Sum of squares optimization

Pablo A. Parrilo

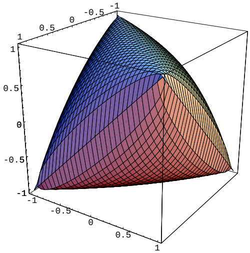
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Summer School on
Numerical Computing in Algebraic Geometry
MPI Leipzig - August 2018

Massachusetts
Institute of
Technology

Topics

- Convexity, non-convexity, and tractability
- Convex sets with algebraic descriptions
- Semidefinite programming and sums of squares
- Unifying idea: convex hull of algebraic varieties
- Examples and applications throughout
- Discuss results, but also open questions
- Computational considerations
- Connections with other areas of mathematics



Outline

- Part I
  - Motivation, Basic notions
  - Convexity vs. non-convexity
  - Linear and Semidefinite programming
- Part II
  - Sums of squares
  - Convex hull of algebraic varieties
  - General constructions and approximation guarantees
- Part III
  - Applications and extensions
  - Rank minimization via nuclear norm
  - Estimation and synchronization over $SO(n)$
  - Algorithmic aspects
  - Recap and conclusions

Motivation: what problems can we solve efficiently?

Many questions in applied mathematics can be formulated in terms of polynomials (sometimes, after nontrivial modeling!)

- Global optimization (e.g., binary, constrained, etc.)
- Stability of dynamical systems (e.g., Lyapunov analysis)
- Quantum information (e.g., entanglement)

Some have "nice" solutions.
Others, we are still struggling with after many years...

Why this difference?
What are the underlying mathematical reasons?

Motivation: what problems can we solve efficiently?

Many questions in applied mathematics can be formulated in terms of polynomials (sometimes, after nontrivial modeling!)

- Global optimization (e.g., binary, constrained, etc.)
- Stability of dynamical systems (e.g., Lyapunov analysis)
- Quantum information (e.g., entanglement)

Some have "nice" solutions.
Others, we are still struggling with after many years...

Why this difference?
What are the underlying mathematical reasons?

Motivation: what problems can we solve efficiently?

Many questions in applied mathematics can be formulated in terms of polynomials (sometimes, after nontrivial modeling!)

- Global optimization (e.g., binary, constrained, etc.)
- Stability of dynamical systems (e.g., Lyapunov analysis)
- Quantum information (e.g., entanglement)

Some have “nice” solutions.
Others, we are still struggling with after many years...

Why this difference?
What are the underlying mathematical reasons?

Motivation: what problems can we solve efficiently?

A (rough) first classification: convex vs. non-convex

Extremely valuable insights! (e.g., Boyd-Vandenberghe)

But, the full answer is a bit more subtle...

- Convexity is "relative," may depend on modeling/parameterization
- If not convex, may perhaps be tractable (e.g., PCA, deep learning)
- Even if convex, may not be efficiently tractable! (e.g., nonnegative polynomials)

Why? Difference between:
"Geometric" vs. "computational" convexity

Need to enrich convexity theory with a computational twist.

Motivation: what problems can we solve efficiently?

A (rough) first classification: convex vs. non-convex

Extremely valuable insights! (e.g., Boyd-Vandenberghe)

But, the full answer is a bit more subtle...

- Convexity is "relative," may depend on modeling/parameterization
- If not convex, may perhaps be tractable (e.g., PCA, deep learning)
- Even if convex, may not be efficiently tractable! (e.g., nonnegative polynomials)

Why? Difference between:
"Geometric" vs. "computational" convexity

Need to enrich convexity theory with a computational twist.

Motivation: what problems can we solve efficiently?

A (rough) first classification: convex vs. non-convex

Extremely valuable insights! (e.g., Boyd-Vandenberghe)

But, the full answer is a bit more subtle...

- Convexity is "relative," may depend on modeling/parameterization
- If not convex, may perhaps be tractable (e.g., PCA, deep learning)
- Even if convex, may not be efficiently tractable! (e.g., nonnegative polynomials)

Why? Difference between:
"Geometric" vs. "computational" convexity

Need to enrich convexity theory with a computational twist.

Motivation: what problems can we solve efficiently?

A (rough) first classification: convex vs. non-convex

Extremely valuable insights! (e.g., Boyd-Vandenberghe)

But, the full answer is a bit more subtle...

- Convexity is "relative," may depend on modeling/parameterization
- If not convex, may perhaps be tractable (e.g., PCA, deep learning)
- Even if convex, may not be efficiently tractable! (e.g., nonnegative polynomials)

Why? Difference between:
"Geometric" vs. "computational" convexity

Need to enrich convexity theory with a computational twist.

Motivation: what problems can we solve efficiently?

A (rough) first classification: convex vs. non-convex

Extremely valuable insights! (e.g., Boyd-Vandenberghe)

But, the full answer is a bit more subtle...

- Convexity is "relative," may depend on modeling/parameterization
- If not convex, may perhaps be tractable (e.g., PCA, deep learning)
- Even if convex, may not be efficiently tractable! (e.g., nonnegative polynomials)

Why? Difference between:
"Geometric" vs. "computational" convexity

Need to enrich convexity theory with a computational twist.

Convex sets: geometry vs. algebra

The geometric theory of convex sets (e.g., Minkowski, Carathéodory, Fenchel) is very rich and well-understood.

Enormous importance in applied mathematics and engineering, in particular in optimization.

But, what if we are concerned with the *representation* of these geometric objects? For instance, basic semialgebraic sets?

How do the *algebraic*, *geometric*, and *computational* aspects interact?

Ex: Convex optimization is not always "easy".

The polyhedral case

Consider first the case of *polyhedra*, which are described by finitely many *linear* inequalities $\{x \in \mathbb{R}^n : a_i^T x \leq b_i\}$.

- Behave well under projections (Fourier-Motzkin)
- Farkas' lemma (or duality) gives emptiness certificates
- Good associated computational techniques
- Optimization over polyhedra is linear programming (LP)

Great. But how to move away from linearity? For instance, if we want convex sets described by polynomial inequalities?

Claim: semidefinite programming is an essential tool.

The polyhedral case

Consider first the case of *polyhedra*, which are described by finitely many *linear* inequalities $\{x \in \mathbb{R}^n \ : \ a_i^T x \leq b_i\}$.

- Behave well under projections (Fourier-Motzkin)
- Farkas' lemma (or duality) gives emptiness certificates
- Good associated computational techniques
- Optimization over polyhedra is linear programming (LP)

Great. But how to move away from linearity? For instance, if we want convex sets described by polynomial inequalities?

Claim: semidefinite programming is an essential tool.

Linear programming

LP in standard (primal) form:

$$\min c^T x \qquad \text{s.t.} \qquad Ax = b, \quad x \geq 0.$$

A geometric view: if $\mathcal{L}$ is an affine subspace of $\mathbb{R}^n$,

$$\min c^T x \qquad \text{s.t.} \quad x \in \mathcal{L} \cap \mathbb{R}^n_+$$

Minimize a linear function, over the intersection of an affine subspace and a polyhedral cone (nonnegative orthant).

In semidefinite programming, x will now be a symmetric matrix...

Linear programming

LP in standard (primal) form:

$$\min c^T x \qquad \text{s.t.} \qquad Ax = b, \quad x \geq 0.$$

A geometric view: if $\mathcal{L}$ is an affine subspace of $\mathbb{R}^n$,

$$\min c^T x \qquad \text{s.t.} \quad x \in \mathcal{L} \cap \mathbb{R}^n_+$$

Minimize a linear function, over the intersection of an affine subspace and a polyhedral cone (nonnegative orthant).

In semidefinite programming, x will now be a symmetric matrix...

Linear programming

LP in standard (primal) form:

$$\min c^T x \qquad \text{s.t.} \qquad Ax = b, \quad x \geq 0.$$

A geometric view: if $\mathcal{L}$ is an affine subspace of $\mathbb{R}^n$,

$$\min c^T x \qquad \text{s.t.} \quad x \in \mathcal{L} \cap \mathbb{R}^n_+$$

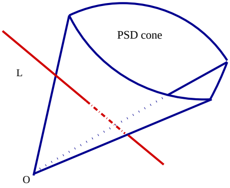
Minimize a linear function, over the intersection of an affine subspace and a polyhedral cone (nonnegative orthant).

In semidefinite programming, x will now be a symmetric matrix...

Semidefinite programming (SDP, LMIs)

A broad generalization of LP to symmetric matrices

$$\min \operatorname{Tr} CX \qquad \text{s.t.} \quad X \in \mathcal{L} \cap \mathcal{S}^n_+$$

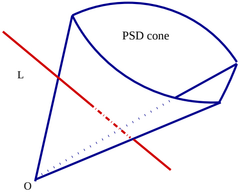


- Intersection of an affine subspace $\mathcal{L}$ and the cone of positive semidefinite matrices.
- Feasible set is called *spectrahedron*
- *Lots* of applications. A true "revolution" in computational methods for engineering applications
- Convex finite dimensional optimization. Nice duality theory.
- Essentially, solvable in polynomial time (interior point, etc.)

SDPs in standard form

Standard (primal) form of a semidefinite program:

$$\min \operatorname{Tr} CX \qquad \text{s.t.} \begin{cases} \operatorname{Tr} A_i X = b_i \\ X \succeq 0, \end{cases}$$



where $X \in \mathbb{R}^{n\times n}$ is the (matrix) decision variable and $A_1, \ldots, A_m \in \mathbb{R}^{n\times n}$ are given symmetric matrices.

The inequality $A \succeq 0$ means that A is positive semidefinite (psd):

$$A \succeq 0 \quad \Leftrightarrow \quad z^T A z \geq 0 \quad \forall z \in \mathbb{R}^n \quad \Leftrightarrow \quad \lambda_i(A) \geq 0 \quad i = 1, \ldots, n.$$

By Sylvester's criterion, also equivalent to nonnegativity of all principal minors:

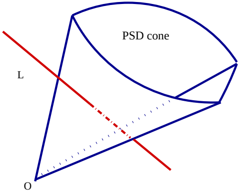
$$\det A_{S,S} \geq 0 \quad \forall S \subset \{1, \ldots, n\}.$$

Thus, feasible set is defined by polynomial inequalities.

SDPs in standard form

Standard (primal) form of a semidefinite program:

$$\min \operatorname{Tr} CX \qquad \text{s.t.} \begin{cases} \operatorname{Tr} A_i X = b_i \\ X \succeq 0, \end{cases}$$



where $X \in \mathbb{R}^{n\times n}$ is the (matrix) decision variable and $A_1, \ldots, A_m \in \mathbb{R}^{n\times n}$ are given symmetric matrices.

The inequality $A \succeq 0$ means that A is positive semidefinite (psd):

$$A \succeq 0 \quad \Leftrightarrow \quad z^T Az \geq 0 \quad \forall z \in \mathbb{R}^n \quad \Leftrightarrow \quad \lambda_i(A) \geq 0 \quad i = 1, \ldots, n.$$

By Sylvester's criterion, also equivalent to nonnegativity of all principal minors:

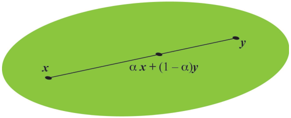
$$\det A_{S,S} \geq 0 \quad \forall S \subset \{1, \ldots, n\}.$$

Thus, feasible set is defined by polynomial inequalities.

Convexity

Recall that a set S is convex if

$$x, y \in S \quad \Rightarrow \quad \alpha x + (1-\alpha) y \in S$$



for all $\alpha \in [0,1]$.

Lemma: The cone of positive semidefinite matrices is convex.

