Performance analysis of approximation hierarchies for polynomial optimization

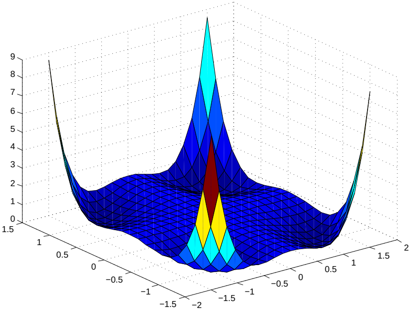
CWI

Tilburg University

Monique Laurent

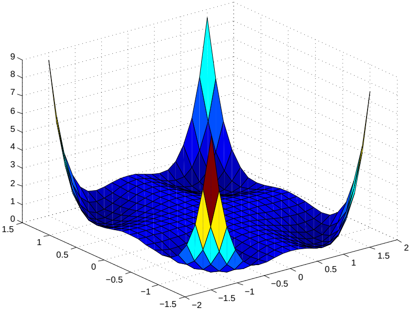
Joint works with Lucas Slot and Etienne de Klerk

Fields Distinguished Lecture Series - May 12, 2021



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

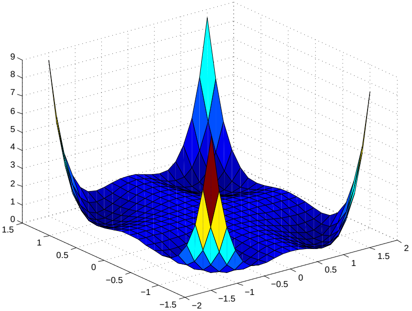
$$f_{\min} = \min_{x \in K} f(x)$$



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

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

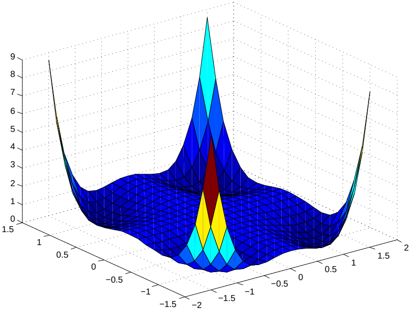
NP-hard problem: it captures hard combinatorial problems
(like computing $\alpha(G)$: the maximum size of a stable set in a graph G)
when K is a hypercube or a simplex and $\deg(f) = 2$,
or K is a sphere and $\deg(f) = 3$



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:

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

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

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

(2) Lasserre *measure-based* **upper bounds:**

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

Common feature:

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

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

This lecture: Main focus on the **error analysis** of these 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

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

When $$K = \{x \in \mathbb{R}^n : \ g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$ 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**:

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

'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 $\boxed{K = \{x \in \mathbb{R}^n : \; g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}}$ 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**:

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

- $f_{(r)} \leq f_{(r+1)} \leq f_{\min}, \quad f_{(r)} \nearrow f_{\min}$ 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] **Improved 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 [Faybusovich 2003], [Doherty-Wehner 2012]

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] **Improved 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 [Faybusovich 2003], [Doherty-Wehner 2012]

There is an intimate link with the **analysis of the upper bounds**

More later!

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 σ is a sum-of-squares density:*

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

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 σ is a sum-of-squares density:*

$$f_{min} = \inf_\sigma \int_K f(x)\sigma(x)\, d\mu \;\; s.t. \;\; \sigma \; SoS, \; \int_K \sigma(x)\, d\mu = 1$$

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

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

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 σ is a sum-of-squares density:*

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

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

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

- $f_{\min} \leq f^{(r+1)} \leq f^{(r)}$, $f^{(r)} \searrow f_{\min}$, $f^{(r)}$ 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 σ is a sum-of-squares density:*

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

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

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

- $f_{\min} \leq f^{(r+1)} \leq f^{(r)}$, $\quad f^{(r)} \searrow f_{\min}$, $\quad f^{(r)}$ 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 σ is a sum-of-squares density:*

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

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

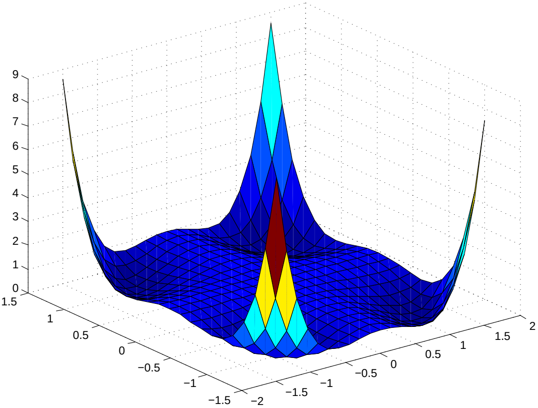
$$f^{(r)} = \inf_\sigma \int_K f(x)\sigma(x)\, d\mu \quad \text{s.t.} \quad \sigma \text{ SoS}, \int_K \sigma(x)\, d\mu = 1, \ \deg(\sigma) \leq 2r$$

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

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

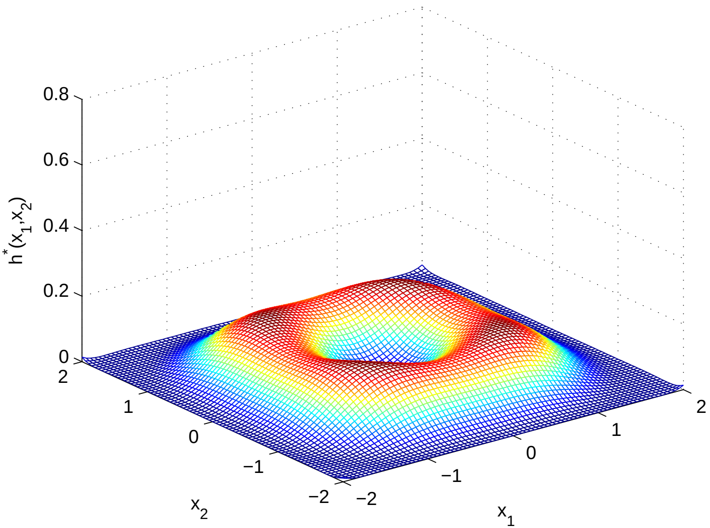
$$f(x_1, x_2) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$$

