $(\mathbf{X}_1, \mathbf{X}_2)$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent. Then $(\mathbf{X}_1 + \mathbf{Z}_1) \perp (\mathbf{X}_2 + \mathbf{Z}_2)$ implies $\mathbf{X}_1 \perp \mathbf{X}_2$.*

Lemma 5 (Lemma 3 of [11]). *Let $\mathbf{X}_1, \mathbf{X}_2$ be real-vector-valued random variables such that $\mathbf{X}_1 \perp \mathbf{X}_2$ and $(\mathbf{X}_1 + \mathbf{X}_2) \perp (\mathbf{X}_1 - \mathbf{X}_2)$. Then $\mathbf{X}_1, \mathbf{X}_2$ are Gaussians having the same covariance matrix.*

The following lemma illustrates the use of rotation argument and sub-additivity to establish mid-point concavity of $v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}$.

Lemma 6. *Let $v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}$ be as defined in Proposition 5. Assume $\epsilon \geq 0$ and $\delta \in \mathbb{R}$. Then for any $0 \leq t \leq 1$,*

$$v^{(1)}_{(\mathcal{K}_{11},\mathcal{K}_{21})} + v^{(1)}_{(\mathcal{K}_{12},\mathcal{K}_{22})} \leq v^{(1)}_{(\mathcal{K}^t_{11},\mathcal{K}^t_{21})} + v^{(1)}_{(\mathcal{K}^t_{12},\mathcal{K}^t_{22})}$$

where

$$\mathcal{K}^t_{11} := t\mathcal{K}_{11} + (1-t)\mathcal{K}_{12}, \quad \mathcal{K}^t_{21} := t\mathcal{K}_{21} + (1-t)\mathcal{K}_{22},$$
$$\mathcal{K}^t_{12} := (1-t)\mathcal{K}_{11} + t\mathcal{K}_{12}, \quad \mathcal{K}^t_{22} := (1-t)\mathcal{K}_{21} + t\mathcal{K}_{22}.$$

Proof. Suppose $(\mathbf{X}^*_{1i}, \mathbf{X}^*_{2i}, V^*_i, W^*_i)$ are random variables satisfying the constraints of $v^{(1)}_{(\mathcal{K}_{1i},\mathcal{K}_{2i})}$, and are independent among $i = 1, 2$. Let

$$(\hat{V}, \hat{W}) := ((V^*_1, V^*_2), (W^*_1, W^*_2))$$

and

$$\begin{pmatrix} \mathbf{X}_{11} \\ \mathbf{X}_{12} \end{pmatrix} := \begin{pmatrix} \sqrt{t}I & \sqrt{1-t}I \\ -\sqrt{1-t}I & \sqrt{t}I \end{pmatrix} \begin{pmatrix} \mathbf{X}^*_{11} \\ \mathbf{X}^*_{12} \end{pmatrix}, \qquad \begin{pmatrix} \mathbf{X}_{21} \\ \mathbf{X}_{22} \end{pmatrix} := \begin{pmatrix} \sqrt{t}I & \sqrt{1-t}I \\ -\sqrt{1-t}I & \sqrt{t}I \end{pmatrix} \begin{pmatrix} \mathbf{X}^*_{21} \\ \mathbf{X}^*_{22} \end{pmatrix}.$$

It is immediate that the distribution of $((\mathbf{X}_{11}, \mathbf{X}_{12}), (\mathbf{X}_{21}, \mathbf{X}_{22}), \hat{V}, \hat{W})$ is in $\mathcal{M}$, and hence by Lemma 1 the distribution of $(\mathbf{X}_{1i}, \mathbf{X}_{2i}, \tilde{V}^{(\delta)}_i, \tilde{W}^{(\delta)}_i)$ is in $\mathcal{M}$ and so satisfies the constraints of $v^{(1)}_{(\mathcal{K}^t_{1i},\mathcal{K}^t_{2i})}$. Now,

$$\begin{aligned}
&\theta^{(\epsilon,\delta)}_{\mu,\alpha}(\mathbf{X}^*_{11}, \mathbf{X}^*_{21}|V^*_1, W^*_1) + \theta^{(\epsilon,\delta)}_{\mu,\alpha}(\mathbf{X}^*_{12}, \mathbf{X}^*_{22}|V^*_2, W^*_2) \\
&\overset{(a)}{=} \theta^{(\epsilon,\delta)}_{\mu,\alpha}((\mathbf{X}^*_{11}, \mathbf{X}^*_{12}), (\mathbf{X}^*_{21}, \mathbf{X}^*_{22})|\hat{V}, \hat{W}) \\
&\overset{(b)}{=} \theta^{(\epsilon,\delta)}_{\mu,\alpha}((\mathbf{X}_{11}, \mathbf{X}_{12}), (\mathbf{X}_{21}, \mathbf{X}_{22})|\hat{V}, \hat{W}) \\
&\overset{(c)}{=} \underbrace{\theta^{(\epsilon,\delta)}_{\mu,\alpha}(\mathbf{X}_{11}, \mathbf{X}_{21}|\tilde{V}^{(\delta)}_1, \tilde{W}^{(\delta)}_1)}_{\leq v^{(1)}_{(\mathcal{K}^t_{11},\mathcal{K}^t_{21})}} + \underbrace{\theta^{(\epsilon,\delta)}_{\mu,\alpha}(\mathbf{X}_{12}, \mathbf{X}_{22}|\tilde{V}^{(\delta)}_2, \tilde{W}^{(\delta)}_2)}_{\leq v^{(1)}_{(\mathcal{K}^t_{12},\mathcal{K}^t_{22})}} \\
&\quad - \Big[(1+\mu(1-\alpha)) \cdot I(\mathbf{Y}_{11}; \mathbf{X}_{22}|\mathbf{X}_{21}, \tilde{\mathbf{Y}}^{(\delta)}_{21}, \tilde{W}^{(\delta)}_1) + (1+\mu(1-\alpha)) \cdot I(\mathbf{Y}_{12}; \mathbf{X}_{21}|\mathbf{X}_{22}, \tilde{\mathbf{Y}}^{(\delta)}_{22}, \tilde{W}^{(\delta)}_2) \\
&\qquad + \epsilon \cdot I(\tilde{\mathbf{T}}^{(\delta)}_{21}; \tilde{\mathbf{T}}^{(\delta)}_{22} + \mathbf{G}_2|\tilde{\mathbf{T}}^{(\delta)}_{21} + \mathbf{G}_1, \hat{V}) + \epsilon \cdot I(\tilde{\mathbf{Y}}^{(\delta)}_{21}; \mathbf{Y}_{12}|\hat{W}) + \epsilon \cdot I(\tilde{\mathbf{T}}^{(\delta)}_{21}, \tilde{\mathbf{Y}}^{(\delta)}_{21}; \mathbf{Y}_{12}|\hat{V}, \hat{W})\Big] \\
&\leq v^{(1)}_{(\mathcal{K}^t_{11},\mathcal{K}^t_{21})} + v^{(1)}_{(\mathcal{K}^t_{12},\mathcal{K}^t_{22})}
\end{aligned}$$

where (a) follows from the additivity of entropy for independent random variables, (b) follows from rotational invariance of entropy, and (c) is established by Proposition 4. □

