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

Let $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$ and suppose $f, g_1, \ldots, g_m \in \mathbb{R}[X]$.

Let $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$ and suppose $f, g_1, \ldots, g_m \in \mathbb{R}[X]$. Consider the optimization problem

$$f^* := \min_{x \in \mathbb{R}^n} \quad f(x)$$
$$s.t. \quad g_1(x) \geq 0, \ldots, g_m(x) \geq 0.$$

Let $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$ and suppose $f, g_1, \ldots, g_m \in \mathbb{R}[X]$. Consider the optimization problem

$$\begin{aligned} f^* := \min_{x \in \mathbb{R}^n} \quad & f(x) \\ s.t. \quad & g_1(x) \geq 0, \ldots, g_m(x) \geq 0. \end{aligned}$$

In other words, f^* is the minimum of f on the basic closed semialgebraic set

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

Let $\mathbb{R}[X] := \mathbb{R}[x_1, \ldots, x_n]$ and suppose $f, g_1, \ldots, g_m \in \mathbb{R}[X]$. Consider the optimization problem

$$\begin{aligned} f^* := \min_{x \in \mathbb{R}^n} \quad & f(x) \\ s.t. \quad & g_1(x) \geq 0, \ldots, g_m(x) \geq 0. \end{aligned}$$

In other words, f^* is the minimum of f on the basic closed semialgebraic set

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

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.

Let $\sum \mathbb{R}[X]^2$ denote the cone of sums of squares in $\mathbb{R}[X]$. For each integer k, let

$$V_k := \{p \in \mathbb{R}[X] \mid \deg p \leq k\}$$

and define the convex cones

$$M_k := \left\{\sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s \mid \sigma_i \in \sum \mathbb{R}[X]^2, \sigma_i g_i \in V_k\right\}$$

$$P_k := \left\{ \sum_{\epsilon \in \{0,1\}^s} \sigma_\epsilon g_1^{\epsilon_1} \ldots g_s{}^{ep_s} \in V_k \mid \sigma_\epsilon \in \sum \mathbb{R}[X]^2 \right\}$$

Let $\sum \mathbb{R}[X]^2$ denote the cone of sums of squares in $\mathbb{R}[X]$. For each integer k, let

$$V_k := \{p \in \mathbb{R}[X] \mid \deg p \leq k\}$$

and define the convex cones

$$M_k := \left\{\sigma_0 + \sigma_1 g_1 + \cdots + \sigma_s g_s \mid \sigma_i \in \sum \mathbb{R}[X]^2, \sigma_i g_i \in V_k\right\}$$

$$P_k := \left\{ \sum_{\epsilon \in \{0,1\}^s} \sigma_\epsilon g_1^{\epsilon_1} \ldots g_s{}^{ep_s} \in V_k \mid \sigma_\epsilon \in \sum \mathbb{R}[X]^2 \right\}$$

Think of polynomials in M_k and P_k as "certifiably nonnegative" on S with "degree k sos certificates".

Consider the following nonconvex quadratic optimization problem:

$$
\begin{aligned}
\min_{x \in \mathbb{R}^2} \quad & f(x) := x_1^2 + x_2^2 \\
s.t. \quad & g_1(x) := x_2^2 - 1 \geq 0 \\
& g_2(x) := x_1^2 - Mx_1x_2 - 1 \geq 0 \\
& g_3(x) := x_1^2 + Mx_1x_2 - 1 \geq 0
\end{aligned}
$$

where M is a positive constant.

Consider the following nonconvex quadratic optimization problem:

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & f(x) := x_1^2 + x_2^2 \\ s.t. \quad & g_1(x) := x_2^2 - 1 \geq 0 \\ & g_2(x) := x_1^2 - Mx_1x_2 - 1 \geq 0 \\ & g_3(x) := x_1^2 + Mx_1x_2 - 1 \geq 0 \end{aligned}$$

where M is a positive constant.
It's easy to see that

$$f^* = \frac{1}{2}(M^2 + M\sqrt{M^2+4}) + 2$$

and the global minimizers are

$$(\pm\frac{1}{2}(M + \sqrt{M^2+4}), 1), \quad ,(\pm\frac{1}{2}(M + \sqrt{M^2+4}), -1).$$

Let $P := P(g_1, g_2, g_3)$, the preorder generated by $\{g_1, g_2, g_3\}$.

Let $P := P(g_1, g_2, g_3)$, the preorder generated by $\{g_1, g_2, g_3\}$.

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 $\{g_1, g_2, g_3\}$.

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 Schmüdgen's Theorem (and the Lasserre method) does not apply.