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



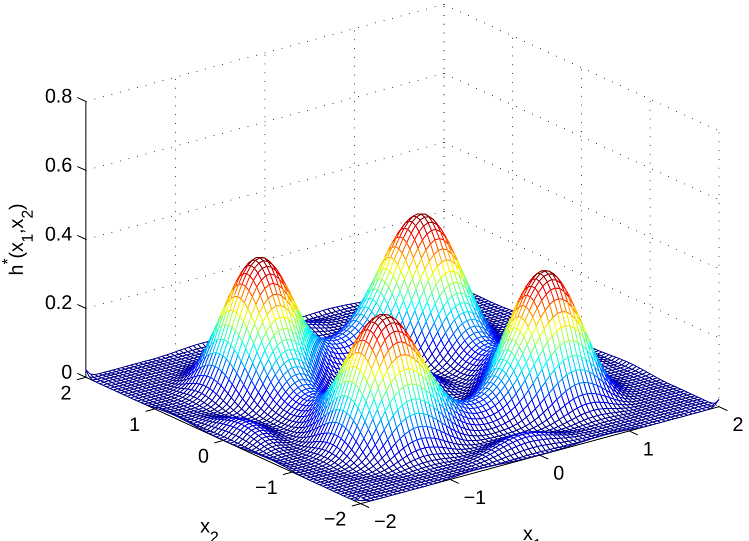
Example: Motzkin polynomial on $[-2,2]^2$ (ctd.)

Optimal SoS density σ of degree 12



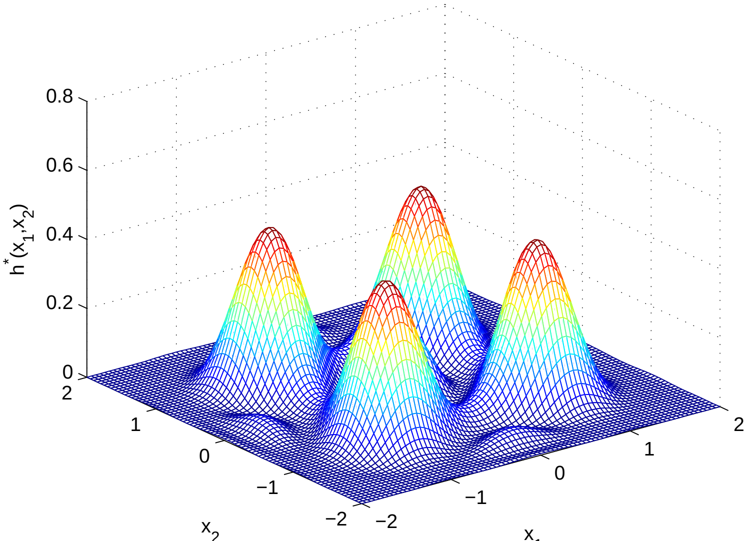
Example: Motzkin polynomial on $[-2, 2]^2$ (ctd.)

Optimal SoS density σ of degree 16



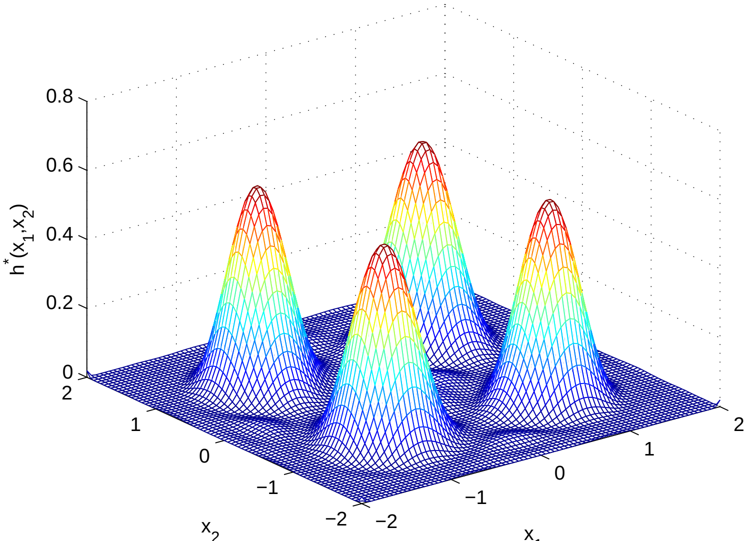
Example: Motzkin polynomial on $[-2,2]^2$ (ctd.)

Optimal SoS density σ of degree 20



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

Optimal SoS density σ 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) Design '*nice*' **SoS polynomial densities**
'*that look like the Dirac delta at a global minimizer*'
and reduce to the **univariate case**
in order to get the $O((\log r)^2/r^2)$ rate for general K

Key proof strategies

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

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

Basic trick: suffices to analyse **polynomials with degree** ≤ 2

Key proof strategies

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

(2) 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

Basic trick: suffices to analyse **polynomials with degree** ≤ 2

(3) Use more 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

Strategy 1: Use SoS approximations of Dirac measures

$\rightsquigarrow O(\frac{\log^2 r}{r^2})$ rate for general K

Step 1: Analyse cheaper (univariate) bounds

Instead of the **multivariate** upper bounds:

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

Step 1: Analyse cheaper (univariate) bounds

Instead of the **multivariate** upper bounds:

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

Consider the **weaker univariate** upper bounds:

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

s **univariate** sum-of-squares

Step 1: Analyse cheaper (univariate) bounds

Instead of the **multivariate** upper bounds:

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

Consider the **weaker univariate** upper bounds:

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

s **univariate** sum-of-squares

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

s **univariate** sum-of-squares

with μ_f is the **push-forward** of μ by f, supported by $[f_{\min}, f_{\max}] \subseteq \mathbb{R}$

Step 1: Analyse cheaper (univariate) bounds

Instead of the **multivariate** upper bounds:

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

Consider the **weaker univariate** upper bounds:

