Convex optimization and quantum information theory

Hamza Fawzi

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

QIP 2021
Tutorial

Convex optimization

- Basics, semidefinite programs
- Algorithms (Newton, interior-point)
- Convex relaxations for polynomial optimization

Quantum information

- Entropies
- Separable states and sums of squares

Convex optimization

- Solve/study optimization problem

$$\boxed{\min_{x \in C} \quad f(x)}$$

where f convex function and C convex set.

- Can be easy or hard depending on f and C

Examples

- Capacity of a cq-channel

$$\max \quad H\left(\sum_{i=1}^{n} p_i \sigma_i\right) - \sum_{i=1}^{n} p_i H(\sigma_i) \quad \text{s.t.} \quad p \in \Delta^{n-1}$$

where $H(\sigma) = -\operatorname{tr}[\sigma \log \sigma]$ von Neumann entropy,
$\Delta^{n-1} = \{p \geq 0, \sum_i p_i = 1\}$.

Examples

- Capacity of a cq-channel

$$\boxed{\max \quad H\left(\sum_{i=1}^{n} p_i\sigma_i\right) - \sum_{i=1}^{n} p_i H(\sigma_i) \quad \text{s.t.} \quad p \in \Delta^{n-1}}$$

where $H(\sigma) = -\operatorname{tr}[\sigma \log \sigma]$ von Neumann entropy,
$\Delta^{n-1} = \{p \geq 0, \sum_i p_i = 1\}$.

- Relative entropy of PPT, for fixed $\rho \in \mathbf{H}^{nm}$

$$\boxed{\min \quad D(\rho\|\sigma) \quad \text{s.t.} \quad \sigma \in \mathrm{PPT}}$$

where $D(\rho\|\sigma) = \operatorname{tr}[\rho(\log \rho - \log \sigma)]$, and

$$\mathrm{PPT} = \{\rho \succeq 0,\ \operatorname{tr}[\rho] = 1,\ \text{and } (I \otimes \mathsf{T})(\rho) \succeq 0\}\,.$$

Examples 2

- Best separable state

$$\boxed{\max \quad \mathrm{tr}[M\sigma] \quad \text{s.t.} \quad \sigma \in \mathrm{Sep}}$$

where $M \in \mathbf{H}^{n^2}$, and $\mathrm{Sep} = \mathrm{conv}\{xx^\dagger \otimes yy^\dagger : x, y \in (\mathbb{C}^n)^2, |x| = |y| = 1\}$.

Examples 2

- Best separable state

$$\boxed{\max \ \operatorname{tr}[M\sigma] \quad \text{s.t.} \quad \sigma \in \mathrm{Sep}}$$

where $M \in \mathbf{H}^{n^2}$, and $\mathrm{Sep} = \operatorname{conv}\{xx^\dagger \otimes yy^\dagger : x, y \in (\mathbb{C}^n)^2, |x| = |y| = 1\}$.

- Nonlocal games, Bell inequalities

$$\boxed{\max \ \sum_{abxy} w(ab|xy)p(ab|xy) \quad \text{s.t.} \quad (p(ab|xy))_{ab|xy} \in C_q}$$

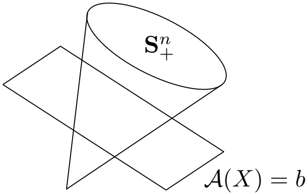
where C_q is the set of quantum correlations

$$C_q = \Bigg\{ p(a,b|x,y) = \langle \psi, (F_a^x \otimes G_b^y)\psi \rangle : d \in \mathbb{N}, \psi \in \mathbb{C}^d, F_a^x \in \mathbf{H}_+^d, G_b^y \in \mathbf{H}_+^d$$

$$\sum_{1 \le a \le m} F_a^x = I, \sum_{1 \le b \le m} G_b^y = I \ \ \forall 1 \le x, y \le n \Bigg\}.$$

Semidefinite programming

$$
\begin{array}{ll}
\min & \langle C, X \rangle \\
\text{s.t.} & X \succeq 0 \\
 & \langle A_i, X \rangle = b_i \ (i = 1, \ldots, m)
\end{array}
$$



Duality

$$p^* = \min \ \langle C, X\rangle \quad \text{s.t.} \quad X \succeq 0, \quad \langle A_i, X\rangle = b_i \ (i = 1, \ldots, m)$$

$$d^* = \max \ \langle b, z\rangle \quad \text{s.t.} \quad C - \sum_{i=1}^{m} z_i A_i \succeq 0.$$

- Weak duality: $p^* \geq d^*$
- Strong duality: $p^* = d^*$, holds assuming e.g., primal or dual problem are strictly feasible (Slater's condition)

Duality

$$p^* = \min \ \mathrm{tr}(CX) \quad \text{s.t.} \quad X \succeq 0, \quad \mathrm{tr}(A_i X) = b_i \ (i = 1, \dots, m)$$

$$d^* = \max \ \langle b, z \rangle \quad \text{s.t.} \quad C - \sum_{i=1}^{m} z_i A_i \succeq 0.$$

KKT conditions of optimality:

$$\begin{cases} X \succeq 0, \ \ \mathrm{tr}(A_i X) = b_i (i = 1, \dots, m) \\ S \succeq 0, \ \ S = C - \sum_{i=1}^{m} z_i A_i \\ XS = 0. \end{cases}$$

Nonlinear system in the unknowns (X, S, z)

Algorithms

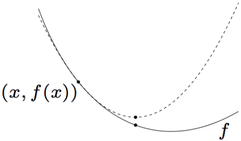
- Newton's method (unconstrained)
- Path-following methods

Newton's method

$$\boxed{\min_{x \in \mathbb{R}^n} f(x)}$$

- At iteration x_k, form quadratic approximation of f

$$f(x_k + h) \approx f(x_k) + \nabla f(x_k)^T h + \frac{1}{2} h^T \nabla^2 f(x_k) h$$



- If f strongly convex then $\nabla^2 f(x_k)$ positive definite, and quadratic approximation has a unique minimum, attained by taking

$$h^* = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Newton's method

$$x_{k+1} = x_k + t_k h^* = x_k - t_k[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$$

where $t_k > 0$ step size (default step size $t_k = 1$).

- Computational complexity: at each step need to solve the linear system

  $$\nabla^2 f(x_k)h = \nabla f(x_k)$$

  $O(n^3)$ flops for unstructured systems. Limiting factor when n large.

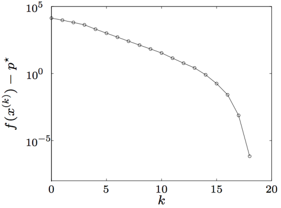
- Compare to gradient method where each iteration takes $O(n)$ flops.

- Method invariant under change of basis $\tilde{x} = Px$ where P invertible.

Convergence of Newton's method

$$x_{k+1} = x_k - t_k[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$$

- Extremely fast when close to optimal solution



- Quadratic convergence: $r_{k+1} \leq r_k^2$ for some measure of the residual r_k. If $r_0 < 1$ this is extremely fast convergence: $r_k \leq (r_0)^{2^k}$.

  Quadratic convergence region is $\{x \in \mathrm{dom}(f) : r_0 < 1\}$.

Quadratic convergence

- *Standard analysis of Newton's method:* Assume f is such that $\nabla^2 f(x) \succeq mI$, and $\nabla^2 f$ is M-Lipschitz (in the operator norm). Then with $r_k = \frac{M}{2m^2}\|\nabla f(x_k)\|_2$ we have $r_{k+1} \leq r_k^2$.

Quadratic convergence

- *Standard analysis of Newton's method:* Assume f is such that $\nabla^2 f(x) \succeq mI$, and $\nabla^2 f$ is M-Lipschitz (in the operator norm). Then with $r_k = \frac{M}{2m^2}\|\nabla f(x_k)\|_2$ we have $r_{k+1} \leq r_k^2$.
- *Nesterov & Nemirovski:* Self-concordant functions f, defined by a Lipschitz condition on $\nabla^2 f$ with respect to local metric

$$\|v\|_x = \sqrt{v^T \nabla^2 f(x) v}$$

Self-concordance: $\|H_x(y) - H_x(x)\|_x \leq \phi(\|y - x\|_x)$

where $\phi(t) = 1/(1-t)^2 - 1$.

