Sums of Squares, Gradient Ideals, and Optimization

Vicki Powers
Dept. of Mathematics and Computer Science
Emory University, Atlanta, USA

January 18, 2007; IMA workshop on Optimization and Control
1. What this talk is about
2. Sums of squares
3. Optimization with gradient ideals
4. Optimization on semialgebraic sets
5. Sums of squares modulo KKT ideals
6. Application to optimization
7. An example
8. Is it practical?
9. A final thought- gradient tentacles
Look at methods for minimizing a real polynomial on a basic closed semialgebraic set using representation theorems from real algebraic geometry.
Look at methods for minimizing a real polynomial on a basic closed semialgebraic set using representation theorems from real algebraic geometry.

The idea is to turn a problem of this type into a question about the existence of a representation involving sums of squares (sos) polynomials and the polynomials defining the semialgebraic set – an sos representation for short. This can then be implemented as a semidefinite program (SDP), and solved numerically.
Look at methods for minimizing a real polynomial on a basic closed semialgebraic set using representation theorems from real algebraic geometry.

The idea is to turn a problem of this type into a question about the existence of a representation involving sums of squares (sos) polynomials and the polynomials defining the semialgebraic set – an sos representation for short. This can then be implemented as a semidefinite program (SDP), and solved numerically.

In cases where the method does not work well, i.e., noncompact semialgebraic sets, combine with traditional optimization methods using gradient ideals/KKT systems to get better results. This is joint work with J. Demmel and J. Nie.
Sums of Squares
Let $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$ and let $\sum \mathbb{R}[X]^2$ denote the cone of sums of squares in $\mathbb{R}[X]$. 

$f \in \sum \mathbb{R}[X]^2$ is sos; $f$ is psd if $f(x) \geq 0$ for all $x \in \mathbb{R}^n$. 

$f$ sos implies $f$ psd, hence the global optimization problem of finding $f^* = \inf_{x \in \mathbb{R}^n} f(x)$ can be relaxed to finding the lower bound $f_{sos} = \max \{ \lambda \in \mathbb{R} | f - \lambda \text{ is sos} \}$ and implemented as a semidefinite programming problem (SDP).
Sums of Squares
Let \( \mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n] \) and let \( \sum \mathbb{R}[X]^2 \) denote the cone of sums of squares in \( \mathbb{R}[X] \).

\( f \in \sum \mathbb{R}[X]^2 \) is \textit{sos}; \( f \) is \textit{psd} if \( f(x) \geq 0 \) for all \( x \in \mathbb{R}^n \).
Sums of Squares

Let $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$ and let $\sum \mathbb{R}[X]^2$ denote the cone of sums of squares in $\mathbb{R}[X]$.

$f \in \sum \mathbb{R}[X]^2$ is sos; $f$ is psd if $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

$f$ sos implies $f$ psd,
**Sums of Squares**

Let $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$ and let $\sum \mathbb{R}[X]^2$ denote the cone of sums of squares in $\mathbb{R}[X]$.

$f \in \sum \mathbb{R}[X]^2$ is sos; $f$ is psd if $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

$f$ sos implies $f$ psd, hence the global optimization problem of finding

$$f^* = \inf_{x \in \mathbb{R}^n} f(x)$$

can be relaxed to finding the lower bound

$$f_{sos} = \max\{\lambda \in \mathbb{R} | f - \lambda \text{ is sos}\}$$

and implemented as a semidefinite programming problem (SDP).
However, this doesn’t always work well.
However, this doesn’t always work well.

Let

\[ M = x^4y^2 + x^2y^4 - 3x^2y^2 + 1 \]

(the Motzkin polynomial), then \( M \) has global minimum 0 but

\[ M_{\text{sos}} = -\infty \]
However, this doesn’t always work well.

Let

\[ M = x^4 y^2 + x^2 y^4 - 3x^2 y^2 + 1 \]

(the Motzkin polynomial), then \( M \) has global minimum 0 but

\[ M_{\text{sos}} = -\infty \]

which gives a not very useful lower bound.
However, this doesn’t always work well.

