Linking optimization with spectral analysis of 3-diagonal (univariate) moment matrices

Jean B. Lasserre*

LAAS-CNRS and Institute of Mathematics, Toulouse, France

SpecFun Seminar
INRIA, Paris, January 2020

⋆ Research funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 666981 TAMING)

Based on arXiv:1907.09784

Let :

- $\Omega \subset \mathbb{R}^n$ be a compact set,
- $f : \Omega \to \mathbb{R}$ be a continuous function,

and consider the optimization problem :

$$\Omega \;:=\; f_* \;=\; \min_{\mathbf{x}} \{ f(\mathbf{x}) \,:\, \mathbf{x} \in \Omega \}$$

Background : A converging hierarchy of upper bounds

• Let $\Sigma[\mathbf{x}]_t$ be the convex cone of Sum-of-Squares polynomials (SOS) of degree at most $2t$.

• Let λ be a Borel measure whose support is EXACTLY Ω, i.e., Ω is the smallest closed set such that $\lambda(\mathbb{R}^n \setminus \Omega) = 0$.

A converging hierarchy of UPPER BOUNDS

For every $t \in \mathbb{N}$, let

$$\rho_t \;:=\; \min_{\sigma} \left\{ \int_{\Omega} f\,\sigma\,d\lambda : \int_{\Omega} \sigma\,d\lambda \;=\; 1; \quad \sigma \in \Sigma[\mathbf{x}]_t \right\}$$

☞ $\rho_t \geq f_*$ because $\sigma\,d\lambda$ is a prob. measure on Ω, and so :

$$f \geq f^* \text{ on } \Omega \Rightarrow \int_{\Omega} f\,\sigma\,d\lambda \geq f_* \underbrace{\int_{\Omega} \sigma\,d\lambda}_{=1} = f_*.$$

Background : A converging hierarchy of upper bounds

• Let $\Sigma[\mathbf{x}]_t$ be the convex cone of Sum-of-Squares polynomials (SOS) of degree at most $2t$.

• Let λ be a Borel measure whose support is EXACTLY Ω, i.e., Ω is the smallest closed set such that $\lambda(\mathbb{R}^n \setminus \Omega) = 0$.

A converging hierarchy of UPPER BOUNDS

For every $t \in \mathbb{N}$, let

$$\rho_t := \min_{\sigma} \left\{ \int_\Omega f\,\sigma\,d\lambda : \int_\Omega \sigma\,d\lambda = 1; \quad \sigma \in \Sigma[\mathbf{x}]_t \right\}$$

☞ $\rho_t \geq f_*$ because $\sigma\,d\lambda$ is a prob. measure on Ω, and so :

$$f \geq f^* \text{ on } \Omega \Rightarrow \int_\Omega f\,\sigma\,d\lambda \geq f_* \underbrace{\int_\Omega \sigma\,d\lambda}_{=1} = f_*.$$

Hence $\rho_{t+1} \geq \rho_t \geq f_*$ for all $t \in \mathbb{N}$.

The dual reads :

$$\rho_t^* := \max_{\theta} \{ \mathbf{M}_t(f\,\lambda) \succeq \theta\, \mathbf{M}_t(\lambda) \}$$

where

- $\mathbf{M}_t(\lambda)$ is the MOMENT matrix of order t, associated with the measure λ
- $\mathbf{M}_t(f\,\lambda))$ is the LOCALIZING matrix of order t, associated with the measure λ and the function f.

☞ Computing ρ_t^* is solving a Generalized Eigenvalue Problem for the pair of matrices $(\mathbf{M}_t(\lambda), \mathbf{M}_t(f\,\lambda))$.

Hence $\rho_{t+1} \geq \rho_t \geq f_*$ for all $t \in \mathbb{N}$.