Quadratic convergence

- *Standard analysis of Newton's method:* Assume f is such that $\nabla^2 f(x) \succeq mI$, and $\nabla^2 f$ is M-Lipschitz (in the operator norm). Then with $r_k = \frac{M}{2m^2}\|\nabla f(x_k)\|_2$ we have $r_{k+1} \leq r_k^2$.

- *Nesterov & Nemirovski:* Self-concordant functions f, defined by a Lipschitz condition on $\nabla^2 f$ with respect to local metric

$$\|v\|_x = \sqrt{v^T \nabla^2 f(x) v}$$

$$\text{Self-concordance: } \|H_x(y) - H_x(x)\|_x \leq \phi(\|y - x\|_x)$$

where $\phi(t) = 1/(1-t)^2 - 1$.

- Advantage of self-concordance: analysis of Newton's method does not depend on any constants, and is invariant under change of coordinates:

$$\lambda(x_{k+1}) \leq \frac{\lambda(x_k)^2}{(1 - \lambda(x_k))^2}$$

where $\lambda(x) = \|g_x(x)\|_x$.

Relationship with Newton-Raphson method

- Newton-Raphson method to solve a system of nonlinear equations, $F(x) = 0$ where $F : \mathbb{R}^n \to \mathbb{R}^n$

- Linearize at x_k
$$F(y) \approx F(x_k) + DF(x_k)(y - x_k)$$

- Equate right-hand side to zero to get
$$x_{k+1} = x_k - (DF(x_k))^{-1} F(x_k)$$

- Newton's method for optimization is Newton-Raphson applied to the system $\nabla f(x) = 0$.

Constrained problems

$$\boxed{\min_{x \in C} \ \langle c, x \rangle}$$

Let $B(x)$ be a convex barrier function for C so that $B(x) \to +\infty$ as $x \to \partial C$.
For example if $C = \{x : \langle a_i, x \rangle \leq b_i\}$ then we choose

$$B(x) = -\sum_i \log(b_i - \langle a_i, x \rangle).$$

We consider, for $t > 0$:

$$\boxed{x^*(t) = \operatorname*{argmin}_x \ \ t\langle c, x \rangle + B(x).}$$

This traces out a path in the interior of C. Can show that

$$\langle c, x^* \rangle \ \ \leq \ \ \langle c, x^*(t) \rangle \ \ \leq \ \ \langle c, x^* \rangle + \theta / t$$

where θ is some parameter that depends on B.

Goal: trace out the path with $t \to \infty$

Central path

$$x^*(t) = \underset{x}{\text{argmin}} \quad B_t(x)$$

where $B_t(x) = t\langle c, x\rangle + B(x)$.

Barrier method, assuming initial point $x^*(t_0)$ for $t_0 > 0$ given:

- `Let` $t = t_0$`,` $x = x^*(t_0)$
- `While not converged (`$t < \theta/\epsilon$`):`
- `Set` $t^+ \leftarrow \alpha t$ `for some` $\alpha > 1$
- `Compute` $x^*(t^+)$ `by using Newton's method starting from` $x^*(t)$
- `Update` $t \leftarrow t^+$

How to choose α? We want:

$x^*(t)$ is in the quadratic convergence region of $B_{\alpha t}$

Illustration

- Theory tells us that $\alpha \approx 1 + 1/\sqrt{\theta}$ works. This leads to a number of iterations $\approx \sqrt{\theta}$
- Practice: such α is too small. Use "predictor-corrector" approach where one predicts at each iteration how large α can you choose.

Application to SDPs

Dual SDP

$$\max \quad \langle b, z\rangle \quad \text{s.t.} \quad C - \sum_{i=1}^{m} z_i A_i \succeq 0.$$

Barrier function $B(z) = \log\det\left(C - \sum_{i=1}^{m} z_i A_i\right)$.

$$\max \quad t\langle b, z\rangle + B(z)$$

Need to compute ∇B_t and $\nabla^2 B_t$. If we let $S = C - \sum_{i=1}^{m} z_i A_i \succ 0$, then

$$[\nabla B_t(z)]_i = tb_i - \langle A_i, S^{-1}\rangle$$
$$[\nabla^2 B_t(z)]_{ij} = -\operatorname{tr}[S^{-1} A_i S^{-1} A_j].$$

Cost of forming the Hessian is $\approx mn^3 + m^2 n^2$, and cost of solving linear system for Newton's method is m^3.

Barrier path and KKT equations

At "time t" of central path we have $\nabla B_t(z) = 0$ i.e.,

$$b_i = \frac{1}{t}\langle A_i, S^{-1}\rangle.$$

If we call $X = \frac{1}{t}S^{-1}$ then the following hold:

$$\begin{cases} X \succeq 0, \ \ \operatorname{tr}(A_i X) = b_i \ (i = 1, \ldots, m) \\ S \succeq 0, \ \ S = C - \sum_{i=1}^{m} z_i A_i \\ XS = \frac{1}{t} I \end{cases}$$

Modified KKT system!

Discussion

- Complexity of solving SDP is $\approx \sqrt{n}(n^4 + m^2 n^2 + m^3)$ floating-point operations
- Well-implemented interior-point method reliable and can give high accurate solutions. Solvers include `Mosek`, `SeDuMi`, `SDPT3`, `SDPA`, `CSDP`, ...
- Recently, research focus on simpler (first-order) methods that scale to large problems, e.g., ADMM.
- Good implementation of ADMM: SCS (Splitting Conic Solver) [O'Donoghue]
- Main drawback of first-order methods is accuracy. Slow convergence/stall at low-medium accuracy.

Polynomial optimization, sums of squares

Polynomial optimization

Let $p \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial.

- Decision question:

$$\text{is } p(x) \geq 0 \quad \forall x \in \mathbb{R}^n?$$

- NP-hard (but decidable, by Tarski's theorem)

- Sufficient condition for nonnegativity of p is that p is a sum-of-squares:

$$p(x) = \sum_i q_i(x)^2$$

  for some polynomials $q_i(x)$.
  - If $\deg p = 2d$ then the q_i are necessarily of degree $\leq d$
  - If p is sos, then it is a sum of squares of finitely many polynomials ($\leq \binom{n+d}{d}$).

Sums of squares and semidefinite programming

Deciding if p is a sum-of-squares is a semidefinite feasibility problem

Hilbert's question

- Are all nonnegative polynomials sums of squares? No! Hilbert (1888) showed that nonnegative polynomials are sos only in the following cases:
    - $n = 1$ (one variable)
    - $2d = 2$ (quadratics)
    - $n = 2, 2d = 4$

  In all the other cases, there are nonnegative polynomials that are not sums of squares

- Motzkin's polynomial $(n = 2, 2d = 6)$ is one such example

$$M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2.$$

- Hilbert's 17th problem: are all nonnegative polynomials sums of squares of rational functions? Question answered positively by Artin (1927)

Optimization

$$\min_{x\in\mathbb{R}^n} p(x) \quad = \quad \max \quad \gamma \quad \text{s.t.} \quad p-\gamma \geq 0.$$

Sum-of-squares relaxation:

$$\max \quad \gamma \quad \text{s.t.} \quad p-\gamma \text{ sum-of-squares}$$

This is an SDP.

