Uniqueness of local maximizers for some non-convex log-determinant optimization problems using information theory

Ken Lau, Chandra Nair, Chaorui Yao

Abstract

Certain families of non-convex optimization problems involving linear combinations of log-determinants of positive definite matrices are shown to have a unique local maximizer. These geometric results are established using information-theoretic arguments. We demonstrate these results for three different families: two of which arise in the study of the capacity region of the vector Gaussian broadcast channel and another one in the study of computing the optimal generalized Brascamp-Lieb constant.

I. Introduction

The results in this paper are motivated by explicit evaluations of capacity regions in multiuser information theory and that of optimal constants in certain families of inequalities. One of the fundamental problems in multiuser information theory is to determine the capacity region of a broadcast channel [1]. For a vector Gaussian broadcast channel with only private message decoding requirements, the capacity region was determined via a very interesting and indirect relaxation (channel enhancement) in [2]. This proof employed the use of Entropy Power Inequality [3], [4]. Entropy power inequality is a fundamental entropic inequality in functional analysis and convex geometry that is usually proved via the calculus of variations where a monotonically increasing path is constructed from a pair of independent distributions to a pair of standard Gaussians. This proof technique of employing the entropy power inequality or of constructing a monotonic path could not be easily employed in the case of the two-receiver vector Gaussian broadcast channel with private and common message decoding requirements.

The capacity region for the private and common message setting was obtained in [5] by using arguments combining sub-additivity and rotations to deduce the Gaussian optimality of a particular information functional. This technique, while demonstrating the Gaussian optimality of an associated functional in a rather direct manner, did not provide (at that time) geometric insights such as the existence of a monotonic path to the maximizer from any point (thereby demonstrating uniqueness of the local maximizer and hence global maximizer). This paper addresses this concern and shows how the very same arguments that were used to deduce the Gaussian optimality, can also be used to recover the uniqueness of local maximizers and a local direction along which the function is always increasing from a non-optimal starting point. This paper was also motivated by the recent result [6], where the authors constructed a different functional in this setting and demonstrated a monotonicity property for that functional.

Apart from the two settings of the vector Gaussian broadcast channel, the paper also considers the family of unified Brascamp-Lieb and Entropy Power Inequalities, [7], and demonstrate a similar uniqueness of the local maximizer for the non-convex problem of computing the optimal constant. Computing the optimal Brascamp-Lieb-inequality constant has garnered attention over the years from math and the computer science communities [8], [9]. In particular the work [8] provides algorithmic guarantees for the computation of the Brascamp-Lieb constant, going beyond the more modest goals of showing the uniqueness of the local maximizer that is attained in this paper. On the other hand, the functional considered in this paper is a unified formulation that generalizes Brascamp-Lieb inequalities and the entropy power inequalities.

Remark 1. *There are many more functionals that can be inferred from the network information theory literature but the examples listed here are chosen because the first two are very relevant to wireless communications and the third one is of broad interest across several communities.*

The use of information theory tools for proving inequalities about determinants of positive semi-definite matrices can be traced back to [10]. Of course, the arguments here are a more involved and rely on fancier proofs of sub-additivity of the associated functionals.

II. Two Receiver vector Gaussian Broadcast Channel with private messages

Consider the following optimization problem defined in the space of positive semi-definite matrices. We will use this as a motivating example for the results of this paper.

Optimization Problem 1. *Let $\Sigma_1, \Sigma_2 \succ 0$ be two positive definite matrices and let $\lambda > 1$ be a constant. Consider*

$$\max_{K: 0 \preceq K \preceq K_0} \log|K + \Sigma_1| - \lambda \log|K + \Sigma_2|.$$

As $\log |K|$ is concave in the space of positive definite matrices, it is evident that Optimization Problem 1 defines a non-convex optimization problem.

Proposition 1. *Let $\lambda > 1$ be a real number and $\Sigma_1, \Sigma_2 \succ 0$ be two $d \times d$ positive definite matrices. Let $K_0 \succeq 0$ be any positive semi-definite matrix. Consider the bounded continuous function*

$$f(K) := \log |K + \Sigma_1| - \lambda \log |K + \Sigma_2|$$

defined on the convex domain $K : 0 \preceq K \preceq K_0$. Then the function, $f(K)$, has a unique local (hence global) maximizer K^* in the domain, and further for any, $K \neq K^*$, $K : 0 \preceq K \preceq K_0$, both the functions $u_a(t) := f(K_{a,t})$ and $u_b(t) = f(K_{b,t})$, satisfy $u_a(1) = f(K^*) > u_a(t) = f(K_{a,t}) > u_a(0) = f(K)$, for any $t \in (0, 1)$, and similarly for $u_b(t)$. Here

$$K_{a,t} = (1-t)K + tK^* - t(1-t)(K - K^*)[tK + (1-t)K^* + \Sigma_2]^{-1}(K - K^*),$$
$$K_{b,t} = (1-t)K + tK^* - t(1-t)(K - K^*)[tK + (1-t)K^* + \Sigma_1]^{-1}(K - K^*).$$

Remark 2. *It turns out (and this is the main point of this paper) this proposition has an elegant information-theoretic proof, and this is not a one-off phenomenon.*

Proof. Let $\mathbf{X} \sim \mathcal{N}(0, K)$ be a zero-mean Gaussian distribution (in $\mathbb{R}^d$) with covariance matrix K satisfying $0 \preceq K \preceq K_0$. Suppose $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are zero-mean Gaussians with positive definite covariance Σ_1 and Σ_2 respectively, independent of $\mathbf{X}$. Define $\mathbf{Y}_1 = \mathbf{X} + \mathbf{Z}_1$ and $\mathbf{Y}_2 = \mathbf{X} + \mathbf{Z}_2$. Let

$$\begin{aligned} g(K) &:= h(\mathbf{Y}_1) - \lambda h(\mathbf{Y}_2) \\ &= h(\mathbf{X} + \mathbf{Z}_1) - \lambda h(\mathbf{X} + \mathbf{Z}_2) \\ &= \frac{1}{2} \log |K + \Sigma_1| - \frac{\lambda}{2} \log |K + \Sigma_2| - \frac{(\lambda - 1)d}{2} \log(2\pi e) \\ &= \frac{1}{2} f(K) - \frac{(\lambda - 1)d}{2} \log(2\pi e). \end{aligned}$$

Observe that $g(K)$ is just a scaled version of $f(K)$ that is shifted by a constant. Therefore it suffices to prove the Proposition with $f(K)$ replaced by $g(K)$.