Proof of Proposition 5 (i). We shall first show

$$\sum_{i=1}^{n} v^{(1)}_{(\mathcal{K}_{1i},\mathcal{K}_{2i})} \leq v^{(n)}_{(\mathcal{K}_{11},\mathcal{K}_{21}),\ldots,(\mathcal{K}_{1n},\mathcal{K}_{2n})}.$$

Suppose $(\mathbf{X}_{1i}^*, \mathbf{X}_{2i}^*, V_i^*, W_i^*)$ are random variables satisfying the constraints of $v^{(1)}_{(\mathcal{K}_{1i},\mathcal{K}_{2i})}$, and are independent among $i = 1, \cdots, n$. Then the random variables defined by

$$(\mathbf{X}_1^n, \mathbf{X}_2^n, \hat{V}, \hat{W}) := ((\mathbf{X}_1^*)_1^n, (\mathbf{X}_2^*)_1^n, (V^*)_1^n, (W^*)_1^n)$$

satisfy the constraints of $v^{(n)}_{(\mathcal{K}_{11},\mathcal{K}_{21}),\ldots,(\mathcal{K}_{1n},\mathcal{K}_{2n})}$, as well as

$$\sum_{i=1}^n \theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_{1i}^*, \mathbf{X}_{2i}^* | V_i^*, W_i^*) \overset{\text{(a)}}{=} \theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n | \hat{V}, \hat{W}) \le v^{(n)}_{(\mathcal{K}_{11},\mathcal{K}_{21}),\ldots,(\mathcal{K}_{1n},\mathcal{K}_{2n})}$$

where (a) follows from the additivity of entropy for independent random variables.

Now we show

$$\sum_{i=1}^n v^{(1)}_{(\mathcal{K}_{1i},\mathcal{K}_{2i})} \ge v^{(n)}_{(\mathcal{K}_{11},\mathcal{K}_{21}),\ldots,(\mathcal{K}_{1n},\mathcal{K}_{2n})}.$$

Suppose $(\mathbf{X}_1^n, \mathbf{X}_2^n, \hat{V}, \hat{W})$ are random variables satisfying the constraints of $v^{(n)}_{(\mathcal{K}_{11},\mathcal{K}_{21}),\ldots,(\mathcal{K}_{1n},\mathcal{K}_{2n})}$. Then

$$\theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n | \hat{V}, \hat{W}) \overset{\text{(a)}}{\le} \sum_{i=1}^n \theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_{1i}, \mathbf{X}_{2i} | \tilde{V}_i^{(\delta)}, \tilde{W}_i^{(\delta)}) \overset{\text{(b)}}{\le} \sum_{i=1}^n v^{(1)}_{(\mathcal{K}_{1i},\mathcal{K}_{2i})}$$

where (a) follows from Proposition 4, and (b) holds since $(\mathbf{X}_{1i}, \mathbf{X}_{2i}, \tilde{V}_i^{(\delta)}, \tilde{W}_i^{(\delta)})$ satisfies the constraints of $v^{(1)}_{(\mathcal{K}_{1i},\mathcal{K}_{2i})}$, as a result of Lemma 1.

□

Proof of Proposition 5 (ii). The existence of maximizer can be justified by Prokhorov theorem through techniques in Appendix II of [14].

Now consider $\epsilon > 0$ and $\delta \neq 0$. Take $(\mathbf{X}_{11}^*, \mathbf{X}_{21}^*, V_1^*, W_1^*)$ and $(\mathbf{X}_{12}^*, \mathbf{X}_{22}^*, V_2^*, W_2^*)$ to be two independent copies of a maximizer $p^*(\mathbf{x}_1, \mathbf{x}_2, v, w)$ for $v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}$. With $t = \frac{1}{2}$ define $((\mathbf{X}_{11}, \mathbf{X}_{12}), (\mathbf{X}_{21}, \mathbf{X}_{22}), \hat{V}, \hat{W})$ as in the proof of Lemma 6. Note that the distribution of $(\mathbf{X}_{1i}, \mathbf{X}_{2i}, \tilde{V}_i^{(\delta)}, \tilde{W}_i^{(\delta)})$ are candidate maximizers for $v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}$. Following the steps of Lemma 6 we have

$$\begin{aligned}
&2v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)} = \theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_{11}^*, \mathbf{X}_{21}^* | V_1^*, W_1^*) + \theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_{12}^*, \mathbf{X}_{22}^* | V_2^*, W_2^*) \\
&\overset{\text{(c)}}{=} \underbrace{\theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_{11}, \mathbf{X}_{21} | \tilde{V}_1^{(\delta)}, \tilde{W}_1^{(\delta)})}_{\le v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}} + \underbrace{\theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_{12}, \mathbf{X}_{22} | \tilde{V}_2^{(\delta)}, \tilde{W}_2^{(\delta)})}_{\le v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}} \\
&\quad - \Big[(1+\mu(1-\alpha)) \cdot I(\mathbf{Y}_{11}; \mathbf{X}_{22} | \mathbf{X}_{21}, \tilde{\mathbf{Y}}_{21}^{(\delta)}, \tilde{W}_1^{(\delta)}) + (1+\mu(1-\alpha)) \cdot I(\mathbf{Y}_{12}; \mathbf{X}_{21} | \mathbf{X}_{22}, \tilde{\mathbf{Y}}_{22}^{(\delta)}, \tilde{W}_2^{(\delta)}) \\
&\qquad + \epsilon \cdot I(\tilde{\mathbf{T}}_{21}^{(\delta)}; \tilde{\mathbf{T}}_{22}^{(\delta)} + \mathbf{G}_2 | \tilde{\mathbf{T}}_{21}^{(\delta)} + \mathbf{G}_1, \hat{V}) + \epsilon \cdot I(\tilde{\mathbf{Y}}_{21}^{(\delta)}; \mathbf{Y}_{12} | \hat{W}) + \epsilon \cdot I(\tilde{\mathbf{T}}_{21}^{(\delta)}, \tilde{\mathbf{Y}}_{21}^{(\delta)}; \mathbf{Y}_{12} | \hat{V}, \hat{W})\Big] \\
&\le 2v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}.
\end{aligned}$$