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

s **univariate** sum-of-squares

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

s **univariate** sum-of-squares

with μ_f is the **push-forward** of μ by f, supported by $[f_{\min}, f_{\max}] \subseteq \mathbb{R}$

Then: $f_{\min} \le f^{(rd)} \le f^{(r)}_{pfm}$ if $d = \deg(f)$

$f^{(r)}_{pfm} \searrow f_{\min}$ [Lasserre 2019]

Step 1: Analyse cheaper (univariate) bounds

Instead of the **multivariate** upper bounds:

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

Consider the **weaker univariate** upper bounds:

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

s **univariate** sum-of-squares

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

s **univariate** sum-of-squares

with μ_f is the **push-forward** of μ by f, supported by $[f_{\min}, f_{\max}] \subseteq \mathbb{R}$

Then: $f_{\min} \le f^{(rd)} \le f^{(r)}_{pfm}$ if $d = \deg(f)$

$f^{(r)}_{pfm} \searrow f_{\min}$ [Lasserre 2019]

Step 2: Use SoS approximations of the Dirac delta

Use the degree $4r$ (half-)**needle polynomials** $s_r^h(t)$ of [Kroó-Swetits'92] ($h > 0$, $r \in \mathbb{N}$, defined as squares of Chebyshev polynomials)

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

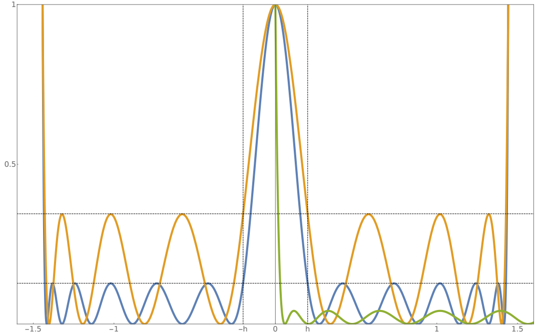
as **univariate SoS density** (with $h = (\log r)^2/r^2$)

Step 2: Use SoS approximations of the Dirac delta

Use the degree $4r$ (half-)**needle polynomials** $s_r^h(t)$ of [Kroó-Swetits'92] ($h > 0$, $r \in \mathbb{N}$, defined as squares of Chebyshev polynomials)

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

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 a* **compact semialgebraic set** *with a* **dense interior** *(aka* **fat***). Then*

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

Theorem (Slot-L 2020)

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

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

- The analysis is **almost tight**

  There can be a **separation** between the multivariate and univariate bounds:

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

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

Theorem (Slot-L 2020)

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

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

- The analysis is **almost tight**

  There can be a **separation** between the multivariate and univariate bounds:

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

$$f_{\min} = 0 \leq f^{(2dr)} = O\left(\frac{(\log r)^{2d}}{r^{2d}}\right) \leq f_{pfm}^{(r)} = \Omega\left(\frac{1}{r^2}\right)$$

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

  Yes *for the* **multivariate bounds** $f^{(r)}$, *for some* **nice sets** K

First basic trick: Suffices to analyze quadratic polynomials

Analyze simpler upper estimators

Lemma

Let $\mathbf{a} \in K$ *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:*

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

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

Analyze simpler upper estimators

Lemma

Let $\mathbf{a} \in K$ *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:*

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

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

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

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

Analyze simpler upper estimators

Lemma

Let $\mathbf{a} \in K$ *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:*

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

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

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

μ given measure with support K

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

μ given measure with support K

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

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

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

μ given measure with support K

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

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

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

$\rightsquigarrow \int_K \sigma \; d\mu = Tr(X)$ as $\int_K \sigma \; d\mu = \sum_{\alpha,\beta} X_{\alpha,\beta} \int_K p_\alpha p_\beta d\mu$

μ given measure with support K

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

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

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

$\rightsquigarrow \int_K \sigma \, d\mu = Tr(X)$ as $\int_K \sigma \, d\mu = \sum_{\alpha,\beta} X_{\alpha,\beta} \int_K p_\alpha p_\beta d\mu$

$\rightsquigarrow \int_K f\sigma \, d\mu = \langle M_r(f), X\rangle$ as $\int_K f\sigma \, d\mu = \sum_{\alpha,\beta} X_{\alpha,\beta} \int_K f p_\alpha p_\beta d\mu$

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

μ given measure with support K

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

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

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

$\rightsquigarrow \int_K \sigma \, d\mu = Tr(X)$ as $\int_K \sigma \, d\mu = \sum_{\alpha,\beta} X_{\alpha,\beta} \int_K p_\alpha p_\beta d\mu$

$\rightsquigarrow \int_K f\sigma \, d\mu = \langle M_r(f), X\rangle$ as $\int_K f\sigma \, d\mu = \sum_{\alpha,\beta} X_{\alpha,\beta} \int_K f p_\alpha p_\beta d\mu$

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

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

[Lasserre 2011]

Analysis in the univariate case: $K = [-1, 1]$

suffices to consider:

f linear, or quadratic

$K = [-1, 1]$, 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}

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

$K = [-1, 1]$, 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***:*

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

$\rightsquigarrow$ *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}

Theorem (de Klerk-L 2020)

For $f(x) = x$:

$$f^{(r)} = \lambda_{\min}(M_r(x))$$

$K = [-1,1]$, 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***:*

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

$\rightsquigarrow$ *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 2020)

For $f(x) = x$:

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

$K = [-1, 1]$, 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***:*

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

$\rightsquigarrow$ *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 2020)

For $f(x) = x$:

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

for the Jacobi measure $d\mu = (1 - x^2)^{\lambda} dx$ with $\lambda > -1$

$K = [-1, 1]$, 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}*

Theorem (de Klerk-L 2020)

For $f(x) = x$:

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

for the Jacobi measure $d\mu = (1 - x^2)^{\lambda} dx$ with $\lambda > -1$

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: $f(x) \le g(x) := kx + 1 \quad \rightsquigarrow \quad E^{(r)}(f) \le E^{(r)}(g) = O(1/r^2)$

