1 Capacity of Rank Decreasing Operators

After Gurvits [5]

A summary written by Jacob Denson

Abstract

We describe the theory of rank-decreasing doubly stochastic positive operators, the capacity of such operators, Gurvit's algorithm to rescale an operator so that it is approximated by a doubly stochastic operator, and a brief idea of the connection of these techniques to the theory of Brascamp-Lieb inequalities.

Recall a theorem of Lieb [7], which shows that any inequality of the form

$$\int_{\mathbf{R}^n} \prod_{i=1}^m |f_i(B_i x)|^{p_i}\ dx \leq \mathrm{BL}(B,p) \cdot \prod_{i=1}^m \|f_i\|_{L^1(\mathbf{R}^{n_i})}^{p_i},$$

has an optimal Brascamp-Lieb constant $\mathrm{BL}(B,p)$ satisfying

$$\mathrm{BL}(B,p)^2 = \sup_{X_1,\dots,X_m \succ 0} \frac{\det(\sum_{i=1}^m p_i B_i^* X_i B_i)}{\prod_{i=1}^m \det(X_i)^{p_i}}. \tag{1}$$

Here and in what follows, for a square matrix A, we write $A \succ 0$ and $A \succeq 0$ to mean A is positive definite or semidefinite. Bennett, Carbery, Christ, and Tao [1] found that $\mathrm{BL}(B,p)$ is finite if and only if $\sum_{j=1}^m p_j n_j = 0$, and

$$\dim(V) \leq \sum_{j=1}^m p_j \dim(B_j V) \qquad \text{for all subspaces } V \subset \mathbf{R}^n. \tag{2}$$

Thus the finiteness of (1) acts as a guarantee for the mapping properties of the matrices $B_1, \dots, B_m$ given in (2), and vice versa.

The focus of these notes is to introduce analysts to a setting, introduced by computing scientists studying combinatorial optimization and quantum information theory, which can provide insight into the computation of the Brascamp-Lieb constant. Using these insights, Garg, Gurvits, and Wigderson [4] formulated polynomial time algorithms for approximating the Brascamp-Lieb constant, a fact that has important theoretical consequences to the general theory of Brascamp-Lieb inequalities independent of practical application to computation of particular Brascamp-Lieb constants. This connection is more fully explored in the subsequent presentation. In this summary, we introduce the original setting studied, only noting some similarities to Brascamp Lieb as we proceed.

2 Operator Scaling

Let $M(N)$ denote the space of all $N \times N$ complex-valued matrices. The main object of study in the setting we aim to introduce are *positive operators* $T : M(N) \to M(N)$, i.e. linear transformations between spaces of matrices such that if $X \succeq 0$, then $T(X) \succeq 0$. There is a rich theory of such maps, connected to representation theory and the theory of free probability (see [2] for a more thorough introduction). A useful example to keep in mind, given any matrices $B_1, \ldots, B_m$ and $p_1, \ldots, p_m > 0$, is the operator

$$T(X) = \sum_{i=1}^{m} p_i B_i^* X B_i \tag{3}$$

With any positive operator T, we associate a quantity $\mathrm{Cap}(T) \geq 0$, known as the *capacity* of T, defined by setting

$$\mathrm{Cap}(T) = \inf_{X \succ 0} \frac{\det(TX)}{\det(X)}. \tag{4}$$

The connection between capacity and Brascamp-Lieb may be hinted at by comparing (4) to (1) when T is of the form given in (3). Being defined by a non-convex optimization, it is difficult to explicitly compute the capacity of a general positive operator. However, the techniques of [5] show that for each $\varepsilon > 0$, given a positive operator T, represented in b bits on a computer, there is an algorithm to compute a value $\mathrm{Cap}_{\mathrm{approx}}(T)$ in $\lesssim (b \log(1/\varepsilon))^{O(1)})$ steps, such that

$$\mathrm{Cap}(T) \leq \mathrm{Cap}_{\mathrm{approx}}(T) \leq (1 + \varepsilon) \cdot \mathrm{Cap}(T).$$

We obtain this approximation by applying the technique of *operator scaling*, which we describe in this section.

