# On Geodesically Convex Formulations for the Brascamp-Lieb Constant

**Suvrit Sra**
Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

**Nisheeth K. Vishnoi**
École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

**Ozan Yıldız**
École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

**Abstract**

We consider two non-convex formulations for computing the optimal constant in the Brascamp-Lieb inequality corresponding to a given datum and show that they are geodesically log-concave on the manifold of positive definite matrices endowed with the Riemannian metric corresponding to the Hessian of the log-determinant function. The first formulation is present in the work of Lieb [15] and the second is new and inspired by the work of Bennett *et al.* [5]. Recent work of Garg *et al.* [12] also implies a geodesically log-concave formulation of the Brascamp-Lieb constant through a reduction to the operator scaling problem. However, the dimension of the arising optimization problem in their reduction depends exponentially on the number of bits needed to describe the Brascamp-Lieb datum. The formulations presented here have dimensions that are polynomial in the bit complexity of the input datum.

**2012 ACM Subject Classification** Theory of computation → Nonconvex optimization, Theory of computation → Convex optimization

**Keywords and phrases** Geodesic convexity, positive definite cone, geodesics, Brascamp-Lieb constant

**Digital Object Identifier** 10.4230/LIPIcs.APPROX-RANDOM.2018.25

**Acknowledgements** We thank Ankit Garg for useful discussions on operator scaling and its connections to Brascamp-Lieb constant.

## 1 Introduction

**The Brascamp-Lieb Inequality.** Brascamp and Lieb [7] presented a class of inequalities that generalize many well-known inequalities and, as a consequence, have played an important role in various mathematical disciplines. Formally, they presented the following class of inequalities where each inequality is described by a "datum", referred to as the Brascamp-Lieb datum.

▶ **Definition 1** (The Brascamp-Lieb Inequality, Datum, Constant)**.** Let $n$, $m$, and $(n_j)_{j\in[m]}$ be positive integers and $p := (p_j)_{j\in[m]}$ be non-negative real numbers. Let $B := (B_j)_{j\in[m]}$ be an $m$-tuple of linear transformations where $B_j$ is a surjective linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^{n_j}$. The corresponding Brascamp-Lieb *datum* is denoted by $(B, p)$. The Brascamp-Lieb inequality states that for each Brascamp-Lieb datum $(B, p)$ there exists a constant $C(B, p)$ (not necessarily finite) such that for any selection of real-valued, non-negative, Lebesgue


Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018).
Editors: Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer; Article No. 25; pp. 25:1–25:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

measurable functions $f_j$ where $f_j : \mathbb{R}^{n_j} \to \mathbb{R}$,

$$\int_{x\in\mathbb{R}^n}\left(\prod_{j\in[m]} f_j(B_jx)^{p_j}\right)dx \le C(B,p)\prod_{j\in[m]}\left(\int_{x\in\mathbb{R}^{n_j}} f_j(x)dx\right)^{p_j}. \tag{1}$$

The smallest constant that satisfies (1) for any choice of $f := (f_j)_{j\in[m]}$ satisfying the properties mentioned above is called the Brascamp-Lieb *constant* and is denoted by $\mathrm{BL}(B,p)$. A Brascamp-Lieb datum $(B,p)$ is called *feasible* if $\mathrm{BL}(B,p)$ is finite, otherwise, it is called *infeasible*. For a given $m$-tuple $B$, the set of real vectors $p$ such that $(B,p)$ is feasible is denoted by $P_B$.

Applications of the Brascamp-Lieb inequality extend beyond functional analysis and appear in convex geometry [3], information theory [8],[16],[17], machine learning [14], and theoretical computer science [11, 10].

**Mathematical Aspects of the Brascamp-Lieb Inequality.** A Brascamp-Lieb inequality is non-trivial only when $(B,p)$ is a feasible Brascamp-Lieb datum. Therefore, it is of interest to characterize feasible Brascamp-Lieb data and compute the corresponding Brascamp-Lieb constant. Lieb [15] showed that one needs to consider only Gaussian functions as inputs for (1). This result suggests the following characterization of the Brascamp-Lieb constant as an optimization problem. For a positive integer $k$, let $\mathbb{S}_+^k$ be the space of real-valued, symmetric, positive semi-definite (PSD) matrices of dimension $k\times k$.

▶ **Theorem 2** (Gaussian maximizers [15]). *Let $(B,p)$ be a Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j\times n}$ for each $j\in[m]$. Let $A := (A_j)_{j\in[m]}$ with $A_j \in \mathbb{S}_+^{n_j}$, and consider the function*

$$\mathrm{BL}(B,p;A) := \left(\frac{\prod_{j\in[m]}\det(A_j)^{p_j}}{\det(\sum_{j\in[m]} p_jB_j^\top A_jB_j)}\right)^{1/2}. \tag{2}$$

*Then, the Brascamp-Lieb constant for $(B,p)$, $\mathrm{BL}(B,p)$ is equal to* $\sup_{A\in\times_{j\in[m]}\mathbb{S}_+^{n_j}} \mathrm{BL}(B,p;A)$.

Bennett *et al.* [5] proved the following necessary and sufficient conditions for the feasibility of a Brascamp-Lieb datum.

▶ **Theorem 3** (Feasibility of Brascamp-Lieb Datum [5], Theorem 1.15). *Let $(B,p)$ be a Brascamp-Lieb datum with $B_j\in\mathbb{R}^{n_j\times n}$ for each $j\in[m]$. Then, $(B,p)$ is feasible if and only if following conditions hold:*

1. $n = \sum_{j\in[m]} p_jn_j$, *and*
2. $\dim(V) \le \sum_{j\in[m]} p_j\dim(B_jV)$ *for any subspace $V$ of $\mathbb{R}^n$.*

Theorem 3 introduces infinitely many linear constraints on $p$ as $V$ varies over different subspaces of $\mathbb{R}^n$. However, there are only finitely many different linear restrictions as $\dim(B_jV)$ can only take integer values from $[n_j]$. Consequently, this theorem implies that $P_B$ is a convex set and, in particular, a polytope. It is referred to as the Brascamp-Lieb polytope, see e.g. [4] the "rank one" case ($n_j = 1$ for all $j$) and [5] for the general case. Some of the above inequality constraints are tight for any $p\in P_B$ such as the inequality constraints induced by $\mathbb{R}^n$ and the trivial subspace, while others can be strict for some $p\in P_B$. If $p$ lies on the boundary of $P_B$, then there should be some non-trivial subspaces $V$ such that the induced inequality constraints are tight for $p$. This leads to the definition of critical subspaces and simple Brascamp-Lieb datums.

▶ **Definition 4** (Critical Subspaces and Simple Brascamp-Lieb Data,[9], [5], Definition 1.12)**.** Let $(B, p)$ be a feasible Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j \times n}$ for each $j \in [m]$. Then, a subspace $V$ of $\mathbb{R}^n$ is called *critical* if

$$\dim(V) = \sum_{j \in [m]} p_j \dim(B_j V).$$

$(B, p)$ is called *simple* if there is no non-trivial proper subspace of $\mathbb{R}^n$ which is critical.