The dual reads :

$$\rho_t^* := \max_{\theta} \{ \mathbf{M}_t(f\,\lambda) \succeq \theta\, \mathbf{M}_t(\lambda) \}$$

where

- $\mathbf{M}_t(\lambda)$ is the MOMENT matrix of order t, associated with the measure λ
- $\mathbf{M}_t(f\,\lambda))$ is the LOCALIZING matrix of order t, associated with the measure λ and the function f.

☞ Computing ρ_t^* is solving a Generalized Eigenvalue Problem for the pair of matrices $(\mathbf{M}_t(\lambda), \mathbf{M}_t(f\,\lambda))$.

Hence $\rho_{t+1} \geq \rho_t \geq f_*$ for all $t \in \mathbb{N}$.

The dual reads :

$$\rho_t^* := \max_{\theta} \{ \mathbf{M}_t(f\,\lambda) \succeq \theta\, \mathbf{M}_t(\lambda) \}$$

where

- $\mathbf{M}_t(\lambda)$ is the MOMENT matrix of order t, associated with the measure λ
- $\mathbf{M}_t(f\,\lambda))$ is the LOCALIZING matrix of order t, associated with the measure λ and the function f.

☞ Computing ρ_t^* is solving a Generalized Eigenvalue Problem for the pair of matrices $(\mathbf{M}_t(\lambda), \mathbf{M}_t(f\,\lambda))$.

• The Moment matrix $\mathbf{M}_t(\lambda)$ associated with λ is real symmetric, with rows & columns indexed by $\alpha \in \mathbb{N}^t$, and with entries

$$\mathbf{M}_t(\lambda)(\alpha,\beta) \,=\, \int_\Omega \mathbf{x}^{\alpha+\beta}\,d\lambda, \quad \alpha,\beta \,\in\, \mathbb{N}^n_t$$

• The Localizing matrix $\mathbf{M}_t(\lambda)$ associated with λ and the function f is real symmetric, with rows & columns indexed by $\alpha \in \mathbb{N}^t$, and with entries

$$\mathbf{M}_t(f\,\lambda)(\alpha,\beta) \,=\, \int_\Omega f(\mathbf{x})\,\mathbf{x}^{\alpha+\beta}\,d\lambda, \quad \alpha,\beta \,\in\, \mathbb{N}^n_t$$

☞ If Ω is a “SIMPLE" set and f is a “POLYNOMI !AL" then ρ_t can be computed easily

• The Moment matrix $\mathbf{M}_t(\lambda)$ associated with λ is real symmetric, with rows & columns indexed by $\alpha \in \mathbb{N}^t$, and with entries

$$\mathbf{M}_t(\lambda)(\alpha,\beta) \;=\; \int_\Omega \mathbf{x}^{\alpha+\beta}\,d\lambda, \quad \alpha,\beta\,\in\,\mathbb{N}^n_t$$

• The Localizing matrix $\mathbf{M}_t(\lambda)$ associated with λ and the function f is real symmetric, with rows & columns indexed by $\alpha \in \mathbb{N}^t$, and with entries

$$\mathbf{M}_t(f\,\lambda)(\alpha,\beta) \;=\; \int_\Omega f(\mathbf{x})\,\mathbf{x}^{\alpha+\beta}\,d\lambda, \quad \alpha,\beta\,\in\,\mathbb{N}^n_t$$

☞ If Ω is a “SIMPLE" set and f is a “POLYNOMI !AL" then ρ_t can be computed easily

Illustrative Example

Let $n = 2$, $\mathbf{B} = [-1, 1]^2$, $f(\mathbf{x}) = x_1 x_2 + x_2^2$, and λ be the Lebesgue measure on $\mathbf{B}$. Then

$$\mathbf{M}_t(\lambda) = 4 \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} ; \mathbf{M}_1(f\,\lambda) = 4 \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{9} & \frac{1}{9} \\ 0 & \frac{1}{9} & \frac{1}{5} \end{bmatrix}.$$

