Analysis of sum-of-squares hierarchies for polynomial optimization

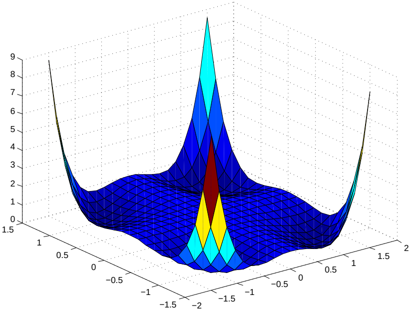
Monique Laurent

CWI

Tilburg University

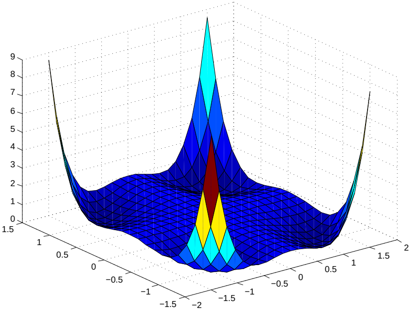
Joint works with Etienne de Klerk and Lucas Slot

One World Optimization Seminar



Minimize a **polynomial** f over a **compact** (semialgebraic) set K

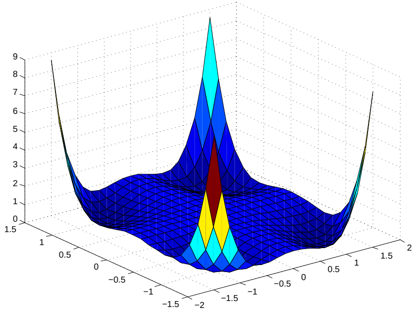
$$f_{\min} = \min_{x \in K} f(x)$$



Minimize a **polynomial** f over a **compact** (semialgebraic) set K

$$f_{\min} = \min_{x \in K} f(x)$$

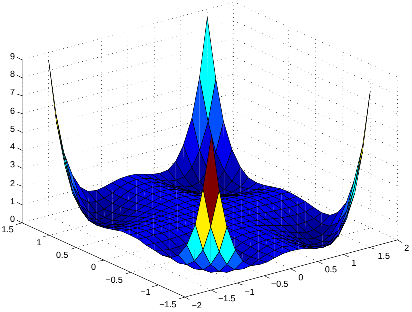
NP-hard problem: it captures hard combinatorial problems
(like computing $\alpha(G)$: the maximum size of a stable set in a graph G)
when K is a hypercube or a simplex and $\deg(f) = 2$,
or K is a sphere and $\deg(f) = 3$



Minimize a **polynomial** f over a **compact** (semialgebraic) set K

$$f_{\min} = \min_{x \in K} f(x)$$

$$\alpha(G) = \max_{x \in [0,1]^n} \sum_{i=1}^{n} x_i - \sum_{ij \in E} x_i x_j \qquad \frac{1}{\alpha(G)} = \min_{x \in \Delta_n} \sum_{i=1}^{n} x_i^2 + 2 \sum_{ij \in E} x_i x_j$$



Minimize a **polynomial** f over a **compact** (semialgebraic) set K

$$f_{\min} = \min_{x \in K} f(x)$$

$$\alpha(G) = \max_{x \in [0,1]^n} \sum_{i=1}^{n} x_i - \sum_{ij \in E} x_i x_j \qquad \frac{1}{\alpha(G)} = \min_{x \in \Delta_n} \sum_{i=1}^{n} x_i^2 + 2 \sum_{ij \in E} x_i x_j$$

$$\frac{2\sqrt{2}}{3\sqrt{3}} \sqrt{1 - \frac{1}{\alpha(G)}} = \max_{(x,y) \in \mathbb{S}^{n+|\overline{E}|-1}} 2 \sum_{ij \in \overline{E}} x_i x_j y_{ij}$$

[Motzkin-Straus'65, Nesterov'03]

Two hierarchies of **lower/upper bounds** for polynomial optimization:

$$f_{\min} = \min_{x \in K} f(x)$$

- Lasserre/Parrilo *sums-of-squares based* **lower bounds:**

$$f_{(r)} \leq f_{\min}$$

- Lasserre *measure-based* **upper bounds:**

$$f_{\min} \leq f^{(r)}$$

Common feature:

- For fixed r the bounds can be computed via semidefinite programming (SDP) with matrix size $O(n^r)$

  (since *sum-of-squares polynomials can be modelled with SDP*)
- the bounds converge asymptotically to $f_{\min}$

This lecture: Main focus on the **error analysis** for the **upper bounds**

Lasserre/Parrilo sums-of-squares based lower bounds

‘Sums-of-squares’ (SoS) lower bounds

$$\textbf{(P)} \quad \boxed{f_{\min} = \min_{x \in K} f(x) = \sup_{\lambda \in \mathbb{R}} \lambda \ \text{ s.t. } \ f(x) - \lambda \geq 0 \ \text{ on } K}$$

'Sums-of-squares' (SoS) lower bounds

$$\textbf{(P)} \quad \boxed{f_{\min} = \min_{x \in K} f(x) = \sup_{\lambda \in \mathbb{R}} \ \lambda \ \text{ s.t. } f(x) - \lambda \geq 0 \ \text{ on } K}$$

When $\boxed{K = \{x \in \mathbb{R}^n : \ g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}}$ with $g_j \in \mathbb{R}[x]$

one can replace the **hard** condition: "$f(x) - \lambda \geq 0$ on K"

by the **easier** condition:

"$f(x) - \lambda$ is a 'weighted sum' of sum-of-squares polynomials"

‘Sums-of-squares’ (SoS) lower bounds

$$\textbf{(P)} \quad \boxed{f_{\min} = \min_{x \in K} f(x) = \sup_{\lambda \in \mathbb{R}} \lambda \ \text{ s.t. } \ f(x) - \lambda \geq 0 \ \text{ on } K}$$

When $\boxed{K = \{x \in \mathbb{R}^n : \ g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}}$ with $g_j \in \mathbb{R}[x]$

one can replace the **hard** condition: “$f(x) - \lambda \geq 0$ on K”

by the **easier** condition:

“$f(x) - \lambda$ is a ‘weighted sum’ of sum-of-squares polynomials”

'Sums-of-squares' (SoS) lower bounds

$$\textbf{(P)} \quad \boxed{f_{\min} = \min_{x\in K} f(x) = \sup_{\lambda\in\mathbb{R}} \lambda \ \text{ s.t. } \ f(x) - \lambda \geq 0 \ \text{ on } K}$$

When $\boxed{K = \{x \in \mathbb{R}^n : \ g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}}$ with $g_j \in \mathbb{R}[x]$

one can replace the **hard** condition: "$f(x) - \lambda \geq 0$ on K"

by the **easier** condition:

"$f(x) - \lambda$ is a 'weighted sum' of sum-of-squares polynomials"

$\rightsquigarrow$ Get the SoS **bounds**:

$$\boxed{f_{(r)} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda = \underbrace{s_0}_{\deg \leq 2r} + \underbrace{s_1 g_1}_{\deg \leq 2r} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2r}, \ s_j \ \text{ SoS}}$$

‘Sums-of-squares’ (SoS) lower bounds

$$\textbf{(P)} \quad \boxed{f_{\min} = \min_{x\in K} f(x) = \sup_{\lambda\in\mathbb{R}} \lambda \ \text{ s.t. } \ f(x)-\lambda \geq 0 \ \text{ on } K}$$

When $\boxed{K = \{x \in \mathbb{R}^n : \ g_1(x) \geq 0, \dots, g_m(x) \geq 0\}}$ with $g_j \in \mathbb{R}[x]$

one can replace the **hard** condition: “$f(x) - \lambda \geq 0$ on K”

by the **easier** condition:

“$f(x) - \lambda$ is a ‘weighted sum’ of sum-of-squares polynomials”

$\rightsquigarrow$ Get the SoS **bounds**:

$$\boxed{f_{(r)} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda = \underbrace{s_0}_{\deg \leq 2r} + \underbrace{s_1 g_1}_{\deg \leq 2r} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2r}, \ s_j \ \text{ SoS}}$$

- $f_{(r)} \leq f_{(r+1)} \leq f_{\min}, \quad f_{(r)} \nearrow f_{\min}$ as $r \to \infty$
- Can compute $f_{(r)}$ with **semidefinite programming**

[Lasserre 2001]

