Sums of Squares in Polynomial Optimization
Lecture 2

João Gouveia

• U C •

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

21st of May 2019 - IPCO Summer School

Section 5

The Moment Approach

The probability measure viewpoint

There is another way of reformulating the unconstrained pop.

Unconstrained POP - v3.0

Given a polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ find

$$p^* = \min_{\mu \in \mathcal{M}(\mathbb{R}^n)} \int p(\mathbf{x}) d\mu,$$

where $\mu \in \mathcal{M}(\mathbb{R}^n)$ is the set of probability distributions in $\mathbb{R}^n$.

In a way, we are just averaging the values of the polynomial with different weights, so the minimum will be attained when we put all mass in the minimizers of p.

But how would we even try to compute this?

Moments

Suppose

$$p(\mathbf{x}) = \sum_{\alpha \in I \subseteq \mathbb{N}^n} p_\alpha \mathbf{x}^\alpha.$$

Then we can think of the integral of p as a sum:

$$\int p(\mathbf{x}) d\mu = \sum_{\alpha \in I} p_\alpha \int x^\alpha d\mu.$$

The sequence $y = (y_\alpha)_{\alpha \in \mathbb{N}^n} = \int x^\alpha d\mu$ is the **moment sequence** of the measure μ. If we call $\tilde{p}$ to the sequence of coefficients of p then

$$\int p(\mathbf{x}) d\mu = \langle \tilde{p}, y \rangle.$$

Let us denote the set of all moment sequences of measures with support contained in a set K by $\mathrm{Mom}(K)$.

The moment viewpoint

We have now yet another way of reformulating the unconstrained pop.

Unconstrained POP - v3.1

Given a polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ find

$$p^* = \min_{y \in \mathrm{Mom}(\mathbb{R}^n)} \langle \tilde{p}, y \rangle.$$

We still need to characterize $\mathrm{Mom}(K)$. This is a very classic (and very hard) problem called the **moment problem**. Only a few simple cases have full solutions.

Observations

1. For $K = \mathbb{R}$ this is the Hamburger moment problem (1921)
2. For $K = \mathbb{R}^+$ this is the Stieltjes moment problem (1894)
3. For $K = [a, b]$ this is the Hausdorff moment problem (1921)

Conditions on moments

It is generally very hard to find necessary and sufficient conditions for a sequence to be a moment sequence, but there is a simple necessary one.

Necessary condition

If y is a moment sequence and p a polynomial, $\langle y, \widetilde{p^2} \rangle \geq 0$.

We can write this in a nicer way.
Let $M(y)$ be the (infinite) matrix indexed by $\mathbb{N}^n$ with $[M(y)]_{\alpha,\beta} = y_{\alpha+\beta}$ then

$$\langle y, \widetilde{p^2} \rangle = \tilde{p}^t M(y) \tilde{p}$$

Necessary condition v2.0

If y is a moment sequence then $M(y) \succeq 0$.

We will denote by $M_d(y)$ the submatrix of $M(y)$ indexed by monomials of degree less or equal to d. $M(y) \succeq 0$ if and only if all truncated matrices $M_d(y) \succeq 0$.

The moment relaxation

We will relax being a measure by this truncated moment condition.

Moment optimization

Given a polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ of degree d find

$$p^{\text{mom}} = \inf_{y \in \mathbb{R}^{\mathbb{N}^n}} \langle y, \tilde{p} \rangle \text{ subject to } M_d(y) \succeq 0$$

Observations

1. This is still a semidefinite program.
2. It is actually dual to the sums of squares version.
3. In fact $p^{\text{mom}} = p^{\text{sos}}$.
4. We can interpret the attained y as approximations for moments of a measure on optimal the solution set. We might recover the minimizers if lucky.
5. Not as nice to provide certificates.