We might consider reducing to the compact case by adding a redundant condition such as

$$R - x_1^2 - x_2^2 \geq 0,$$

where R is a sufficiently large positive number.

We might consider reducing to the compact case by adding a redundant condition such as

$$R - x_1^2 - x_2^2 \geq 0,$$

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

$$
\begin{aligned}
2(1-\lambda_2-\lambda_3)x_1+(\lambda_2-\lambda_3)Mx_2 &= 0\\
2(1-\lambda_1)x_1+(\lambda_2-\lambda_3)Mx_1 &= 0\\
(x_2^2-1)\lambda_1 &= 0\\
(x_1^2-Mx_1x_2-1)\lambda_2 &= 0\\
(x_1^2+Mx_1x_2-1)\lambda_3 &= 0.
\end{aligned}
$$

We use instead the optimization method based on the KKT system.
The KKT system for the problem is

$$\begin{aligned}
2(1-\lambda_2-\lambda_3)x_1+(\lambda_2-\lambda_3)Mx_2 &= 0\\
2(1-\lambda_1)x_1+(\lambda_2-\lambda_3)Mx_1 &= 0\\
(x_2^2-1)\lambda_1 &= 0\\
(x_1^2-Mx_1x_2-1)\lambda_2 &= 0\\
(x_1^2+Mx_1x_2-1)\lambda_3 &= 0.
\end{aligned}$$

Using Macaulay 2 we check that the KKT ideal I_{KKT} in this case is radical.

Now let $q(x) =$

$$\rho_1 \left(x_1^2 - \frac{1}{4}(M + \sqrt{M^2+4})^2)^2 \right)^2 \lambda_1^2 \left((x_1^2 + Mx_1x_2 - 1)\lambda_2^2 + (x_1^2 - Mx_1x_2\right.$$

$$\rho_2 \lambda_1^2 \left(\left(\frac{2\lambda_1}{2+M^2} - 1 \right)^2 - \left(\frac{M^2}{M^2+4} \right)^2 \right)^2 (x_1^2 + Mx_1x_2 - 1) +$$

$$\rho_3 (4\lambda_1\lambda_2)^2 (x_2^2 - 1) + \rho_4 (x_1^2 + Mx_1x_2 - 1)^2 (x_1^2 + Mx_1x_2 - 1)^2 (x_2^2 -$$

$$(\lambda_2(1-2\lambda_2))^2 \left(\rho_5 \left(\sqrt{M^2+1} x_1^2 + 1 \right) (x_2^2 - 1) + \rho_6 \left(\sqrt{M^2+1} x_1^2 - \right.\right.$$

$$(\lambda_3(1-2\lambda_3))^2 \left(\rho_5 \left(\sqrt{M^2+1} x_1^2 + 1 \right) (x_2^2 - 1) + \rho_6 \left(\sqrt{M^2+1} x_1^2 - \right.\right.$$

where the ρ_i are suitably chosen constants.

Now let $q(x) =$

$$\rho_1 \left(x_1^2 - \frac{1}{4}(M + \sqrt{M^2+4})^2)^2 \right)^2 \lambda_1^2 \left((x_1^2 + Mx_1x_2 - 1)\lambda_2^2 + (x_1^2 - Mx_1x_2 \right.$$

$$\rho_2 \lambda_1^2 \left(\left(\frac{2\lambda_1}{2+M^2} - 1 \right)^2 - \left(\frac{M^2}{M^2+4} \right)^2 \right)^2 (x_1^2 + Mx_1x_2 - 1) +$$

$$\rho_3 (4\lambda_1\lambda_2)^2 (x_2^2 - 1) + \rho_4 (x_1^2 + Mx_1x_2 - 1)^2 (x_1^2 + Mx_1x_2 - 1)^2 (x_2^2 -$$

$$(\lambda_2(1 - 2\lambda_2))^2 \left(\rho_5 \left(\sqrt{M^2+1}x_1^2 + 1 \right) (x_2^2 - 1) + \rho_6 \left(\sqrt{M^2+1}x_1^2 - \right.\right.$$

$$(\lambda_3(1 - 2\lambda_3))^2 \left(\rho_5 \left(\sqrt{M^2+1}x_1^2 + 1 \right) (x_2^2 - 1) + \rho_6 \left(\sqrt{M^2+1}x_1^2 - \right.\right.$$

where the ρ_i are suitably chosen constants. Then $q(x)$ is visibly in M_{KKT} and hence in P_{KKT}.

It can be shown, e.g. using Macaulay 2, that

$$f(x) - f^* \equiv q(x) \qquad \text{mod} \quad I_{5,KKT}.$$

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.