Since $f(K)$ is a bounded continuous function on the compact set $K : 0 \preceq K \preceq K_0$, it (and correspondingly $g(K)$) attains its maximum value at some point (not necessarily unique) in the set, say K^*. For $K \neq K^*$, $K : 0 \preceq K \preceq K_0$, Let $\mathbf{X}$ and $\mathbf{X}^*$ be independent Gaussians with distributions $\mathcal{N}(0, K)$ and $\mathcal{N}(0, K^*)$ respectively. Let $\mathbf{Z}_{11}$ and $\mathbf{Z}_{12}$ be two i.i.d. copies of $\mathbf{Z}_1$. Similarly, let $\mathbf{Z}_{21}$ and $\mathbf{Z}_{22}$ be two i.i.d. copies of $\mathbf{Z}_2$. Here the collection $(\mathbf{X}, \mathbf{X}^*, \mathbf{Z}_{11}, \mathbf{Z}_{12}, \mathbf{Z}_{21}, \mathbf{Z}_{22})$ are assumed to be mutually independent. Define $\mathbf{Y}_1 = \mathbf{X} + \mathbf{Z}_{11}$, $\mathbf{Y}_2 = \mathbf{X} + \mathbf{Z}_{21}$, $\mathbf{Y}_1^* = \mathbf{X}^* + \mathbf{Z}_{12}$ and $\mathbf{Y}_2^* = \mathbf{X}^* + \mathbf{Z}_{22}$.

Define $(\mathbf{X}_{t+}, \mathbf{X}_{t-})$ to be the rotated versions of $(\mathbf{X}, \mathbf{X}^*)$ as follows,

$$\begin{bmatrix} \mathbf{X}_{t+} \\ \mathbf{X}_{t-} \end{bmatrix} = \begin{bmatrix} \sqrt{1-t} & \sqrt{t} \\ \sqrt{t} & -\sqrt{1-t} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{X}^* \end{bmatrix}$$

We define $\mathbf{Y}_{1t+}, \mathbf{Y}_{1t-}, \mathbf{Y}_{2t+}, \mathbf{Y}_{2t-}$ similarly.

Let v and V^* denote the value of $g(K)$ and $g(K^*)$, we have

$$\begin{aligned} v + V^* &= g(K) + g(K^*) \\ &= h(\mathbf{Y}_1) - \lambda h(\mathbf{Y}_2) + h(\mathbf{Y}_1^*) - \lambda h(\mathbf{Y}_2^*) \\ &= h(\mathbf{Y}_1, \mathbf{Y}_1^*) - \lambda h(\mathbf{Y}_2, \mathbf{Y}_2^*) \\ &\overset{(a)}{=} h(\mathbf{Y}_{1t+}, \mathbf{Y}_{1t-}) - \lambda h(\mathbf{Y}_{2t+}, \mathbf{Y}_{2t-}) \\ &= h(\mathbf{Y}_{1t+}|\mathbf{Y}_{2t-}) - \lambda h(\mathbf{Y}_{2t+}|\mathbf{Y}_{2t-}) + h(\mathbf{Y}_{1t-}|\mathbf{Y}_{1t+}) - \lambda h(\mathbf{Y}_{2t-}|\mathbf{Y}_{1t+}) - (\lambda - 1)I(\mathbf{Y}_{1t+}; \mathbf{Y}_{2t-}), \end{aligned}$$

where (a) follows since rotations preserve entropies. Note that$\mathbf{Z}_{1t+} := \sqrt{1-t}\mathbf{Z}_{11} + \sqrt{t}\mathbf{Z}_{12}$ is independent of $\mathbf{Z}_{1t-} := \sqrt{t}\mathbf{Z}_{11} - \sqrt{1-t}\mathbf{Z}_{12}$, and $\mathbf{X}_+$ can be expressed as $\mathbf{X}_{t+} = \sqrt{t(1-t)}(K - K^*)[tK + (1-t)K^* + \Sigma_2]^{-1}\mathbf{Y}_{2t-} + \mathbf{X}_{t+}^\dagger$, where $\mathbf{X}_{t+}^\dagger$ is independent of $\mathbf{Y}_{2t-}$. It is immediate that $\mathbf{Z}_{1t+}, \mathbf{Z}_{2t+}, \mathbf{X}_{t+}^\dagger$, and $\mathbf{Y}_{2t-}$ are mutually independent. Further $\mathbf{X}_{t+}^\dagger \sim \mathcal{N}(0, K_{a,t})$ where $K_{a,t} = tK + (1-t)K^* - t(1-t)(K - K^*)[tK + (1-t)K^* + \Sigma_2]^{-1}(K - K^*)$. Therefore

$$h(\mathbf{Y}_{1t+}|\mathbf{Y}_{2t-}) - \lambda h(\mathbf{Y}_{2t+}|\mathbf{Y}_{2t-}) = g(K_{a,t}).$$

In a similar fashion we can write $\mathbf{X}_{t-} = \sqrt{t(1-t)}(K - K^*)[tK + (1-t)K^* + \Sigma_2]^{-1}\mathbf{Y}_{1t+} + \mathbf{X}_{t-}^\ddagger$ where $\mathbf{Z}_{1t-}, \mathbf{Z}_{2t-}, \mathbf{X}_{t-}^\ddagger$, and $\mathbf{Y}_{1t+}$ are mutually independent. Further $\mathbf{X}_{t-}^\ddagger \sim \mathcal{N}(0, K_{b,1-t})$ where $K_{b,t} = tK + (1-t)K^* - t(1-t)(K - K^*)[tK + (1-t)K^* + \Sigma_1]^{-1}(K - K^*)$. Therefore

$$h(\mathbf{Y}_{1t-}|\mathbf{Y}_{1t+}) - \lambda h(\mathbf{Y}_{2t-}|\mathbf{Y}_{1t+}) = g(K_{b,1-t}).$$

Putting this together we have

$$\begin{aligned}v + V^* &= g(K) + g(K^*)\\ &= g(K_{a,t}) + g(K_{b,1-t}) - (\lambda - 1)I(\mathbf{Y}_{1t+}; \mathbf{Y}_{2t-}).\end{aligned}$$