Non-negativity of mutual information forces the Markov chains:

$$\tilde{\mathbf{T}}_{21}^{(\delta)} \to (\tilde{\mathbf{T}}_{21}^{(\delta)} + \mathbf{G}_1, \hat{V}) \to \tilde{\mathbf{T}}_{22}^{(\delta)} + \mathbf{G}_2, \tag{a}$$

$$\tilde{\mathbf{Y}}_{21}^{(\delta)} \to \hat{W} \to \mathbf{Y}_{12}, \tag{b}$$

$$(\tilde{\mathbf{T}}_{21}^{(\delta)}, \tilde{\mathbf{Y}}_{21}^{(\delta)}) \to (\hat{V}, \hat{W}) \to \mathbf{Y}_{12} \tag{c}$$

where (a) together with the Markov chain

$$\tilde{\mathbf{T}}_{21}^{(\delta)} + \mathbf{G}_1 \to (\tilde{\mathbf{T}}_{21}^{(\delta)}, \hat{V}) \to \tilde{\mathbf{T}}_{22}^{(\delta)} + \mathbf{G}_2$$

implies by double Markovity (Lemma 3) that

$$(\tilde{\mathbf{T}}_{21}^{(\delta)}, \tilde{\mathbf{T}}_{21}^{(\delta)} + \mathbf{G}_1) \to \hat{V} \to \tilde{\mathbf{T}}_{22}^{(\delta)} + \mathbf{G}_2 \tag{a'}$$

forms a Markov chain. Applying Lemma 4 to (a'), (b), (c) respectively gives the Markov chains (using the fact that $\delta \neq 0$)

$$\mathbf{X}_{21} \to \hat{V} \to \mathbf{X}_{22}, \qquad \mathbf{X}_{11} \to \hat{W} \to \mathbf{X}_{12}, \qquad \mathbf{X}_{11} \to (\hat{V}, \hat{W}) \to \mathbf{X}_{12}$$

which, by Lemma 5, imply that for any $v_1^*, v_2^*, w_1^*, w_2^*$ each of the pairs of distributions

$$(\mathbf{X}_{21}^*|_{V_1^*=v_1^*}, \mathbf{X}_{22}^*|_{V_2^*=v_2^*}), \qquad (\mathbf{X}_{11}^*|_{W_1^*=w_1^*}, \mathbf{X}_{12}^*|_{W_2^*=w_2^*}), \qquad (\mathbf{X}_{11}^*|_{V_1^*=v_1^*, W_1^*=w_1^*}, \mathbf{X}_{12}^*|_{V_2^*=v_2^*, W_2^*=w_2^*})$$

consists of Gaussians with the same covariance matrix. Since $v_1^*, v_2^*, w_1^*, w_2^*$ are arbitrary, we can conclude that $p^*(\mathbf{x}_1|w)$, $p^*(\mathbf{x}_1|v,w)$ and $p^*(\mathbf{x}_2|v)$ are Gaussians with covariance matrices independent of choice of v and w.

Finally we show the case $\epsilon = \delta = 0$. Let $\tilde{v}^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}$ be defined in the same way as $v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}$ but with the additional constraint that $p(\mathbf{x}_1^n|\hat{w})$, $p(\mathbf{x}_1^n|\hat{v},\hat{w})$ and $p(\mathbf{x}_2^n|\hat{v})$ are Gaussians with covariance matrices independent of choice of $\hat{v}$ and $\hat{w}$. Then we have $\tilde{v}^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)} = v^{(1)}_{(\mathcal{K}_1,\mathcal{K}_2)}$ for $\epsilon > 0$ and $\delta \neq 0$. One can take the limit $(\epsilon,\delta) \to (0,0)$ and apply Lemma 2 to get the result.

□

D Proof of Lemma 2

The proof applies two lemmas. Lemma 7 is a statement of asymptotic bound on entropy by power. Lemma 8 says one can interchange limit and maximization under uniform convergence.

Lemma 7. *Let $\mathbf{Z}$ be a Gaussian random variable in $\mathbb{R}^d$ ($d \geq 1$) with an invertible covariance matrix. Then there exists $c \geq 0$ depending only on the covariance matrix of $\mathbf{Z}$ such that*

$$0 \leq h(\mathbf{X}+\mathbf{Z}|U) - h(\mathbf{Z}) \leq c \cdot \mathbb{E}[\|\mathbf{X}\|^2]$$

for any random variables $(\mathbf{X}, U) \perp \mathbf{Z}$ where $\mathbf{X}$ is in $\mathbb{R}^d$.

Proof. We have

$$\begin{aligned} h(\mathbf{X}+\mathbf{Z}|U) - h(\mathbf{Z}) &= h(\mathbf{X}+\mathbf{Z}|U) - h(\mathbf{X}+\mathbf{Z}|\mathbf{X},U) \\ &= I(\mathbf{X};\mathbf{X}+\mathbf{Z}|U) \\ &\geq 0. \end{aligned}$$

On the other hand, with

$$K := \operatorname{Cov}(\mathbf{Z})^{-1/2}\operatorname{Cov}(\mathbf{X})\operatorname{Cov}(\mathbf{Z})^{-1/2}$$

we have

$$\begin{aligned} h(\mathbf{X}+\mathbf{Z}|U) - h(\mathbf{Z}) &\leq h(\mathbf{X}+\mathbf{Z}) - h(\mathbf{Z}) \\ &\overset{\text{(a)}}{\leq} \frac{1}{2}\log\left|2\pi e(\operatorname{Cov}(\mathbf{X})+\operatorname{Cov}(\mathbf{Z}))\right| - \frac{1}{2}\log\left|2\pi e\operatorname{Cov}(\mathbf{Z})\right| \\ &= \frac{1}{2}\log|K+I| \\ &= \frac{1}{2}\sum_{i=1}^{d}\log\left[1+\lambda_i(K)\right] \\ &\overset{\text{(b)}}{\leq} \frac{d}{2}\log\left[1+\sum_{i=1}^{d}\frac{1}{d}\lambda_i(K)\right] \\ &= \frac{d}{2}\log\left[1+\frac{1}{d}\operatorname{tr}(K)\right] \\ &\overset{\text{(c)}}{\leq} \frac{1}{2}\operatorname{tr}(K) \\ &= \frac{1}{2}\operatorname{tr}(\operatorname{Cov}(\mathbf{X})\operatorname{Cov}(\mathbf{Z})^{-1}) \\ &\overset{\text{(d)}}{\leq} \frac{\operatorname{tr}(\operatorname{Cov}(\mathbf{X}))}{2\lambda_{\min}(\operatorname{Cov}(\mathbf{Z}))} \\ &\leq \frac{\mathbb{E}[\|\mathbf{X}\|^2]}{2\lambda_{\min}(\operatorname{Cov}(\mathbf{Z}))} \end{aligned}$$

where (a) holds since Gaussian maximizes entropy, (b) follows from Jensen's inequality, (c) holds since $\log(1+x) \leq x$ for $x \geq 0$, (d) follows from von Neumann's trace inequality, and $\lambda_i(\cdot)$ (respectively $\lambda_{\min}(\cdot)$) denotes the i-th largest (respectively smallest) eigenvalue functional.

□

Lemma 8. *Let f and f_n ($n \in \mathbb{N}$) be real-valued functions defined on the same set. Suppose*

$$\lim_{n\to\infty}\sup_x |f_n(x) - f(x)| = 0.$$

Then

$$\lim_{n\to\infty} \sup_x f_n(x) = \sup_x f(x).$$

Proof. Let $\epsilon_n := \sup_x |f_n(x) - f(x)|$. Note that ϵ_n is bounded for sufficiently large n since ϵ_n converges. We have

$$f(x) - \epsilon_n \le f_n(x) \le f(x) + \epsilon_n$$

for any x and sufficiently large n. Taking supremum over x gives

$$\sup_x f(x) - \epsilon_n \le \sup_x f_n(x) \le \sup_x f(x) + \epsilon_n$$

and then the result follows by squeezing. □

Proof of Lemma 2. Recall that by definition we have

$$(\tilde{\mathbf{T}}_2^{(\delta)})_1^n = (\mathbf{T}_2^n, \delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n), \qquad (\tilde{\mathbf{Y}}_2^{(\delta)})_1^n = (\mathbf{Y}_2^n, \delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n).$$

