Polyhedral approximations of the positive semidefinite cone

Hamza Fawzi

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

Tilburg, 2019

Conic programming

$$\begin{array}{ll} \text{min.} & \langle c, x \rangle \\ \text{s.t.} & A(x) = b, \ \ x \in K \end{array}$$

K is a convex cone, and A linear map

- Three major cones
  - Linear programming (LP): $K = \mathbb{R}^m_+$
  - Second-order cone programming (SOCP): $K = \mathcal{Q}^{m_1} \times \cdots \times \mathcal{Q}^{m_k}$ where

    $$\mathcal{Q}^m = \{(x, t) \in \mathbb{R}^m \times \mathbb{R} : \|x\|_2 \leq t\}.$$

  - Semidefinite programming (SDP): $K = \mathbf{S}^m_+$ ($m \times m$ symmetric positive semidefinite matrices)

Reduction between conic programs

It is known that

$$\text{“}LP \subsetneq SOCP \subsetneq SDP\text{”}$$

- $LP \neq SOCP, SDP$ because $\mathcal{Q}^m$ and $\mathbf{S}_+^m$ are nonpolyhedral
- Not possible to express $\mathbf{S}_+^3$ using second-order cones! [F17]

Reduction between conic programs

It is known that

$$\text{“}LP \subsetneq SOCP \subsetneq SDP\text{”}$$

- $LP \neq SOCP, SDP$ because $\mathcal{Q}^m$ and $\mathbf{S}_+^m$ are nonpolyhedral
- Not possible to express $\mathbf{S}_+^3$ using second-order cones! [F17]

What about approximations?

- Ben-Tal & Nemirovski: $\mathcal{Q}^m$ can be ϵ-approximated using an LP of size $O(1)m\log(1/\epsilon)$. More precisely:

Theorem (Ben-Tal & Nemirovski, 2001)

For any $\epsilon > 0$, there is a polytope P with extension complexity $\leq O(1)m\log(1/\epsilon)$ such that $(1-\epsilon)B^m \subset P \subset B^m$, where B^m is the unit Euclidean ball in $\mathbb{R}^m$

Reduction between conic programs

It is known that

$$\text{“}LP \subsetneq SOCP \subsetneq SDP\text{”}$$

- $LP \neq SOCP, SDP$ because $\mathcal{Q}^m$ and $\mathbf{S}_+^m$ are nonpolyhedral
- Not possible to express $\mathbf{S}_+^3$ using second-order cones! [F17]

What about approximations?

- Ben-Tal & Nemirovski: $\mathcal{Q}^m$ can be ϵ-approximated using an LP of size $O(1)m\log(1/\epsilon)$. More precisely:

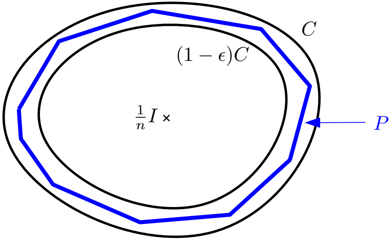
Theorem (Ben-Tal & Nemirovski, 2001)

For any $\epsilon > 0$, there is a polytope P with extension complexity $\leq O(1)m\log(1/\epsilon)$ such that $(1-\epsilon)B^m \subset P \subset B^m$, where B^m is the unit Euclidean ball in $\mathbb{R}^m$

- What about the PSD cone?

$$
\begin{aligned}
C &= \text{convex set of } n \times n \text{ real symmetric} \\
&\qquad \text{positive semidefinite matrices of trace 1} \\
&= \operatorname{conv}\left\{xx^T : x \in \mathbb{R}^n, \|x\|_2 = 1\right\}.
\end{aligned}
$$

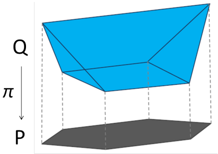
Question: How well can we approximate C using polyhedra?



