1 Capacity of Rank Decreasing Operators

After Gurvits [5]

A summary written by Jacob Denson

Abstract

We describe the theory of the capacity of rank-decreasing positive operators, and Gurvit's operator rescaling algorithm to compute the capacity of such operators, with a brief discussion of the connection between this theory and the computation of Brascamp-Lieb constants.

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

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

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

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

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.

The focus of these notes is to introduce analysts to a setting, originally studied by computing scientists to understand problems in combinatorial optimization and quantum information theory, which can provide insight into the computation of the Brascamp-Lieb constant. Utilizing this theory, Garg, Gurvits, and Wigderson [4] formulated a polynomial time algorithm 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 following talk. 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 these notes 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 these maps, connected to representation theory and free probability theory. For more background, the textbook [2] provides a thorough introduction to this theory. 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

$$T(X) = \sum_{i=1}^{m} p_i (B_i^* X B_i). \tag{2}$$

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

$$\mathrm{Cap}(T) = \inf_{X \succ 0} \frac{\det(TX)}{\det(X)}. \tag{3}$$

The connection between capacity and Brascamp-Lieb may be hinted at by comparing (3) to (1) when T is of the form given in (2). Being defined by a non-convex optimization, it seems difficult to explicitly compute capacity. In this talk, we discuss a result of Gurvits [5], who found an efficient algorithm to compute, for each $\varepsilon > 0$, a value $\mathrm{Cap}_{\mathrm{approx}}(T)$ 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 invertible matrices $A, B \in M(N)$, and any positive operator $T : M(N) \to M(N)$, we define a *scaled operator* $T_{A,B}(X) = BT(AXA^*)B^*$. For any $X \succ 0$, if we write $Y = AXA^*$, then

$$\frac{\det(T_{A,B}X)}{\det(X)} = \det(A)^2 \det(B)^2 \cdot \frac{\det(TY)}{\det(Y)}. \tag{4}$$

Taking infima over both sides of (4) for all choices of input X shows that $\mathrm{Cap}(T_{A,B}) = \det(A)^2 \det(B^2) \mathrm{Cap}(T)$. Thus computing $\mathrm{Cap}(T_{A,B})$ immediately gives $\mathrm{Cap}(T)$. The idea of operator is to rescale an operator into

something whose capacity we can more easily compute. This 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 $(X, Y) \mapsto \text{Tr}(XY^*)$. What interests us about this definition is that $\text{Cap}(T) = 1$ for *any* doubly stochastic operator T. The proof of this statement can be found as Theorem 4.32 of [8], with relevant background information about doubly stochastic matrices found in Section 8.5 of [9].

Thus, given a positive operator T, if we can find $A, B \in M(N)$ such that $T_{A,B}$ is doubly stochastic, then it follows that $\text{Cap}(T) = \det(A)^{-2} \det(B)^{-2}$. This is clearly only possible if $\text{Cap}(T) > 0$, but not *quite* possible for all such T. In fact, in [3], it is shown that this rescaling is only possible if the equation defining capacity has extremizers, i.e. $\text{Cap}(T) = \det(TX)/\det(X)$ for some particular input X. Nonetheless, if $\text{Cap}(T) > 0$ then we will find doubly stochastic operators approximating T arbitrarily closely. Since capacity is a continuous function of the input, this suffices to approximate $\text{Cap}(T)$ up to an arbitrary error.

Gurvits' approximation algorithm is quite simple, an application of a similar method first utilized by Sinkhorn [10]. Given a positive operator T, it is easy to scale the operator to an operator S with $S(I) = I$; we simply consider the operator $T_{I,T(I)^{-1/2}}$. 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 approximately true. Sinkhorn's trick is to iteratively apply each of the rescalings to our operators, obtaining a sequence of matrices $T_0, T_1, T_2, \ldots$ with $T_0 = T$, $T_i(I) = I$ for odd i, and $T_i^*(I) = I$ for even i. If this sequence converges, the limit will be doubly stochastic, as desired.