Revisited example

Let us revisit yet again the problem

$$\min_{x\in\mathbb{R}} p(x) = 3x^4 - 4x^3 + 12x^2 - 24x + 10.$$

Step 1: Establish the sdp: $\min 3y_4 - 4y_3 + 12y_2 - 24y_1 + 10$ s.t.

$$M_2(y) = \begin{bmatrix} 1 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0.$$

Step 2: Solve the sdp:

```
sdpvar y1 y2 y3 y4
optimize([1,y1,y2;y1,y2,y3;y2,y3,y4]>=0,3*y4-4*y3+12*y2-24*
```

We get -3.0000. **But what is the minimizer?**

$$M_2(y) = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 \\ 1.0000 & 1.0000 & 1.0000 \\ 1.0000 & 1.0000 & 1.0000 \end{bmatrix}$$

This is a rank 1 matrix. y_1 should be the average of x with respect to optimal measure. If we plug 1 for x we get in fact -3.

A larger example

Let us think about $p(x,y) = x^4 - 4x^3y + 7x^2y^2 - 4xy^3 - 4xy + y^4$.

Minimize $y_{40} - 4y_{31} + 7y_{22} - 4y_{13} - 4y_{11} + y_{04}$ subject to

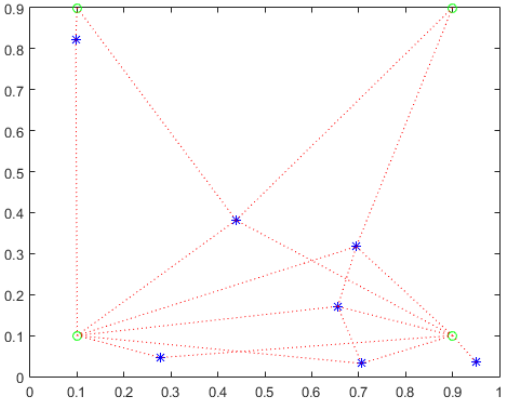
$$\begin{bmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix} \succeq 0$$

We get -4 as the minimum and the rank 2 matrix

$$M_2(y) = \begin{bmatrix} 1 & 0 & 0 & 2 & 2 & 2 \\ 0 & 2 & 2 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 4 & 4 & 4 \\ 2 & 0 & 0 & 4 & 4 & 4 \\ 2 & 0 & 0 & 4 & 4 & 4 \end{bmatrix} \succeq 0$$

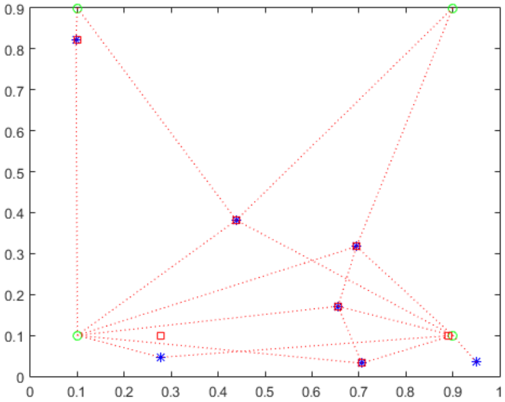
This is the average combination of the moments of $\pm(\sqrt{2}, \sqrt{2})$, the two minimizers.
This is not always possible!

The graph realization problem



$$\min_{x \in \mathbb{R}^{2 \times n}} \sum_{\{i,j\} \in E} (\|x_i - x_j\|^2 - d_{ij}^2)^2$$

The graph realization problem



$$\min_{x \in \mathbb{R}^{2 \times n}} \sum_{\{i,j\} \in E} (\|x_i - x_j\|^2 - d_{ij}^2)^2$$

Section 6

Nonnegative certificates over the nonnegative orthant

Nonnegativity over $\mathbb{R}^n_+$

To warm up for the general constrained polynomial optimization, let us study a very particular case.