Plugging into the definition of $\theta_{\mu,\alpha}^{(\epsilon,\delta)}$, with $\mathbf{G}_i, \hat{\mathbf{G}}_i$'s being independent Gaussian random variables with identity covariance matrix, we have that for any $p(\mathbf{x}_1^n, \mathbf{x}_2^n, \hat{v}, \hat{w})$,

$$\begin{aligned}
&\left|\theta_{\mu,\alpha}^{(\epsilon,\delta)}(\mathbf{X}_1^n, \mathbf{X}_2^n|\hat{V},\hat{W}) - \theta_{\mu,\alpha}^{(0,0)}(\mathbf{X}_1^n, \mathbf{X}_2^n|\hat{V},\hat{W})\right| \\
&\le \mu\alpha \cdot \left|I(\mathbf{X}_2^n; \mathbf{T}_2^n, \delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\hat{V}) - I(\mathbf{X}_2^n; \mathbf{T}_2^n, \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\hat{V})\right| \\
&\quad + \mu(1-\alpha) \cdot \left|h(\mathbf{Y}_2^n, \delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\hat{W}) - h(\mathbf{Y}_2^n, \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\hat{W})\right| \\
&\quad + (1+\mu(1-\alpha)) \cdot \left|h(\mathbf{Y}_2^n, \delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\mathbf{X}_2^n, \hat{W}) - h(\mathbf{Y}_2^n, \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\mathbf{X}_2^n, \hat{W})\right| \\
&\quad + |\epsilon| \cdot \left|I(\mathbf{X}_2^n; \mathbf{T}_2^n + \mathbf{G}^n, \delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n + \hat{\mathbf{G}}^n|\hat{V})\right| + |\epsilon| \cdot \left|h(\mathbf{X}_1^n + \mathbf{Z}_1^n|\hat{W})\right| + |\epsilon| \cdot \left|h(\mathbf{X}_1^n + \mathbf{Z}_1^n|\hat{V},\hat{W})\right| \\
&\overset{\text{(a)}}{=} \mu\alpha \cdot \left|\left[h(\delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\mathbf{T}_2^n, \hat{V}) - h(\hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n)\right] - \left[h(\delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\mathbf{X}_2^n, \mathbf{T}_2^n, \hat{V}) - h(\hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n)\right]\right| \\
&\quad + \mu(1-\alpha) \cdot \left|h(\delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\mathbf{Y}_2^n, \hat{W}) - h(\hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n)\right| \\
&\quad + (1+\mu(1-\alpha)) \cdot \left|h(\delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n|\mathbf{Y}_2^n, \mathbf{X}_2^n, \hat{W}) - h(\hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n)\right| \\
&\quad + |\epsilon| \cdot \Big|\left[h(\delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n + \hat{\mathbf{G}}^n|\mathbf{T}_2^n + \mathbf{G}^n, \hat{V}) - h(\hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n + \hat{\mathbf{G}}^n)\right] \\
&\qquad - \left[h(\delta\mathbf{X}_2^n + \hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n + \hat{\mathbf{G}}^n|\mathbf{T}_2^n + \mathbf{G}^n, \mathbf{X}_2^n, \hat{V}) - h(\hat{\mathbf{Z}}_1^n + \hat{\mathbf{Z}}_2^n + \hat{\mathbf{G}}^n)\right] \\
&\qquad + \left[h(\mathbf{T}_2^n + \mathbf{G}^n|\hat{V}) - h(\mathbf{Z}_1^n + \mathbf{Z}_2^n + \mathbf{G}^n)\right] - \left[h(\mathbf{T}_2^n + \mathbf{G}^n|\mathbf{X}_2^n, \hat{V}) - h(\mathbf{Z}_1^n + \mathbf{Z}_2^n + \mathbf{G}^n)\right]\Big| \\
&\quad + |\epsilon| \cdot \left|h(\mathbf{X}_1^n + \mathbf{Z}_1^n|\hat{W}) - h(\mathbf{Z}_1^n) + h(\mathbf{Z}_1^n)\right| + |\epsilon| \cdot \left|h(\mathbf{X}_1^n + \mathbf{Z}_1^n|\hat{V},\hat{W}) - h(\mathbf{Z}_1^n) + h(\mathbf{Z}_1^n)\right| \\
&\overset{\text{(b)}}{\le} c_1\delta^2\mathbb{E}[\|\mathbf{X}_2^n\|^2] + |\epsilon|\Big(c_2\delta^2\mathbb{E}[\|\mathbf{X}_2^n\|^2] + c_3\mathbb{E}[\|\mathbf{X}_2^n\|^2] + c_4\mathbb{E}[\|\mathbf{X}_1^n\|^2] + 2|h(\mathbf{Z}_1^n)|\Big)
\end{aligned}$$

where (a) follows from chain rule, and (b) follows from Lemma 7, with $c_1, c_2, c_3, c_4 \ge 0$ being constants that only depends μ, α and on the distributions of $\hat{\mathbf{Z}}_1^n, \hat{\mathbf{Z}}_2^n, \mathbf{Z}_1^n, \mathbf{Z}_2^n$. Recall that $\mathbf{G}_i, \hat{\mathbf{G}}_i$'s are zero mean Gaussian random variables with identity covariance matrix.

This shows

$$\sup_{\substack{p(\mathbf{x}_1^n, \mathbf{x}_2^n, \hat{v}, \hat{w}) \\ p(\mathbf{x}_1^n, \mathbf{x}_2^n) \in \mathcal{D}}} \left|(\theta_{\mu,\alpha}^{(\epsilon,\delta)} - \theta_{\mu,\alpha}^{(0,0)})(\mathbf{X}_1^n, \mathbf{X}_2^n|\hat{V},\hat{W})\right| \to 0$$

as $(\epsilon, \delta) \to (0,0)$. This further implies that

$$\sup_{p(\mathbf{x}_1^n, \mathbf{x}_2^n) \in \mathcal{D}} \left|(\Theta_{\mu,\alpha}^{(\epsilon,\delta)} - \Theta_{\mu,\alpha}^{(0,0)})(\mathbf{X}_1^n, \mathbf{X}_2^n)\right| \to 0$$

as $(\epsilon, \delta) \to (0,0)$ which, by Lemma 8, implies the result. □

E Proof of Theorem 2

Lemma 9. *Let $A, B, K, \tilde{K}$ be square matrices such that $A \succeq B \succeq 0$ and $\tilde{K} \succeq K \succeq 0$. Then*

$$\log|K + B| - \log|K + A| \le \log|\tilde{K} + B| - \log|\tilde{K} + A|.$$

Proof. One can assume without loss of generality that $B = 0$, $A = I$ and $K \succ 0$. Since $\tilde{K} \succeq K$ we have $\tilde{K}^{-1} + I \preceq K^{-1} + I$ and hence

$$\log |\tilde{K}^{-1} + I| \leq \log |K^{-1} + I|$$

or equivalently

$$\log |K| - \log |K + I| \leq \log |\tilde{K}| - \log |\tilde{K} + I|.$$

Alternatively, the inequality is equivalent with $I(X; X + Z) \geq I(X; X + Z + \hat{Z})$ for mutually independent Gaussian random variables $X \sim \mathcal{N}(0, A - B)$, $Z \sim \mathcal{N}(0, K + B)$, $\hat{Z} \sim \mathcal{N}(0, \tilde{K} - K)$, and follows from the data processing inequality for mutual information. □