Observe that if $K \neq K^*$, then $\mathbf{Y}_{1t+}$ and $\mathbf{Y}_{2t-}$ are not independent for any $t \in (0,1)$, as $E(\mathbf{Y}_{1t+}\mathbf{Y}_{2t-}^T) = \sqrt{t(1-t)}(K - K^*) \neq 0$. Therefore if K^* is a global maximizer and as $K \neq K^*$ is a point in the domain $K : 0 \preceq K \preceq K_0$, we have

$$\begin{aligned}g(K) + g(K^*) &= g(K_{a,t}) + g(K_{b,1-t}) - (\lambda - 1)I(\mathbf{Y}_{1t+}; \mathbf{Y}_{2t-})\\ &< g(K_{a,t}) + g(K_{b,1-t})\\ &\leq 2g(K^*),\end{aligned}$$

where the last inequality being due to K^* is a global maximizer, and that $K_{a,t}$ and $K_{b,1-t}$ belong to the domain $K : 0 \preceq K \preceq K_0$. Therefore $g(K) < g(K^*)$, implying the uniqueness of the global maximizer. Consequently, for $t \in (0,1)$ since $K_{b,1-t} \neq K^*$, we have

$$0 < g(K^*) - g(K_{b,1-t}) < g(K_{a,t}) - g(K).$$

Therefore $g(K_{a,t}) > g(K)$ for any $t \in (0,1)$. Similarly, we have

$$0 < g(K^*) - g(K_{a,t}) < g(K_{b,1-t}) - g(K).$$

Hence $g(K_{b,t}) > g(K)$ for any $t \in (0,1)$. This shows that $u_a(t)$ and $u_b(t)$ have the desired properties as claimed in the Proposition.

Note that as $t \to 0$, we have $K_{a,t}, K_{b,t} \to K$, and hence $g(K)$ cannot be a local maximizer. Since this is true for any $K : 0 \preceq K \preceq K_0$, and $K \neq K^*$, the function $f(K)$ has a unique local maximizer in the domain as required. □

Remark 1. The two possible paths $K_{a,t}$ and $K_{b,t}$ in the Proposition 1 yield gradients at each point along which the function is strictly increasing. The gradients at K are respectively: $(K^* - K) - (K^* - K)(K^* + \Sigma_2)^{-1}(K^* - K)$ and $(K^* - K) - (K^* - K)(K^* + \Sigma_1)^{-1}(K^* - K)$. This induces, informally, a gradient flow on the space on the feasible set of positive semi-definite matrices, that can be used to deduce a monotonically increasing flow from any point to the global maximizer.

III. Gaussian Broadcast Channel with Common Message

In this section, we will consider an optimization problem associated with the computation of the capacity region of the two-receiver Gaussian broadcast channel with private and common messages [5].

Optimization Problem 2. *Let $\Sigma_1, \Sigma_2 \succ 0$ be two positive definite matrices and let $\lambda_0 > \lambda_2 > 1$ and $\alpha \in [0,1]$ be constants. Consider*

$$\max_{\substack{K,\hat{K} \succeq 0\\ K+\hat{K} \preceq K_0}} (\lambda_2 - \lambda_0\bar{\alpha}) \log |K + \hat{K} + \Sigma_2| - \lambda_0\alpha \log |K + \hat{K} + \Sigma_1| + \log |K + \Sigma_1| - \lambda_2 \log |K + \Sigma_2|,$$

where $\bar{\alpha} := 1 - \alpha$.

In spite of the above function being non-convex, we will show, as in the previous section, that the function has a unique local maximizer (hence global maximizer).

Proposition 2. *Let $\lambda_0 > \lambda_2 > 1$ and $\alpha \in [0,1]$ be real numbers and $\Sigma_1, \Sigma_2 \succ 0$ be two $d \times d$ positive definite matrices. Let $K_0 \succeq 0$ be any positive semi-definite matrix. Consider the bounded continuous function*

$$f(K, \hat{K}) := (\lambda_2 - \lambda_0\bar{\alpha}) \log |K + \hat{K}_1 + \Sigma_2| - \lambda_0\alpha \log |K + \hat{K}_1 + \Sigma_1| + \log |K + \Sigma_1| - \lambda_2 \log |K + \Sigma_2|$$

defined on the convex domain $(K, \hat{K}_1) : K, \hat{K}_1 \succeq 0; K + \hat{K}_1 \preceq K_0$. Then the function, $f(K, \hat{K})$, has a unique local maximizer $(K^*, \hat{K}^*)$ in the domain, and further for any, $(K, \hat{K}) \neq (K^*, \hat{K}^*)$, $(K, \hat{K}) : K, \hat{K} \succeq 0, K + \hat{K} \preceq K_0$, both the functions $u_a(t) := f(K_{a,t}, \hat{K}_{a,t})$ and $u_b(t) = f(K_{b,t}, \hat{K}_{b,t})$, satisfy $u_a(1) = f(K^*, \hat{K}^*) > u_a(t) = f(K_{a,t}, \hat{K}_{a,t}) > u_a(0) = f(K, \hat{K})$, for any $t \in (0,1)$, and similarly for $u_b(t)$. Here

$$\begin{aligned}K_{a,t} &= (1-t)K + tK^* - t(1-t)(K - K^*)[tK + (1-t)K^* + \Sigma_2]^{-1}(K - K^*),\\ K_{a,t} + \hat{K}_{a,t} &= (1-t)(K + \hat{K}) + t(K^* + \hat{K}^*)\\ &\quad - t(1-t)[(K + \hat{K}) - (K^* + \hat{K}^*)][t(K + \hat{K}) + (1-t)(K^* + \hat{K}^*) + \Sigma_2]^{-1}[(K + \hat{K}) - (K^* + \hat{K}^*)].\end{aligned}$$

The alternate path is

$$K_{b,t} = (1-t)K + tK^* - t(1-t)(K - K^*)[tK + (1-t)K^* + \Sigma_1]^{-1}(K - K^*),$$

$$K_{b,t}+\hat{K}_{b,t}=(1-t)(K+\hat{K})+t(K^*+\hat{K}^*)$$
$$-t(1-t)[(K+\hat{K})-(K^*+\hat{K}^*)][t(K+\hat{K})+(1-t)(K^*+\hat{K}^*)+\Sigma_1]^{-1}[(K+\hat{K})-(K^*+\hat{K}^*)].$$

Proof. Let $\mathbf{U}\sim\mathcal{N}(0,K)$ and $\mathbf{V}\sim\mathcal{N}(0,\hat{K})$ be zero-mean Gaussian distributions with covariance matrix K and $\hat{K}$ satisfying $0\preceq K,\hat{K}$ and $K+\hat{K}\preceq K_0$. Suppose $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are zero-mean Gaussians with positive definite covariance Σ_1 and Σ_2 respectively.