For a fixed $B$, simple Brascamp-Lieb data correspond to points $p$ that lie in the relative interior of the Brascamp-Lieb polytope $P_B$. One important property of simple Brascamp-Lieb data is that there exists a maximizer for $\mathrm{BL}(B, p; A)$. This was proved by Bennett *et al.* [5] by analyzing Lieb's formulation (2). This analysis also leads to a characterization of maximizers of $\mathrm{BL}(B, p; A)$.

▶ **Theorem 5** (Characterization of Maximizers[5], Theorem 7.13)**.** *Let $(B, p)$ be a Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j \times n}$ and $p_j > 0$ for all $j \in [m]$. Let $A := (A_j)_{j \in [m]}$ be an $m$-tuple of positive semidefinite matrices with $A_j \in \mathbb{R}^{n_j \times n_j}$ and let $M := \sum_{j \in [m]} p_j B_j^\top A_j B_j$. Then, the following statements are equivalent,*

1. *$A$ is a global maximizer for $\mathrm{BL}(B, p; A)$ as in (2).*
2. *$A$ is a local maximizer for $\mathrm{BL}(B, p; A)$.*
3. *$M$ is invertible and $A_j^{-1} = B_j M^{-1} B_j^\top$ for each $j \in [m]$.*

*Furthermore, the global maximizer $A$ for $\mathrm{BL}(B, p; A)$ exists and is unique up to scalar if and only if $(B, p)$ is simple.*

**Computational Aspects of the Brascamp-Lieb Inequality.** One of the computational questions concerning the Brascamp-Lieb inequality is: Given a Brascamp-Lieb datum $(B, p)$, can we compute $\mathrm{BL}(B, p)$ in time that is polynomial in the number of bits required to represent the datum? Since computing $\mathrm{BL}(B, p)$ exactly may not be possible due to the fact that this number may not be rational even if the datum $(B, p)$ is, one seeks an arbitrarily good approximation. Formally, given the entries of $B$ and $p$ in binary, and an $\varepsilon > 0$, compute a number $Z$ such that

$$\mathrm{BL}(B, p) \leq Z \leq (1 + \varepsilon)\, \mathrm{BL}(B, p)$$

in time that is polynomial in the combined bit lengths of $B$ and $p$ and $\log 1/\varepsilon$.

There are a few obstacles to this problem: (a) Checking if a given Brascamp-Lieb datum is feasible is not known to be in **P**. (b) The formulation of the Brascamp-Lieb constant by Lieb [15] as in (2) is neither concave nor log-concave. Thus, techniques developed in the context of linear and convex optimization do not seem to be directly applicable.

A step towards the computability of $\mathrm{BL}(B, p)$ was taken recently by Garg *et al.* [12] where they presented a pseudo-polynomial time algorithm for (a) and a pseudo-polynomial time algorithm to compute $\mathrm{BL}(B, p)$. The running time of this algorithm to compute $\mathrm{BL}(B, p)$ up to multiplicative error $1+\varepsilon$ has a polynomial dependency to $\varepsilon^{-1}$ and the *magnitude* of the denominators in the components of $p$ rather than the number of bits required to represent them. Garg *et al.* presented a reduction of the problem of computing $\mathrm{BL}(B, p)$ to the problem of computing the "capacity" in an "operator scaling" problem considered by Gurvits [13]. Roughly, in the operator scaling problem, given a representation of a linear mapping from PSD matrices to PSD matrices, the goal is to compute the minimum "distortion" of this mapping; see Section A for their reduction from Brascamp-Lieb to operator scaling. The operator scaling problem is also not a concave or log-concave optimization problem. However, operator scaling is known to be "geodesically" log-concave; see [1].

Geodesic convexity is an extension of convexity with respect to straight lines in Euclidean spaces to geodesics in Riemannian manifolds. Since all the problems mentioned so far are defined on positive definite matrices, the natural manifold to consider is the space of positive definite matrices with a particular Riemannian metric: the Hessian of the $-\log\det$ function; see section 2. Geodesics are analogs of straight lines on a manifold and, roughly, a function $f$ on is said to be geodesically convex if the average of its values at the two endpoints of any geodesic is at least its value at its mid-point.

The reduction of Garg *et al.* [12], thus, leads to a geodesically log-concave formulation to compute $\mathrm{BL}(B, p)$. However, this construction does not lead to an optimization problem whose dimension is polynomial in the input bit length as the size of constructed positive linear operator in the operator scaling problem depends exponentially on the bit lengths of the entries of $p$. More precisely, if $p_j = {}^{c_j}/_{c}$ for integers $(c_j)_{j\in[m]}$ and $c$, then the aforementioned construction results in operators over the space of real-valued, symmetric, positive definite (PD) matrices of dimension $(nc) \times (nc)$, $\mathbb{S}^{nc}_{++}$.

**Our Contribution.** Our first result is that Lieb's formulation presented in Theorem 2 is jointly geodesically log-concave with respect to inputs $(A_j)_{j\in[m]}$.

▶ **Theorem 6** (Geodesic Log-Concavity of Lieb's Formulation). *Let $(B, p)$ be a feasible Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j\times n}$ for each $j \in [m]$. Then, $\mathrm{BL}(B, p; A)$ is jointly geodesically log-concave with respect to $A := (A_j)_{j\in[m]}$ where $\mathrm{BL}(B,\ p;\ A)$ is defined in (2).*

This formulation leads to a geodesically convex optimization problem on $\bigtimes_{j\in[m]} \mathbb{S}^{n_j}_{++}$ that captures the Brascamp-Lieb constant.

Subsequently, we present a new formulation for the Brascamp-Lieb constant by combining Lieb's result with observations made by Bennett *et al.* [5] about maximizers of $\mathrm{BL}(B,\ p;\ A)$; see Theorem 5. [5] showed that if $A = (A_j)_{j\in[m]}$ is a maximizer to (2), then $A_j = (B_j M^{-1} B_j^\top)^{-1}$ for each $j \in [m]$, where

$$M := \sum_{j\in[m]} p_j B_j^\top A_j B_j.$$

Thus, we can write each $A_j$ as a function of $M$ and obtain, $2\log(\mathrm{BL}(B, p; A(M)))$ equals

$$\sum_{j\in[m]} p_j \log\det((B_j M^{-1} B_j^\top)^{-1}) - \log\det\left(\sum_{j\in[m]} p_j B_j^\top (B_j M^{-1} B_j^\top)^{-1} B_j\right). \tag{3}$$

One can show that the expressions $\log\det((B_j M^{-1} B_j^\top)^{-1})$ for each $j \in [m]$ and $\log\det\left(\sum_{j\in[m]} p_j B_j^\top (B_j M^{-1} B_j^\top)^{-1} B_j\right)$ are geodesically concave functions of $M$ in the positive definite cone. However, the expression in (3) being a difference, is not geodesically concave with respect to $M$ in general. However, if $A$ is a global maximizer of $\mathrm{BL}(B, p; A)$, then we also have that

$$M = \sum_{j\in[m]} p_j B_j^\top (B_j M^{-1} B_j^\top)^{-1} B_j.$$

Combining these two observations, we obtain the following geodesically concave optimization problem for computing the Brascamp-Lieb constant.