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

(2) Minimizer in **interior**: Then, $f^{(r)} = \lambda_{\min}(M_r(f))$ 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: $\quad f(x) \leq g(x) := kx + 1 \quad \rightsquigarrow \quad E^{(r)}(f) \leq E^{(r)}(g) = O(1/r^2)$

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

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

$$M_r(f) = \left(\begin{array}{cc|ccccccc}
1/2 & \frac{k}{\sqrt{2}} & \frac{4}{\sqrt{2}} & & & & & & \\
\frac{k}{\sqrt{2}} & 3/4 & \frac{k}{2} & \frac{1}{4} & & & & & \\
\hline
\frac{4}{\sqrt{2}} & \frac{k}{2} & a & b & c & & & & \\
 & \frac{1}{4} & b & a & b & c & & & \\
 & & c & b & \ddots & \ddots & \ddots & & \\
 & & & c & \ddots & \ddots & \ddots & \ddots & \\
 & & & & \ddots & \ddots & \ddots & \ddots & c \\
 & & & & & \ddots & \ddots & \ddots & b \\
 & & & & & & c & b & a
\end{array}\right), \qquad a = \frac{1}{2},\; b = \frac{k}{2},\; c = \frac{1}{4}$$

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:

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

Theorem (de Klerk-L 2020)

For the **Chebyshev measure** $\prod_i (1-x_i^2)^{-1/2}$ *on* $[-1,1]^n$ *and for any polynomial* f*:*

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

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 2020)

For the **Chebyshev measure** $\prod_i (1-x_i^2)^{-1/2}$ *on* $[-1,1]^n$ *and for any polynomial* f*:*

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

Next: extend to the Jacobi measure $(1-x^2)^{\lambda}$ on $[-1,1]$ with $\lambda \ge -1/2$ and to other sets

Extension:
$O(\frac{1}{r^2})$ convergence rate for the sphere

using an integration trick

Key steps

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

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

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

Key steps

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

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

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

Key steps

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

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

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

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:

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

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 $\sigma(x_1)$ 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:

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

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

Key steps

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

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

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

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

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

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

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

Key steps

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

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

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

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

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

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

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

[de Klerk-L 2020]

Extension:

$O(\frac{1}{r^2})$ convergence rate for

box, ball, simplex, round convex body

using ‘local similarity’ trick

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

Lemma (Slot-L 2020)

Let $\mathbf{a} \in K$ 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)

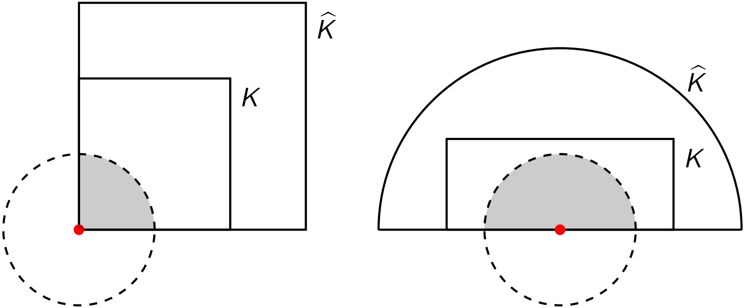
Lemma (Slot-L 2020)

Let $\mathbf{a} \in K$ *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***:*

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



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

Lemma (Slot-L 2020)

Let $\mathbf{a} \in K$ *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***:*

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

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

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

(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 $\mathbf{a} \in K$ *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***:*

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

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

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

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

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

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

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

Lemma (Slot-L 2020)

Let $\mathbf{a} \in K$ *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***:*

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

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

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

(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* $\mathbf{a}$*, satisfying*

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

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

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

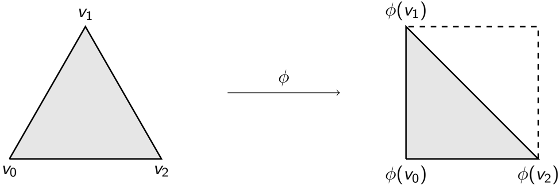
[using Chebyshev weight $\widehat{w}(x) = (1-x^2)^{-1/2}$], to $K = [-1,1]^n$

(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 $\widehat{K} = [-1,1]^n$ 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 $\widehat{K} = [-1,1]^n$ 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 $\widehat{w} = 1$]

Back to analyzing the lower bounds

for the unit sphere

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup\ \lambda\ \text{ s.t. } f(x) - \lambda = \sigma(x)\ \text{ on } \mathbb{S}^{n-1}, \text{ where } \sigma \text{ SoS}, \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup\ \lambda\ \text{ s.t. }\ f(x) - \lambda = \sigma(x)\ \text{ on } \mathbb{S}^{n-1}, \ \text{ where } \sigma \text{ SoS},\ \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Strategy: Construct a **'nice' polynomial kernel** $K(x, y)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

$\rightsquigarrow$ **kernel operator K** : $p \in \mathcal{P} \mapsto \mathbf{K}p(x) = \int_{\mathbb{S}^{n-1}} p(y) K(x, y) d\mu(y) \ \in \mathcal{P}$

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup \ \lambda \ \text{ s.t. } \ f(x) - \lambda = \sigma(x) \ \text{ on } \mathbb{S}^{n-1}, \ \text{ where } \sigma \text{ SoS}, \ \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Strategy: Construct a **'nice' polynomial kernel** $K(x, y)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

$\rightsquigarrow$ **kernel operator K** : $p \in \mathcal{P} \mapsto \mathbf{K}p(x) = \int_{\mathbb{S}^{n-1}} p(y) K(x, y) d\mu(y) \ \in \mathcal{P}$

(A1) $\mathbf{K}1 = 1$: $\int_{-1}^{1} K(x, y) d\mu(y) = 1 \quad \forall x \in \mathbb{S}^{n-1}$

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup\ \lambda \ \text{ s.t. } f(x) - \lambda = \sigma(x) \ \text{ on } \mathbb{S}^{n-1}, \ \text{ where } \sigma \text{ SoS}, \ \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Strategy: Construct a **'nice' polynomial kernel** $K(x,y)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