Define $\mathbf{Y}_1=\mathbf{U}+\mathbf{V}+\mathbf{Z}_1$, $\mathbf{Y}_2=\mathbf{U}+\mathbf{V}+\mathbf{Z}_2$, $\hat{\mathbf{Y}}_1=\mathbf{U}+\mathbf{Z}_1$ and $\hat{\mathbf{Y}}_2=\mathbf{U}+\mathbf{Z}_2$. Let

$$\begin{aligned}g(K,\hat{K})&:=(\lambda_2-\lambda_0\bar{\alpha})h(\mathbf{Y}_2)-\lambda_0\alpha h(\mathbf{Y}_1)+h(\hat{\mathbf{Y}}_1)-\lambda_2 h(\hat{\mathbf{Y}}_2)\\&=\frac{1}{2}(\lambda_2-\lambda_0\bar{\alpha})\log|K+\hat{K}+\Sigma_2|-\frac{\lambda_0}{2}\alpha\log|K+\hat{K}+\Sigma_1|\\&\quad+\frac{1}{2}\log|K+\Sigma_1|-\frac{\lambda_2}{2}\log|K+\Sigma_2|-\frac{(\lambda_0-1)d}{2}\log(2\pi e)\\&=\frac{1}{2}f(K,\hat{K})-\frac{(\lambda_0-1)d}{2}\log(2\pi e).\end{aligned}$$

As in the earlier section, it suffices to prove the Proposition with $f(K,\hat{K})$ replaced by $g(K,\hat{K})$.

Since $f(K,\hat{K})$ is a bounded continuous function on the compact set $(K,\hat{K}):K,\hat{K}\succeq 0,K+\hat{K}\preceq K_0$, it (and correspondingly $g(K,\hat{K})$) attains its maximum value at some point (not necessarily unique) in the set, say $(K^*,\hat{K}^*)$.

For $(K,\hat{K})\neq(K^*,\hat{K}^*)$, $(K,\hat{K}):K,\hat{K}\succeq 0,K+\hat{K}\preceq K_0$. Let $\mathbf{U},\mathbf{V}$ and $\mathbf{U}^*,\mathbf{V}^*$ be independent Gaussians with distribution $\mathcal{N}(0,K),\mathcal{N}(0,\hat{K})$ and $\mathcal{N}(0,K^*),\mathcal{N}(0,\hat{K}^*)$ respectively. Let $\mathbf{Z}_{11}$ and $\mathbf{Z}_{12}$ be two i.i.d. copies of $\mathbf{Z}_1$. Similarly, let $\mathbf{Z}_{21}$ and $\mathbf{Z}_{22}$ be two i.i.d. copies of $\mathbf{Z}_2$. Here the collection $(\mathbf{U},\mathbf{V},\mathbf{U}^*,\mathbf{V}^*,\mathbf{Z}_{11},\mathbf{Z}_{12},\mathbf{Z}_{21},\mathbf{Z}_{22})$ are assumed to be mutually independent. Define $\mathbf{Y}_{11}=\mathbf{U}+\mathbf{V}+\mathbf{Z}_{11}$, $\mathbf{Y}_{21}=\mathbf{U}+\mathbf{V}+\mathbf{Z}_{21}$, $\hat{\mathbf{Y}}_{11}=\mathbf{U}+\mathbf{Z}_{11}$ and $\hat{\mathbf{Y}}_{21}=\mathbf{U}+\mathbf{Z}_{21}$. Also, we define $\mathbf{Y}_{12}=\mathbf{U}^*+\mathbf{V}^*+\mathbf{Z}_{12}$, $\mathbf{Y}_{22}=\mathbf{U}^*+\mathbf{V}^*+\mathbf{Z}_{22}$, $\hat{\mathbf{Y}}_{12}=\mathbf{U}^*+\mathbf{Z}_{11}$ and $\hat{\mathbf{Y}}_{22}=\mathbf{U}^*+\mathbf{Z}_{22}$.

Define $(\mathbf{U}_{t+},\mathbf{U}_{t-})$ to be the rotated versions of $(\mathbf{U},\mathbf{U}^*)$ as follows,

$$\begin{bmatrix}\mathbf{U}_{t+}\\\mathbf{U}_{t-}\end{bmatrix}=\begin{bmatrix}\sqrt{1-t}&\sqrt{t}\\\sqrt{t}&-\sqrt{1-t}\end{bmatrix}\begin{bmatrix}\mathbf{U}\\\mathbf{U}^*\end{bmatrix}$$

We define $\mathbf{V}_{t+}$, $\mathbf{V}_{t-}$, $\mathbf{Y}_{1t+}$, $\mathbf{Y}_{1t-}$, $\mathbf{Y}_{2t+}$, $\mathbf{Y}_{2t-}$, $\hat{\mathbf{Y}}_{1t+}$, $\hat{\mathbf{Y}}_{1t-}$, $\hat{\mathbf{Y}}_{2t+}$, $\hat{\mathbf{Y}}_{2t-}$ similarly.

Let v and V^* denote the value of $g(K,\hat{K})$ and $g(K^*,\hat{K}^*)$.

$$\begin{aligned}v+V^*&=g(K,\hat{K})+g(K^*,\hat{K}^*)\\&=(\lambda_2-\lambda_0\bar{\alpha})h(\mathbf{Y}_{21})-\lambda_0\alpha h(\mathbf{Y}_{11})+h(\hat{\mathbf{Y}}_{11})-\lambda_2 h(\hat{\mathbf{Y}}_{21})\\&\quad+(\lambda_2-\lambda_0\bar{\alpha})h(\mathbf{Y}_{22})-\lambda_0\alpha h(\mathbf{Y}_{12})+h(\hat{\mathbf{Y}}_{12})-\lambda_2 h(\hat{\mathbf{Y}}_{22})\\&=(\lambda_2-\lambda_0\bar{\alpha})h(\mathbf{Y}_{21},\mathbf{Y}_{22})-\lambda_0\alpha h(\mathbf{Y}_{11},\mathbf{Y}_{12})\\&\quad+h(\hat{\mathbf{Y}}_{11},\hat{\mathbf{Y}}_{12})-\lambda_2 h(\hat{\mathbf{Y}}_{21},\hat{\mathbf{Y}}_{22})\\&\overset{(a)}{=}(\lambda_2-\lambda_0\bar{\alpha})h(\mathbf{Y}_{2t+},\mathbf{Y}_{2t-})-\lambda_0\alpha h(\mathbf{Y}_{1t+},\mathbf{Y}_{1t-})\\&\quad+h(\hat{\mathbf{Y}}_{1t+},\hat{\mathbf{Y}}_{1t-})-\lambda_2 h(\hat{\mathbf{Y}}_{2t+},\hat{\mathbf{Y}}_{2t-})\\&=(\lambda_2-\lambda_0\bar{\alpha})h(\mathbf{Y}_{2t+}|\mathbf{Y}_{2t-})-\lambda_0\alpha h(\mathbf{Y}_{1t+}|\mathbf{Y}_{2t-})\\&\quad+h(\hat{\mathbf{Y}}_{1t+}|\hat{\mathbf{Y}}_{2t-})-\lambda_2 h(\hat{\mathbf{Y}}_{2t+}|\hat{\mathbf{Y}}_{1t-})\\&\quad+(\lambda_2-\lambda_0\bar{\alpha})h(\mathbf{Y}_{2t-}|\mathbf{Y}_{1t+})-\lambda_0\alpha h(\mathbf{Y}_{1t-}|\mathbf{Y}_{1t+})\\&\quad+h(\hat{\mathbf{Y}}_{1t-}|\hat{\mathbf{Y}}_{1t+})-\lambda_2 h(\hat{\mathbf{Y}}_{2t-}|\hat{\mathbf{Y}}_{1t+})\\&\quad-(\lambda_0-\lambda_2)I(\mathbf{Y}_{1t+};\mathbf{Y}_{2t-})-(\lambda_2-1)I(\hat{\mathbf{Y}}_{1t+};\hat{\mathbf{Y}}_{2t-}).\end{aligned}$$