▶ **Definition 7** (A Geodesically Log-Concave Formulation for Computation of the Brascsamp-Lieb Constant)**.** Let $(B, p)$ be a feasible Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j \times n}$ for each $j \in [m]$. Let $F_{B,\,p}(X) : \mathbb{S}^n_{++} \to \mathbb{R}$ be defined as follows,

$$F_{B,p}(X) := \log\det(X) - \sum_{j\in[m]} p_j \log\det(B_j X B_j^\top). \tag{4}$$

The following theorem establishes the geodesic concavity of $F_{B,p}$ and its equivalence to $\mathrm{BL}(B, p; A)$.

▶ **Theorem 8** (Properties of $F_{B,p}$)**.** *Let $(B, p)$ be a feasible Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j\times n}$ for each $j \in [m]$. The function $F_{B,p}(X)$ as defined in (4) has following properties;*

1. *$F_{B,p}$ is geodesically concave.*
2. *If $(B, p)$ is simple, then $\sup_{X\in\mathbb{S}^n_{++}} F_{B,p}(X)$ is attained. Moreover, if $X^\star$ is a maximizer of $F_{B,p}$, then $\exp(\frac{1}{2}F_{B,p}(X^\star)) = \mathrm{BL}(B, p)$ and $A^\star = ((B_j X^\star B_j^\top)^{-1})_{j\in[m]}$ maximizes $\mathrm{BL}(B, p;\, A^\star)$.*

Our results lead to a natural question: is there a polynomial time algorithm based on techniques from geodesic optimization to compute the Brascamp-Lieb constant. For the case when $n_j = 1$ for all $j$, or the rank-one case, a polynomial time algorithm to compute the Brascamp-Lieb constant is known; see [23, 25]. This algorithm relies on the observation that the dual of the problem to compute the Brascamp-Lieb constant is the problem of optimizing an entropy maximizing probability distribution on the vertices of the Brascamp-Lieb polytope where the marginals of the probability distribution should correspond to the given point $p$. This algorithm computes $\mathrm{BL}(B, p)$ up to multiplicative error $1 + \varepsilon$ in $\mathrm{poly}(m, \langle B, p\rangle, \log \varepsilon^{-1})$ where $\langle B, p\rangle$ denotes the bit complexity of $B, p$. The main technical result is to show that for any $p$ in the polytope, an $\varepsilon$-approximate solution to a function like $F_{B,p}$ can be found in a ball of radius that is polynomial in the bit-complexity of $B$ and $p$. To extend this rank-one result to a higher rank already seems non-trivial due to a couple of reasons. (a) Lack of a separation oracle to the Brascamp-Lieb polytope in general, and (b) lack of an interpretation of $F_{B,p}$ as an optimization problem over the Brascamp-Lieb polytope. The hope is that our formulation might lead to such an optimization interpretation of the Brascamp-Lieb constant over the Brascamp-Lieb polyhedron and, consequently, lead to polynomial time algorithms following the general approach of [25].

## 2 The Positive Definite Cone, its Riemannian Geometry, and Geodesic Convexity

**The Metric.** Consider the set of positive definite matrices $\mathbb{S}^d_{++}$ as a subset of $\mathbb{R}^{d\times d}$ with the inner product $\langle X, Y\rangle := \mathrm{Tr}(X^\top Y)$ for $X, Y \in \mathbb{R}^{d\times d}$. At any point $X \in \mathbb{S}^d_{++}$, the tangent space consists of all $d \times d$ real symmetric matrices. There is a natural metric $g$ on this set that gives it a Riemannian structure: For $X \in \mathbb{S}^d_{++}$ and two symmetric matrices $\nu, \xi$

$$g_X(\nu, \xi) := \mathrm{Tr}(X^{-1}\nu X^{-1}\xi). \tag{5}$$

It is an exercise in differentiation to check that this metric arises as the Hessian of the following function $\varphi : \mathbb{S}^d_{++} \to \mathbb{R}$:

$$\varphi(X) := -\log\det X.$$

Hence, $\mathbb{S}^d_{++}$ endowed with the metric $g$ is not only a Riemannian manifold but a Hessian manifold [21]. The study of this metric on $\mathbb{S}^d_{++}$ goes back at least to Siegel [22]; see also the book of Bhatia [6].

**Geodesics on $\mathbb{S}^d_{++}$.** If $X, Y \in \mathbb{S}^d_{++}$ and $\gamma : [0,1] \to \mathbb{S}^d_{++}$ is a smooth curve between $X$ and $Y$, then the arc-length of $\gamma$ is given by the (action) integral

$$L(\gamma) := \int_0^1 \sqrt{g_{\gamma(t)}\left(\frac{d\gamma(t)}{dt}, \frac{d\gamma(t)}{dt}\right)} dt. \tag{6}$$

The geodesic between $X$ and $Y$ is the unique smooth curve between $X$ and $Y$ with the smallest arc-length [26]. The following theorem asserts that between any two points in $\mathbb{S}^d_{++}$, there is a geodesic that connects them. In other words, $\mathbb{S}^d_{++}$ is a geodesically convex set. Moreover, there exists a closed form expression for the geodesic between two points, a formula that is useful for calculations.

▶ **Theorem 9** (Geodesics on $\mathbb{S}^d_{++}$ [6], Theorem 6.1.6)**.** *For $X, Y \in \mathbb{S}^d_{++}$, the exists a unique geodesic between $X$ and $Y$, and this geodesic is parametrized by the following equation:*

$$X \#_t Y := X^{1/2}(X^{-1/2} Y X^{-1/2})^t X^{1/2} \tag{7}$$

*for $t \in [0,1]$.*

**Geodesic Convexity.** One definition of convexity of a function $f$ in a Euclidean space is that the average of the function at the endpoints of each line in the domain is at least the value of the function at the average point on the line. Geodesic convexity is a natural extension of this notion of convexity from Euclidean spaces to Riemannian manifolds that are geodesically convex. A set in the manifold is said to be geodesically convex if, for every pair of points in the set, the geodesic combining these points lies entirely in the set.

▶ **Definition 10** (Geodesically Convex Sets)**.** A set $S \subseteq \mathbb{S}^d_{++}$ is called geodesically convex if for any $X, Y \in S$ and $t \in [0,1]$, $X \#_t Y \in S$.

A function defined on a geodesically convex set is said to be geodesically convex if the average of the function at the endpoints of any geodesic in the domain is at least the value of the function at the average point on the geodesic.

▶ **Definition 11** (Geodesically Convex Functions)**.** Let $S \subseteq \mathbb{S}^d_{++}$ be a geodesically convex set. A function $f : S \to \mathbb{R}$ is called geodesically convex if for any $X, Y \in \mathbb{S}^d_{++}$ and $t \in [0,1]$,

$$f(X \#_t Y) \leq (1-t) f(X) + t f(Y). \tag{8}$$

$f$ is called geodesically concave if $-f$ is geodesically convex.

An important point regarding geodesic convexity is that a non-convex function might be geodesically convex or vice-versa. In general, one cannot convert a geodesically convex function to a convex function by a change of variables. A well-known example for this is the $\log\det(X)$ function whose concavity a classical result from the matrix calculus. On the other hand, a folklore result is that $\log\det(X)$ is both geodesically convex and geodesically concave on the space of positive definite matrices with the metric (5).

