Certificates, convex optimization, and their applications

Pablo A. Parrilo
Institut für Automatik
ETH Zürich

http://www.aut.ee.ethz.ch/~parrilo

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Outline

- Convex optimization, semidefinite programming.
- Nonnegativity of polynomials.
- Applications:
  - Global optimization, Lyapunov functions, Quantum entanglement.
- Emptiness of sets. The role of certificates.
- Sums of squares and the P-satz. A convex approach.
- Applications:
  - Robust bifurcation, combinatorial optimization.
  - System analysis, geometric theorem proving.
- Conclusions.

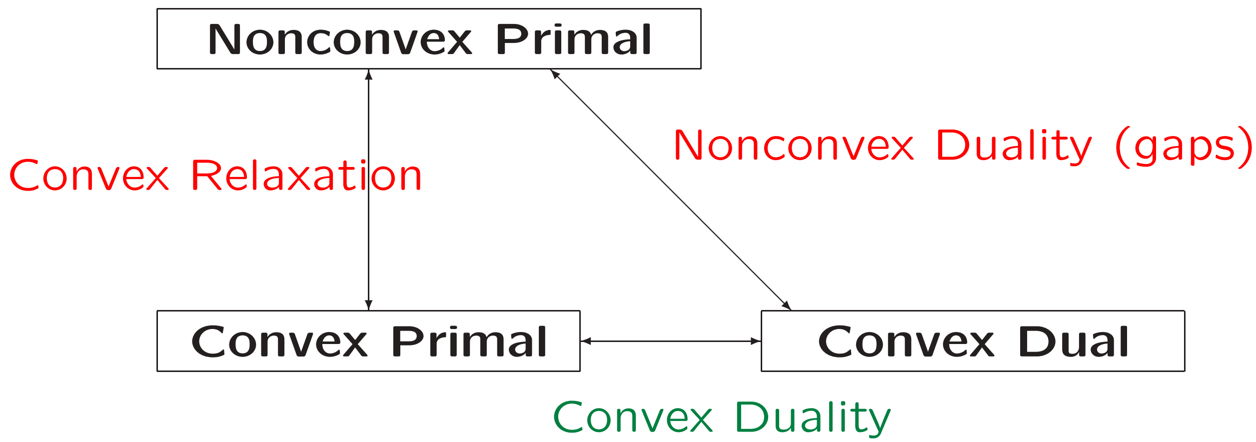
Semialgebraic problems

- Semialgebraic: a finite number of polynomial equalities and inequalities.
- Ubiquitous in systems engineering (and elsewhere).
- Surprising expressive power.
- Mix continuous and discrete variables (ex. hybrid systems).
- In particular:
  - Optimization problems with polynomial objective and constraints.
  - Quadratic, linear, Boolean programming.

Extremely broad class of problems, and clearly NP-hard in general.

Our claim: by combining ideas from real algebra and convex optimization, very effective *algorithms* can be obtained.

Relaxations



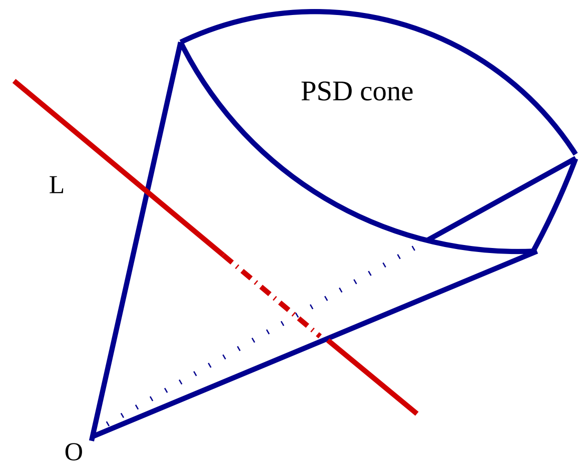
- Make the problems "simpler," by modifying the constraints.
- The relaxed problem provides bounds, or even the exact answer.
- The results can be used directly, or combined with other schemes.
- A fundamental technique in many existing results.
- In the last few years, semidefinite relaxations (LMIs).

Semidefinite programming - background

- A semidefinite program takes the form:

$$M(z) := M_0 + \sum_{i=1}^{m} z_i M_i > 0,$$

  where $z \in \mathbb{R}^m$ is the variable and $M_i \in \mathbb{R}^{n\times n}$ are given symmetric matrices.
- The intersection of an affine subspace L and the self-dual cone of positive definite matrices.