Certifying nonnegative over the orthant

Given a **homogeneous** polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ find if

$$p(\chi) \geq 0 \text{ for all } \chi \in \mathbb{R}^n_+.$$

Why do we care? Because it is fun, but also:

Copositive matrices

A symmetric matrix $M \in \mathbb{R}^{n \times n}$ is **copositive** if and only if $\chi^t M \chi \geq 0$ for all $\chi \in \mathbb{R}^n_+$.

Optimizing over the cone of copositive matrices can be used to reformulate almost everything.

Reducing to sums of squares

A simple trick relies on the following:

$$p(x) \geq 0, \forall x \in \mathbb{R}^n_+ \quad \textbf{if and only if} \quad p(x^2) \geq 0, \forall x \in \mathbb{R}+^n$$

where $x^2 = (x_1^2, x_2^2, \ldots, x_n^2)$.

We can now use sums of squares, to search for certificates of the type

$$(x_1^2 + x_2^2 + \cdots + x_n^2)^r p(\mathbf{x}^2) \in \Sigma[\mathbf{x}].$$

When applied to check copositivity this is usually called the *Parrilo hierarchy*.

Observation

We can use this trick to optimize over any polynomial image of an affine space.
If $C = \varphi(\mathbb{R}^n)$ where $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ is a polynomial map, then $p \geq 0$ in C if and only if $p \circ \varphi \geq 0$ in $\mathbb{R}^n$, and we can use sums of squares.

Pólya's Certificates

Let Δ^n be the standard simplex $\{x \in \mathbb{R}^n_+ \mid \sum x_i = 1\}$ as usual.
For homogeneous polynomials nonnegative over $\mathbb{R}^n_+$ and Δ^n is equivalent.

Trivial observation

If $p(x)$ only has nonnegative coefficients it is nonnegative over the simplex Δ^n.

Not enough:

$$x^2 + xy - xz + y^2 - yz + z^2$$

in positive over Δ^3.

Theorem (Pólya - 1928)

$p(\mathbf{x})$ is **positive** *over Δ^n if and only if there exists k such that*

$$(x_1 + x_2 + \cdots + x_n)^k p(\mathbf{x})$$

has only **positive** *coefficients.*

Example

$$p(x,y,z) = x^2 + xy - xz + y^2 - yz + z^2$$

$$(x+y+z)p(x,y,z) = x^3 + 2x^2y + 2xy^2 - xyz + y^3 + z^3$$

$$(x+y+z)^2p(x,y,z) = x^4 + 3x^3y + x^3z + 4x^2y^2 + x^2yz + 3xy^3 + xy^2z - xyz^2 + xz^3 + y^4 + y^3z + yz^3 + z^4$$

$$(x+y+z)^3p(x,y,z) = x^5 + 4x^4y + 2x^4z + 7x^3y^2 + 5x^3yz + x^3z^2 + 7x^2y^3 + 6x^2y^2z + x^2z^3 + 4xy^4 + 5xy^3z + xyz^3 + 2xz^4 + y^5 + 2y^4z + y^3z^2 + y^2z^3 + 2yz^4 + z^5$$

Observation

We can use this to generate **linear programming** relaxations of strict copositivity of matrices, among other things.

What is actually happening

The proof boils down to a simple fact. Given a monomial $\mathbf{x}^\alpha$ define

$$\mathbf{x}_\varepsilon^\alpha = \prod_{i=1}^{n} \prod_{j=0}^{\alpha_i - 1} (x_i - j\varepsilon).$$

For example $(x^3y^2)_{0.1} = x(x-0.1)(x-0.2)y(y-0.1)$.

We denote by $p_\varepsilon(\mathbf{x})$ the polynomial obtained by applying this to each monomial.

Then, the coefficient of $\mathbf{x}^\alpha$ in $(x_1 + x_2 + \cdots + x_n)^k p(\mathbf{x})$, where d is the degree of p anda $\tilde{\alpha} = \frac{\alpha}{|\alpha|} \in \Delta^n$ is

$$\left(\frac{k!|\alpha|^d}{\alpha_1!\alpha_2!\cdots\alpha_n!} \right) p_{\frac{1}{|\alpha|}}(\tilde{\alpha})$$

Since $p_{\frac{1}{|\alpha|}} \to p$ uniformly in Δ^n we get the result.

