Nonnegative polynomials modulo their gradient ideal

About the article
"Minimizing Polynomials via Sum of Squares over the Gradient Ideal"
from Demmel, Nie and Sturmfels.

Richard Leroy

6th May 2005
Universität Konstanz

Introduction : motivation and notations

Introduction : motivation and notations

- Goal : Use semidefinite programming (SDP) to solve the problem of minimizing a real polynomial over $\mathbb{R}^n$.

Introduction : motivation and notations

- Goal : Use semidefinite programming (SDP) to solve the problem of minimizing a real polynomial over $\mathbb{R}^n$.

- Idea : If $u \in \mathbb{R}^n$ is a minimizer of a polynomial $f \in \mathbb{R}[X] := \mathbb{R}[X_1, \ldots, X_n]$, then $\nabla f(u) = 0$, i.e.

$$\forall i = 1, \ldots, n, \ \frac{\partial f}{\partial x_i}(u) = 0$$

Introduction : motivation and notations

- Goal : Use semidefinite programming (SDP) to solve the problem of minimizing a real polynomial over $\mathbb{R}^n$.

- Idea : If $u \in \mathbb{R}^n$ is a minimizer of a polynomial $f \in \mathbb{R}[X] := \mathbb{R}[X_1, \ldots, X_n]$, then $\nabla f(u) = 0$, i.e.

$$\forall i = 1, \ldots, n, \ \frac{\partial f}{\partial x_i}(u) = 0$$

- Tools :

Introduction : motivation and notations

- Goal : Use semidefinite programming (SDP) to solve the problem of minimizing a real polynomial over $\mathbb{R}^n$.

- Idea : If $u \in \mathbb{R}^n$ is a minimizer of a polynomial $f \in \mathbb{R}[X] := \mathbb{R}[X_1, \ldots, X_n]$, then $\nabla f(u) = 0$, i.e.

$$\forall i = 1, \ldots, n, \; \frac{\partial f}{\partial x_i}(u) = 0$$

- Tools :
  - (Real) algebraic geometry : Sos representation of a nonnegative polynomial modulo its gradient ideal

Introduction : motivation and notations

- Goal : Use semidefinite programming (SDP) to solve the problem of minimizing a real polynomial over $\mathbb{R}^n$.

- Idea : If $u \in \mathbb{R}^n$ is a minimizer of a polynomial $f \in \mathbb{R}[X] := \mathbb{R}[X_1, \dots, X_n]$, then $\nabla f(u) = 0$, i.e.

$$\forall i = 1, \dots, n, \ \frac{\partial f}{\partial x_i}(u) = 0$$

- Tools :
  - (Real) algebraic geometry : Sos representation of a nonnegative polynomial modulo its gradient ideal
  - SDP : duality theory (sos representation / moment approach)

Polynomials over their gradient varieties

Polynomials over their gradient varieties

- Notations :

Polynomials over their gradient varieties

- Notations :
  - Gradient varieties :

$$V_{grad}(f) := \{u \in \mathbb{C}^n : \nabla f(u) = 0\} \subset \mathbb{C}^n$$

$$V^{\mathbb{R}}_{grad}(f) := \{u \in \mathbb{R}^n : \nabla f(u) = 0\} \subset \mathbb{R}^n$$

  - Gradient ideal :

$$I_{grad}(f) := \langle \nabla f(X) \rangle = \left\langle \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right\rangle \subset \mathbb{R}[X]$$

Polynomials over their gradient varieties

- Notations :
  - Gradient varieties :

$$V_{grad}(f) := \{u \in \mathbb{C}^n : \nabla f(u) = 0\} \subset \mathbb{C}^n$$

$$V^{\mathbb{R}}_{grad}(f) := \{u \in \mathbb{R}^n : \nabla f(u) = 0\} \subset \mathbb{R}^n$$

  - Gradient ideal :

$$I_{grad}(f) := \langle \nabla f(X) \rangle = \left\langle \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right\rangle \subset \mathbb{R}[X]$$

