Sums of Squares in Polynomial Optimization
Lecture 1

João Gouveia

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

20th of May 2019 - IPCO Summer School

Section 1

Unconstrained Polynomial Optimization and Nonnegativity

The problem

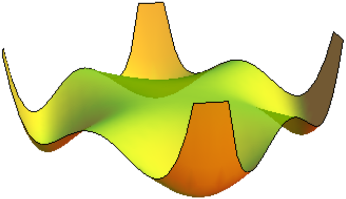
We will start by studying the simplest polynomial optimization problem.

Unconstrained POP

Given a polynomial $p(\mathbf{x}) = p(x_1, \cdots, x_n) \in \mathbb{R}[\mathbf{x}]$ find

$$p^* = \inf_{\chi \in \mathbb{R}^n} p(\chi).$$

How hard can that be?



The univariate case - a classic approach

Lets keep it simple. What can we do when p is a univariate polynomial?

It is almost equivalent to detecting real roots, which is somewhat easy, but not trivial. We have tools for that.

Sturm's Sequence

Given a univariate polynomial p of degree d, we define the Sturm's sequence of the polynomials by:

$$P_0 = P$$

$$P_1 = P'$$

$$P_{i+1} = -\text{remainder of the division of } P_{i-1} \text{ by } P_i$$

for $i = 1, \ldots, d-1$.

For $P_0 = p = 3x^4 - 4x^3 + 12x^2 - 24x + 10$ we get

$P_1 = 12x^3 - 12x^2 + 24x - 24,$ $P_2 = -5x^2 + 16x - 8,$

$P_3 = \frac{-2232x+1656}{25}$ $P_4 = \frac{-1075}{961}.$

The univariate case (continuation)

This is enough to locate all roots:

Theorem (Sturm's Theorem (1829))

Given a univariate polynomial of degree d, p, and $\chi \in \mathbb{R}$ denote by $w(\chi)$ the number of sign changes in the sequence

$$P_0(\chi), P_1(\chi), \cdots, P_d(\chi)$$

then the number of distinct real zeros of p in the interval $(a, b]$ equals $w(a) - w(b)$, this extends to $a = +\infty$ and $b = -\infty$ by looking at the signs of leading monomials.

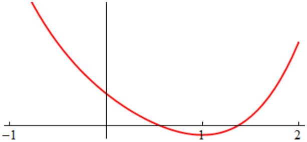
Recall, we had $P_0 = 3x^4 - 4x^3 + 12x^2 - 24x + 10$ and

$$P_1 = 12x^3 - 12x^2 + 24x - 24, \qquad P_2 = -5x^2 + 16x - 8,$$
$$P_3 = \frac{-2232x + 1656}{25} \qquad P_4 = \frac{-1075}{961}.$$

Evaluating at 0 we get $(+, -, -, +, -)$ so $w(0) = 3$.
Evaluating at $+\infty$ we get $(+, +, -, -, -)$ so $w(+\infty) = 1$.
There are 2 roots in $(0, +\infty)$.
Evaluating at $-\infty$ we get $(+, -, -, +, -)$ so $w(-\infty) = 3$.
There are no more roots.

The univariate case (continuation)

There are two distinct real roots.



A polynomial p is positive if and only if $w(+\infty) = w(-\infty)$ and $p(0) > 0$.
This is what is called a **certificate of positivity**.

Observations

1. We could locate the zeros of the derivative and test them.
2. We can also search for the maximal $\lambda \in \mathbb{R}$ such that $p - \lambda$ has no real roots by using the bisection method, for instance.
3. In practice, versions implementing variants of the Descartes rule of signs are faster and more robust.

The general case

The general case is not as simple.

Theorem

For polynomials of degree 4 in n variables, deciding if $p^* = 0$ is NP-hard.

