Algorithmic Aspects of the Brascamp Lieb Inequality

Jacob Denson

University of Wisconsin Madison

September 29, 2021

- *Classical Complexity and Quantum Entanglement*
  Gurvits, 2004.
- *Algorithmic and Optimization Aspects of Brascamp-Lieb Inequalities, Via Operator Scaling*
  Garg, Gurvits, Oliveira, Wigderson, 2016.
- *A Deterministic Polynomial Time Algorithm For Non-Commutative Rational Identity Testing*
  Garg, Gurvits, Oliveira, Wigderson, 2016.

Lieb's Theorem

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

- (Lieb, 1990) To calculate BL(B,p), plug in Gaussians.
- If $f_i(x) = e^{-\pi (A_i x)\cdot x}$ for $A_i \succ 0$, then

$$\int_{\mathbf{R}^n} \prod_{i=1}^m |f_i(B_i x)|^{p_i} \, dx = \det\Big(\sum p_i B_i^* A_i B_i\Big)^{-1/2}$$

$$\prod_{i=1}^m \|f_i\|_{L^1(\mathbf{R}^{n_i})}^{p_i} = \det\Big(\prod_{i=1}^m \det(A_i)^{p_i}\Big)^{-1/2}.$$

Thus

$$\Big(\sum p_i B_i^* A_i B_i\Big)^{-1/2} \le \mathrm{BL}(B,p) \cdot \Big(\prod_{i=1}^m \det(A_i)^{p_i}\Big)^{-1/2}$$

Lieb's Theorem

- Rearranging gives

$$\mathrm{BL}(B, p) = \left(\sup_{A_1,\dots,A_m \succ 0} \frac{\prod_i \det(A_i)^{p_i}}{\det(\sum p_i B_i^* A_i B_i)} \right)^{1/2}.$$

- Non convex objective, so tricky to optimize (NP hard).
- Checking Finiteness efficiently also non-obvious.
  - BCCT gives a family of conditions to check finiteness.

$$\dim(V) \leq \sum p_i \dim(B_i V).$$

    Since $0 \leq \dim(V) \leq n$ and $0 \leq \dim(B_i V) \leq n_i$, only finitely many linear conditions on the exponents p_i.
  - Can be exponentially many constraints, so inefficient.

Geometric Brascamp Lieb

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

- A Brascamp-Lieb inequality is *geometric* if
  - (Projection Property): $B_i B_i^* = I$.
  - (Isotropy Property): $\sum_i p_i B_i^* B_i = I$.
- (Barthe, 1998), generalizing (Ball, 1989), showed that if (B, p) is geometric, $\mathrm{BL}(B, p) = 1$.

Operator Rescaling

- What happens to $\mathrm{BL}(B, p)$ if we 'change coordinates'.
- Fix invertible matrices M and $M_1, \ldots, M_m$, and consider the Brascamp-Lieb inequality with matrices $B_i' = M_i B_i M$,

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

$$\begin{aligned}
\mathrm{BL}(B', p) &= \sup_{A_1, \ldots, A_m \succ 0} \frac{\prod_i \det(A_i)^{p_i}}{\det(\sum_i p_i M^* B_i^* M_i^* A_i M_i B_i M)} \\
&= \det(M)^{-2} \prod_i \det(M_i)^{-2p_i} \cdot \mathrm{BL}(B, p).
\end{aligned}$$

Operator Rescaling

$$\mathrm{BL}(B', p) = \det(M)^{-2} \prod_i \det(M_i)^{-2p_i} \cdot \mathrm{BL}(B, p)$$

- If (B', p) is geometric, $\mathrm{BL}(B', p) = 1$, so

$$\mathrm{BL}(B, p) = \det(M)^{2} \prod_i \det(M_i)^{2p_i}.$$

- (BCCT) Geometric rescaling possible iff extremizers exist.