Let

\[ M = x^4 y^2 + x^2 y^4 - 3x^2 y^2 + 1 \]

(the Motzkin polynomial), then \( M \) has global minimum 0 but

\[ M_{\text{sos}} = -\infty \]

which gives a not very useful lower bound.

The problem, of course, is that not every psd polynomial is sos.
Global optimization with gradient ideals

In 2005, Demmel, Nie, and Sturmfels proved that if \( f > 0 \) on \( \mathbb{R}^n \), then \( f \) is sos modulo its gradient ideal, i.e., the ideal generated by the partial derivatives.
Global optimization with gradient ideals
In 2005, Demmel, Nie, and Sturmfels proved that if $f > 0$ on $\mathbb{R}^n$, then $f$ is sos modulo its gradient ideal, i.e., the ideal generated by the partial derivatives.

They proposed using sos representations of positive polynomials modulo their gradient ideals to solve the global optimization problem.
Global optimization with gradient ideals
In 2005, Demmel, Nie, and Sturmfels proved that if $f > 0$ on $\mathbb{R}^n$, then $f$ is sos modulo its gradient ideal, i.e., the ideal generated by the partial derivatives.

They proposed using sos representations of positive polynomials modulo their gradient ideals to solve the global optimization problem. They also showed that if the gradient ideal is radical, then $f \geq 0$ on $\mathbb{R}^n$ implies that $f$ is sos modulo its gradient ideal.
Global optimization with gradient ideals

In 2005, Demmel, Nie, and Sturmfels proved that if \( f > 0 \) on \( \mathbb{R}^n \), then \( f \) is sos modulo its gradient ideal, i.e., the ideal generated by the partial derivatives.

They proposed using sos representations of positive polynomials modulo their gradient ideals to solve the global optimization problem.

They also showed that if the gradient ideal is radical, then \( f \geq 0 \) on \( \mathbb{R}^n \) implies that \( f \) is sos modulo its gradient ideal.

In the case where \( f \) attains its minimum, they construct a sequence of SDP’s of increasing size which converge to \( f^* \).
Global optimization with gradient ideals

In 2005, Demmel, Nie, and Sturmfels proved that if \( f > 0 \) on \( \mathbb{R}^n \), then \( f \) is sos modulo its gradient ideal, i.e., the ideal generated by the partial derivatives.

They proposed using sos representations of positive polynomials modulo their gradient ideals to solve the global optimization problem.

They also showed that if the gradient ideal is radical, then \( f \geq 0 \) on \( \mathbb{R}^n \) implies that \( f \) is sos module its gradient ideal.

In the case where \( f \) attains its minimum, they construct a sequence of SDP’s of increasing size which converge to \( f^* \).

Numerical experiments show that this method improves prior sos methods, e.g., it gives a better bound for \( M \).
Optimization on semialgebraic sets
Consider the optimization problem

\[
    f^* := \inf_{x \in \mathbb{R}^n} f(x)
\]

s.t. \( g_1(x) \geq 0, \ldots, g_m(x) \geq 0 \).
Optimization on semialgebraic sets

Consider the optimization problem

\[ f^* := \inf_{x \in \mathbb{R}^n} f(x) \]

s.t. \( g_1(x) \geq 0, \ldots, g_m(x) \geq 0. \)

In other words, \( f^* \) is the infimum of \( f \) on the basic closed semialgebraic set

\[ S := S(g_1, \ldots, g_m) = \{ x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0 \}. \]
Optimization on semialgebraic sets
Consider the optimization problem

\[ f^* := \inf_{x \in \mathbb{R}^n} f(x) \]

s.t. \( g_1(x) \geq 0, \ldots, g_m(x) \geq 0. \)

In other words, \( f^* \) is the infimum of \( f \) on the basic closed semialgebraic set

\[ S := S(g_1, \ldots, g_m) = \{ x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0 \}. \]