- Convex finite dimensional optimization problem.
- A broad generalization of linear programming. Nice duality theory.
- Solvable in polynomial time (interior point, etc.).
- *Many* applications.

Nonnegativity of polynomials

Polynomials of degree d in n variables:

$$F(x_1, x_2, \ldots, x_n) = \sum_{k_1+k_2+\cdots+k_n \leq d} a_{k_1 k_2 \ldots k_n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$$

How to check if a given F (of even degree) is globally nonnegative?

$$F(x_1, x_2, \ldots, x_n) \geq 0, \quad \forall x \in \mathbb{R}^n$$

- For $d = 2$, easy (check eigenvalues). What happens in general?
- Decidable, but NP-hard when $d \geq 4$.
- Possible approaches: Decision algebra, Tarski-Seidenberg, quantifier elimination, etc. Very powerful, but bad complexity properties.
- Numerous applications. We'll see some later...
- Want "low" complexity, at the cost of possibly being conservative.

A sufficient condition

A “simple” sufficient condition: a sum of squares (SOS) decomposition:

$$F(x) = \sum_i f_i^2(x)$$

If $F(x)$ can be written as above, for some polynomials f_i, then $F(x) \geq 0$.

Is this condition conservative? Can we quantify this?

- In some cases (for example, polynomials in one variable), it is exact.
- Known counterexamples, but perhaps “rare” (ex. Motzkin, Reznick 99, etc.)

Can we compute it efficiently?

- Yes, using semidefinite programming.

Checking the SOS condition

Given $F(x)$, degree $2d$.

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

Let z be a suitably chosen vector of monomials (in the dense case, all monomials of degree $\leq d$).

Then, F is SOS iff:

$$F(x) = z^T Q z, \qquad Q \succeq 0$$

- Comparing terms, obtain linear equations for the elements of Q.
- Can be solved as a semidefinite program (with equality constraints).
- Factorize $Q = L^T L$. The SOS is given by $f = Lz$.

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

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

Using SOSTOOLS: `[Q,Z]=findsos(2*x^4+5*y^4-x^2*y^2+2*x^3*y)`

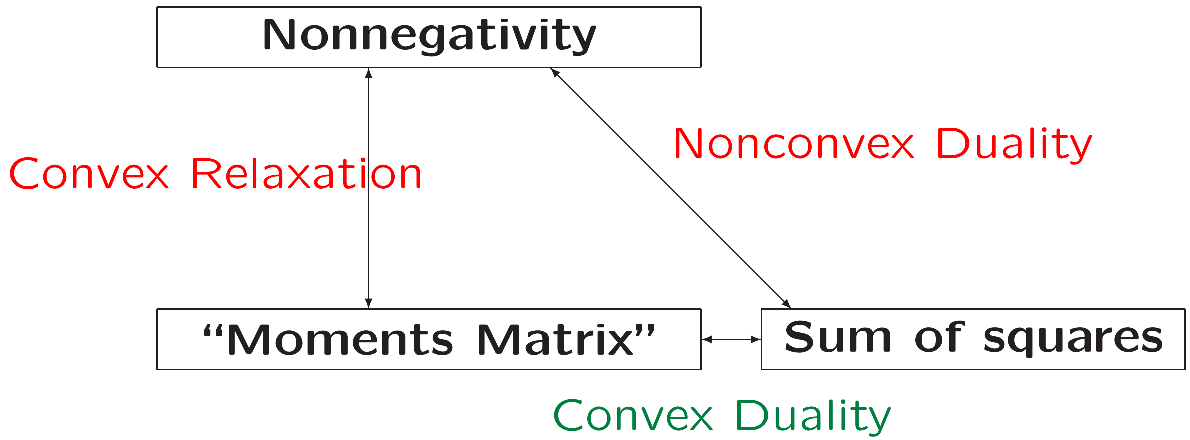
SOS are nice

Nonnegativity is hard

Sums of squares are *much* easier!

But surprisingly, not too different.

Relaxations - SOS



- The "moments matrix" suggests candidate points where the polynomial is negative.
- The sums of squares certify or prove polynomial nonnegativity.