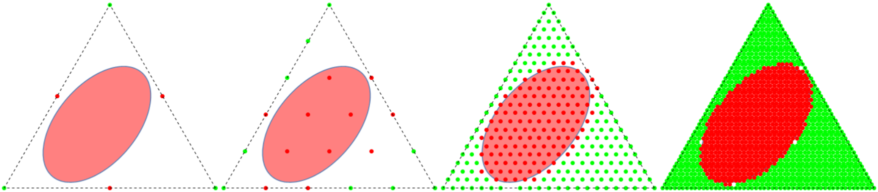
What is actually happening - Part 2

What we actually proved:

Sketch of Proposition

If we plot the points $\tilde{\alpha}$ in the simplex colored green or red depending on if the coefficients are positive or negative in $(x_1 + x_2 + \cdots + x_n)^k p(\mathbf{x})$ then the green points converge to the region of positivity and the red ones to that of negativity as $k \to \infty$.

$$p(x, y, z) = 5x^2 - 6xy + 2y^2 - 4xz - 2yz + z^2$$



Section 7

Constrained Polynomial Optimization

The general formulation

In general we are interested in constrained optimization.

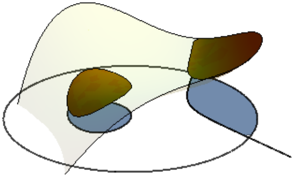
Semialgebraically constrained POP

Given polynomials $p(\mathbf{x}), g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_m(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ we want to find

$$p^* = \inf_{\chi \in S} p(\chi).$$

where $S = \{\chi \in \mathbb{R}^n \mid g_1(\chi) \geq 0, \ldots, g_m(\chi) \geq 0\}$.

In other words, we want to optimize a polynomial over a *basic closed algebraic set*.



$$\min x^3 + y^2 - x^2 - xy + 3 \text{ s.t. } 4 - x^2 - y^2 \geq 0 \text{ and } x^3 - y^2 - x \geq 0$$

Nonnegativity certificates over semialgebraic sets

We can now leverage our sum of squares idea to this case. To guarantee nonnegativity of $p(\mathbf{x})$ over S we can just ask for a certificate of the type

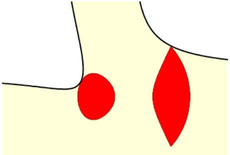
$$p(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sigma_1(\mathbf{x})g_1(\mathbf{x}) + \sigma_2(\mathbf{x})g_2(\mathbf{x}) + \cdots + \sigma_m(\mathbf{x})g_m(\mathbf{x})$$

where $\sigma_i \in \Sigma[\mathbf{x}]$. We call the set of all such polynomials $Q_S[\mathbf{x}]$.
Disclaimer: we are conflating S with its defining polynomials $g_1, \ldots, g_m$. It is not as innocent as it appears.
Example: For

$$S = \{(x, y) \mid 4 - x^2 - y^2 \geq 0, x^3 - y^2 - x \geq 0\}$$
$$p(x, y) = -y^2x^2 - 4yx^2 + 3x^2 - 2y^2x + 6x + y^2 + 4$$

we have

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



Putinar's Positivstellensatz

In principle, this seems to work.

Putinar's Positivstellensatz - 1993

If S is archimedean then any $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ such that $p(\mathbf{x}) > 0$ on S is in $Q_S[\mathbf{x}]$.

Here *archimedean* just means certifiably compact, more precisely, $N - \sum x_i^2 \in Q_S[\mathbf{x}]$ for big enough N.

There is a catch!