One idea for finding \( f^* \) is to use theorems from real algebraic geometry to reduce to a question involving sums of squares, implement this as a semidefinite program (SDP) and then solve numerically.
Fix the basic closed semialgebraic set \( S = \{g_1 \geq 0, \ldots, g_m \geq 0\} \) and let \( M \) be the quadratic module generated by the \( g_i \)'s:

\[
M = \left\{ \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s \mid \sigma_i \in \sum \mathbb{R}[X]^2 \right\}
\]
Fix the basic closed semialgebraic set $S = \{g_1 \geq 0, \ldots, g_m \geq 0\}$ and let $M$ be the quadratic module generated by the $g_i$'s:

$$M = \left\{ \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s \mid \sigma_i \in \sum R[X]^2 \right\}$$

and $P$ the preorder generated by the $g_i$'s:

$$P = \left\{ \sum_{\epsilon \in \{0,1\}^s} \sigma_\epsilon g_1^{\epsilon_1} \cdots g_m^{\epsilon_m} \mid \sigma_\epsilon \in \sum R[X]^2 \right\}$$
A couple of beautiful theorems
A couple of beautiful theorems

**Theorem** [Schmüdgen] If $S$ is compact, then $f > 0$ on $S$ implies $f \in P$. 
A couple of beautiful theorems

**Theorem** [Schmüdgen] If $S$ is compact, then $f > 0$ on $S$ implies $f \in P$.

In general, even with the assumption that $S$ is compact, this does not hold if we replace $P$ by $M$, nor if we assume only that $f \geq 0$ on $S$. 
A couple of beautiful theorems

**Theorem** [Schmüdgen] If $S$ is compact, then $f > 0$ on $S$ implies $f \in P$.

In general, even with the assumption that $S$ is compact, this does not hold if we replace $P$ by $M$, nor if we assume only that $f \geq 0$ on $S$.

$M$ is **archimedean** if there exists $p \in M$ such that the set 
\[
\{ x \in \mathbb{R}^n : p(x) \geq 0 \}
\]
is compact, equivalently, if there exists $N \in \mathbb{N}$ such that $N - \sum_{i=1}^{m} x_i^2 \in M$. Note that if $M$ or $P$ is archimedean, then $S$ is compact.
A couple of beautiful theorems

**Theorem** [Schmüdgen] If $S$ is compact, then $f > 0$ on $S$ implies $f \in P$.

In general, even with the assumption that $S$ is compact, this does not hold if we replace $P$ by $M$, nor if we assume only that $f \geq 0$ on $S$.

$M$ is *archimedean* if there exists $p \in M$ such that the set \( \{ x \in \mathbb{R}^n : p(x) \geq 0 \} \) is compact, equivalently, if there exists $N \in \mathbb{N}$ such that $N - \sum_{i=1}^{m} x_i^2 \in M$. Note that if $M$ or $P$ is archimedean, then $S$ is compact.

**Theorem**[Putinar] Suppose $M$ is archimedean, then $f > 0$ on $S$ implies $f \in M$. 
A couple of beautiful theorems

**Theorem** [Schmüdgen] If $S$ is compact, then $f > 0$ on $S$ implies $f \in P$.

In general, even with the assumption that $S$ is compact, this does not hold if we replace $P$ by $M$, nor if we assume only that $f \geq 0$ on $S$.

$M$ is *archimedean* if there exists $p \in M$ such that the set \{ $x \in \mathbb{R}^n : p(x) \geq 0$ \} is compact, equivalently, if there exists $N \in \mathbb{N}$ such that $N - \sum_{i=1}^{m} x_i^2 \in M$. Note that if $M$ or $P$ is archimedean, then $S$ is compact.

**Theorem** [Putinar] Suppose $M$ is archimedean, then $f > 0$ on $S$ implies $f \in M$.

If $S$ is compact, using the representation theorems, we can get a sequence of relaxations of the original problem which converge to the minimum — the Lasserre method.
The Noncompact Case.
The Noncompact Case.

Putinar and Schmüdgen theorems do not hold in the case where $S$ is not compact, so the Lasserre method may not work.
The Noncompact Case.

Putinar and Schm{"u}dgen theorems do not hold in the case where $S$ is not compact, so the Lasserre method may not work.