Error analysis in terms of the relaxation order r

- [Nie-Schweighofer 2007] Let K semi-algebraic compact (+ technical condition). There exists a constant $c = c_K$ such that for any degree d polynomial f:

$$f_{\min} - f_{(r)} \leq 6d^3 n^{2d} L_f \ \frac{1}{\sqrt[c]{\log \frac{r}{c}}} \qquad \text{for all } r \geq c \cdot e^{(2d^2 n^d)^c}$$

Error analysis in terms of the relaxation order r

- [Nie-Schweighofer 2007] Let K semi-algebraic compact (+ technical condition). There exists a constant $c = c_K$ such that for any degree d polynomial f:

$$f_{\min} - f_{(r)} \leq 6d^3 n^{2d} L_f \, \frac{1}{\sqrt[c]{\log \frac{r}{c}}} \qquad \text{for all } r \geq c \cdot e^{(2d^2 n^d)^c}$$

- [Fang-Fawzi 2020] **Better error analysis** in $O(1/r^2)$ for the **unit sphere** $K = \mathbb{S}^{n-1}$, for f homogeneous with degree $2d$:

$$f_{\min} - f_{(r)} \leq (f_{\max} - f_{\min}) \frac{C_d^2 n^2}{r^2} \quad \text{for } r \geq C_d \cdot n$$

This improves the earlier $O(1/r)$ result of [Doherty-Wehner 2012]

Lasserre measure-based upper bounds

Basic observation: identify **points** $x \in K$ with **Dirac measures on K**

$$f_{\min} = \min_{x \in K} f(x) = \min_{\nu \text{ probability measure on } K} \int_K f(x) d\nu(x)$$

Basic observation: identify **points** $x \in K$ with **Dirac measures on** K

$$f_{\min} = \min_{x \in K} f(x) = \min_{\nu \text{ probability measure on } K} \int_K f(x) d\nu(x)$$

Theorem (Lasserre 2011)

For K compact, one may restrict to $d\nu(x) = h(x) d\mu(x)$, where μ is a **fixed** *measure with support K and h is a sum-of-squares density:*

$$f_{min} = \inf_h \int_K f(x) h(x) \, d\mu \quad s.t. \quad h \text{ SoS}, \ \int_K h(x) \, d\mu = 1$$

Basic observation: identify **points** $x \in K$ with **Dirac measures on** K

$$f_{\min} = \min_{x \in K} f(x) = \min_{\nu \text{ probability measure on } K} \int_K f(x) d\nu(x)$$

Theorem (Lasserre 2011)

*For K compact, one may restrict to $d\nu(x) = h(x) d\mu(x)$, where μ is a **fixed** measure with support K and h is a sum-of-squares density:*

$$f_{min} = \inf_h \int_K f(x) h(x)\ d\mu \quad s.t. \quad h \ SoS, \int_K h(x)\ d\mu = 1$$

Bound degree: $\deg(h) \leq 2r \rightsquigarrow$ **upper bounds** $f^{(r)}$ converging to $f_{\min}$:

$$f^{(r)} = \inf_h \int_K f(x) h(x)\ d\mu \ \text{ s.t. } \ h \text{ SoS}, \int_K h(x)\ d\mu = 1, \ \deg(h) \leq 2r$$

Basic observation: identify **points** $x \in K$ with **Dirac measures on** K

$$f_{\min} = \min_{x \in K} f(x) = \min_{\nu \text{ probability measure on } K} \int_K f(x) d\nu(x)$$

Theorem (Lasserre 2011)

For K compact, one may restrict to $d\nu(x) = h(x) d\mu(x)$, where μ is a **fixed** *measure with support K and h is a sum-of-squares density:*

$$f_{min} = \inf_h \int_K f(x) h(x)\, d\mu \quad s.t. \quad h \text{ SoS}, \ \int_K h(x)\, d\mu = 1$$

Bound degree: $\deg(h) \leq 2r \rightsquigarrow$ **upper bounds** $f^{(r)}$ converging to $f_{\min}$:

$$f^{(r)} = \inf_h \int_K f(x) h(x)\, d\mu \ \text{ s.t. } \ h \text{ SoS}, \ \int_K h(x)\, d\mu = 1, \ \deg(h) \leq 2r$$

- $f_{\min} \leq f^{(r+1)} \leq f^{(r)}$, $\quad f^{(r)} \searrow f_{\min}$, $\quad f^{(r)}$ can be computed via SDP

Basic observation: identify **points** $x \in K$ with **Dirac measures on** K

$$f_{\min} = \min_{x \in K} f(x) = \min_{\nu \text{ probability measure on } K} \int_K f(x) d\nu(x)$$

Theorem (Lasserre 2011)

*For K compact, one may restrict to $d\nu(x) = h(x)d\mu(x)$, where μ is a **fixed** measure with support K and h is a sum-of-squares density:*

$$f_{min} = \inf_h \int_K f(x)h(x)\ d\mu \quad s.t. \quad h\ SoS,\ \int_K h(x)\ d\mu = 1$$

Bound degree: $\deg(h) \leq 2r \rightsquigarrow$ **upper bounds** $f^{(r)}$ converging to $f_{\min}$:

$$f^{(r)} = \inf_h \int_K f(x)h(x)\ d\mu \ \text{ s.t. } \ h \text{ SoS},\ \int_K h(x)\ d\mu = 1,\ \deg(h) \leq 2r$$

- $f_{\min} \leq f^{(r+1)} \leq f^{(r)}$, $\quad f^{(r)} \searrow f_{\min}$, $\quad f^{(r)}$ can be computed via SDP
- **but** one needs to know the **moments** of μ: $\quad m_\alpha = \int_K x^\alpha d\mu(x)$ to compute $\int_K f(x) d\mu = \int_K (\sum_\alpha f_\alpha x^\alpha) d\mu = \sum_\alpha f_\alpha m_\alpha$

Basic observation: identify **points** $x \in K$ with **Dirac measures on** K

$$f_{\min} = \min_{x \in K} f(x) = \min_{\nu \text{ probability measure on } K} \int_K f(x) d\nu(x)$$

Theorem (Lasserre 2011)

For K compact, one may restrict to $d\nu(x) = h(x) d\mu(x)$, where μ is a **fixed** *measure with support K and h is a sum-of-squares density:*

$$f_{min} = \inf_h \int_K f(x) h(x)\ d\mu \quad s.t. \quad h \ SoS, \ \int_K h(x)\ d\mu = 1$$

Bound degree: $\deg(h) \le 2r \rightsquigarrow$ **upper bounds** $f^{(r)}$ converging to $f_{\min}$:

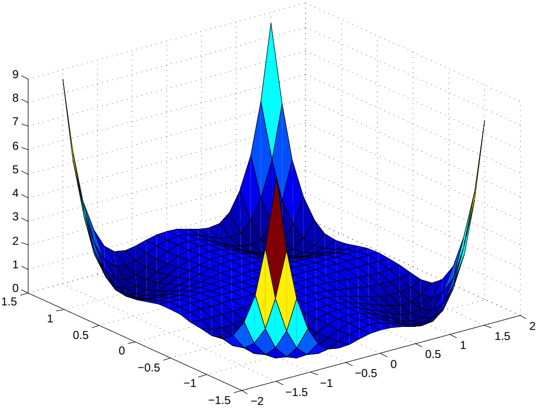
$$f^{(r)} = \inf_h \int_K f(x) h(x)\ d\mu \ \text{ s.t. } \ h \text{ SoS}, \ \int_K h(x)\ d\mu = 1, \ \deg(h) \le 2r$$

- $f_{\min} \le f^{(r+1)} \le f^{(r)}$, $\quad f^{(r)} \searrow f_{\min}$, $\quad f^{(r)}$ can be computed via SDP
- **but** one needs to know the **moments** of μ: $\quad m_\alpha = \int_K x^\alpha d\mu(x)$
- m_α known if μ Lebesgue on cube, ball, simplex; Haar on sphere,...