▶ **Proposition 12** (Geodesic Linearity of $\log\det$)**.** *The $\log\det(X)$ function is geodesically linear, i.e, it is both geodesically convex and geodesically concave over $\mathbb{S}^n_{++}$.*

**Proof.** Let $X, Y \in \mathbb{S}^n_{++}$ and $t \in [0,1]$. Then,

$$\log\det(X\#_t Y) \overset{\text{Theorem 9}}{=} \log\det(X^{1/2}(X^{-1/2}YX^{-1/2})^t X^{1/2}) \\ =(1-t)\log\det(X) + t\log\det(Y).$$

Therefore, $\log\det(X)$ is a geodesically linear function over the positive definite cone with respect to the metric in (5). ◀

Henceforth, when we mention geodesic convexity, it is with respect to the metric in (5). Geodesically convex functions share some properties with usual convex functions. One such property is the relation between local and global minimizers.

▶ **Theorem 13** (Minimizers of Geodesically Convex Functions [20], Theorem 6.1.1)**.** *Let $S \subseteq \mathbb{S}^d_{++}$ be a geodesically convex set and $f : S \to \mathbb{R}$ be a geodesically convex function. Then, any local minimum point of $f$ is also a global minimum of $f$. More precisely, if $x^\star := \arg\min_{x\in O} f(x)$ for some open geodesically convex subset $O$ of $S$, then $f(x^\star) = \inf_{x\in S} f(x)$.*

**Geometric Mean of Matrices and Linear Maps.** While the function $\log\det(P)$ is geodesically linear, our proof of Theorem 6 relies on the geodesic convexity of $\log\det(\sum_{j\in[m]} p_j B_j^\top A_j B_j)$. A simple but important observation is that, if $(B, p)$ is feasible, then $p_j B_j^\top A_j B_j$ is a strictly positive linear map for each $j$ as proved below.

▶ **Lemma 14** (Strictly Linear Maps Induced by Feasible Brascamp-Lieb Datums)**.** *Let $(B,\, p)$ be a Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j\times n}$ for each $j \in [m]$ such that $\sum_{j\in[m]} p_j n_j = n$ and $\dim(\mathbb{R}^n) \le \sum_{j\in[m]} p_j \dim(B_j\mathbb{R}^n)$. Then, $\Phi_j(X) := B_j^\top X B_j$ is a strictly positive linear map for each $j \in [m]$.*

Theorem 3 shows that any feasible Brascamp-Lieb datum satisfies both conditions. Furthermore non-feasible Brascamp-Lieb data can also satisfy these conditions as second condition on Theorem 3 enforced for only the $\mathbb{R}^n$.

**Proof.** Let us assume that for some $j_0 \in [m]$, $\Phi_{j_0}(X)$ is not a strictly positive linear map. Then, there exists $X_0 \in \mathbb{S}^n_{++}$ such that $\Phi_{j_0}(X_0)$ is not positive definite. Thus, there exists $v \in \mathbb{R}^{n_{j_0}}$ such that $v^\top\Phi_{j_0}(X_0)v \le 0$. Equivalently, $(B_{j_0}^\top v)^\top X_0(B_{j_0}^\top v) \le 0$. Since $X_0$ is positive definite, we get $B_{j_0}^\top v = 0$. Hence, $v^\top B_{j_0} B_{j_0}^\top v = 0$. Consequently, the rank of $B_{j_0}$ is at most $n_{j_0} - 1$ and $\dim(B_{j_0}\mathbb{R}^n) < n_{j_0}$. Therefore,

$$n = \dim(\mathbb{R}^n) \le \sum_{j\in[m]} p_j \dim(B_j\mathbb{R}^n) < \sum_{j\in[m]} p_j n_j = n,$$

by the hypothesis, a contradiction. Consequently, for any $j \in [m]$, $\Phi_j(X) := B_j^\top X B_j$ is strictly positive linear whenever $(B, p)$ satisfies conditions $\sum_{j\in[m]} p_j n_j = n$ and $\dim(\mathbb{R}^n) \le \sum_{j\in[m]} p_j \dim(B_j\mathbb{R}^n)$. ◀

The joint geodesic convexity of $\log\det(\sum_{j\in[m]} p_j B_j^\top A_j B_j)$ follows from a more general observation (that we prove) that asserts that if $\Phi_j$s are strictly positive linear maps from $\mathbb{S}^{n_j}_+$ to $\mathbb{S}^n_+$, then $\log\det(\sum_{j\in[m]} \Phi_j(A_j))$ is geodesically convex. Sra and Hosseini [24] observed this when $m = 1$. Their result also follows from a result of Ando [2] about "geometric means" that is also important for us and we explain it next.

The geometric mean of two matrices was introduced by Pusz and Woronowicz [19]. If $P, Q \in \mathbb{S}^d_{++}$, then the geometric mean of $P$ and $Q$ is defined as

$$P\#_{1/2}Q = P^{1/2}(P^{-1/2}QP^{-1/2})^{1/2}P^{1/2}. \tag{9}$$

By abuse of notation, we drop $1/2$ and denote geometric mean by $P\#Q$. Recall that, the geodesic convexity of a function $f : \mathbb{S}^d_{++} \to \mathbb{R}$ is equivalent to for any $P, Q \in \mathbb{S}^d_{++}$ and $t \in [0, 1]$,

$$f(P\#_t Q) \leq (1-t)f(P) + tf(Q).$$

If $f$ is continuous, then the geodesic convexity of $f$ can be deduced from the following:

$$\forall P, Q \in \mathbb{S}^d_{++}, \quad f(P\#Q) \leq 1/2 f(P) + 1/2 f(Q).$$

Ando proved the following result about the effect of a strictly positive linear map on the geometric mean of two matrices.

▶ **Theorem 15** (Effect of a Linear Map over Geometric Mean [2], Theorem 3). *Let $\Phi : \mathbb{S}^d_+ \to \mathbb{S}^{d'}_+$ be a strictly positive linear map. If $P, Q \in \mathbb{S}^d_{++}$, then*

$$\Phi(P\#Q) \preceq \Phi(P)\#\Phi(Q). \tag{10}$$

The monotonicity of logdet ($P \preceq Q$ implies $\log\det(P) \leq \log\det(Q)$) and the multiplicativity of the determinant, combined with Theorem 15, imply the following result.

▶ **Corollary 16** (Geodesic Convexity of the Logarithm of Linear Maps[24], Corollary 12). *If $\Phi : \mathbb{S}^d_+ \to \mathbb{S}^{d'}_+$ is a strictly positive linear map, then $\log\det(\Phi(P))$ is geodesically convex.*

While the proof of Theorem 8 uses Theorem 16, it is not enough for the proof of Theorem 6. Instead of geodesic convexity of $\log\det(\Phi(P))$, the joint geodesic convexity of $\log\det(\sum_{j\in[m]} \Phi_j(P_j))$ is needed where $\Phi_j : \mathbb{S}^{n_j}_+ \to \mathbb{S}^n_+$ is a linear map for each $j \in [m]$. We conclude this section with the following two results on a maximal characterization of the geometric mean and the effect of positive linear maps on positive definiteness of block diagonal matrices.

