Nodal Domains and Diffusion Processes

Jacob Denson

University of Wisconsin Madison

October 4, 2022

- Georgiev, Mukherjee, *Nodal Geometry, Heat Diffusion, and Brownian Motion*, Anal. PDE. **12** (2017), 133-148.
- Steinerberger, *Lower Bounds on Nodal Sets of Eigenfunctions via the Heat Flow*, Comm. Partial Differential Equations. **39** (2014), 2240-2261.
- Øksendal, *Stochastic Differential Equations*, Springer, 2003.
- Chung, *Green, Brown, and Probability and Brownian Motion*, World Scientific Publishing Company, 2002.

Nodal Domains

Goal

Study 'asymptotic geometry' of D_λ as $\lambda \to \infty$.

Main Result

- **Theorem**: There is $c_M > 0$ such that for any 'good' k-dimensional submanifold Σ of M, then

$$N(\Sigma, c_M/\lambda) = \{x \in M : d(x,\Sigma) < c_M/\lambda\}$$

doesn't contain D_λ.

- Consider the radius $1/\lambda$ tubular neighborhood

$$T_{1/\lambda}\Sigma = \bigcup_{x\in\Sigma}\{v \in (T_x\Sigma)^\perp : |v|_g \leq 1/\lambda\}.$$

The submanifold Σ is 'good' if the geodesic map $T_{1/\lambda}\Sigma \to N(\Sigma, 1/\lambda)$ is an embedding.

  - Local condition: All principal curvatures of Σ are $\lesssim \lambda$.
  - But no cheating globally!

Main Result

- **Theorem**: There is $c_M > 0$ such that for any 'good' k-dimensional submanifold Σ of M, then

$$N(\Sigma, c_M/\lambda) = \{x \in M : d(x, \Sigma) < c_M/\lambda\}$$

  doesn't contain D_λ.
- Can replace Σ with a finite union of $\Omega(1/\lambda)$ separated 'good' submanifolds. Or allow finite unions with 'transverse enough' intersections.
- There is $C_M > 0$ such that $D_\lambda \subset N(Z_\lambda, C_M/\lambda)$.
- Heuristic: Elliptic methods work for $O(1/\lambda)$ localized results. We study stochastic diffusions, which provide cool tools to analyze eigenfunctions!

Uncertainty Principle on Manifolds?

- What would an analogous result look like on $\mathbb{R}^d$?
- **Theorem**: Let D_λ be a nodal domain in $\mathbb{R}^d$. Then there is $c_d > 0$ such that if Σ is a finite union of $O(1/\lambda)$-separated k dimensional planes, then D_λ is not contained in $N(\Sigma, c_d/\lambda)$.
- Stronger Result: D_λ contains a ball of radius $O(1/\lambda)$.
- Version on Manifolds: Paper proves for any $\varepsilon > 0$, there is $r_0 > 0$ such that if $x_0 \in D_\lambda$ maximizes $|e_\lambda(x_0)|$ in D_λ, then D_λ contains $1 - \varepsilon_0$ percent of $B(x_0, r_0\lambda^{-1/2})$.

Continuous Stochastic Processes

- Here are three ways to define continuous stochastic processes:
  - As a Borel-measurable function
$$X : \Omega \to C([0, \infty), M).$$
  - As a family of correlated random variables
$$\{X_t : \Omega \to M : t \in [0, \infty)\}.$$
  - As a law predicting future behaviour from present behaviour, i.e. by defining quantities such as
$$\mathbb{E}^x(f(X)) = \mathbb{E}[f(X)|X_0 = x]$$
$$\mathbb{P}^x(P(X)) = \mathbb{P}(P(X)|X_0 = x).$$

Brownian Motion on $\mathbb{R}^d$

- A stochastic process $\{B_t\}$ such that:
  - For any $I = [t, s]$, given $B_t = x$, the random variable $d_I B = B_s - B_t$ is normally distributed with mean x and variance $s - t$.
  - For any family of disjoint intervals $I_1, \ldots, I_N \subset [0, \infty)$, with $I_k = [t_k, s_k]$, the random variables $d_{I_k} B$ are independent from one another.