Main Result

- (Garg, Gurvits, Oliveira, Wigderson, 2018) If $\mathrm{BL}(B, p) < \infty$, we can rescale to be *w*ithin a ε of a geometric Brascamp-Lieb equation.
  - We can do this algorithmically, i.e. if p_i are rationals, with common denominator d, a computer can compute a ε-approximate geometric rescaling in $\mathrm{Poly}(\mathrm{Bits}(B, p), d, 1/\varepsilon)$ computations.
  - Conversely, we can determine if $\mathrm{BL}(B, p) = \infty$, and find a violating subspace V to the condition $\dim(V) \leq \sum p_i \dim(B_i V)$, in $\mathrm{Poly}(\mathrm{Bits}(B, p), d)$ computations.
  - Open Problem: Can we can improve this to $\mathrm{Poly}(\mathrm{Bits}(B, p), d, \log(1/\varepsilon))$ computations?

Analytic Consequences

- Typing a single digit is a single unit of computation.
  - Thus given (B, p), either $\mathrm{BL}(B, p) < \infty$, or

$$2^{-\mathrm{Poly}(\mathrm{Bits}(B,p),d,\log(1/\varepsilon))} \leq \mathrm{BL}(B, p)) \leq 2^{\mathrm{Poly}(\mathrm{Bits}(B,p),d,\log(1/\varepsilon))}.$$

- Calculation is also stable.
  - Thus we conclude a *continuity bound*:
  - For each BL input (B, p) with $\mathrm{BL}(B, p) < \infty$, and each $\varepsilon > 0$, if

$$\|B_1 - B\| \leq \exp\left(-\frac{\mathrm{Poly}(\mathrm{Bits}(B, p), d)}{\varepsilon^3}\right),$$

then

$$(1-\varepsilon)\mathrm{BL}(B, p) \leq \mathrm{BL}(B_1, p) \leq (1+\varepsilon)\mathrm{BL}(B, p).$$

Computing Permanents (An Analogous Problem)

- Given an $n \times n$ matrix A with non-negative entries, define

$$\mathrm{Perm}(A) = \sum_{\sigma \in S_n} \prod_i A_{i\sigma(i)}.$$

- This is (surprisingly) NP hard to compute.
  - If

$$R = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \gamma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \gamma_n \end{pmatrix},$$

  then

$$\begin{aligned} \mathrm{Perm}(RAC) &= \mathrm{Perm}(\lambda_i A_{ij} \gamma_j) \\ &= (\lambda_1 \dots \lambda_n)(\gamma_1 \dots \gamma_n)\mathrm{Perm}(A) \\ &= \det(R)\det(C)\mathrm{Perm}(A). \end{aligned}$$

Computing Permanents (An Analogous Problem)

$$\mathrm{Perm}(RAC) = \det(R)\det(C)\mathrm{Perm}(A).$$

- (Egorychev, 1981), (Falikman, 1981) If A is *doubly stochastic*, $e^{-n} \leq \mathrm{Perm}(A) \leq 1$.
- If RAC is doubly stochastic, then

$$\mathrm{Perm}(A) \leq \det(R)^{-1}\det(C)^{-1} \leq e^{n} \cdot \mathrm{Perm}(A),$$

so $\mathrm{Perm}(A) \approx \det(R)^{-1}\det(C)^{-1}$.

- Given a non-negative matrix A_0, alternatively apply row and column summation, obtaining a sequence

$$A_0 \to A_1 \to A_2 \to \dots.$$

- Claim: If $\mathrm{Per}(A_0) > 0$, $d(A_i, \mathbf{S}) \to 0$, where $\mathbf{S}$ is the family of stochastic matrices.
- For even (odd) i, let γ_{ij} be the jth row (column) sum of A_i, so that $\mathrm{Per}(A_{i+1}) = (\gamma_{i1} \dots \gamma_{in})^{-1} \cdot \mathrm{Per}(A_i)$.
- **Two Key Facts Ensuring Convergence**:
  (1) $\mathrm{Per}(A) \leq 1$ if A is partially normalized.
  (2) If $\Delta_i = \sum_j (\gamma_{ij} - 1)^2$, $\mathrm{Per}(A_{i+1}) \geq (1 + C\Delta_i) \cdot \mathrm{Per}(A_i)$.
- Thus $\mathrm{Per}(A_i)$ is bounded, monotonic, converges to $P \leq 1$.
- If $\mathrm{Per}(A_i) \geq P - \varepsilon$ for $\varepsilon \ll 1$, then

$$P \geq \mathrm{Per}(A_{i+1}) \geq (1+C\cdot\Delta_i)\cdot\mathrm{Per}(A_i) \geq (1+C\cdot\Delta_i)(P-\varepsilon).$$