Proof: A number of hard combinatorial problems can be translated into this decision problem. For instance, the **partition problem** asks if given a set $\{a_1, \ldots, a_n\}$ of positive integers one can partition it into two sets of equal sum. This is NP-hard. Equivalently this is the same as asking if

$$\min_{\chi \in \mathbb{R}^n} \left(\sum_{i=1}^{n} a_i \chi_i \right)^2 + \sum_{i=1}^{n} (\chi_i^2 - 1)^2 = 0.$$

Observation

1. For odd degree, $p^* - \infty$.
2. For degree 2 solve a linear system and check the Hessian.

Why should we solve this?

This is a hard problem. So before attempting to tackle it, we should make a case of why should we do it. Polynomials are incredibly versatile tools, and can capture an array of important problems.

Distance Graph Realization Problem

Given a graph $G = ([n], E)$ and some distance information d_{ij} for all $\{i, j\} \in E$, there is a realization in $\mathbb{R}^k$ with those distances if and only if

$$\min_{\chi \in \mathbb{R}^{k \times n}} \sum_{\{i,j\} \in E} (\|\chi_i - \chi_j\|^2 - d_{ij}^2)^2 = 0.$$

Independence number of a graph [via Motzkin-Straus]

Given a graph $G = ([n], E)$ its **independence number** verifies $\alpha(G) \geq t$ if and only if for the matrix $M^t = t(A_G + I) - J$ we have

$$\min_{\chi \in \mathbb{R}^n} \sum_{i=1}^{n} \sum_{j=1}^{n} \chi_i^2 \chi_j^2 M_{ij}^t = 0.$$

Re-framing the POP

Lets rewrite the unconstrained POP in a trivial way. Denote by $\mathcal{P}[\mathbf{x}]$ the set of globally nonnegative polynomials on $\mathbf{x}$.

Unconstrained POP - v2.0

Given a polynomial $p(\mathbf{x}) = p(x_1, \cdots, x_n) \in \mathbb{R}[\mathbf{x}]$ find

$$p^* = \sup_{\lambda \in \mathbb{R}} \{\lambda \mid p(x) - \lambda \in \mathcal{P}[\mathbf{x}]\}.$$

Advantages: Certificates of nonnegativity can be leveraged into optimization schemes. If nothing else, by using bisection methods as we did with Sturm root counting certificate for univariate polynomials.

Moreover, if the certificates are nice enough, we can replace $\mathcal{P}[\mathbf{x}]$ by the set of certifiably nonnegative polynomials, and maybe directly optimize over that, attaining directly a lower bound for p^*.

Section 2

Sums of Squares and Nonnegativity

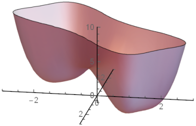
Sums of squares

A simple certificate for nonnegativity of a polynomial $p \in \mathbb{R}[\mathbf{x}]$ is being a **sum of squares** of other polynomials, i.e,

$$p(\mathbf{x}) = \sum_{i=1}^{t} (h_i(\mathbf{x}))^2,$$

for some $h_i \in \mathbb{R}[\mathbf{x}]$. In that case we say p is a sum of squares or sos, and we denote the set of all such polynomials by $\Sigma[\mathbf{x}]$. Clearly $\Sigma[\mathbf{x}] \subseteq \mathcal{P}[\mathbf{x}]$.

Example: Consider the polynomial $p(x,y) = x^4 - 4x^3y + 7x^2y^2 - 4xy^3 - 4xy + y^4 + 4$.



$$p(x,y) = (x-y)^4 + (xy-2)^2 \in \Sigma[x,y]$$

Univariate polynomials revisited

Once again, univariate polynomials turn out to be very nice.

Proposition

A univariate polynomial is nonnegative if and only if it is sos.

Proof: Suppose p is a nonnegative univariate polynomial.
Any real root of a nonnegative polynomial must have even multiplicity, as the zero must be a local minimum. Hence

$$p(x) = c^2(x - r_1)^{2m_1} \cdots (x - r_k)^{2m_k}((x - a_1)^2 + b_1^2) \cdots ((x - a_l)^2 + b_l^2)$$

where $a_j \pm ib_j$ are the complex roots of p. Distributing the sums we get a sum of 2^l squares.

