Minimizing Polynomials on Semialgebraic Sets

Vicki Powers
Emory University

http://www.mathcs.emory.edu/~vicki

vicki@mathcs.emory.edu
To Begin

Special thanks to the organizers for....

- holding the conference during my spring break
- letting me speak first ....
... so I can have wine with lunch and dinner for 5 days

Disclaimer!

For every topic mentioned in this talk, there is at least one conference participant who knows more about it than the speaker.
Notation

- We work in $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$. $f \in \mathbb{R}[X]$ is positive semidefinite (psd) if $f$ takes only non-negative values, we write $\sum \mathbb{R}[X]^2$ for the set of sums of squares in $\mathbb{R}[X]$ and say $f \in \mathbb{R}[X]^2$ is sos.

- Given $F = \{g_1, \ldots, g_r\} \subseteq \mathbb{R}[X]$, let $S(F)$ be the basic closed semialgebraic set generated by $F$, i.e., $S(F) = \{\alpha \in \mathbb{R}^n | g_i(\alpha) \geq 0, i = 1, \ldots, r\}$.

We have two algebraic objects in $\mathbb{R}[X]$ associated to $F$:

- The **quadratic module** generated by $F$, $M(F) := \{s_0 + s_1 g_1 + \cdots + s_r g_r | \text{each } s_i \text{ is sos}\}$.

- The **preorder** generated by $F$, $PO(F) := \left\{ \sum_{\epsilon \in \{0,1\}^n} s_\epsilon g_1^{\epsilon_1} \cdots g_r^{\epsilon_r} | s_\epsilon \text{ is sos}\right\}$. 

Positive Polynomials, Luminy, March 2005 – p.3/29
The Basic Problem

Given $S = S(F)$ and $f \in \mathbb{R}[X]$, find

$$f^* = \text{minimum of } f \text{ on } S$$

We assume (without proof!) that

- this is a hard problem
- this is an interesting problem

An observation: $f^* = \text{maximum } \{ \lambda \in \mathbb{R} \mid f - \lambda \geq 0 \text{ on } S \}$.

Another observation: If $f \in M(F)$ or $f \in PO(F)$, then $f \geq 0$ on $S$. Furthermore, a representation of $f$ in either $M(F)$ or $PO(F)$ is an explicit witness to the non-negativity of $f$ on $S$. 
Some definitions

Throughout, we will fix \( F = \{g_1, \ldots, g_r\} \subseteq \mathbb{R}[X] \) and let \( S = S(F) \), \( M = M(F) \), and \( P = PO(F) \).

Following the above observations, it makes sense to define

\[
    f_{sos} := \text{maximum} \ \{\lambda \mid f - \lambda \in M\}
\]

\[
    \hat{f}_{sos} := \text{maximum} \ \{\lambda \mid f - \lambda \in P\}
\]

Clearly, we have

\[
    f_{sos} \leq \hat{f}_{sos} \leq f^*
\]

We’ll call this an f \( \text{SOS relaxation} \) of the original problem.
What’s the point?

Thanks to the magic of linear algebra, the question “is $h$ sos?” is more computationally tractable than “is $h$ psd”:

“$h(X)$ is sos” is equivalent to the existence of a psd matrix $A$ such that

$$h(X) = V^T \cdot A \cdot V,$$

where $V$ is the vector of monomials of degree $\leq \frac{1}{2} \deg h$.

Using this fact, the computation of $f_{sos}$ and $\hat{f}_{sos}$ can be implemented as a semidefinite programming problem (SDP) and hence solved numerically!

...not precisely the truth, but close enough at this point in the talk!
The Global Case

Suppose we take $S = \mathbb{R}^n$, so that $f^*$ is the global minimum of $f$. Then $f_{sos} = \hat{f}_{sos} = \max\{\lambda \mid f - \lambda \text{ is sos}\}$.

The idea of using $f_{sos}$ to approximate $f^*$ in this case goes back to N. Z. Shor in the 1950’s. Computation of $f_{sos}$ gives a lower bound for $f^*$ and when the degree of $f$ is fixed the bound $f_{sos}$ can be computed in polynomial time using an SDP.

However, the bound may not be useful! For example, if $m$ is the Motzkin polynomial, then $m^* = 0$ but $m - \lambda$ is not sos for any $\lambda$, so that $m_{sos} = -\infty$.