Itô Diffusions

- Brownian Motion where diffusion is not radially symmetric.
- For each $x \in \mathbb{R}^d$, let $A(x)$ be a $d \times d$ positive semidefinite matrix. Then we have an Itô diffusion $\{X_t\}$ given in law by the 'Stochastic differential equation' $dX = A(X)dB$.
- For practical purposes, we have

$$X_{t+\delta} - X_t \approx A(X_t)[B_{t+\delta} - B_t]$$

  where the difference between the LHS and RHS is a random variable with mean $o(\delta)$, and variance $O(\delta)$.
- Diffuses faster in directions where A has large eigenvalues.

Itô Diffusions

- Can define Itô diffusions on compact Riemannian manifolds M given a section $A : M \to \mathrm{Hom}(TM)$ of positive definite matrices.
- We can define Brownian motion on a Riemannian manifold such that Brownian motion locally diffuses along geodesics at unit speed.

Connection to Elliptic Operators

- For any diffusion X, we can associate a semielliptic operator L, the *generator* of X, such that for $f \in C^\infty(M)$,

$$Lf(x) = \partial_t\{\mathbb{E}^x[f(X_t)]\}|_{t=0} = \lim_{t\to 0^+} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}.$$

- Second order because paths of X are 'half differentiable'.
- For Brownian motion (on $\mathbb{R}^d$ or a manifold M), $L = \Delta/2$.
- 'Morally' apply the Fundamental Theorem of Calculus to get *Dynkin's Formula*

$$\mathbb{E}^x[f(X_T)] = f(x) + \mathbb{E}^x\left[\int_0^T (Lf)(X_s)\, ds\right].$$

Application: Escape Times

- In Dynkin's formula, T can be a 'stopping time', i.e. any $[0, \infty)$ valued function of X which doesn't 'predict the future', i.e. if T stops at a time t, it must only stop because of the properties of X on $[0, T]$, and not behaviour on (T, ∞).
- Given an open, bounded set U, let

$$T_U = \inf\{t : X_t \notin U\}$$

  be the *escape time* of U.
- If B is Brownian motion on $\mathbb{R}^d$, and U is the escape time of a ball of radius $R^{1/2}$ centered at x, $\mathbb{E}^x[T_U] = R/n$.
- If B is Brownian motion on M, escape time will be slower if volume expands (negative curvature) and faster if volume contracts (positive curvature). But irrelevant for the values R we care about.

Feynman Kac Formula

- Reverses Dynkin's Formula: Solves PDEs via Diffusions.
- Physically Intuitive Situations:
  - (1) If $\partial_t u = Lu$ on M with $u_0 = f$, then
$$u(x,t) = \mathbb{E}^x[f(X_t)].$$
  - (2) $\partial_t u = Lu$ on $D \subset M$ with $u_0 = f$ and $u = 0$ on ∂M,
$$u(x,t) = \mathbb{E}^x[f(X_t)\chi_t],$$
where $\chi_t = \mathbb{I}(T_D > t)$ *kills* paths absorbed by ∂D.
  - (3) If $Lu = 0$ on $D \subset M$ with $u = \phi$ on ∂D, then
$$u(x) = \mathbb{E}^x\left[\phi(X_{T_D})\right].$$
  - Can also solve $\partial_t u = Lu$ with $\partial u / \partial \eta = 0$ on ∂D using 'reflection on Brownian motion', but a little more technical with singularities.

The Proof

- **Theorem**: There is $c_M > 0$ such that for any 'good' k-dimensional submanifold Σ of M, then

$$N(\Sigma, c_M/\lambda) = \{x \in M : d(x, \Sigma) < c_M/\lambda\}$$

  doesn't contain D_λ.
- Assume $e_\lambda \geq 0$ on D_λ. Let $x^* = \mathrm{argmax}\{e_\lambda(x)\}$.
- Let $p(x,t)$ and $u(x,t)$ solve $\partial_t = \Delta$ with initial / boundary conditions:
    - $p_0 = 0$ and $p = 1$ on ∂D_λ.
    - $u_0 = e_\lambda$, and $u = 0$ on ∂D_λ.

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

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

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

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

- Claim: If $\operatorname{Per}(A_0) > 0$, $d(A_i, \mathbf{S}) \to 0$, where $\mathbf{S}$ is the family of doubly stochastic matrices.