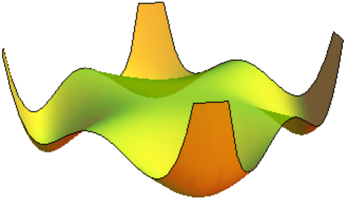
Observation

By noting $(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 - (ad - bc)^2$ one can do it with two squares only.

Motzkin Example - 1967

Proposition

Motzkin's polynomial $p(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ *is nonnegative but not a sum of squares.*



Motzkin Example (cont)

Proposition

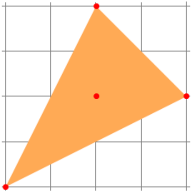
Motzkin's polynomial $p(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ *is nonnegative.*

Proof: By AM/GM inequality we have

$$\frac{x^4y^2 + x^2y^4 + 1}{3} \geq \sqrt[3]{x^4y^2 \times x^2y^4 \times 1} = x^2y^2,$$

so p is in fact nonnegative.

To show that Motzkin is not sos, we will need an auxiliary Lemma. Recall that the Newton Polytope of a polynomial p, $\mathrm{N}(p)$, is the convex hull of the vectors of exponents of the monomials of p.



Motzkin Example (cont.)

Lemma

If $p = \sum h_i^2$ then for every i, we have $\mathrm{N}(p) = \operatorname{conv}\left(\bigcup_i \mathrm{N}(h_i^2)\right)$ *in particular we have the polytope inclusion*

$$2\mathrm{N}(h_i) \subseteq \mathrm{N}(p).$$

Sketch of proof: By contradiction take a vertex $x^{2\alpha}$ of the convex hull of the union that is not in $\mathrm{N}(p)$. Since it was a squared monomial it appears with positive coefficient, hence to be canceled there should be $c_\beta x^\beta - c_\gamma x^\gamma$ in other h_i with $\beta + \gamma = 2\alpha$. But then 2α is in the segment $[2\beta, 2\gamma]$ and was not a vertex.

Proposition

Motzkin's polynomial $p(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ *is not sos.*

Proof: By the previous Lemma, if it was sos the squares would be of the form

$$(a + bxy + cx^2y + cxy^2)^2$$

none of which can have negative coefficients for x^2y^2.

Hilbert's Theorem

When do sums of squares work perfectly? Let $\mathcal{P}_d^n$ and Σ_d^n be the sets of respectively nonnegative and sos polynomials with n variables and degree d.

Theorem (Hilbert's Theorem - 1888)

We have that $\mathcal{P}_{2d}^n = \Sigma_{2d}^n$ in exactly the following cases:

1. $n = 1$ *(univariate polynomials)*
2. $d = 1$ *(quadratic polynomials)*
3. $n = 2$ *and* $d = 2$ *(bivariate quartic polynomials)*

We saw a counterexample for $n = 2$ and $d = 3$, and a similar argument can be used to find a counterexample for $n = 3$ and $d = 2$. **[Exercise]**

Observations

1. In fact Hilbert proved that every bivariate quartic polynomial is the sum of at most three squares.
2. For some deeper history and many examples check Bruce Reznicks's awesome paper *Some concrete aspects of Hilbert's 17th problem*.

Blekherman's Theorem

Unfortunately, it mostly does not work.

Theorem (Blekherman's Theorem - 2006)

For an even fixed degree $d \geq 4$, we have

$$\lim_{n \to \infty} \sqrt[N]{\frac{\operatorname{vol}\left(\widehat{\Sigma_d^n}\right)}{\operatorname{vol}\left(\widehat{\mathcal{P}_d^n}\right)}} = 0$$

where $\widehat{K}$ just means we intersect K with polynomials with integral 1 on the unit ball, and N is the dimension of $\mathcal{P}_d^n$.

Observations

1. The result is actually much more precise, this is a coarse simplification.
2. It relies on a clever application of Urysohn's inequality.

There are almost no sums of squares. So why use it?