Little is known about when this works, how good the bound is, etc.
On the one hand, Parrilo and Sturmfels computed a family of examples of a particular form – with leading form \( \sum x_i^{2d} \) – and found that in all cases tested \( f_{sos} \) was finite and very close to \( f^* \) (within the range of numerical error).

On the other hand, Blekherman showed that for large \( n \), there are many more psd polynomials than sos polynomials. So the method shouldn’t work well!

**Problem**: Explain the experimental results. Find results about when the method works well and/or gives explicit bounds.
Fix $F = \{g_1, \ldots, g_r\} \subseteq \mathbb{R}[X]$, $S = S(F)$, $M = M(F)$, and $P = PO(F)$. As above, we have

$$f_{sos} \leq \hat{f}_{sos} \leq f^*.$$  

As in the global case, we want an SOS relaxation to the problem of finding $f^*$ and hence a computationally feasible (via SDPs) method for approximating $f^*$.

Two important issues:

- When do we have $f_{sos} = f^*$ or $\hat{f}_{sos} = f^*$?
  $\Rightarrow$ Representation Theorems

- Can we define an SDP whose solution is $f_{sos}$ or $\hat{f}_{sos}$?
  $\Rightarrow$ Stability of quadratic modules/preorders.
A quadratic module $M$ is archimedean if one of the following (non-trivially) equivalent conditions holds:

- There is $p \in M$ such that $\{p \geq 0\}$ is compact.
- There is $N \in \mathbb{N}$ such that $N - \sum x_i^2 \in M$.
- For all $p \in M$, there is $N \in \mathbb{N}$ such that $N - p \in M$.

From the work of Jacobi and Prestal, there are other nice criteria.

\begin{advertisement}
Alex Prestal will be talking about archimedean modules later today. Don’t miss it!
\end{advertisement}
Notice that if $M$ is archimedean, then $S$ must be compact, however, the converse is not true in general.

However in the preorder case, $S$ compact implies that $P$ is archimedean. This follows work of Schmüdgen; there is a nice algebraic proof due to Wörmann.

**Theorem (Putinar).** If $M$ is archimedean and $g > 0$ on $S$, then $g \in M$.

**Theorem (Schmüdgen).** If $S$ is compact, then $P$ is archimedean. Hence $S$ compact, then $g > 0$ on $S$ implies $g \in P$.

This means that if $S$ is compact, then $\hat{f}_{sos} = f^*$ and if $M$ is archimedean, then $f_{sos} = f^*$. That’s the good news.

Now for the bad news...
Stability

Suppose \( \deg p = 2d \) and \( p \) is sos, say

\[
p = h_1^2 + \cdots + h_k^2.
\]

Then comparing leading terms on both sides of the equation, we see that \( \deg h_i \leq d \) for all \( i \): No “leading term cancellation" is possible. Thus we can implement “maximize \( \lambda \) subject to \( f - \lambda \) sos" directly as an SDP.

However, given \( p \in M \), say

\[
p = s_0 + s_1 g_1 + \cdots + s_r g_r,
\]

in general there is no \textit{a priori} bound in terms of \( \deg p \) on the \( s_i \)'s. There can be lots of leading term cancellation in this case!
Quadratic modules where there is such a bound are called stable.

It turns out that $M$ is rarely stable. In fact, very roughly speaking, Scheiderer shows that in almost all cases, if SOS representations exist for all polynomials positive on a semialgebraic set, then the corresponding quadratic module won’t be stable.

This is an example of the TANSTAAFL principal: There Ain’t No Such Thing As A Free Lunch!

\begin{advertisement}
Claus Scheiderer will be talking about this tomorrow. Don’t miss it!
\end{advertisement}
The Lasserre Method

If $M$ is not stable, then we can’t implement the problem
Maximize $\lambda$ subject to $f - \lambda \in M$
directly as an SDP. Instead we will have to approximate by a series of SDPs corresponding to bounding the degree of a representation.

For $d \in \mathbb{N}$, let

$$M_d = \{s_0 + s_1 g_1 + \cdots + s_r g_r \in M \mid \deg(s_i g_i) \leq d\},$$

and let

$$f_{sos}^d := \max\{\lambda \mid f - \lambda \in M_d\}.$$

**Theorem (Lasserre).** Suppose $M$ is archimedean, then $\{f_{sos}^d\}$ is an increasing sequence that converges to $f^*$. 
As mentioned above, $M$ is rarely stable. In particular, if $S$ is compact, then $M$ is not stable. Thus even if $f^* = f_{sos}^d$ for some $d$, there is in general no bound on $d$ which does not depend on $f^*$ itself!