Let A and B be psd, and $C = \alpha A + (1-\alpha) B$ with $\alpha \in [0,1]$.
Then, for all $z \in \mathbb{R}^n$,

$$z^T C z = z^T (\alpha A + (1-\alpha) B) z = \alpha \underbrace{z^T A z}_{\geq 0} + (1-\alpha) \underbrace{z^T B z}_{\geq 0} \geq 0,$$

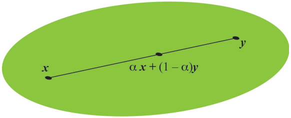
and thus C is also psd.

Convexity

Recall that a set S is convex if

$$x, y \in S \quad \Rightarrow \quad \alpha x + (1-\alpha)y \in S$$

for all $\alpha \in [0, 1]$.



Lemma: The cone of positive semidefinite matrices is convex.

Let A and B be psd, and $C = \alpha A + (1-\alpha)B$ with $\alpha \in [0,1]$.
Then, for all $z \in \mathbb{R}^n$,

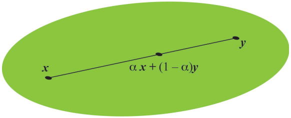
$$z^T Cz = z^T(\alpha A + (1-\alpha)B)z = \alpha \underbrace{z^T Az}_{\geq 0} + (1-\alpha) \underbrace{z^T Bz}_{\geq 0} \geq 0,$$

and thus C is also psd.

Convexity

Recall that a set S is convex if

$$x, y \in S \quad \Rightarrow \quad \alpha x + (1-\alpha) y \in S$$



for all $\alpha \in [0, 1]$.

Lemma: The cone of positive semidefinite matrices is convex.

Let A and B be psd, and $C = \alpha A + (1-\alpha) B$ with $\alpha \in [0, 1]$.
Then, for all $z \in \mathbb{R}^n$,

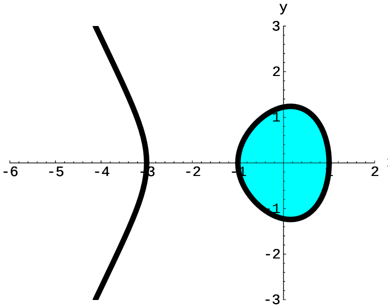
$$z^T C z = z^T(\alpha A + (1-\alpha) B) z = \alpha \underbrace{z^T A z}_{\geq 0} + (1-\alpha) \underbrace{z^T B z}_{\geq 0} \geq 0,$$

and thus C is also psd.

Example (I)

Consider the spectrahedron given by the SDP:

$$\begin{bmatrix} x & 0 & y \\ 0 & 1 & -x \\ y & -x & 1 \end{bmatrix} \succeq 0.$$



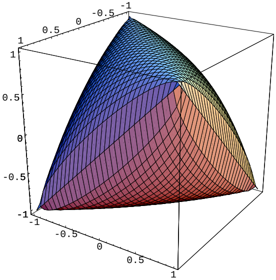
- Convex, but not necessarily polyhedral
- In general, boundary is piecewise-smooth
- Determinant vanishes on the boundary

In this example, the determinant is the elliptic curve $x - x^3 = y^2$.

Example (II)

Consider the spectrahedron:

$$X_{ii} = 1 \qquad X \succeq 0$$



- PSD matrices of unit diagonal
- Interpretation: set of correlation matrices
- Known as the elliptope.

$$M = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix} \succeq 0$$

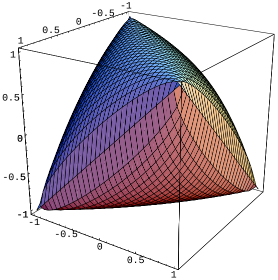
Boundary is the Cayley cubic

$$\det M = 1 - (a^2 + b^2 + c^2) + 2abc$$

Example (II)

Consider the spectrahedron:

$$X_{ii} = 1 \qquad X \succeq 0$$



- PSD matrices of unit diagonal
- Interpretation: set of correlation matrices
- Known as the elliptope.

$$M = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix} \succeq 0$$

Boundary is the Cayley cubic

$$\det M = 1 - (a^2 + b^2 + c^2) + 2abc$$

Symbolic vs. numerical computation

An ongoing discussion. Clearly, both have advantages/disadvantages.

- "Exact solutions" vs. "approximations"
- "Input data often inexact"
- "Global" vs. "local". One vs. all solutions.
- Computational models: bits vs. reals. Encoding of solutions.

"Best" method depends on the context. Hybrid symbolic-numeric methods are an interesting possibility.

SDP bring some interesting new twists.

- For LP, "numerical" algorithms (ellipsoid, interior-point) are polytime, while "symbolic" or "combinatorial" ones (e.g. simplex) are not.
- Worse for SDP.

Algebraic aspects of SDP

In LPs with rational data, the optimal solution is rational. Not so for SDP.

- Optimal solutions of relatively small SDPs generically have minimum defining polynomials of very high degree.
- Example (von Bothmer and Ranestad): For $n = 20$, $m = 105$, the algebraic degree of the optimal solution is $\approx 1.67 \times 10^{41}$.
- Explicit algebraic representations are absolutely impossible to compute (even without worrying about coefficient size!).
- Nevertheless, interior point methods yield arbitrary precision numerical approximations!

SDP provides an efficient, and numerically convenient *encoding*.
Representation does not pay the price of high algebraic complexity.

Application: binary quadratic optimization

Consider the maximization problem

$$\max_x x^T Q x \qquad \text{s.t.} \qquad x_i \in \{-1, 1\}.$$

A quadratic function, on the vertices of the hypercube. Difficult in theory (NP-hard), and also in practice. Very important in applications.

Can we produce "strong" upper bounds on the optimal value $q^\star$? (e.g., for branch and bound)

Let $\gamma_\star$ be the optimal value of the SDP:

$$\min \operatorname{Tr} D \qquad Q \preceq D, \quad D \text{ diagonal.}$$

Then, for any $x \in \{-1, 1\}^n$, and any feasible D we have:

$$x^T Q x \leq x^T D x = \sum_{i=1}^n D_{ii} x_i^2 = \operatorname{Tr} D$$

and thus $q^\star \leq \gamma_\star$. The upper bound $\gamma_\star$ can be efficiently computed.

Application: binary quadratic optimization

Consider the maximization problem

$$\max_x x^T Q x \qquad \text{s.t.} \qquad x_i \in \{-1, 1\}.$$

A quadratic function, on the vertices of the hypercube. Difficult in theory (NP-hard), and also in practice. Very important in applications.

Can we produce "strong" upper bounds on the optimal value $q^\star$? (e.g., for branch and bound)

Let $\gamma_\star$ be the optimal value of the SDP:

$$\min \operatorname{Tr} D \qquad Q \preceq D, \quad D \text{ diagonal.}$$

Then, for any $x \in \{-1, 1\}^n$, and any feasible D we have:

$$x^T Q x \leq x^T D x = \sum_{i=1}^n D_{ii} x_i^2 = \operatorname{Tr} D$$

and thus $q^\star \leq \gamma_\star$. The upper bound $\gamma_\star$ can be efficiently computed.

Application: binary quadratic optimization

Consider the maximization problem

$$\max_x x^T Q x \qquad \text{s.t.} \qquad x_i \in \{-1, 1\}.$$

A quadratic function, on the vertices of the hypercube. Difficult in theory (NP-hard), and also in practice. Very important in applications.

Can we produce "strong" upper bounds on the optimal value $q^\star$? (e.g., for branch and bound)

Let $\gamma_\star$ be the optimal value of the SDP:

$$\min \operatorname{Tr} D \qquad Q \preceq D, \quad D \text{ diagonal}.$$

Then, for any $x \in \{-1, 1\}^n$, and any feasible D we have:

$$x^T Q x \leq x^T D x = \sum_{i=1}^{n} D_{ii} x_i^2 = \operatorname{Tr} D$$

and thus $q^\star \leq \gamma_\star$. The upper bound $\gamma_\star$ can be efficiently computed.

Application: binary quadratic optimization

Consider the maximization problem

$$\max_x x^T Q x \qquad \text{s.t.} \qquad x_i \in \{-1, 1\}.$$

A quadratic function, on the vertices of the hypercube. Difficult in theory (NP-hard), and also in practice. Very important in applications.

Can we produce "strong" upper bounds on the optimal value $q^\star$? (e.g., for branch and bound)

Let $\gamma_\star$ be the optimal value of the SDP:

$$\min \operatorname{Tr} D \qquad Q \preceq D, \quad D \text{ diagonal}.$$

Then, for any $x \in \{-1, 1\}^n$, and any feasible D we have:

$$x^T Q x \leq x^T D x = \sum_{i=1}^{n} D_{ii} x_i^2 = \operatorname{Tr} D$$

and thus $q^\star \leq \gamma_\star$. The upper bound $\gamma_\star$ can be efficiently computed.

Application: binary quadratic optimization (II)

How good is this upper bound? How to quantify this?

Different cases, depending on properties of the cost function Q:

- Q is diagonally dominant ($Q_{ii} \geq \sum_{i \neq j} Q_{ij}$). This is the case of MAX-CUT, where Q is the Laplacian of a graph. Goemans and Williamson showed that

$$0.878\,\gamma_\star \leq f^* \leq \gamma_\star$$

- Q is positive semidefinite ($Q \succeq 0$). By results of Nesterov (and earlier, in very different form, Grothendieck)

$$\frac{2}{\pi}\gamma_\star \leq f^* \leq \gamma_\star$$

- Q has a bipartite structure. Then, Grothendieck's inequality (and Krivine's bound) yields

$$0.5611\,\gamma_\star \leq f^* \leq \gamma_\star$$

Application: binary quadratic optimization (II)

How good is this upper bound? How to quantify this?

Different cases, depending on properties of the cost function Q:

- Q is diagonally dominant ($Q_{ii} \geq \sum_{i \neq j} Q_{ij}$). This is the case of MAX-CUT, where Q is the Laplacian of a graph. Goemans and Williamson showed that

$$0.878\,\gamma_\star \leq f^* \leq \gamma_\star$$

- Q is positive semidefinite ($Q \succeq 0$). By results of Nesterov (and earlier, in very different form, Grothendieck)

$$\frac{2}{\pi}\gamma_\star \leq f^* \leq \gamma_\star$$

- Q has a bipartite structure. Then, Grothendieck's inequality (and Krivine's bound) yields

$$0.5611\,\gamma_\star \leq f^* \leq \gamma_\star$$

Semidefinite representations

Natural question in convex optimization:

What sets can be represented using semidefinite programming?

Equivalently, can I solve this problem using SDP?

In the LP case, well-understood question: finite number of extreme points/rays (polyhedral sets, Minkowski-Weyl)

Are there "obstructions" to SDP representability?

Semidefinite representations

Natural question in convex optimization:

What sets can be represented using semidefinite programming?

Equivalently, can I solve this problem using SDP?

In the LP case, well-understood question: finite number of extreme points/rays (polyhedral sets, Minkowski-Weyl)

Are there "obstructions" to SDP representability?

Semidefinite representations

Natural question in convex optimization:

What sets can be represented using semidefinite programming?

Equivalently, can I solve this problem using SDP?

In the LP case, well-understood question: finite number of extreme points/rays (polyhedral sets, Minkowski-Weyl)

Are there "obstructions" to SDP representability?

Semidefinite representations

Natural question in convex optimization:

What sets can be represented using semidefinite programming?

Equivalently, can I solve this problem using SDP?

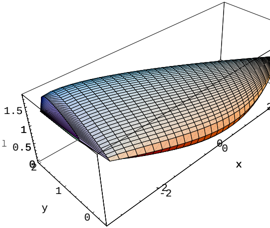
In the LP case, well-understood question: finite number of extreme points/rays (polyhedral sets, Minkowski-Weyl)

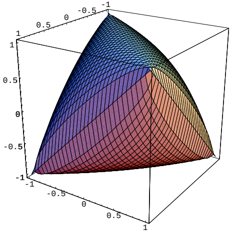
Are there "obstructions" to SDP representability?