Extension complexity of polytopes

A polytope P has extension complexity N if it can be written as

$$P = \pi(Q) \quad \text{where} \quad \begin{cases} Q = \{x \in \mathbb{R}^N : x \geq 0 \text{ and } Ax = b\} \\ \pi \text{ linear (projection) map} \end{cases}$$



Note:

- $\mathrm{xc}(P) \leq \#\textit{vertices}(P)$ and $\mathrm{xc}(P) \leq \#\textit{facets}(P)$.
- Also $\mathrm{xc}(P) = \mathrm{xc}(P^*)$ (invariant under duality)

Main theorem

Theorem (Fawzi)

Assume P is a polytope such that $\frac{2}{3}C \subset P \subset C$. Then $\mathrm{xc}(P) \geq \exp(c\sqrt{n})$ *where $c > 0$ is an absolute constant.*

Main theorem

Theorem (Fawzi)

Assume P is a polytope such that $\frac{2}{3}C \subset P \subset C$. Then $\mathrm{xc}(P) \geq \exp(c\sqrt{n})$ where $c > 0$ is an absolute constant.

Comments/previous work:

- Best construction known has size $\exp(cn)$ [Aubrun & Szarek] $\rightarrow$ (random) discretization of the sphere

Main theorem

Theorem (Fawzi)

Assume P is a polytope such that $\frac{2}{3}C \subset P \subset C$. Then $\mathrm{xc}(P) \geq \exp(c\sqrt{n})$ where $c > 0$ is an absolute constant.

Comments/previous work:

- Best construction known has size $\exp(cn)$ [Aubrun & Szarek] $\rightarrow$ (random) discretization of the sphere
- Braun, Fiorini, Pokutta, Steurer 2013: There exists a spectrahedral shadow S in $\mathbb{R}^{n^2}$ such that $\mathrm{rank}_{psd}(S) \leq n+1$ and whenever $S \subset P \subset S + \epsilon B_1$ then $\mathrm{xc}(P) \geq e^{cn}$, where B_1 is the ℓ_1 ball.

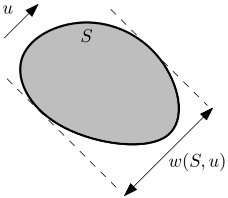
Gaussian width

Given a direction u, the *width* of a set S in this direction is:

$$w(S,u) = \max_{x \in S} \langle u, x \rangle - \min_{x \in S} \langle u, x \rangle.$$

The *Gaussian width* of S is

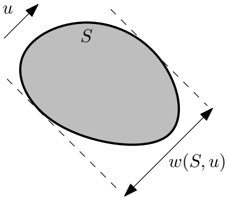
$$w_G(S) = \mathbb{E}_{u \sim N(0,1)}[w(S,u)].$$



Gaussian width

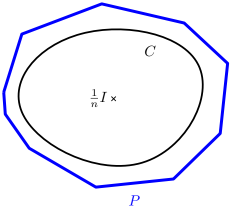
Given a direction u, the *width* of a set S in this direction is:

$$w(S, u) = \max_{x \in S} \langle u, x \rangle - \min_{x \in S} \langle u, x \rangle.$$



The *Gaussian width* of S is

$$w_G(S) = \mathbb{E}_{u \sim N(0,1)}[w(S, u)].$$

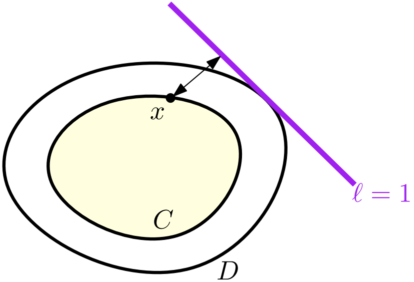


Theorem

Assume P is a polytope such that $C \subset P$ and $w_G(P) \leq 2 w_G(C)$. Then $\mathrm{xc}(P) \geq e^{cn^{1/3}}$.

Extension complexity: Yannakakis theorem

Controlling extension complexity is much more difficult than # vertices. Very useful tool is Yannakakis theorem...
Let $C \subset D$ be two nested convex sets.



- Slack matrix of D wrt C is an (infinite) matrix S

$$S[x, \ell] = 1 - \ell(x)$$

  where
  - $x =$ extreme point of C
  - $\ell =$ extreme point of D° (polar)

Yannakakis theorem

Theorem (Yannakakis 1991)

Let $C \subset D$ compact convex sets. There exists a polytope P with $\mathrm{xc}(P) = N$ s.t. $C \subset P \subset D$ iff the slack matrix S has a nonnegative factorization of size N.