where (a) follows since rotations preserve entropies. Note that $\mathbf{Z}_{1t+}:=\sqrt{1-t}\mathbf{Z}_{11}+\sqrt{t}\mathbf{Z}_{12}$ is independent of $\mathbf{Z}_{1t-}:=\sqrt{t}\mathbf{Z}_{11}-\sqrt{1-t}\mathbf{Z}_{12}$, and $\mathbf{U}_{t+}$ can be expressed as $\mathbf{U}_{t+}=\sqrt{t(1-t)}(K-K^*)[tK+(1-t)K^*+\Sigma_2]^{-1}(\hat{\mathbf{Y}}_{2t-})+\mathbf{U}_{t+}^{\dagger}$, where $\mathbf{U}_{t+}^{\dagger}$ is independent of $\hat{\mathbf{Y}}_{2t-}$. It is immediate that $\mathbf{Z}_{1t+},\mathbf{Z}_{2t+},\mathbf{U}_{t+}^{\dagger}$, and $\hat{\mathbf{Y}}_{2t-}$ are mutually independent. Further $\mathbf{U}_{t+}^{\dagger}\sim\mathcal{N}(0,K_{a,t})$ where $K_{a,t}=(1-t)K+tK^*-t(1-t)(K-K^*)[tK+(1-t)K^*+\Sigma_2]^{-1}(K-K^*)$. In a similar fashion we can write $\mathbf{U}_{t+}+\mathbf{V}_{t+}=\sqrt{t(1-t)}[(K+\hat{K})-(K^*+\hat{K}^*)][t(K+\hat{K})+(1-t)(K^*+\hat{K}^*)+\Sigma_2]^{-1}(\mathbf{Y}_{2t-})+(\mathbf{U}_{t+}^{\dagger}+\mathbf{V}_{t+}^{\dagger})$ where $\mathbf{Z}_{1t+},\mathbf{Z}_{2t+},\mathbf{V}_{t+}^{\dagger}$, and $\mathbf{Y}_{2t-}$ are mutually independent. Further $\mathbf{U}_{t+}^{\dagger}+\mathbf{V}_{t+}^{\dagger}\sim\mathcal{N}(0,K_{a,t}+\hat{K}_{a,t})$ where

$$K_{a,t}+\hat{K}_{a,t}=(1-t)(K+\hat{K})+t(K^*+\hat{K}^*)$$
$$-t(1-t)[(K+\hat{K})-(K^*+\hat{K}^*)][t(K+\hat{K})+(1-t)(K^*+\hat{K}^*)+\Sigma_2]^{-1}[(K+\hat{K})-(K^*+\hat{K}^*)].$$

Let $E(\text{Cov}(\mathbf{X}|\mathbf{Y})) := E((\mathbf{X} - E(\mathbf{X}|\mathbf{Y}))(\mathbf{X} - E(\mathbf{X}|\mathbf{Y}))^T)$. Note that

$$\begin{aligned}
E(\text{Cov}(\mathbf{X}|\mathbf{Y})) &= E((\mathbf{X} - E(\mathbf{X}|\mathbf{Y}))(\mathbf{X} - E(\mathbf{X}|\mathbf{Y}))^T) \\
&= E((\mathbf{X} - E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) + E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) - E(\mathbf{X}|\mathbf{Y}))(\mathbf{X} - E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) + E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) - E(\mathbf{X}|\mathbf{Y}))^T) \\
&\overset{(a)}{=} E((\mathbf{X} - E(\mathbf{X}|\mathbf{Y},\mathbf{Z}))(\mathbf{X} - E(\mathbf{X}|\mathbf{Y},\mathbf{Z}))^T) + E((E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) - E(\mathbf{X}|\mathbf{Y}))(E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) - E(\mathbf{X}|\mathbf{Y}))^T) \\
&\succeq E((\mathbf{X} - E(\mathbf{X}|\mathbf{Y},\mathbf{Z}))(\mathbf{X} - E(\mathbf{X}|\mathbf{Y},\mathbf{Z}))^T) \\
&= E(\text{Cov}(\mathbf{X}|\mathbf{Y},\mathbf{Z})).
\end{aligned}$$

where (a) uses the orthogonality property of conditional expectation which implies that $E((\mathbf{X} - E(\mathbf{X}|\mathbf{Y},\mathbf{Z}))(E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) - E(\mathbf{X}|\mathbf{Y}))^T) = 0$ since $E(\mathbf{X}|\mathbf{Y},\mathbf{Z}) - E(\mathbf{X}|\mathbf{Y})$ is measurable w.r.t $\sigma(\mathbf{Y},\mathbf{Z})$.

Observe that

$$\begin{aligned}
K_{a,t} &= E(\text{Cov}(\mathbf{U}_{t+}|\mathbf{U}_{t_-} + \mathbf{Z}_{2t-})) \\
&= E(\text{Cov}(\mathbf{U}_{t+}|\mathbf{U}_{t_-} + \mathbf{Z}_{2t-}, \mathbf{V}_{t+}, \mathbf{V}_{t-})) && \because (\mathbf{V}_{t+}, \mathbf{V}_{t-}) \perp (\mathbf{U}_{t+}, \mathbf{U}_{t-}, \mathbf{Z}_{2t-}) \\
&= E(\text{Cov}(\mathbf{U}_{t+} + \mathbf{V}_{t+}|\mathbf{U}_{t_-} + \mathbf{Z}_{2t-}, \mathbf{V}_{t+}, \mathbf{V}_{t-})) \\
&= E(\text{Cov}(\mathbf{U}_{t+} + \mathbf{V}_{t+}|\mathbf{U}_{t_-} + \mathbf{Z}_{2t-} + \mathbf{V}_{t-}, \mathbf{V}_{t+}, \mathbf{V}_{t-})) \\
&\preceq E(\text{Cov}(\mathbf{U}_{t+} + \mathbf{V}_{t+}|\mathbf{U}_{t_-} + \mathbf{Z}_{2t-} + \mathbf{V}_{t-})) \\
&= K_{a,t} + \hat{K}_{a,t}.
\end{aligned}$$

This implies that $\hat{K}_{a,t} \succeq 0$.