- Given a non-negative matrix A_0, alternatively apply row and column normalization, 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 doubly stochastic matrices.
- For even (odd) i, let γ_{ij} be the jth row (column) sum of A_i, so that $\mathrm{Per}(A_{i+1}) = (\gamma_{i1} \dots \gamma_{in})^{-1} \cdot \mathrm{Per}(A_i)$.

- Given a non-negative matrix A_0, alternatively apply row and column normalization, 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 doubly 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**:

- Given a non-negative matrix A_0, alternatively apply row and column normalization, 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 doubly 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.

- Given a non-negative matrix A_0, alternatively apply row and column normalization, 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 doubly 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 $\Delta_i = \sum_j (\gamma_{ij} - 1)^2$, $\mathrm{Per}(A_{i+1}) \geq (1 + C\Delta_i) \cdot \mathrm{Per}(A_i)$.

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

$$A_0 \to A_1 \to A_2 \to \ldots .$$

- Claim: If $\mathrm{Per}(A_0) > 0$, $d(A_i, \mathbf{S}) \to 0$, where $\mathbf{S}$ is the family of doubly stochastic matrices.
- For even (odd) i, let γ_{ij} be the jth row (column) sum of A_i, so that $\mathrm{Per}(A_{i+1}) = (\gamma_{i1} \ldots \gamma_{in})^{-1} \cdot \mathrm{Per}(A_i)$.
- **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 $P \leq 1$.

- Given a non-negative matrix A_0, alternatively apply row and column normalization, 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 doubly 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 $\Delta_i \lesssim \varepsilon$. Taking $\varepsilon \to 0$ shows $\Delta_i \to 0$.

- **Proof that Per**$(A_i) \leq 1$**:**

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 $\text{Per}(B) \leq 1$.

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)**:**

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)**:**
    - Really just more robust form of AGM inequality.

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 $\text{Per}(B) \leq 1$.
- **Proof that Per**$(A_{i+1}) \geq (1 + C\Delta_i) \cdot$ **Per**(A_i)**:**
  - Really just more robust form of AGM inequality.
  - If γ_{ij} are the row sums, then because the column sums are all one,

$$\frac{1}{n} \sum_j \gamma_{ij} = 1.$$

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):**
  - Really just more robust form of AGM inequality.
  - If γ_{ij} are the row sums, then because the column sums are all one,
$$\frac{1}{n}\sum_j \gamma_{ij} = 1.$$
  - AGM implies $\gamma_{i1} \ldots \gamma_{in} \geq 1$, and monotonicity follows from
$$\mathrm{Per}(A_{i+1}) = (\gamma_{i1} \ldots \gamma_{in})^{-1}\mathrm{Per}(A_i).$$

And now, back to our regularly scheduled programming

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

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

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) $\sum p_i B_i^* B_i = I$.

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:

Iteration of Operator Rescaling

- Sinkhorn: Alternately apply the following two procedures:
  - (Isotropy Normalization)

Iteration of Operator Rescaling

- Sinkhorn: Alternately apply the following two procedures:
  - (Isotropy Normalization)
    - Let $M = \sum_i p_i B_i^* B_i$.

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 $B_i' = B_i M^{-1/2}$.

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.

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)

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 $M_i = B_i B_i^*$.

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 $B_i' = M_i^{-1/2} B_i$.

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.

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

- **Two Key Facts Ensuring Convergence:**

Iteration of Operator Rescaling

- **Two Key Facts Ensuring Convergence:**
  (1) $\mathrm{BL}(B, p) \geq 1$ if (B, p) is partially normalized.

Iteration of Operator Rescaling

- **Two Key Facts Ensuring Convergence:**
  (1) $\mathrm{BL}(B, p) \geq 1$ if (B, p) is partially normalized.
  (2) For some $r \leq \mathrm{Poly}(\mathrm{Bits}(B), d, n)$, if

$$\mathrm{BL}(B_i, p) \geq 1 + \varepsilon,$$

  then

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

Iteration of Operator Rescaling

- **Two Key Facts Ensuring Convergence:**
  (1) $\mathrm{BL}(B, p) \geq 1$ if (B, p) is partially normalized.
  (2) For some $r \leq \mathrm{Poly}(\mathrm{Bits}(B), d, n)$, if

  $$\mathrm{BL}(B_i, p) \geq 1 + \varepsilon,$$

  then

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

- Thus convergence occurs as with Sinkhorn iteration provided that $\mathrm{BL}(B, p) < \infty$.