Nonnegative factorization:

$$S[x, \ell] = \sum_{i=1}^{N} a_i(x) b_i(\ell)$$

where $a_i(x), b_i(\ell) \geq 0$.

Our slack matrix

$$C = \text{conv}\left\{xx^T : \|x\|_2 = 1\right\}$$

- Can easily show that slack matrix of C wrt $(1-\epsilon)C$ is

$$S[x,y] = (1-\epsilon)n(x^T y)^2 + \epsilon \quad \forall x, y \in S^{n-1}.$$

- A nonnegative factorization of S of size N:

$$S[x,y] = \sum_{i=1}^{N} f_i(x) g_i(y) \quad \forall x, y \in S^{n-1}.$$

  where $f_i, g_i \geq 0$

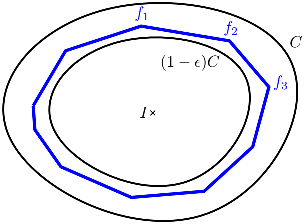
- Useful normalization: $\int_{S^{n-1}} f_i = 1$ and $\sum_{i=1}^{N} g_i \equiv 1$. ($S$ is a "column-stochastic" matrix)

Interpretation

$$\underbrace{(1-\epsilon)n(x^T y)^2 + \epsilon}_{\tilde{Q}_y(x)} \quad = \quad \sum_{i=1}^{N} f_i(x) g_i(y) \quad \forall x, y \in S^{n-1}$$

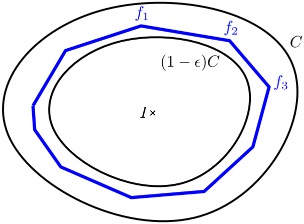
- $\tilde{Q}_y$ is a quadratic form with $\tilde{Q}_y(x) \geq \epsilon$ for $x \in S^{n-1}$ and $\int \tilde{Q}_y(x) dx = 1$
- Interpretation of nonnegative factorization: All $\tilde{Q}_y$'s are convex combinations of the functions $\{f_1, \ldots, f_N\}$.
- To prove a lower bound, we need to show that we need many functions f_i's to cover all the $\tilde{Q}_y$'s ($y \in S^{n-1}$)
- Important: the functions f_i need not be quadratic!

Warmup: when f_i are quadratics



Assume f_i are quadratic s.t. $\tilde{Q}_y \in \mathrm{conv}(f_1, \ldots, f_N)$ for all $y \in S^{n-1}$. Then $N \geq \exp(cn)$.

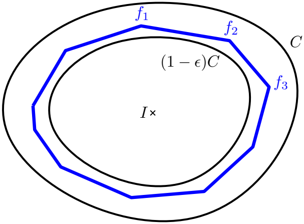
Warmup: when f_i are quadratics



Assume f_i are quadratic s.t. $\tilde{Q}_y \in \mathrm{conv}(f_1, \ldots, f_N)$ for all $y \in S^{n-1}$. Then $N \geq \exp(cn)$.

- Can assume that $f_i(x) = n(a_i^T x)^2$ for some $a_i \in S^{n-1}$

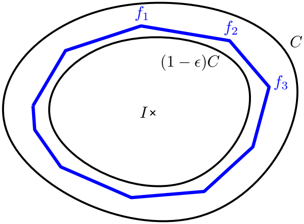
Warmup: when f_i are quadratics



Assume f_i are quadratic s.t. $\tilde{Q}_y \in \mathrm{conv}(f_1, \ldots, f_N)$ for all $y \in S^{n-1}$. Then $N \geq \exp(cn)$.

- Can assume that $f_i(x) = n(a_i^T x)^2$ for some $a_i \in S^{n-1}$
- Evaluate at $x = y$: $\tilde{Q}_y(y) = (1-\epsilon)n + \epsilon \in n\,\mathrm{conv}((a_1^T y)^2, \ldots, (a_N^T y)^2)$.

Warmup: when f_i are quadratics



Assume f_i are quadratic s.t. $\tilde{Q}_y \in \text{conv}(f_1, \ldots, f_N)$ for all $y \in S^{n-1}$. Then $N \geq \exp(cn)$.