In general, it seems that little is known about bounds, convergence, etc., However, work of Schweighofer yields a bound in the case where $S = S(\{g\})$ is defined by a single inequality. In this case there is a constant $c \in \mathbb{N}$ depending on $\deg f$ and $g$ and a constant $b \in \mathbb{N}$ depending on $g$ such that

$$f^* - f_{sos}^d \leq \frac{c}{b \sqrt{d}}$$

for large $d$. 
Quadratic Modules vs. Preorders

In the case where \( S \) is compact, why not use the preorder since it is automatically archimedean? The problem is that the generators of the preorder are all possible products of the elements in \( F \) and hence if the set \( F \) has \( r \) elements, then the preorder has \( 2^r \) generators.

Recall that \( M \) is archimedean if there is some \( N \in \mathbb{N} \) such that \( N - \sum x_i^2 \in M \). In practical applications we might know, or can compute, some \( N \) so that \( S \) is contained in a ball of radius \( \sqrt{N} \) about \( 0 \). Then we can simply add the polynomial \( N - \sum x_i^2 \) to \( F \) and use the quadratic module.
We assume again that $S = \mathbb{R}^n$ so that $f^*$ is the global minimum of $f$.

The real gradient ideal of $f$, $I_{\text{grad}}(f)$, is the ideal in $\mathbb{R}[X]$ generated by the partial derivatives of $f$.

We also consider the gradient variety of $f$, $V_{\text{grad}}(f) = V(I_{\text{grad}}(f)) \subseteq \mathbb{C}^n$ and

the real gradient variety $V_{\text{grad}}^\mathbb{R}(f) = V_{\mathbb{R}}(I_{\text{grad}}(f)) \subseteq \mathbb{R}^n$.

All local and global minima of $f$ occur at points in the real gradient variety of $f$ and so to compute $f^*$ we can minimize $f$ on the real variety $V_{\text{grad}}^\mathbb{R}(f)$. 
There are several recent works on minimizing polynomials via the gradient ideal, among them:

- Hanzon and Jibetean apply perturbations of $f$ to produce a sequence $f_\gamma$ with the property that $V_{\text{grad}}(f_\gamma)$ is finite and the minima $f_\gamma^*$ converge to $f^*$.

- Jibetean and Laurent give a method for computing $f^*$ via the gradient ideal in the case where the gradient variety is zero-dimensional.
In recent work of Demmel, Nie, and Sturmfels, a method is proposed for minimizing a polynomial via an SOS relaxation over its gradient ideal, without assuming that the gradient variety is finite.

**Theorem (Demmel, Nie, Sturmfels).** Suppose $I_{\text{grad}}(f)$ is a radical ideal and $f$ is psd, then $f$ is sos modulo $I_{\text{grad}}(f)$. In other words, there is an sos $s \in \mathbb{R}[X]$ and $h \in I_{\text{grad}}(f)$ such that $f = s + h$.

There exist counterexamples in the case where $I_{\text{grad}}(f)$ is not radical.

However, if $f > 0$ on $\mathbb{R}^n$, then the assumption that $I_{\text{grad}}(I)$ is radical can be dropped.
Now suppose we want to find the global minimum $f^*$ of $f$, then since $f^*$ occurs at a point in the real gradient variety, fix $N \in \mathbb{N}$ and define the SOS relaxation

Maximize $\lambda$ subject to

$$f(x) - \gamma - \sum_{i=1}^{n} a_i(x) \frac{\partial f}{\partial x_i} \text{ is sos and } \deg a_i(x) \leq N$$

This problem can be translated into an SDP, and computed numerically. Let $f_N^*$ be the optimal value. If $f(x)$ attains its global minimum $f^*$ then $\lim_{N \to \infty} f_N^* = f^*$. Further, if the gradient ideal is radical, then $f^* = f_N^*$ for some $N$. 
Note that this will always give a finite lower bound for $f^*$. Numerical experiments by Demmel, Nie and Sturmfels suggest that this method outperforms the "pure" SOS method.
Generalization to Semialgebraic Sets

Very recently, in joint work with Demmel and Nie, we have generalized the above ideas from the global case to the case of minimizing $f$ on a semialgebraic set.