Iteration of Operator Rescaling

- **Two Key Facts Ensuring Convergence:**
  (1) $\mathrm{BL}(B, p) \geq 1$ if (B, p) is partially normalized.
  (2) For some $r \leq \mathrm{Poly}(\mathrm{Bits}(B), d, n)$, if

  $$\mathrm{BL}(B_i, p) \geq 1 + \varepsilon,$$

  then

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

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

Positive Operators

- A linear map $T : M_n \to M_n$ is *completely positive* if there are $n \times n$ matrices $B_1, \ldots, B_K$ and $p_i > 0$ such that

$$T(A) = \sum p_i B_i A B_i^*.$$

Positive Operators

- A linear map $T : M_n \to M_n$ is *completely positive* if there are $n \times n$ matrices $B_1, \ldots, B_K$ and $p_i > 0$ such that

$$T(A) = \sum p_i B_i A B_i^*.$$

- $T : M_n \to M_n$ is *positive* if $A \succeq 0$ implies $T(A) \succeq 0$.

Positive Operators

- A linear map $T : M_n \to M_n$ is *completely positive* if there are $n \times n$ matrices $B_1, \ldots, B_K$ and $p_i > 0$ such that

$$T(A) = \sum p_i B_i A B_i^*.$$

- $T : M_n \to M_n$ is *positive* if $A \succeq 0$ implies $T(A) \succeq 0$.
- Can reduce the study of non-negative matrices to 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.

Positive Operators

- A linear map $T : M_n \to M_n$ is *completely positive* if there are $n \times n$ matrices $B_1, \ldots, B_K$ and $p_i > 0$ such that

$$T(A) = \sum p_i B_i A B_i^*.$$

- $T : M_n \to M_n$ is *positive* if $A \succeq 0$ implies $T(A) \succeq 0$.
- Can reduce the study of non-negative matrices to 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 $T^*(A) = \sum p_i B_i^* A B_i$.

Further Connections

- For simplicity, look at Brascamp Lieb where all spaces have the same dimension (all B_i are square matrices).

Further Connections

- For simplicity, look at Brascamp Lieb where all spaces have the same dimension (all B_i are square matrices).
- $\mathrm{BL}(B, p) < \infty$ can only hold if $\sum p_i = 1$.

Further Connections

- For simplicity, look at Brascamp Lieb where all spaces have the same dimension (all B_i are square matrices).
- $\mathrm{BL}(B,p) < \infty$ can only hold if $\sum p_i = 1$.
- Consider optimizing the quantity

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

analogous to

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

if all A_i are equal.

Capacity of Operators

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

Capacity of Operators

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

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

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

Capacity of Operators

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

- (Gurvits, 2004) The *capacity* of $T : M_n \to M_n$ 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 $\mathrm{Cap}(T) = 1/\mathrm{BL}(B, p)^{2k}$.

Capacity of Operators

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

- (Gurvits, 2004) The *capacity* of $T : M_n \to M_n$ 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 $T(A) = \sum p_i B_i^* A B_i$.

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 $T(I) = I$.

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 $T(A) = B_i^* A B_i$.

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 $T^*(I) = I$.

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 $T : M_n \to M_n$, then (B, p) is geometric if and only if T is *doubly stochastic*, i.e. $T(I) = I$ and $T^*(I) = I$.

Operator Rescaling for Positive Operators

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

Operator Rescaling for Positive Operators

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

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

then $\text{Cap}(T_{M_1,M_2}) = \det(M_1)^2 \det(M_2)^2 \cdot \text{Cap}(T)$.

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

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

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

Upper Bounds For Capacity

- To guarantee efficiency, we need to show that for partially normalized T, $\mathrm{Cap}(T) \geq 1/e^{\mathrm{Poly}(\mathrm{Bits}(B))}$.

Upper Bounds For Capacity

- To guarantee efficiency, we need to show that for partially normalized T, $\text{Cap}(T) \geq 1/e^{\text{Poly}(\text{Bits}(B))}$.
- (Gurvits, 2004) If $T(A) = \sum B_i A B_i^*$, and $\det(\sum B_i) \neq 0$, then $\text{Cap}(T) \gtrsim (\text{Bits}(B) \cdot n)^{-O(n)}$.