Hence

$$\rho_1^* = \max\left\{\theta : \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{9} & \frac{1}{9} \\ 0 & \frac{1}{9} & \frac{1}{5} \end{bmatrix} \succeq \theta \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}\right\}$$

$$f^* = 0 \leq \rho_1^* \approx 0.22$$

☞ Typical examples of such "simple sets" Ω are :

- Box $[a,b]^n$ and Simplex $\{\mathbf{x} : \mathrm{e}^T\mathbf{x} \,\leq\, 1\}$,

- ellipsoid $\{\mathbf{x} : \mathbf{x}^T\mathbf{Q}\mathbf{x} \,\leq\, 1\}$ for $\mathbf{Q} \succ 0$, and sphere,

- Hypercube $\{-1,1\}^n$

- $\mathbb{R}^n$ (with λ the Gaussian measure), positive orthant $\mathbb{R}^n_+$ (with λ the exponential measure)

as well as their affine transformations.

Theorem (Lass (2011))

Let Ω be compact with nonempty interior. Then $\rho_t^* = \rho_t \geq f_*$ for all t. In addition the sequence $(\rho_t)_{t\in\mathbb{N}}$ is monotone decreasing and converges to f_*, that is, $\rho_t \downarrow f_*$ as $t \to \infty$.

☞ If $\mathbf{M}_t(\lambda)$ and $\mathbf{M}_t(f\,\lambda)$ are expressed in the basis of polynomials $(T_\alpha)_{\alpha\in\mathbb{N}^n}$ orthonormal w.r.t. λ, then :

$$\rho_t \;=\; \lambda_{\min}(\mathbf{M}_t(f\,\lambda)).$$

☞ However one still has to compute the smallest eigenvalue of a real symmetric matrix of size $\binom{n+t}{n}$

☞ Lass (2011) : A new look at nonnegativity on closed sets and polynomial optimization, SIAM J. Optim. 21, pp. 864–885.

Theorem (Lass (2011))

Let Ω be compact with nonempty interior. Then $\rho_t^* = \rho_t \geq f_*$ for all t. In addition the sequence $(\rho_t)_{t\in\mathbb{N}}$ is monotone decreasing and converges to f_*, that is, $\rho_t \downarrow f_*$ as $t \to \infty$.

☞ If $\mathbf{M}_t(\lambda)$ and $\mathbf{M}_t(f\,\lambda)$ are expressed in the basis of polynomials $(T_\alpha)_{\alpha\in\mathbb{N}^n}$ orthonormal w.r.t. λ, then :

$$\rho_t = \lambda_{\min}(\mathbf{M}_t(f\,\lambda)).$$

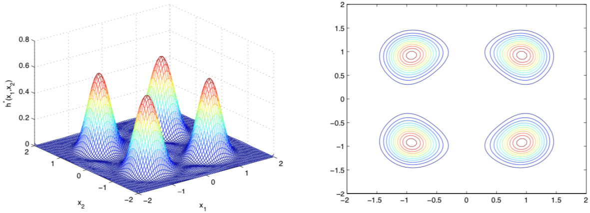
☞ However one still has to compute the smallest eigenvalue of a real symmetric matrix of size $\binom{n+t}{n}$

☞ Lass (2011) : A new look at nonnegativity on closed sets and polynomial optimization, SIAM J. Optim. 21, pp. 864–885.

As an illustrative example consider the bivariate Motzkin-like polynomial

$$\mathbf{x} \mapsto f(\mathbf{x}) \;:=\; x_1^3 4 x_2^2 + x_1^2 x_2^4 - 3\, x_1^2 x_2^2 + 1,$$

whig has 4 global minimizers. Below is the optimal SOS density σ^* of degree 24.



In a relatively recent series of papers E. De Klerk and M. Laurent (Netherlands) and collaborators have provided detailed analysis of the convergence $\rho_t \downarrow f_*$ as $t \to \infty$.

In a number of interesting cases where :

- Ω is a simple set (e.g., box, sphere), and
- λ is an appropriate well-known measure (Lebesgue, Chebyshev, rotation invariant, etc.)