A more traditional approach in numerical optimization methods is to use the Karush-Kuhn-Tucker (KKT) system in the constrained case. Our idea was to combine the two methods to give a procedure for approximating $f^*$ in the case where the semialgebraic set is not necessarily compact.
The Noncompact Case.

Putinar and Schm"udgen theorems do not hold in the case where $S$ is not compact, so the Lasserre method may not work.

A more traditional approach in numerical optimization methods is to use the Karush-Kuhn-Tucker (KKT) system in the constrained case. Our idea was to combine the two methods to give a procedure for approximating $f^*$ in the case where the semialgebraic set is not necessarily compact.

We need a generalization of the Demmel-Nie-Sturmfels theorem for polynomials positive/nonnegative on a (not necessarily compact) basic closed semialgebraic set.
The *KKT system* associated to the optimization problem is

\[
\nabla f - \sum_{j=1}^{m} \lambda_j \nabla g_j = 0
\]

(1)

\[
g_j \geq 0, \quad \lambda_j g_j = 0, \quad j = 1, \ldots, m
\]

(2)

where \( \nabla f \) denotes the gradient of \( f \), \( \lambda_j \) are the Lagrange multipliers.
The *KKT system* associated to the optimization problem is

\[
\nabla f - \sum_{j=1}^{m} \lambda_j \nabla g_j = 0 \tag{1}
\]

\[
g_j \geq 0, \quad \lambda_j g_j = 0, \quad j = 1, \ldots, m \tag{2}
\]

where \(\nabla f\) denotes the gradient of \(f\), \(\lambda_j\) are the Lagrange multipliers.

Under certain regularity conditions, for example if the gradients of the \(g_j\)'s are linearly independent, the local (including global) minimizers of \(f\) on \(S\) satisfy the KKT system above. A point is said to be a KKT point if the KKT system holds at that point.
The KKT system associated to the optimization problem is

\[ \nabla f - \sum_{j=1}^{m} \lambda_j \nabla g_j = 0 \]  

(1)

\[ g_j \geq 0, \lambda_j g_j = 0, j = 1, \ldots, m \]  

(2)

where \( \nabla f \) denotes the gradient of \( f \), \( \lambda_j \) are the Lagrange multipliers.

Under certain regularity conditions, for example if the gradients of the \( g_j \)'s are linearly independent, the local (including global) minimizers of \( f \) on \( S \) satisfy the KKT system above. A point is said to be a KKT point if the KKT system holds at that point.

We do not include the condition that the Lagrange multipliers \( \lambda_j \) are nonnegative, as is usual. It turns out that we don’t need it and it adds unnecessary complication.
Now let $\mathbb{R}[X, \lambda] := \mathbb{R}[x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_s]$. 
Now let $\mathbb{R}[X, \lambda] := \mathbb{R}[x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_s]$.

Let $F_i = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{s} \lambda_j \frac{\partial g_j}{\partial x_i}$ and define the **KKT ideal** $I_{KKT}$ and the real variety associated with KKT system as follows:
Now let $\mathbb{R}[X, \lambda] := \mathbb{R}[x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_s]$.

Let $F_i = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{s} \lambda_j \frac{\partial g_j}{\partial x_i}$ and define the **KKT ideal** $I_{KKT}$ and the real variety associated with KKT system as follows:

$$I_{KKT} = \langle F_1, \cdots, F_n, \lambda_1 g_1, \cdots, \lambda_s g_s \rangle \subseteq \mathbb{R}[X, \lambda]$$

$$V_{KKT}^{\mathbb{R}} = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^s : p(x, \lambda) = 0, \ \forall p \in I_{KKT}\}$$
Keeping in mind that we are now working in the larger polynomial ring, we use $P$, resp. $M$, to denote the preorder, resp. quadratic module, in $\mathbb{R}[X, \lambda]$ generated by $g_1, \ldots, g_m$. 
Keeping in mind that we are now working in the larger polynomial ring, we use $P$, resp. $M$, to denote the preorder, resp. quadratic module, in $\mathbb{R}[X, \lambda]$ generated by $g_1, \ldots, g_m$.