Thus $\Delta_i \lesssim_{C,P} \varepsilon$. Taking $\varepsilon \to 0$ shows $\Delta_i \to 0$.

Sinkhorn Iteration

- **Proof that Per**$(A_i) \leq 1$**:**
  - Simple inductive argument: the hypothesis is that if each row of a matrix B all sum up to less than one, then $\mathrm{Per}(B) \leq 1$.
- **Proof that Per**$(A_{i+1}) \geq (1 + C\Delta_i) \cdot$ **Per**(A_i)**:**
  - Write $\gamma_i = 1 + \delta_i$.
    - Then $\sum \delta_i^2 = \Delta_i$, and $\sum \delta_i = 0$.
  - Since $1 + t \leq \exp(t - t^2/2 + t^3/3)$,

$$\begin{aligned}
\mathrm{Per}(A_i)/\mathrm{Per}(A_{i+1}) &= \gamma_1 \ldots \gamma_n \\
&= (1+\delta_1) \ldots (1+\delta_n) \\
&\leq \exp\left(\sum \delta_i - \sum \delta_i^2/2 + \sum \delta_i^3/3\right) \\
&\leq \exp(0 - \Delta_i/2 + \Delta_i^{3/2}/3) \\
&= 1 - \Delta_i/2 + O(\Delta_i^{3/2}).
\end{aligned}$$

And now, back to our regularly scheduled programming

$$\mathrm{BL}(B,p) = \sqrt{\sup_{A_1,\ldots,A_m \succ 0} \frac{\prod_i \det(A_i)^{p_i}}{\det(\sum p_i \cdot B_i^* A_i B_i)}}.$$

- Goal: Rescale our inputs so that
  - (Isotropy) $\sum p_i B_i^* B_i = I$.
  - (Projection) $B_i B_i^* = I$ for each i.

Iteration of Operator Rescaling

- Sinkhorn: Alternately apply the following two procedures:
  - (Isotropy Normalization)
    - Let $M = \sum_i p_i B_i^* B_i$.
    - Replace B_i with $B_i' = B_i M^{-1/2}$.
    - Then $\sum p_i (B_i')^* B_i' = 1$, i.e. isotropy holds.
  - (Projection Normalization)
    - Let $M_i = B_i B_i^*$.
    - Replace B_i with $B_i' = M_i^{-1/2} B_i$.
    - Then $(B_i')^* B_i' = I$ for each i.
  - We obtain a sequence $B \to B_1 \to B_2 \to \dots$.

Iteration of Operator Rescaling

- **Three Key Facts Ensuring Convergence:**
  (1) $\mathrm{BL}(B, p) \geq 1$ if (B, p) is partially normalized.
  (2) If $1 + \varepsilon \leq \mathrm{BL}(B_i, p) \leq 2$, then

$$\mathrm{BL}(B_{i+1}, p) \leq (1 - C\varepsilon^s)\mathrm{BL}(B_i, p)$$

  (3) If isotropy or projection holds, then

$$\|\sum p_i B_i^* B_i - I\| + \sum_i \|B_i^* B_i - I\| \lesssim \log(\mathrm{BL}(B, p)),$$

  and so $d(B, \mathbf{G}_p) \lesssim \log(\mathrm{BL}(B, p))$.
- Thus convergence occurs as with Sinkhorn iteration provided that $\mathrm{BL}(B, p) < \infty$.
- (1), (2), and (3) follow from techniques in the study of *positive operators*.