We denote $\mathrm{Cov}(\mathbf{X}|U) := \mathbb{E}[\mathbf{X}\mathbf{X}^T|U] - \mathbb{E}[\mathbf{X}|U]\mathbb{E}[\mathbf{X}|U]^T$. Therefore, using the tower property of conditional expectation we have

$$\begin{aligned}\mathbb{E}[\mathrm{Cov}(\mathbf{X}|U)] &= \mathbb{E}[\mathbf{X}\mathbf{X}^T] - \mathbb{E}[\mathbb{E}[\mathbf{X}|U]\mathbb{E}[\mathbf{X}|U]^T] \\ &= \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}|U])(\mathbf{X} - \mathbb{E}[\mathbf{X}|U])^T].\end{aligned}$$

Lemma 10. *Let $U, \mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2$ be random variables where $\mathbf{X}$ is real-vector-valued, $\mathbf{Z}_1, \mathbf{Z}_2$ are Gaussians and $(U, \mathbf{X})$, $\mathbf{Z}_1$, $\mathbf{Z}_2$ are mutually independent. Then*

$$\begin{aligned}&h(\mathbf{X} + \mathbf{Z}_1|U) - h(\mathbf{X} + \mathbf{Z}_1 + \mathbf{Z}_2|U) \\ &\leq \frac{1}{2}\Big[\log\big|\mathbb{E}[\mathrm{Cov}(\mathbf{X}|U)] + \mathrm{Cov}(\mathbf{Z}_1)\big| - \log\big|\mathbb{E}[\mathrm{Cov}(\mathbf{X}|U)] + \mathrm{Cov}(\mathbf{Z}_1) + \mathrm{Cov}(\mathbf{Z}_2)\big|\Big]\end{aligned}$$

Remark 6. *The scalar version of this result is rather well-known. This is Exercise 9.21 in [15], for instance. It is also possible that the vector case is known but we present a short proof here for completeness.*

Proof. We have

$$\begin{aligned}&h(\mathbf{X} + \mathbf{Z}_1|U) - h(\mathbf{X} + \mathbf{Z}_1 + \mathbf{Z}_2|U) \\ &\leq h(\mathbf{X} - \mathbb{E}[\mathbf{X}|U] + \mathbf{Z}_1|U) - h(\mathbf{X} - \mathbb{E}[\mathbf{X}|U] + \mathbf{Z}_1 + \mathbf{Z}_2|U) \\ &\leq \sup_{\substack{p(v,\tilde{\mathbf{x}}) \\ \mathbb{E}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T) \preceq \mathbb{E}(\mathrm{Cov}(\mathbf{X}|U)) \\ (V,\tilde{\mathbf{X}}) \perp (\mathbf{Z}_1, \mathbf{Z}_2)}} \Big[h(\tilde{\mathbf{X}} + \mathbf{Z}_1|V) - h(\tilde{\mathbf{X}} + \mathbf{Z}_1 + \mathbf{Z}_2|V)\Big] \\ &\overset{(a)}{=} \sup_{0 \preceq K \preceq \mathbb{E}(\mathrm{Cov}(\mathbf{X}|U))} \frac{1}{2}\Big[\log\big|K + \mathrm{Cov}(\mathbf{Z}_1)\big| - \log\big|K + \mathrm{Cov}(\mathbf{Z}_1) + \mathrm{Cov}(\mathbf{Z}_2)\big|\Big] \\ &\overset{(b)}{=} \frac{1}{2}\Big[\log\big|\mathbb{E}(\mathrm{Cov}(\mathbf{X}|U)) + \mathrm{Cov}(\mathbf{Z}_1)\big| - \log\big|\mathbb{E}(\mathrm{Cov}(\mathbf{X}|U)) + \mathrm{Cov}(\mathbf{Z}_1) + \mathrm{Cov}(\mathbf{Z}_2)\big|\Big]\end{aligned}$$

where (a) is a consequence of Theorem 1 of [14], and (b) follows from Lemma 9. □

Lemma 11. *Let $U, \mathbf{X}$ be random variables where $\mathbf{X}$ is real-vector-valued. Then*

$$h(\mathbf{X}|U) \leq \frac{1}{2}\log\big|2\pi e\mathbb{E}[\mathrm{Cov}(\mathbf{X}|U)]\big|.$$

Proof. We have

$$\begin{aligned}h(\mathbf{X}|U) &\overset{(a)}{\leq} \mathbb{E}\big[\frac{1}{2}\log\big|2\pi e\,\mathrm{Cov}(\mathbf{X}|U)\big|\big] \\ &\overset{(b)}{\leq} \frac{1}{2}\log\big|2\pi e\mathbb{E}[\mathrm{Cov}(\mathbf{X}|U)]\big|\end{aligned}$$

where (a) holds since Gaussian maximizes entropy and (b) follows from Jensen's inequality and concavity of log-determinant. □

Lemma 12. *Let $U, \mathbf{X}$ be random variables where $\mathbf{X}$ is real-vector-valued. Then*

$$\mathrm{Cov}(\mathbf{X}) - \mathbb{E}[\mathrm{Cov}(\mathbf{X}|U)] \succeq 0.$$

Proof. This follows from the law of total variance. □

Proof of Theorem 2. Let us call the left hand side and the right hand side v_L and v_R respectively.

We first show $v_L \leq v_R$. By Proposition 5 (ii) v_L admits a maximizing distribution $(\mathbf{X}_1, \mathbf{X}_2, V, W)$ such that $p(\mathbf{x}_1|w)$, $p(\mathbf{x}_1|v,w)$ and $p(\mathbf{x}_2|v)$ are Gaussians with covariance matrices independent of choice of v and w. Let

$$\begin{aligned} A_2 &:= \mathrm{Cov}(\mathbf{X}_2|V), \\ \begin{pmatrix} B_1 & \Sigma \\ \Sigma^T & B_2 \end{pmatrix} &:= \mathbb{E}\left[\mathrm{Cov}\left(\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\middle| W\right)\right] - \begin{pmatrix} 0 & 0 \\ 0 & A_2 \end{pmatrix}, \\ C_1 &:= \mathrm{Cov}(\mathbf{X}_1) - B_1, \\ C_2 &:= \mathrm{Cov}(\mathbf{X}_2) - B_2 - A_2. \end{aligned}$$

Note that since the variance of $\mathbf{X}_2$ given $V = v$ is independent of choice of v, A_2 is a deterministic matrix. Moreover, we utilized the notation defined earlier that

$$\mathrm{Cov}\left(\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\middle| W\right) = \mathbb{E}\left[\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}^T \middle| W\right] - \mathbb{E}\left[\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\middle| W\right]\mathbb{E}\left[\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\middle| W\right]^T$$

is a function of the random variable W, and

$$\begin{pmatrix} B_1 & \Sigma \\ \Sigma^T & A_2 + B_2 \end{pmatrix}$$

is set to be its expected value.