The associated KKT preorder $P_{KKT}$ and KKT quadratic module $M_{KKT}$ in $\mathbb{R}[X, \lambda]$ are defined as

$$P_{KKT} = P + I_{KKT}$$
$$M_{KKT} = M + I_{KKT}.$$  

Finally, let $\mathcal{H}$ be the set satisfying the optimization constraints:

$$\mathcal{H} = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^s : g_j(x) \geq 0, \ j = 1, \ldots, s\}.$$
**Proposition.** Assume $I_{KKT}$ is zero-dimensional and radical. If $f \in \mathbb{R}[X, \lambda]$ is nonnegative on $V_{KKT}^R \cap \mathcal{H}$, then $f \in M_{KKT}$. 
**Proposition.** Assume $I_{KKT}$ is zero-dimensional and radical. If $f \in \mathbb{R}[X, \lambda]$ is nonnegative on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in M_{KKT}$.

**Theorem.** Assume $I_{KKT}$ is radical. If $f \geq 0$ on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in P_{KKT}$.
Proposition. Assume $l_{KKT}$ is zero-dimensional and radical. If $f \in \mathbb{R}[X, \lambda]$ is nonnegative on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in M_{KKT}$.

Theorem. Assume $l_{KKT}$ is radical. If $f \geq 0$ on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in P_{KKT}$.

Theorem. If $f > 0$ on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in P_{KKT}$.
Proposition. Assume $I_{KKT}$ is zero-dimensional and radical. If $f \in \mathbb{R}[X, \lambda]$ is nonnegative on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in M_{KKT}$.

Theorem. Assume $I_{KKT}$ is radical. If $f \geq 0$ on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in P_{KKT}$.

Theorem. If $f > 0$ on $V_{KKT}^\mathbb{R} \cap \mathcal{H}$, then $f \in P_{KKT}$.

We have examples to show that we cannot replace $P_{KKT}$ by $M_{KKT}$ in the theorems.
Remarks on the proofs.
The proofs are fairly straightforward generalizations of the proofs in the global case (Demmel, Nie, Sturmfels results).
Remarks on the proofs.
The proofs are fairly straightforward generalizations of the proofs in the global case (Demmel, Nie, Sturmfels results).

They use classical algebraic geometry results such as the fact that irreducible varieties over \( \mathbb{C} \) are connected.
Remarks on the proofs.
The proofs are fairly straightforward generalizations of the proofs in the global case (Demmel, Nie, Sturmfels results).

They use classical algebraic geometry results such as the fact that irreducible varieties over $\mathbb{C}$ are connected.

Note that proofs of the Schmüdgen and Putinar theorems also need results from algebraic geometry, e.g., the classic Positivstellensatz.
Application to optimization

Using the theorems above, we combine the Lasserre method with the KKT system approach to give a procedure for approximating $f^*$ in the case where the semialgebraic set is not necessarily compact.
Application to optimization

Using the theorems above, we combine the Lasserre method with the KKT system approach to give a procedure for approximating $f^*$ in the case where the semialgebraic set is not necessarily compact.

Let $f_{KKT}^*$ be the minimum of $f$ over the KKT system defined above. **Assume the KKT system holds at at least one global optimizer.** Then $f^* = f_{KKT}^*$. 
Application to optimization

Using the theorems above, we combine the Lasserre method with the KKT system approach to give a procedure for approximating $f^*$ in the case where the semialgebraic set is not necessarily compact.

Let $f^*_{KKT}$ be the minimum of $f$ over the KKT system defined above. **Assume the KKT system holds at at least one global optimizer.** Then $f^* = f^*_{KKT}$.

Just as in the Lasserre method for the compact case, we define a series of relaxations of increasing size for the problem of finding $f^*_{KKT}$. 