Some properties

- The resulting problem is polynomially sized (in n).
- SDPs can be efficiently solved in practice. Approximate solutions in provable polynomial time. Exact complexity not fully understood yet.
- A most important feature: the problem is still a SDP if the coefficients of F are variable, and the dependence is affine.

$$F(x, \alpha) = \alpha_1 F_1(x) + \cdots + \alpha_m F_m(x)$$

- Can optimize over SOS polynomials in affinely described families.
- By properly choosing the monomials, can exploit structure (sparsity, symmetries, ideal structure).

Let's see some concrete applications...

Global optimization

Consider for example:

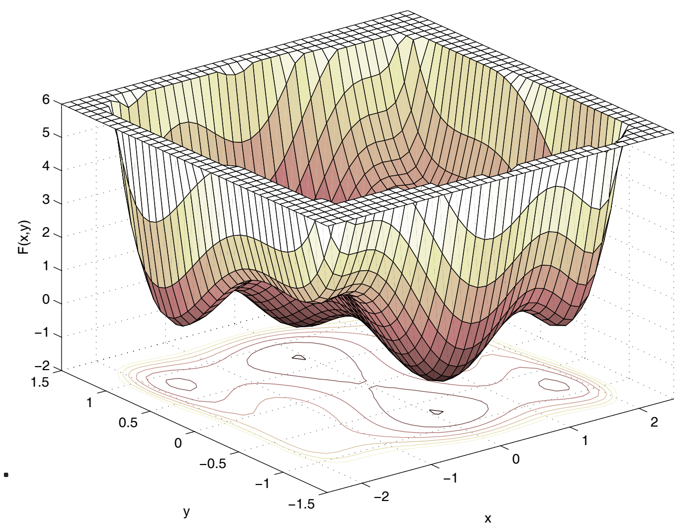
$$\min_{x,y} F(x,y)$$

with $F(x,y) := 4x^2 - \frac{21}{10}x^4 + \frac{1}{3}x^6 + xy - 4y^2 + 4y^4.$

- Not convex. Many local minima. NP-hard.
- Find the largest γ s.t.

$$F(x,y) - \gamma \text{ is SOS.}$$

- Essentially due to Shor (1987).
- A semidefinite program (convex!).
- If exact, can recover optimal solution.
- Surprisingly effective.



Solving, the maximum γ is -1.0316. Exact value.
Many more details in (P. & Sturmfels, 2001).

Why does this work?

Three independent facts, theoretical and experimental:

- The existence of efficient algorithms for SDP.
- The size of the SDPs grows much slower than the Bézout number μ.
  - A bound on the number of (complex) critical points.
  - A reasonable estimate of complexity.
  - The bad news: $\mu = (2d-1)^n$ (for dense polynomials).
  - Almost all (exact) algebraic techniques scale as μ.
- The lower bound f^{SOS} very often coincides with f^*. (why? what does *often* mean?)

SOS provides *short proofs*, even though they're not guaranteed to exist.

Lyapunov stability analysis

- To prove asymptotic stability of $\dot{x} = f(x)$,

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

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

- For linear systems $\dot{x} = Ax$, quadratic Lyapunov functions $V(x) = x^T P x$

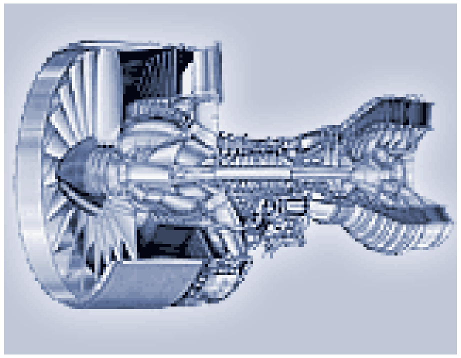
$$P > 0, \qquad A^T P + PA < 0.$$

- With an affine family of candidate polynomial V, $\dot{V}$ is also affine.
- Instead of checking nonnegativity, use a SOS condition.
- Many variations possible: nonlinear $\mathcal{H}_\infty$ analysis, parameter dependent Lyapunov functions, stochastic versions, etc.

Lyapunov stability - Example

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

$$\begin{aligned}\dot{x} &= -y + \frac{3}{2}x^2 - \frac{1}{2}x^3 \\ \dot{y} &= 3x - y;\end{aligned}$$



Try a generic 4th order polynomial Lyapunov function.