Example: Motzkin polynomial on $K = [-2, 2]^2$

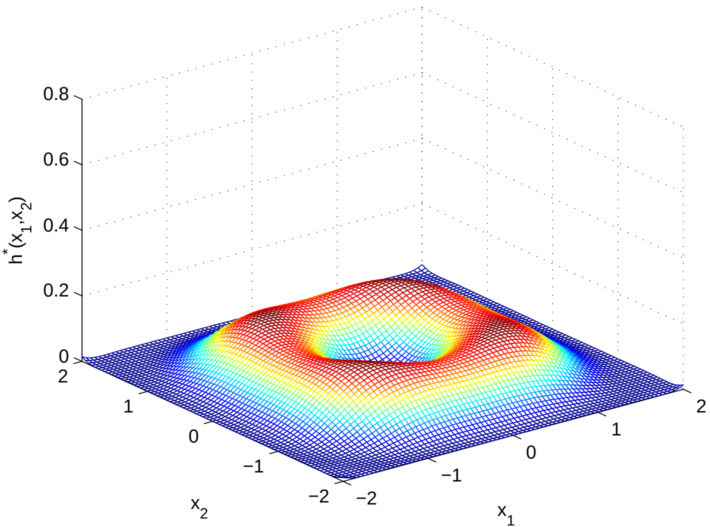
$$f(x_1, x_2) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$$

Four global minimizers: $(-1, -1)$, $(-1, 1)$, $(1, -1)$, $(1, 1)$



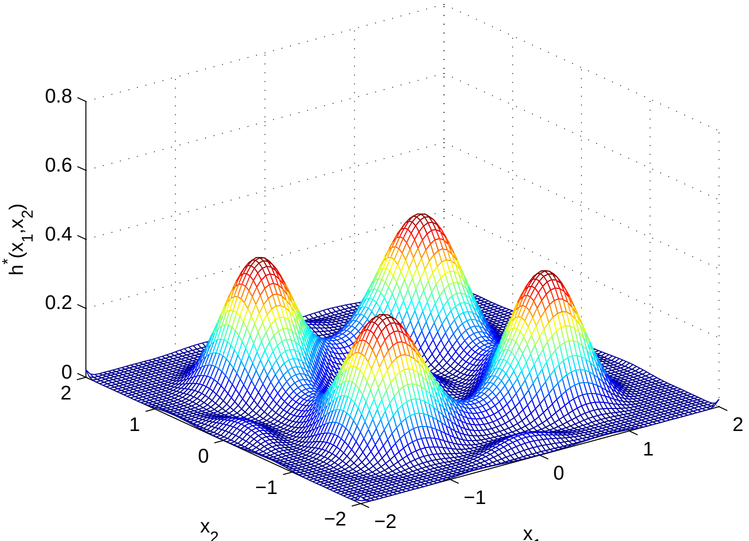
Example: Motzkin polynomial on $[-2, 2]^2$ (ctd.)

Optimal SoS density h of degree 12



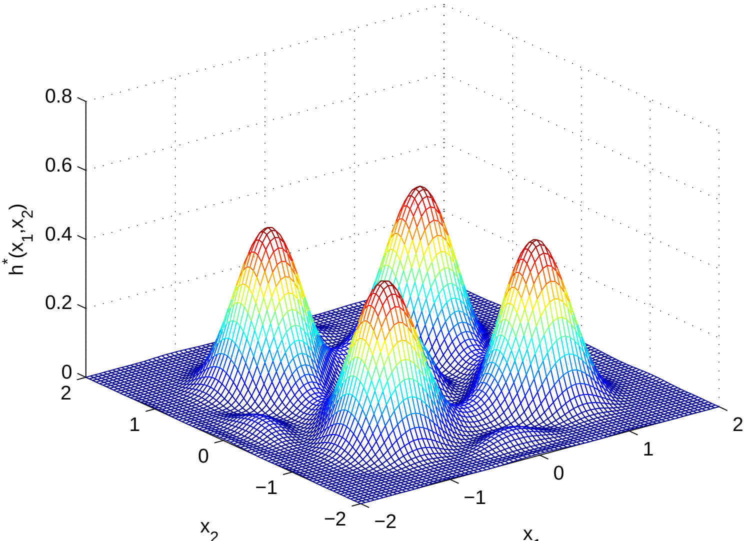
Example: Motzkin polynomial on $[-2, 2]^2$ (ctd.)

Optimal SoS density h of degree 16



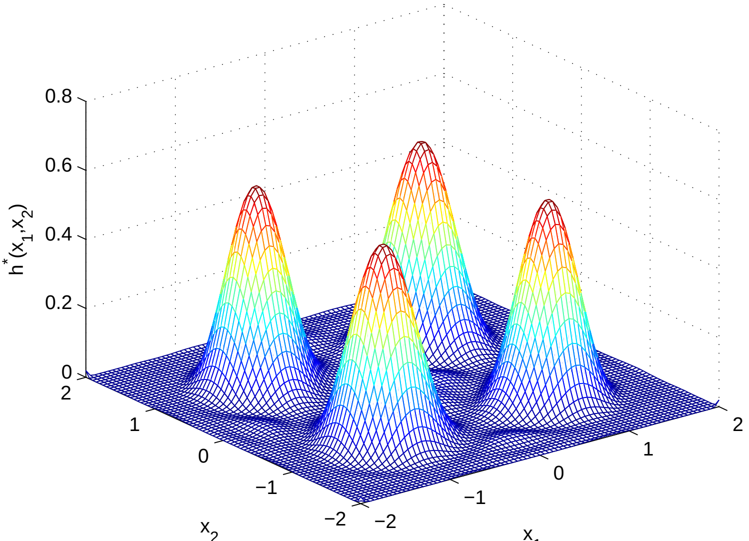
Example: Motzkin polynomial on $[-2, 2]^2$ (ctd.)

Optimal SoS density h of degree 20



Example: Motzkin polynomial on $[-2, 2]^2$ (ctd.)

Optimal SoS density h of degree 24



Goal: Analyze rate of convergence of error range:

$$E^{(r)}(f) = E^{(r)}_{\mu,K}(f) := f^{(r)} - f_{\min}$$

Goal: Analyze rate of convergence of error range:

$$E^{(r)}(f) = E^{(r)}_{\mu,K}(f) := f^{(r)} - f_{\min}$$

compact K	$E^{(r)}(f)$	μ	
Hypercube			
f linear	$\Theta(1/r^2)$	$(1-x^2)^{\lambda}$, $\lambda > -1$	de Klerk-L 2020
any f	$O(1/r^2)$	Chebyshev: $\lambda = -1/2$	" "
any f	$O(1/r^2)$	$\lambda \geq -1/2$	Slot-L 2020
Sphere			
f homogeneous	$O(1/r)$	Haar	Doherty-Wehner'12
any f	$O(1/r^2)$	Haar	de Klerk-L 2020
Ball			
any f	$O(1/r^2)$	$(1-\|x\|^2)^{\lambda}$, $\lambda \geq 0$	Slot-L 2020
Simplex, 'round' convex body	$O(1/r^2)$	Lebesgue	Slot-L 2020
Convex body, fat semialgebraic	$O((\log r)^2/r^2)$	Lebesgue	Slot-L 2020

Key proof strategies

(1) Reformulate $f^{(r)}$ as an **eigenvalue problem** and
relate $f^{(r)}$ to **extremal roots of orthogonal polynomials**

$\rightsquigarrow$ $O(1/r^2)$ rate for the Chebyshev measure on $[-1,1]$,
and other measures (with Jacobi weight) for **linear** polynomials

Key proof strategies

(1) Reformulate $f^{(r)}$ as an **eigenvalue problem** and
relate $f^{(r)}$ to **extremal roots of orthogonal polynomials**

$\rightsquigarrow$ $O(1/r^2)$ rate for the Chebyshev measure on $[-1, 1]$,
and other measures (with Jacobi weight) for **linear** polynomials

() Use tricks (**Taylor approx.**, **integration**, **'local similarity'**) to transport the $O(1/r^2)$ rate for $[-1, 1]$ to more sets (and measures): hypercube, simplex, ball, sphere, 'round' convex bodies

Key proof strategies

(1) Reformulate $f^{(r)}$ as an **eigenvalue problem** and
relate $f^{(r)}$ to **extremal roots of orthogonal polynomials**
$\rightsquigarrow$ $O(1/r^2)$ rate for the Chebyshev measure on $[-1,1]$,
and other measures (with Jacobi weight) for **linear** polynomials

() Use tricks (**Taylor approx.**, **integration**, **'local similarity'**) to
transport the $O(1/r^2)$ rate for $[-1,1]$ to more sets (and measures):
hypercube, simplex, ball, sphere, ‘round’ convex bodies

(2) Design ‘*nice*’ **SoS polynomial densities**
‘*that look like the Dirac delta at a global minimizer*’
and use **push-up measures** to reduce to the **univariate case**
in order to get the $O((\log r)^2/r^2)$ rate for general K

First basic trick: reduction to the analysis of quadratic polynomials

Analyze simpler upper estimators

Lemma

Let $a \in K$ be a global minimizer of f in K.

Set $\gamma = \max_{x \in K} \|\nabla^2 f(x)\|$.