Given any nonsingular matrices $A, B \in M(N)$, and any positive operator $T : M(N) \to M(N)$, we define a *scaled operator* $T_{A,B}(X) = B \cdot T(AXA^*) \cdot B^*$. For any $X \succeq 0$, if we write $Y = AXA^*$, then

$$\frac{\det(T_{A,B}X)}{\det(X)} = \frac{\det(B) \cdot T(Y) \cdot \det(B^*)}{\det(X)} = \det(A)^2 \det(B)^2 \frac{T(Y)}{\det(Y)}.$$

Taking infima over both sides of the equation thus gives

$$\mathrm{Cap}(T_{A,B}) = \det(A)^2 \det(B^2) \mathrm{Cap}(T).$$

Thus computing $\mathrm{Cap}(T_{A,B})$ immediately gives $\mathrm{Cap}(T)$. The idea of operator scaling is to rescale an operator to something whose capacity we can easily compute, which motivates the introduction of *doubly stochastic operators*.

A positive operator T is *doubly stochastic* if $T(I) = T^*(I) = I$, where T^* is the adjoint of T with respect to the inner product $(X, Y) \mapsto \mathrm{Tr}(XY^*)$. What interests us about these operators is that $\mathrm{Cap}(T) = 1$ for any doubly stochastic operator T. The proof is somewhat technical from our perspective, reducing the question to the study of doubly stochastic matrices. A proof can be found in Theorem 4.32 of [8], with relevant background about doubly stochastic matrices found in Section 8.5 of [9].

3 Capacity of Rank-Decreasing Operators

A useful property of a positive operator T is that it is *rank non-decreasing*, i.e. for any $X \succeq 0$,

$$\mathrm{Rank}(TX) \geq \mathrm{Rank}(X). \tag{5}$$

Condition (5) seems somewhat similar to (2). And indeed, in [5] it is shown that one has an equivalence between the non-vanishing of (4) and (5), analogous to the equivalence between the finiteness of (1) and (2).

Theorem 1. *T is rank non-decreasing if and only if $\mathrm{Cap}(T) > 0$.*

Proof. A simple family of positive operators are those of the form

$$TX = X_{11}A_1 + \cdots + X_{NN}A_N, \tag{6}$$

where $A_1, \ldots, A_N \succeq 0$. For such an operator, we can write

$$\mathrm{Cap}(T) = \inf_{\gamma_1,\ldots,\gamma_N>0} \frac{\det(\sum_{j=1}^N \gamma_j A_j)}{\gamma_1 \cdots \gamma_N}.$$

Results from a previous paper of Gurvits and Samorodnitsky [6] imply Theorem 1 in the special case of an operator defined by (6). Assuming this result, we indicate how this implies the general case.

For each orthonormal basis $U = \{u_1, \ldots, u_N\}$, we define the decoherence operator $D_U(X) = \sum \langle Xu_i, u_i \rangle \cdot u_i u_i^*$, and then consider the operator

$$T_U(X) = (T \circ D_U)(X) = \sum_{i=1}^N \langle Xu_i, u_i \rangle \cdot T(u_i u_i^*).$$

This operator is, up to a change of basis in $M(N)$, described in the form (6). Thus T_U is rank non-decreasing if and only if $\mathrm{Cap}(T_U) > 0$. The theorem then follows from the following two properties of this construction:

1. T is rank non-decreasing if and only if T_U is as well, for all bases U.

2. $\mathrm{Cap}(T) = \inf_U \mathrm{Cap}(T_U)$.

If $\mathrm{Cap}(T) > 0$, then Property 2 implies $\mathrm{Cap}(T_U) > 0$ for all U, so T_U is rank non-decreasing for all U, and thus Property 1 implies T is rank non-decreasing. The converse is similar.