Known SDP-representable sets

- Many interesting sets are known to be SDP-representable (e.g., polyhedra, convex quadratics, matrix norms, etc.)
- Preserved by "natural" properties: affine transformations, convex hull, polarity, etc.
- Several known structural results (e.g., facial exposedness)

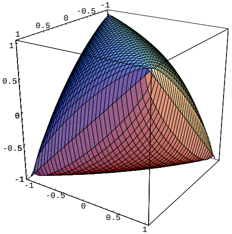
Work of Nesterov-Nemirovski, Ben-Tal, Ramana, Tunçel, Güler, Renegar, Chua, etc.





A few examples of SDP-representable sets

- Sums of eigenvalues of symmetric matrices
- Convex envelope of univariate polynomials
- Multivariate polynomials that are sums of squares
- Unit ball of matrix operator and nuclear norms
- Geometric and harmonic means
- (Some) orbitopes – convex hulls of group orbits
- Lyapunov functions of (non)linear systems
- Optimal decentralized controllers (under certain information structures)



Often, clever and non-obvious reformulations.

Existing results

Obvious necessary conditions: $\mathcal{S}$ must be convex and semialgebraic.

Several versions of the problem:

- *Exact* vs. *approximate* representations.
- "Direct" (non-lifted) representations: no additional variables.

$$x \in \mathcal{S} \quad \Leftrightarrow \quad A_0 + \sum_i x_i A_i \succeq 0$$

- "Lifted" representations: can use extra variables (projection)

$$x \in \mathcal{S} \quad \Leftrightarrow \quad \exists y \text{ s.t. } A_0 + \sum_i x_i A_i + \sum y_j B_j \succeq 0$$

Projection helps a lot!

Existing results

Obvious necessary conditions: $\mathcal{S}$ must be convex and semialgebraic.

Several versions of the problem:

- *Exact* vs. *approximate* representations.
- "Direct" (non-lifted) representations: no additional variables.

$$x \in \mathcal{S} \quad \Leftrightarrow \quad A_0 + \sum_i x_i A_i \succeq 0$$

- "Lifted" representations: can use extra variables (projection)

$$x \in \mathcal{S} \quad \Leftrightarrow \quad \exists y \text{ s.t. } A_0 + \sum_i x_i A_i + \sum y_j B_j \succeq 0$$

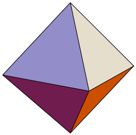
Projection helps a lot!

Liftings and projections

Often, "simpler" descriptions of convex sets from higher-dimensions.

Ex: The n-dimensional crosspolytope (ℓ_1 unit ball). Requires 2^n linear inequalities, of the form

$$\pm x_1 \pm x_2 \pm \cdots \pm x_n \leq 1.$$



However, can efficiently represent it as a *projection*:

$$\{(x, y) \in \mathbb{R}^{2n}, \quad \sum_{i=1}^{n} y_i = 1, \quad -y_i \leq x_i \leq y_i \quad i = 1, \ldots, n\}$$

Only $2n$ variables, and $2n+1$ constraints!

In convexity, elimination is *not* always a good idea.

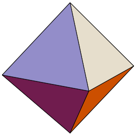
Quite the opposite, it is often advantageous to go to higher-dimensional spaces, where descriptions (can) become simpler.

Liftings and projections

Often, "simpler" descriptions of convex sets from higher-dimensions.

Ex: The n-dimensional crosspolytope (ℓ_1 unit ball). Requires 2^n linear inequalities, of the form

$$\pm x_1 \pm x_2 \pm \cdots \pm x_n \leq 1.$$



However, can efficiently represent it as a *projection*:

$$\{(x, y) \in \mathbb{R}^{2n}, \quad \sum_{i=1}^{n} y_i = 1, \quad -y_i \leq x_i \leq y_i \quad i = 1, \ldots, n\}$$

Only $2n$ variables, and $2n+1$ constraints!
In convexity, elimination is *not* always a good idea.
Quite the opposite, it is often advantageous to go to higher-dimensional spaces, where descriptions (can) become simpler.

Aside: representability of convex sets

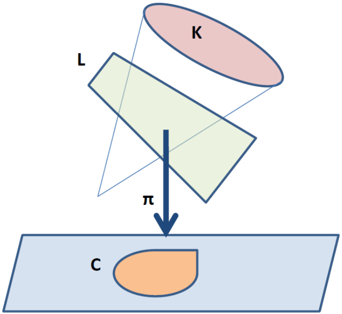
Existence and efficiency:

- When is a convex set representable by conic optimization?
- How to quantify the number of additional variables that are needed?

Given a convex set C, is it possible to represent it as

$$C = \pi(K \cap L)$$

where K is a cone, L is an affine subspace, and π is a linear map?



A beatiful story, much of it in progress. But not today...

Aside: representability of convex sets

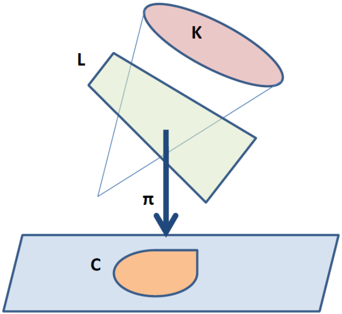
Existence and efficiency:

- When is a convex set representable by conic optimization?
- How to quantify the number of additional variables that are needed?

Given a convex set C, is it possible to represent it as

$$C = \pi(K \cap L)$$

where K is a cone, L is an affine subspace, and π is a linear map?



A beatiful story, much of it in progress. But not today...

Example: facility location and k-ellipses

Consider the *facility location* problem: given k customer locations $(u_i, v_i) \in \mathbb{R}^2$, decide where to build a new facility in such a way that total shipping costs are minimized:

$$\min_{(x,y)} \sum_{i=1}^{k} f((x,y),(u_i,v_i)),$$

where $f(\cdot,\cdot)$ models the shipping costs.

If f is the Euclidean distance, this is the classical Fermat-Weber problem.

Simple and natural SOCP/SDP representation (w/extra variables):

$$\min \sum_{i=1}^{k} d_i \qquad \text{s.t.} \qquad \begin{bmatrix} x-u_i+d_i & y-v_i \\ y-v_i & x-u_i-d_i \end{bmatrix} \succeq 0$$

(constraints are $\|(x,y)-(u_i,v_i)\| \leq d_i$)

Example: facility location and k-ellipses

Consider the *facility location* problem: given k customer locations $(u_i, v_i) \in \mathbb{R}^2$, decide where to build a new facility in such a way that total shipping costs are minimized:

$$\min_{(x,y)} \sum_{i=1}^{k} f((x, y), (u_i, v_i)),$$

where $f(\cdot, \cdot)$ models the shipping costs.

If f is the Euclidean distance, this is the classical Fermat-Weber problem.

Simple and natural SOCP/SDP representation (w/extra variables):

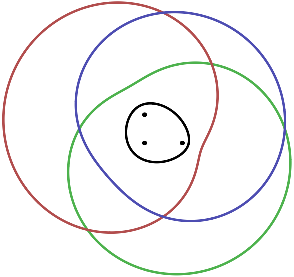
$$\min \sum_{i=1}^{k} d_i \qquad \text{s.t.} \qquad \begin{bmatrix} x - u_i + d_i & y - v_i \\ y - v_i & x - u_i - d_i \end{bmatrix} \succeq 0$$

(constraints are $\|(x, y) - (u_i, v_i)\| \leq d_i$)

Example: k-ellipse

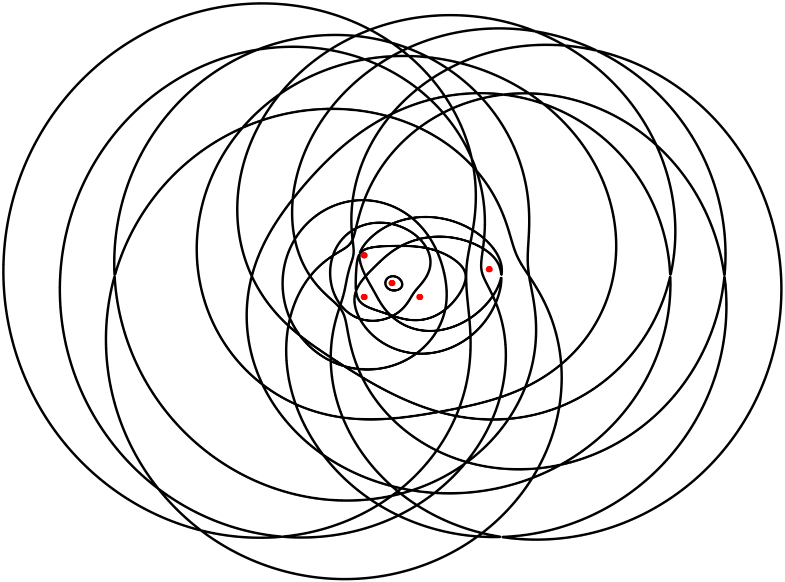
Fix a positive real number d and fix k distinct points (u_i, v_i) in $\mathbb{R}^2$. The *k-ellipse* with *foci* (u_i, v_i) and *radius* d is the following curve in the plane:

$$\left\{ (x, y) \in \mathbb{R}^2 \, : \, \sum_{i=1}^{k} \sqrt{(x - u_i)^2 + (y - v_i)^2} \, = \, d \right\}.$$



Thm:(Nie-P.-Sturmfels 07) The k-ellipse has degree 2^k if k is odd and degree $2^k - \binom{k}{k/2}$ if k is even. It has an explicit $2^k \times 2^k$ SDP representation.

5-ellipse



Results on exact SDP representations

- Direct representations:
  - Necessary condition: rigid convexity. Helton & Vinnikov (2004) showed that in $\mathbb{R}^2$, rigid convexity is also sufficient.
  - Related to hyperbolic polynomials and the Lax conjecture (Güler, Renegar, Lewis-P.-Ramana 2005)
  - For higher dimensions the problem is open.
- Lifted representations:
  - Does *every* convex basic SA set have a lifted exact SDP representation?
  - (Helton & Nie 2007): Under strict positive curvature assumptions on the boundary, this is true.
  - No known nontrivial obstructions – until recently...

Results on exact SDP representations

- Direct representations:
  - Necessary condition: rigid convexity. Helton & Vinnikov (2004) showed that in $\mathbb{R}^2$, rigid convexity is also sufficient.
  - Related to hyperbolic polynomials and the Lax conjecture (Güler, Renegar, Lewis-P.-Ramana 2005)
  - For higher dimensions the problem is open.
- Lifted representations:
  - Does *every* convex basic SA set have a lifted exact SDP representation?
  - (Helton & Nie 2007): Under strict positive curvature assumptions on the boundary, this is true.
  - No known nontrivial obstructions – until recently...

Exact representations: direct case

$$x \in \mathcal{S} \quad \Leftrightarrow \quad A_0 + \sum_i x_i A_i \succeq 0$$

Necessary condition: "rigid convexity." Every line through the set must intersect the Zariski closure of the boundary a constant number of times (equal to the degree of the curve).

[Assume $A_0 \succ 0$, and let $x_i = t\beta_i$. Then the univariate polynomial $q(t) := \det(A_0 + \sum x_i A_i) = \det(A_0 + t \cdot \sum \beta_i A_i)$ has all its d roots real.]

Can use it to show, for instance, that the convex set $\mathcal{S} = \{(x, y) : x^4 + y^4 \leq 1\}$ does not have a direct SDP representation.