By Taylor's theorem, f has a **quadratic, separable** *upper estimator:*

$$f(x) \leq f(a) + \langle \nabla f(a), x - a \rangle + \gamma \|x - a\|^2 := g(x),$$

where $f(a) = g(a) \quad \rightsquigarrow \quad f_{min} = g_{\min}$.

Hence, for all $r \in \mathbb{N}$,

$$E^{(r)}(f) \leq E^{(r)}(g).$$

Analyze simpler upper estimators

Lemma

Let $a \in K$ be a global minimizer of f in K.

Set $\gamma = \max_{x \in K} \|\nabla^2 f(x)\|$.

By Taylor's theorem, f has a **quadratic, separable** *upper estimator:*

$$f(x) \leq f(a) + \langle \nabla f(a), x - a \rangle + \gamma \|x - a\|^2 := g(x),$$

where $f(a) = g(a)$ $\quad \rightsquigarrow \quad f_{min} = g_{\min}$.

Hence, for all $r \in \mathbb{N}$,

$$E^{(r)}(f) \leq E^{(r)}(g).$$

$\rightsquigarrow$ It suffices to analyze **quadratic** (separable) polynomials

and sometimes we may even obtain a **linear** upper estimator!
(e.g. for the sphere)

Eigenvalue reformulation & application to the univariate case: $K = [-1, 1]$

μ given measure with support K

$$f^{(r)} = \min \int_K f h \, d\mu \quad \text{s.t.} \quad h \text{ SoS}, \ \int_K h \, d\mu = 1, \ \deg(h) \leq 2r$$

μ given measure with support K

$$f^{(r)} = \min \int_K f h \, d\mu \quad \text{s.t.} \quad h \text{ SoS}, \; \int_K h \, d\mu = 1, \; \deg(h) \leq 2r$$

Choose an **orthonormal basis** $\{p_\alpha : |\alpha| \leq 2r\}$ of $\mathbb{R}[x]_{2r}$ w.r.t. μ and set

$$M_r(f) := \left(\int_K f \, p_\alpha p_\beta \, d\mu \right)_{|\alpha|,|\beta| \leq r}$$ (moment) Hankel-type matrix

μ given measure with support K

$$f^{(r)} = \min \int_K f h \, d\mu \quad \text{s.t.} \quad h \text{ SoS}, \ \int_K h \, d\mu = 1, \ \deg(h) \le 2r$$

Choose an **orthonormal basis** $\{p_\alpha : |\alpha| \le 2r\}$ of $\mathbb{R}[x]_{2r}$ w.r.t. μ and set

$$M_r(f) := \left(\int_K f \, p_\alpha p_\beta \, d\mu\right)_{|\alpha|,|\beta| \le r}$$ (moment) Hankel-type matrix

Note: h SoS $\iff$ $h = ((p_\alpha)_{|\alpha|\le r})^{\mathsf{T}} X (p_\alpha)_{|\alpha|\le r}$ for some $(X_{\alpha,\beta}) \succeq 0$

μ given measure with support K

$$f^{(r)} = \min \int_K fh \, d\mu \quad \text{s.t.} \quad h \text{ SoS}, \ \int_K h \, d\mu = 1, \ \deg(h) \leq 2r$$

Choose an **orthonormal basis** $\{p_\alpha : |\alpha| \leq 2r\}$ of $\mathbb{R}[x]_{2r}$ w.r.t. μ and set

$$M_r(f) := \left(\int_K f \, p_\alpha p_\beta \, d\mu\right)_{|\alpha|,|\beta| \leq r}$$ (moment) Hankel-type matrix

Note: h SoS $\iff h = ((p_\alpha)_{|\alpha| \leq r})^{\mathsf{T}} X (p_\alpha)_{|\alpha| \leq r}$ for some $(X_{\alpha,\beta}) \succeq 0$

$$\rightsquigarrow \quad \int_K fh \, d\mu = \langle M_r(f), X \rangle, \quad \int_K h \, d\mu = Tr(X)$$

$$f^{(r)} = \min \left\{ \langle M_r(f), X \rangle \ \text{ s.t. } \ \mathrm{Tr}(X) = 1, \ X \succeq 0 \right\} = \lambda_{\min}(M_r(f))$$

μ given measure with support K

$$f^{(r)} = \min \int_K f h \, d\mu \quad \text{s.t.} \quad h \text{ SoS}, \int_K h \, d\mu = 1, \deg(h) \leq 2r$$

Choose an **orthonormal basis** $\{p_\alpha : |\alpha| \leq 2r\}$ of $\mathbb{R}[x]_{2r}$ w.r.t. μ and set

$$M_r(f) := \left(\int_K f \, p_\alpha p_\beta \, d\mu\right)_{|\alpha|,|\beta| \leq r}$$ (moment) Hankel-type matrix

Note: h SoS $\iff h = ((p_\alpha)_{|\alpha| \leq r})^{\mathsf{T}} X (p_\alpha)_{|\alpha| \leq r}$ for some $(X_{\alpha,\beta}) \succeq 0$

$$\rightsquigarrow \quad \int_K f h \, d\mu = \langle M_r(f), X \rangle, \quad \int_K h \, d\mu = Tr(X)$$

$$f^{(r)} = \min \left\{ \langle M_r(f), X \rangle \text{ s.t. } \mathrm{Tr}(X) = 1, \ X \succeq 0 \right\} = \lambda_{\min}(M_r(f))$$

For $K = [-1, 1]$, can analyze $f^{(r)}$ for Chebyshev measure $d\mu = (1 - x^2)^{-1/2} dx$,
and for *any Jacobi measure* $d\mu = (1 - x^2)^{\lambda} dx$ $(\lambda > -1)$ when f is *linear*

Recall it is enough to deal with f **quadratic**: $f(x) = x$, $f(x) = x^2 + kx$

$K = [-1, 1]$, linear case: $f(x) = x$

$K = [-1, 1]$, linear case: $f(x) = x$

Theorem (classical theory of orthogonal polynomials)

Let $\{p_0, p_1, p_2, \ldots\}$ be a (graded) orthonormal basis of $\mathbb{R}[x]$ w.r.t. μ. Then the polynomials p_k satisfy a **3-term recurrence***:*

$$xp_k = a_k p_{k+1} + b_k p_k + a_{k-1} p_{k-1} \quad \text{for } k \geq 0, \quad p_0 \text{ constant}$$

$\rightsquigarrow$ *the (Jacobi) matrix* $M_r(x) = \left(\int_{-1}^{1} x p_i p_j \, d\mu\right)_{i,j=0}^{r}$ *is tri-diagonal and its eigenvalues are the roots of* p_{r+1}

$K = [-1, 1]$, linear case: $f(x) = x$

Theorem (classical theory of orthogonal polynomials)

Let $\{p_0, p_1, p_2, \ldots\}$ be a (graded) orthonormal basis of $\mathbb{R}[x]$ w.r.t. μ. Then the polynomials p_k satisfy a **3-term recurrence***:*

$$xp_k = a_k p_{k+1} + b_k p_k + a_{k-1} p_{k-1} \quad \text{for } k \geq 0, \quad p_0 \text{ constant}$$

$\rightsquigarrow$ *the (Jacobi) matrix $M_r(x) = \left(\int_{-1}^{1} x p_i p_j \; d\mu\right)_{i,j=0}^{r}$ is tri-diagonal and its eigenvalues are the roots of p_{r+1}*

$$M_r(x) = \begin{pmatrix} b_0 & a_0 & & & & & \\ a_0 & b_1 & a_1 & & & & \\ & a_1 & b_2 & a_2 & & & \\ & & a_2 & b_3 & a_3 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & a_{r-2} & b_{r-1} & a_{r-1} \\ & & & & & a_{r-1} & b_r \end{pmatrix}$$

$K = [-1, 1]$, linear case: $f(x) = x$

Theorem (classical theory of orthogonal polynomials)

Let $\{p_0, p_1, p_2, \ldots\}$ be a (graded) orthonormal basis of $\mathbb{R}[x]$ w.r.t. μ.
Then the polynomials p_k satisfy a **3-term recurrence***:*

$$xp_k = a_k p_{k+1} + b_k p_k + a_{k-1} p_{k-1} \quad \text{for } k \geq 0, \quad p_0 \text{ constant}$$

$\rightsquigarrow$ *the (Jacobi) matrix* $M_r(x) = \left(\int_{-1}^{1} x p_i p_j \, d\mu \right)_{i,j=0}^{r}$ *is tri-diagonal and its eigenvalues are the roots of* p_{r+1}

Theorem (de Klerk-L 2019)