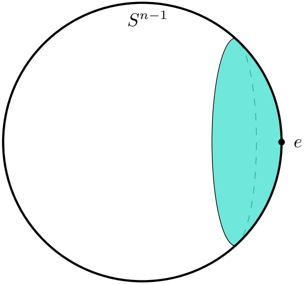
- Can assume that $f_i(x) = n(a_i^T x)^2$ for some $a_i \in S^{n-1}$
- Evaluate at $x = y$: $\tilde{Q}_y(y) = (1-\epsilon)n + \epsilon \in n\,\text{conv}((a_1^T y)^2, \ldots, (a_N^T y)^2)$.
- Implies: for any $y \in S^{n-1}$, there exists $i \in \{1, \ldots, N\}$ s.t. $(a_i^T y)^2 \geq 1 - \epsilon = 2/3$
- Hence $N \geq \exp(cn)$

Discretizing the sphere S^{n-1}

Spherical cap around $a \in S^{n-1}$:

$$\{x \in S^{n-1} : \langle a, x \rangle \geq 2/3\}.$$

Surface area of this spherical cap is $\leq e^{-cn}$!



Need exponentially many such spherical caps to cover sphere!

Functions on S^{n-1}

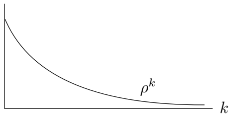
- Fourier decomposition on S^{n-1}: any function $f : S^{n-1} \to \mathbb{R}$ has a Fourier decomposition
$$f = Y_0 + Y_1 + Y_2 + \cdots = \sum_{k=0}^{\infty} Y_k$$
(similar to Fourier decomposition on the cube)

- The Y_k are *harmonic polynomials* of degree k.

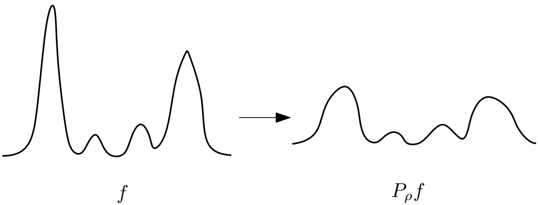
A "low-pass" filter

If $f = \sum_{k=0}^{\infty} Y_k$ we can apply a "low-pass filter" (smoothing operation) to f:

$$P_\rho f = \sum_{k=0}^{\infty} \rho^k Y_k$$



where $0 \leq \rho < 1$ (Poisson kernel).



We want to quantify how much P_ρ *flattens* functions.

L^p norms

- Given $f : S^{n-1} \to \mathbb{R}$ and $p \geq 1$ define

$$\|f\|_p = \int_{S^{n-1}} |f|^p d\sigma.$$

Fractional L^p norms quantify how flat a function f

L^p norms

- Given $f : S^{n-1} \to \mathbb{R}$ and $p \geq 1$ define

$$\|f\|_p = \int_{S^{n-1}} |f|^p d\sigma.$$

Fractional L^p norms quantify how flat a function f

- Hölder: $\|f\|_1 \leq \|f\|_2 \leq \cdots \leq \|f\|_\infty$.
- Ratio $\frac{\|f\|_q}{\|f\|_p}$ for $q > p$ tells us how flat/spiked f is.

L^p norms

- Given $f : S^{n-1} \to \mathbb{R}$ and $p \geq 1$ define

$$\|f\|_p = \int_{S^{n-1}} |f|^p d\sigma.$$

Fractional L^p norms quantify how flat a function f

- Hölder: $\|f\|_1 \leq \|f\|_2 \leq \cdots \leq \|f\|_\infty$.
- Ratio $\frac{\|f\|_q}{\|f\|_p}$ for $q > p$ tells us how flat/spiked f is.

