Unconstrained Polynomial Optimization

Monique Laurent

CWI, Amsterdam

ADONET Doctoral school

Optimization over Polynomials and Semidefinite Programming

Klagenfurt

September 12-16, 2005

Unconstrained Polynomial Minimization

$$\boxed{p_{\min} = \inf_{x \in \mathbb{R}^n} p(x)}$$

where $\deg(p) = 2d$

As there is *no constraint*, the relaxation scheme just gives *one bound*:

$$\boxed{\rho_t^* = p_t^* = \rho_d^* = p_d^* \leq p_{\min} \quad \text{for all } t \geq d}$$

with equality iff $p(x) - p_{\min}$ is SOS

Idea

Get better bounds by transforming the *unconstrained* problem into a *constrained* problem

If p has a minimum:

$$\boxed{p_{\min} = p^*_{\text{grad}} := \inf_{x \in V^{\mathbb{R}}_{grad}} p(x)}$$

where $\boxed{V^{\mathbb{R}}_{\text{grad}} := \{x \in \mathbb{R}^n \mid \frac{\partial p}{\partial x_i} = 0 \ (i = 1, \ldots, n)\}}$

If, moreover, a bound R is known on the norm of a global minimizer:

$$\boxed{p_{\min} = p^*_{\text{ball}} := \inf_{R^2 - \sum_i x_i^2 \geq 0} p(x)}$$

When p attains its minimum

The 'ball approach':

• p^*_{ball} can be approximated via Lasserre's relaxation scheme (asymptotic convergence, as Putinar's assumption holds!)
• seems to work well only if the radius R of the ball is not too large ...

The 'gradient variety' approach:

Representation result: [Demmel, Nie, Sturmfels 2004]

$$\boxed{p > 0 \text{ on } V^{\mathbb{R}}_{\text{grad}} \Longrightarrow p = s_0 + \sum_{i=1}^n u_i \frac{\partial p}{\partial x_i}}$$

$$\boxed{p \geq 0 \text{ on } V^{\mathbb{R}}_{\text{grad}} \Longrightarrow p = s_0 + \sum_{i=1}^n u_i \frac{\partial p}{\partial x_i}} \quad \text{if } I_{\text{grad}} \text{ radical}$$

where s_0 SOS, $u_i \in \mathbb{R}[x_1, \ldots, x_n]$

Convergence Result [Demmel, Nie, Sturmfels 2004]

Asymptotic Convergence Result: There is *asymptotic convergence* of the SOS and moment bounds to p^*_{grad}

Finite Convergence Result: When $I_{\text{grad}} = \langle \frac{\partial p}{\partial x_1}, \ldots, \frac{\partial p}{\partial x_n} \rangle$ is a radical ideal, there is *finite convergence* to p^*_{grad}

Hence, when p attains its minimum, we have a converging hierarchy of SDP bounds for $p_{\min}$

Note: [Marshall 2003]
p attains its minimum if its *highest degree homogeneous component* $\displaystyle\sum_{|\alpha|=2d} p_\alpha x^\alpha$ is positive on $\mathbb{R}^n \setminus \{0\}$

Sketch of Proof of the DNS Representation Result

Proposition 1: p takes finitely many distinct values on $V_{\text{grad}} := V(I_{\text{grad}})$; say, $\underbrace{a_1, \ldots, a_r}_{\in\mathbb{R}}, \underbrace{b_1, \ldots, b_s}_{\in\mathbb{C}\setminus\mathbb{R}}$

$$W_i := \{x \in V_{\text{grad}} \mid p(x) = a_i\} \quad (i = 1, \ldots, r)$$
$$W_0 := \{x \in V_{\text{grad}} \mid p(x) \in \mathbb{C} \setminus \mathbb{R}\}$$

$\rightsquigarrow$ partition $V_{\text{grad}} = W_0 \cup W_1 \cup \ldots \cup W_r$

(1) $W_0 \cap \mathbb{R}^n = \emptyset$; $W_1 \cup \ldots \cup W_r = V_{\text{grad}} \cap \mathbb{R}^n = V^{\mathbb{R}}_{\text{grad}}$