- Theorem 1

$$\left.\begin{array}{l} f \geq 0 \text{ on } V^{\mathbb{R}}_{grad}(f) \\ I_{grad}(f) \text{ radical} \end{array}\right\} \Rightarrow f \text{ sos modulo } I_{grad}(f) :$$

$$\exists q_i, \phi_j \in \mathbb{R}[X],\ f = \sum_{i=1}^{s} q_i^2 + \sum_{j=1}^{n} \phi_j \frac{\partial f}{\partial x_j}$$

Proof

The proof is based on the following two lemmas :

Proof

The proof is based on the following two lemmas :

- lemma 1.1

$$V_1, \ldots, V_r \text{ pairwise disjoint varieties in } \mathbb{C}^n$$

$$\Downarrow$$

$$\exists p_1, \ldots, p_r \in \mathbb{R}[X], \forall i, j,\ \ p_i(V_j) = \delta_{ij}$$

Proof

The proof is based on the following two lemmas :

- lemma 1.1

$$V_1, \dots, V_r \text{ pairwise disjoint varieties in } \mathbb{C}^n$$

$$\Downarrow$$

$$\exists p_1, \dots, p_r \in \mathbb{R}[X], \forall i, j,\ p_i(V_j) = \delta_{ij}$$

- lemma 1.2

$$W \text{ irreducible subvariety of } V_{grad}(f) \text{ s.t. } W \cap \mathbb{R}^n \neq \emptyset$$

$$\Downarrow$$

$$f \equiv \text{ const on } W$$

In case $I_{grad}(f)$ is not radical

In case $I_{grad}(f)$ is not radical

- Example :

$$f(x,y,z) := x^8 + y^8 + z^8 + \underbrace{x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2}_{\text{Motzkin polynomial } M(x,y,z)}$$

In case $I_{grad}(f)$ is not radical

- Example :

$$f(x,y,z) := x^8 + y^8 + z^8 + \underbrace{x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2}_{\text{Motzkin polynomial } M(x,y,z)}$$

  - Fact 1 : $f \equiv \frac{1}{4}M \pmod{I_{grad}(f)}$

In case $I_{grad}(f)$ is not radical

- Example :

$$f(x,y,z) := x^8 + y^8 + z^8 + \underbrace{x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2}_{\text{Motzkin polynomial } M(x,y,z)}$$

  - Fact 1 : $f \equiv \frac{1}{4} M \pmod{I_{grad}(f)}$
  - Fact 2 : M is not a sos in $\mathbb{R}[x,y,z]/I_{grad}(f)$

In case $I_{grad}(f)$ is not radical

- Example :

$$f(x,y,z) := x^8 + y^8 + z^8 + \underbrace{x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2}_{\text{Motzkin polynomial } M(x,y,z)}$$

  - Fact 1 : $f \equiv \frac{1}{4} M \pmod{I_{grad}(f)}$
  - Fact 2 : M is not a sos in $\mathbb{R}[x,y,z]/_{I_{grad}(f)}$
  - Fact 3 : Ask Claus Scheiderer for more details

In case $I_{grad}(f)$ is not radical

- Example :

$$f(x,y,z) := x^8 + y^8 + z^8 + \underbrace{x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2}_{\text{Motzkin polynomial } M(x,y,z)}$$

  - Fact 1 : $f \equiv \frac{1}{4} M \pmod{I_{grad}(f)}$
  - Fact 2 : M is not a sos in $\mathbb{R}[x,y,z]/_{I_{grad}(f)}$
  - Fact 3 : Ask Claus Scheiderer for more details

- Theorem 2

$$f > 0 \text{ on } V^{\mathbb{R}}_{grad}(f) \Rightarrow f \text{ sos modulo } I_{grad}(f)$$

Unconstrained optimization

Unconstrained optimization

- Notations :

Unconstrained optimization

- Notations :
  - $\deg(f) = d$ even

Unconstrained optimization

- Notations :
  - $\deg(f) = d$ even
  - $f_i := \dfrac{\partial f}{\partial x_i}$