Constrained polynomial optimization: the case of the sphere

Let $p(x)$ be a (homogeneous) polynomial of degree $2d$.

$$p_{\min} := \min \quad p(x) \quad \text{s.t.} \quad \sum_{i=1}^{n} x_i^2 = 1.$$

Define

$$\begin{aligned} p_\ell = \quad & \max_{\gamma, s(x), g(x)} && \gamma \\ & \text{s.t.} && p(x) - \gamma = s(x) + g(x)(|x|^2 - 1) && \text{Equality of polynomials} \\ &&& s(x) \text{ is a sum-of-squares}, \deg s \leq 2\ell && \text{Semidefinite constraint} \\ &&& g(x) \in \mathbb{R}[x], \deg g \leq 2\ell - 2 \end{aligned}$$

- Each p_ℓ is a lower bound on $p_{\min}$

$$\cdots \leq p_{\ell-1} \leq p_\ell \leq p_{\min}$$

- Note that $\deg s$ can be larger than $\deg p$! (Cancellation)

Relaxations based on linear programming

- Any *certificates of nonnegativity* give us a way to construct a relaxation
- Example (positivity on $\mathbb{R}^n_+$): Given $p \in \mathbb{R}[x]$

$$\text{is } p(x) \geq 0 \quad \forall x = (x_1, \ldots, x_n) > 0\ ?$$

- Sufficient condition 1: coefficients of p are ≥ 0
- Sufficient condition 2: there is $N \in \mathbb{N}$ such that

$$(\sum_{i=1}^{N} x_i)^n p(x) \text{ has nonnegative coefficients.}$$

- Polya's theorem: if $p > 0$ on $\mathbb{R}^n_+$ then sufficient condition 2 is also necessary.

Linear programming hierarchy

Polya's theorem suggests a hierarchy of linear programs to compute

$$p_{\min} \;=\; \min \;\; p(x) \;\text{ s.t. }\; x \geq 0,\;\; \sum_{i=1}^{n} x_i = 1.$$

If p is homogeneous of degree d, consider

$$\begin{array}{lll} p_N = & \max & \gamma \\ & \text{s.t.} & (\sum_{i=1}^n x_i)^N (p(x) - \gamma(\sum_{i=1}^n x_i)^d) \text{ has nonnegative coefficients} \end{array}$$

- For each N, $p_N \leq p_{\min}$, and Polya's theorem guarantees that $p_N \to p_{\min}$.
- Computing p_N is a linear program, with a number of inequalities equal to $\binom{n+N+d}{n}$ = dimension of space of polynomials of degree at most $N + d$ in n variables.

Quantum information

- Entropies
- The set of separable states

Entropies

- Relative entropy, jointly convex in (ρ, σ) [Lieb-Ruskai]

$$D(\rho\|\sigma) = \mathrm{tr}[\rho(\log\rho - \log\sigma)]$$

- Petz Rényi divergence, concave for $\alpha \in [0, 1]$ and convex for $\alpha \in [1, 2]$ [Lieb]

$$Q_\alpha(\rho\|\sigma) = \mathrm{tr}[\rho^\alpha \sigma^{1-\alpha}]$$

- Geometric Rényi divergence [Matsumoto]

$$\hat{Q}_\alpha(\rho\|\sigma) = \mathrm{tr}[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha].$$

Optimizing entropies

Optimization problems with Rényi divergences are very common in QI

To solve such optimization problems, either:

1. Implement your own optimization algorithm, using expression for the gradient (and Hessian) of these functions.

   `Pros: can exploit problem structure with custom algorithm`

   `Cons: have to implement your algorithm (deal with choices of step sizes, scaling, ...), time-consuming`

2. Or, try to formulate such problems as *semidefinite programs*, and rely on existing solvers for SDPs.

   `Pros: very easy to use/compose with other constraints via interfaces such as CVX, Yalmip`
   `Cons: restriction to SDPs means we have to do approximations (more on this later)`

We'll focus on semidefinite representations

Interfaces to convex solvers

Example: distance to PPT in the relative entropy sense

$$\min\ D(\rho\|\tau)\ \text{ s.t. }\ \tau \in \mathrm{PPT}$$

```
cvx_begin sdp
  variable tau(na*nb,na*nb) hermitian;
  minimize    (quantum_rel_entr(rho,tau));
  subject to  tau >= 0; trace(tau) == 1;
              Tx(tau,2,[na nb]) >= 0;  % Positive partial transpose
cvx_end
```

Semidefinite representations

- Concave function f has a *semidefinite representation* if:

$$f(x) \geq t \qquad \Longleftrightarrow \qquad \mathcal{S}(x,t) \succeq 0$$

for some affine function $\mathcal{S} : \mathbb{R}^{n+1} \to \mathbf{H}^d$

- Key fact: if f has a semidefinite representation then can solve optimisation problems involving f using semidefinite solvers.

Semidefinite representations

- Concave function f has a *semidefinite representation* if:

$$f(x) \geq t \qquad \iff \exists u \in \mathbb{R}^m : \mathcal{S}(x,t,u) \succeq 0$$

for some affine function $\mathcal{S} : \mathbb{R}^{n+1+m} \to \mathbf{H}^d$

- Key fact: if f has a semidefinite representation then can solve optimisation problems involving f using semidefinite solvers.

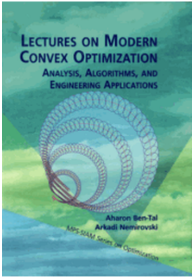
Semidefinite representations

- Concave function f has a *semidefinite representation* if:

$$f(x) \geq t \qquad \Longleftrightarrow \exists u \in \mathbb{R}^m : \mathcal{S}(x, t, u) \succeq 0$$

for some affine function $\mathcal{S} : \mathbb{R}^{n+1+m} \to \mathbf{H}^d$

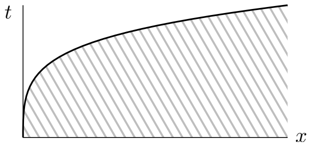
- Key fact: if f has a semidefinite representation then can solve optimisation problems involving f using semidefinite solvers.



- Book by Ben-Tal and Nemirovski gives semidefinite representations of many convex/concave functions.

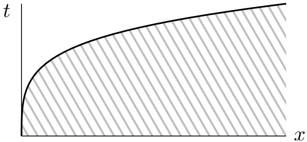
Examples of semidefinite formulation

$$\sqrt{x} \geq t \quad \Leftrightarrow \quad \begin{bmatrix} x & t \\ t & 1 \end{bmatrix} \succeq 0$$

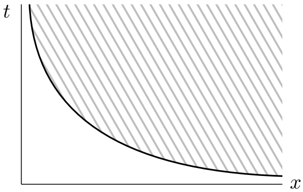


Examples of semidefinite formulation

$$\sqrt{x} \geq t \quad \Leftrightarrow \quad \begin{bmatrix} x & t \\ t & 1 \end{bmatrix} \succeq 0$$



$$\frac{1}{x} \leq t \quad \Leftrightarrow \quad \begin{bmatrix} x & 1 \\ 1 & t \end{bmatrix} \succeq 0$$



Matrix geometric mean

Geometric mean of $A, B \succ 0$ is [Kubo-Ando]

$$A\#B = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2}$$

- Homogeneous, $(rA)\#(rB) = r(A\#B)$
- Jointly operator concave in (A, B)

$$(A_1 + A_2)\#(B_1 + B_2) \succeq A_1\#B_1 + A_2\#B_2$$

- Symmetric: $A\#B = B\#A$.

SDP representation of geometric mean

$$A \# B = \max_{X \succeq 0} \left\{ X : \begin{bmatrix} A & X \\ X & B \end{bmatrix} \succeq 0 \right\}$$

Proof

SDP representation of geometric mean (2)

$$A \# B \succeq T \iff \exists X \text{ s.t. } \begin{bmatrix} A & X \\ X & B \end{bmatrix} \succeq 0 \text{ and } X \succeq T.$$

SDP representation of geometric means

The t-geometric mean of (A, B) is [Kubo-Ando]

$$A \#_t B = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^t A^{1/2}$$

Jointly concave in (A, B) for $t \in [0, 1]$ / convex for $t \in [-1, 0] \cup [1, 2]$.
Semidefinite representation:

- Use composition property

$$A \#_{1/4} B = A \# (A \# B)$$

  $+$ monotonicity of $\#$ in 2nd argument, to get:

$$A \#_{1/4} B \succeq T \iff \exists Z : \begin{cases} A \# B \succeq Z \\ A \# Z \succeq T. \end{cases}$$

- Can get $A \#_{3/4} B$ using $A \#_{3/4} B = B \#_{1/4} A$.
- $\Rightarrow$ SDP representation of $A \#_t B$ for all dyadic numbers $t \in [0, 1]$.