$\rightsquigarrow$ **kernel operator K** : $p \in \mathcal{P} \mapsto \mathbf{K}p(x) = \int_{\mathbb{S}^{n-1}} p(y)K(x,y)d\mu(y) \ \in \mathcal{P}$

(A1) $\mathbf{K}1 = 1$: $\quad \int_{-1}^{1} K(x,y)d\mu(y) = 1 \quad \forall x \in \mathbb{S}^{n-1}$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$: $\quad \mathbf{K}\mathcal{P}_d = \mathcal{P}_d$

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup \ \lambda \ \text{ s.t. } \ f(x) - \lambda = \sigma(x) \ \text{ on } \mathbb{S}^{n-1}, \ \text{ where } \sigma \text{ SoS}, \ \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Strategy: Construct a **'nice' polynomial kernel** $K(x,y)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

$\rightsquigarrow$ **kernel operator K** : $p \in \mathcal{P} \mapsto \mathbf{K}p(x) = \int_{\mathbb{S}^{n-1}} p(y) K(x,y) d\mu(y) \ \in \mathcal{P}$

(A1) $\mathbf{K}1 = 1$: $\quad \int_{-1}^{1} K(x,y) d\mu(y) = 1 \quad \forall x \in \mathbb{S}^{n-1}$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$: $\quad \mathbf{K}\mathcal{P}_d = \mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\quad \|\mathbf{K} - I\| := \sup_{p \in \mathcal{P}_d : \|p\|_\infty = 1} \|\mathbf{K}p - p\|_\infty \leq \epsilon$

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup \ \lambda \ \text{ s.t. } f(x) - \lambda = \sigma(x) \ \text{ on } \mathbb{S}^{n-1}, \ \text{ where } \sigma \text{ SoS}, \ \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Strategy: Construct a **'nice' polynomial kernel** $K(x,y)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

$\rightsquigarrow$ **kernel operator K** : $p \in \mathcal{P} \mapsto \mathbf{K}p(x) = \int_{\mathbb{S}^{n-1}} p(y) K(x,y) d\mu(y) \ \in \mathcal{P}$

(A1) $\mathbf{K}1 = 1$: $\quad \int_{-1}^{1} K(x,y) d\mu(y) = 1 \quad \forall x \in \mathbb{S}^{n-1}$

(A2) **K** preserves $\mathcal{P}_d$: $\quad \mathbf{K}\mathcal{P}_d = \mathcal{P}_d$

(A3) **K** close to I: $\quad \|\mathbf{K} - I\| := \sup_{p \in \mathcal{P}_d : \|p\|_\infty = 1} \|\mathbf{K}p - p\|_\infty \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\quad \|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup \ \lambda \ \text{ s.t. } \ f(x) - \lambda = \sigma(x) \ \text{ on } \mathbb{S}^{n-1}, \ \text{ where } \sigma \text{ SoS}, \ \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Strategy: Construct a **'nice' polynomial kernel** $K(x,y)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

$\rightsquigarrow$ **kernel operator K** : $p \in \mathcal{P} \mapsto \mathbf{K}p(x) = \int_{\mathbb{S}^{n-1}} p(y)K(x,y)d\mu(y) \ \in \mathcal{P}$

(A1) $\mathbf{K}1 = 1$: $\int_{-1}^{1} K(x,y)d\mu(y) = 1 \quad \forall x \in \mathbb{S}^{n-1}$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$: $\mathbf{K}\mathcal{P}_d = \mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| := \sup_{p \in \mathcal{P}_d : \|p\|_\infty = 1} \|\mathbf{K}p - p\|_\infty \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) for fixed $y \in \mathbb{S}^{n-1}$, $\mathbf{K}(x,y)$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Polynomial kernel approach for the unit sphere

Goal: Let $f \in \mathcal{P}_d$: polynomial of degree d on $\mathbb{S}^{n-1}$

$$f_{(r)} = \sup \ \lambda \ \text{ s.t. } f(x) - \lambda = \sigma(x) \ \text{ on } \mathbb{S}^{n-1}, \ \text{ where } \sigma \text{ SoS}, \ \deg(\sigma) \leq 2r$$

Theorem: [Fang-Fawzi 2020] $f_{\min} - f_{(r)} = O(\frac{1}{r^2})$

Strategy: Construct a **'nice' polynomial kernel** $K(x,y)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

$\rightsquigarrow$ **kernel operator K** : $p \in \mathcal{P} \mapsto \mathbf{K}p(x) = \int_{\mathbb{S}^{n-1}} p(y) K(x,y) d\mu(y) \ \in \mathcal{P}$

(A1) $\mathbf{K}1 = 1$: $\quad \int_{-1}^{1} K(x,y) d\mu(y) = 1 \quad \forall x \in \mathbb{S}^{n-1}$

(A2) **K** preserves $\mathcal{P}_d$: $\quad \mathbf{K}\mathcal{P}_d = \mathcal{P}_d$

(A3) **K** close to I: $\quad \|\mathbf{K} - I\| := \sup_{p \in \mathcal{P}_d : \|p\|_\infty = 1} \|\mathbf{K}p - p\|_\infty \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\quad \|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) for fixed $y \in \mathbb{S}^{n-1}$, $\mathbf{K}(x,y)$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

$\rightsquigarrow$ $p \geq 0$ on $\mathbb{S}^{n-1} \implies \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Deriving the analysis of $f_{(r)}$

(A1) $\mathbf{K}1 = 1$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) $p \geq 0$ on $\mathbb{S}^{n-1} \implies \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Theorem: $f_{\min} - f_{(r)} \leq 3\epsilon\,(f_{\max} - f_{\min})$

Deriving the analysis of $f_{(r)}$

(A1) $\mathbf{K}1 = 1$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) $p \geq 0$ on $\mathbb{S}^{n-1} \implies \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Theorem: $f_{\min} - f_{(r)} \leq 3\epsilon\,(f_{\max} - f_{\min})$

Proof: Wlog $f_{\min} = 0$, $f_{\max} = 1$, so $\|f\|_\infty = 1$.