▶ **Theorem 17** (Maximal Characterization of the Geometric Mean, see e.g. [6], Theorem 4.1.1). *Let $P, Q \in \mathbb{S}^d_{++}$. The geometric mean of $P$ and $Q$ can be characterized as follows,*

$$P\#Q = \max\left\{ Y \in \mathbb{S}^d_{++} \;\middle|\; \begin{bmatrix} P & Y \\ Y & Q \end{bmatrix} \succeq 0 \right\},$$

*where the maximal element is with respect to Loewner partial order.*

▶ **Proposition 18** (Effect of Positive Linear Maps, see e.g. [6], Exercise 3.2.2). *Let $\Phi : \mathbb{S}^d_+ \to \mathbb{S}^{d'}_+$ be a strictly positive linear map and $P, Q, R \in \mathbb{S}^d_+$. If*

$$\begin{bmatrix} P & R \\ R & Q \end{bmatrix} \succeq 0, \text{ then } \begin{bmatrix} \Phi(P) & \Phi(R) \\ \Phi(R) & \Phi(Q) \end{bmatrix} \succeq 0.$$

## 3 Proof of Theorem 6

Let $(B, p)$ be a feasible Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j \times n}$ and $A := (A_j)_{j\in[m]}$ with $A_j \in \mathbb{S}^{n_j}_{++}$ be the input of $\mathrm{BL}(B, p; A)$ as defined in (2). To prove the joint geodesic convexity of $\mathrm{BL}(B, p; A)$ with respect to $A$ we extend Theorem 16 and Theorem 15 from linear maps to "jointly linear maps". We use the term jointly linear maps to refer to multivariable functions of the form $\sum_{j\in[m]} \Phi_j(P_j)$ where each $\Phi_j$ is a strictly positive linear map for each $j \in [m]$. In particular, the term $\sum_{j\in[m]} p_j B_j^\top A_j B_j$ in (2) is a jointly linear map.

The extension of Theorem 15 is presented in Theorem 20 and its proof is based on the maximal characterization of geometric mean (Theorem 17) and the effect of positive linear maps on the positive definiteness of block matrices (Theorem 18). We follow the proof of Theorem 15 for each $\Phi_j$, but instead of concluding $\Phi_j(P_j \# Q_j) \preceq \Phi_j(P_j) \# \Phi_j(Q_j)$ from the maximality of geometric mean, we sum the resulting inequalities. Subsequently, Theorem 20 follows from the maximality of geometric mean. Lemma 21 is an extension of Theorem 16 and follows directly from Theorem 20.

▶ **Definition 19** (Jointly linear map). Let $\Phi : \mathbb{S}_+^{n_1} \times \cdots \times \mathbb{S}_+^{n_m} \to \mathbb{S}_+^n$. We say that $\Phi$ is a jointly linear map if there exist strictly positive linear maps $\Phi_j : \mathbb{S}_+^{n_j} \to \mathbb{S}_+^n$ such that

$$\Phi(P_1, \ldots, P_k) := \sum_{j \in [k]} \Phi_j(P_j). \tag{11}$$

Now, we state the extension of Theorem 15.

▶ **Theorem 20** (Effect of Jointly Linear Maps over Geometric Means). *Let $\Phi : \mathbb{S}_+^{n_1} \times \cdots \times \mathbb{S}_+^{n_m} \to \mathbb{S}_+^n$ be a jointly linear map. Then,*

$$\Phi(G) \preceq \Phi(P) \# \Phi(Q)$$

*where $P := (P_j)_{j \in [m]}$, $Q := (Q_j)_{j \in [m]}$, and $G := (G_j)_{j \in [m]}$ with $P_j, Q_j \in \mathbb{S}_{++}^{n_j}$ and $G_j := P_j \# Q_j$.*

The following is a corollary of the theorem above and a generalization of Theorem 16.

▶ **Corollary 21** (Joint Geodesic Convexity of Logarithm of Jointly Linear Maps). *If $\Phi : \mathbb{S}_+^{n_1} \times \cdots \times \mathbb{S}_+^{n_m} \to \mathbb{S}_+^n$ is a jointly linear map, then*

$$g(P_1, \ldots, P_m) := \log \det(\Phi(P_1, \ldots, P_m)) \tag{12}$$

*is jointly geodesically convex in $\mathbb{S}_{++}^{n_1} \times \cdots \times \mathbb{S}_{++}^{n_m}$.*

**Proof of Corollary 21.** We show that $g$ is jointly geodesically mid-point convex. Theorem 20 implies that

$$\Phi(G) \preceq \Phi(P) \# \Phi(Q) \tag{13}$$

for any $P := (P_j)_{j \in [m]}$ and $Q := (Q_j)_{j \in [m]}$ with $P_j, Q_j \in \mathbb{S}_{++}^{n_j}$. Therefore,

$$\begin{aligned} g(G) &\overset{(12)}{=} \log \det(\Phi(G)) \\ &\leq \log \det(\Phi(P) \# \Phi(Q)) && \text{(monotonicity of } \log\det \text{ and (13))} \\ &\leq {}^1\!/_2 \log \det(\Phi(P)) + {}^1\!/_2 \log \det(\Phi(Q)) && \text{(multiplicativity of det)} \\ &\overset{(12)}{=} {}^1\!/_2 \left( g(P) + g(Q) \right). \end{aligned}$$

Thus, $g$ satisfies mid-point geodesic convexity. Consequently, we establish the geodesic convexity of $g$ using the continuity of $g$. ◀

The proof of Theorem 6 is a simple application of Corollary 21 and Theorem 12.

**Proof of Theorem 6.** We show that $\mathrm{BL}(B, p; A)$ is jointly geodesically mid-point log-concave with respect to $A$. In other words, we show that for arbitrary $P = (P_j)_{j \in [m]}$, $Q = (Q_j)_{j \in [m]}$

$$-\log \mathrm{BL}(B, p; G) \leq -{}^1\!/_2 \left( \log \mathrm{BL}(B, p; P) + \log \mathrm{BL}(B, p; Q) \right)$$

where $G = (G_j)_{j\in[m]}$ with $G_j := P_j \# Q_j$, being the midpoint of geodesic combining $P_j$ to $Q_j$. This implies that $\mathrm{BL}(B, p; A)$ is jointly geodesically log-concave with respect to $A$ due to the continuity of $\mathrm{BL}(B, p; A)$ with respect to $A$.