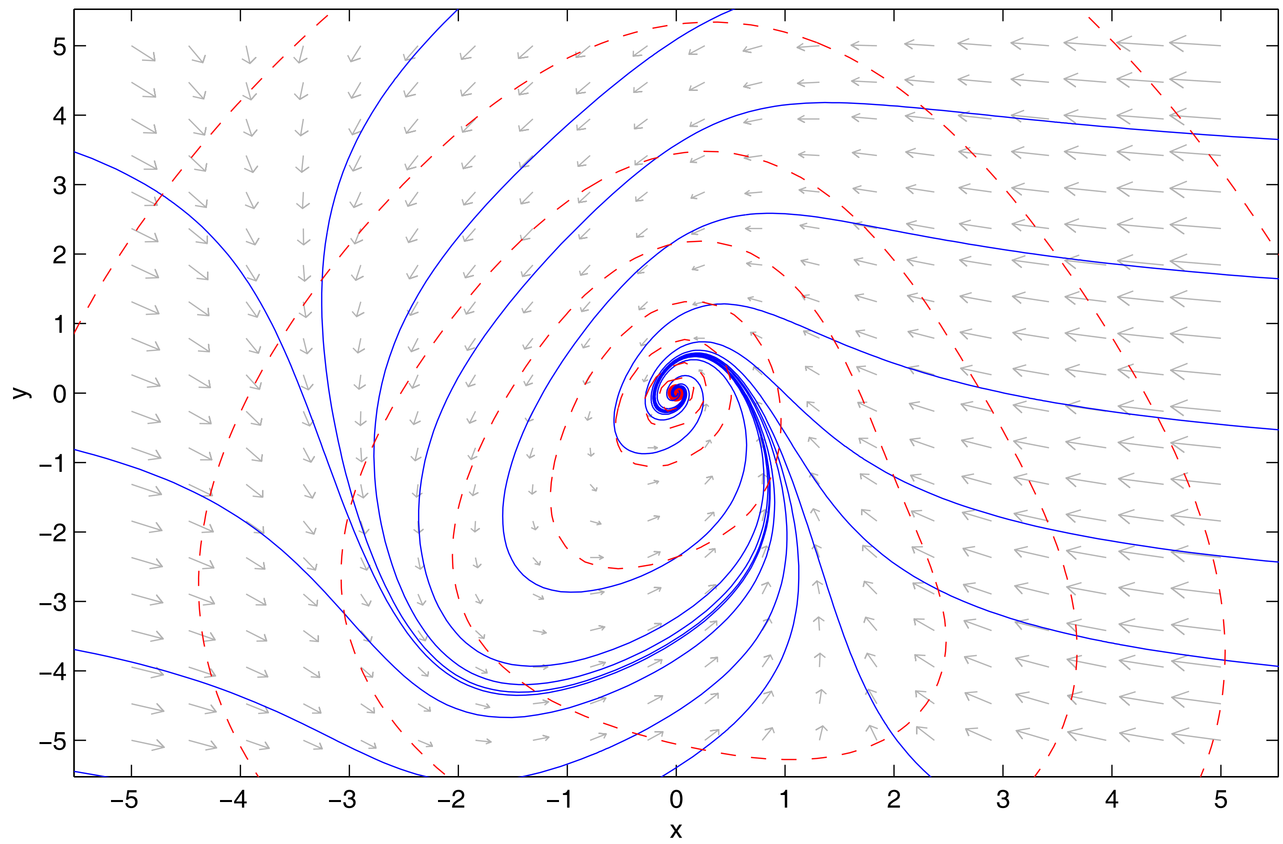
Find a $V(x, y)$ that satisfies the conditions:

- $V(x, y)$ is SOS.
- $-\dot{V}(x, y)$ is SOS.

Can easily do this using SOS/SDP techniques . . .

Lyapunov stability (cont.)

After solving the SDPs, we obtain a Lyapunov function.



Deciding quantum entanglement

A bipartite mixed quantum state ρ is *separable* (not *entangled*) if

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_i\rangle\langle\phi_i| \qquad \sum p_i = 1,$$

for some ψ_i, ϕ_i. Given ρ, how to decide if it is entangled?

- Joint work with A. Doherty and F. Spedalieri (PRL 88, May 2002).
- A hierarchy of SDP-based tests providing entanglement witnesses
- The first level, corresponds to a well-known criterion (PPT).
- The second level, detects all entangled quantum states tried!

Andrew will talk about this in more detail tomorrow...

Toward the general case

- Everything described earlier deals with *global* properties.
  - But, also want local results (*constrained* optimization).
  - For example, to handle *discrete* variables (or mixtures).

How do we generalize these ideas, while keeping everything computable?