Deriving the analysis of $f_{(r)}$

(A1) $\mathbf{K}1 = 1$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) $p \geq 0$ on $\mathbb{S}^{n-1} \Longrightarrow \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Theorem: $f_{\min} - f_{(r)} \leq 3\epsilon\,(f_{\max} - f_{\min})$

Proof: Wlog $f_{\min} = 0$, $f_{\max} = 1$, so $\|f\|_\infty = 1$.

By (A3): $\|(\mathbf{K}^{-1} - I)f\|_\infty \leq 3\epsilon$

Deriving the analysis of $f_{(r)}$

(A1) $\mathbf{K}1 = 1$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) $p \geq 0$ on $\mathbb{S}^{n-1} \implies \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Theorem: $f_{\min} - f_{(r)} \leq 3\epsilon \, (f_{\max} - f_{\min})$

Proof: Wlog $f_{\min} = 0$, $f_{\max} = 1$, so $\|f\|_\infty = 1$.

By (A3): $\|(\mathbf{K}^{-1} - I)f\|_\infty \leq 3\epsilon \implies \mathbf{K}^{-1}f - f \geq -3\epsilon$ on $\mathbb{S}^{n-1}$

Deriving the analysis of $f_{(r)}$

(A1) $\mathbf{K}1 = 1$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) $p \geq 0$ on $\mathbb{S}^{n-1} \implies \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Theorem: $f_{\min} - f_{(r)} \leq 3\epsilon\,(f_{\max} - f_{\min})$

Proof: Wlog $f_{\min} = 0$, $f_{\max} = 1$, so $\|f\|_\infty = 1$.

By (A3): $\|(\mathbf{K}^{-1} - I)f\|_\infty \leq 3\epsilon \implies \mathbf{K}^{-1}f - f \geq -3\epsilon$ on $\mathbb{S}^{n-1}$

$$\implies \mathbf{K}^{-1}f + 3\epsilon \geq f \geq 0 \text{ on } \mathbb{S}^{n-1}$$

Deriving the analysis of $f_{(r)}$

(A1) $\mathbf{K}1 = 1$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) $p \geq 0$ on $\mathbb{S}^{n-1} \implies \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Theorem: $f_{\min} - f_{(r)} \leq 3\epsilon\,(f_{\max} - f_{\min})$

Proof: Wlog $f_{\min} = 0$, $f_{\max} = 1$, so $\|f\|_\infty = 1$.

By (A3): $\|(\mathbf{K}^{-1} - I)f\|_\infty \leq 3\epsilon \implies \mathbf{K}^{-1}f - f \geq -3\epsilon$ on $\mathbb{S}^{n-1}$

$$\implies \mathbf{K}^{-1}f + 3\epsilon \geq f \geq 0 \text{ on } \mathbb{S}^{n-1}$$

By (A1), (A4): $f + 3\epsilon = \mathbf{K}(\mathbf{K}^{-1}f + 3\epsilon)$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Deriving the analysis of $f_{(r)}$

(A1) $\mathbf{K}1 = 1$

(A2) $\mathbf{K}$ preserves $\mathcal{P}_d$

(A3) $\mathbf{K}$ close to I: $\|\mathbf{K} - I\| \leq \epsilon$

$\rightsquigarrow$ $\mathbf{K}^{-1}$ close to I: $\|\mathbf{K}^{-1} - I\| \leq 3\epsilon$

(A4) $p \geq 0$ on $\mathbb{S}^{n-1} \implies \mathbf{K}p$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Theorem: $f_{\min} - f_{(r)} \leq 3\epsilon\,(f_{\max} - f_{\min})$

Proof: Wlog $f_{\min} = 0$, $f_{\max} = 1$, so $\|f\|_\infty = 1$.

By (A3): $\|(\mathbf{K}^{-1} - I)f\|_\infty \leq 3\epsilon \implies \mathbf{K}^{-1}f - f \geq -3\epsilon$ on $\mathbb{S}^{n-1}$

$\implies \mathbf{K}^{-1}f + 3\epsilon \geq f \geq 0$ on $\mathbb{S}^{n-1}$

By (A1), (A4): $f + 3\epsilon = \mathbf{K}(\mathbf{K}^{-1}f + 3\epsilon)$ is SoS with degree $2r$ on $\mathbb{S}^{n-1}$

Next: Construct such kernel $K(x, y)$ with $\epsilon = O(1/r^2)$ using *Fourier analysis* and reducing to the *upper bound approach* (in univariate case)

- Select $K(x, y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$

- Select $K(x, y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k \geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x, y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right)$

- Select $K(x,y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k\geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x,y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^{\mathsf{T}} y)$

- Select $K(x, y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k \geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x, y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^\mathsf{T} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\rightsquigarrow$ (A1), (A2) hold

- Select $K(x,y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k \geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x,y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^\mathsf{T} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\rightsquigarrow$ (A1), (A2) hold

- If $p \in \mathcal{P}_d$ with $p = \sum_{k=0}^{d} p_k$, then $\mathbf{K}p = \sum_{k=0}^{d} \lambda_k p_k$

- Select $K(x,y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k\geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x,y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^\mathsf{T} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\rightsquigarrow$ (A1), (A2) hold

- If $p \in \mathcal{P}_d$ with $p = \sum_{k=0}^{d} p_k$, then $\mathbf{K}p = \sum_{k=0}^{d} \lambda_k p_k$
- $\|\mathbf{K}p - p\|_\infty = \| \sum_{k=0}^{d} (\lambda_k - 1) p_k \|_\infty \leq \sum_{k=0}^{d} \|p_k\|_\infty |\lambda_k - 1|$

- Select $K(x, y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k\geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x, y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^\mathsf{T} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\rightsquigarrow$ (A1), (A2) hold

- If $p \in \mathcal{P}_d$ with $p = \sum_{k=0}^{d} p_k$, then $\mathbf{K}p = \sum_{k=0}^{d} \lambda_k p_k$
- $\|\mathbf{K}p - p\|_\infty = \|\sum_{k=0}^{d} (\lambda_k - 1) p_k\|_\infty \leq \sum_{k=0}^{d} \|p_k\|_\infty |\lambda_k - 1|$

$$\leq \sum_{k=0}^{d} |\lambda_k - 1| \cdot \|p\|_\infty C_d$$

- Select $K(x, y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k \geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x, y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^{\mathsf{T}} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\quad\rightsquigarrow$ (A1), (A2) hold