Let $\Phi_j(P_j) := p_j B_j^\top P_j B_j$. $\Phi_j$ is strictly positive linear map by Lemma 14. Then, $\Phi(P) := \sum_{j\in[m]} p_j B_j^\top P_j B_j$ is jointly linear, as $\Phi(P) = \sum_{j\in[m]} \Phi_j(P_j)$. Hence, $\log\det(\Phi(P))$ is jointly geodesically convex by Corollary 21. Also, $\log\det(X)$ is geodesically linear (Theorem 12). Thus, for any $P := (P_j)_{j\in[m]}$, $Q := (Q_j)_{j\in[m]}$ and $G := (G_j)_{j\in[m]}$ with $G_j := P_j \# Q_j$ we have

$$
\begin{aligned}
-\log \mathrm{BL}(B,p;G) &\overset{(2)}{=} 1/2(\log\det(\Phi(G)) - \sum_{j\in[m]} p_j \log\det(G_j)) \\
&\overset{(12)}{\leq} 1/2(1/2(\log\det(\Phi(P)) + \log\det(\Phi(Q))) - \sum_{j\in[m]} p_j \log\det(G_j)) \\
&= 1/2(1/2(\log\det(\Phi(P)) - \sum_{j\in[m]} p_j \log\det(P_j)) \qquad \text{(Theorem 12)} \\
&\qquad + 1/2(\log\det(\Phi(Q)) - \sum_{j\in[m]} p_j \log\det(Q_j))) \\
&\overset{(2)}{=} -1/2(\log\det \mathrm{BL}(B,p;P) + \log\det \mathrm{BL}(B,p;Q)).
\end{aligned}
$$

This concludes the proof. ◀

Now we prove Theorem 20. This proof is based on the proof of Theorem 15 and depends on the maximality of geometric mean (Theorem 17) and effects of positive linear maps on block matrices (Theorem 18).

**Proof of Theorem 20.** $\Phi$ is a jointly linear map by the assumption. Thus, there exist linear maps $\Phi_j : \mathbb{S}_+^{n_j} \to \mathbb{S}_+^n$ such that $\Phi(P) = \sum_{j\in[m]} \Phi_j(P_j)$. Theorem 17 implies for each $j \in [m]$,

$$
0 \preceq \begin{bmatrix} P_j & G_j \\ G_j & Q_j \end{bmatrix}.
$$

Since $\Phi_j$'s are strictly positive linear maps, Theorem 18 implies that for each $j \in [m]$,

$$
0 \preceq \begin{bmatrix} \Phi_j(P_j) & \Phi_j(G_j) \\ \Phi_j(G_j) & \Phi_j(Q_j) \end{bmatrix}.
$$

The dimension of these block matrices is $2n \times 2n$ for each $j \in [m]$. Thus we can sum these inequalities and the summation leads to

$$
0 \preceq \sum_{j\in[m]} \begin{bmatrix} \Phi_j(P_j) & \Phi_j(G_j) \\ \Phi_j(G_j) & \Phi_j(Q_j) \end{bmatrix} = \begin{bmatrix} \sum_{j\in[m]} \Phi_j(P_j) & \sum_{j\in[m]} \Phi_j(G_j) \\ \sum_{j\in[m]} \Phi_j(G_j) & \sum_{j\in[m]} \Phi_j(Q_j) \end{bmatrix} \overset{(11)}{=} \begin{bmatrix} \Phi(P) & \Phi(G) \\ \Phi(G) & \Phi(Q) \end{bmatrix}. \tag{14}
$$

Theorem 17 and (14) imply that $\Phi(G) \preceq \Phi(P) \# \Phi(Q)$. ◀

## 4 Proof of Theorem 8

Let $(B, p)$ be a feasible Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j \times n}$. Let $A := (A_j)_{j\in[m]}$ with $A_j \in \mathbb{S}_{++}^{n_j}$ be the input of $\mathrm{BL}(B, p; A)$ as defined in (2). The proof of Theorem 8 first establishes the geodesic concavity of $F_{B,p}$ as defined in (4) when $(B, p)$ is feasible. Next, it establishes the relation between global maximizers of $F_{B,p}(X)$ and global maximizers of $\mathrm{BL}(B,p;A)$, as well as the relation between $\sup_{X\in\mathbb{S}^n_{++}} F_{B,p}(X)$ and $\mathrm{BL}(B,p)$ when $(B,p)$ is simple.

Feasibility of $(B,p)$ implies the linear maps $B_j X B_j^\top$ are strictly positive linear for each $j\in[m]$ ( Lemma 14). Consequently, $-\log\det(B_j X B_j^\top)$ is geodesically concave by Theorem 16 for each $j\in[m]$. Also, $\log\det(X)$ is geodesically concave by Theorem 12. Thus, $F_{B,p}(X)$ is geodesically concave as a sum of geodesically concave functions with non-negative coefficients.

The geodesic concavity of $F_{B,p}$ implies that any local maximum is also a global maximum (Theorem 13). Consequently, we investigate the points where all directional derivatives of $F_{B,p}$ vanish, the critical points of $F_{B,p}$. A simple calculation involving the first derivative shows that any critical point $X$ of $F_{B,p}$ should satisfy

$$X^{-1} = \sum_{j\in[m]} p_j B_j^\top (B_j X B_j^\top)^{-1} B_j.$$

Theorem 5 implies that we can construct a global maximizer of $\mathrm{BL}(B,p;A)$ from $X$ by setting $A_j := (B_j X B_j^\top)^{-1}$. Furthermore, we can construct a critical point of $F_{B,p}$ using a global maximizer of $\mathrm{BL}(B,p;A)$ by setting $X := (\sum_{j\in[m]} p_j B_j^\top A_j B_j)^{-1}$. Theorem 5 guarantees the existence of a global maximizer of $\mathrm{BL}(B,p;A)$ if $(B,p)$ is simple. Thus, if $(B,p)$ is simple, then $\sup_X F_{B,p}(X)$ should be attained. We can deduce

$$\sup_X F_{B,p}(X) = 2\log \mathrm{BL}(B,p)$$

from the construction of $F_{B,p}$ and the relation between maximizers of $F_{B,p}$ and $\mathrm{BL}(B,p;A)$.

The second part of the proof Theorem 8 depends on well-known identities from matrix calculus. We present these identities for the convenience of the reader and refer the interested readers to the matrix cookbook [18] for more details.

▶ **Proposition 22.** *Let $X(t), Y(t)$ be differentiable functions from $\mathbb{R}$ to $d\times d$ invertible symmetric matrices. Let $U\in\mathbb{R}^{d'\times d}$, $V\in R^{d\times d''}$, $W\in\mathbb{R}^{d\times d}$ be matrices which do not depend on $t$. Then, the following identities hold:*

$$\frac{d\log\det(X(t))}{dt} = \mathrm{Tr}\left(X(t)^{-1}\frac{dX(t)}{dt}\right) \tag{15}$$

$$\frac{dUX(t)V}{dt} = U\frac{dX(t)}{dt}V \tag{16}$$

$$\frac{dtW}{dt} = W. \tag{17}$$

**Proof of Theorem 8.** We start by showing that $F_{B,p}(X)$ is geodesically concave. The feasibility of $(B,p)$ implies $B_j X B_j^\top$ is a strictly positive linear map for each $j\in[m]$ ( Lemma 14). Thus, Theorem 16 yields that for any $X,Y\in\mathbb{S}^n_{++}$,

$$\log\det(B_j(X\#Y)B_j^\top) \le {}^1\!/_2\log\det(B_j X B_j^\top) + {}^1\!/_2\log\det(B_j Y B_j^\top). \tag{18}$$

Combining this with the geodesic linearity of $\log\det(X)$ (Theorem 12), we obtain

$$\begin{aligned}F_{B,p}(X\#Y) =& \log\det(X\#Y) - \sum_{j\in[m]} p_j \log\det(B_j(X\#Y)B_j^\top)\\ \ge& {}^1\!/_2(\log\det(X) - \sum_{j\in[m]} p_j\log\det(B_j X B_j^\top))\\ &+ {}^1\!/_2(\log\det(Y) - \sum_{j\in[m]} p_j\log\det(B_j Y B_j^\top))\\ =& {}^1\!/_2 F_{B,p}(X) + {}^1\!/_2 F_{B,p}(Y)\end{aligned}$$

Therefore, $F_{B,p}(X)$ is geodesically concave.