The case for sums of squares

Why use sums of squares certificates?

1. They are simple, and very versatile.
2. If one is found, it gives a short, easy to verify, proof of nonnegativity.
3. They can be strengthened.
4. **Most important**: We can actually find them “efficiently”.

In what follows I will try to convince you of all these points, starting from the last one.

Section 3

Sums of Squares and Semidefinite Programming

Semidefinite matrices

Recall that a symmetric matrix $M \in \mathbb{R}^{n\times n}$ is **positive semidefinite** (psd), denoted by $M \succeq 0$, if and only if $x^t M x \geq 0$ for all $x \in \mathbb{R}^n$. Equivalently

- All eigenvalues of M are nonnegative;
- $M = UU^t$ for some $U \in \mathbb{R}^{n\times m}$;
- $M = \sum_{i=1}^k v_i v_i^t$ for some $v_i \in \mathbb{R}^n$.

Application: $d = 2$

A quadratic polynomial is nonnegative if and only if it is sos.

Note that a quadratic $q(\mathbf{x})$ can be written as

$$q(\mathbf{x}) = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}^t Q \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}$$

for some real Q symmetric. The quadratic q being nonnegative is equivalent (almost) by definition to $Q \succeq 0$. By the condition above that means

$$q(\mathbf{x}) = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}^t \left(\sum_{i=1}^n v_i v_i^t \right) \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}$$

hence $q(\mathbf{x}) = \sum_{i=1}^n \langle v_i, (1, \mathbf{x}) \rangle^2$ and is sos.

Semidefinite matrices and sums of squares

The second part of this idea still works for general sums of squares.

Proposition

Let $p(\mathbf{x})$ be an n variable degree $2d$ polynomial, and $\mathbf{x}_d$ the vector of all monomials of degree at most d. Then p is sos if and only if there exists a semidefinite matrix Q such that $p(\mathbf{x}) = \mathbf{x}_d^t Q \mathbf{x}_d$.

Proof: Suppose $p(\mathbf{x}) = \sum_{i=1}^k h_i(\mathbf{x})^2$.
Each $h_i(x)$ has degree at most d and can be written as $h_i(\mathbf{x}) = \langle \tilde{h}_i, \mathbf{x}_d \rangle = \tilde{h}_i^t \mathbf{x}_d$. Hence

$$p(\mathbf{x}) = \sum_{i=1}^k h_i(\mathbf{x})^2 = \sum_{i=1}^k \mathbf{x}_d^t \tilde{h}_i \tilde{h}_i^t \mathbf{x}_d = \mathbf{x}_d^t \left(\sum_{i=1}^k \tilde{h}_i \tilde{h}_i^t \right) \mathbf{x}_d.$$

Calling $Q = \sum_{i=1}^k \tilde{h}_i \tilde{h}_i^t$ we get the result.
The other direction is the same reasoning in reverse order.

Example

Consider the polynomial $p(x,y) = x^4 - 4x^3y + 7x^2y^2 - 4xy^3 - 4xy + y^4 + 4$. It is sos if there exists $Q \succeq 0$ such that

$$\begin{bmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{bmatrix}^t \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} & q_{15} & q_{16} \\ q_{12} & q_{22} & q_{23} & q_{24} & q_{25} & q_{26} \\ q_{13} & q_{23} & q_{33} & q_{34} & q_{35} & q_{36} \\ q_{14} & q_{24} & q_{34} & q_{44} & q_{45} & q_{46} \\ q_{15} & q_{25} & q_{35} & q_{45} & q_{55} & q_{56} \\ q_{16} & q_{26} & q_{36} & q_{46} & q_{56} & q_{66} \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{bmatrix} = p(x,y)$$

In other words, $Q \succeq 0$ verifying

$$\begin{array}{ccccc} 1 = q_{44}, & -4 = 2q_{45}, & 7 = q_{55} + 2q_{46}, & -4 = 2q_{56} & -4 = 2q_{15} + 2q_{23} \\ 1 = q_{66}, & 4 = q_{11}, & 0 = 2q_{12} & 0 = 2q_{13} & 0 = q_{22} + 2q_{14} \\ 0 = q_{33} + 2q_{16} & 0 = 2q_{24} & 0 = 2q_{26} + 2q_{35} & 0 = 2q_{36} & 0 = 2q_{34} + 2q_{25} \end{array}$$

15 monomials of degree less or equal 4 $\longleftrightarrow$ 15 linear equations

Example (cont.)