One can then verify

$$\begin{aligned} h(\mathbf{X}_2 + \mathbf{Z}_1 + \mathbf{Z}_2|V) - h(\mathbf{Z}_1 + \mathbf{Z}_2) &= \frac{1}{2}\Big[\log|A_2 + N_1 I + N_2 I| - \log|N_1 I + N_2 I|\Big], \\ h(\mathbf{X}_2 + \mathbf{X}_1 + \mathbf{Z}_1 + \mathbf{Z}_2|W) - h(\mathbf{X}_1 + \mathbf{Z}_1|W) &\overset{(a)}{\leq} \frac{1}{2}\Big[\log|B_1 + B_2 + \Sigma + \Sigma^T + A_2 + N_1 I + N_2 I| - \log|B_1 + N_1 I|\Big], \end{aligned}$$

and

$$\begin{aligned} &h(\mathbf{X}_1 + \mathbf{Z}_1|\mathbf{X}_2, W) - h(\mathbf{X}_1 + \mathbf{Z}_1 + \mathbf{Z}_2|\mathbf{X}_2, W) \\ &\overset{(b)}{\leq} \frac{1}{2}\Big[\log|\mathbb{E}[\mathrm{Cov}(\mathbf{X}_1|\mathbf{X}_2, W)] + N_1 I| - \log|\mathbb{E}[\mathrm{Cov}(\mathbf{X}_1|\mathbf{X}_2, W)] + N_1 I + N_2 I|\Big] \\ &\overset{(c)}{\leq} \frac{1}{2}\Big[\log|B_1 - \Sigma(A_2 + B_2)^{-1}\Sigma^T + N_1 I| - \log|B_1 - \Sigma(A_2 + B_2)^{-1}\Sigma^T + N_1 I + N_2 I|\Big], \end{aligned}$$

where (a) follows from Lemma 11 and the assumption that $p(\mathbf{x}_1|w)$ is Gaussian, (b) follows from Lemma 10, and (c) follows from Lemma 9 as shown below. Observe that the orthogonality property of conditional expectation (implying also the vector extension of minimum mean square error property) yields the following:

$$\begin{aligned} \mathbb{E}[\mathrm{Cov}(\mathbf{X}_1|\mathbf{X}_2, W)] &= \mathbb{E}[(\mathbf{X}_1 - \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2, W])(\mathbf{X}_1 - \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2, W])^T] \\ &\preceq \mathbb{E}[(\mathbf{X}_1 - \tilde{\mathbf{X}}_1)(\mathbf{X}_1 - \tilde{\mathbf{X}}_1)^T] \end{aligned}$$

for any $\tilde{\mathbf{X}}_1$ that is $\sigma(\mathbf{X}_2, W)$ measurable and $\mathbb{E}[\|\tilde{\mathbf{X}}_1\|^2] < \infty$. Set

$$\tilde{\mathbf{X}}_1 := \mathbb{E}[\mathbf{X}_1|W] + \mathbb{E}[(\mathbf{X}_1 - \mathbb{E}[\mathbf{X}_1|W])(\mathbf{X}_2 - \mathbb{E}[\mathbf{X}_2|W])^T]\mathbb{E}[(\mathbf{X}_2 - \mathbb{E}[\mathbf{X}_2|W])(\mathbf{X}_2 - \mathbb{E}[\mathbf{X}_2|W])^T]^{-1}(\mathbf{X}_2 - \mathbb{E}[\mathbf{X}_2|W]).$$

Now observe that

$$\mathbf{X}_1 - \tilde{\mathbf{X}}_1 = \hat{\mathbf{X}}_1 - \mathbb{E}[\hat{\mathbf{X}}_1\hat{\mathbf{X}}_2^T]\mathbb{E}[\hat{\mathbf{X}}_2\hat{\mathbf{X}}_2^T]^{-1}\hat{\mathbf{X}}_2$$

where $\hat{\mathbf{X}}_1 := \mathbf{X}_1 - \mathbb{E}[\mathbf{X}_1|W]$ and $\hat{\mathbf{X}}_2 := \mathbf{X}_2 - \mathbb{E}[\mathbf{X}_2|W]$. Therefore

$$\begin{aligned} \mathbb{E}[(\mathbf{X}_1 - \tilde{\mathbf{X}}_1)(\mathbf{X}_1 - \tilde{\mathbf{X}}_1)^T] &= \mathbb{E}[\hat{\mathbf{X}}_1\hat{\mathbf{X}}_1^T] - \mathbb{E}[\hat{\mathbf{X}}_1\hat{\mathbf{X}}_2^T]\mathbb{E}[\hat{\mathbf{X}}_2\hat{\mathbf{X}}_2^T]^{-1}\mathbb{E}[\hat{\mathbf{X}}_2\hat{\mathbf{X}}_1^T] \\ &= B_1 - \Sigma(A_2 + B_2)^{-1}\Sigma^T. \end{aligned}$$

Putting these together gives

$$f_{\mu,\alpha}(\mathbf{X}_1, \mathbf{X}_2|V, W) \leq g_{\mu,\alpha}(A_2, B_1, B_2, \Sigma).$$

Now we verify $A_2, B_1, B_2, C_1, C_2, \Sigma$ satisfy the constraints of v_R. We have

$$\mathbb{E}\left[\mathrm{Cov}\left(\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\middle| W\right)\right] \overset{(a)}{\succeq} \mathbb{E}\left[\mathrm{Cov}\left(\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\middle| V, W\right)\right] \overset{(b)}{=} \begin{pmatrix}\mathrm{Cov}(\mathbf{X}_1|V,W) & 0 \\ 0 & \mathrm{Cov}(\mathbf{X}_2|V)\end{pmatrix} \succeq \begin{pmatrix} 0 & 0 \\ 0 & A_2 \end{pmatrix}$$

where (a) follows from Lemma 12 and (b) holds since

$$\mathbf{X}_1 \to (V, W) \to \mathbf{X}_2, \qquad W \to V \to \mathbf{X}_2$$

form Markov chains. This gives

$$\begin{pmatrix} B_1 & \Sigma \\ \Sigma^T & B_2 \end{pmatrix} \succeq 0.$$

We also have

$$\begin{aligned} \begin{pmatrix} C_1 & -\Sigma \\ -\Sigma^T & C_2 \end{pmatrix} &= \begin{pmatrix} \mathrm{Cov}(\mathbf{X}_1) & 0 \\ 0 & \mathrm{Cov}(\mathbf{X}_2) \end{pmatrix} - \mathbb{E}\left[\mathrm{Cov}\left(\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \middle| W\right)\right] \\ &\overset{(a)}{=} \mathrm{Cov}\left(\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}\right) - \mathbb{E}\left[\mathrm{Cov}\left(\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \middle| W\right)\right] \overset{(b)}{\succeq} 0 \end{aligned}$$

where (a) holds since $\mathbf{X}_1 \perp \mathbf{X}_2$ and (b) follows from Lemma 12. The remaining constraints, namely,

$$A_2 \succeq 0, \qquad B_1 + C_1 \in \mathcal{K}_1, \qquad A_2 + B_2 + C_2 \in \mathcal{K}_2,$$

are obvious.