Example: If f poly of degree d then $\|f\|_p \leq (p-1)^{d/2}\|f\|_2$ for $p \geq 2$

Properties of low-pass filter P_ρ

$$P_\rho f = \sum_{k=0}^{\infty} \rho^k Y_k$$

Contractivity: $P_\rho f$ is no more spiked than f:

$$\|P_\rho f\|_p \leq \|f\|_p, \text{ for all } p \geq 1.$$

Properties of low-pass filter P_ρ

$$P_\rho f = \sum_{k=0}^{\infty} \rho^k Y_k$$

Contractivity: $P_\rho f$ is no more spiked than f:

$$\|P_\rho f\|_p \leq \|f\|_p, \text{ for all } p \geq 1.$$

Theorem (Hypercontractivity of Poisson kernel (Beckner))

For any $p \geq 1$, $\|P_\rho f\|_q \leq \|f\|_p$ for $q = 1 + \rho^{-2}(p-1) > p$.

Properties of low-pass filter P_ρ

$$P_\rho f = \sum_{k=0}^{\infty} \rho^k Y_k$$

Contractivity: $P_\rho f$ is no more spiked than f:

$$\|P_\rho f\|_p \leq \|f\|_p, \text{ for all } p \geq 1.$$

Theorem (Hypercontractivity of Poisson kernel (Beckner))

For any $p \geq 1$, $\|P_\rho f\|_q \leq \|f\|_p$ for $q = 1 + \rho^{-2}(p-1) > p$.

- Hypercontractivity: Nelson 1960s for a smoothing kernel in Gaussian space.
- Gross 1975: logarithmic Sobolev inequalities
- Generalization to other smoothing kernels. Applications in computer science. [de Wolf, O'Donnell, Klartag-Regev].

A lemma

Lemma

Let $f : S^{n-1} \to \mathbb{R}$ *s.t.* $f \geq 0$, $\int f = 1$ *and* $\max f \leq e^{\sqrt{n}}$. *Let* $\rho = \sqrt{5/n}$ *(smoothing kernel parameter). Then*

$$\sigma\left\{P_\rho f \geq 4\right\} \leq c^{-\sqrt{n}}$$

for some absolute constant $c > 1$.

Proof.

Markov's inequality + hypercontractivity with good choices of q, p. □

Proof of lower bound

- Assume that $f_1, \ldots, f_N : S^{n-1} \to \mathbb{R}_+$, $\int f_i = 1$ are such that:

$$\forall e \in S^{n-1},\ (1-\epsilon) Q_e + \epsilon \in \mathrm{conv}(f_1, \ldots, f_N)$$

where $Q_e(x) = n(e^T x)^2$ [extreme rays of the psd cone].

Proof of lower bound

- Assume that $f_1, \ldots, f_N : S^{n-1} \to \mathbb{R}_+$, $\int f_i = 1$ are such that:

$$\forall e \in S^{n-1},\ (1-\epsilon) Q_e + \epsilon \in \operatorname{conv}(f_1, \ldots, f_N)$$

  where $Q_e(x) = n(e^T x)^2$ [extreme rays of the psd cone].

- Apply P_ρ on both sides (simple calculation $P_\rho Q_e = 1 + \rho^2(Q_e - 1)$)

$$\forall e \in S^{n-1},\ (1-\epsilon)(1 + \rho^2(Q_e - 1)) + \epsilon \in \operatorname{conv}(P_\rho f_1, \ldots, P_\rho f_N)$$

Proof of lower bound

- Assume that $f_1, \ldots, f_N : S^{n-1} \to \mathbb{R}_+$, $\int f_i = 1$ are such that:

$$\forall e \in S^{n-1},\ (1-\epsilon) Q_e + \epsilon \in \mathrm{conv}(f_1, \ldots, f_N)$$

  where $Q_e(x) = n(e^T x)^2$ [extreme rays of the psd cone].

- Apply P_ρ on both sides (simple calculation $P_\rho Q_e = 1 + \rho^2(Q_e - 1)$)

$$\forall e \in S^{n-1},\ (1-\epsilon)(1 + \rho^2(Q_e - 1)) + \epsilon \in \mathrm{conv}(P_\rho f_1, \ldots, P_\rho f_N)$$

- Evaluate at $x = e$:

$$\forall e \in S^{n-1},\quad (1-\epsilon)(1 + \rho^2(n - 1)) + \epsilon \ \in\ \mathrm{conv}(P_\rho f_1(e), \ldots, P_\rho f_N(e))$$

Proof of lower bound

- Assume that $f_1, \ldots, f_N : S^{n-1} \to \mathbb{R}_+$, $\int f_i = 1$ are such that:

$$\forall e \in S^{n-1},\ (1-\epsilon) Q_e + \epsilon \in \mathrm{conv}(f_1, \ldots, f_N)$$

where $Q_e(x) = n(e^T x)^2$ [extreme rays of the psd cone].