For $N \in \mathbb{N}$, define the truncated KKT ideal

$$I_{N,KKT} = \left\{ \sum_{k=1}^{n} \phi_k F_k + \sum_{j=1}^{s} \psi_j \lambda_j g_j \bigg| \deg(\phi_k F_k), \deg(\psi_j \lambda_j g_j) \leq 2N \right\}.$$ 

and the truncated preorder

$$P_{N,KKT} = \left\{ \sum_{\theta \in \{0,1\}^s} \sigma_{\theta} g_1^{\theta_1} g_2^{\theta_2} \cdots g_s^{\theta_s} \bigg| \deg(\sigma_{\theta} g_1^{\theta_1} \cdots g_s^{\theta_s}) \leq 2N \right\} + I_{N,KKT}$$
For \( N \in \mathbb{N} \), define the truncated KKT ideal

\[
I_{N,KKT} = \left\{ \sum_{k=1}^{n} \phi_k F_k + \sum_{j=1}^{s} \psi_j \lambda_j g_j \bigg| \text{deg}(\phi_k F_k), \text{deg}(\psi_j \lambda_j g_j) \leq 2N \right\}.
\]

and the truncated preorder

\[
P_{N,KKT} = \left\{ \sum_{\theta \in \{0,1\}^s} \sigma_{\theta} g_{\theta_1} g_{\theta_2} \cdots g_{\theta_t} \bigg| \text{deg}(\sigma_{\theta} g_{\theta_1} \cdots g_{\theta_s}) \leq 2N \right\} + I_{N,KKT}.
\]

Finally, define

\[
f^*_N = \max_{\gamma \in \mathbb{R}} \gamma \quad \text{s.t.} \quad f(x) - \gamma \in P_{N,KKT}.
\]

Finding \( f^*_N \) can be implemented as an SDP.
**Theorem.** Assume $f(x)$ has a minimum $f^* := f(x^*)$ at one KKT point $x^*$. Then $\lim_{N \to \infty} f_N^* = f^*$. Furthermore, if $I_{KKT}$ is radical, then there exists some $N \in \mathbb{N}$ such that $f_N^* = f^*$, i.e., the sos relaxations converge in finitely many steps.
**Theorem.** Assume \( f(x) \) has a minimum \( f^* := f(x^*) \) at one KKT point \( x^* \). Then \( \lim_{N \to \infty} f^*_N = f^* \). Furthermore, if \( I_{KKT} \) is radical, then there exists some \( N \in \mathbb{N} \) such that \( f^*_N = f^* \), i.e., the sos relaxations converge in finitely many steps.

We have examples to show that the assumption that \( f \) has a minimum at a KKT point is nontrivial and cannot be removed.
**Theorem.** Assume $f(x)$ has a minimum $f^* := f(x^*)$ at one KKT point $x^*$. Then $\lim_{N \to \infty} f_N^* = f^*$. Furthermore, if $I_{KKT}$ is radical, then there exists some $N \in \mathbb{N}$ such that $f_N^* = f^*$, i.e., the sos relaxations converge in finitely many steps.

We have examples to show that the assumption that $f$ has a minimum at a KKT point is nontrivial and cannot be removed.

One might ask: Do you really gain anything?
An example
Consider the following nonconvex quadratic optimization problem:

\[ \min_{x \in \mathbb{R}^2} f := x^2 + y^2 \]

\[ s.t. \quad g_1 := y^2 - 1 \geq 0 \]
\[ g_2 := x^2 - Mxy - 1 \geq 0 \]
\[ g_3(x) := x^2 + Mxy - 1 \geq 0 \]

where \( M \) is a positive constant.
An example
Consider the following nonconvex quadratic optimization problem:

\[
\min_{x \in \mathbb{R}^2} \quad f := x^2 + y^2 \\
\text{s.t.} \quad g_1 := y^2 - 1 \geq 0 \\
\quad g_2 := x^2 - Mxy - 1 \geq 0 \\
\quad g_3(x) := x^2 + Mxy - 1 \geq 0
\]

where \( M \) is a positive constant.
It’s easy to see that

\[
f^* = \frac{1}{2} \left( M^2 + M\sqrt{M^2 + 4} \right) + 2
\]

and the global minimizers are

\[
\left( \pm \frac{1}{2} (M + \sqrt{M^2 + 4}), 1 \right), \left( \pm \frac{1}{2} (M + \sqrt{M^2 + 4}), -1 \right).
\]
Let $P := P(g_1, g_2, g_3)$, the preorder generated by $\{g_1, g_2, g_3\}$. Then $f - 2 = g_2 + g_3 \in P$. 
Let $P := P(g_1, g_2, g_3)$, the preorder generated by $\{g_1, g_2, g_3\}$. Then $f - 2 = g_2 + g_3 \in P$.