Exact representations: direct case

$$x \in \mathcal{S} \quad \Leftrightarrow \quad A_0 + \sum_i x_i A_i \succeq 0$$

Necessary condition: "rigid convexity." Every line through the set must intersect the Zariski closure of the boundary a constant number of times (equal to the degree of the curve).

[Assume $A_0 \succ 0$, and let $x_i = t\beta_i$. Then the univariate polynomial $q(t) := \det(A_0 + \sum x_i A_i) = \det(A_0 + t \cdot \sum \beta_i A_i)$ has all its d roots real.]

Can use it to show, for instance, that the convex set $\mathcal{S} = \{(x, y) : x^4 + y^4 \leq 1\}$ does not have a direct SDP representation.

Exact representations: direct case

$$x \in \mathcal{S} \quad \Leftrightarrow \quad A_0 + \sum_i x_i A_i \succeq 0$$

Necessary condition: "rigid convexity." Every line through the set must intersect the Zariski closure of the boundary a constant number of times (equal to the degree of the curve).

[Assume $A_0 \succ 0$, and let $x_i = t\beta_i$. Then the univariate polynomial $q(t) := \det(A_0 + \sum x_i A_i) = \det(A_0 + t \cdot \sum \beta_i A_i)$ has all its d roots real.]

Can use it to show, for instance, that the convex set $\mathcal{S} = \{(x, y) \, : \, x^4 + y^4 \leq 1\}$ does not have a direct SDP representation.

Exact representations: lifted case

Very active research topic, both in qualitative and quantitative flavors. Recent exciting progress, on several fronts!

- Gouveia-P.-Thomas (`arXiv:1111.3164`): PSD rank of a convex set, extension of Yannakakis' LP theory to SDP extension complexity.
- Lee-Raghavendra-Steurer (`arXiv:1411.6317`, STOC2015): exponential lower bounds on SDP relaxations of cut polytope.
- Fawzi (`arXiv:1610.04901`): S_+^3 is not SOCP-representable. SOCP-rank + Turán's theorem
- Scheiderer (`arXiv:1612.07048`): nonnegative polynomials (for $d \geq 4$, $n \geq 2$) are not SDP-representable. Real algebraic-geometric tools (real spectrum, Tarski's transfer principle, ...)

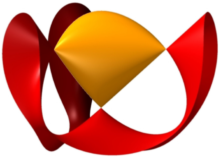
Summary

- SDP is a natural generalization of linear programming
- Rich algebraic-geometric structure
- Many applications, efficient numerical solvers
- Some fundamental aspects not fully understood yet

End of Part I

Summary

- SDP is a natural generalization of linear programming
- Rich algebraic-geometric structure
- Many applications, efficient numerical solvers
- Some fundamental aspects not fully understood yet



End of Part I

Part II

Recap

- Motivation: convexity, with computational content
- SDP as a natural generalization of LP
- Understanding the power and limitations of SDP

Lyapunov stability analysis

- A linear system $\dot{x} = Ax$, quadratic Lyapunov function $V(x) = x^T P x$

$$P \succ 0, \qquad A^T P + PA \prec 0$$

- More generally, for a nonlinear system $\dot{x} = f(x)$,

$$V(x) > 0 \quad x \neq 0, \qquad \dot{V}(x) = \left(\frac{\partial V}{\partial x}\right)^T f(x) < 0, \quad x \neq 0$$

(locally, or globally if V is radially unbounded).

- Many variations: $\mathcal{H}_2$ and/or $\mathcal{H}_\infty$ analysis, parameter-dependent Lyapunov functions, etc.

- The problem is clearly convex in $V(x)$. But, how to solve this?

Lyapunov stability analysis

- A linear system $\dot{x} = Ax$, quadratic Lyapunov function $V(x) = x^T P x$

$$P \succ 0, \qquad A^T P + PA \prec 0$$

- More generally, for a nonlinear system $\dot{x} = f(x)$,

$$V(x) > 0 \quad x \neq 0, \qquad \dot{V}(x) = \left(\frac{\partial V}{\partial x}\right)^T f(x) < 0, \quad x \neq 0$$

(locally, or globally if V is radially unbounded).

- Many variations: $\mathcal{H}_2$ and/or $\mathcal{H}_\infty$ analysis, parameter-dependent Lyapunov functions, etc.

- The problem is clearly convex in $V(x)$. But, how to solve this?

Lyapunov stability analysis

- A linear system $\dot{x} = Ax$, quadratic Lyapunov function $V(x) = x^T P x$

$$P \succ 0, \qquad A^T P + PA \prec 0$$

- More generally, for a nonlinear system $\dot{x} = f(x)$,

$$V(x) > 0 \quad x \neq 0, \qquad \dot{V}(x) = \left(\frac{\partial V}{\partial x}\right)^T f(x) < 0, \quad x \neq 0$$

  (locally, or globally if V is radially unbounded).

- Many variations: $\mathcal{H}_2$ and/or $\mathcal{H}_\infty$ analysis, parameter-dependent Lyapunov functions, etc.

- The problem is clearly convex in $V(x)$. But, how to solve this?

Lyapunov stability analysis

- A linear system $\dot{x} = Ax$, quadratic Lyapunov function $V(x) = x^T Px$

$$P \succ 0, \qquad A^T P + PA \prec 0$$

- More generally, for a nonlinear system $\dot{x} = f(x)$,

$$V(x) > 0 \quad x \neq 0, \qquad \dot{V}(x) = \left(\frac{\partial V}{\partial x}\right)^T f(x) < 0, \quad x \neq 0$$

  (locally, or globally if V is radially unbounded).

- Many variations: $\mathcal{H}_2$ and/or $\mathcal{H}_\infty$ analysis, parameter-dependent Lyapunov functions, etc.

- The problem is clearly convex in $V(x)$. But, how to solve this?

Nonnegativity is hard

What is the issue?

- Structure of V is unclear in general. Differentiable? Algebraic? Polynomial?
- More importantly, even if we nicely parameterize $V(x)$ (e.g., polynomials), how to verify the nonnegativity conditions?

$$V(x) > 0 \quad x \neq 0, \qquad \dot{V}(x) = \left(\frac{\partial V}{\partial x}\right)^T f(x) < 0, \quad x \neq 0$$

- Unfortunately, given a polynomial $p(x_1, \ldots, x_n)$, verifying if

$$p(x_1, \ldots, x_n) \geq 0 \qquad \forall \mathbf{x} \in \mathbb{R}^n$$

  is NP-hard (and also difficult in practice)

What to do about this?

Sum of squares

A multivariate polynomial $p(x)$ is a sum of squares (SOS) if

$$p(x) = \sum_i q_i^2(x), \quad q_i(x) \in \mathbb{R}[x].$$

- If $p(x)$ is SOS, then clearly $p(x) \geq 0$ for all $x \in \mathbb{R}^n$.
- Converse not true, in general (Hilbert). Counterexamples exist.
- For univariate or quadratics, nonnegativity is equivalent to SOS.

Let $P_{n,2d}$ be the set of nonnegative polynomials in n variables of degree less than or equal to $2d$, and $\Sigma_{n,2d}$ the corresponding set of SOS. Clearly,

$$\Sigma_{n,2d} \subseteq P_{n,2d}$$

Sum of squares

A multivariate polynomial $p(x)$ is a sum of squares (SOS) if

$$p(x) = \sum_i q_i^2(x), \quad q_i(x) \in \mathbb{R}[x].$$

- If $p(x)$ is SOS, then clearly $p(x) \geq 0$ for all $x \in \mathbb{R}^n$.
- Converse not true, in general (Hilbert). Counterexamples exist.
- For univariate or quadratics, nonnegativity is equivalent to SOS.

Let $P_{n,2d}$ be the set of nonnegative polynomials in n variables of degree less than or equal to $2d$, and $\Sigma_{n,2d}$ the corresponding set of SOS. Clearly,

$$\Sigma_{n,2d} \subseteq P_{n,2d}$$

Nonnegativity and sum of squares

In 1888, Hilbert showed that $P_{n,2d} = \Sigma_{n,2d}$ iff:

- $2d = 2$. Quadratic forms. SOS decomposition follows from eigenvalue/eigenvector, square root, or Cholesky decomposition.
- $n = 2$. Equivalent to polynomials in one variable.
- $2d = 4, n = 3$. Quartic forms in three variables.

Also, a nonconstructive proof of the nonequivalence in all other cases.

Years later, Motzkin gave an explicit counterexample:

$$M(x, y, z) = x^2y^4 + x^4y^2 + z^6 - 3x^2y^2z^2$$

- Is positive semidefinite. Apply the AGI to (x^2y^4, x^4y^2, z^6).
- Is *not* a sum of squares.

How do we check the sums of squares condition?

Checking the SOS condition

Basic "Gram matrix" method (Shor 87, Choi-Lam-Reznick 95, Powers-Wörmann 98, Nesterov, Lasserre, P., etc.)

A polynomial $F(x)$ is SOS if and only if

$$\boxed{F(x) = w(x)^T Q w(x),}$$

where $w(x)$ is a vector of monomials, and $Q \succeq 0$.

(If $F \in \mathbb{R}[x]_{n,2d}$, it is sufficient to choose for $w(x)$ all $\binom{n+d}{d}$ monomials of degree less than or equal to d.)

This is a semidefinite program!
Let's see an example, and then the general formulation...

Checking the SOS condition

Basic "Gram matrix" method (Shor 87, Choi-Lam-Reznick 95, Powers-Wörmann 98, Nesterov, Lasserre, P., etc.)

A polynomial $F(x)$ is SOS if and only if

$$\boxed{F(x) = w(x)^T Q w(x),}$$

where $w(x)$ is a vector of monomials, and $Q \succeq 0$.

(If $F \in \mathbb{R}[x]_{n,2d}$, it is sufficient to choose for $w(x)$ all $\binom{n+d}{d}$ monomials of degree less than or equal to d.)

This is a semidefinite program!
Let's see an example, and then the general formulation...

SOS Example

$$
\begin{aligned}
F(x,y) &= 2x^4 + 5y^4 - x^2y^2 + 2x^3y \\
&= \begin{bmatrix} x^2 \\ y^2 \\ xy \end{bmatrix}^T \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} \begin{bmatrix} x^2 \\ y^2 \\ xy \end{bmatrix} \\
&= q_{11}x^4 + q_{22}y^4 + (q_{33} + 2q_{12})x^2y^2 + 2q_{13}x^3y + 2q_{23}xy^3
\end{aligned}
$$

An SDP with equality constraints. Solving, we obtain:

$$
Q = \begin{bmatrix} 2 & -3 & 1 \\ -3 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix} = L^T L, \qquad L = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \end{bmatrix}
$$

And therefore $F(x,y) = \frac{1}{2}(2x^2 - 3y^2 + xy)^2 + \frac{1}{2}(y^2 + 3xy)^2$

Checking SOS via SDP

Let $F(x) = \sum f_\alpha x^\alpha$. Index rows and columns of Q by monomials. Then,

$$F(x) = w(x)^T Q w(x) \qquad \Leftrightarrow \qquad f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}$$

Thus, we have the SDP feasibility problem

$$\boxed{f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}, \qquad Q \succeq 0}$$

- Factorize $Q = L^T L$. The SOS is given by $F(x) = \|Lw(x)\|^2$.

(Can exploit sparsity, symmetry, etc. — more on this later).

And, we can actually search over such polynomials!

Checking SOS via SDP

Let $F(x) = \sum f_\alpha x^\alpha$. Index rows and columns of Q by monomials. Then,

$$F(x) = w(x)^T Q w(x) \qquad \Leftrightarrow \qquad f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}$$

Thus, we have the SDP feasibility problem

$$\boxed{f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}, \qquad Q \succeq 0}$$

- Factorize $Q = L^T L$. The SOS is given by $F(x) = \|Lw(x)\|^2$.