Therefore

$$(\lambda_2 - \lambda_0\bar{\alpha})h(\mathbf{Y}_{2t+}|\mathbf{Y}_{2t-}) - \lambda_0\alpha h(\mathbf{Y}_{1t+}|\mathbf{Y}_{2t-}) + h(\hat{\mathbf{Y}}_{1t+}|\hat{\mathbf{Y}}_{2t-}) - \lambda_2 h(\hat{\mathbf{Y}}_{2t+}|\hat{\mathbf{Y}}_{1t-}) = g(K_{a,t}, \hat{K}_{a,t}).$$

By a similar procedure, we may also conclude that

$$\begin{aligned}
&(\lambda_2 - \lambda_0\bar{\alpha})h(\mathbf{Y}_{2t-}|\mathbf{Y}_{1t+}) - \lambda_0\alpha h(\mathbf{Y}_{1t-}|\mathbf{Y}_{1t+}) + \\
&\quad + h(\hat{\mathbf{Y}}_{1t-}|\hat{\mathbf{Y}}_{1t+}) - \lambda_2 h(\hat{\mathbf{Y}}_{2t-}|\hat{\mathbf{Y}}_{1t+}) = g(K_{b,1-t}, \hat{K}_{b,1-t}).
\end{aligned}$$

Putting this together we have

$$\begin{aligned}
v + V^* &= g(K, \hat{K}) + g(K^*, \hat{K}^*) \\
&= g(K_{a,t}, \hat{K}_{a,t}) + g(K_{b,1-t}, \hat{K}_{b,1-t}) - (\lambda_0 - \lambda_2)I(\mathbf{Y}_{1t+}; \mathbf{Y}_{2t-}) - (\lambda_2 - 1)I(\hat{\mathbf{Y}}_{1t+}; \hat{\mathbf{Y}}_{2t-}).
\end{aligned}$$

Observe that if $(K, \hat{K}) \neq (K^*, \hat{K}^*)$, then either $\hat{\mathbf{Y}}_{1t+}$ and $\hat{\mathbf{Y}}_{2t-}$ are not independent, or $\mathbf{Y}_{1t+}$ and $\mathbf{Y}_{2t-}$ are not independent for any $t \in (0,1)$. This follows since $E(\hat{\mathbf{Y}}_{1t+}\hat{\mathbf{Y}}_{2t-}^T) = \sqrt{t(1-t)}(K - K^*)$ and $E(\mathbf{Y}_{1t+}\mathbf{Y}_{2t-}^T) = \sqrt{t(1-t)}((K + \hat{K}) - (K^* + \hat{K}^*))$. Therefore if $K^*, \hat{K}^*$ is a global maximizer and $(K, \hat{K}) \neq (K^*, \hat{K}^*)$ is a point in the domain $(K, \hat{K}) : K, \hat{K} \succeq 0, K + \hat{K} \preceq K_0$, then we have

$$\begin{aligned}
g(K, \hat{K}) + g(K^*, \hat{K}^*) &= g(K_{a,t}, \hat{K}_{a,t}) + g(K_{b,1-t}, \hat{K}_{b,1-t}) - (\lambda_0 - \lambda_2)I(\mathbf{Y}_{1t+}; \mathbf{Y}_{2t-}) \\
&\quad - (\lambda_2 - 1)I(\hat{\mathbf{Y}}_{1t+}; \hat{\mathbf{Y}}_{2t-}) \\
&< g(K_{a,t}, \hat{K}_{a,t}) + g(K_{b,1-t}, \hat{K}_{b,1-t}) \\
&\leq 2g(K^*, \hat{K}^*),
\end{aligned}$$

where the last inequality being due to $K_{a,t}, K_{b,t}$ and $K_{b,1-t}, \hat{K}_{b,1-t}$ belong to the domain $(K, \hat{K}) : K, \hat{K} \succeq 0, K + \hat{K} \preceq K_0$. Therefore $g(K, \hat{K}) < g(K^*, \hat{K}^*)$, implying the uniqueness of the global maximizer. Consequently, for $t \in (0,1)$ since $K_{b,1-t}, \hat{K}_{b,1-t} \neq (K^*, \hat{K}^*)$, we have

$$0 < g(K^*, \hat{K}^*) - g(K_{b,1-t}, \hat{K}_{b,1-t}) < g(K_{a,t}, \hat{K}_{a,t}) - g(K, \hat{K}).$$

Therefore $g(K_{a,t}, \hat{K}_{a,t}) > g(K, \hat{K})$ for any $t \in (0,1)$. Similarly, we have

$$0 < g(K^*, \hat{K}^*) - g(K_{a,t}, \hat{K}_{a,t}) < g(K_{b,1-t}, \hat{K}_{b,1-t}) - g(K, \hat{K}).$$

Hence $g(K_{b,t}, \hat{K}_{b,t}) > g(K)$ for any $t \in (0,1)$. This shows that $u_a(t)$ and $u_b(t)$ have the desired properties as claimed in the Proposition.

Note that as $t \to 0$, we have $(K_{a,t}, \hat{K}_{a,t}) \to (K, \hat{K})$, and hence $g(K, \hat{K})$ cannot be a local maximizer. Since this is true for any $(K, \hat{K}) : K, \hat{K} \succeq 0, K + \hat{K} \preceq K_0$, and $(K, \hat{K}) \neq (K^*, \hat{K}^*)$, the function $f(K, \hat{K})$ has a unique local maximizer in the domain as required. $\square$

IV. Generalized Brascamp-Lieb Inequality

Optimization Problem 3. *Let $\{d_i\}_{i=1}^k, \{c_j\}_{j=1}^m$ be positive real numbers. Let $\epsilon > 0$ be a constant. Let $A_j : \mathbb{R}^n \mapsto \mathbb{R}^{m_j}, j = 1, 2, .., m$ be surjective linear maps. Let $n_1, .., n_k$ be positive natural numbers such that $\sum_{i=1}^k n_i = n$.*

$$\sup_{\{K_i\}_{i=1}^k : 0 \preceq K_i \preceq K_{0i}} \sum_{i=1}^k d_i \log |K_i| - \sum_{j=1}^m c_j \log |A_j K A_j^T + \epsilon I_{m_j}|.$$

Here $K = diag(K_1, K_2, .., K_k)$ is a block diagonal positive definite matrix.