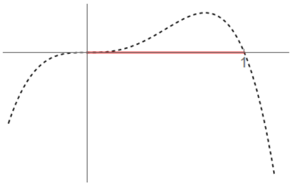
Checking membership in $Q_S[\mathbf{x}]$ is hard:

$$p(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sigma_1(\mathbf{x})g_1(\mathbf{x}) + \sigma_2(\mathbf{x})g_2(\mathbf{x}) + \cdots + \sigma_m(\mathbf{x})g_m(\mathbf{x})$$

implies no degree bounds on the σ_i...

A bad example

Consider $S = \{x \mid x^3(1-x) \geq 0\} = [0,1]$ and $p_\varepsilon(x) = x + \varepsilon^2$.



Claim

p_ε is nonnegative in S for all ε but

$$p_0(x) = x \notin Q_S[\mathbf{x}]$$

Assume it is, i.e., $x = \sum h_i(x)^2 + x^3(1-x)\sum g_i^2$

Evaluating at $x = 0$ we get $0 = \sum h_i(0)^2$ which implies $h_i(0) = 0$ for all i.

Differentiating we get $1 = 2\sum h_i(x)h_i'(x) + (x^3(1-x)\sum g_i^2)'$.

Evaluating at $x = 0$ we get $1 = 0$.

Observation

In fact, it can be shown that as $\varepsilon \to 0$, we need higher and higher degrees of h_i and g_i to certify the nonnegativity of $x + \varepsilon^2$.

Degree limits

In order to be able to search for such certificates, we need to bound the degrees.

Truncated quadratic module

We define $Q_S^d[\mathbf{x}]$ to be the set of all polynomials of the type:

$$\sigma_0(\mathbf{x}) + \sigma_1(\mathbf{x})g_1(\mathbf{x}) + \sigma_2(\mathbf{x})g_2(\mathbf{x}) + \cdots + \sigma_m(\mathbf{x})g_m(\mathbf{x})$$

where all σ are sums of squares and the degree of σ_0 and $\sigma_i g_i$ for all i is at most $2d$.

We can now work with this.

Observation

1. Searching for certificates in $Q_S^d[\mathbf{x}]$ is now a semidefinite program.
2. $Q_S^d[\mathbf{x}]$ is closed if S has nonempty interior.

The Lasserre hierarchy

We are now ready to establish a relaxation for constrained POP.

Lasserre Hierarchy

Given polynomials $p(\mathbf{x}), g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_m(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ and

$$S = \{\chi \in \mathbb{R}^n \mid g_1(\chi) \geq 0, \ldots, g_m(\chi) \geq 0\},$$

we define the d-th Lasserre hierarchy relaxation as

$$p_d^{\text{sos}} = \sup \lambda \text{ such that } p(\mathbf{x}) - \lambda \in Q_S^d[\mathbf{x}].$$

This can still be solved efficiently by semidefinite programming.

Properties

1. $p_1^{\text{sos}} \leq p_2^{\text{sos}} \leq \cdots \leq p^*$
2. If S is archimedean then $p_d^{\text{sos}} \to p^*$.

Toy Example

Recall the problem of minimizing x subject to $x^3(1-x) \geq 0$.
Lets compute p_2^{sos}. We want to maximize λ such that

$$x - \lambda = \sigma_0(x) + \sigma_1(x)x^3(1-x) = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}^t Q_0 \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} + q_1 x^3(1-x)$$

and $Q_0 \succeq 0$ and $q_1 \geq 0$.

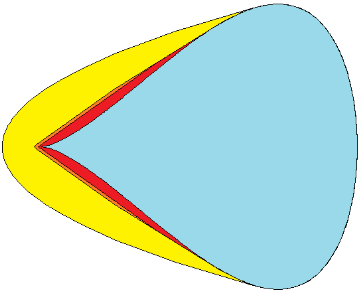
```
Q=sdpvar(3);
sdpvar l q x
F=[Q>=0,q>=0,coefficients([1,x,x^2]*Q*[1;x;x^2]
                                 +q*x^3*(1-x)-x+l,x)==0]
solvesdp(F,-l)
```

We get $p_2^{\text{sos}} = 0.125$ and in fact $x + \frac{1}{8} = \frac{1}{2}(1 + x + 2x^2)^2 + 4x^3(1-x)$.
Increasing d, $p_3^{\text{sos}} \approx -0.0416667$, $p_4^{\text{sos}} \approx -0.0208397$, $p_5^{\text{sos}} \approx -0.0127555$,...

and it kind of gets stuck there...