(Can exploit sparsity, symmetry, etc. — more on this later).

And, we can actually search over such polynomials!

Checking SOS via SDP

Let $F(x) = \sum f_\alpha x^\alpha$. Index rows and columns of Q by monomials. Then,

$$F(x) = w(x)^T Q w(x) \qquad \Leftrightarrow \qquad f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}$$

Thus, we have the SDP feasibility problem

$$\boxed{f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}, \qquad Q \succeq 0}$$

- Factorize $Q = L^T L$. The SOS is given by $F(x) = \|Lw(x)\|^2$.

(Can exploit sparsity, symmetry, etc. — more on this later).

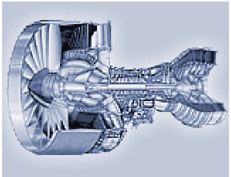
And, we can actually search over such polynomials!

Nonlinear Lyapunov

For $\dot{x} = f(x)$, a Lyapunov function must satisfy

$$V(x) \geq 0, \quad \left(\frac{\partial V}{\partial x}\right)^T f(x) \leq 0$$

Jet engine model (derived from Moore-Greitzer), with controller:



$$\dot{x} = -y + \frac{3}{2}x^2 - \frac{1}{2}x^3$$
$$\dot{y} = 3x - y$$

Postulate a generic 4th order polynomial Lyapunov function:

$$V(x,y) = \sum_{0 \leq j+k \leq 4} c_{jk}\, x^j y^k$$

Find a $V(x,y)$ by solving the SOS program:

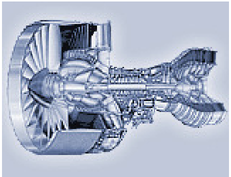
$$V(x,y) \text{ is SOS}, \qquad -\nabla V(x,y) \cdot f(x,y) \text{ is SOS}.$$

Nonlinear Lyapunov

For $\dot{x} = f(x)$, a Lyapunov function must satisfy

$$V(x) \geq 0, \quad \left(\frac{\partial V}{\partial x}\right)^T f(x) \leq 0$$

Jet engine model (derived from Moore-Greitzer), with controller:



$$\dot{x} = -y + \frac{3}{2}x^2 - \frac{1}{2}x^3$$
$$\dot{y} = 3x - y$$

Postulate a generic 4th order polynomial Lyapunov function:

$$V(x,y) = \sum_{0 \leq j+k \leq 4} c_{jk}\, x^j y^k$$

Find a $V(x,y)$ by solving the SOS program:

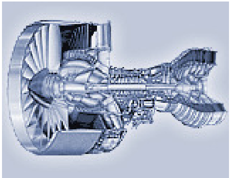
$$V(x,y) \text{ is SOS}, \qquad -\nabla V(x,y) \cdot f(x,y) \text{ is SOS}.$$

Nonlinear Lyapunov

For $\dot{x} = f(x)$, a Lyapunov function must satisfy

$$V(x) \geq 0, \quad \left(\frac{\partial V}{\partial x}\right)^T f(x) \leq 0$$

Jet engine model (derived from Moore-Greitzer), with controller:



$$\dot{x} = -y + \frac{3}{2}x^2 - \frac{1}{2}x^3$$
$$\dot{y} = 3x - y$$

Postulate a generic 4th order polynomial Lyapunov function:

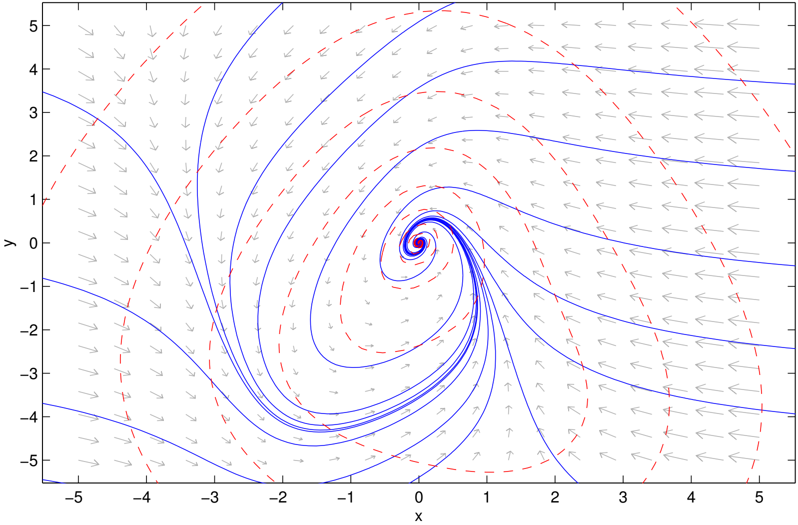
$$V(x, y) = \sum_{0 \leq j+k \leq 4} c_{jk}\, x^j y^k$$

Find a $V(x, y)$ by solving the SOS program:

$$V(x, y) \text{ is SOS}, \qquad -\nabla V(x, y) \cdot f(x, y) \text{ is SOS}.$$

Lyapunov example (cont.)

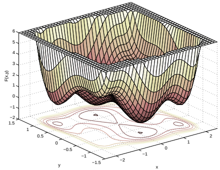
After solving, we obtain a Lyapunov function.



From feasibility to optimization

SOS directly yields lower bounds for optimization!

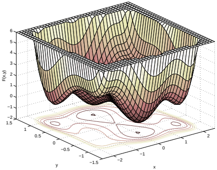
$$F(x) - \gamma \text{ is SOS} \quad \Rightarrow \quad F(x) \geq \gamma \text{ for all } x$$



- Finding the best such γ is also an SDP
- Typically, very high-quality bounds
- If exact, can recover exact solution
- Natural extensions to constrained case

From feasibility to optimization

SOS directly yields lower bounds for optimization!



$$F(x) - \gamma \text{ is SOS} \quad \Rightarrow \quad F(x) \geq \gamma \ \text{ for all } x$$

- Finding the best such γ is also an SDP
- Typically, very high-quality bounds
- If exact, can recover exact solution
- Natural extensions to constrained case

Polynomial systems over $\mathbb{R}$

- When do equations and inequalities have real solutions?
- A remarkable answer: the Positivstellensatz.
- Centerpiece of real algebraic geometry (Stengle 1974).
- Common generalization of Hilbert's Nullstellensatz and LP duality.
- Guarantees the existence of algebraic infeasibility certificates for real solutions of systems of polynomial equations.
- Sums of squares are a fundamental ingredient.

How does it work?

P-satz and SOS

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$, want to *prove* that it is empty.

Define:

$$\mathrm{Cone}(f_i) = \sum s_i \cdot (\prod_j f_j), \qquad \mathrm{Ideal}(h_i) = \sum t_i \cdot h_i,$$

where the $s_i, t_i \in \mathbb{R}[x]$ and the s_i are sums of squares.

What is this? What's the idea?

Want to capture the algebraic structure of the allowable operations among constraints (alternatively, how to generate new constraints from old ones):

- If $f_i(x) \geq 0$, $f_j(x) \geq 0$, then $f_i(x)f_j(x) \geq 0$.
- If $f_i(x) \geq 0$, then $s(x)f_i(x) \geq 0$, where $s(x)$ is SOS.
- If $h_i(x) = 0$, then $t(x)h_i(x) = 0$.

Describes *all* valid constraints that can be "easily" generated.

P-satz and SOS

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$, want to *prove* that it is empty. Define:

$$\mathrm{Cone}(f_i) = \sum s_i \cdot (\prod_j f_j), \qquad \mathrm{Ideal}(h_i) = \sum t_i \cdot h_i,$$

where the $s_i, t_i \in \mathbb{R}[x]$ and the s_i are sums of squares.

What is this? What's the idea?

Want to capture the algebraic structure of the allowable operations among constraints (alternatively, how to generate new constraints from old ones):

- If $f_i(x) \geq 0$, $f_j(x) \geq 0$, then $f_i(x)f_j(x) \geq 0$.
- If $f_i(x) \geq 0$, then $s(x)f_i(x) \geq 0$, where $s(x)$ is SOS.
- If $h_i(x) = 0$, then $t(x)h_i(x) = 0$.

Describes *all* valid constraints that can be "easily" generated.

P-satz and SOS

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$, want to *prove* that it is empty.
Define:

$$\text{Cone}(f_i) = \sum s_i \cdot (\prod_j f_j), \qquad \text{Ideal}(h_i) = \sum t_i \cdot h_i,$$

where the $s_i, t_i \in \mathbb{R}[x]$ and the s_i are sums of squares.
What is this? What's the idea?
Want to capture the algebraic structure of the allowable operations among constraints (alternatively, how to generate new constraints from old ones):

- If $f_i(x) \geq 0$, $f_j(x) \geq 0$, then $f_i(x)f_j(x) \geq 0$.
- If $f_i(x) \geq 0$, then $s(x)f_i(x) \geq 0$, where $s(x)$ is SOS.
- If $h_i(x) = 0$, then $t(x)h_i(x) = 0$.

Describes *all* valid constraints that can be "easily" generated.

P-satz and SOS

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$, want to *prove* that it is empty.
Define:

$$\text{Cone}(f_i) = \sum s_i \cdot (\prod_j f_j), \qquad \text{Ideal}(h_i) = \sum t_i \cdot h_i,$$

where the $s_i, t_i \in \mathbb{R}[x]$ and the s_i are sums of squares.
What is this? What's the idea?
Want to capture the algebraic structure of the allowable operations among constraints (alternatively, how to generate new constraints from old ones):

- If $f_i(x) \geq 0$, $f_j(x) \geq 0$, then $f_i(x) f_j(x) \geq 0$.
- If $f_i(x) \geq 0$, then $s(x) f_i(x) \geq 0$, where $s(x)$ is SOS.
- If $h_i(x) = 0$, then $t(x) h_i(x) = 0$.

Describes *all* valid constraints that can be "easily" generated.

P-satz and SOS

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$, want to *prove* that it is empty.
Define:

$$\mathrm{Cone}(f_i) = \sum s_i \cdot (\prod_j f_j), \qquad \mathrm{Ideal}(h_i) = \sum t_i \cdot h_i,$$

where the $s_i, t_i \in \mathbb{R}[x]$ and the s_i are sums of squares.
To prove infeasibility, find $f \in \mathrm{Cone}(f_i), h \in \mathrm{Ideal}(h_i)$ such that

$$f + h = -1.$$

- Can find certificates by solving SOS programs!
- Complete SOS hierarchy, by certificate degree (P. 2000).
- Directly provides hierarchies of bounds for optimization.

P-satz and SOS

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$, want to *prove* that it is empty.
Define:

$$\mathrm{Cone}(f_i) = \sum s_i \cdot (\prod_j f_j), \qquad \mathrm{Ideal}(h_i) = \sum t_i \cdot h_i,$$

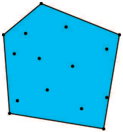
where the $s_i, t_i \in \mathbb{R}[x]$ and the s_i are sums of squares.
To prove infeasibility, find $f \in \mathrm{Cone}(f_i), h \in \mathrm{Ideal}(h_i)$ such that

$$f + h = -1.$$

- Can find certificates by solving SOS programs!
- Complete SOS hierarchy, by certificate degree (P. 2000).
- Directly provides hierarchies of bounds for optimization.

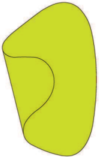
Convex hulls of algebraic varieties

Back to SDP representations...
Focus here on a specific, but very important case.



Given a set $S \subset \mathbb{R}^n$, we can define its *convex hull*

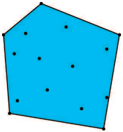
$$\mathrm{conv} S := \left\{ \sum_i \lambda_i x_i \,:\, x_i \in S, \sum_i \lambda_i = 1, \lambda_i \geq 0 \right\}$$



Our interest: S is a real algebraic variety and the approximation is an SDP-representable set.