- Apply P_ρ on both sides (simple calculation $P_\rho Q_e = 1 + \rho^2(Q_e - 1)$)

$$\forall e \in S^{n-1},\ (1-\epsilon)(1 + \rho^2(Q_e - 1)) + \epsilon \in \mathrm{conv}(P_\rho f_1, \ldots, P_\rho f_N)$$

- Evaluate at $x = e$:

$$\forall e \in S^{n-1},\ \underbrace{(1-\epsilon)(1 + \rho^2(n-1)) + \epsilon}_{\geq 4 \text{ for } \rho = \sqrt{5/n} \text{ and } \epsilon = 1/3} \in \mathrm{conv}(P_\rho f_1(e), \ldots, P_\rho f_N(e))$$

Proof of lower bound

- Assume that $f_1, \ldots, f_N : S^{n-1} \to \mathbb{R}_+$, $\int f_i = 1$ are such that:

$$\forall e \in S^{n-1},\ (1-\epsilon) Q_e + \epsilon \in \mathrm{conv}(f_1, \ldots, f_N)$$

where $Q_e(x) = n(e^T x)^2$ [extreme rays of the psd cone].

- Apply P_ρ on both sides (simple calculation $P_\rho Q_e = 1 + \rho^2 (Q_e - 1)$)

$$\forall e \in S^{n-1},\ (1-\epsilon)(1 + \rho^2 (Q_e - 1)) + \epsilon \in \mathrm{conv}(P_\rho f_1, \ldots, P_\rho f_N)$$

- Evaluate at $x = e$:

$$\forall e \in S^{n-1},\quad \underbrace{(1-\epsilon)(1+\rho^2(n-1)) + \epsilon}_{\geq 4 \text{ for } \rho = \sqrt{5/n} \text{ and } \epsilon = 1/3} \ \in\ \mathrm{conv}(P_\rho f_1(e), \ldots, P_\rho f_N(e))$$

- Consequence: for any $e \in S^{n-1}$ there is at least one i s.t. $(P_\rho f_i)(e) \geq 4$

Proof of lower bound

- Assume that $f_1, \ldots, f_N : S^{n-1} \to \mathbb{R}_+$, $\int f_i = 1$ are such that:

$$\forall e \in S^{n-1},\ (1-\epsilon) Q_e + \epsilon \in \mathrm{conv}(f_1, \ldots, f_N)$$

where $Q_e(x) = n(e^T x)^2$ [extreme rays of the psd cone].

- Apply P_ρ on both sides (simple calculation $P_\rho Q_e = 1 + \rho^2(Q_e - 1)$)

$$\forall e \in S^{n-1},\ (1-\epsilon)(1 + \rho^2(Q_e - 1)) + \epsilon \in \mathrm{conv}(P_\rho f_1, \ldots, P_\rho f_N)$$

- Evaluate at $x = e$:

$$\forall e \in S^{n-1},\quad \underbrace{(1-\epsilon)(1 + \rho^2(n - 1)) + \epsilon}_{\geq 4 \text{ for } \rho = \sqrt{5/n} \text{ and } \epsilon = 1/3} \ \in\ \mathrm{conv}(P_\rho f_1(e), \ldots, P_\rho f_N(e))$$

- Consequence: for any $e \in S^{n-1}$ there is at least one i s.t. $(P_\rho f_i)(e) \geq 4$
- Previous lemma (+some technical details) tells us $N \geq \exp(c\sqrt{n})$.

Conclusion

- Can we identify slices of the positive semidefinite cone that can be well approximated using LPs? Can be potentially useful for mixed-integer SDPs

- Dependence on ϵ, i.e., regime $\epsilon \to 0$?

Conclusion

- Can we identify slices of the positive semidefinite cone that can be well approximated using LPs? Can be potentially useful for mixed-integer SDPs

- Dependence on ϵ, i.e., regime $\epsilon \to 0$?

Thank you!