SDP representation of geometric mean for $t \in [-1, 0]$

For $t \in [-1, 0]$ we use $A \#_{-t} B = A \#_{-1} (A \#_t B)$ to get

$$A \#_{-t} B \preceq T \iff \exists Z : \begin{cases} A \#_t B \succeq Z \\ A \#_{-1} Z \preceq T \end{cases}$$

and

$$A \#_{-1} Z \preceq T \iff \begin{bmatrix} T & A \\ A & Z \end{bmatrix} \succeq 0.$$

(Schur complement)

Petz divergence

$$Q_\alpha(\rho\|\sigma) = \mathrm{tr}[\rho^\alpha \sigma^{1-\alpha}].$$

Concave for $\alpha \in [0, 1]$ and convex for $\alpha \in [1, 2]$ [Lieb].

Petz divergence

$$Q_\alpha(\rho\|\sigma) = \mathrm{tr}[\rho^\alpha \sigma^{1-\alpha}].$$

Concave for $\alpha \in [0,1]$ and convex for $\alpha \in [1,2]$ [Lieb].

Proof: [Ando]

1. $Q_\alpha(\rho\|\sigma) = \langle \Phi, (\rho^\alpha \otimes \bar{\sigma}^{1-\alpha})\Phi \rangle$ where $\Phi = \sum_{i=1}^d e_i \otimes e_i \in \mathbb{C}^{d^2}$.
   More generally: $\langle \Phi, (M \otimes N)\Phi \rangle = \mathrm{tr}[M\bar{N}]$

Petz divergence

$$Q_\alpha(\rho\|\sigma) \geq t \iff \exists T \in \mathbf{Herm}(d^2) : \begin{cases} (\rho \otimes I)\#_{1-\alpha}(I \otimes \bar{\sigma}) \succeq T \\ \langle \Phi, T\Phi \rangle \geq t. \end{cases}$$

Relative entropy

$$\boxed{D(\rho\|\sigma) = \mathrm{tr}[\rho(\log\rho - \log\sigma)]}$$

- We know that

$$D(\rho\|\sigma) = \lim_{h\to 0}\left(Q_{1+h}(\rho\|\sigma) - \mathrm{tr}[\rho]\right)/h$$

  Follows from $\log(x) = \lim_{h\to 0}(x^h - 1)/h$

- Yields SDP approximations of $D(\rho\|\sigma)$ with approximation quality $O(h)$.

- To follow: a different approach to get more accurate approximation, namely in $O(h^m)$ for any choice of m.

Better approximation of log

Approximation

$$\log(x) \approx \frac{x^h - 1}{h}$$

is a composition of two things:

1. $\log(x) \approx x - 1$
2. $\log(x) = \frac{1}{h}\log(x^h)$

Goal: use a rational approximation $\log(x) \approx r(x)$ instead of $\log(x) \approx x - 1$.
Then use second point to improve it

$$\log(x) \approx \frac{1}{h} r(x^h).$$

Matrix logarithm

- Integral representation of log:

$$\log(X) = \int_0^1 (X - I)(I + s(X - I))^{-1} ds$$

Matrix logarithm

- Integral representation of log:

$$\log(X) = \int_0^1 (X - I)(I + s(X - I))^{-1} ds$$

- Key fact: integrand is **operator concave** and semidefinite rep. for any fixed s (use Schur complements)

$$(X - I)(I + s(X - I))^{-1} \succeq T \quad \Leftrightarrow \quad \begin{bmatrix} I + s(X - I) & I \\ I & I - sT \end{bmatrix} \succeq 0$$

Matrix logarithm

- Integral representation of log:

$$\log(X) = \int_0^1 (X - I)(I + s(X - I))^{-1} ds$$

- Key fact: integrand is **operator concave** and semidefinite rep. for any fixed s (use Schur complements)

$$(X - I)(I + s(X - I))^{-1} \succeq T \quad \Leftrightarrow \quad \begin{bmatrix} I + s(X - I) & I \\ I & I - sT \end{bmatrix} \succeq 0$$

- Get semidefinite approximation of matrix log using quadrature:

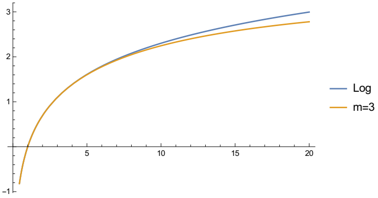
$$\log(X) \approx \sum_{j=1}^{m} w_j \frac{X - 1}{1 + s_j(X - 1)}$$

Right-hand side is semidefinite representable

Rational approximation

$$\log(x) \approx \underbrace{\sum_{j=1}^{m} w_j \frac{x-1}{1+s_j(x-1)}}_{r_m(x)}$$

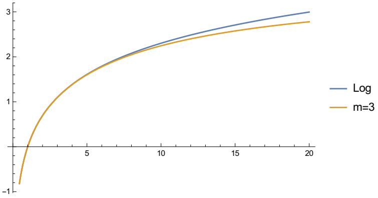
$r_m = m$'th diagonal Padé approximant of log at $x = 1$ (matches the first $2m$ Taylor coefficients).



Rational approximation

$$\log({\color{red}x}) \approx \underbrace{\sum_{j=1}^{m} w_j \frac{{\color{red}x}-1}{1+s_j({\color{red}x}-1)}}_{r_m(x)}$$

$r_m = m$'th diagonal Padé approximant of log at $x = 1$ (matches the first $2m$ Taylor coefficients).



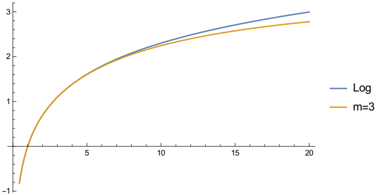
- Improve approximation by bringing x closer to 1 and using $\log(x) = \frac{1}{h}\log(x^h)$ $(0 < h < 1)$:

$$r_{m,h}(x) := \frac{1}{h} r_m(x^h)$$

Rational approximation

$$\log(x) \approx \underbrace{\sum_{j=1}^{m} w_j \frac{x-1}{1+s_j(x-1)}}_{r_m(x)}$$

$r_m = m$'th diagonal Padé approximant of log at $x = 1$ (matches the first $2m$ Taylor coefficients).



- Improve approximation by bringing x closer to 1 and using $\log(x) = \frac{1}{h}\log(x^h)$ $(0 < h < 1)$:

$$r_{m,h}(x) := \frac{1}{h} r_m(x^h)$$

- $r_{m,h}$ is still concave and semidefinite representable!

From (matrix) logarithm to (matrix) relative entropy

$$\log(X) \approx r_{m,h}(X)$$

- This allows us to approximate the *operator relative entropy*, which is the operator perspective of $-\log$.

$$\boxed{D_{op}(A\|B) = A^{1/2}\log(A^{1/2}B^{-1}A^{1/2})A^{1/2}}$$

From (matrix) logarithm to (matrix) relative entropy

$$\log(X) \approx r_{m,h}(X)$$

- This allows us to approximate the *operator relative entropy*, which is the operator perspective of $-\log$.

$$D_{op}(A\|B) = A^{1/2}\log(A^{1/2}B^{-1}A^{1/2})A^{1/2}$$

- Finally we use the fact that

$$D(\rho\|\sigma) = \langle \Phi, D_{op}(\rho \otimes I \| I \otimes \bar{\sigma})\Phi \rangle$$

where $\Phi = \sum_{i=1}^{d} e_i \otimes e_i$.