- If $p \in \mathcal{P}_d$ with $p = \sum_{k=0}^{d} p_k$, then $\mathbf{K}p = \sum_{k=0}^{d} \lambda_k p_k$
- $\|\mathbf{K}p - p\|_\infty = \| \sum_{k=0}^{d} (\lambda_k - 1) p_k \|_\infty \leq \sum_{k=0}^{d} \|p_k\|_\infty |\lambda_k - 1|$

$$\leq \sum_{k=0}^{d} |\lambda_k - 1| \cdot \|p\|_\infty C_d$$

So $\|\mathbf{K} - I\| \leq \sum_{k=1}^{d} |\lambda_k - 1| \cdot C_d$

- Select $K(x, y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k \geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x, y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^\mathsf{T} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\rightsquigarrow$ (A1), (A2) hold

- If $p \in \mathcal{P}_d$ with $p = \sum_{k=0}^{d} p_k$, then $\mathbf{K}p = \sum_{k=0}^{d} \lambda_k p_k$
- $\|\mathbf{K}p - p\|_\infty = \| \sum_{k=0}^{d} (\lambda_k - 1) p_k \|_\infty \leq \sum_{k=0}^{d} \|p_k\|_\infty |\lambda_k - 1|$

$$\leq \sum_{k=0}^{d} |\lambda_k - 1| \cdot \|p\|_\infty C_d$$

So $\|\mathbf{K} - I\| \leq \sum_{k=1}^{d} |\lambda_k - 1| \cdot C_d$

Therefore: It suffices to select $\lambda_0 = 1$, λ_k s.t.
(1) $\sum_{k=1}^{d} |1 - \lambda_k|$ $(= \epsilon)$ is small $\rightsquigarrow$ (A1), (A2), (A3) hold

- Select $K(x,y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k\geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x,y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x) e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^\mathsf{T} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\rightsquigarrow$ (A1), (A2) hold

- If $p \in \mathcal{P}_d$ with $p = \sum_{k=0}^{d} p_k$, then $\mathbf{K}p = \sum_{k=0}^{d} \lambda_k p_k$
- $\|\mathbf{K}p - p\|_\infty = \| \sum_{k=0}^{d} (\lambda_k - 1) p_k \|_\infty \leq \sum_{k=0}^{d} \|p_k\|_\infty |\lambda_k - 1|$
$\leq \sum_{k=0}^{d} |\lambda_k - 1| \cdot \|p\|_\infty C_d$

So $\|\mathbf{K} - I\| \leq \sum_{k=1}^{d} |\lambda_k - 1| \cdot C_d$

Therefore: It suffices to select $\lambda_0 = 1$, λ_k s.t.

(1) $\sum_{k=1}^{d} |1 - \lambda_k|$ $(= \epsilon)$ is small $\rightsquigarrow$ (A1), (A2), (A3) hold

(2) $K(x,y) = (q(x^\mathsf{T} y))^2$, where q univariate of degree r $\rightsquigarrow$ (A4) holds

- Select $K(x,y)$ invariant under action of $O(n) \times O(n)$ on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$
- Harmonic decomposition: $\mathcal{P} = \oplus_{k\geq 0} \mathrm{Harm}_k = \mathrm{span}(e_{k,i}) : i \in [h_k]\}$

Select: $K(x,y) = \sum_{k=0}^{2r} \lambda_k \left(\sum_{i=1}^{h_k} e_{ki}(x)e_{ki}(y)\right) = \sum_{k=0}^{2r} \lambda_k \; C_k^n(x^\mathsf{T} y)$
with $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_d \neq 0$ $\rightsquigarrow$ (A1), (A2) hold

- If $p \in \mathcal{P}_d$ with $p = \sum_{k=0}^{d} p_k$, then $\mathbf{K}p = \sum_{k=0}^{d} \lambda_k p_k$
- $\|\mathbf{K}p - p\|_\infty = \|\sum_{k=0}^{d}(\lambda_k - 1)p_k\|_\infty \leq \sum_{k=0}^{d} \|p_k\|_\infty |\lambda_k - 1|$
$\leq \sum_{k=0}^{d} |\lambda_k - 1| \cdot \|p\|_\infty C_d$

So $\|\mathbf{K} - I\| \leq \sum_{k=1}^{d} |\lambda_k - 1| \cdot C_d$

Therefore: It suffices to select $\lambda_0 = 1$, λ_k s.t.

(1) $\sum_{k=1}^{d} |1 - \lambda_k|$ $(= \epsilon)$ is small $\rightsquigarrow$ (A1), (A2), (A3) hold

(2) $K(x,y) = (q(x^\mathsf{T} y))^2$, where q univariate of degree r $\rightsquigarrow$ (A4) holds

Write: $q(t)^2 = \sum_{k=0}^{2r} \lambda_k C_k^n(t)$, where $C_k^n(t)$ are the Gegenbauer orthogonal polynomials on $[-1,1]$ w.r.t. $d\nu(t) = (1-t^2)^{(n-3)/2} dt$

Final step: select λ_k via the 'upper bound' approach

Want: $q(t)^2 = \sum_{k=0}^{2r} \lambda_k C_k^n(t)$, where $\lambda_0 = 1$ and $\sum_{k=1}^{d}(1-\lambda_k)$ is small
and $C_k^n(t)$ orthonormal polynomials on $[-1,1]$ for $d\nu(t)$

Final step: select λ_k via the 'upper bound' approach

Want: $q(t)^2 = \sum_{k=0}^{2r} \lambda_k C_k^n(t)$, where $\lambda_0 = 1$ and $\sum_{k=1}^{d}(1-\lambda_k)$ is small
and $C_k^n(t)$ orthonormal polynomials on $[-1,1]$ for $d\nu(t)$