Now, we can show the second part of the theorem. The geodesic concavity of $F_{B,p}$ implies any local maximum of $F_{B,p}$ is a global maximum of $F_{B,p}$. A local maximum of $F_{B,p}$ is achieved at $X$ if it is a critical point of $F_{B,p}$. If $X$ is a critical point of $F_{B,p}$, then for any symmetric matrix $Q$, the directional derivative of $F_{B,p}$ at $X$ in the direction of $Q$ should be 0. In other words, if $\zeta(t) := X + tQ$ and $f(t) := F_{B,p}(\zeta(t))$, then $\frac{df}{dt}\Big|_{t=0}$ should be 0 for any $Q$. Let us compute $\frac{df}{dt}$,

$$\begin{aligned}\frac{df}{dt} &\overset{(4)}{=} \frac{d}{dt}\log\det(\zeta(t)) - \sum_{j\in[m]} p_j \frac{d}{dt}\log\det(B_j\zeta(t)B_j^\top)\\ &\overset{(15)}{=}\mathrm{Tr}\left(\zeta(t)^{-1}\frac{d\zeta(t)}{dt}\right) - \sum_{j\in[m]} p_j \mathrm{Tr}\left((B_j\zeta(t)B_j^\top)^{-1}\frac{dB_j\zeta(t)B_j^\top}{dt}\right)\\ &\overset{(16)}{=}\mathrm{Tr}\left(\zeta(t)^{-1}\frac{d\zeta(t)}{dt}\right) - \sum_{j\in[m]} p_j \mathrm{Tr}\left((B_j\zeta(t)B_j^\top)^{-1}B_j\frac{d\zeta(t)}{dt}B_j^\top\right)\\ &\overset{(17)}{=}\mathrm{Tr}(\zeta(t)^{-1}Q) - \sum_{j\in[m]} p_j \mathrm{Tr}((B_j\zeta(t)B_j^\top)^{-1}B_jQB_j^\top).\end{aligned}$$

Hence, the directional derivative of $F_{B,p}(X)$ in the direction of $Q$, $\frac{df}{dt}\Big|_{t=0}$ is

$$\mathrm{Tr}(X^{-1}Q) - \sum_{j\in[m]} p_j \mathrm{Tr}((B_jXB_j^\top)^{-1}B_jQB_j^\top). \tag{19}$$

If directional derivates of $F_{B,p}$ vanish at $X$, then (19) should be 0 for any symmetric matrix $Q$. Consequently,

$$\mathrm{Tr}(Q[X^{-1} - \sum_{j\in[m]} p_jB_j^\top(B_jXB_j^\top)^{-1}B_j]) = 0.$$

This observation leads to

$$X^{-1} = \sum_{j\in[m]} p_jB_j^\top(B_jXB_j^\top)^{-1}B_j. \tag{20}$$

If $(B,p)$ is simple, then there exists an input $A^\star = (A_j)_{j\in[m]}$ such that $A_j^{-1} = B_jM^{-1}B_j^\top$ where $M = \sum_{j\in[m]} p_jB_j^\top A_jB_j$ by Theorem 5. Consequently, $M$ satisfies

$$M = \sum_{j\in[m]} p_jB_j^\top(B_jM^{-1}B_j^\top)^{-1}B_j. \tag{21}$$

If $X^\star := M^{-1}$, then $X^\star$ satisfies (20) due to (21). Thus, $X^\star$ is a critical point of $F_{B,p}(X)$ and $F_{B,p}$ attains its maximal value at $X^\star$. Furthermore, the maximizer $A^\star$ of $\mathrm{BL}(B,p;A)$ is equal to $((B_jX^\star B_j^\top)^{-1})_{j\in[m]}$. Finally,

$$\begin{aligned}F_{B,p}(X^\star) &\overset{(4)}{=} \log\det(X^\star) - \sum_{j\in[m]} p_j\log\det(B_jX^\star B_j^\top)\\ &= \log\det((\sum_{j\in[m]} p_jB_j^\top(B_jX^\star B_j^\top)^{-1}B_j)^{-1}) - \sum_{j\in[m]} p_j\log\det(B_jX^\star B_j^\top)\\ &= \log\det((\sum_{j\in[m]} p_jB_j^\top A_jB_j)^{-1}) - \sum_{j\in[m]} p_j\log\det(A_j^{-1})\\ &= \sum_{j\in[m]} p_j\log\det(A_j) - \log\det(\sum_{j\in[m]} p_jB_j^\top A_jB_j).\end{aligned}$$

Therefore, $\mathrm{BL}(B,p;A^\star) = \exp(\frac{1}{2}F_{B,p}(X^\star))$. ◀

## References


1 Zeyuan Allen-Zhu, Garg Ankit, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing, Los Angeles, CA, USA, June 25-29*, 2018. To appear. arXiv:arXiv:1804.01076.

2 Tsuyoshi Ando. Concavity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra and its Applications*, 26:203–241, 1979. doi:10.1016/0024-3795(79)90179-4.

3 Keith Ball. Volumes of sections of cubes and related problems. In *Geometric Aspects of Functional Analysis*, pages 251–260. Springer, 1989.

4 Franck Barthe. On a reverse form of the Brascamp-Lieb inequality. *Inventiones mathematicae*, 134(2):335–361, 1998.

5 Jonathan Bennett, Anthony Carbery, Michael Christ, and Terence Tao. The Brascamp-Lieb inequalities: finiteness, structure and extremals. *Geometric and Functional Analysis*, 17(5):1343–1415, 2008.

6 Rajendra Bhatia. *Positive definite matrices*. Princeton University Press, 2009.

7 Herm Jan Brascamp and Elliott H Lieb. Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Advances in Mathematics*, 20(2):151–173, 1976.

8 Eric A Carlen and Dario Cordero-Erausquin. Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities. *Geometric and Functional Analysis*, 19(2):373–405, 2009.

9 Eric A. Carlen, Elliott. H. Lieb, and Michael Loss. A sharp analog of Young's inequality on SN and related entropy inequalities. *The Journal of Geometric Analysis*, 14(3):487–520, Sep 2004. doi:10.1007/BF02922101.

10 Zeev Dvir, Ankit Garg, Rafael Mendes de Oliveira, and József Solymosi. Rank bounds for design matrices with block entries and geometric applications. *Discrete Analysis*, 2018:5, Mar 2018. doi:10.19086/da.3118.

11 Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Breaking the quadratic barrier for 3-LCC's over the reals. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing, Los Angeles, CA, USA, May 31- June 3, 2014*, pages 784–793, New York, NY, USA, 2014. doi:10.1145/2591796.2591818.