(2) Each W_i is a variety; $W_i = \overline{W}_i$ $(i = 0, \ldots, r)$

(3) $p(x) = a_i \geq 0 \ \ \forall x \in W_i \quad (i = 1, \ldots, r)$

Lemma 1: There exist *real polynomials* $p_0, p_1, \ldots, p_r$ such that
$p_i(W_j) = \delta_{ij} \quad (i, j = 0, 1, \ldots, r)$

Proof: Set $I_i := I(W_i)$. Then, $W_i = V(I_i)$ and

$$V(I_0 + I_1 \cap \ldots \cap I_r) = V(I_0) \cap V(I_1 \cap \ldots \cap I_r)$$

$$= V_0 \cap (V_1 \cup \ldots \cup V_r) = \emptyset$$

By the (weak) **Nullstellensatz**, $1 \in I_0 + I_1 \cap \ldots \cap I_r$

That is, $1 = q_0 + p_0$, where $q_0 \in I_0$, $p_0 \in I_1 \cap \ldots \cap I_r$

Lemma 2: $p = s_0 \mod I(W_0)$
where s_0 is a sum of squares of polynomials

Proof: Apply the **Real Nullstellensatz** to $I := I(W_0)$, $f := p$

Real Nullstellensatz:
Let $I \subseteq \mathbb{R}[x_1, \ldots, x_n]$ be an ideal.

$V(I) \cap \mathbb{R}^n = \emptyset \Longrightarrow -1$ is SOS $\mod I$
$\Longrightarrow$ any polynomial f is SOS $\mod I$

Proof of the DNS Representation Result when I_{grad} is *radical*

Set $q := s_0 p_0^2 + \sum_{i=1}^{r} a_i p_i^2$

Then, q is SOS

Moreover, $p - q$ vanishes on $V_{\text{grad}} = W_0 \cup \ldots \cup W_r$

That is, $p - q \in I(V_{\text{grad}}) = I(V(I_{\text{grad}})) = I_{\text{grad}}$ [I_{grad} radical]

Therefore, p is SOS modulo I_{grad}

General Case: Assume $p > 0$ on $V^{\mathbb{R}}_{\mathbf{grad}}$, i.e., $a_1, \ldots, a_r > 0$

Write $I_{\text{grad}} = J_0 \cap \ldots \cap J_r$, where $V(J_i) = W_i$

Then, $J_i + J_j = \mathbb{R}[x_1, \ldots, x_n] \;\; \forall i \neq j = 0, \ldots, r$

This gives an *isomorphism*:

$$\boxed{\mathbb{R}[x_1, \ldots, x_n]/I_{\text{grad}} \longrightarrow \mathbb{R}[x_1, \ldots, x_n]/J_0 \times \ldots \times \mathbb{R}[x_1, \ldots, x_n]/J_r}$$

Now, $p = \text{SOS} \mod J_0$ [by Real Nullstellensatz]

$p = \text{SOS} \mod J_i \quad (i = 1, \ldots, r)$ [to show]

$\Longrightarrow p = \text{SOS} \mod I_{\text{grad}}$

Show: $p = \mathbf{SOS} \mod J_i \quad (i = 1, \ldots, r)$

$$p(x)/a_i - 1 \text{ vanishes on } W_i = V(J_i) \Longrightarrow p(x)/a_i - 1 \in \sqrt{J_i}$$

$$\Longrightarrow (p(x)/a_i - 1)^m \in J_i \text{ for some integer } m$$

$$\Longrightarrow p(x)/a_i = \text{SOS} \mod J_i$$

Write: $p(x)/a_i = 1 + \underbrace{(p(x)/a_i - 1)}_{=t}$ and use the fact:

$$\boxed{1 + t = \left(\sum_{k=0}^{m-1} \binom{1/2}{k} t^k \right)^2 \mod t^m}$$

Proof: Expand $(\sum_{k=0}^{m-1} \binom{1/2}{k} t^k)^2 = \sum_{j=0}^{2m-2} c_j t^j$ and note:

$$c_j = \sum_{h=0}^{j} \binom{1/2}{h} \binom{1/2}{j-h} = \begin{cases} 1 & \text{if } j = 0, 1 \\ 0 & \text{if } j = 2, \ldots, m-1 \end{cases}$$

Remains to show Proposition 1

p takes finitely many values on its gradient variety
$V_{\text{grad}} = V(I_{\text{grad}})$