Implementation

- These formulations are implemented in the Matlab package `CVXQUAD`

  `https://www.github.com/hfawzi/cvxquad/`

- Augments Matlab's CVX [Grant, Boyd] with functions

  `quantum_rel_entr, lieb_ando, quantum_cond_entr, ...`

- Please let me know if there are bugs :)

Separable states and semidefinite hierarchies

Separable states

$$\mathrm{Sep}(n, m) = \mathrm{conv}\left\{ \ (x \otimes y)(\bar{x} \otimes \bar{y})^T \quad : x \in \mathbb{C}^n, y \in \mathbb{C}^m \ \ |x| = |y| = 1 \right\}.$$

- Convex set living in $\{\rho \in \mathbf{Herm}(nm) : \mathrm{Tr}[\rho] = 1\} \simeq \mathbb{C}^{n^2 m^2 - 1}$
- Sep = set of *non-entangled* bipartite states on $\mathbb{C}^n \otimes \mathbb{C}^m$

Linear optimization on Sep

$$\mathrm{Sep}(n, m) = \mathrm{conv}\left\{ \; (x \otimes y)(\bar{x} \otimes \bar{y})^T \quad : x \in \mathbb{C}^n, y \in \mathbb{C}^m \;\; |x| = |y| = 1 \right\}.$$

Cost vector $M \in \mathbf{Herm}(nm)$:

$$\max_{\rho \in \mathrm{Sep}(n,m)} \mathrm{tr}[M\rho]$$

Linear optimization on Sep

$$\mathrm{Sep}(n, m) = \mathrm{conv}\left\{ \ (x \otimes y)(\bar{x} \otimes \bar{y})^T \quad : x \in \mathbb{C}^n, y \in \mathbb{C}^m \ \ |x| = |y| = 1 \right\}.$$

Cost vector $M \in \mathbf{Herm}(nm)$:

$$\max_{\rho \in \mathrm{Sep}(n,m)} \mathrm{tr}[M\rho] \ = \ \max_{|x|=|y|=1} \mathrm{tr}\left[M(x \otimes y)(\bar{x} \otimes \bar{y})^T\right]$$

Linear optimization on Sep

$$\mathrm{Sep}(n,m) = \mathsf{conv}\left\{ \ (x \otimes y)(\bar{x} \otimes \bar{y})^T \ \ : x \in \mathbb{C}^n, y \in \mathbb{C}^m \ \ |x| = |y| = 1 \right\}.$$

Cost vector $M \in \textbf{Herm}(nm)$:

$$\max_{\rho \in \mathrm{Sep}(n,m)} \mathrm{tr}[M\rho] \ = \ \max_{|x|=|y|=1} \underbrace{\mathrm{tr}\left[M(x \otimes y)(\bar{x} \otimes \bar{y})^T\right]}_{\sum_{ijkl} M_{ij,kl} x_i \bar{x}_k y_j \bar{y}_l}$$

Linear optimization on Sep

$$\mathrm{Sep}(n,m) = \mathrm{conv}\left\{ \ (x \otimes y)(\bar{x} \otimes \bar{y})^T \ : x \in \mathbb{C}^n, y \in \mathbb{C}^m \ \ |x| = |y| = 1 \right\}.$$

Cost vector $M \in \mathbf{Herm}(nm)$:

$$\max_{\rho \in \mathrm{Sep}(n,m)} \mathrm{tr}[M\rho] \ = \ \max_{|x|=|y|=1} \underbrace{\mathrm{tr}\left[M(x \otimes y)(\bar{x} \otimes \bar{y})^T\right]}_{\sum_{ijkl} M_{ij,kl} x_i \bar{x}_k y_j \bar{y}_l}$$

Linear optimization on $\mathrm{Sep}(n,m)$ $\leftrightarrow$ Optimizing a (Hermitian) polynomial on a product of two spheres ($S_{\mathbb{C}^n} \times S_{\mathbb{C}^m}$).

Optimization on the sphere

Polynomial optimization on the sphere:

$$\max \quad p(x_1, \dots, x_n) \quad : \quad \sum_{i=1}^{n} x_i^2 = 1.$$

Hard in general (Nesterov).

- Stable set problem on a graph can be written as a maximizing a degree-4 polynomial on the sphere (Motzkin formulation of the stable set problem)

$$1 - \frac{1}{\alpha(G)} = \max_{x \in S^{n-1}} 2 \sum_{ij \notin E} x_i^2 x_j^2$$

- $2 \to 4$ norm of a matrix A

$$\max \quad \|Ax\|_4^4 \quad \text{s.t.} \quad \|x\|_2^2 = 1.$$

Hardness of Sep

- Deciding membership in $\mathrm{Sep}(n, m)$ is NP-hard in general [Gurvits]

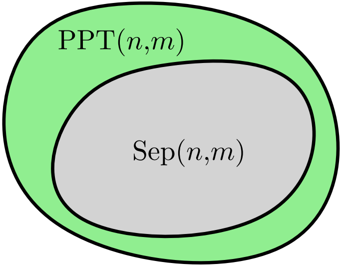
- Semidefinite relaxations

PPT relaxation (positive partial transpose)

With $\mathrm{T} : \mathbb{C}^{m\times m} \to \mathbb{C}^{m\times m}$ = transpose map, let

$$\mathrm{PPT}(n,m) = \{\rho \in \textbf{Herm}(nm) : \rho \succeq 0, \mathrm{tr}[\rho] = 1, \text{ and } (I \otimes \mathrm{T})(\rho) \geq 0\}$$

(Check that $\mathrm{Sep} \subset \mathrm{PPT}$: $(I \otimes \mathrm{T})(xx^\dagger \otimes yy^\dagger) = xx^\dagger \otimes \bar{y}\bar{y}^\dagger \succeq 0$)



Størmer–Woronowicz [60/70's]: $\mathrm{Sep}(n,m) = \mathrm{PPT}(n,m)$ iff $n + m \leq 5$

Duality

$$\mathbf{Sep}(n,m) = \mathrm{conv}\left\{\ (x\otimes y)(\bar{x}\otimes\bar{y})^T \ : x\in\mathbb{C}^n, y\in\mathbb{C}^m\right\}.$$

- Dual of $\mathbf{Sep}$:

$$\mathbf{Sep}^* \stackrel{def}{=} \{M\in\mathbf{Herm}(nm) : \mathrm{tr}[M\rho]\geq 0\ \forall\rho\in\mathbf{Sep}\}$$

Duality

$$\mathbf{Sep}(n,m) = \operatorname{conv}\left\{ \ (x\otimes y)(\bar{x}\otimes\bar{y})^T \ : x\in\mathbb{C}^n, y\in\mathbb{C}^m\right\}.$$

- Dual of $\mathbf{Sep}$:

$$\begin{aligned}\mathbf{Sep}^* &\stackrel{def}{=} \{M\in\mathbf{Herm}(nm) : \operatorname{tr}[M\rho]\geq 0\ \forall\rho\in\mathbf{Sep}\}\\ &= \{M\in\mathbf{Herm}(nm) : p_M(x,y) \text{ is nonnegative}\}\end{aligned}$$

where

$$p_M(x,y) = \sum_{ijkl} M_{ij,kl}x_i\bar{x}_k y_j\bar{y}_l.$$

Duality

$$\mathbf{Sep}(n,m) = \mathrm{conv}\left\{ \ (x \otimes y)(\bar{x} \otimes \bar{y})^T \ : x \in \mathbb{C}^n, y \in \mathbb{C}^m \right\}.$$

- Dual of $\mathbf{Sep}$:

$$\begin{aligned} \mathbf{Sep}^* &\overset{def}{=} \{M \in \mathbf{Herm}(nm) : \mathrm{tr}[M\rho] \geq 0 \ \forall \rho \in \mathbf{Sep}\} \\ &= \{M \in \mathbf{Herm}(nm) : p_M(x,y) \text{ is nonnegative}\} \end{aligned}$$

  where

$$p_M(x,y) = \sum_{ijkl} M_{ij,kl} x_i \bar{x}_k y_j \bar{y}_l.$$

- $\mathbf{PPT}^* = \{M \in \mathbf{Herm}(nm) : p_M \text{ is sum-of-squares}\}$

Duality

$$p_M(x, \bar{x}, y, \bar{y}) = \sum_{ijkl} M_{ijkl} x_i \bar{x}_k y_j \bar{y}_l$$

$$\begin{array}{ccc} \textbf{Sep} & \subset & \textbf{PPT} \\ \text{(duality)} \updownarrow & & \updownarrow \text{(duality)} \\ \text{Nonnegative polynomials } p_M & \supset & \text{Sum-of-squares polynomials } p_M \end{array}$$

Sums of squares

- A polynomial $f(z, \bar{z})$ in $z \in \mathbb{C}^d$ is Hermitian if it is real-valued.