Unconstrained optimization

- Notations :
  - $\deg(f) = d$ even
  - $f_i := \dfrac{\partial f}{\partial x_i}$
  - $\forall k \geq d,\ f \in \mathbb{R}[X]_k : f = \sum f_\alpha x^\alpha \leftrightsquigarrow f \in \mathbb{R}^{\nu_{n,k}}$

    where $\nu_{n,k} = \begin{pmatrix} n+k \\ k \end{pmatrix}$

Unconstrained optimization

- Notations :
  - $\deg(f) = d$ even
  - $f_i := \dfrac{\partial f}{\partial x_i}$
  - $\forall k \geq d,\ f \in \mathbb{R}[X]_k : f = \sum f_\alpha x^\alpha \leftrightsquigarrow f \in \mathbb{R}^{\nu_{n,k}}$

    where $\nu_{n,k} = \begin{pmatrix} n+k \\ k \end{pmatrix}$
  - $\forall N,\ mon_N(x) = {}^t(1, x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_n^N) \in \mathbb{R}^{\nu_{n,N}}$

Unconstrained optimization

- Notations :
  - $\deg(f) = d$ even
  - $f_i := \dfrac{\partial f}{\partial x_i}$
  - $\forall k \geq d,\ f \in \mathbb{R}[X]_k : f = \sum f_\alpha x^\alpha \leftrightsquigarrow f \in \mathbb{R}^{\nu_{n,k}}$

    where $\nu_{n,k} = \begin{pmatrix} n+k \\ k \end{pmatrix}$
  - $\forall N,\ mon_N(x) = {}^t(1, x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_n^N) \in \mathbb{R}^{\nu_{n,N}}$

- Restrictive hypothesis (H) :

$$f \text{ attains its infimum } f^* \text{ over } \mathbb{R}^n$$

SDP relaxations

SDP relaxations

- Primal SDP : moment formulation

$$(P) : \begin{cases} f^*_{N,mom} := \quad \inf\limits_{y} \quad {}^t f y = \sum f_\alpha y_\alpha \\ \qquad\qquad\qquad s.t. \quad \begin{cases} \forall i,\ M_{N-\frac{d}{2}}(f_i * y) = 0 \\ M_N(y) \succeq 0 \\ y_0 = 1 \end{cases} \end{cases}$$

SDP relaxations

- Primal SDP : moment formulation

$$(P):\begin{cases} f^*_{N,mom} := \inf\limits_{y} \quad {}^t f y = \sum f_\alpha y_\alpha \\ \quad s.t. \begin{cases} \forall i,\ M_{N-\frac{d}{2}}(f_i * y) = 0 \\ M_N(y) \succeq 0 \\ y_0 = 1 \end{cases} \end{cases}$$

- Dual SDP : sos formulation

$$(D):\begin{cases} f^*_{N,grad} := \sup\limits_{\gamma \in R} \quad \gamma \\ \quad s.t. \begin{cases} f - \gamma = \sigma + \sum\limits_{j=1}^{n} \phi_j \frac{\partial f}{\partial x_j} \\ \sigma \in \sum \left(\mathbb{R}[X]_N\right)^2 \\ \phi_j \in \mathbb{R}[X]_{2N-d+1} \end{cases} \end{cases}$$

SDP relaxations

SDP relaxations

- Theorem 3

  Under the assumption (H), the following holds :

SDP relaxations

- Theorem 3

  Under the assumption (H), the following holds :

  - $\lim_N f^*_{N,grad} = \lim_N f^*_{N,mom} = f^*$

SDP relaxations

- Theorem 3

  Under the assumption (H), the following holds :

  - $\lim_N f^*_{N,grad} = \lim_N f^*_{N,mom} = f^*$
  - $I_{grad}(f)$ radical $\Rightarrow \exists N_0,\ f^*_{\mathbf{N_0},grad} = f^*_{\mathbf{N_0},mom} = f^*$

SDP relaxations

- Theorem 3

  Under the assumption (H), the following holds :

  - $\lim_N f^*_{N,grad} = \lim_N f^*_{N,mom} = f^*$
  - $I_{grad}(f)$ radical $\Rightarrow \exists N_0,\ f^*_{\mathbf{N_0},grad} = f^*_{\mathbf{N_0},mom} = f^*$

- Extracting solutions

  In practice, Lasserre and Henrion's technique :
  If, for some N, and some optimal primal solution y^*, we have

  $$\text{rank } M_N(y^*) = \text{rank } M_{N-d/2}(y^*)$$

  then we have reached the global minimum f^*, and one can extract global minimizers (implemented in Gloptipoly).

What's next ?