12 Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. Algorithmic and optimization aspects of Brascamp-Lieb inequalities, via operator scaling. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing, Montreal, QC, Canada, June 19-23, 2017*, pages 397–409, 2017. doi:10.1145/3055399.3055458.

13 Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004. Special Issue on STOC 2003. doi:10.1016/j.jcss.2004.06.003.

14 Moritz Hardt and Ankur Moitra. Algorithms and hardness for robust subspace recovery. In *Conference on Learning Theory*, pages 354–375, 2013.

15 Elliott H Lieb. Gaussian kernels have only Gaussian maximizers. *Inventiones Mathematicae*, 102(1):179–208, 1990.

16 Jingbo Liu, Thomas A. Courtade, Paul W. Cuff, and Sergio Verdú. Smoothing Brascamp-Lieb inequalities and strong converses for common randomness generation. In *IEEE International Symposium on Information Theory, Barcelona, Spain, July 10-15, 2016*, pages 1043–1047. IEEE, 2016. doi:10.1109/ISIT.2016.7541458.

17 Jingbo Liu, Thomas A. Courtade, Paul W. Cuff, and Sergio Verdú. Information-theoretic perspectives on Brascamp-Lieb inequality and its reverse. *CoRR*, abs/1702.06260, 2017. arXiv:1702.06260.

18 Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.

19 W. Pusz and S.L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. *Reports on Mathematical Physics*, 8(2):159–170, 1975. doi:10.1016/0034-4877(75)90061-0.

20 Tamás Rapcsák. *Geodesic convex functions*, pages 61–86. Springer US, Boston, MA, 1997. doi:10.1007/978-1-4615-6357-0_6.

21 Hirohiko Shima. *The geometry of Hessian structures*. World Scientific Publishing, 2007.

22 Carl Ludwig Siegel. Symplectic geometry. *American Journal of Mathematics*, 65(1):1–86, 1943. URL: http://www.jstor.org/stable/2371774.

23 Mohit Singh and Nisheeth K Vishnoi. Entropy, optimization and counting. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 50–59. ACM, 2014.

24 Suvrit Sra and Reshad Hosseini. Conic geometric optimization on the manifold of positive definite matrices. *SIAM Journal on Optimization*, 25(1):713–739, 2015.

25 Damian Straszak and Nisheeth K. Vishnoi. Computing maximum entropy distributions everywhere. *ArXiv e-prints*, 2017. arXiv:1711.02036.

26 Nisheeth K Vishnoi. Geodesic convex optimization: Differentiation on manifolds, geodesics, and convexity, Jun 2018. URL: https://nisheethvishnoi.files.wordpress.com/2018/06/geodesicconvexity.pdf.


# A An Exponential-Sized Geodesically Convex Formulation from Operator Scaling

In this section, we describe the operator scaling problem and the reduction of Garg *et al.* [12] from the computation of the Brascasmp-Lieb constant to the computation of the "capacity" of a positive operator.

**The Operator Scaling Problem and its Geodesic Convexity.** In the operator scaling problem [13], one is given a linear operator $T(X) := \sum_{j\in[m]} T_j^\top X T_j$ through the tuple of matrices $T_j$s and the goal is to find square matrices $L$ and $R$ such that

$$\sum_{j\in[m]} \hat{T}_j^\top \hat{T}_j = I \qquad \text{and} \qquad \sum_{j\in[m]} \hat{T}_j \hat{T}_j^\top = I, \tag{22}$$

where $\hat{T}_j := LT_jR$. The matrices $L$ and $R$ can be computed by solving the following optimization problem.

▶ **Definition 23** (Operator Capacity). Let $T : \mathbb{S}^d_{++} \to \mathbb{S}^{d'}_{++}$ be a linear operator, then the capacity of $T$ is

$$\operatorname{cap}(T) := \inf_{\det(X)=1} \det\left(\frac{d'}{d} T(X)\right). \tag{23}$$

In particular, if $X_T^\star$ is a minimizer of (23) and $Y_T^\star = T(X_T^\star)^{-1}$, then (22) holds if we let $L := (Y_T^\star)^{1/2}$ and $R := (X_T^\star)^{1/2}$; see [13] for details. Operator capacity is known to be geodesically log-convex, see e.g. [1], Lemma C.1.

**The Reduction.** Let $(B, p)$ be a Brascamp-Lieb datum with $B_j \in \mathbb{R}^{n_j \times n}$ for each $j \in [m]$. Garg *et al.* [12] proved that if the exponent $p = (p_j)_{j \in [m]}$ is a rational vector, then one can construct an operator scaling problem from $(B, p)$. Let $p_j = {}^{c_j}/_{c}$ where $c_j$s are non-negative integers and $c$ is a positive integer, the common denominator for all the $p_j$s. Their reduction, outlined below, results in an operator $T_{B,p} : \mathbb{S}^{nc}_{++} \to \mathbb{S}^{n}_{++}$ with the property that $\mathrm{cap}(T_{B,p}) = {}^{1}/_{\mathrm{BL}(B,p)^2}$.

The operator $T_{B,p}$ is constructed with $c_j$ copies of the matrix $B_j$ for each $j \in [m]$. In order to easily refer these copies, let us define $m' := \sum_{j \in [m]} c_j$, and the function $\delta : [m'] \to [m]$. $\delta(i)$ is defined as the integer $j$ such that,

$$\sum_{k<j} c_k < i \leq \sum_{k \leq j} c_k.$$

Let $Z_{ij}$ be an $n_{\delta(i)} \times n$ matrix all of whose entries are zero when $\delta(i) \neq j$ and $B_{\delta(i)}$ if $\delta(i) = j$, for $i, j \in [m']$. Define $nc \times n$ matrices $T_j$ for $j \in [m]$ as follows:

$$T_j := \begin{bmatrix} Z_{1j} \\ \vdots \\ Z_{m'j} \end{bmatrix},$$

and define the linear operator $T_{B,p} : \mathbb{S}^{nc}_{++} \to \mathbb{S}^{n}_{++}$ as

$$T_{B,p}(X) := \sum_{j \in [m']} T_j^\top X T_j. \tag{24}$$

▶ **Theorem 24** (Reduction from Brascamp-Lieb to Operator Scaling[12], Lemma 4.4.). *Let $(B,\ p)$ be a Brascamp-Lieb a datum with $B_j \in \mathbb{R}^{n_j \times n}$ and $p_j = c_j/c$ where $c, c_j \in \mathbb{Z}_+$ for each $j \in [m]$. The capacity of the operator $T_{B,p}$ defined in (24) satisfies $\mathrm{cap}(T_{B,p}) = {}^{1}/_{\mathrm{BL}(B,p)^2}$.*

While this reduction gives a geodesically log-concave formulation to compute the Brascamp-Lieb constant, the dimension of the optimization problem is exponentially large in the bit complexity of the Brascamp-Lieb datum. Consequently, a truly polynomial time algorithm for the computation of the Brascamp-Lieb constant does not follow from any black-box optimization method for geodesically convex functions or polynomial time algorithms for operator capacity; e.g. the algorithm presented in [1].