Remark 2. When $k = 1$, $\epsilon = 0$, and the upper bound K_0 is removed, the above optimization problem reduces to that of computing the optimal Brascamp-Lieb constants. In the form written above, this corresponds to a unified inequality studied in [7].

Proposition 3. *Let $\{d_i\}_{i=1}^k, \{c_j\}_{j=1}^m$ be positive real numbers. Let $\epsilon > 0$ be a constant. Let $A_j : \mathbb{R}^n \mapsto \mathbb{R}^{m_j}, j = 1, 2, .., m$ be surjective linear maps. Let $n_1, .., n_k$ be positive natural numbers such that $\sum_{i=1}^k n_i = n$. Let $\{K_{0i}\}_{i=1}^k$ be given positive definite matrices. Consider the bounded continuous function*

$$f(K_1, .., K_k) := \sum_{i=1}^k d_i \log |K_i| - \sum_{j=1}^m c_j \log |A_j K A_j^T + \epsilon I_{m_j}|,$$

where $K = diag(K_1, K_2, .., K_k)$, defined on the convex domain $K_i : 0 \preceq K_i \preceq K_{0i}$. The function, $f(K_1, ..., K_k)$, has a unique local (hence global) maximizer $K^* = (K_1^*, K_2^*, ..., K_k^*)$ in the domain, and further for any, $K \neq K^*$, both the functions $u_a(t) := f(K_{a,t})$ and $u_b(t) = f(K_{b,t})$, satisfy $u_a(1) = f(K^*, \hat{K}^*) > u_a(t) = f(K_{a,t}, \hat{K}_{a,t}) > u_a(0) = f(K, \hat{K})$, for any $t \in (0, 1)$, and similarly for $u_b(t)$. Here

$$\begin{aligned}
&K_{a,t} = (1-t)K + tK^*, \\
&K_{b,t} = (1-t)K + tK^* - t(1-t)(K - K^*)[tK + (1-t)K^*]^{-1}(K - K^*).
\end{aligned}$$

Proof. Let $\mathbf{X}_i \sim \mathcal{N}(0, K_i), i = 1, .., k$ be mutually independent Gaussian random variables with covariance matrices K_i satisfying $0 \prec K_i \preceq K_{0i}$. Denote $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k)$. Let $\mathbf{Z}_j \sim \mathcal{N}(0, I_{m_j})$ be a collection of mutually independent standard Gaussians, that are also independent of $\mathbf{X}$. Define $\mathbf{Y}_j = A_j \mathbf{X} + \sqrt{\epsilon} \mathbf{Z}_j$ for $j = 1, .., m$. Define

$$\begin{aligned}
g(K_1, .., K_k) &:= \sum_{i=1}^k d_i h(\mathbf{X}_i) - \sum_{j=1}^m c_j h(\mathbf{Y}_j) \\
&= \frac{1}{2}\left(\sum_{i=1}^k d_i \log |K_i| - \sum_{j=1}^m c_j \log |A_j K A_j^T + \epsilon I_{m_j}|\right) + \frac{1}{2}\left(\sum_{i=1}^k d_i n_i - \sum_{j=1}^m c_j m_j\right) \log(2\pi e) \\
&= \frac{1}{2} f(K_1, .., K_k) + \frac{1}{2}\left(\sum_{i=1}^k d_i n_i - \sum_{j=1}^m c_j m_j\right) \log(2\pi e).
\end{aligned}$$

Since $f(K_1, \dots, K_k)$ is a bounded continuous function on the compact set $K_i : 0 \preceq K_i \preceq K_{0i}$, it (and correspondingly $g(K_1, \dots, K_k)$) attains its maximum value at some point (not necessarily unique) in the set, say $\{K_i^*\}_{i=1}^k$. For any feasible $(K_1, ..., K_k) \neq (K_1^*, .., K_k^*)$, let $\mathbf{X}_i$ and $\mathbf{X}_i^*$ be independent Gaussians with distribution $\mathcal{N}(0, K_i)$ and $\mathcal{N}(0, K_i^*)$ respectively. Let $\mathbf{Z}_{j1}$ and $\mathbf{Z}_{j2}$ be two i.i.d. copies of $\mathbf{Z}_j$. Here the collection $(\{\mathbf{X}_i\}, \{\mathbf{X}_i^*\}, \{\mathbf{Z}_{j1}\}, \{\mathbf{Z}_{j2}\}$ are assumed to be mutually independent. Define $\mathbf{Y}_j = A_j \mathbf{X} + \mathbf{Z}_{j1}$, and $\mathbf{Y}_j^* = A_j \mathbf{X}^* + \mathbf{Z}_{j2}$.

Define $(\mathbf{X}_{t+}, \mathbf{X}_{t-})$ to be the rotated versions of $(\mathbf{X}, \mathbf{X}^*)$ as follows,

$$\begin{bmatrix} \mathbf{X}_{it+} \\ \mathbf{X}_{it-} \end{bmatrix} = \begin{bmatrix} \sqrt{1-t} & \sqrt{t} \\ \sqrt{t} & -\sqrt{1-t} \end{bmatrix} \begin{bmatrix} \mathbf{X_i} \\ \mathbf{X_i}^* \end{bmatrix}$$

We define $\mathbf{Y}_{1t+}, \mathbf{Y}_{1t-}, \mathbf{Y}_{2t+}, \mathbf{Y}_{2t-}$ similarly. Let v and V^* denote the value of $g(K_1, .., K_k)$ and $g(K_1^*, .., K_k^*)$. Now, observe that

$$\begin{aligned}
g(K_1, .., K_k) + g(K_1^*, .., K_k^*) &= \sum_{i=1}^k d_i h(\mathbf{X}_i, \mathbf{X}_i^*) - \sum_{j=1}^m c_j h(\mathbf{Y}_j, \mathbf{Y}_j^*) \\
&= \sum_{i=1}^k d_i h(\mathbf{X}_{it+}, \mathbf{X}_{it-}) - \sum_{j=1}^m c_j h(\mathbf{Y}_{jt+}, \mathbf{Y}_{jt-})
\end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{=} \sum_{i=1}^{k} d_i h(\mathbf{X}_{it+}) - \sum_{j=1}^{m} c_j h(\mathbf{Y}_{jt+}) \\
&\quad + \sum_{i=1}^{k} d_i h(\mathbf{X}_{it-}|\mathbf{X}_{t+}) - \sum_{j=1}^{m} c_j h(\mathbf{Y}_{jt-}|\mathbf{X}_{t+}) - \sum_{j=1}^{m} c_j I(\mathbf{Y}_{jt-};\mathbf{X}_{t+}|\mathbf{Y}_{jt+}) \\
&= g(K_{a1,t},...,K_{ak,t}) + g(K_{b1,1-t},...,K_{bk,1-t}) - \sum_{j=1}^{m} c_j I(\mathbf{Y}_{jt-};\mathbf{X}_{t+}|\mathbf{Y}_{jt+}),
\end{aligned}$$

where $K_{ai,t} = (1-t)K_i + tK_i^*$ and $K_{bi,t} = (1-t)K_i + tK_i^* - t(1-t)(K_i - K_i^*)[tK_i + (1-t)K_i^*]^{-1}(K_i - K_i^*)$. The equality (a) uses the Markov chain $\mathbf{Y}_{jt-} \to \mathbf{X}_{t+} \to \mathbf{Y}_{jt+}$.