- Hermitian polynomial $f(z, \bar{z})$ is a sum of squares if

$$f(z, \bar{z}) = \sum_i g_i(z, \bar{z})^2$$

for some Hermitian polynomials $g_i(z, \bar{z})$

Sums of squares

- A polynomial $f(z,\bar{z})$ in $z \in \mathbb{C}^d$ is Hermitian if it is real-valued.

- Hermitian polynomial $f(z,\bar{z})$ is a sum of squares if

$$f(z,\bar{z}) = \sum_i g_i(z,\bar{z})^2$$

for some Hermitian polynomials $g_i(z,\bar{z})$

- $f(z,\bar{z})$ is a (complex) sum of squares if

$$f(z,\bar{z}) = \sum_i |h_i(z)|^2$$

for some polynomials $h_i(z) \in \mathbb{C}[z]$

The two notions are different: $f(z,\bar{z}) = (z+\bar{z})^2$ is real sos but not complex sos.

Proof $\mathbf{PPT}^* \leftrightarrow$ *sos*

Outline

- Sum-of-squares hierarchy for the set of separable states

- Semidefinite lifts of $\mathrm{Sep}(n, m)$?

SDP hierarchy

We have seen

- $\mathbf{Sep}^* = \{M \in \textbf{Herm}(nm) : p_M \text{ is nonnegative}\}$.
- $\mathbf{PPT}^* = \{M \in \textbf{Herm}(nm) : p_M \text{ is sos}\}$

where

$$p_M(x, \bar{x}, y, \bar{y}) = \sum_{ijkl} M_{ijkl} x_i \bar{x}_k y_j \bar{y}_l$$

SDP hierarchy

We have seen

- $\mathbf{Sep}^* = \{M \in \textbf{Herm}(nm) : p_M \text{ is nonnegative}\}$.
- $\mathbf{PPT}^* = \{M \in \textbf{Herm}(nm) : p_M \text{ is sos}\}$

where

$$p_M(x, \bar{x}, y, \bar{y}) = \sum_{ijkl} M_{ijkl} x_i \bar{x}_k y_j \bar{y}_l$$

The *Doherty-Parrilo-Spedalieri* hierarchy is (from the dual point of view):

$$\mathbf{DPS}_\ell^* = \left\{ M \in \textbf{Herm}(nm) : |y|^{2(\ell-1)} p_M \text{ is sos} \right\}.$$

SDP hierarchy

We have seen

- $\mathbf{Sep}^* = \{M \in \textbf{Herm}(nm) : p_M \text{ is nonnegative}\}$.
- $\mathbf{PPT}^* = \{M \in \textbf{Herm}(nm) : p_M \text{ is sos}\}$

where

$$p_M(x, \bar{x}, y, \bar{y}) = \sum_{ijkl} M_{ijkl} x_i \bar{x}_k y_j \bar{y}_l$$

The *Doherty-Parrilo-Spedalieri* hierarchy is (from the dual point of view):

$$\mathbf{DPS}_\ell^* = \left\{ M \in \textbf{Herm}(nm) : |y|^{2(\ell-1)} p_M \text{ is sos} \right\}.$$

We have

$$\mathbf{Sep} \subset \mathbf{PPT} = \mathbf{DPS}_1 \subset \mathbf{DPS}_2 \subset \cdots \subset \mathbf{DPS}_\ell$$

Convergence of the hierarchy? (Note: $\mathbf{DPS}_\ell$ has an SDP representation of size $\approx d^\ell$)

Convergence

We prove a general convergence result for the sum-of-squares hierarchy on the sphere

Theorem (Fang-Fawzi)

Let $p(x_1, \ldots, x_d)$ *homogeneous polynomial such that*

$$\epsilon \leq p(x) \leq 1 \qquad \forall x \in S^{d-1}$$

Then p *is* ℓ*-sos with* $\ell \gtrsim d/\sqrt{\epsilon}$*.*

- Best previous result gives convergence in d/ϵ instead of $d/\sqrt{\epsilon}$ [Reznick 95, Doherty-Wehner 12]

Convergence

We prove a general convergence result for the sum-of-squares hierarchy on the sphere

Theorem (Fang-Fawzi)

Let $p(x_1, \ldots, x_d)$ homogeneous polynomial such that

$$\epsilon \leq p(x) \leq 1 \qquad \forall x \in S^{d-1}$$

Then p is ℓ-sos with $\ell \gtrsim d/\sqrt{\epsilon}$.

- Best previous result gives convergence in d/ϵ instead of $d/\sqrt{\epsilon}$ [Reznick 95, Doherty-Wehner 12]
- Applied to the DPS hierarchy we get that, for $M \succeq 0$

$$h_{\mathrm{Sep}}(M) \leq h_{\mathrm{DPS}_\ell}(M) \leq \left(1 + c\left(\frac{d}{\ell}\right)^2\right) h_{\mathrm{Sep}}(M)$$

where $h_C(M) = \max_{\rho \in C} \mathrm{tr}[M\rho]$. Recovers result of Navascués-Owari-Plenio [2009] based on quantum information tools.

Overview of proof

- Let $p(x)$ real polynomial such that $0 < \epsilon \leq p \leq 1$ on S^{n-1}.

  Goal: write p as a sum of squares

- Define integral transform

$$(Kp)(x) = \int_{y \in S^{d-1}} \phi\left(x^T y\right) p(y) d\sigma(y)$$

  where $\phi : [-1, 1] \to \mathbb{R}$ is a univariate function.

- Observations:
  - If $\phi(t) = \delta(t-1)$ [Dirac delta] then $Kp = p$.
  - If $\phi(t) = h(t)^2$ where $\deg h \leq \ell$ then Kp is an (integral) sum of squares of degree ℓ polynomials.

- Strategy of proof: we write $p = K(K^{-1}p)$. If $p \geq \epsilon > 0$, and if $K \approx$ identity then we hope that $K^{-1}p \geq 0$. In this case $p = K(K^{-1}p)$ is a sum of squares.

Fourier analysis

$$(Kp)(x) = \int_{y \in S^{d-1}} \phi\left(x^T y\right) p(y) d\sigma(y)$$

- Need to understand how close K is to the identity operator
- Fourier analysis: there is a decomposition $\mathbb{R}[x] = H_0 \oplus H_1 \oplus H_2 \oplus \dots$ (*harmonic spaces*) such that if $p = p_0 + p_1 + \dots$ then

$$Kp = \lambda_0 p_0 + \lambda_1 p_1 + \dots$$

where (λ_i) are coefficients of ϕ in a basis of Gegenbauer polynomials.

$$\|K^{-1}p - p\|_\infty = \|\sum_i (\lambda_i^{-1} - 1) p_i\|_\infty \leq \|p\|_\infty C \sum_i |\lambda_i^{-1} - 1|.$$

Continued

To recap: if we can find $\phi(t) = h(t)^2$ with $\deg h \leq \ell$ such that $C \sum_i |\lambda_i^{-1} - 1| \leq \epsilon$, then $K^{-1}p \geq 0$, and thus $p = K(K^{-1}p)$ is ℓ-sos on the sphere.

- Some analysis with orthogonal polynomials tells us that we can find such a $\phi(t) = h(t)^2$ with $\ell \approx d/\sqrt{\epsilon}$.

- Idea not new, rediscovered many times: Reznick, Doherty-Wehner, Parrilo. Main differences is choice of kernel/Fourier analysis part

Semidefinite representations of Sep?

Semidefinite programming lifts

Spectrahedra A *spectrahedron* is a convex set of the form

$$S = \left\{ w \in \mathbb{R}^d : \mathcal{A}(w) \succeq 0 \right\}$$

where $\mathcal{A} : \mathbb{R}^d \to \mathbf{Herm}(N)$ is a linear map.