Upper Bounds For Capacity

- To guarantee efficiency, 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.

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

Upper Bounds For Capacity

- To guarantee efficiency, 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

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

Upper Bounds For Capacity

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

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

- '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

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

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

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

Upper Bounds For Capacity

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

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

- '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

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

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

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

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

The Invariant Theory

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

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

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 $(M, N) \in \mathrm{SL}_n \times \mathrm{SL}_n$ and all tuples B.

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 $(M, N) \in \mathrm{SL}_n \times \mathrm{SL}_n$ and all tuples B.
- Note: $f_C(B) = \det(\sum C_i \otimes B_i)$ is an invariant homogeneous 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, for all $d > 0$.

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, for all $d > 0$.
  - (1890) 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 $d \leq d_0$. Killed the field for 100 years.

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, for all $d > 0$.
  - (1890) 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 $d \leq d_0$. Killed the field for 100 years.
  - (Ivanyos, Qiao, Subrahmanyam, 2015) $d_0 \lesssim n^4[(n+1)!]^2$.

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, for all $d > 0$.
  - (1890) 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 $d \leq d_0$. Killed the field for 100 years.
  - (Ivanyos, Qiao, Subrahmanyam, 2015) $d_0 \lesssim n^4[(n+1)!]^2$.
  - Suppose there are $d \times d$ matrices C_i (with d minimal) such that $f_C(B) \neq 0$.

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, for all $d > 0$.
  - (1890) 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 $d \leq d_0$. Killed the field for 100 years.
  - (Ivanyos, Qiao, Subrahmanyam, 2015) $d_0 \lesssim n^4[(n+1)!]^2$.
  - Suppose there are $d \times d$ matrices C_i (with d minimal) such that $f_C(B) \neq 0$.
  - Find families of matrices $C(1), \ldots, C(n)$ of dimension at most d_0 such that

$$f_C(B) = \sum c_\alpha f_{C(1)}(B)^{\alpha_1} \ldots f_{C(n)}(B)^{\alpha_n}.$$

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, for all $d > 0$.
  - (1890) 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 $d \leq d_0$. Killed the field for 100 years.
  - (Ivanyos, Qiao, Subrahmanyam, 2015) $d_0 \lesssim n^4[(n+1)!]^2$.
  - Suppose there are $d \times d$ matrices C_i (with d minimal) such that $f_C(B) \neq 0$.
  - Find families of matrices $C(1), \ldots, C(n)$ of dimension at most d_0 such that

$$f_C(B) = \sum c_\alpha f_{C(1)}(B)^{\alpha_1} \ldots f_{C(n)}(B)^{\alpha_n}.$$

  - Since $f_C(B) \neq 0$, there must exist i with $f_{C(i)}(B) \neq 0$.

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, for all $d > 0$.
  - (1890) 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 $d \leq d_0$. Killed the field for 100 years.
  - (Ivanyos, Qiao, Subrahmanyam, 2015) $d_0 \lesssim n^4[(n+1)!]^2$.
  - Suppose there are $d \times d$ matrices C_i (with d minimal) such that $f_C(B) \neq 0$.
  - Find families of matrices $C(1), \ldots, C(n)$ of dimension at most d_0 such that

$$f_C(B) = \sum c_\alpha f_{C(1)}(B)^{\alpha_1} \ldots f_{C(n)}(B)^{\alpha_n}.$$

  - Since $f_C(B) \neq 0$, there must exist i with $f_{C(i)}(B) \neq 0$.
  - Thus $d \leq d_0$

Rank Decreasing Operators

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

Rank Decreasing Operators

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

- (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 $V \subset \mathbb{R}^n$,

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

Rank Decreasing Operators

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

- (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 $V \subset \mathbb{R}^n$,

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

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

Rank Decreasing Operators

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

- (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 $V \subset \mathbb{R}^n$,

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

- 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 $\mathrm{Cap}(T) > 0$.

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.

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* $D_U(A) = \sum \langle Au_i, u_i \rangle \cdot u_i u_i^*$.

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 $T_U = D_U \circ T$.

Proof Idea

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

- 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:

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.

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

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!