Sketch of proof:

• Decompose V_{grad} into finitely many *path-connected components*: $V_{\text{grad}} = \cup_{i=1}^{s} V_i$

• Then, *p is constant on each V_i*

For $x, y \in V_i$, let $\varphi : [0,1] \longrightarrow V_i$ continuous differentiable function with $\varphi(0) = x$, $\varphi(1) = y$

Then, $\exists t^* \in (0,1)$ s.t.
$p(y) - p(x) = p(\varphi(1)) - p(\varphi(0)) = \nabla p(\varphi(t^*))^T \varphi'(t^*) = 0$
[as $\varphi(t^*) \in V_{\text{grad}}$]

End of proof of the DNS representation result

What if p is not known to have a minimum?

Strategy: Perturb the polynomial p

[Hanzon-Jibetean 2003] [Jibetean-Laurent 2004]

$$\boxed{p_\epsilon(x) := p(x) + \epsilon\left(\sum_{i=1}^{n} x_i^{2d+2}\right)}$$ for small $\epsilon > 0$

- p_ϵ *has a minimum* and $\boxed{\lim_{\epsilon \to 0} (p_\epsilon)_{\min} = p_{\min}}$
- the gradient variety of p_ϵ is *finite*
  $\rightsquigarrow$ one can compute $(p_\epsilon)_{\min}$ via the 'gradient variety' approach
  $\rightsquigarrow$ *finite convergence* of the moment bounds to $(p_\epsilon)_{\min}$ (in $2nd$ steps)
- the limits of global minimizers of p_ϵ (as $\epsilon \to 0$) are global minimizers of p (if any)

Relaxation Scheme

For $\epsilon > 0$, $t \geq d+1$

$$\boxed{p^*_{\epsilon,t} := \inf y^T p \;\text{ s.t. }\; M_t(y) \succeq 0,\; y_0 = 1,\; M_{t-d-1}\left(\frac{\partial p_\epsilon}{\partial x_i} * y\right) = 0 \; (i \leq n)}$$

Then, $p_{\min} \leq p^*_{\epsilon,t}$ [if rank $M_t(y) =$ rank $M_{t-1}(y)$]
$\leq p_{\min} + \epsilon \sum_i (x_i^*)^{2d+2}$ [if $p_{\min} = p(x^*)$]

Note: One can use the gradient equations:

$$\boxed{\frac{\partial p_\epsilon}{\partial x_i} = (2d+2)x_i^{2d+1} + \frac{\partial p}{\partial x_i} = 0 \;\; (i = 1, \ldots, n)}$$

to reduce # variables ($\rightsquigarrow$ only variables y_α with $\alpha_i \leq 2d \; \forall i$) and express $p_{\epsilon,t}$ via a SDP with a *single* LMI

Examples

Motzkin polynomial: $p = x^2y^2(x^2 + y^2 - 3) + 1$

Then, $p_{\min} = 0$, attained at $(\pm 1, \pm 1)$, and p is *not* SOS

ϵ	order t	(r_t, r_{t-1}, r_{t-2})	$(p_\epsilon)^*_t$	extracted solutions
	3	(10,6,3)	$-3.5070\ 10^4$	none
10^{-1}	4	(4,4,4)	0.0316	$(\pm 0.9453, \pm 0.9453)$
10^{-2}	4	(4,4,4)	$5.0177\ 10^{-4}$	$(\pm 0.9935, \pm 0.9935)$
10^{-3}	5	(4,4,4)	$5.2951\ 10^{-6}$	$(\pm 0.9993, \pm 0.9993)$
10^{-4}	5	(4,4,4)	$5.3170\ 10^{-8}$	$(\pm 0.9999, \pm 0.9999)$

Example:

$p = (xy - 1)^2 + y^2$ Then, $p_{\min} = 0$ is *not* attained

ϵ	order t	(r_t, r_{t-1}, r_{t-2})	$(p_\epsilon)_t^*$	extracted solutions
	2	(4,2,1)	$6.8791\ 10^{-5}$	none
10^{-2}	3	(2,2,2)	0.3385	$\pm(1.3981, 0.4729)$
10^{-3}	3	(2,2,2)	0.2082	$\pm(1.9499, 0.4060)$
10^{-4}	3	(2,2,2)	0.1232	$\pm(2.6674, 0.3287)$
10^{-5}	3	(2,2,2)	0.0713	$\pm(3.6085, 0.2574)$
10^{-6}	3	(2,2,2)	0.0408	$\pm(4.8511, 0.1977)$
10^{-7}	3	(3,2,2)	0.0231	$\pm(6.4986, 0.1503)$
10^{-8}	3	(3,2,2)	0.0131	$\pm(8.6882, 0.1136)$
10^{-9}	3	(8,4,2)	0.0074	none
10^{-10}	3	(7,4,2)	0.0041	none

Testing copositivity of a matrix M

M copositive $\Longrightarrow p_{\min} = 0$ for $p(x) := \sum_{i,j=1}^{n} x_i^2 x_j^2 M_{ij}$

Example 1:

$$M = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}$$

M is copositive, e.g., since $(\sum_i x_i^2)p$ is SOS [Parrilo 2000]

ϵ	order t	(r_t, r_{t-1}, r_{t-2})	$(p_\epsilon)^*_t$	extracted solutions
	2	(5,15)	$-1.4955\ 10^6$	none
10^{-2}	3	(18,6,3)	-1.5407	none
10^{-2}	4	(4,4,4)	$1.3854\ 10^{-5}$	$(\pm 0.7058, 0, 0, \pm 0.7058)$
10^{-4}	4	(9,7,5)	$1.5544\ 10^{-7}$	none

Example 2:

$$M = \begin{pmatrix} 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}$$

ϵ	order t	(r_t, r_{t-1}, r_{t-2})	$(p_\epsilon)_t^*$	extracted solutions
1	3	(8,7,4)	-1.3333	none
1	4	(8,8,7)	-1.3333	$\pm(0.8165, 0.8165, 0.8165, 0, 0)$
10^{-1}	3	(8,7,4)	-133.3333	none
10^{-1}	4	(8,8,7)	-133.3333	$\pm(2.5820, 2.5820, 2.5820, 0, 0)$
10^{-2}	3	(8,7,4)	$-1.3333\ 10^4$	none

$\rightsquigarrow$ certificate that M is *not* copositive

Recall Lasserre's Representation Result: Another Type of Perturbation

Theorem: [Lasserre 2004]

Assume $p \geq 0$ on $K = \{x \in \mathbb{R}^n \mid h_1(x) = 0, \ldots, h_m(x) = 0\}$

Then, $\forall \epsilon > 0 \; \exists t \in \mathbb{N}, \exists \lambda_1, \ldots, \lambda_m \geq 0$ for which

$$\underbrace{p + \epsilon \left(\sum_{k=0}^{t} \sum_{i=1}^{n} \frac{x_i^{2k}}{k!} \right)}_{p_{\epsilon,t}} + \sum_{j=1}^{m} \lambda_j h_j^2 \text{ is SOS}$$

$\rightsquigarrow$ New relaxation scheme for $\boxed{p_{\min} = \inf_{x \in K} p(x)}$

$$\boxed{p^*_{\epsilon,t} := \inf \; y^T p_{t,\epsilon} \;\text{ s.t. } M_t(y) \succeq 0, \; y_0 = 1, \; y^T(h_j^2) \leq 0 \; (j = 1, \ldots, m)}$$

For $\epsilon > 0$, $p_{\min} \leq p^*_{\epsilon,t}$ [for t large enough] $\leq p_{\min} + \epsilon \sum_{i=1}^n e^{x_i^*}$ [if $p_{\min} = p(x^*)$]

Another Approach to Matrix Copositivity Based on Pólya's Result [Parrilo 2000]

Recall:

M *copositive* if $p_M(x) := \sum_{i,j=1}^{n} M_{ij} x_i^2 x_j^2 \geq 0 \ \ \forall x \in \mathbb{R}^n$

M *strictly copositive* if $p_M(x) > 0 \ \ \forall x \neq 0$

Pólya's Representation Result: M strictly copositive

$$\Longrightarrow \exists t \in \mathbb{N} \text{ s.t. } \left(\sum_{i=1}^{n} x_i\right)^t x^T M x \text{ has nonnegative coefficients}$$

$$\Longrightarrow \exists t \in \mathbb{N} \text{ s.t. } \left(\sum_{i=1}^{n} x_i^2\right)^t p_M \text{ is SOS}$$