Another Viewpoint: Positive Operators

- For simplicity, assume that all spaces have the same ambient dimension (all B_i are square matrices).
- $\mathrm{BL}(B, p) < \infty$ can only hold if $\sum p_i = 1$.
- Also assume all A_i are equal, and let us consider optimizing the quantity

$$\inf_{A \succ 0} \frac{\det(\sum p_i B_i^* A B_i)}{\det(A)}$$

analogous to

$$\mathrm{BL}(B, p) = \sup_{A_1, \ldots, A_m \succ 0} \sqrt{\frac{\prod_i \det(A_i)^{p_i}}{\det(\sum p_i \cdot B_i^* A_i B_i)}}.$$

Positive Operators

- A linear map $T : M_{n_1} \to M_{n_2}$ is *completely positive* if there are $n_2 \times n_1$ matrices $B_1, \ldots, B_K$ and $p_i > 0$ such that
$$T(A) = \sum p_i B_i A B_i^*.$$
For simplicity, focus on the case $n_1 = n_2$.
- $T : M_n \to M_n$ is *positive* if $A \succeq 0$ implies $T(A) \succeq 0$.
- Another example of positive operators: For a non-negative matrix S, $T(A)$ is the diagonal matrix whose entries are precisely the vector Sa, where a is the vector formed from the diagonal entries of A.
- Given T, we have $T^*(A) = \sum B_i^* A B_i$.

Capacity of Operators

$$\inf_{A \succ 0} \frac{\det(\sum p_i B_i^* A B_i)}{\det(A)}$$

- (Gurvits, 2004) The *capacity* of a positive operator $T : M_n \to M_n$ is

$$\mathrm{Cap}(T) = \inf_{A \succ 0} \frac{\det(TA)}{\det(A)}.$$

- For any Brascamp-Lieb data (B, p), there exists a positive $T : M_n \to M_n$ and k such that $\mathrm{Cap}(T) = 1/\mathrm{BL}(B, p)^{2k}$.
- Positive operators are well studied in the quantum information theory literature, so reduction of BL to this theory is useful.

Doubly Stochastic Positive Operators

- (Isotropy) Let $T(A) = \sum p_i B_i^* A B_i$.
  - $\sum p_i B_i^* B_i = I$ holds iff $T(I) = I$.
- (Projection) Let $T(A) = B_i^* A B_i$.
  - $B_i B_i^* = I$ if and only if $T^*(I) = I$.
- If (B, p) is a Brascamp-Lieb datum with associated operator $T : M_n \to M_m$, then (B, p) is geometric if and only if T is *doubly stochastic*. For $n = m$ this means $T(I) = I$ and $T^*(I) = I$.

Operator Rescaling for Positive Operators

- If T is doubly stochastic, $\mathrm{Cap}(T) = 1$.
- We can rescale. If

  $$T_{M_1 M_2}(A) = M_2^* T(M_1^* A M_1) M_2,$$

  then $\mathrm{Cap}(T_{M_1,M_2}) = \det(M_1)^2 \det(M_2)^2 \cdot \mathrm{Cap}(T)$.
- Sinkhorn says to iterate

  $$T \mapsto T_{I, T(I)^{-1/2}} \quad \text{and} \quad T \mapsto T_{T^*(I)^{-1/2}, I}.$$

- If $\mathrm{Cap}(T) > 0$, iteration yields a rescaling arbirarily close to a doubly stochastic operator, in $\mathrm{Poly}(\mathrm{Bits}(B), 1/\varepsilon)$ time.