$\mathrm{PPT}(n, m)$ is a spectrahedron where $\mathcal{A}(\rho) = \left[\begin{smallmatrix} \rho & 0 \\ 0 & (I \otimes \mathsf{T})(\rho) \end{smallmatrix}\right]$

Semidefinite programming lifts

Spectrahedra A *spectrahedron* is a convex set of the form

$$S = \{w \in \mathbb{R}^d : \mathcal{A}(w) \succeq 0\}$$

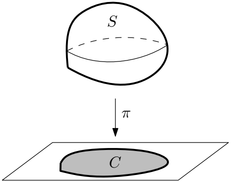
where $\mathcal{A} : \mathbb{R}^d \to \mathbf{Herm}(N)$ is a linear map.

$\mathrm{PPT}(n, m)$ is a spectrahedron where $\mathcal{A}(\rho) = \begin{bmatrix} \rho & 0 \\ 0 & (I \otimes \mathsf{T})(\rho) \end{bmatrix}$

Projection We say that C has an *SDP lift* of size N if

$$C = \pi(S)$$

where S is a spectrahedron as above, and π is a linear map.
More expressive!



Lifting

- Can show that
$$C = \{(x, y) : x^4 + y^4 \leq 1\}$$
has a SDP representation, but is *not* a spectrahedron.

Lifting

- Can show that
$$C = \{(x, y) : x^4 + y^4 \le 1\}$$
has a SDP representation, but is *not* a spectrahedron.

- If C has a SDP representation, then optimizing a linear function on C is a semidefinite program:
$$\min_{x \in C} \ell(x) = \min_{\mathcal{A}(y) \succeq 0} \ell \circ \pi(y).$$

Lifting

- Can show that

$$C = \{(x, y) : x^4 + y^4 \leq 1\}$$

has a SDP representation, but is *not* a spectrahedron.

- If C has a SDP representation, then optimizing a linear function on C is a semidefinite program:

$$\min_{x \in C} \ell(x) = \min_{\mathcal{A}(y) \succeq 0} \ell \circ \pi(y).$$

- Lifting can be very helpful from a complexity point of view

Other lifting examples

- Permutahedron

$$\mathrm{conv}\,\{(\sigma(1), \ldots, \sigma(n)) : \sigma \in S_n\}$$

has $n!$ vertices and $\sim 2^n$ facets. Can express it as the projection of the convex polytope of doubly stochastic matrices

$$DS_n = \{M \in \mathbb{R}^{n \times n} : M_{ij} \geq 0\ \forall ij \text{ and } M\mathbf{1} = \mathbf{1}^T M = 1\}$$

Other lifting examples

- Permutahedron

$$\operatorname{conv}\{(\sigma(1),\ldots,\sigma(n)) : \sigma \in S_n\}$$

has $n!$ vertices and $\sim 2^n$ facets. Can express it as the projection of the convex polytope of doubly stochastic matrices

$$DS_n = \{M \in \mathbb{R}^{n\times n} : M_{ij} \geq 0 \; \forall ij \text{ and } M\mathbf{1} = \mathbf{1}^T M = 1\}$$

- For perfect graphs Lovász showed

$$STAB(G) = \left\{ x \in \mathbb{R}^n : \exists X \text{ s.t. } \begin{bmatrix} 1 & x^T \\ x & X \end{bmatrix} \succeq 0,\; X_{ii} = x_i, X_{ij} = 0 \; \forall ij \in E \right\}.$$

SDP lifts

Which convex sets C have an SDP lift? A necessary condition is that C is *semialgebraic* (Tarski)

Semialgebraic geometry

- A set is *semialgebraic* if it is a boolean combination (union, intersection, complement) of sets defined using polynomials equalities and inequalities
- Tarski's quantifier elimination (1940s): the projection of any semialgebraic set is semialgebraic

SDP lifts

Which convex sets C have an SDP lift? A necessary condition is that C is *semialgebraic* (Tarski)

Semialgebraic geometry

- A set is *semialgebraic* if it is a boolean combination (union, intersection, complement) of sets defined using polynomials equalities and inequalities
- Tarski's quantifier elimination (1940s): the projection of any semialgebraic set is semialgebraic

- Nemirovski (ICM 2006): does any convex *semialgebraic* set C have a semidefinite lift?

SDP lifts

Which convex sets C have an SDP lift? A necessary condition is that C is *semialgebraic* (Tarski)

Semialgebraic geometry

- A set is *semialgebraic* if it is a boolean combination (union, intersection, complement) of sets defined using polynomials equalities and inequalities
- Tarski's quantifier elimination (1940s): the projection of any semialgebraic set is semialgebraic

- Nemirovski (ICM 2006): does any convex *semialgebraic* set C have a semidefinite lift?
- Helton-Nie (2009): if boundary of C is smooth with positive curvature then it has SDP lift. They conjectured a positive answer to Nemirovski.

SDP lifts

Which convex sets C have an SDP lift? A necessary condition is that C is *semialgebraic* (Tarski)

Semialgebraic geometry

- A set is *semialgebraic* if it is a boolean combination (union, intersection, complement) of sets defined using polynomials equalities and inequalities
- Tarski's quantifier elimination (1940s): the projection of any semialgebraic set is semialgebraic

- Nemirovski (ICM 2006): does any convex *semialgebraic* set C have a semidefinite lift?
- Helton-Nie (2009): if boundary of C is smooth with positive curvature then it has SDP lift. They conjectured a positive answer to Nemirovski.
- Scheiderer (2012): convex semialgebraic sets in the plane have SDP lift

SDP lifts

Which convex sets C have an SDP lift? A necessary condition is that C is *semialgebraic* (Tarski)

Semialgebraic geometry

- A set is *semialgebraic* if it is a boolean combination (union, intersection, complement) of sets defined using polynomials equalities and inequalities
- Tarski's quantifier elimination (1940s): the projection of any semialgebraic set is semialgebraic

- Nemirovski (ICM 2006): does any convex *semialgebraic* set C have a semidefinite lift?
- Helton-Nie (2009): if boundary of C is smooth with positive curvature then it has SDP lift. They conjectured a positive answer to Nemirovski.
- Scheiderer (2012): convex semialgebraic sets in the plane have SDP lift
- Scheiderer (2016): there are (many) convex semialgebraic sets that do not have an SDP representation

No exact SDP representations exist for Sep in general

Theorem (Fawzi)

If $Sep(n,m) \neq PPT(n,m)$ then $Sep(n,m)$ has no SDP lift. In other words, $Sep(3,3)$ and $Sep(4,2)$ have no SDP lift.

- Horodecki's formulation of $\mathrm{Sep}(n,m)$:

$$\mathrm{Sep}(n,m) = \{\rho \in \mathbf{Herm}(nm) : (I \otimes \Phi)(\rho) \geq 0 \;\; \forall \Phi : \mathrm{M}_m \to \mathrm{M}_n \text{ positive}\}.$$

  Skowronek (2016) showed that for $\mathrm{Sep}(3,3)$ it is not possible to reduce the quantifier $\forall\Phi$ to a finite number of maps $\Phi_1, \ldots, \Phi_k$.

- Result also includes as a special that the DPS (Doherty-Parrilo-Spedalieri) hierarchy does not converge in a finite number of levels when $n + m > 5$.

Conclusion

- Use tools from semidefinite programming, polynomial optimization and sums of squares to study problems in quantum information.

- Things I did not talk about: noncommutative polynomial optimization (NPA hierarchy), tomography problems and compressed sensing, ...

Thank you!