Toy Example 2

Take $S = \{(x, y) \mid -x^4 + x^3 - y^2 \geq 0\}$. We can draw the regions cut out by all the linear inequalities whose nonnegativity over S is certified by $Q_S^d[x, y]$, for different d.



Results for $d = 2, 3$ and 4.

This is a relaxation for the convex hull of S. It is precisely the set of all relaxations of first order moments in the moment approach.

Dealing with equalities

Consider the alternative formulation for the general POP.

Semialgebraically constrained POP v2.0

Given polynomials $p(\mathbf{x}), g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}), h_1(\mathbf{x}), \ldots, h_l(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ we want to find

$$p^* = \inf_{\chi \in S} p(\chi).$$

where $S = \{\chi \in \mathbb{R}^n \mid g_1(\chi) \geq 0, \ldots, g_m(\chi) \geq 0, h_1(\chi) = \cdots = h_l(\chi) = 0\}$.

Clearly this equivalent to the previous one as we can replace $h_i(\chi) = 0$ by $h_i(\chi) \geq 0$ and $-h_i(\chi) \geq 0$. But it is helpful to think separately of the equalities.

Example (MaxCut)

Given some symmetric $Q \in \mathbb{R}^{n \times n}$,

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j=1}^{n} q_{ij} x_i x_j \\ \text{subject to} \quad & x_i^2 = 1, \quad i = 1, \ldots, n \end{aligned}$$

Dealing with equalities - Part 2

There are two equivalent ways of thinking about equalities.

Commutative algebra free way

If S is defined as previously, one can certify nonnegativity by writing

$$p(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sigma_1(\mathbf{x})g_1(\mathbf{x}) + \cdots + \sigma_m(\mathbf{x})g_m(\mathbf{x}) + f_1(\mathbf{x})h_1(\mathbf{x}) + \cdots + f_l(\mathbf{x})h_l(\mathbf{x})$$

where $\sigma_i \in \Sigma[\mathbf{x}]$ and $f_j \in \mathbb{R}[\mathbf{x}]$.

For those more commutative algebra inclined, one can consider the ideal I generated by the polynomials $h_1(\mathbf{x}), \ldots, h_l(\mathbf{x})$ and think of working modulo it.

Commutative algebra way

If S is defined as previously, one can certify nonnegativity by writing

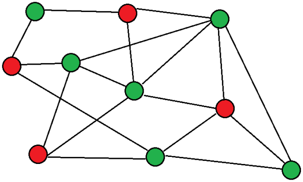
$$p(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sigma_1(\mathbf{x})g_1(\mathbf{x}) + \cdots + \sigma_m(\mathbf{x})g_m(\mathbf{x}) \mod I$$

where $\sigma_i \in \Sigma[\mathbf{x}]$.

These are of course precisely the same!!! But result in different degree restrictions.

Example - The stable set problem

Given a graph $G = ([n], E)$ we want to



$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} x_i \\ \text{subject to} \quad & x_i^2 = x_i, \;\; i = 1, \ldots, n \\ & x_i x_j = 0, \;\; \{i,j\} \in E \end{aligned}$$

Taking as our ideal

$$I_G = \langle x_1^2 - x_1, \ldots, x_n^2 - x_n, x_i x_j \mid \text{ for all } \{i,j\} \in E \rangle$$

we get as p_1^{sos} the value of minimizing λ such that for some $Q \succeq 0$ we have

$$\lambda - \sum_{i=1}^{n} x_i \equiv \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}^t Q \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix} \quad \bmod I_G.$$

This just means that we set all $x_i x_j$ to zero if $\{i,j\} \in E$ and all x_i^2 to x_i.
This is precisely Lovász *theta* number.