For the Jacobi measure $d\mu = (1 - x^2)^\lambda dx$ with $\lambda > -1$, and $f(x) = x$:

$$f^{(r)} = \lambda_{\min}(M_r(x)) = \text{smallest root of } p_{r+1} = -1 + \Theta(1/r^2) = f_{\min} + \Theta(1/r^2)$$

Chebyshev measure on $K = [-1, 1]$, $f(x) = x^2 + kx$

Chebyshev measure on $K = [-1, 1]$, $f(x) = x^2 + kx$

(1) Minimizer on **boundary** (i.e., $k \notin [-2, 2]$): Then f has a **linear** upper estimator: $f(x) \leq g(x) := kx + 1 \quad \rightsquigarrow \quad E^{(r)}(f) \leq E^{(r)}(g) = O(1/r^2)$

NB: This holds for *any Jacobi measure* $(1 - x^2)^\lambda dx$, $\lambda > -1$

Chebyshev measure on $K = [-1, 1]$, $f(x) = x^2 + kx$

(1) Minimizer on **boundary** (i.e., $k \notin [-2, 2]$): Then f has a **linear** upper estimator: $f(x) \leq g(x) := kx + 1 \quad \rightsquigarrow \quad E^{(r)}(f) \leq E^{(r)}(g) = O(1/r^2)$

NB: This holds for *any Jacobi measure* $(1 - x^2)^{\lambda} dx$, $\lambda > -1$

(2) Minimizer in **interior**: Then, $f^{(r)} = \lambda_{\min}(M_r(f))$ where

Chebyshev measure on $K = [-1, 1]$, $f(x) = x^2 + kx$

(1) Minimizer on **boundary** (i.e., $k \notin [-2, 2]$): Then f has a **linear** upper estimator: $\quad f(x) \le g(x) := kx + 1 \quad \rightsquigarrow \quad E^{(r)}(f) \le E^{(r)}(g) = O(1/r^2)$

NB: This holds for *any Jacobi measure* $(1 - x^2)^\lambda dx$, $\lambda > -1$

(2) Minimizer in **interior**: Then, $\; f^{(r)} = \lambda_{\min}(M_r(f)) \quad$ where

$M_r(f) = \left(\int_{-1}^{1} (x^2 + kx) p_i p_j d\mu \right)_{i,j=0}^{r}$ is 5-diagonal **'almost' Toeplitz**:

$$M_r(f) = \left(\begin{array}{cccccccccc}
1/2 & \frac{k}{\sqrt{2}} & | & \frac{4}{\sqrt{2}} & & & & & & \\
\frac{k}{\sqrt{2}} & 3/4 & | & \frac{k}{2} & \frac{1}{4} & & & & & \\
- & - & - & - & - & - & - & - & - & \\
\frac{4}{\sqrt{2}} & \frac{k}{2} & | & a & b & c & & & & \\
 & \frac{1}{4} & | & b & a & b & c & & & \\
 & & | & c & b & \ddots & \ddots & \ddots & & \\
 & & | & & c & \ddots & \ddots & \ddots & \ddots & \\
 & & | & & & \ddots & \ddots & \ddots & \ddots & c \\
 & & | & & & & \ddots & \ddots & \ddots & b \\
 & & & & & & & c & b & a
\end{array} \right), \quad a = \frac{1}{2},\ b = \frac{k}{2},\ c = \frac{1}{4}$$

Write $M_r(f) = \begin{pmatrix} * & * & \cdots \\ * & * & \cdots \\ \vdots & \vdots & B \end{pmatrix}$, with B 5-diagonal **Toeplitz** of size $r-1$

Write $M_r(f) = \begin{pmatrix} * & * & \cdots \\ * & * & \cdots \\ \vdots & \vdots & B \end{pmatrix}$, with B 5-diagonal **Toeplitz** of size $r-1$

Embed B in a **symmetric circulant** matrix C of size $r+1$

Write $M_r(f) = \begin{pmatrix} * & * & \cdots \\ * & * & \cdots \\ \vdots & \vdots & B \end{pmatrix}$, with B 5-diagonal **Toeplitz** of size $r-1$

Embed B in a **symmetric circulant** matrix C of size $r+1$

By interlacing of eigenvalues:

$$\lambda_{\min}(M_r(f)) \ \leq \ \lambda_{\min}(B) \ \leq \ \lambda_3(C)$$

Write $M_r(f) = \begin{pmatrix} * & * & \cdots \\ * & * & \cdots \\ \vdots & \vdots & B \end{pmatrix}$, with B 5-diagonal **Toeplitz** of size $r-1$

Embed B in a **symmetric circulant** matrix C of size $r+1$

By interlacing of eigenvalues:

$$\lambda_{\min}(M_r(f)) \ \leq \ \lambda_{\min}(B) \ \leq \ \lambda_3(C) \ = \ -\frac{k^2}{4} + O(1/r^2) = f_{\min} + O(1/r^2)$$

Write $M_r(f) = \begin{pmatrix} * & * & \cdots \\ * & * & \cdots \\ \vdots & \vdots & B \end{pmatrix}$, with B 5-diagonal **Toeplitz** of size $r-1$

Embed B in a **symmetric circulant** matrix C of size $r+1$

By interlacing of eigenvalues:

$$\lambda_{\min}(M_r(f)) \le \lambda_{\min}(B) \le \lambda_3(C) = -\frac{k^2}{4} + O(1/r^2) = f_{\min} + O(1/r^2)$$

Theorem (de Klerk-L 2019)

For the **Chebyshev measure** *on* $[-1,1]^n$ *and any polynomial* f*:*

$$f^{(r)} - f_{min} = O(1/r^2)$$

$O\left(\frac{1}{r^2}\right)$ CONVERGENCE RATE FOR THE SPHERE

Key steps

(1) Reduce to the case when f is **linear**:

By Taylor, f has a **quadratic** upper estimator:

$f(x) \leq f(a) + \nabla f(a)^T (x - a) + \gamma \|x - a\|^2$

Key steps

(1) Reduce to the case when f is **linear**:

By Taylor, f has a **linear** upper estimator:

$f(x) \leq f(a) + \nabla f(a)^T (x - a) + \gamma \, (2 - 2x^T a)$

Key steps

(1) Reduce to the case when f is **linear**:

By Taylor, f has a **linear** upper estimator:

$f(x) \leq f(a) + \nabla f(a)^T(x - a) + \gamma\ (2 - 2x^T a)$

Up to rotation and translation, we may assume $f(x) = x_1$

Key steps

(1) Reduce to the case when f is **linear**:

By Taylor, f has a **linear** upper estimator:

$f(x) \leq f(a) + \nabla f(a)^T(x - a) + \gamma\ (2 - 2x^T a)$

Up to rotation and translation, we may assume $f(x) = x_1$

(2) Reduce to the analysis for the interval $[-1, 1]$:

Key fact: Let $h(x_1)$ be a degree $2r$ **univariate optimal** SoS density for the univariate problem $\min_{x_1 \in [-1,1]} x_1$ (with $d\mu = (1 - x_1^2)^{(n-3)/2} dx_1$)

Key steps

(1) Reduce to the case when f is **linear**:

By Taylor, f has a **linear** upper estimator:

$f(x) \leq f(a) + \nabla f(a)^T(x - a) + \gamma\,(2 - 2x^T a)$

Up to rotation and translation, we may assume $f(x) = x_1$

(2) Reduce to the analysis for the interval $[-1, 1]$:

Key fact: Let $h(x_1)$ be a degree $2r$ **univariate optimal** SoS density for the univariate problem $\min_{x_1 \in [-1,1]} x_1$ (with $d\mu = (1 - x_1^2)^{(n-3)/2} dx_1$)

Then $h(x_1)$ (rescaled) gives a (good) SoS density for the multivariate problem: $\min_{x \in \mathbb{S}^{n-1}} x_1$ (with μ Haar measure)

Key steps

(1) Reduce to the case when f is **linear**:

By Taylor, f has a **linear** upper estimator:

$f(x) \leq f(a) + \nabla f(a)^T(x-a) + \gamma\,(2 - 2x^T a)$

Up to rotation and translation, we may assume $f(x) = x_1$

(2) Reduce to the analysis for the interval $[-1, 1]$:

Key fact: Let $h(x_1)$ be a degree $2r$ **univariate optimal** SoS density for the univariate problem $\min_{x_1 \in [-1,1]} x_1$ (with $d\mu = (1-x_1^2)^{(n-3)/2} dx_1$)

Then $h(x_1)$ (rescaled) gives a (good) SoS density for the multivariate problem: $\min_{x \in \mathbb{S}^{n-1}} x_1$ (with μ Haar measure)