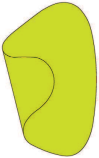
Convex hulls of algebraic varieties

Back to SDP representations...
Focus here on a specific, but very important case.



Given a set $S \subset \mathbb{R}^n$, we can define its *convex hull*

$$\mathrm{conv} S := \left\{ \sum_i \lambda_i x_i \ : \ x_i \in S, \sum_i \lambda_i = 1, \ \lambda_i \geq 0 \right\}$$



Our interest: S is a real algebraic variety and the approximation is an SDP-representable set.

Why?

Many interesting problems require or boil down *exactly* to understanding and describing convex hulls of algebraic varieties.

- Nonnegative polynomials and optimization
- Polynomial games
- Convex relaxations for minimum-rank
- Convex hull of rotation matrices

We'll discuss some of these in detail...

Polynomial optimization

Consider the unconstrained minimization of a multivariate polynomial

$$p(x) = \sum_{\alpha \in S} p_\alpha x^\alpha,$$

where $x \in \mathbb{R}^n$ and S is a given set of monomials (e.g., all monomials of total degree less than or equal to $2d$, in the dense case).

Define the (real, toric) algebraic variety $V_S \subset \mathbb{R}^{|S|}$:

$$V_S := \left\{ (x^{\alpha_1}, \ldots, x^{\alpha_{|S|}}) \,:\, x \in \mathbb{R}^n \right\}.$$

This is the image of $\mathbb{R}^n$ under the monomial map (e.g., in the homogeneous case, the Veronese embedding).
Want to study the *convex hull* of V_S. Extends to the constrained case.

Polynomial optimization

Consider the unconstrained minimization of a multivariate polynomial

$$p(x) = \sum_{\alpha \in S} p_\alpha x^\alpha,$$

where $x \in \mathbb{R}^n$ and S is a given set of monomials (e.g., all monomials of total degree less than or equal to $2d$, in the dense case).

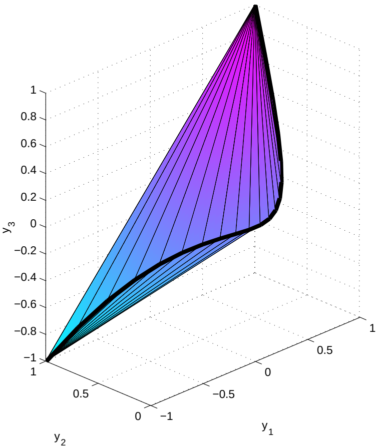
Define the (real, toric) algebraic variety $V_S \subset \mathbb{R}^{|S|}$:

$$V_S := \left\{ \left(x^{\alpha_1}, \ldots, x^{\alpha_{|S|}}\right) \,:\, x \in \mathbb{R}^n \right\}.$$

This is the image of $\mathbb{R}^n$ under the monomial map (e.g., in the homogeneous case, the Veronese embedding).
Want to study the *convex hull* of V_S. Extends to the constrained case.

Univariate case

Convex hull of the rational normal curve $(1, t, \dots, t^d)$.
Not polyhedral.
Known geometry (Karlin-Shapley)



"Simplicial": every supporting hyperplane yields a simplex.
Related to cyclic polytopes.

Polynomial optimization

We have then (almost trivially):

$$\inf_{x \in \mathbb{R}^n} p(x) = \inf\{p^T y \,:\, y \in \mathrm{conv}\, V_S\}$$

Optimizing a nonconvex polynomial is equivalent to linear optimization over a convex set (!)

Unfortunately, in general, it is NP-hard to check membership in $\mathrm{conv}\, V_S$. Nevertheless, we can turn this around, and use SOS relaxations to obtain "good" approximate SDP descriptions of the convex hull V_S.

Polynomial optimization

We have then (almost trivially):

$$\inf_{x \in \mathbb{R}^n} p(x) = \inf\{p^T y \,:\, y \in \operatorname{conv} V_S\}$$

Optimizing a nonconvex polynomial is equivalent to linear optimization over a convex set (!)

Unfortunately, in general, it is NP-hard to check membership in $\operatorname{conv} V_S$. Nevertheless, we can turn this around, and use SOS relaxations to obtain "good" approximate SDP descriptions of the convex hull V_S.

A geometric interlude

How is this possible? Convex optimization for solving nonconvex problems?

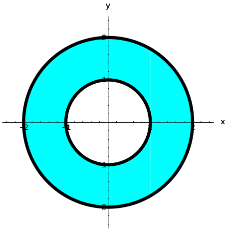
Convexity is *relative*. Every problem can be trivially "lifted" to a convex setting (in general, infinite dimensional).

Ex: mixed strategies in games, "relaxed" controls, Fokker-Planck, etc.

Interestingly, however, often a finite (and small) dimension is enough.

Consider the set defined by

$$1 \leq x^2 + y^2 \leq 2$$



Clearly non-convex.

Can we use convex optimization?

A geometric interlude

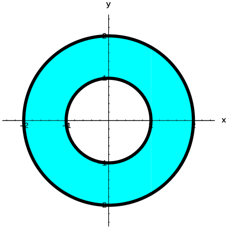
How is this possible? Convex optimization for solving nonconvex problems?

Convexity is *relative*. Every problem can be trivially "lifted" to a convex setting (in general, infinite dimensional).
Ex: mixed strategies in games, "relaxed" controls, Fokker-Planck, etc.
Interestingly, however, often a finite (and small) dimension is enough.

Consider the set defined by

$$1 \leq x^2 + y^2 \leq 2$$

Clearly non-convex.
Can we use convex optimization?



A geometric interlude

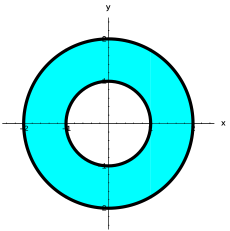
How is this possible? Convex optimization for solving nonconvex problems?

Convexity is *relative*. Every problem can be trivially "lifted" to a convex setting (in general, infinite dimensional).
Ex: mixed strategies in games, "relaxed" controls, Fokker-Planck, etc.
Interestingly, however, often a finite (and small) dimension is enough.

Consider the set defined by

$$1 \leq x^2 + y^2 \leq 2$$

Clearly non-convex.
Can we use convex optimization?



Geometric interpretation

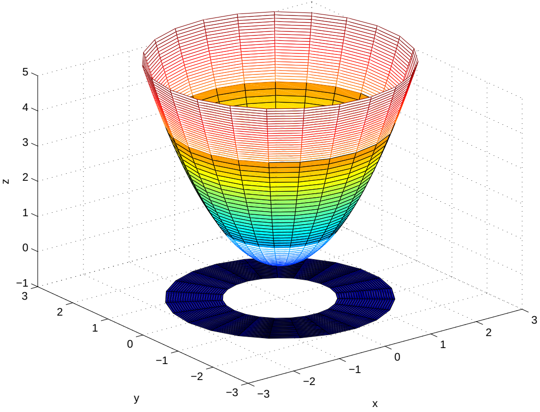
A polynomial “lifting” to a higher dimensional space:

$$(x, y) \mapsto (x, y, x^2 + y^2)$$

The nonconvex set is the projection of the extreme points of a convex set.

In particular, the convex set defined by

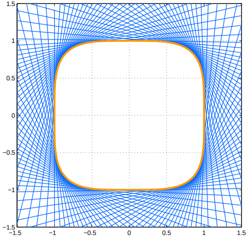
$$x^2 + y^2 \leq z$$
$$1 \leq z \leq 4$$



Convex hull of varieties: a "polar" viewpoint

How to describe a convex hull?

Any convex set $\mathcal{S}$ is uniquely defined by its supporting hyperplanes.



Thus, if we can optimize a *linear function* over a set using SDP, we effectively have an SDP representation.

Need to solve (or approximate)

$$\min c^T x \qquad \text{s.t. } x \in \mathcal{S}$$

If $\mathcal{S}$ is defined by polynomial equations/inequalities, can use SOS.

Theta bodies

We define the k-th *theta body* of a real variety (Gouveia-P.-Thomas 08).

Let V be an algebraic variety, and $I = I(V) \subseteq \mathbb{R}[x_1, \ldots, x_n]$ the associated polynomial ideal. The polynomial f is k-**sos modulo the ideal** I if

$$f = \sum_i q_i^2 \quad \forall x \in V, \qquad \deg(q_i) \leq k.$$

If f is k-sos mod I, then clearly f is nonnegative on V.

Recall the characterization of the (closed) convex hull of a set S as the intersection of all half-spaces that contain S:

$$\overline{\operatorname{conv}(S)} = \{\mathbf{p} \in \mathbb{R}^n \,:\, f(\mathbf{p}) \geq 0 \text{ for all } f \text{ affine and nonnegative on } S\}$$

Next, we will do the same, but replacing nonnegativity with k-sos.

Theta bodies

We define the k-th *theta body* of a real variety (Gouveia-P.-Thomas 08).

Let V be an algebraic variety, and $I = I(V) \subseteq \mathbb{R}[x_1, \ldots, x_n]$ the associated polynomial ideal. The polynomial f is k-**sos modulo the ideal** I if

$$f = \sum_i q_i^2 \quad \forall x \in V, \qquad \deg(q_i) \leq k.$$

If f is k-sos mod I, then clearly f is nonnegative on V.

Recall the characterization of the (closed) convex hull of a set S as the intersection of all half-spaces that contain S:

$$\overline{\mathrm{conv}(S)} = \{\mathbf{p} \in \mathbb{R}^n \,:\, f(\mathbf{p}) \geq 0 \text{ for all } f \text{ affine and nonnegative on } S\}$$

Next, we will do the same, but replacing nonnegativity with k-sos.

Theta bodies

We define the k-th *theta body* of a real variety (Gouveia-P.-Thomas 08).

Let V be an algebraic variety, and $I = I(V) \subseteq \mathbb{R}[x_1, \ldots, x_n]$ the associated polynomial ideal. The polynomial f is k-**sos modulo the ideal** I if

$$f = \sum_i q_i^2 \quad \forall x \in V, \qquad \deg(q_i) \leq k.$$

If f is k-sos mod I, then clearly f is nonnegative on V.

Recall the characterization of the (closed) convex hull of a set S as the intersection of all half-spaces that contain S:

$$\overline{\mathrm{conv}(S)} = \{\mathbf{p} \in \mathbb{R}^n \;:\; f(\mathbf{p}) \geq 0 \text{ for all } f \text{ affine and nonnegative on } S\}$$

Next, we will do the same, but replacing nonnegativity with k-sos.

Theta bodies (II)

Def: Given an ideal $I \subset \mathbb{R}[x_1, \ldots, x_n]$ we define its k-th theta body:

$$\mathrm{TH}_k(I) := \{\mathbf{p} \in \mathbb{R}^n \,:\, f(\mathbf{p}) \geq 0 \text{ for all } f \text{ affine and } k\text{-sos mod } I\}$$

Replace all halfspaces with "k-sos certifiable" halfspaces.

Since

$$\overline{\mathrm{conv}(S)} = \{\mathbf{p} \in \mathbb{R}^n \,:\, f(\mathbf{p}) \geq 0 \text{ for all } f \text{ affine and nonnegative on } S\}$$

We have then

$$\overline{\mathrm{conv}(V_{\mathbb{R}}(I))} \subseteq \cdots \subseteq \mathrm{TH}_k(I) \subseteq \mathrm{TH}_{k-1}(I) \subseteq \cdots \subseteq \mathrm{TH}_1(I).$$

Under mild conditions, $\lim_{k\to\infty} TH_k(I)$ gives the desired convex hull.
Often, finite convergence.