Suppose $K = (K_1,..,K_k)$ was another global maximizer, then we would have

$$\begin{aligned}
2V^* &= g(K_{a1,t},...,K_{ak,t}) + g(K_{b1,1-t},...,K_{bk,1-t}) \\
&\quad - \sum_{j=1}^{m} c_j I(\mathbf{Y}_{jt-};\mathbf{X}_{t+}|\mathbf{Y}_{jt+}).
\end{aligned}$$

As $K_{a,t}$ and $K_{b,(1-t)}$ are feasible choices, this necessitates that $V^* = g(K_{a1,t},...,K_{ak,t}) = g(K_{b1,1-t},...,K_{bk,1-t})$ and $I(\mathbf{Y}_{jt-};\mathbf{X}_{t+}|\mathbf{Y}_{jt+}) = 0$ for each $j = 1,...,m$. Therefore we have the Markov chain $\mathbf{Y}_{jt-} \to \mathbf{Y}_{jt+} \to \mathbf{X}_{t+}$. Since we also have the Markov chain $\mathbf{Y}_{jt-} \to \mathbf{X}_{t+} \to \mathbf{Y}_{jt+}$, this induces a double Markovity condition (see Exercise 16.25 in [11] and Remark 19 in [12]) and consequently (crucially using the fact that $\mathbf{Y}_{jt+}$ has an independent additive Gaussian $\mathbf{Z}_{jt+}$ that renders the support condition in Remark 19 of [12]) we can conclude that $\mathbf{Y}_{jt-}$ is independent of $(\mathbf{X}_{t+}, \mathbf{Y}_{jt+})$. This implies that

$$0 = \mathrm{E}\left(A_j \mathbf{X}_{t-} (A_j \mathbf{X}_{t+})^T\right) = t(1-t) A_j (K - K^*) A_j^T.$$

Hence $\sum_{j=1}^m c_j h(\mathbf{Y}_{jt+})$ does not depend on t, implying since $V^* = g(K_{a1,t},...,K_{ak,t})$ that $\sum_{i=1}^k d_i h(\mathbf{X}_{it+})$, also does not depend on t. From the strict concavity of $\log|K|$ on the space of positive definite matrices and as $K_{at,i} = (1-t)K_i + tK_i^*$, we must have $K_i = K_i^*$ for $i = 1,...,k$. This establishes the uniqueness of the global maximizer.

Now, as in the previous sections, if $K \neq K^*$, then we have

$$\begin{aligned}
v + V^* &= g(K_{a1,t},...,K_{ak,t}) + g(K_{b1,1-t},...,K_{bk,1-t}) - \sum_{j=1}^{m} c_j I(\mathbf{Y}_{jt-};\mathbf{X}_{t+}|\mathbf{Y}_{jt+}) \\
&\leq g(K_{a1,t},...,K_{ak,t}) + g(K_{b1,1-t},...,K_{bk,1-t}).
\end{aligned}$$

Now from the uniqueness of the global maximizer, for any $t \in (0,1)$, we have that both $g(K_{a1,t},...,K_{ak,t})$, $g(K_{b1,1-t},...,K_{bk,1-t}) < V^*$, implying that both $g(K_{a1,t},...,K_{ak,t})$, $g(K_{b1,1-t},...,K_{bk,1-t}) > v$, showing the uniqueness of the local maximizer as desired. $\square$

Remark 3. *Note that the path $K_{a,t}$ is the one employed by Stam in his proof of the entropy power inequality [13]. However the path in [9] is different from both $K_{a,t}$ and $K_{b,t}$*

CONCLUSION

Designing efficient algorithms using the geometric insights for the first two settings, in spite of the problem being non-convex, is an important problem for future research. From a mathematical perspective, identifying the gradient flow and then perhaps using that in discrete probability distributions may help yield insights for optimality of Marton's region for the two receiver discrete memoryless broadcast channel.

ACKNOWLEDGEMENTS

Chandra Nair would like to thank Jun Chen for discussions about the rotation arguments and the optimization aspects of the rotation method.

REFERENCES

[1] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, Jan 1972.

[2] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept 2006.

[3] C. E. Shannon, "A mathermatical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October, 1948.

[4] A. Stam, "Some inequalities satisfied by the quantities of information of fisher and shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019995859903481

[5] 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.

[6] Y. Xu and G. Chen, “New proofs of extremal inequalities with applications,” *CoRR*, vol. abs/2109.00377, 2021. [Online]. Available: https://arxiv.org/abs/2109.00377
[7] V. Anantharam, V. Jog, and C. Nair, “Unifying the brascamp-lieb inequality and the entropy power inequality,” *CoRR*, vol. abs/1901.06619, 2019. [Online]. Available: http://arxiv.org/abs/1901.06619v1
[8] A. Garg, L. Gurvits, R. M. de Oliveira, and A. Wigderson, “Algorithmic aspects of brascamp-lieb inequalities,” *CoRR*, vol. abs/1607.06711, 2016. [Online]. Available: http://arxiv.org/abs/1607.06711
[9] S. Sra, N. K. Vishnoi, and O. Yildiz, “On Geodesically Convex Formulations for the Brascamp-Lieb Constant,” in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), E. Blais, K. Jansen, J. D. P. Rolim, and D. Steurer, Eds., vol. 116. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, pp. 25:1–25:15. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2018/9429
[10] T. Cover and A. Gamal, “An information - theoretic proof of hadamard’s inequality (corresp.),” *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 930–931, 1983.
[11] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge University Press, 1 2011.
[12] A. Gohari and C. Nair, “Outer bounds for multiuser settings: The auxiliary receiver approach,” *IEEE Transactions on Information Theory*, vol. 68, no. 2, pp. 701–736, 2022.
[13] A. Stam, “Some inequalities satisfied by the quantities of information of fisher and shannon,” *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0019995859903481