Next we show $v_L \geq v_R$. Suppose $A_2, B_1, B_2, C_1, C_2, \Sigma$ are matrices that satisfy the constraints of v_R. Let

$$\begin{pmatrix} \mathbf{A}_2 \\ \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{C}_1 \\ \mathbf{C}_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} A_2 & & & & \\ & B_1 & \Sigma & & \\ & \Sigma^T & B_2 & & \\ & & & C_1 & -\Sigma \\ & & & -\Sigma^T & C_2 \end{pmatrix}\right)$$

and let

$$\mathbf{X}_1 := \mathbf{B}_1 + \mathbf{C}_1, \qquad \mathbf{X}_2 := \mathbf{A}_2 + \mathbf{B}_2 + \mathbf{C}_2, \qquad V := \mathbf{B}_2 + \mathbf{C}_2, \qquad W := (\mathbf{C}_1, \mathbf{C}_2).$$

Then one can readily verify

$$f_{\mu,\alpha}(\mathbf{X}_1, \mathbf{X}_2 | V, W) = g_{\mu,\alpha}(A_2, B_1, B_2, \Sigma)$$

and that $(\mathbf{X}_1, \mathbf{X}_2, V, W)$ satisfies the constraints of v_L.

□

F Proof of Proposition 3

We will need Lemma 13 to justify the continuity of $G^{(1)}$.

Lemma 13. *Let (X, d) be a metric space. Suppose $f : X \to \mathbb{R}$ is Lipschitz, i.e. there exists a constant $C \in \mathbb{R}$ such that*

$$|f(x) - f(y)| \leq C \cdot d(x, y)$$

for any $x, y \in X$. Suppose S and S_n ($n \in \mathbb{N}$) are subsets of X such that S_n converges to S in Hausdorff distance, i.e.

$$\lim_n \max\left\{\sup_{x \in S_n} \inf_{y \in S} d(x, y), \sup_{x \in S} \inf_{y \in S_n} d(x, y)\right\} = 0.$$

Then

$$\lim_n \sup_{x \in S_n} f(x) = \sup_{x \in S} f(x).$$

Proof. For any $x \in S_n$ and $y \in S$ it holds that

$$f(x) \leq f(y) + C \cdot d(x, y)$$

and hence

$$\sup_{x \in S_n} f(x) \leq \sup_{y \in S} f(y) + C \cdot \sup_{x \in S_n} \inf_{y \in S} d(x, y).$$

Taking limit yields

$$\limsup_n \sup_{x \in S_n} f(x) \leq \sup_{x \in S} f(x).$$

Similarly for any $x \in S$ and $y \in S_n$ it holds that

$$f(x) \leq f(y) + C \cdot d(x, y)$$

and hence

$$\sup_{x\in S} f(x) \le \sup_{y\in S_n} f(y) + C \cdot \sup_{x\in S} \inf_{y\in S_n} d(x,y).$$

Taking limit yields

$$\sup_{x\in S} f(x) \le \liminf_n \sup_{x\in S_n} f(x).$$

□

Proof of Proposition 3 (i). Proposition 5 (i) (with $\mathcal{K}_{1i} := \{K_1 : \mathrm{tr}(K_1) \le P_1\}$ and $\mathcal{K}_{2i} := \{K_2 : \mathrm{tr}(K_2) \le P_2\}$) implies that

$$G^{(n)}(P_1,P_2) \ge n \cdot G^{(1)}(P_1,P_2)$$

as well as (with $\mathcal{K}_{1i} := \{K_1 : \mathrm{tr}(K_1) \le Q_{1i}\}$ and $\mathcal{K}_{2i} := \{K_2 : \mathrm{tr}(K_2) \le Q_{2i}\}$) that

$$G^{(n)}(P_1,P_2) \le \sup_{\substack{Q_{1i},Q_{2i}\ge 0\\ \sum_{i=1}^n Q_{1i}\le nP_1\\ \sum_{i=1}^n Q_{2i}\le nP_2}} \sum_{i=1}^n G^{(1)}(Q_{1i},Q_{2i}).$$

It then suffices to show that $(P_1,P_2)\mapsto G^{(1)}(P_1,P_2)$ is concave. Indeed Lemma 6 (with $\mathcal{K}_{ji} := \{K : \mathrm{tr}(K) \le P_{ji}\}$ and $t := \frac{1}{2}$) implies

$$G^{(1)}(P_{11},P_{21}) + G^{(1)}(P_{12},P_{22}) \le 2\cdot G^{(1)}(\frac{P_{11}+P_{12}}{2}, \frac{P_{21}+P_{22}}{2})$$

for any $P_{1i}, P_{2i} \ge 0$ $(i=1,2)$, i.e. $G^{(1)}$ is midpoint-concave. This together with the fact that $G^{(1)}$ is continuous, which can be shown by Lemma 13 by considering the matrix expression in Theorem 2, implies that $G^{(1)}$ is concave.

□

Proof of Proposition 3 (ii). In view of (i) it suffices to show the scalar case, i.e. $n=1$. Applying Lemma 9 one can get

$$g_{\mu,\alpha}(A_2,B_1,B_2,\Sigma) \le \tilde{g}_{\mu,\alpha}(A_2,B_1,B_2,\Sigma)$$

for scalars, where $g_{\mu,\alpha}$ is defined as in Theorem 2. This gives an upper bound for the matrix expression in Theorem 2:

$$G^{(1)}(P_1,P_2) \le \sup_{\substack{A_2,B_1,B_2,C_1,C_2\ge 0,\,\Sigma\in\mathbb{R}\\ \Sigma^2\le B_1B_2\\ \Sigma^2\le C_1C_2\\ B_1+C_1\le P_1\\ A_2+B_2+C_2\le P_2}} \tilde{g}_{\mu,\alpha}(A_2,B_1,B_2,\Sigma).$$

Now we simplify this maximization. The variables C_1, C_2 can be eliminated:

$$\sup_{\substack{A_2,B_1,B_2\ge 0,\,\Sigma\in\mathbb{R}\\ \Sigma^2\le B_1B_2\\ \Sigma^2\le (P_1-B_1)(P_2-A_2-B_2)\\ B_1\le P_1\\ A_2+B_2\le P_2}} \tilde{g}_{\mu,\alpha}(A_2,B_1,B_2,\Sigma).$$

Since the objective is increasing in Σ, we can put its maximum value:

$$\sup_{\substack{A_2,B_1,B_2\ge 0\\ B_1\le P_1\\ A_2+B_2\le P_2}} \tilde{g}_{\mu,\alpha}(A_2,B_1,B_2,\sqrt{\min\{B_1B_2,(P_1-B_1)(P_2-A_2-B_2)\}}).$$

We can further assume $B_1B_2 \le (P_1-B_1)(P_2-A_2-B_2)$ (otherwise we could increase the objective by increasing A_2 while fixing A_2+B_2):

$$\sup_{\substack{A_2,B_1,B_2\ge 0\\ B_1\le P_1\\ A_2+B_2\le P_2\\ B_1B_2\le (P_1-B_1)(P_2-A_2-B_2)}} \tilde{g}_{\mu,\alpha}(A_2,B_1,B_2,\sqrt{B_1B_2}).$$