This is based on the **integration trick**:

$$\int_{-1}^{1} h(x_1)(1-x_1^2)^{\frac{n-3}{2}} dx_1 = C \int_{S^{n-1}} h(x_1) d\mu$$

$$\int_{-1}^{1} x_1 h(x_1)(1-x_1^2)^{\frac{n-3}{2}} dx_1 = C \int_{S^{n-1}} x_1 h(x_1) d\mu$$

Key steps

(1) Reduce to the case when f is **linear**:

By Taylor, f has a **linear** upper estimator:

$f(x) \leq f(a) + \nabla f(a)^T(x-a) + \gamma\,(2 - 2x^T a)$

Up to rotation and translation, we may assume $f(x) = x_1$

(2) Reduce to the analysis for the interval $[-1, 1]$:

Key fact: Let $h(x_1)$ be a degree $2r$ **univariate optimal** SoS density for the univariate problem $\min_{x_1 \in [-1,1]} x_1$ (with $d\mu = (1-x_1^2)^{(n-3)/2} dx_1$)

Then $h(x_1)$ (rescaled) gives a (good) SoS density for the multivariate problem: $\min_{x \in \mathbb{S}^{n-1}} x_1$ (with μ Haar measure)

This is based on the **integration trick**:

$$1 = \int_{-1}^{1} h(x_1)(1-x_1^2)^{\frac{n-3}{2}} dx_1 = C \int_{S^{n-1}} h(x_1) d\mu$$

$$-1 + O\left(\frac{1}{r^2}\right) = \int_{-1}^{1} x_1 h(x_1)(1-x_1^2)^{\frac{n-3}{2}} dx_1 = C \int_{S^{n-1}} x_1 h(x_1) d\mu$$

[de Klerk-L 2020]

The bound $1/r^2$ is tight

Theorem (de Klerk-L 2020)

For any polynomial $f(x) = (-1)^d (c^T x)^d$, *the analysis is* **tight***:*

$$E^{(r)}(f) = \Omega\left(\frac{1}{r^2}\right)$$

This relies on the following **link to cubature rules**:

Fact (Martinez et al. 2019)

Let $\{(x^{(i)}, w_i) : i \in [N]\}$ *be a* **positive cubature rule** *on* K *that is* **exact for integrating polynomials of degree** $d + 2r$*. If* f *has degree* d

$$f^{(r)} = \int_K fh d\mu = \sum_{i=1}^{N} w_i f(x^{(i)}) h(x^{(i)}) \geq \min_{i \in [N]} f(x^{(i)}) \overbrace{\sum_i w_i h(x^{(i)})}^{=1} \geq f_{\min}$$

For $K = \mathbb{S}^{n-1}$, use cubature rule from the roots of Gegenbauer polys.

‘Local similarity’ trick

&

Application to box, ball, simplex, round convex body

'Local similarity': lift results from $(\widehat{K}, \widehat{w})$ to (K, w)

Lemma (Slot-L 2020)

Let $a \in K$ be a global minimizer of f in K. Assume:

$K \subseteq \widehat{K}$, w a weight function on K, $\widehat{w}$ weight function on $\widehat{K}$ satisfy:

'Local similarity': lift results from $(\widehat{K}, \widehat{w})$ to (K, w)

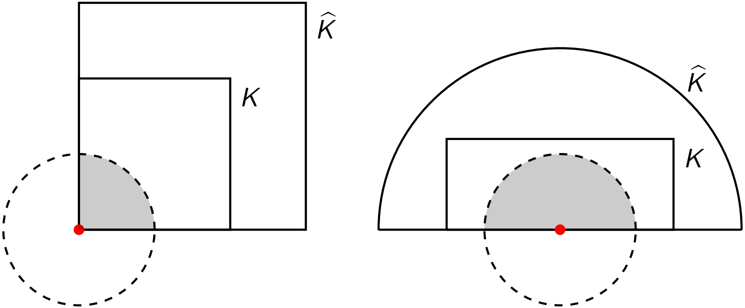
Lemma (Slot-L 2020)

Let $a \in K$ be a global minimizer of f in K. Assume:

$K \subseteq \widehat{K}$, w a weight function on K, $\widehat{w}$ weight function on $\widehat{K}$ satisfy:

(1) $K, \widehat{K}$ *are* **'locally similar' at** *a:*

$$K \cap B_\epsilon(a) = \widehat{K} \cap B_\epsilon(a) \quad \textit{for some } \epsilon > 0.$$



'Local similarity': lift results from $(\widehat{K}, \widehat{w})$ to (K, w)

Lemma (Slot-L 2020)

Let $a \in K$ be a global minimizer of f in K. Assume:

$K \subseteq \widehat{K}$, w a weight function on K, $\widehat{w}$ weight function on $\widehat{K}$ satisfy:

(1) *$K, \widehat{K}$ are* **'locally similar' at** *a:*

$$K \cap B_\epsilon(a) = \widehat{K} \cap B_\epsilon(a) \quad \text{for some } \epsilon > 0.$$

(2) *$w, \widehat{w}$ are* **'locally similar' at** *a:*

$$m \cdot \widehat{w}(x) \leq w(x) \;\text{ on }\; \mathrm{int}(K) \cap B_\epsilon(a) \quad \text{for some } \epsilon, m > 0.$$

'Local similarity': lift results from $(\widehat{K}, \widehat{w})$ to (K, w)

Lemma (Slot-L 2020)

Let $a \in K$ be a global minimizer of f in K. Assume:

$K \subseteq \widehat{K}$, *w a weight function on K, $\widehat{w}$ weight function on $\widehat{K}$ satisfy:*

(1) $K, \widehat{K}$ *are* **'locally similar' at** *a:*

$$K \cap B_\epsilon(a) = \widehat{K} \cap B_\epsilon(a) \quad \textit{for some } \epsilon > 0.$$

(2) $w, \widehat{w}$ *are* **'locally similar' at** *a:*

$$m \cdot \widehat{w}(x) \leq w(x) \;\; \textit{on} \;\; \mathrm{int}(K) \cap B_\epsilon(a) \quad \textit{for some } \epsilon, m > 0.$$

(3) $w(x) \leq \widehat{w}(x)$ *for all $x \in \mathrm{int}(K)$.*

'Local similarity': lift results from $(\widehat{K}, \widehat{w})$ to (K, w)

Lemma (Slot-L 2020)

Let $a \in K$ be a global minimizer of f in K. Assume:

$K \subseteq \widehat{K}$, w a weight function on K, $\widehat{w}$ weight function on $\widehat{K}$ satisfy:

(1) $K, \widehat{K}$ *are* **'locally similar' at** *a:*

$$K \cap B_\epsilon(a) = \widehat{K} \cap B_\epsilon(a) \quad \text{for some } \epsilon > 0.$$

(2) $w, \widehat{w}$ *are* **'locally similar' at** *a:*

$$m \cdot \widehat{w}(x) \leq w(x) \;\text{ on }\; \mathrm{int}(K) \cap B_\epsilon(a) \quad \text{for some } \epsilon, m > 0.$$

(3) $w(x) \leq \widehat{w}(x)$ *for all $x \in \mathrm{int}(K)$.*

Then, f has an upper estimator g on $\widehat{K}$, exact at a, satisfying

$$E^{(r)}_{K,w}(f) \leq E^{(r)}_{\widehat{K},\widehat{w}}(g).$$

'Local similarity': lift results from $(\widehat{K}, \widehat{w})$ to (K, w)

Lemma (Slot-L 2020)

Let $a \in K$ be a global minimizer of f in K. Assume:

$K \subseteq \widehat{K}$, w a weight function on K, $\widehat{w}$ weight function on $\widehat{K}$ satisfy:

(1) $K, \widehat{K}$ *are* **'locally similar' at** *a:*

$$K \cap B_\epsilon(a) = \widehat{K} \cap B_\epsilon(a) \quad \text{for some } \epsilon > 0.$$

(2) $w, \widehat{w}$ *are* **'locally similar' at** *a:*

$$m \cdot \widehat{w}(x) \leq w(x) \;\; \text{on} \;\; \mathrm{int}(K) \cap B_\epsilon(a) \quad \text{for some } \epsilon, m > 0.$$