Example - Projecting tensors to rank 1 tensors

Given a tensor $\mathcal{F} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_m}$ a very important problem with many applications is to find a rank one approximation. More precisely:

Best rank-1 approximation tensor

Given $\mathcal{F}$ as above, find $x^i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, m$ such that

$$\|\mathcal{F} - x^1 \otimes x^2 \otimes \cdots \otimes x^n\|$$

is minimal.

This can be made into an equivalent polynomial optimization problem.

Best rank-1 approximation tensor

$$\begin{aligned} \text{maximize} \quad & \left(\sum_{i_1,\ldots,i_m} \mathcal{F}_{i_1,\ldots,i_m} x^1_{i_1} \cdots x^m_{i_m} \right)^2 \\ \text{subject to} \quad & \|x^i\|^2 = 1, i = 1, \ldots, m \end{aligned}$$

Relaxations actually work fine. (see Nie and Wang *Semidefinite Relaxations for Best Rank-1 Approximations*

Section 8

A few more nonnegativity certificates

A well known nonnegativity certificate

An important case of constrained POP is that of linear programming:

Linear Programming

Given affine polynomials $p(\mathbf{x}), g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_m(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ we want to find

$$p^* = \inf_{\chi \in S} p(\chi).$$

where $S = \{\chi \in \mathbb{R}^n \mid g_1(\chi) \geq 0, \ldots, g_m(\chi) \geq 0\}$.

We are naturally not suggesting using polynomial optimization techniques to solve this problem, but note that p_1^{sos} restricts to

$$p_1^{\text{sos}} = \sup \lambda \text{ such that } p(\mathbf{x}) - \lambda = a_0 + \sum_{i=1}^{m} a_i g_i(\mathbf{x})$$

for $a_i \geq 0$. This is the certificate guaranteed to exist by Farkas' Lemma. Hence $p^P = p_1^{\text{sos}}$ in this case.

Linear Constrains

Somewhere between the LP and the full-fledged POP, sits the case of optimizing a general polynomial under affine constrains.

Linear constrained POP

Given a polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ and affine polynomials $l_1(\mathbf{x}), l_2(\mathbf{x}), \dots, l_m(\mathbf{x})$ we want to find

$$p^* = \inf_{\chi \in P} p(\chi).$$

where $P = \{\chi \in \mathbb{R}^n \mid l_1(\chi) \geq 0, \dots, l_m(\chi) \geq 0\}$ is a polytope.

In this case we have a simple certificate.

Theorem (Handelman - 1988)

A polynomial $p(\mathbf{x})$ is positive over a polytope P if and only if there exists a finite subset $\mathcal{I} \subseteq \mathbb{N}^m$ and $\lambda_I \geq 0$ for all $I \in \mathcal{I}$ such that

$$p(\mathbf{x}) = \sum_{I \in \mathcal{I}} \lambda_I l_1(\mathbf{x})^{I_1} l_2(\mathbf{x})^{I_2} \cdots l_m(\mathbf{x})^{I_m}.$$

This can be used to derive an LP hierarchy of approximations, on the same spirit of Lasserre's.

Schmüdgen's certificates

Recall that in our nonnegativity certificates over a basic closed semialgebraic set

$$S = \{\chi \in \mathbb{R}^n \mid g_1(\chi) \geq 0, \dots, g_m(\chi) \geq 0\}$$

we try to represent

$$p(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sigma_1(\mathbf{x})g_1(\mathbf{x}) + \sigma_2(\mathbf{x})g_2(\mathbf{x}) + \cdots + \sigma_m(\mathbf{x})g_m(\mathbf{x})$$

where all σ are sums of squares.