Theta bodies (II)

Def: Given an ideal $I \subset \mathbb{R}[x_1, \ldots, x_n]$ we define its k-th theta body:

$$\mathrm{TH}_k(I) := \{\mathbf{p} \in \mathbb{R}^n \,:\, f(\mathbf{p}) \geq 0 \text{ for all } f \text{ affine and } k\text{-sos mod } I\}$$

Replace all halfspaces with "k-sos certifiable" halfspaces.

Since

$$\overline{\mathrm{conv}(S)} = \{\mathbf{p} \in \mathbb{R}^n \,:\, f(\mathbf{p}) \geq 0 \text{ for all } f \text{ affine and nonnegative on } S\}$$

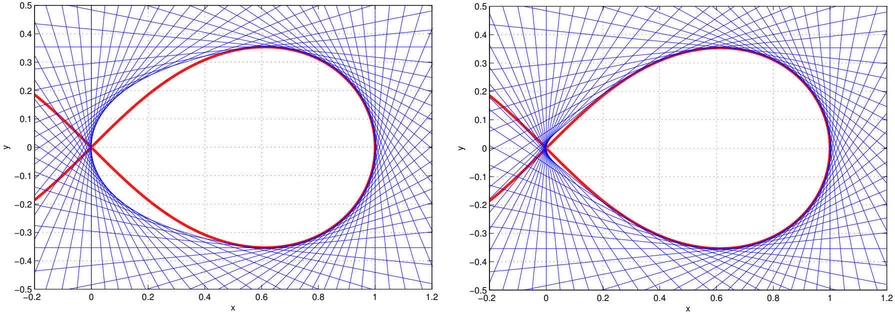
We have then

$$\overline{\mathrm{conv}(V_{\mathbb{R}}(I))} \subseteq \cdots \subseteq \mathrm{TH}_k(I) \subseteq \mathrm{TH}_{k-1}(I) \subseteq \cdots \subseteq \mathrm{TH}_1(I).$$

Under mild conditions, $\lim_{k\to\infty} TH_k(I)$ gives the desired convex hull. Often, finite convergence.

Example

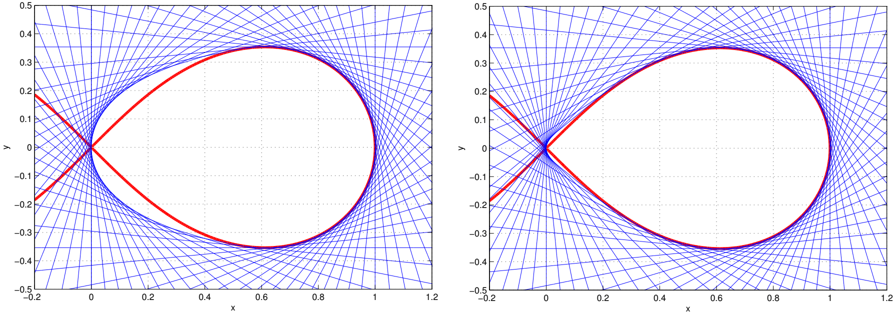
SDP approximations to a (half) lemniscate:



Many variations (e.g., Putinar-Lasserre, LP-based, sparse SOS, etc.), same basic ideas.

Example

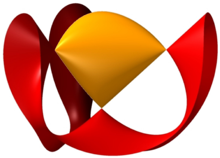
SDP approximations to a (half) lemniscate:



Many variations (e.g., Putinar-Lasserre, LP-based, sparse SOS, etc.), same basic ideas.

Summary

- Sum of squares allows the use of SDP for polynomial problems
- Through the P-satz, extend to constrained problems
- Convexity properties depend on description
- Convex hulls can be nicely approximated by SDP



End of Part II

Summary

- Sum of squares allows the use of SDP for polynomial problems
- Through the P-satz, extend to constrained problems
- Convexity properties depend on description
- Convex hulls can be nicely approximated by SDP

End of Part II

Part III

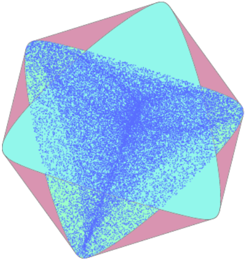
Example: orthogonal matrices

Consider $O(3)$, the group of 3×3 orthogonal matrices. It has two connected components (determinant is ± 1). Rotation matrices have *determinant one* (preserve orientation).

Can use the double-cover of $SO(3)$ with $SU(2)$ (equivalently, quaternions) to provide an exact SDP representation of the convex hull of $SO(3)$:

$$\begin{bmatrix} Z_{11}+Z_{22}-Z_{33}-Z_{44} & 2Z_{23}-2Z_{14} & 2Z_{24}+2Z_{13} \\ 2Z_{23}+2Z_{14} & Z_{11}-Z_{22}+Z_{33}-Z_{44} & 2Z_{34}-2Z_{12} \\ 2Z_{24}-2Z_{13} & 2Z_{34}+2Z_{12} & Z_{11}-Z_{22}-Z_{33}+Z_{44} \end{bmatrix}, \quad Z \succeq 0, \quad \operatorname{Tr} Z = 1.$$

This is a convex set in $\mathbb{R}^9$.
Here is a two-dimensional projection.



Generalizations to SO(n) via Clifford algebras
(Saunderson-P.-Willsky, arXiv:1403.4914,
SIAM J. Optim. 25:3, 1314–1343, 2015.)

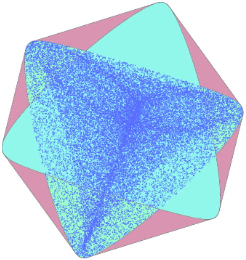
Example: orthogonal matrices

Consider $O(3)$, the group of 3×3 orthogonal matrices. It has two connected components (determinant is ± 1). Rotation matrices have *determinant one* (preserve orientation).

Can use the double-cover of $SO(3)$ with $SU(2)$ (equivalently, quaternions) to provide an exact SDP representation of the convex hull of $SO(3)$:

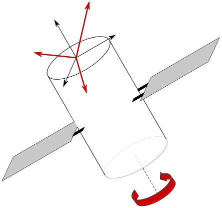
$$\begin{bmatrix} Z_{11}+Z_{22}-Z_{33}-Z_{44} & 2Z_{23}-2Z_{14} & 2Z_{24}+2Z_{13} \\ 2Z_{23}+2Z_{14} & Z_{11}-Z_{22}+Z_{33}-Z_{44} & 2Z_{34}-2Z_{12} \\ 2Z_{24}-2Z_{13} & 2Z_{34}+2Z_{12} & Z_{11}-Z_{22}-Z_{33}+Z_{44} \end{bmatrix}, \quad Z \succeq 0, \quad \mathrm{Tr}\, Z = 1.$$

This is a convex set in $\mathbb{R}^9$.
Here is a two-dimensional projection.



Generalizations to SO(n) via Clifford algebras (Saunderson-P.-Willsky, `arXiv:1403.4914`, SIAM J. Optim. 25:3, 1314–1343, 2015.)

Joint spin-rate and attitude estimation



- unknown initial 'attitude' Q
- spinning at unknown rate ω around known (in body frame) axis

Data: sequence of noisy measurements (in body frame) of reference directions (sun, stars, magnetic field, etc) known in inertial frame

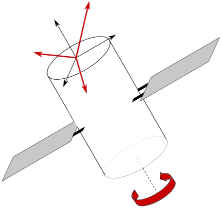
Problem: Estimate initial attitude Q and spin-rate ω from data.

$$\max_{\substack{Q \in SO(3) \\ \omega \in [-\pi, \pi)}} \sum_{n=0}^{N} \left[\langle A_n, Q \cos(n\omega) \rangle + \langle B_n, Q \sin(n\omega) \rangle \right]$$

Representation allow us to exactly solve this problem with SDP!

(arXiv:1410.2841, J. Guidance, Control, and Dynamics, 39:1, 118-127, 2016.)

Joint spin-rate and attitude estimation



- unknown initial 'attitude' Q
- spinning at unknown rate ω around known (in body frame) axis

Data: sequence of noisy measurements (in body frame) of reference directions (sun, stars, magnetic field, etc) known in inertial frame

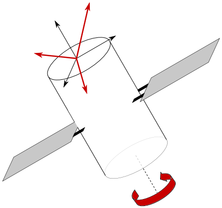
Problem: Estimate initial attitude Q and spin-rate ω from data.

$$\max_{\substack{Q \in SO(3) \\ \omega \in [-\pi, \pi)}} \sum_{n=0}^{N} \left[\langle A_n, Q \cos(n\omega) \rangle + \langle B_n, Q \sin(n\omega) \rangle \right]$$

Representation allow us to exactly solve this problem with SDP!

(arXiv:1410.2841, J. Guidance, Control, and Dynamics, 39:1, 118-127, 2016.)

Joint spin-rate and attitude estimation



- unknown initial 'attitude' Q
- spinning at unknown rate ω around known (in body frame) axis

Data: sequence of noisy measurements (in body frame) of reference directions (sun, stars, magnetic field, etc) known in inertial frame

Problem: Estimate initial attitude Q and spin-rate ω from data.

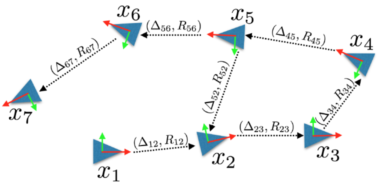
$$\max_{\substack{Q \in SO(3) \\ \omega \in [-\pi, \pi)}} \sum_{n=0}^{N} \left[\langle A_n, Q \cos(n\omega) \rangle + \langle B_n, Q \sin(n\omega) \rangle \right]$$

Representation allow us to exactly solve this problem with SDP!
(`arXiv:1410.2841`, J. Guidance, Control, and Dynamics, 39:1, 118-127, 2016.)

Pose graph optimization

e.g., Bandeira-Kennedy-Singer, Carlone-Rosen-Calafiore-Leonard-Dellaert, ...

- Collection of rigid bodies (e.g., drones w/cameras, SLAM)
- (Few) measurements of pairwise relative positions M_{ij}
- Estimate the position of *all* bodies



(figure from Calafiore *et al.*)

$$\min_{\{R_i\} \in SO(3)} \sum_{(ij) \in \mathcal{M}} \|M_{ij} - R_i R_j^T\|^2$$

Natural semidefinite relaxation:

$$\min \sum_{(ij) \in \mathcal{M}} \|M_{ij} - R_{ij}\|^2 \qquad \text{s.t.} \begin{bmatrix} I_3 & R_{12} & \dots & R_{1n} \\ R_{12}^T & I_3 & \dots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1n}^T & R_{2n}^T & \dots & I_3 \end{bmatrix} \succeq 0, \quad R_{ij} \in \text{conv} SO(3).$$

Minimum rank and convex relaxations

Consider the rank minimization problem

$$\text{minimize } \operatorname{rank} X \qquad \text{subject to } \mathcal{A}(X) = b,$$

where $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is a linear map.

Find the minimum-rank matrix in a given subspace. In general, NP-hard.

Let's see a few applications...

Minimum rank and convex relaxations

Consider the rank minimization problem

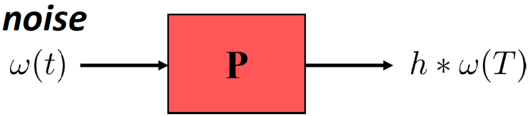
$$\text{minimize } \operatorname{rank} X \qquad \text{subject to } \mathcal{A}(X) = b,$$

where $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is a linear map.

Find the minimum-rank matrix in a given subspace. In general, NP-hard.

Let's see a few applications...