(3) $w(x) \leq \widehat{w}(x)$ *for all $x \in \mathrm{int}(K)$.*

Then, f has an upper estimator g on $\widehat{K}$, exact at a, satisfying

$$E^{(r)}_{K,w}(f) \leq E^{(r)}_{\widehat{K},\widehat{w}}(g).$$

Note: (1),(2) clearly hold if $a \in \mathrm{int}(K)$

Lift known $O(1/r^2)$ rate for $\widehat{K} = [-1,1], \lambda = -\frac{1}{2}$

(1) to $K = [-1,1]$, with $w(x) = (1-x^2)^{\lambda}$, $\lambda \geq -1/2$, any f

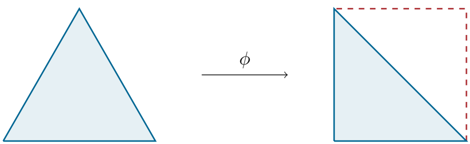
[using Chebyshev weight $\widehat{w}(x) = (1-x^2)^{-1/2}$], to $K = [-1,1]^n$

(2) to any K, with $w = 1$, when minimizer a lies in the **interior** of K

[using $K \subseteq \widehat{K} = [-1,1]^n$ with $\widehat{w} = 1$]

(3) to K simplex, with $w = 1$, when minimizer lies on the **boundary**

[after applying affine mapping and using $\widehat{K} = [0,1]^n$ with $\widehat{w} = 1$]



Lift known $O(1/r^2)$ rate for $\widehat{K} = [-1,1], \lambda = -\frac{1}{2}$

(1) to $K = [-1,1]$, with $w(x) = (1-x^2)^{\lambda}$, $\lambda \geq -1/2$, any f

[using Chebyshev weight $\widehat{w}(x) = (1-x^2)^{-1/2}$], to $K = [-1,1]^n$

(2) to any K, with $w = 1$, when minimizer a lies in the **interior** of K

[using $K \subseteq \widehat{K} = [-1,1]^n$ with $\widehat{w} = 1$]

(3) to K simplex, with $w = 1$, when minimizer lies on the **boundary**

[after applying affine mapping and using $\widehat{K} = [0,1]^n$ with $\widehat{w} = 1$]

(4) to K ball, with $w(x) = (1 - \|x\|^2)^{\lambda}$, $\lambda \geq 0$

[using a linear upper estimator and an integration trick, when the minimizer lies on the **boundary**]

(5) to K 'round' convex body, with $w = 1$ (i.e., K has inscribed and circumscribed tangent balls at any boundary point)

[using the result for the ball $\widehat{K}$ with $\widehat{w} = 1$]

SoS approximations of Dirac measures & application to convex bodies and to fat compact semialgebraic sets

Cheaper bounds using the 'push-forward measure'

- μ measure supported by K (e.g., Lebesgue measure)

$\rightsquigarrow$ μ_f **push-forward** of μ by f, supported by $f(K) = [f_{\min}, f_{\max}] \subseteq \mathbb{R}$:

$$\int_{f(K)} \varphi(t) d\mu_f(t) = \int_K \varphi(f(x)) d\mu(x) \quad \text{for } \varphi : \mathbb{R} \to \mathbb{R}$$

Cheaper bounds using the ‘push-forward measure’

- μ measure supported by K (e.g., Lebesgue measure)

$\rightsquigarrow$ μ_f **push-forward** of μ by f, supported by $f(K) = [f_{\min}, f_{\max}] \subseteq \mathbb{R}$:

$$\int_{f(K)} \varphi(t) d\mu_f(t) = \int_K \varphi(f(x)) d\mu(x) \quad \text{for } \varphi : \mathbb{R} \to \mathbb{R}$$

- This motivates defining the *weaker* **‘univariate’ upper bounds**:

$$f^{(r)}_{pfm} = \min \int_K f(x) s(f(x)) d\mu(x) \quad \text{s.t.} \quad \int_K s(f(x)) d\mu(x) = 1, \; \deg(s) \leq 2r$$

s **univariate** sum-of-squares

Cheaper bounds using the 'push-forward measure'

- μ measure supported by K (e.g., Lebesgue measure)

$\rightsquigarrow$ μ_f **push-forward** of μ by f, supported by $f(K) = [f_{\min}, f_{\max}] \subseteq \mathbb{R}$:

$$\int_{f(K)} \varphi(t) d\mu_f(t) = \int_K \varphi(f(x)) d\mu(x) \quad \text{for } \varphi : \mathbb{R} \to \mathbb{R}$$

- This motivates defining the *weaker* **'univariate' upper bounds**:

$$f_{pfm}^{(r)} = \min \int_K f(x) s(f(x)) d\mu(x) \ \text{ s.t. } \ \int_K s(f(x)) d\mu(x) = 1, \ \deg(s) \leq 2r$$

s **univariate** sum-of-squares

$$= \min \int_{f(K)} t \cdot s(t) d\mu_f(t) \ \text{ s.t. } \ \int_{f(K)} s(t) d\mu_f(t) = 1, \ \deg(s) \leq 2r$$

s **univariate** sum-of-squares

Cheaper bounds using the 'push-forward measure'

- μ measure supported by K (e.g., Lebesgue measure)

$\rightsquigarrow$ μ_f **push-forward** of μ by f, supported by $f(K) = [f_{\min}, f_{\max}] \subseteq \mathbb{R}$:

$$\int_{f(K)} \varphi(t) d\mu_f(t) = \int_K \varphi(f(x)) d\mu(x) \quad \text{for } \varphi : \mathbb{R} \to \mathbb{R}$$

- This motivates defining the *weaker* **'univariate' upper bounds**:

$$f_{pfm}^{(r)} = \min \int_K f(x) s(f(x)) d\mu(x) \ \text{ s.t. } \int_K s(f(x)) d\mu(x) = 1, \ \deg(s) \leq 2r$$

s **univariate** sum-of-squares

$$= \min \int_{f(K)} t \cdot s(t) d\mu_f(t) \text{ s.t. } \int_{f(K)} s(t) d\mu_f(t) = 1, \ \deg(s) \leq 2r$$

s **univariate** sum-of-squares

$f_{\min} \leq f^{(rd)} \leq f_{pfm}^{(r)}$ if $d = \deg(f)$; $f_{pfm}^{(r)} \searrow f_{\min}$ [Lasserre 2019]

Cheaper bounds using the 'push-forward measure'

- μ measure supported by K (e.g., Lebesgue measure)

$\rightsquigarrow$ μ_f **push-forward** of μ by f, supported by $f(K) = [f_{\min}, f_{\max}] \subseteq \mathbb{R}$:

$$\int_{f(K)} \varphi(t) d\mu_f(t) = \int_K \varphi(f(x)) d\mu(x) \quad \text{for } \varphi : \mathbb{R} \to \mathbb{R}$$

- This motivates defining the *weaker* **'univariate' upper bounds**:

$$f_{pfm}^{(r)} = \min \int_K f(x) s(f(x)) d\mu(x) \ \text{ s.t. } \ \int_K s(f(x)) d\mu(x) = 1, \ \deg(s) \leq 2r$$

s **univariate** sum-of-squares

$$= \min \int_{f(K)} t \cdot s(t) d\mu_f(t) \ \text{ s.t. } \ \int_{f(K)} s(t) d\mu_f(t) = 1, \ \deg(s) \leq 2r$$

s **univariate** sum-of-squares

$f_{\min} \leq f^{(rd)} \leq f_{pfm}^{(r)}$ if $d = \deg(f)$; $f_{pfm}^{(r)} \searrow f_{\min}$ [Lasserre 2019]

Error rate: $f_{pfm}^{(r)} - f_{\min} = O\left(\frac{(\log r)^2}{r^2}\right)$ [Slot-L 2020]

Cheaper bounds using the 'push-forward measure'

- μ measure supported by K (e.g., Lebesgue measure)

$\rightsquigarrow$ μ_f **push-forward** of μ by f, supported by $f(K) = [f_{\min}, f_{\max}] \subseteq \mathbb{R}$:

$$\int_{f(K)} \varphi(t) d\mu_f(t) = \int_K \varphi(f(x)) d\mu(x) \quad \text{for } \varphi : \mathbb{R} \to \mathbb{R}$$

- This motivates defining the *weaker* **'univariate' upper bounds**:

$$f_{pfm}^{(r)} = \min \int_K f(x) s(f(x)) d\mu(x) \ \text{ s.t. } \ \int_K s(f(x)) d\mu(x) = 1, \ \deg(s) \le 2r$$

s **univariate** sum-of-squares

$$= \min \int_{f(K)} t \cdot s(t) d\mu_f(t) \ \text{ s.t. } \ \int_{f(K)} s(t) d\mu_f(t) = 1, \ \deg(s) \le 2r$$

s **univariate** sum-of-squares

$f_{pfm}^{(r)}$ = **smallest root** of orthogonal polynomial p_{r+1} w.r.t. measure μ_f, but this is **not known** in general! $\rightsquigarrow$ need another approach