Proof Sep has no semidefinite lift

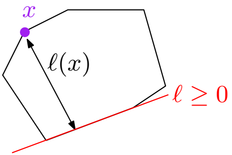
Semidefinite lifts

- Definition of SDP lift is hard to work with. Need a more algebraic way of thinking about it

- Given convex cone $C = \text{conv}(X)$, we associate a *slack matrix* $\boldsymbol{S}$ (potentially infinite) defined as follows:

$$\boldsymbol{S}(x, \ell) = \ell(x) \geq 0 \qquad \forall x \in X, \ell \in C^*$$

- If C polytope, then slack matrix $\boldsymbol{S}$ has size $\#$*vertices* $\times$ $\#$*facets*



Factorization theorem

SDP lift of C $\quad\Leftrightarrow\quad$ Factorization of $\boldsymbol{S}$

Factorization theorem

SDP lift of $C \quad \Leftrightarrow \quad$ Factorization of $\boldsymbol{S}$

Theorem (Gouveia, Parrilo, Thomas)

$C = \mathrm{conv}(X)$ *has an SDP lift of size* N *iff one can find maps* $A : X \to \mathbf{Herm}_+^N$ *and* $B : C^* \to \mathbf{Herm}_+^N$ *such that we have the factorization*

$$\boldsymbol{S}(x, \ell) = \mathrm{Tr}[A(x)B(\ell)] \qquad \forall x \in X, \ell \in C^*$$

Generalizes a result of Yannakakis 1991 (LPs) to SDPs

SDP lifts and sums of squares

A corollary of the previous theorem is

Theorem

Assume $C = \mathrm{conv}(X)$ has an SDP lift of size N. Then there is a subspace $\mathcal{V}$ of functions on X of dimension at most N^2 s.t. for any $\ell \in C^*$

$$\ell|_X = \sum_k h_k^2 \qquad \textit{where} \quad h_k \in \mathcal{V}.$$

Proof: write
$\ell(x) = \mathrm{tr}[A(x)B(\ell)] = \mathrm{tr}[F(x)F(x)^T G(\ell)G(\ell)^T] = \|F(x)^T G(\ell)\|_F^2 = \sum_k h_k(x)^2$

General result in the real case

Theorem (Main, real case)

Let $p \in \mathbb{R}[\boldsymbol{x}]$ be a nonnegative polynomial that is not sos. Let

$$A = \{\alpha \in \mathbb{N}^n : \alpha \leq \beta \text{ for some } \beta \in \operatorname{supp}(p)\}$$

be the "staircase" under $\operatorname{supp}(p)$. *Then*

$$C_A = \operatorname{conv}\{(\boldsymbol{x}^\alpha)_{\alpha \in A} : \boldsymbol{x} \in \mathbb{R}^n\}$$

has no semidefinite representation.

Proof

$$C_A = \text{conv}\left\{(x^\alpha)_{\alpha\in A} : x \in \mathbb{R}^n\right\}$$

- Linear functions nonnegative on $C_A \leftrightarrow$ nonnegative polynomials supported on A

Characterization of SDP lifts using sum-of-squares:

Theorem

C_A has an SDP representation iff there are functions $f_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, k)$ such that any nonnegative polynomial supported on A can be written as a sum of squares of functions from $\text{span}(f_1, \ldots, f_k)$.

Proof

$$C_A = \operatorname{conv}\{(x^\alpha)_{\alpha\in A} : x \in \mathbb{R}^n\}$$

- Linear functions nonnegative on $C_A \leftrightarrow$ nonnegative polynomials supported on A

Characterization of SDP lifts using sum-of-squares:

Theorem

C_A has an SDP representation iff there are semialgebraic functions $f_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, k)$ such that any nonnegative polynomial supported on A can be written as a sum of squares of functions from $\operatorname{span}(f_1, \ldots, f_k)$.

Proof

$$C_A = \operatorname{conv}\{(x^\alpha)_{\alpha\in A} : x \in \mathbb{R}^n\}$$

- Linear functions nonnegative on $C_A \leftrightarrow$ nonnegative polynomials supported on A

Characterization of SDP lifts using sum-of-squares:

Theorem

C_A has an SDP representation iff there are semialgebraic *functions $f_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \dots, k)$ such that any nonnegative polynomial supported on A can be written as a sum of squares of functions from* $\operatorname{span}(f_1, \dots, f_k)$.

- $f : \mathbb{R}^n \to \mathbb{R}$ is *semialgebraic* if its graph $\{(x, f(x)) : x \in \mathbb{R}^n\}$ is a semialgebraic subset of $\mathbb{R}^{n+1}$

Proof

$$C_A = \text{conv}\,\{(x^\alpha)_{\alpha \in A} : x \in \mathbb{R}^n\}$$

- Linear functions nonnegative on $C_A \leftrightarrow$ nonnegative polynomials supported on A

Characterization of SDP lifts using sum-of-squares:

Theorem

C_A has an SDP representation iff there are semialgebraic functions $f_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, k)$ such that any nonnegative polynomial supported on A can be written as a sum of squares of functions from $\text{span}(f_1, \ldots, f_k)$.

- $f : \mathbb{R}^n \to \mathbb{R}$ is *semialgebraic* if its graph $\{(x, f(x)) : x \in \mathbb{R}^n\}$ is a semialgebraic subset of $\mathbb{R}^{n+1}$
- **Semialgebraic functions are tame:** They are smooth (C^∞) almost everywhere (except on a set of measure 0)

Proof of main theorem

p nonnegative polynomial not sos, $A =$ staircase under supp(p)

$$C_A = \text{conv}\left\{(x^\alpha)_{\alpha \in A} : x \in \mathbb{R}^n\right\}.$$

- Assume C_A has an SDP representation, and let $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ be the semialgebraic functions associated to this representation
- Since the $(f_i)_{i=1,\ldots,k}$ are smooth almost everywhere, there is a point $a \in \mathbb{R}^n$ such that the f_i are all smooth at a
- Since A is the staircase under support(p), the polynomial $p(x+a)$ is supported on A, and since it is nonnegative, it must be a sum-of-squares from span$(f_1, \ldots, f_k)$. Shifting by a, this means that p is a sum of squares from span$(\tilde{f}_1, \ldots, \tilde{f}_k)$ where $\tilde{f}_i(x) = f_i(x-a)$

Smooth sums of squares

Proposition

Assume p is a homogeneous polynomial such that $p = \sum_j f_j^2$ for some arbitrary functions f_j that are C^∞ at the origin. Then p is a sum of squares of polynomials.

Proof: Taylor expansion

Smooth sums of squares

Proposition

Assume p is a homogeneous polynomial such that $p = \sum_j f_j^2$ for some arbitrary functions f_j that are C^∞ at the origin. Then p is a sum of squares of polynomials.

Proof: Taylor expansion

- Proves theorem when p is homogeneous

Smooth sums of squares

Proposition

Assume p is a homogeneous polynomial such that $p = \sum_j f_j^2$ for some arbitrary functions f_j that are C^∞ at the origin. Then p is a sum of squares of polynomials.

Proof: Taylor expansion

- Proves theorem when p is homogeneous
- Additional technical argument based on Puiseux expansions is needed for general p

Main result, complex case

Theorem (Main, complex case)

Let p be a nonnegative Hermitian polynomial that is not sos. Let

$$A = \{(\alpha, \alpha') \in \mathbb{N}^n \times \mathbb{N}^n : (\alpha, \alpha') \leq (\beta, \beta'),\ \textit{for some } (\beta, \beta') \in \mathrm{supp}(p)\}$$

be the "staircase" under $\mathrm{supp}(p)$*. Then*

$$C_A = \mathrm{conv}\left\{(z^\alpha \bar{z}^{\alpha'})_{(\alpha,\alpha')\in A} : z \in \mathbb{C}^n\right\}$$

has no semidefinite representation.

Main result, complex case

Theorem (Main, complex case)

Let p be a nonnegative Hermitian polynomial that is not sos. Let

$$A = \{(\alpha,\alpha') \in \mathbb{N}^n \times \mathbb{N}^n : (\alpha,\alpha') \leq (\beta,\beta'),\ \textit{for some}\ (\beta,\beta') \in \mathrm{supp}(p)\}$$

be the "staircase" under $\mathrm{supp}(p)$*. Then*

$$C_A = \mathrm{conv}\left\{(z^\alpha \bar{z}^{\alpha'})_{(\alpha,\alpha')\in A} : z \in \mathbb{C}^n\right\}$$

has no semidefinite representation.

- If $\mathrm{Sep}(n,m) \neq \mathrm{PPT}(n,m)$, apply theorem above with $p =$ (dehomogenized) nonnegative Hermitian biquadratic on (n,m) variables that is not sos
- For $\mathrm{Sep}(3,3)$ use the Choi polynomial. For $\mathrm{Sep}(4,2)$ use a polynomial exhibited by Woronowicz and further studied by Ha and Kye.