We assume once again that we have $S = S(F)$, $M = M(F)$, and $P = PO(M)$ and we are trying to find the minimum $f^*$ of $f \in \mathbb{R}[X]$ on $S$.

**Theorem.** Suppose $I_{grad}(f)$ is radical and $f \geq 0$ on $S$. Then $f$ is in the preorder $P$ modulo $I_{grad}(f)$, i.e., there is $p \in P$ and $h \in I_{grad}(f)$ so that $f = p + h$.

Again, we can drop the assumption the $I_{grad}(f)$ is radical if we assume $f > 0$ on $S$. 
Note we do not assume that $S$ is compact.

On the other hand, we have counterexamples to show that the theorem is not true in general if we replace $P$ by the quadratic module $M$.

If $M$ is archimedean, then the theorem does hold for $M$. 
Application to Optimization

The Lasserre method for optimizing on a compact semialgebraic set \( S \) is based on representation theorems for positive polynomials.

On the other hand, the traditional local methods in optimization often use the Karush-Kuhn-Tucker (KKT) system for finding \( f^* \).

Why not combine these two methods?
Using a KKT System

Our optimization problem is

\[ f^* = \min f \text{ subject to } g_i(x) \geq 0, \quad i = 1, \ldots, r \]

The KKT system of this is

\[ H \triangleq \nabla f(x) - \sum_{j=1}^{r} \nu_j \nabla g_j(x) = 0, \]
\[ g_j(x) \geq 0, \quad \nu_j g_j(x) = 0, \quad j = 1, \ldots, r, \]

\( \nu = (\nu_1, \ldots, \nu_r) \) are called Lagrange multipliers. Note that we do not require \( \nu \geq 0 \).
Under certain regularity conditions, if $f^*$ exists and occurs at an optimizer $x^*$, then the KKT system holds at $x^*$. Then $f^*$ is the minimum of $f$ over the KKT system.

In $\mathbb{R}[x_1, \ldots, x_n, \nu_1, \ldots, \nu_r]$, define the KKT ideal $I_{KKT}$, KKT preorder $P_{KKT}$, and KKT quadratic module $M_{KKT}$ as follows:

$$I_{KKT} = \langle H_1, \ldots, H_n, \nu_1 g_1, \ldots, \nu_r g_r \rangle,$$

where $H = (H_1, \ldots, H_n)$.

$$P_{KKT} := PO(F) + I_{KKT}$$

$$M_{KKT} := M(F) + I_{KKT}$$
We also define the finite-dimensional analogues of these: \( I_{n,KKT} \) are elements of \( I_{KKT} \) where each term in the sum has degree \( \leq d \), then \( M_{d,KKT} = M_d + I_{d,KKT} \) and similarly \( P_{d,KKT} = P_d + I_{n,KKT} \).

Finally, we define \( f_d^* \) to be the maximum \( \lambda \) such that \( f - \lambda \in M_{d,KKT} \) and \( p_d^* \) to be the maximum \( \lambda \) such that \( f - \lambda \in P_{d,KKT} \).

**Theorem.** Assume \( f^* \) exists and the optimizer \( x^* \) satisfies the KKT system above. The \( p_d^* \) converges to \( f^* \). Further, if \( I_{KKT} \) is a radical ideal, then there is some \( N \) such that \( p_N^* = f^* \).

The theorem does not hold in general if we replace \( p_d^* \) by \( f_d^* \). If the semialgebraic set \( S \) is archimedean, then it does hold.
In general, the SOS relaxations are difficult to solve when there are many constraints, which introduces many Lagrange multipliers. In special cases, for example if the semialgebraic set is contained in the non-negative orthant $\mathbb{R}^n_+$, Lagrange multipliers can be removed and the KKT system simplifies considerably, making the computations more efficient.

Future research could focus on generalizations of this or more general methods for removing Lagrange multipliers.

Also, it’s not always true that the KKT condition holds for the minimizer $x^*$. We would like to remove the KKT conditions, or find ones which always hold.
Conclusions

- SOS relaxations are a powerful tool for solving optimization problems.
- Deep results in Real Algebraic Geometry, such as representation theorems, can lead to practical algorithms for SOS relaxations.
- Understanding the “gap” between positive polynomials and sums of squares (representations) should lead to better algorithms, information on bounds, etc.

I’m Vicki Powers and I approve of this message!