A straightforward argument shows that

$$\max\{\gamma \mid f - \gamma \in P\} = 2.$$ 

This means that using the Lasserre method we get the lower bound 2, regardless of the degree of the relaxation, and the ratio of the lower bound to the true global minimum tends to zero as $M$ goes to infinity.
Let $P := P(g_1, g_2, g_3)$, the preorder generated by $\{g_1, g_2, g_3\}$. Then $f - 2 = g_2 + g_3 \in P$.

A straightforward argument shows that

$$\max\{\gamma \mid f - \gamma \in P\} = 2.$$ 

This means that using the Lasserre method we get the lower bound 2, regardless of the degree of the relaxation, and the ratio of the lower bound to the true global minimum tends to zero as $M$ goes to infinity.

Of course, the feasible set is noncompact in this case, hence Putinar/Schmüdgen theorems don’t apply.
We might consider reducing to the compact case by adding a redundant condition such as

\[ R - x^2 - y^2 \geq 0, \]

where \( R \) is a sufficiently large positive number.
We might consider reducing to the compact case by adding a redundant condition such as

$$R - x^2 - y^2 \geq 0,$$

where $R$ is a sufficiently large positive number.

We implemented this approach using SOSTOOLS and found that the lower bounds obtained this way are still very bad. The bigger $M$ is, the worse the bound.
We use instead the optimization method based on the KKT system.
We use instead the optimization method based on the KKT system. The KKT system for the problem is

\[2(1 - \lambda_2 - \lambda_3)x + (\lambda_2 - \lambda_3)My = 0\]
\[2(1 - \lambda_1)x + (\lambda_2 - \lambda_3)Mx = 0\]
\[(y^2 - 1)\lambda_1 = 0\]
\[(x^2 - Mxy - 1)\lambda_2 = 0\]
\[(x^2 + Mxy - 1)\lambda_3 = 0.\]
We use instead the optimization method based on the KKT system.

The KKT system for the problem is

\[
2(1 - \lambda_2 - \lambda_3)x + (\lambda_2 - \lambda_3)My = 0 \\
2(1 - \lambda_1)x + (\lambda_2 - \lambda_3)Mx = 0 \\
(y^2 - 1)\lambda_1 = 0 \\
(x^2 - Mxy - 1)\lambda_2 = 0 \\
(x^2 + Mxy - 1)\lambda_3 = 0.
\]

Using Macaulay 2 we check that the KKT ideal \( I_{KKT} \) in this case is radical.
We can construct an explicit $q(x) \in M_{KKT} \subseteq P_{KKT}$ of degree 10 such that

$$f(x) - f^* \equiv q(x) \mod l_{5,KKT}.$$ 

This implies that $f_5^* = f^*$, hence we converge to the exact solution for $N = 5$. 

Vicki Powers
Dept. of Mathematics and Computer Science Emory University, Atlanta, USA
Sums of Squares, Gradient Ideals, and Optimization
We can construct an explicit $q(x) \in M_{KKT} \subseteq P_{KKT}$ of degree 10 such that

$$f(x) - f^* \equiv q(x) \mod l_{5,KKT}.$$ 

This implies that $f_5^* = f^*$, hence we converge to the exact solution for $N = 5$.

Thus the KKT system plays a crucial role in this example.
Is it really practical?
Is it really practical?