We can further assume $B_1B_2 = (P_1-B_1)(P_2-A_2-B_2)$ (otherwise we would also have $A_2+B_2 < P_2$ and we could increase the objective by increasing A_2 or B_2 since the objective is increasing in both A_2 and B_2). If $B_1 \ne P_1$ then it means

$$A_2 = P_2 - \frac{P_1B_2}{P_1-B_1}. \tag{8}$$

If $B_1 = P_1$ then it implies $B_1 B_2 = 0$ and a maximizer is given by $A_2 = P_2$ and $B_2 = 0$. In both cases, (8) is satisfied. From equation (8) the constraint $A_2 + B_2 \leq P_2$ is automatically satisfied and the constraint $A_2 \geq 0$ is equivalent to $\frac{B_1}{P_1} + \frac{B_2}{P_2} \leq 1$. So we have:

$$\sup_{\substack{B_1, B_2 \geq 0 \\ \frac{B_1}{P_1} + \frac{B_2}{P_2} \leq 1}} \tilde{g}_{\mu,\alpha}(P_2 - \frac{P_1 B_2}{P_1 - B_1}, B_1, B_2, \sqrt{B_1 B_2})$$

where $\frac{0}{0}$ is understood to be 0. □

Proof of Proposition 3 (iii). The maximization is attained by $B_1 = B_2 = 0$ if and only if for any B_1, B_2 satisfying the constraints we have

$$\begin{aligned}
&\frac{1}{2}\Big[\mu\alpha \cdot \big[\log(P_2 - \frac{P_1 B_2}{P_1 - B_1} + N_1 + N_2) - \log(N_1 + N_2)\big] \\
&\quad + \mu(1-\alpha) \cdot \log(B_1 + B_2 + 2\sqrt{B_1 B_2} + P_2 - \frac{P_1 B_2}{P_1 - B_1} + N_1 + N_2) \\
&\quad + \log(B_1 + N_1) - (1 + \mu(1-\alpha)) \cdot \log(B_1 + N_1 + N_2)\Big] \\
\leq\ &\frac{1}{2}\Big[\mu\alpha \cdot \big[\log(P_2 + N_1 + N_2) - \log(N_1 + N_2)\big] \\
&\quad + \mu(1-\alpha) \cdot \log(P_2 + N_1 + N_2) + \log(N_1) - (1 + \mu(1-\alpha)) \cdot \log(N_1 + N_2)\Big]
\end{aligned}$$

or equivalently

$$\begin{aligned}
&\mu\alpha \cdot \log(1 - \frac{P_1 B_2}{P_1 - B_1} \frac{1}{P_2 + N_1 + N_2}) \\
&\quad + \mu(1-\alpha) \cdot \log(1 + \frac{\frac{B_1 + 2\sqrt{B_1 B_2} - \frac{B_1 B_2}{P_1 - B_1}}{P_2 + N_1 + N_2}(N_1 + N_2) - B_1}{B_1 + N_1 + N_2}) + \log(1 + \frac{B_1 N_2}{(B_1 + N_1 + N_2) N_1}) \leq 0.
\end{aligned}$$

By dividing the above inequality by $\mu + 1$ and utilizing concavity of $x \mapsto \log(1 + x)$, this is implied by

$$-\mu\alpha \cdot \frac{P_1 B_2}{P_1 - B_1} \frac{1}{P_2 + N_1 + N_2} + \mu(1-\alpha) \cdot \frac{\frac{B_1 + 2\sqrt{B_1 B_2} - \frac{B_1 B_2}{P_1 - B_1}}{P_2 + N_1 + N_2}(N_1 + N_2) - B_1}{B_1 + N_1 + N_2} + \frac{B_1 N_2}{(B_1 + N_1 + N_2) N_1} \leq 0$$

or equivalently

$$-\mu\alpha \frac{P_1(B_1 + N_1 + N_2)}{(P_1 - B_1)(N_1 + N_2)} B_2 + \mu(1-\alpha)\left[2\sqrt{B_1}\sqrt{B_2} - \frac{B_1}{P_1 - B_1} B_2 - \frac{B_1 P_2}{N_1 + N_2}\right] + \frac{B_1 N_2 (P_2 + N_1 + N_2)}{N_1 (N_1 + N_2)} \leq 0.$$

Since the left hand side is quadratic in $\sqrt{B_2}$ with negative leading coefficient, this is implied by

$$\mu^2(1-\alpha)^2 B_1 + \left[\mu\alpha \frac{P_1(B_1 + N_1 + N_2)}{(P_1 - B_1)(N_1 + N_2)} + \mu(1-\alpha)\frac{B_1}{P_1 - B_1}\right]\left[\frac{B_1 N_2 (P_2 + N_1 + N_2)}{N_1 (N_1 + N_2)} - \mu(1-\alpha)\frac{B_1 P_2}{N_1 + N_2}\right] \leq 0$$

or equivalently

$$\mu(1-\alpha)^2 (P_1 - B_1) + \left[\alpha(P_1 + \frac{P_1}{N_1 + N_2} B_1) + (1-\alpha) B_1\right]\left[\frac{N_2}{N_1}(1 + \frac{P_2}{N_1 + N_2}) - \mu(1-\alpha)\frac{P_2}{N_1 + N_2}\right] \leq 0.$$

Since left hand side is linear in B_1, this is true for all $B_1 \geq 0$ if the linear coefficient and the constant term are ≤ 0, that is

$$-\mu(1-\alpha)^2 + \left[\alpha \frac{P_1}{N_1 + N_2} + (1-\alpha)\right]\left[\frac{N_2}{N_1}(1 + \frac{P_2}{N_1 + N_2}) - \mu(1-\alpha)\frac{P_2}{N_1 + N_2}\right] \leq 0, \tag{9}$$

$$\mu(1-\alpha)^2 P_1 + \alpha P_1 \left[\frac{N_2}{N_1}(1 + \frac{P_2}{N_1 + N_2}) - \mu(1-\alpha)\frac{P_2}{N_1 + N_2}\right] \leq 0. \tag{10}$$

Inequality (10) implies

$$\frac{N_2}{N_1}(1 + \frac{P_2}{N_1 + N_2}) - \mu(1-\alpha)\frac{P_2}{N_1 + N_2} \leq 0.$$

which implies (9). Thus, it suffices to satisfy (10) since it implies (9). This is equivalent to

$$\frac{N_2}{N_1}(1+\frac{P_2}{N_1+N_2}) \leq \mu \cdot \left[(1-\alpha)\frac{P_2}{N_1+N_2} - \frac{(1-\alpha)^2}{\alpha}\right].$$

It can be shown that the right hand side is ≥ 0 if and only if $\alpha \geq \frac{N_1+N_2}{P_2+N_1+N_2}$, and is maximized by $\alpha = \sqrt{\frac{N_1+N_2}{P_2+N_1+N_2}}$. Putting this maximizing α and rearranging we get

$$\mu \geq \frac{N_2}{N_1}\frac{1}{\left(1-\sqrt{\frac{N_1+N_2}{P_2+N_1+N_2}}\right)^2} = \mu_0.$$

To conclude, if $\mu \geq \mu_0$ then letting $\alpha = \sqrt{\frac{N_1+N_2}{P_2+N_1+N_2}}$ gives that the maximization on the right hand side of (ii) is attained by $B_1 = B_2 = 0$. □

Proof of Proposition 3 (iv). This is immediate from (ii) and (iii). □