Plugging in the equations we want to find

$$Q = \begin{bmatrix} 4 & 0 & 0 & q_{14} & q_{15} & q_{16} \\ 0 & -2q_{14} & -2 - q_{15} & 0 & q_{25} & q_{26} \\ 0 & -2 - q_{15} & -2q_{16} & -q_{25} & -q_{26} & 0 \\ q_{14} & 0 & -q_{25} & 1 & -2 & q_{46} \\ q_{15} & q_{25} & -q_{26} & -2 & 7 - 2q_{46} & -2 \\ q_{16} & q_{26} & 0 & q_{46} & -2 & 1 \end{bmatrix} \succeq 0$$

For instance

$$Q = \begin{bmatrix} 4 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ -2 & 0 & 0 & -2 & 5 & -2 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix} \succeq 0.$$

In this case

$$p(x,y) = \langle (2,0,0,0,-1,0), \mathbf{x}_d \rangle^2 + \langle (0,0,0,1,-2,1), \mathbf{x}_d \rangle^2 = (xy-2)^2 + (x^2 - 2xy + y^2)^2.$$

Observations

1. Q is not unique.
2. $\operatorname{rank}(Q)$ corresponds to number of squares.
3. Finding Q is a feasibility problem on a **semidefinite program**.

Semidefinite programming

Semidefinite programs come in primal-dual pairs. They are of the form:

Primal problem

$$\begin{aligned} p^* = \min_X \quad & \langle A_0, X\rangle \\ \text{s.t.} \quad & \langle A_i, X\rangle = b_i,\, i = 1, \ldots, m \\ & X \succeq 0. \end{aligned}$$

Dual problem

$$\begin{aligned} d^* = \max_y \quad & \langle b, y\rangle \\ \text{s.t.} \quad & A_0 - \textstyle\sum_{i=1}^m y_i A_i \succeq 0. \end{aligned}$$

where $A_0, \ldots, A_m$ are real symmetric matrices.

Semidefinite programming is a generalization of linear programming. Using interior point methods they can *efficiently* be solved to arbitrary precision.

We saw:

Proposition

Σ_d^n is the feasible set of a semidefinite programming of size $\binom{n+d}{d}$.

In other words, we can efficiently optimize over the set of sums of squares polynomials, but complexity grows with n^d.

The sum of squares relaxation

We can now replace the unconstrained POP by the sos relaxation.

SOS optimization

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

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

We know that $p^{\text{sos}} \leq p^*$ so we always get a lower bound.

It can be trivial:
If p is the Motzkin polynomial, we proved that $p^{\text{sos}} = -\infty$.

It can be perfect:
Always if p is univariate, quadratic or a bivariate quartic, but also in other cases if we are lucky.

Another example

Let us revisit the problem

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

Step 1: max λ such that $p(x) - \lambda \in \Sigma[x]$
Step 2: max λ such that $Q \succeq 0$ and

$$p(x) - \lambda = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}^t \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}$$

Step 3: max λ such that there exists γ with

$$\begin{bmatrix} 10 - \lambda & -12 & \gamma \\ -12 & 12 - 2\gamma & -2 \\ \gamma & -2 & 3 \end{bmatrix} \succeq 0$$

Step 4: Solve the sdp:

```
sdpvar  l g
optimize([10-l,-12,g;-12,12-2*g,-2;g,-2,3]>=0,-l)
```

We get the optimum at $\lambda = -3.0000$. **But what is the minimizer?**

Parentheses: From theory to practice

1. Many solvers for semidefinite programas are available:

   `sdpt3, mosek, sedumi, ...`

2. A few toolboxes have implemented sums of squares automatically

   `sostools, gloptipoly, yalmip, ...`

3. While theoretically the solution to an SDP can be attained in algebraic form, its degree would be impractical (see *The Algebraic Degree of Semidefinite Programming* by Nie, Ranestad and Sturmfels) so we must make do with floating point approximations.