• May assume $f(K) = [0,1]$ (up to affine transformation)

• Use the (half-)**needle polynomials** $s_r^h(t)$ of [Kroó-Swetits 1992] ($h > 0$, $r \in \mathbb{N}$, defined as squares of Chebyshev polynomials) with degree $4r$ and satisfying

$$s_r^h(t) \begin{cases} = 1 & \text{at } t = 0 \\ \leq 1 & \text{at } t \in [0,1] \\ \leq 4e^{-\frac{1}{2}\sqrt{h}r} & \text{at } t \in [h,1] \end{cases}$$

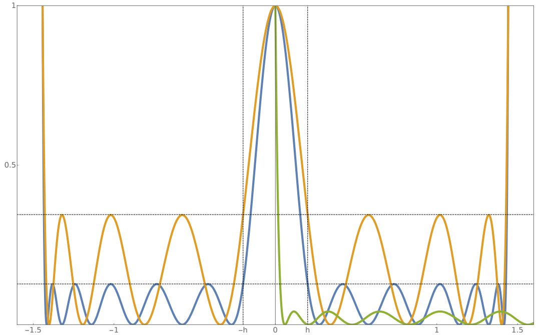
as **univariate SoS density** (with $h = (\log r)^2/r^2$)

- May assume $f(K) = [0,1]$ (up to affine transformation)

- Use the (half-)**needle polynomials** $s_r^h(t)$ of [Kroó-Swetits 1992] ($h > 0$, $r \in \mathbb{N}$, defined as squares of Chebyshev polynomials) with degree $4r$ and satisfying

$$s_r^h(t) \begin{cases} = 1 & \text{at } t = 0 \\ \leq 1 & \text{at } t \in [0,1] \\ \leq 4e^{-\frac{1}{2}\sqrt{h}r} & \text{at } t \in [h,1] \end{cases}$$

as **univariate SoS density** (with $h = (\log r)^2/r^2$)



In green, the half-needle polynomial with $h = 1/5$

Theorem (Slot-L 2020)

Assume K is a **convex body**, *or K is compact,* **fat** *(with dense interior) and* **semialgebraic**. *Then*

$$f_{pfm}^{(r)} - f_{min} = O\Big(\frac{(\log r)^2}{r^2}\Big)$$

- The analysis is **almost tight** and there can be a **separation** between the multivariate and univariate bounds:

  For $f(x) = x^{2k}$ and $K = [-1, 1]$:

$$f_{\min} = 0 \le f^{(2kr)} = O(\frac{(\log r)^{2k}}{r^{2k}}) \le f_{pfm}^{(r)} = \Omega(\frac{1}{r^2})$$

- **Open question:** Can one get rid of the factor $(\log r)^2$?

Concluding remarks (1)

- Can compute $f^{(r)}$ as smallest eigenvalue of a matrix with size $O(n^r)$, and the bounds $f_{pfm}^{(r)}$ as smallest eigenvalue of a matrix of size $r+1$

  ... but computing its entries is more expensive

  since one needs to integrate powers of f

Concluding remarks (1)

- Can compute $f^{(r)}$ as smallest eigenvalue of a matrix with size $O(n^r)$, and the bounds $f^{(r)}_{pfm}$ as smallest eigenvalue of a matrix of size $r+1$
  ... but computing its entries is more expensive
  since one needs to integrate powers of f

- There is a **link to simulated annealing:** [de Klerk-L 2018]
  Instead of sampling w.r.t. Boltzman distribution $e^{-f(x)/T}$ with temperature $T > 0$, use the Taylor expansion of $t \to e^{-t/T}$ truncated at degree $r \sim 1/T$ as univariate SoS density (to analyze $f^{(r)}_{pfm}$ and thus $f^{(r)}$)

  But, while the measure-based upper bound has error $O(\frac{(\log r)^2}{r^2})$, it is known that the simulated annealing bound has error $O(1/r)$ for convex f [Kalai-Vempala 2006], which is tight for linear f

Concluding remarks (1)

- Can compute $f^{(r)}$ as smallest eigenvalue of a matrix with size $O(n^r)$, and the bounds $f^{(r)}_{pfm}$ as smallest eigenvalue of a matrix of size $r+1$
  ... but computing its entries is more expensive
  since one needs to integrate powers of f

- There is a **link to simulated annealing:** [de Klerk-L 2018]
  Instead of sampling w.r.t. Boltzman distribution $e^{-f(x)/T}$ with temperature $T > 0$, use the Taylor expansion of $t \to e^{-t/T}$ truncated at degree $r \sim 1/T$ as univariate SoS density
  (to analyze $f^{(r)}_{pfm}$ and thus $f^{(r)}$)

  But, while the measure-based upper bound has error $O(\frac{(\log r)^2}{r^2})$, it is known that the simulated annealing bound has error $O(1/r)$ for convex f [Kalai-Vempala 2006], which is tight for linear f

- The error analysis for $f^{(r)}$ extends to **rational** functions f [dK-L'19]

Concluding remarks (1)

- Can compute $f^{(r)}$ as smallest eigenvalue of a matrix with size $O(n^r)$, and the bounds $f^{(r)}_{pfm}$ as smallest eigenvalue of a matrix of size $r+1$

  ... but computing its entries is more expensive

  since one needs to integrate powers of f

- There is a **link to simulated annealing:** [de Klerk-L 2018]

  Instead of sampling w.r.t. Boltzman distribution $e^{-f(x)/T}$ with temperature $T > 0$, use the Taylor expansion of $t \to e^{-t/T}$ truncated at degree $r \sim 1/T$ as univariate SoS density (to analyze $f^{(r)}_{pfm}$ and thus $f^{(r)}$)

  But, while the measure-based upper bound has error $O(\frac{(\log r)^2}{r^2})$, it is known that the simulated annealing bound has error $O(1/r)$ for convex f [Kalai-Vempala 2006], which is tight for linear f

- The error analysis for $f^{(r)}$ extends to **rational** functions f [dK-L'19] and can be adapted to the **general problem of moments**

  [de Klerk-Postek-Kuhn'19]

Concluding remarks (2)

- Comparison to **grid-point search**: When optimizing over all grid points in $K = [0,1]^n$ with denominator r one gets an upper bound with error in $O(1/r^2)$

  ... but this requires r^n function evaluations, compared to solving an SDP with matrix size $O(n^r)$ for the bounds $f^{(r)}$

Concluding remarks (2)

- Comparison to **grid-point search**: When optimizing over all grid points in $K = [0,1]^n$ with denominator r one gets an upper bound with error in $O(1/r^2)$

  ... but this requires r^n function evaluations, compared to solving an SDP with matrix size $O(n^r)$ for the bounds $f^{(r)}$

- How to get an improved analysis for the **lower bounds** $f_{(r)}$?

  - An analysis in $O(\frac{1}{r^2})$ is shown by [Fang-Fawzi 2020] for the **unit sphere**, which interestingly uses the analysis for the **upper bounds** (for a related univariate problem, obtained by symmetry reduction), and the polynomial kernel method
  - Extension to the case of the binary hypercube $K = \{0,1\}^n$ [Slot-L 2021]
  - **Open question:** Extension to more general sets K?

Some references

- K. Fang, H. Fawzi: *The sum-of-squares hierarchy on the sphere and application to quantum information theory*. Math. Prog., 2020
- E. de Klerk, M. Laurent: *Comparison of Lasserre's measure-based hierarchy of upper bunds for polynomial optimization to bounds obtained by simulated annealing.* Math. of OR, 2018
- - - *Worst-case examples for Lasserre's measure-based hierarchy for polynomial optimization on the hypercube.* Math. of OR, 2020
- - - *Convergence analysis of upper bounds for polynomial minimization on the sphere.* Math. Prog., 2020
- E. de Klerk, K. Postek, D. Kuhn: *Distributionally robust optimization with polynomialdensities: theory, models and algorithms*. Math. Prog., 2020
- J.B. Lasserre: *A new look at nonnegativity on closed sets and polynomial optimization*. SIOPT, 2011
- - - *Connecting optimization with spectral analysis of tri-diagonal Hankel matrices.* arXiv:1907.09784v4, 2019
- L. Slot, M. Laurent: *Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets.* Math. Prog., 2020
- - - *Near-optimal analysis of univariate moment bounds for polynomial optimization.* Math. Prog., 2020

Thank you!