- A model problem: checking emptiness of semialgebraic sets.
  - Many interesting questions can be cast in this form.

  Example: γ is a lower bound of

$$\min_{x \in S} F(x)$$

  iff $\{x \in S, F(x) < \gamma\}$ is empty.

What can we do within this framework? *Lots* of things...

Proving emptiness

- There is a fundamental asymmetry between establishing that:
  - A set has at least one element.
  - The set is empty.
- In optimization, finding feasible points vs. bounds.
- Roughly speaking, the difference between NP and co-NP.

For existence, it is enough to produce an *instance*. These are always "simple".

For emptiness, we need a *certificate*, that *could* potentially be "complicated".

Equivalent terms: *witnesses* and *proofs*.

What certificates of emptiness do we know?

Linear programming duality

Certificates nonexistence of real solutions of linear equations.

$$\left\{ \begin{array}{rcl} Ax + b & \geq & 0 \\ Cx + d & = & 0 \end{array} \right\} = \emptyset \quad \Longleftrightarrow \quad \exists\, \lambda, \nu \;\text{ s.t. } \left\{ \begin{array}{rcl} \lambda^T A + \nu^T C & = & 0 \\ \lambda^T b + \nu^T d & = & -1 \\ \lambda & \geq & 0 \end{array} \right.$$

- Finding certificates is also a linear programming problem.
- Also known as Farkas' lemma.
- Primal and dual are polynomial time solvable.
- Relies on convexity.

Well known, but there are more...

LP duality (II)

$$\left\{ \begin{array}{rcl} Ax+b & \geq & 0 \\ Cx+d & = & 0 \end{array} \right\} = \emptyset \quad \Longleftrightarrow \quad \exists\, \lambda, \nu \text{ s.t. } \left\{ \begin{array}{rcl} \lambda^T A + \nu^T C & = & 0 \\ \lambda^T b + \nu^T d & = & -1 \\ \lambda & \geq & 0 \end{array} \right.$$

Proof: ($\Leftarrow$) Assume the system is feasible (i.e., there exists an x).
Now, let's multiply the equations by λ^T, ν^T:

$$0 \leq \lambda^T(Ax+b) + \nu^T(Cx+d) = \underbrace{(\lambda^T A + \nu^T C)}_{0} x + \underbrace{(\lambda^T b + \nu^T d)}_{-1}$$

A contradiction!

The set has to be empty.

Well known, but there are more...

Hilbert's Nullstellensatz

Certificates nonexistence of complex solutions of polynomial equations.

$$\{z \in \mathbb{C}^n \mid f_i(z) = 0\} = \emptyset \quad \iff \quad \begin{array}{c} 1 \in \text{ Ideal } (f_i) \\ \text{or} \\ \exists g_i(x) \text{ s.t. } \sum_i f_i(z) g_i(z) = 1 \end{array}$$

- Cornerstone of algebraic geometry, establish a correspondence between geometric ideas and algebraic objects.

  affine varieties $\Leftrightarrow$ *polynomial ideals*

- The "canonical" NP-complete problem in the real model of computation.
- For fixed degree of the g_i can solve using linear algebra.
- In control, appears as the Bézout equation (factorizations).

How to generalize this?

Degree \ Field	Complex	Real
Linear	Kernel/range Thm	LP duality
Polynomial	Nullstellensatz	?????

Can we get the best of both worlds?

General polynomial equations, as in the Nullstellensatz.

And real solutions, so we can handle inequalities?

HOW?

The search for P-proofs

- Look for "obvious" algebraic proofs, of bounded complexity.
- Example:

$$\text{Is } \{f(x) \geq 0, g(x) \geq 0, h(x) = 0\} \text{ empty?}$$

- If we can find polynomials s_i, t_i, with s_i SOS such that:

$$s_1 + s_2 \cdot f + s_3 \cdot g + s_4 \cdot f \cdot g + t_1 \cdot h = -1$$

  then the set has to be empty. Why?

- Condition is affine in s_i, t_i. Important later.

P-proofs

- Recall our example:

$$\text{Is } \{f(x) \geq 0, g(x) \geq 0, h(x) = 0\} \text{ empty?}$$

- We have polynomials s_i, t_i, with s_i SOS such that:

$$s_1 + s_2 \cdot f + s_3 \cdot g + s_4 \cdot f \cdot g + t_1 \cdot h = -1$$

  Then, the set is empty. Why?