**Algebraicist's answer:** Who cares, the representation theorem is the interesting part.
Is it really practical?

**Algebraicist’s answer:** Who cares, the representation theorem is the interesting part.

**Optimized answer:** Without the (sometimes very restrictive) assumption that one of the global minimizers satisfies the KKT system, we might get a wrong answer.
Is it really practical?

**Algebraicist’s answer:** Who cares, the representation theorem is the interesting part.

**Optimized answer:** Without the (sometimes very restrictive) assumption that one of the global minimizers satisfies the KKT system, we might get a wrong answer.

Also, in general, the sos relaxations are very hard to solve when there are many constraints, since this introduces many Lagrange multipliers.
Is it really practical?

Algebraicist’s answer: Who cares, the representation theorem is the interesting part.

Optimized answer: Without the (sometimes very restrictive) assumption that one of the global minimizers satisfies the KKT system, we might get a wrong answer.

Also, in general, the sos relaxations are very hard to solve when there are many constraints, since this introduces many Lagrange multipliers.

It would be nice if we had $M_{KKT}$ in the theorems instead of $P_{KKT}$. 
Is it really practical?

**Algebraicist’s answer:** Who cares, the representation theorem is the interesting part.

**Optimized answer:** Without the (sometimes very restrictive) assumption that one of the global minimizers satisfies the KKT system, we might get a wrong answer.

Also, in general, the sos relaxations are very hard to solve when there are many constraints, since this introduces many Lagrange multipliers.

It would be nice if we had $M_{KKT}$ in the theorems instead of $P_{KKT}$.

The structure of the particular problem should be exploited to improve the efficiency of the method.
A final thought - gradient tentacles

For the case of finding the global infimum of a real polynomial, M. Schweighofer proposed replacing the gradient variety by gradient tentacles, an ascending sequence of larger semialgebraic sets. For polynomials bounded below, the global infimum equals the infimum on almost every gradient tentacle, hence this allows computation of the infimum even in the case where $f$ does not attain a minimum.

There are still problems with this method, e.g. how do you decide if a polynomial is bounded below? Also, polynomials that do not attain their minimum globally could be quite ill-conditioned with respect to finding the infimum.

However, it seems to be an improvement on the gradient ideal method.
A final thought - gradient tentacles

For the case of finding the global infimum of a real polynomial, M. Schweighofer proposed replacing the gradient variety by \textbf{gradient tentacles}, an ascending sequence of larger semialgebraic sets. For polynomials bounded below, the global infimum equals the infimum on almost every gradient tentacle, hence this allows computation of the infimum even in the case where $f$ does not attain a minimum.
A final thought - gradient tentacles

For the case of finding the global infimum of a real polynomial, M. Schweighofer proposed replacing the gradient variety by gradient tentacles, an ascending sequence of larger semialgebraic sets. For polynomials bounded below, the global infimum equals the infimum on almost every gradient tentacle, hence this allows computation of the infimum even in the case where \( f \) does not attain a minimum.

There are still problems with this method, e.g. how do you decide if a polynomial is bounded below? Also, polynomials that do not attain their minimum globally could be quite ill-conditioned with respect to finding the infimum.
A final thought - gradient tentacles

For the case of finding the global infimum of a real polynomial, M. Schweighofer proposed replacing the gradient variety by gradient tentacles, an ascending sequence of larger semialgebraic sets. For polynomials bounded below, the global infimum equals the infimum on almost every gradient tentacle, hence this allows computation of the infimum even in the case where $f$ does not attain a minimum.

There are still problems with this method, e.g. how do you decide if a polynomial is bounded below? Also, polynomials that do not attain their minimum globally could be quite ill-conditioned with respect to finding the infimum.

However, it seems to be an improvement on the gradient ideal method.
Perhaps there is a generalization of the gradient tentacles to the case of finding the infimum of a polynomial on a non-compact semialgebraic set. This could lead to a way to optimize without the assumption that one of the global minimizers satisfies the KKT system.
Perhaps there is a generalization of the gradient tentacles to the case of finding the infimum of a polynomial on a non-compact semialgebraic set. This could lead to a way to optimize without the assumption that one of the global minimizers satisfies the KKT system.

Thanks for listening!