$\mathcal{C}_n$:= all $n \times n$ copositive matrices

$$\mathcal{K}_n^{(t)} := \{M \mid \left(\sum_{i=1}^n x_i^2\right)^t p_M \text{ SOS}\}$$

$$\boxed{\text{int}(\mathcal{C}_n) \subseteq \bigcup_{t=0}^{\infty} \mathcal{K}_n^{(t)} \subseteq \mathcal{C}_n}$$

$$\mathcal{K}_n^{(0)} = \text{PSD}_n + \mathbb{R}_+^{n \times n}$$

$\mathcal{K}_n^{(0)} = \mathcal{C}_n$ for $n \leq 4$ [Diananda]

$\bigcup_{t=0}^{\infty} \mathcal{K}_n^{(t)} \neq \mathcal{C}_n$ for $n \geq 5$ [Parrilo]

Application to the Stability Number [de Klerk-Pasechnik 2002]

Theorem: [Motzkin-Straus 1965]
$G = (V, E)$ graph with *stability number* $\alpha(G)$, *adjacency matrix* A_G. Then,

$$\boxed{1/\alpha(G) = \min \ x^T(I + A_G)x \ \text{ s.t. } \sum_{i=1}^{n} x_i = 1, \ x \geq 0}$$

That is, $\boxed{\alpha(G) = \min \ \lambda \text{ s.t. } \lambda(I + A_G) - J \text{ is copositive}}$

Define: $\boxed{\vartheta^{(t)}(G) := \min \ \lambda \text{ s.t. } \lambda(I + A_G) - J \in \mathcal{K}_n^{(t)}}$

Thus, $\vartheta^{(t)}(G) \downarrow \alpha(G)$ as $t \to \infty$

Some Results

Theorem: [dKP] $\lfloor \vartheta^{(t)}(G) \rfloor = \alpha(G)$ if $t \geq \alpha(G)^2$

Conjecture: [dKP] $\vartheta^{(t)}(G) = \alpha(G)$ if $t \geq \alpha(G) - 1$

Motivation: This rate of convergence holds for other SDP hierarchies, e.g., by Lovász-Schrijver, Lasserre

Note: $\alpha(G)^2$ steps are *necessary* for the analogous *linear* hierarchy

Theorem: [Gvozdenović-L 2004]
(1) The conjecture holds for graphs with $\alpha(G) \leq 8$
(2) Lasserre's hierarchy refines the new hierarchy $\vartheta^{(t)}(G)$

- Many technical details for (1)
- Use a dual formulation in terms of moment matrices for (2)

What is known about Error Analysis ? Not much ...

Optimization over simplex:
$$\boxed{p_{\min} := \min_{x \geq 0,\ \sum_i x_i = 1} p(x)}$$

That is, $\boxed{p_{\min} = \sup \rho \ \text{ s.t. } p(x) - \rho(\sum_{i=1}^n x_i)^d \geq 0 \ \ \forall x \geq 0}$

Note: W.l.o.g. p is a *homogeneous* polynomial of degree d

Relaxation scheme: For $t \in \mathbb{N}$

$$\boxed{p_t^* := \sup \rho \ \text{ s.t. } \left(p(x) - \rho(\sum_{i=1}^n x_i)^d\right)\left(\sum_{i=1}^n x_i\right)^t \ \text{ nonneg. coeff.}}$$

Then, $p_t^* \uparrow p_{\min}$ as $t \to \infty$ [by Pólya's theorem]

Pólya's Theorem: For q homogeneous polynomial

Assume $q(x) > 0$ on Δ

Then, $\exists t \in \mathbb{N}$ s.t. $q(x)(\sum_{i=1}^{n} x_i)^t$ has nonnegative coefficients

Bounds on t by [Powers-Reznick 2001]

Error estimates: [de Klerk-L-Parrilo 2004][Bomze-de Klerk 2002 for $d = 2$]

$$\boxed{p_{\min} - p_t^* \leq O(\tfrac{1}{t})(p_{\max} - p_{\min})}$$

where $p_{\max} := \max_{x \in \Delta} p(x)$

$\rightsquigarrow$ *PTAS* for minimizing a polynomial of *fixed* degree over the simplex