Assume it is not, and plug a feasible point x_0 in the expression above:

$$\underbrace{s_1(x_0) + s_2(x_0) \cdot f(x_0) + s_3(x_0) \cdot g(x_0) + s_4(x_0) \cdot f(x_0) \cdot g(x_0)}_{\geq 0} + \underbrace{t_1(x_0) \cdot h(x_0)}_{0} = -1$$

A contradiction. The set has to be empty!

Now, for the theorem...

Positivstellensatz

Certificates for real solutions of systems of polynomial equations!

$$\left\{ \begin{array}{rcl} x & \in & \mathbb{R}^n \\ f_i(x) & \geq & 0 \\ h_i(x) & = & 0 \end{array} \right\} = \emptyset \qquad \Longleftrightarrow \qquad \exists f, h \quad \left\{ \begin{array}{l} f + 1 + h = 0 \\ f \in \mathrm{Cone}(f_i) \\ h \in \mathrm{Ideal}(h_i) \end{array} \right.$$

- A fundamental theorem in real algebraic geometry (Stengle 1974).
- A common generalization of Hilbert's Nullstellensatz and LP duality.
- Provides infeasibility certificates.
- Unless NP=co-NP, the certificates cannot *always* be polynomially sized.
- Sums of squares are a fundamental ingredient.

How does it work?

P-satz and SDP

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$, decide whether it is empty.

What is the algebraic structure of the allowable operations among constraints?

Define

- The cone (or preorder) corresponding to the inequalities.
- The ideal generated by the equality constraints.

To prove infeasibility, find $f \in \mathrm{Cone}(f_i), h \in \mathrm{Ideal}(h_i)$ such that

$$f + 1 + h = 0.$$

- Equations are affine. Can find certificates by solving SDPs!
- A explicit SDP hierarchy, given by certificate degree (P. 2000).
- Tons of applications:
  optimization, dynamical systems, quantum mechanics...

A (brief) overview of applications

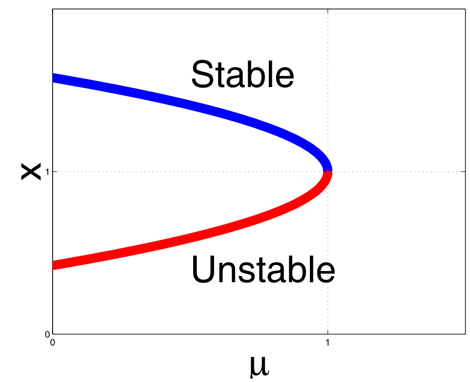
- Systems and control.
  - Lyapunov functions.
  - Robust bifurcation analysis.
  - Hybrid/uncertain system analysis.
- Matrix copositivity: Is $x^T A x \geq 0$ for all $x \geq 0$?
- Higher order relaxations for quadratic programming.
  - Natural generalization of the standard SDP relaxation.
- Combinatorial optimization: MAX-CUT, 3SAT, etc.
- Geometric theorem proving.

Robust bifurcation analysis

- $\dot{x} = f(x, \mu)$ has a fixed point bifurcation when the flow around a fixed point x_0 changes qualitatively, when μ crosses some critical value μ_0.
- *Local* bifurcations can be simply characterized. For saddle-node:

$$\begin{array}{ll} f = 0 & w^* D_\mu f \neq 0 \\ w^* D_x f = 0 & w^* D_x^2 f(v, v) \neq 0 \end{array}$$

  where v, w are the right and left eigenvectors of $J := D_x f$. The normal form is $\mu - x^2$.

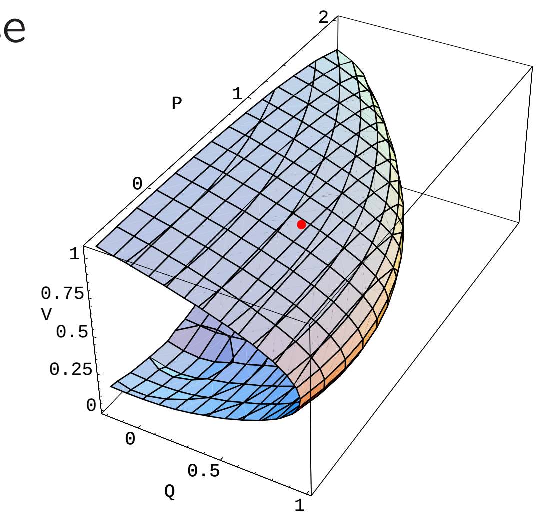


- Given an equilibrium, what is the maximum variation in the parameters?
- Want to *guarantee a minimum distance* (or safety margin) to the hypersurface where bifurcations occur. *Global* information.