Upper Bounds For Capacity

- To guarantee efficiently, we need to show that for partially normalized T, $\mathrm{Cap}(T) \geq 1/e^{\mathrm{Poly}(\mathrm{Bits}(B))}$.
- (Gurvits, 2004) If $T(A) = \sum B_i A B_i^*$, and $\det(\sum B_i) \neq 0$, then $\mathrm{Cap}(T) \gtrsim (\mathrm{Bits}(B) \cdot n)^{-O(n)}$.
- If $\mathrm{Cap}(T) > 0$, there is $d > 0$ and $d \times d$ matrices C_i s.t.

$$\det(\sum C_i \otimes B_i) \neq 0 \quad \text{and} \quad \mathrm{Bits}(C) \leq \mathrm{Poly}(d, \mathrm{Bits}(B)).$$

- 'Fairly simple' to show that if $S(A) = \sum C_i A C_i^*$, and $(S \otimes T)(A) = \sum (C_i \otimes B_i) A (C_i \otimes B_i)^*$, then

$$\mathrm{Cap}(S \otimes T) \leq \mathrm{Cap}(T)^d \mathrm{Cap}(L)^n.$$

- Since $\mathrm{Cap}(L) \leq 1$, it follows that

$$\mathrm{Cap}(T) \geq \mathrm{Cap}(S)^{1/d} \gtrsim (\mathrm{Poly}(d, Bits(B))n)^{-O(n)}.$$

- Invariant theory shows we can choose $d \leq n^4[(n+1)!]^2$.

The Invariant Theory

- We have a group action of $\mathrm{SL}_n \times \mathrm{SL}_n$ on tuples $B = (B_1, \ldots, B_m)$, such that

$$(M, N) \circ B = (MB_1N, \ldots, MB_mN).$$

- Invariant Theory: Find the ring R of all 'invariant polynomials' $f(B)$ such that

$$f((M, N) \circ B) = f(B)$$

  for all $(M, N) \in \mathrm{SL}_n \times \mathrm{SL}_n$.
- Note: $f_C(B) = \det(\sum C_i \otimes B_i)$ is an invariant polynomial under this action for any C_i.

The Invariant Theory

- (Nayak and Subrahmanyam, 2010) R is a ring generated by the homogeneous polynomials $f_C(A) = \det(\sum C_i \circ A_i)$, for $d \times d$ matrices C_i.
  - Hilbert showed that for a fairly general family of group actions, R is finitely generated, so we should expect there is some d_0 such that R is generated by the polynomials above for $d \leq d_0$.
  - (Ivanyos, Qiao, Subrahmanyam, 2015) shows $d_0 \lesssim n^4[(n+1)!]^2$.
  - Thus if there exists $d \times d$ matrices C_i such that $f_C(B) \neq 0$, then we can choose $d \leq d_0$.

Rank Decreasing Operators

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

- (Bennett et al, 2008) implies that $\mathrm{BL}(B,p) < \infty$ if and only if $\sum p_i n_i = n$, and for any subspace $V \subset \mathbf{R}^n$,

$$\dim(V) \leq \sum p_i \dim(B_i V).$$

- An operator $T : M_n \to M_n$ is *rank non-decreasing* if for any $A \succeq 0$, $\mathrm{Rank}(TA) \geq \mathrm{Rank}(A)$.
- (Gurvits, 2004) $T : M_n \to M_n$ is rank non-decreasing if and only if $\mathrm{Cap}(T) > 0$.

Proof Idea

(Gurvits, 2004) $T : M_n \to M_n$ is rank non-decreasing if and only if $\mathrm{Cap}(T) > 0$.

- Results from (Gurvits and Samorodnitsky, 2002) show the result is true if $T(X) = \sum X_{ii} A_i$, where $A_i \succeq 0$. The general case can be reduced to this case.
- Given an orthonormal basis $U = \{u_1, \ldots, u_N\}$, we define the *decoherence operator* $D_U(A) = \sum \langle Au_i, u_i \rangle \cdot u_i u_i^*$.
- Define $T_U = D_U \circ T$.
- Result follows from the following two facts:
  (1) T is rank non-decreasing if and only if T_U is rank non-decreasing for all U.
  (2) $\mathrm{Cap}(T) = \inf_U \mathrm{Cap}(T_U)$.
- To prove (1) and (2), use a simple trick: Given $A \succeq 0$, find U diagonalizing $T(A)$. Then $T(A) = T_U(A)$.

Thanks For Listening!