Convergence to a single stochastic matrix does not occur for all T with $\text{Cap}(T) > 0$, but we will show that if $\text{Cap}(T) > 0$, then the distance between the elements of the sequence to the family of *all* stochastic operators converges 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 Gurvits first defined the capacity in [5]). To determine how close we are to a doubly stochastic matrix, we use the measure

$$\text{DS}(T) = \|T(I) - I\|^2 + \|T^*(I) - I\|^2.$$

The following properties then hold:

1. If $T(I) = I$ or $T^*(I) = I$, then $A_N \lesssim_N \mathrm{Cap}(T) \leq 1$ for some constant $A_N > 0$ depending on N. Moreover, for $\mathrm{Cap}(T) \geq 1/2$,

$$\mathrm{DS}(T_n) \leq 6 \log(1/\mathrm{Cap}(T_n)).$$

2. $\mathrm{Cap}(T_n)$ is increasing in $n > 0$. More precisely,

$$\mathrm{Cap}(T_{n+1}) \geq e^{\min(1, \mathrm{DS}(T_n))/6} \cdot \mathrm{Cap}(T_n).$$

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

We claim that from these properties, we can detect in $M_0 = 6(N+1)\log(1/A_N)$ iterations of the algorithm whether $\mathrm{Cap}(T) > 0$. Indeed, if $\mathrm{Cap}(T) > 0$, then Property 1 implies that $\mathrm{Cap}(T_1) \geq A_N$. For any δ and M, repeated applications of Property 3 imply that there either exists $i < M$ such that $\mathrm{DS}(T_i) \leq \delta$, or $\mathrm{Cap}(T_M) \geq e^{\varepsilon M/6} A_N$. In particular, if $\delta = 1/(N+1)$, and $M = M_0$, then this calculation, combined with Property 1, implies that if $\mathrm{Cap}(T) > 0$, then there must exist $i \leq M_0$ with $\mathrm{DS}(T_i) < 1/(N+1)$. On the other hand, Property 3 implies that if $\mathrm{Cap}(T) = 0$, then $\mathrm{DS}(T_n) \geq 1/(N+1)$ for all n. Checking $\mathrm{DS}(T_i)$ for $i \leq M_0$ thus gives a simple way to check whether $\mathrm{Cap}(T) > 0$. Using very similar techniques, we leave it as an exercise to check these properties imply that if $\mathrm{Cap}(T) > 0$, then in $(6/\varepsilon)\log(1/A_N)$ iterations of the algorithm, one can find i such that $\mathrm{DS}(T_i) \leq \varepsilon$. Thus we have a reliable way to approximate the capacity of a matrix, which [4] extends to find efficient ways to approximate Brascamp-Lieb constants. More advanced techniques, given in [3], show that we only actually need to run the algorithm for $\mathrm{Poly}(N, \log(1/\varepsilon))$ iterations to obtain a ε-approximation.

3 Capacity of Rank-Decreasing Operators

We conclude these notes by discussing an analogy between Brascamp-Lieb and the study of capacity. Bennett, Carbery, Christ, and Tao [1] showed that $\mathrm{BL}(B, p)$ is finite in (1) if and only if $\sum_{j=1}^{m} p_j n_j = 0$, and

$$\dim(V) \leq \sum_{j=1}^{m} p_j \dim(B_j V) \qquad \text{for all subspaces } V \subset \mathbf{R}^n. \tag{5}$$

Thus the finiteness of (1) acts as a guarantee for the mapping properties of the matrices $B_1, \ldots, B_m$ given in (5), and vice versa. A useful property of a positive operator T is that it is *rank non-decreasing*, i.e. for any $X \succeq 0$,

$$\text{Rank}(TX) \geq \text{Rank}(X). \tag{6}$$

Condition (6) seems somewhat similar to (5). And indeed, analogous to the equivalence between (5) and the finiteness of (1), one has an equivalence between (6) and the non-vanishing of (3).

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

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

$$TX = X_{11}A_1 + \cdots + X_{NN}A_N, \tag{7}$$

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

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

Results from a previous paper of Gurvits and Samorodnitsky [6] imply Theorem 1 in the special case of an operator defined by (7). 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 defined such that $T_U(X) = (T \circ D_U)(X) = \sum_{i=1}^{N} \langle Xu_i, u_i \rangle \cdot T(u_i u_i^*)$. This operator is, up to a change of basis in $M(N)$, described in the form (7). Thus T_U is rank non-decreasing if and only if $\text{Cap}(T_U) > 0$. 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. $\text{Cap}(T) = \inf_U \text{Cap}(T_U)$.

If $\text{Cap}(T) > 0$, then Property 2 implies $\text{Cap}(T_U) > 0$ 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 (all doubly stochastic operators are rank non-decreasing). $\square$

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.

[10] Richard Sinkhorn, *A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices*, Ann. Math. Statist. **35** (1964), 876–879.

Jacob Denson, UW Madison
email: jcdenson@wisc.edu