Note: $\lambda_k = \int_{-1}^{1} q(t)^2 C_k^n(t) d\nu(t), \quad 1 = \lambda_0 = \int_{-1}^{1} q(t)^2 d\nu(t)$

Final step: select λ_k via the 'upper bound' approach

Want: $q(t)^2 = \sum_{k=0}^{2r} \lambda_k C_k^n(t)$, where $\lambda_0 = 1$ and $\sum_{k=1}^{d}(1-\lambda_k)$ is small and $C_k^n(t)$ orthonormal polynomials on $[-1,1]$ for $d\nu(t)$

Note: $\lambda_k = \int_{-1}^{1} q(t)^2 C_k^n(t) d\nu(t), \quad 1 = \lambda_0 = \int_{-1}^{1} q(t)^2 d\nu(t)$

Hence: $\sum_{k=1}^{d} 1 - \lambda_k = \int_{-1}^{1} q(t)^2 \underbrace{\Big(d - \sum_{k=1}^{d} C_k^n(t) \Big)}_{F(t)} d\nu(t)$

Final step: select λ_k via the 'upper bound' approach

Want: $q(t)^2 = \sum_{k=0}^{2r} \lambda_k C_k^n(t)$, where $\lambda_0 = 1$ and $\sum_{k=1}^{d}(1-\lambda_k)$ is small and $C_k^n(t)$ orthonormal polynomials on $[-1,1]$ for $d\nu(t)$

Note: $\lambda_k = \int_{-1}^{1} q(t)^2 C_k^n(t) d\nu(t), \quad 1 = \lambda_0 = \int_{-1}^{1} q(t)^2 d\nu(t)$

Hence: $\sum_{k=1}^{d} 1 - \lambda_k = \int_{-1}^{1} q(t)^2 \underbrace{\left(d - \sum_{k=1}^{d} C_k^n(t) \right)}_{F(t)} d\nu(t)$

So we arrive at the problem:

$$F^{(r)} = \min_{q \in \mathbb{R}[t]_r} \int_{-1}^{1} q(t)^2 \; F(t) \; d\nu(t) \text{ s.t. } \int_{-1}^{1} q(t)^2 d\nu(t) = 1$$

Final step: select λ_k via the 'upper bound' approach

Want: $q(t)^2 = \sum_{k=0}^{2r} \lambda_k C_k^n(t)$, where $\lambda_0 = 1$ and $\sum_{k=1}^{d}(1-\lambda_k)$ is small and $C_k^n(t)$ orthonormal polynomials on $[-1,1]$ for $d\nu(t)$

Note: $\lambda_k = \int_{-1}^{1} q(t)^2 C_k^n(t) d\nu(t), \quad 1 = \lambda_0 = \int_{-1}^{1} q(t)^2 d\nu(t)$

Hence: $\sum_{k=1}^{d} 1 - \lambda_k = \int_{-1}^{1} q(t)^2 \Big(\underbrace{d - \sum_{k=1}^{d} C_k^n(t)}_{F(t)} \Big) d\nu(t)$

So we arrive at the problem:

$$F^{(r)} = \min_{q \in \mathbb{R}[t]_r} \int_{-1}^{1} q(t)^2 \; F(t) \; d\nu(t) \text{ s.t. } \int_{-1}^{1} q(t)^2 d\nu(t) = 1$$

By the analysis for the *upper bounds* (univariate case): $F^{(r)} = O(1/r^2)$

Final step: select λ_k via the 'upper bound' approach

Want: $q(t)^2 = \sum_{k=0}^{2r} \lambda_k C_k^n(t)$, where $\lambda_0 = 1$ and $\sum_{k=1}^{d}(1-\lambda_k)$ is small and $C_k^n(t)$ orthonormal polynomials on $[-1,1]$ for $d\nu(t)$

Note: $\lambda_k = \int_{-1}^{1} q(t)^2 C_k^n(t) d\nu(t), \quad 1 = \lambda_0 = \int_{-1}^{1} q(t)^2 d\nu(t)$

Hence: $\sum_{k=1}^{d} 1 - \lambda_k = \int_{-1}^{1} q(t)^2 \underbrace{\Big(d - \sum_{k=1}^{d} C_k^n(t) \Big)}_{F(t)} d\nu(t)$

So we arrive at the problem:

$$F^{(r)} = \min_{q \in \mathbb{R}[t]_r} \int_{-1}^{1} q(t)^2 \; F(t) \; d\nu(t) \text{ s.t. } \int_{-1}^{1} q(t)^2 d\nu(t) = 1$$

By the analysis for the *upper bounds* (univariate case): $F^{(r)} = O(1/r^2)$

$\rightsquigarrow$ optimal q gives desired λ_k $\rightsquigarrow$ desired kernel $K(x,y)$

$\rightsquigarrow$ desired rate $O(1/r^2)$ for SoS lower bounds $f_{(r)}$

Concluding remarks

- Interesting interplay between the **lower** and **upper** bounds
- An analogous technique can be applied to analyse the **lower bounds** $f_{(r)}$ when minimizing f on the Boolean cube $\{0,1\}^n$

  [Slot-L 2021]
- For the box $[-1,1]^n$, one can derive an analysis in $O(1/r^2)$ for the **lower bounds** based on the *preordering* (instead of the *quadratic module*)
- Open question: Can one get an improved analysis for the **lower bounds** based on the quadratic module for the box, the ball, etc. ?
- The error analysis for the **upper bounds** $f^{(r)}$ extends to **rational** functions f [dK-L'19]

  and can be adapted to the **general problem of moments**

  [de Klerk-Postek-Kuhn'19]

Some references

- J. Nie and M. Schweighofer. On the complexity of Putinar's Positivstellensatz. J. of Complexity, 2007
- J.B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. SIAM Journal on Optimization, 2011
- E. de Klerk and M. Laurent. Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube. Math. of Operations Research, 2020
- – – Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. Math. Programming, 2020
- L. Slot and M. Laurent. Near-optimal analysis of univariate moment bounds for polynomial optimization. Math. Programming, 2020
- – – Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets. Math. Programming, 2020