4. Recovering rigorous sos certificates from the sdp is an art as well as a science (see *A Macaulay 2 package for computing sum of squares decompositions of polynomials with rational coefficients* by Peyrl and Parrilo and follow up work by Kaltofen, Zhi, El Din and others).

Section 4

Strengthening Sums of Squares

So sums of squares did not work, now what?

One can try a stronger certificate.

Theorem (Artin - 1927 - *Hilbert's 17th problem*)

Every nonnegative polynomial can be written as a sum of squares of rational functions.

In other words, $p(\mathbf{x}) \geq 0$ iff there are $h_i(\mathbf{x})$ and $g_i(\mathbf{x})$, $i = 1, \ldots, t$ such that

$$p(\mathbf{x}) = \left(\frac{h_1(\mathbf{x})}{g_1(\mathbf{x})}\right)^2 + \left(\frac{h_2(\mathbf{x})}{g_2(\mathbf{x})}\right)^2 + \cdots + \left(\frac{h_t(\mathbf{x})}{g_t(\mathbf{x})}\right)^2 .$$

Checking this is not easy though.

Idea: Consider a uniform denominator: Fix a nonnegative polynomial $q(x)$ and try to write $q(x)p(x)$ as a sum of squares.

A usual candidate would be $q(x) = (1 + x_1^2 + x_2^2 + \cdots + x_n^2)$.

Using multipliers

Lets try it with Motzkin.

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

$$(1+x^2+y^2)p(x,y) = (x^2-y)^2+(xy^2-x)^2+(x^2y^2-1)^2+\frac{1}{4}(xy^3-x^3y)^2+\frac{3}{4}(xy^3+x^3y-2xy)^2$$

It works!

In fact it kind of always does...

Theorem (Reznick - 1995)

If $\inf p(\mathbf{x}) > 0$ *then there exists* r *such that*

$$(1 + x_1^2 + x_2^2 + \cdots + x_n^2)^r p(\mathbf{x}) \in \Sigma[\mathbf{x}].$$

Bounds on r are not great. But it works surprisingly well in many applications.

A multiplier based hierarchy

We can weaponize this certificate into a hierarchy of relaxations.

SOS optimization hierarchy

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

$$p_r^{\text{sos}} = \sup \lambda \text{ such that } (1 + x_1^2 + x_2^2 + \cdots + x_n^2)^r (p(\mathbf{x}) - \lambda) \in \Sigma[\mathbf{x}].$$

This is an increasing sequence of lower bounds attained by semidefinite programming. Reznick's theorem can be seen as a convergence result.

Corollary

For any polynomial $p(\mathbf{x})$, we have

$$\lim_{r\to\infty} p_r^{\text{sos}} = p^*.$$

A simple example

Consider an example from Leep-Star via Reznick

$$p(x,y) = 3x^4y^2 - 2x^3y^3\frac{1}{2}x^2y^4 + 4x^3y^3 + x^2y^3 + 10x^2y^2 + 2xy^2 - 8x^2y + 4xy.$$

Compute p^{sos}:

```
sdpvar x y t
p=3*x^4*y^2-2*x^3*y^3+(x^2*y^4)/2+4*x^3*y^3+x^2*y^3
                  +10*x^2*y^2+2*x*y^2-8*x^2*y+4*x*y
solvesos(sos(p-t),-t,[],t)
```

We get $t = -40.9412$. Repeat for p_1^{sos}

```
solvesos(sos((x^2+y^2)*(p-t)),-t,[],t)
```

We get $t = -5.765$. After that it stabilizes. It is actually the true answer.