The proof of Properties 1 and 2 both rely on a simple trick. We will prove Property 1 here. Given any $X \succeq 0$, we can find an orthonormal basis U diagonalizing X, and then for such U we have $T(X) = T_U(X)$. This immediately implies T is rank non-decreasing if T_U is rank non-decreasing for all U. The converse follows because the composition of rank non-decreasing operators is rank non-decreasing, and D_U is rank non-decreasing because it is *doubly stochastic*, a family of operators we will study shortly. □

4 Operator Scaling

Being defined by a non-convex optimization procedure, it is difficult to explicitly compute the capacity of a positive operator. However, the techniques described in this section show that for each $\varepsilon > 0$, given a positive operator T (with rational entries described using b bits), there is an algorithm to compute a value $\mathrm{Cap}_{\mathrm{approx}}(T) \in [0, \infty)$ in $\lesssim (b \log(1/\varepsilon))^{O(1)}$ steps, such that

$$\mathrm{Cap}(T) \leq \mathrm{Cap}_{\mathrm{approx}}(T) \leq (1+\varepsilon) \cdot \mathrm{Cap}(T).$$

We obtain this approximation by the technique of *operator scaling*, which we describe in this section.

Given any nonsingular matrices $A, B \in M(N)$, and any positive operator T, we define a *scaled operator* $T_{AB}(X) = B \cdot T(AXA^*) \cdot B^*$. For any $X \succeq 0$ with $\det(X) = 1$, if $Y = AXA^*$, then

$$\det(T_{AB}X) = \det(B)^2 \det(T(AXA^*)) = \det(B)^2 \det(T(Y)).$$

Since $\det(Y) = \det(A)^2$, taking infima over both sides of the equation gives

$$\mathrm{Cap}(T_{AB}) = \det(A)^2 \det(B)^2 \cdot \mathrm{Cap}(T).$$

Thus operator scaling predictably changes the capacity. The idea is now to scale an operator to another operator whose capacity is more simple to compute. Our goal is to rescale our operators to a *doubly stochastic* operator. A positive operator T is called *doubly stochastic* if $T(I) = T^*(I) = I$, where T^* is the adjoint of T with respect to the inner product $(X, Y) \mapsto \text{Tr}(XY^*)$. One property of these operators is that $\text{Tr}(T(X)) = \text{Tr}(X)$ for any input X.

Lemma 2. *If T is doubly stochastic, then $Cap(T) = 1$.*

Proof. Since $T(I) = I$, it is clear that $\text{Cap}(T) \leq 1$. The converse is less trivial, reducing the question to the study of doubly stochastic matrices. The proof can be found in Theorem 4.32 of [8], with relevant background about doubly stochastic matrices found in Section 8.5 of [9]. $\square$

Thus, given an operator T, if we can find nonsingular $A, B \in M(N)$ such that $T_{A,B}$ is doubly stochastic, then it follows from Theorem 2 that $\text{Cap}(T) = \det(A)^2 \det(B)^2$. This is not quite possible to do for all T with $\text{Cap}(T) > 0$. In fact, it is only possible if the infinum defining $\text{Cap}(T)$ is attained. Nonetheless, if $\text{Cap}(T) > 0$ then we can find doubly stochastic operators approximating T arbitrarily closely. The capacity is a continuous function of the input, so this is good enough to approximate $\text{Cap}(T)$ rather than calculating $\text{Cap}(T)$ exactly.

Our approximation algorithm is quite simple, an application of a similar method first used by Sinkhorn to reduce a matrix to something resembling a doubly stochastic process. Given a positive operator T it is easy to scale the operator to an operator S with $S(I) = I$. We simply consider $T_{I,T(I)^{-1/2}}$. Similarily, we can scale T to an operator S with $S^*(I) = I$ by taking the scaling $T_{T*(I)^{-1/2},I}$. The challenge is to obtain scalings for which both properties are true. Sinkhorn's trick is to iteratively rescale our operators , obtaining a sequence of matrices $T_0, T_1, T_2, \dots$ with $T_i(I) = I$ for odd i, and $T_i^*(I) = I$ for even i. If this sequence converges, continuity of the equations imply the limit will be a doubly stochastic operator as desired.