This is not the most powerful form one could think of. For $J \subseteq \{1, \ dots, m\}$ let $g_J(\mathbf{x}) = \prod_{j \in J} g_j(\mathbf{x})$. All these are nonnegative over S so we could search for certificates

$$p(\mathbf{x}) = \sum_{J \subseteq \{1,\dots,m\}} \sigma_J(\mathbf{x}) g_J(\mathbf{x}).$$

We denote by $T_S[\mathbf{x}]$ the set of all such polynomials,and by $T_S^d[\mathbf{x}]$ those for which $\sigma_J g_J$ has degree at most $2d$ for all J.

Schmüdgen's certificates - Part 2

We can now establish another relaxation for constrained POP.

Another sos hierarchy

Given polynomials $p(\mathbf{x}), g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_m(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ and

$$S = \{\chi \in \mathbb{R}^n \mid g_1(\chi) \geq 0, \ldots, g_m(\chi) \geq 0\},$$

we define the $\overline{p}_d^{\text{sos}}$ as

$$\overline{p}_d^{\text{sos}} = \sup \lambda \text{ such that } p(\mathbf{x}) - \lambda \in T_S^d[\mathbf{x}].$$

Note that this is till and SDP, just a larger one with 2^m psd constrains.
We also have $\overline{p}_d^{\text{sos}} \geq p_d^{\text{sos}}$ for all p since $Q_S^d[\mathbf{x}] \subseteq T_S^d[\mathbf{x}]$.

Theorem (Schmüdgen's Positivstellensatz - 1991)

If S is compact and $p(\mathbf{x})$ is positive over S then $p(\mathbf{x}) \in T_S[\mathbf{x}]$.

Hence $\overline{p}_d^{\text{sos}} \to p^*$ if S is compact. It is harder to compute but it might conceivably converge much faster in some instances.

Sums of Binomial Squares

Sometimes SDP is still too hard. We can try to compromise by limiting our certificates.
We say that $p(\mathbf{x})$ is a sum of binomial squares (sobs), if it can be written as

$$p(\mathbf{x}) = \sum_{i=1}^{t} (a_i x^{\alpha_i} + b_i x^{\beta_i})^2.$$

Clearly sobs implies sos. So it is a weaker certificate. Why bother?

Observation

A polynomial $p(\mathbf{x})$ of degree $2d$ is sobs if and only if

$$p(\mathbf{x}) = \mathbf{x}_d^t Q \mathbf{x}_d$$

for some $Q = UU^t$ where every column of U has at most two nonzero entries.

Such Q are known as factor width 2 matrices. They are precisely the scaled diagonally dominant matrices.
Checking if Q is of that form is an SOCP!

Sums of Binomial Squares - Part 2

It is possible to create an SOCP out of that.

SDSOS optimization - Ahmadi-Majumdar 2017

Given a polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ find

$$p^{\text{sobs}} = \sup \lambda \text{ such that } p(\mathbf{x}) - \lambda \text{ is sobs.}$$

Or we can make a hierarchy of SOCP's.

SDSOS hierarchy - Ahmadi-Majumdar 2017

Given a polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ find

$$p_r^{\text{sobs}} = \sup \lambda \text{ such that } (1 + x_1^2 + \cdots + x_n^2)^r (p(\mathbf{x}) - \lambda) \text{ is sobs.}$$

What do we know about this?

1. It works nicely in many large control applications.
2. But it might not converge to the optimum!!!
3. It can be improved in several ways.

Section 9

Final Remarks

To know more:

- Blekherman, G., Parrilo, P. A., & Thomas, R. R. (Eds.). (2012). *Semidefinite optimization and convex algebraic geometry*. SIAM.
- Lasserre, J. (2015). *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge Texts in Applied Mathematics.
- Lasserre, J. (2009). *Moments, positive polynomials and their applications*. World Scientific.
- Laurent, M. (2009). *Sums of squares, moment matrices and optimization over polynomials*. In *Emerging applications of algebraic geometry* (pp. 157-270). Springer.
- Marshall, M. (2008). *Positive polynomials and sums of squares* (No. 146). AMS.