Application: Voltage collapse in power systems

- In power systems, saddle-node bifurcations cause *voltage collapse* (Dobson 1993).

$$
\begin{aligned}
0 &= -4V\sin\alpha - P \\
0 &= -4V^2 + 4V\cos\alpha - Q
\end{aligned}
$$

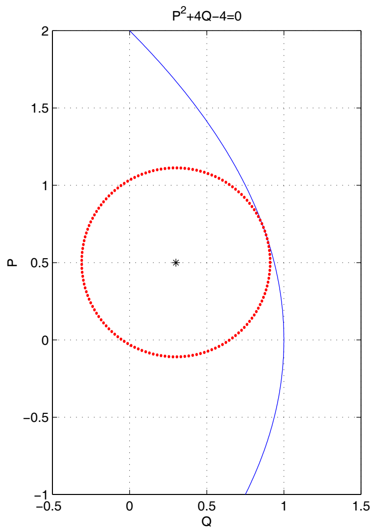


- Nominal equilibrium $(P, Q) = (0.5, 0.3)$.
- Want bounds on the maximum allowable loads.

Minimize the function $J(P, Q) := (P-0.5)^2 + (Q-0.3)^2$ subject to:

$$
\begin{aligned}
f_1 &:= x^2 + y^2 - 1 = 0 \\
f_2 &:= -4Vx - P = 0 \\
f_3 &:= -4V^2 + 4Vy - Q = 0 \\
f_4 &:= \det J/(-16V) = x^2 + y^2 - 2Vy = 0
\end{aligned}
$$

Bifurcation example (continued)



- For simplicity, eliminate the variables (x, y, V) that do not appear in the objective.
- Compute the *elimination ideal*

$$\langle f_1, f_2, f_3, f_4 \rangle \cap \mathbb{R}[P, Q] = \langle P^2 + 4Q - 4 \rangle$$

using Gröbner basis. All the constraints that only include P and Q.

- Find the maximum γ^2 that verifies the condition:

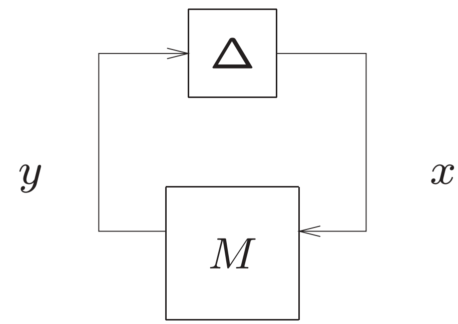
$(P - 0.5)^2 + (Q - 0.3)^2 - \gamma^2 + \lambda(P, Q)(P^2 + 4Q - 4)$ is a sum of squares.

In this case, it is sufficient to pick $\lambda(P, Q)$ constant, optimal value of $\gamma^2 \approx 0.3735$, with $\lambda \approx -0.2883$.

- In this case, the bound is *exact*.

Example - structured singular value μ

- A central paradigm in robust control.
- Structured singular value μ and related problems: provides better upper bounds.
- μ is a measure of robustness: how big can a structured perturbation Δ be, without losing stability.

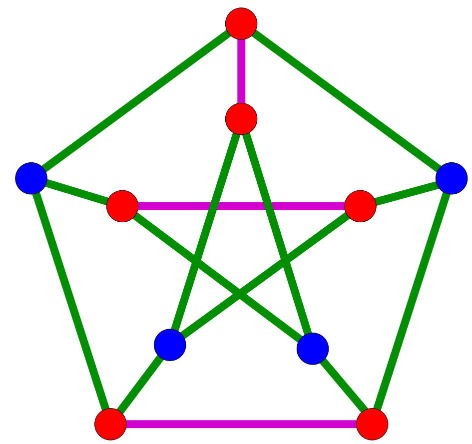


- A standard semidefinite relaxation: the μ upper bound.
  - Morton and Doyle's counterexample with four scalar blocks.
  - Exact value: approx. 0.8723
  - Standard μ upper bound: 1
  - New bound: 0.895

New MAX CUT relaxations

- Partition the nodes of a graph in two disjoint sets, maximizing the number of edges between sets.
- Practical applications (optimal circuit layout, etc.), but NP-complete.
- As boolean optimization:

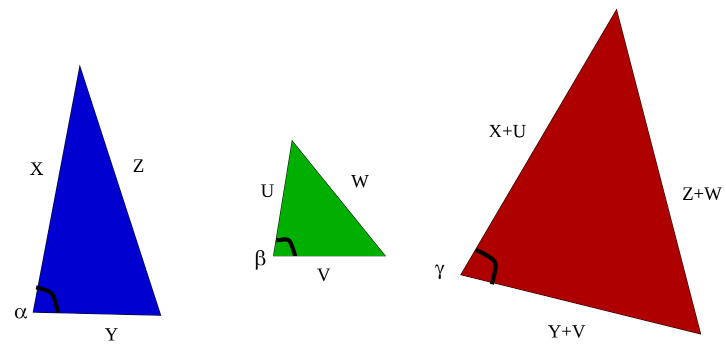
$$\max_{y_i \in \{-1,1\}} \frac{1}{2} \sum_{i,j} w_{ij}(1 - y_i y_j),$$



- A well-known semidefinite relaxation (basis of Goemans-Williamson).
- For some cases (n-cycle, Petersen graph) the new conditions are exact. The standard relaxation is not.
- Petersen graph: Standard relaxation: 12.5, New relaxation: 12.

Geometric theorem proving

- A geometric inequality arising from circle packings (Ronen Peretz):



$$\alpha \cdot (X + Y - Z) + \beta \cdot (U + V - W) \leq \gamma \cdot ((X + U) + (Y + V) - (Z + W))$$

- Not easy to prove. *Not* semialgebraic, in the standard form.
- The inequality holds if certain polynomial expression is nonnegative.
- Using SOS/SDP, we will obtain a very concise proof.

Proof length and complexity

- P-satz is a complete algebraic proof system.
- Certificate size (proof length) is *crucial*.
- Depends on the problem, no "uniformly best" system is known (ex: resolution vs. cutting-plane).
- Only want proofs of bounded complexity (for practical reasons).
- The strategy in our methods:
  - Shoot for best possible result, fixing the P-satz proof length.
  - Potentially generate all the valid constraints.
  - Search over combinations using SDP, until a contradiction is found.

The P-satz is nice because (like SOS) usually gives short certificates.

Exploiting structure

Crucial for good performance. What algebraic properties can we profit of?

- Sparseness: few nonzero coefficients.
  - Newton polytopes techniques.
- Symmetries: invariance under a group of transformations.
  - Appear quite frequently in practice.
  - Representation- and invariant-theoretic methods (Gatermann and P.).
  - Enabling factor in applications.
- Ideal structure: equality constraints.
  - Compute in the coordinate ring.
  - Quotient bases (Gröbner).
  - Zero dimensional case is particularly interesting.

Key issues, longer term

- System analysis *should be* automatic theorem proving (that works!).
- We've been doing it somehow, but need more sophisticated techniques.
- "Extend and embrace," to incorporate proven techniques from other domains:
  - From AI: selection of proof strategies.
  - Use of abstractions.
  - Randomization: good for analysis (NP, coNP), not clear for synthesis (Π_2, Σ_2).
- What does shortest proof length tells us?
  - Connections to sensitivity issues, Lagrange multipliers, etc.

Conclusions and future research

- Constructive methodology for practically relevant questions.
- A broad generalization of known successful techniques.
- Tradeoff between accuracy vs. computation time.
- Practicalities. How big are the problems that we can solve?
- Can combine with other techniques, e.g. symmetry reduction.

- How can we exploit the problem structure for more efficient solutions?
- What are the computational complexity implications?

SOSTOOLS: sums of squares toolbox

Handles the general problem:

$$\begin{array}{ll} \text{min}_{u_i} & c_1 u_1 + \cdots + c_n u_n \\ \text{s.t} & P_i(x,u) := A_{i0}(x) + A_{i1}(x) u_1 + \cdots + A_{in}(x) u_n \quad \text{are SOS} \end{array}$$

- MATLAB toolbox, freely available.
- Requires MATLAB's symbolic toolbox, and SeDuMi (SDP solver).
- Natural syntax, efficient implementation.
- Developed by Stephen Prajna, Antonis Papachristodoulou, and PP.
- Includes customized functions for several problems.

Get it from: `http://www.aut.ee.ethz.ch/~parrilo/sostools`
`http://www.cds.caltech.edu/sostools`