Unfortunately, this sequence does not necessarily converge for any positive operator T (it converges if and only if the infinum defining $\text{Cap}(T)$ is attained by a particular input). But provided that $\text{Cap}(T) > 0$, the distance between the elements of the sequence to the set of all stochastic operators *will* converge to zero, which is sufficient for our purposes. To analyze the convergence, we rely on the capacity as a *potential* for the analysis of the algorithm (this was the main reason Gurvit's introduced the capacity in [5]). To determine how close we are to a doubly stochastic matrix, we use the measure $\mathrm{DS}(T) = \|T(I) - I\|^2 + \|T^*(I) - I\|^2$. The following properties then hold:

1. If $T(I) = I$ or $T^*(I) = I$, then $1 \lesssim_N \mathrm{Cap}(T) \leq 1$. Moreover, for $\mathrm{Cap}(T) \geq 1/2$,
$$\mathrm{DS}(T_n) \leq 6 \log(1/\mathrm{Cap}(T_n)),$$
so the closer the capacity is to one, the closer a matrix is to being doubly stochastic.

2. $\mathrm{Cap}(T_n)$ is increasing in $n > 0$. Indeed, $\mathrm{Tr}(T_n(I)) = \mathrm{Tr}(T_n^*(I)) = N$, and the arithmetic-geometric mean inequality (applied to the sum and product of the eigenvalues of a matrix) imply that
$$\mathrm{Det}(T_n(I)), \mathrm{Det}(T_n^*(I)) \leq 1,$$
which implies the capacity is increasing. In fact, we have a more precise bound, namely
$$\mathrm{Cap}(T_{n+1}) \geq e^{\min(1, \mathrm{DS}(T_n))/6} \cdot \mathrm{Cap}(T_n).$$

3. If $T(I) = I$ or $T^*(I) = I$ and $\mathrm{DS}(T) < 1/(N+1)$, then $\mathrm{Cap}(T) > 0$.

Property 2 implies that the capacities of the operators in the sequence increase faster the further away the operators are from being stochastic. Property 1 shows that the distance from being stochastic grows smaller the closer $\mathrm{Cap}(T_n)$ is from being equal to one. Thus regardless of our input operator $T = T_0$, provided that $\mathrm{Cap}(T) > 0$, the algorithm will eventually scale T to be as close as possible to a doubly stochastic operator, providing us with a way to approximate $\mathrm{Cap}(T)$. On the other hand, if $\mathrm{Cap}(T) = 0$, then Property 3 implies that we will be able to check this by observing whether $\mathrm{DS}(T_n) > 1/(N+1)$ after a certain number of iterations of the algorithm. Thus we have a reliable way to approximate the capacity of a matrix, which can be extended to give us an efficient way to approximate Brascamp-Lieb constants.

References

[1] J.M. Bennett, A. Carbery, M. Christ and T. Tao, *The Brascamp-Lieb Inequalities: Finiteness, Structure and Extremals*, Geom. Funct. Anal. **17** (2008), 1343-1415.

[2] Rajendra Bhatia, *Positive Definite Matrices*, Princeton University Press, 2007.

[3] Ankit Garg, Leonid Gurvits, Rafael Oliveira, Avi Wigderson, *Operator Scaling: Theory and Applications*, Foundations of Computational Mathematics, **20** (2020), 223–290.

[4] Ankit Garg, Leonid Gurvits, Avi Wigderson, *Algorithmic and Optimization Aspects of Brascamp-Lieb Inequalities, via Operator Scaling*, Geom. Funct. Anal, **28** (2018), 100–145.

[5] Leonid Gurvits, *Classical Complexity and Quantum Entanglement*, J. Comput. Syst. Sci. **69** (2004), 448–484.

[6] L. Gurvits, A. Samorodnitsky, *a Deterministic Algorithm Approximating the Mixed Discriminant and Mixed Volume, and a Combinatorial Corollary*, Discrete Comput. Geom. **27** (2002), 531–550.

[7] E.H. Lieb, *Gaussian Kernels have only Gaussian Maximizers* Invent. Math. **102** (1990), 179–208.

[8] John Watrous, *the Theory of Quantum Information* Cambridge University Press, 2018.

[9] Denis Serre, *Matrices*, Springer, 2010.

JACOB DENSON, UW MADISON
email: `jcdenson@wisc.edu`