Application: System identification

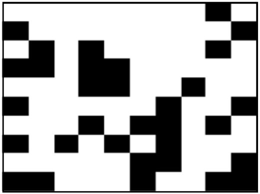


- Measure response at time T (e.g., for random input)
- Response at time T is linear in impulse response h.

$$\text{hank}(h) := \begin{bmatrix} h(0) & h(1) & \cdots & h(N) \\ h(1) & h(2) & \cdots & h(N+1) \\ \vdots & \vdots & \ddots & \vdots \\ h(N) & h(N+1) & \cdots & h(2N) \end{bmatrix}$$

- Complexity of $P \approx \text{rank}(\text{hank}(h))$.

Application: Matrix completion

$M =$ 

M_{ij} known for black cells
M_{ij} unknown for white cells

- Partially specified matrix, known pattern
- Often, random sampling of entries
- Applications:
  - Partially specified covariances (PSD case)
  - Collaborative prediction (e.g., Rennie-Srebro'05, "Netflix problem", Candés-Recht'09)

Minimum rank and convex relaxations

Consider the rank minimization problem

$$\text{minimize } \operatorname{rank} X \qquad \text{subject to } \mathcal{A}(X) = b,$$

where $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is a linear map.
Find the minimum-rank matrix in a given subspace. In general, NP-hard.

Since rank is hard, let's use instead its *convex envelope*, the nuclear norm. The nuclear norm of a matrix (alternatively, Schatten 1-norm, Ky Fan r-norm, or trace class norm) is the sum of its singular values, i.e.,

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X).$$

Minimum rank and convex relaxations

Consider the rank minimization problem

$$\text{minimize } \operatorname{rank} X \qquad \text{subject to } \mathcal{A}(X) = b,$$

where $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map.
Find the minimum-rank matrix in a given subspace. In general, NP-hard.

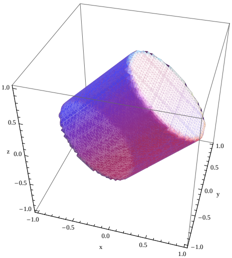
Since rank is hard, let's use instead its *convex envelope*, the nuclear norm.
The nuclear norm of a matrix (alternatively, Schatten 1-norm, Ky Fan r-norm, or trace class norm) is the sum of its singular values, i.e.,

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X).$$

Convex hulls and nuclear norm

Nuclear norm ball is convex hull of rank one matrices:

$$B = \{X \in \mathbb{R}^{m\times n} \; : \; \|X\|_* \leq 1\}$$
$$= \mathrm{conv}\{uv^T \; : \; u \in \mathbb{R}^m, v \in \mathbb{R}^n, \|u\|^2 = 1, \|v\|^2 = 1\}$$



Exactly SDP-characterizable!

$$B = \left\{ X \in \mathbb{R}^{m\times n} : \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0, \qquad \mathrm{Tr}\, W_1 + \mathrm{Tr}\, W_2 = 2 \right\}$$

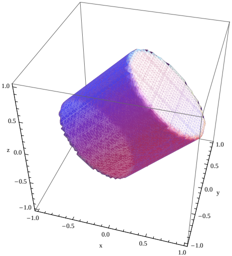
Under certain conditions (e.g., if subspace defined by $\mathcal{A}$ is "random"), optimizing the nuclear norm yields the *true minimum rank solution*.

For details, see Recht-Fazel-P., "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, 2010.

Convex hulls and nuclear norm

Nuclear norm ball is convex hull of rank one matrices:

$$B = \{X \in \mathbb{R}^{m\times n} \,:\, \|X\|_* \leq 1\}$$
$$= \mathrm{conv}\{uv^T \,:\, u \in \mathbb{R}^m, v \in \mathbb{R}^n, \|u\|^2 = 1, \|v\|^2 = 1\}$$



Exactly SDP-characterizable!

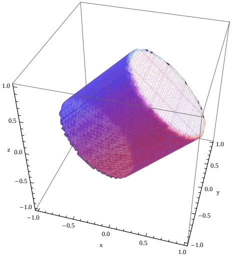
$$B = \left\{ X \in \mathbb{R}^{m\times n} : \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0, \qquad \mathrm{Tr}\, W_1 + \mathrm{Tr}\, W_2 = 2 \right\}$$

Under certain conditions (e.g., if subspace defined by $\mathcal{A}$ is "random"), optimizing the nuclear norm yields the *true minimum rank solution*.

For details, see Recht-Fazel-P., "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, 2010.

Convex hulls and nuclear norm

Nuclear norm ball is convex hull of rank one matrices:



$$
\begin{aligned}
B &= \{X \in \mathbb{R}^{m\times n} \,:\, \|X\|_* \leq 1\} \\
&= \mathrm{conv}\{uv^T \,:\, u \in \mathbb{R}^m, v \in \mathbb{R}^n, \|u\|^2 = 1, \|v\|^2 = 1\}
\end{aligned}
$$

Exactly SDP-characterizable!

$$
B = \left\{ X \in \mathbb{R}^{m\times n} : \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0, \qquad \mathrm{Tr}\, W_1 + \mathrm{Tr}\, W_2 = 2 \right\}
$$

Under certain conditions (e.g., if subspace defined by $\mathcal{A}$ is "random"), optimizing the nuclear norm yields the *true minimum rank solution*.

For details, see Recht-Fazel-P., "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, 2010.

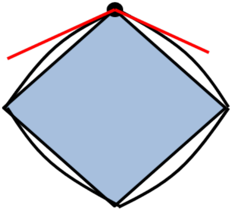
Rank, sparsity, and beyond: atomic norms

Exactly the same constructions can be applied to more general situations: atomic norms.

Structure-inducing regularizer is convex hull of atom set, e.g., low-rank matrices/tensors, permutation matrices, cut matrices, etc.

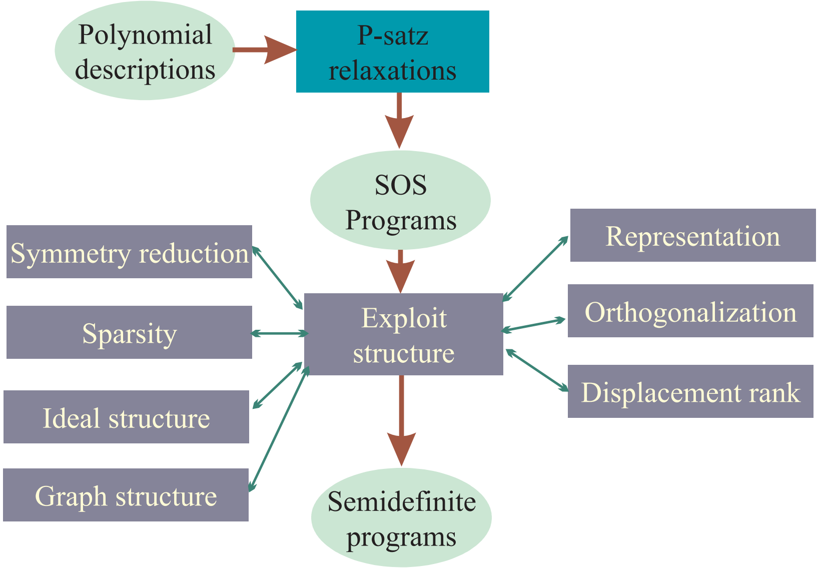
Generally NP-hard to compute, but good SDP approximations.

Statistical guarantees for recovery based on *Gaussian width of tangent cones*. Interesting interplay between computational and sample complexities.



For details, see Chandrasekaran-Recht-P.-Willsky, "The convex geometry of linear inverse problems," *Found. Comp. Math.*, 2012.

Exploiting structure



Algebraic structure

- Algebraic sparsity: polynomials with few nonzero coefficients.
  - Newton polytopes techniques.
- Ideal structure: equality constraints.
  - SOS on *quotient rings* $\mathbb{R}[x]/I$.
  - Compute in the coordinate ring. Quotient bases.
- Graphical structure:
  - Dependency graph among the variables
  - Chordality/treewidth techniques
- Symmetries: invariance under a group
  - SOS on *invariant rings*
  - Representation theory and invariant-theoretic methods.
  - Enabling factor in applications (e.g., Markov chains)

Numerical structure

- Rank one SDPs.
  - Dual coordinate change makes all constraints rank one
  - Efficient computation of Hessians and gradients
- Representations
  - Interpolation representation
  - Orthogonalization
- Displacement rank
  - Fast solvers for search direction
- Alternatives to interior-point methods?
  - E.g., factorization approaches (Burer-Monteiro)?

Related work

(very incomplete/partial list!)

- Related basic work: N.Z. Shor, Nesterov, Lasserre, etc.
- Systems and control (Prajna, Rantzer, Hol-Scherer, Henrion, etc.)
- Sparse optimization (Waki-Kim-Kojima-Muramatsu, Lasserre, Nie-Demmel, etc.)
- Approximation algorithms (de Klerk-Laurent-P.)
- Filter design (Alkire-Vandenberghe, Hachez-Nesterov, etc.)
- Stability number of graphs (Laurent, Peña, Rendl)
- Quantum information theory (Doherty-Spedalieri-P., Childs-Landahl-P.)
- Joint spectral radius (P.-Jadbabaie, Legat-Jungers)
- Game theory (Stein-Ozdaglar-P.)
- Theoretical computer science (Barak, Kelner, Steurer, Lee, Raghavendra)

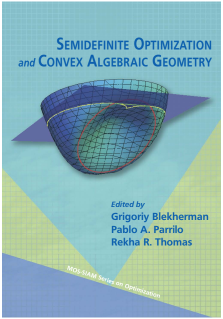
Connections

Many fascinating links to other areas of mathematics:

- Probability (moments, exchangeability and de Finetti, etc)
- Operator theory (via Gelfand-Neimark-Segal)
- Harmonic analysis on semigroups
- Noncommutative probability and quantum information
- Complexity and proof theory (degrees of certificates)
- Graph theory (perfect graphs)
- Tropical geometry (SDP over more general fields)

Summary

- A very rich class of optimization problems
- Methods have enabled many new applications
- Interplay of many branches of mathematics
- Structure must be exploited for reliability and efficiency
- Combination of numerical and algebraic techniques.



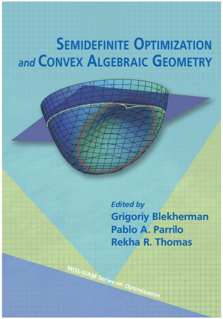
If you want to know more:

- Papers, slides, lecture notes, software, etc.: www.mit.edu/~parrilo
- NSF FRG: "SDP and convex algebraic geometry"
- NSF AF: "Algebraic proof systems, convexity, and algorithms"

Thanks for your attention!

Summary

- A very rich class of optimization problems
- Methods have enabled many new applications
- Interplay of many branches of mathematics
- Structure must be exploited for reliability and efficiency
- Combination of numerical and algebraic techniques.



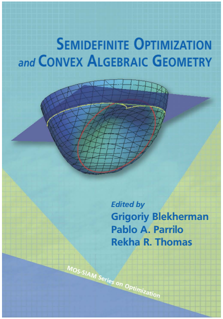
If you want to know more:

- Papers, slides, lecture notes, software, etc.: `www.mit.edu/~parrilo`
- NSF FRG: "SDP and convex algebraic geometry"
- NSF AF: "Algebraic proof systems, convexity, and algorithms"

Thanks for your attention!

Summary

- A very rich class of optimization problems
- Methods have enabled many new applications
- Interplay of many branches of mathematics
- Structure must be exploited for reliability and efficiency
- Combination of numerical and algebraic techniques.



If you want to know more:

- Papers, slides, lecture notes, software, etc.: `www.mit.edu/~parrilo`
- NSF FRG: “SDP and convex algebraic geometry”
- NSF AF: “Algebraic proof systems, convexity, and algorithms”

Thanks for your attention!