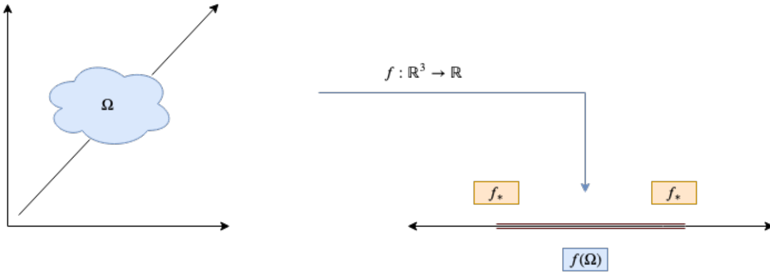
they could prove $O(1/t^2)$ rates of convergence.

☞ De Klerk, Laurent, Sun (2017) Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization, Math. Program. 162, 1, p. 363-392
☞ de Klerk, Laurent (2018) Worst-case examples for Lasserre's measure-based hierarchy for polynomial optimization on the hypercube, Math. Oper. Res.

A new approach via a simple transformation

Let the measure $\#\lambda$ on $\mathbb{R}$ be the pushforward of λ by the mapping $f : \Omega \to \mathbb{R}$. That is :

$$\#\lambda(B) \;=\; \lambda(f^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

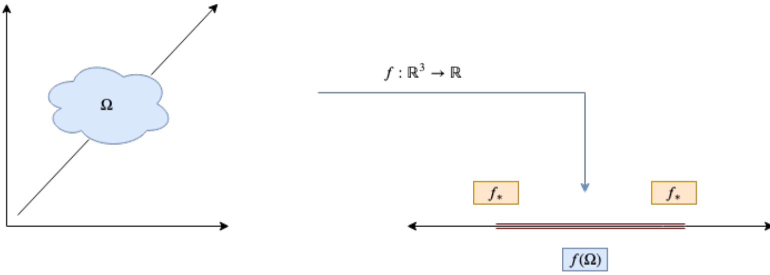


$f_* = \min \{ f(\mathbf{x}) : \mathbf{x} \in \Omega \}$ $\qquad$ $f^* = \max \{ f(\mathbf{x}) : \mathbf{x} \in \Omega \}$

A new approach via a simple transformation

Let the measure $\#\lambda$ on $\mathbb{R}$ be the pushforward of λ by the mapping $f : \Omega \to \mathbb{R}$. That is :

$$\#\lambda(B) \,=\, \lambda(f^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$



$f_* = \min\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$ $\qquad$ $f^* = \max\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$

All moments of $\#\lambda$ are obtained by :

$$\#\lambda_j := \int_{\mathbb{R}} z^j \, d\#\lambda(z) = \int_{\Omega} f(\mathbf{x})^j \, \lambda(d\mathbf{x}), \quad j \in \mathbb{N}$$

☞ If f is a POLYNOMIAL and Ω is a “SIMPLE" set, then the moments $(\#\lambda_j)_{j\in\mathbb{N}}$ are obtained in closed-form.

For instance Ω is a box, simplex, ellipsoid, ... and their affine transformations.

All moments of $\#\lambda$ are obtained by :

$$\#\lambda_j := \int_{\mathbb{R}} z^j \, d\#\lambda(z) = \int_{\Omega} f(\mathbf{x})^j \, \lambda(d\mathbf{x}), \quad j \in \mathbb{N}$$

☞ If f is a POLYNOMIAL and Ω is a “SIMPLE" set, then the moments $(\#\lambda_j)_{j\in\mathbb{N}}$ are obtained in closed-form.

For instance Ω is a box, simplex, ellipsoid, ... and their affine transformations.

A typical example : quadratic 0/1 problems

The 0/1 problem

$$\min \{ f(\mathbf{x}) : \quad \mathbf{A}\,\mathbf{x} = \mathbf{b}; \quad \mathbf{x} \in \{0,1\}^n \}$$

with $f \in \mathbb{R}[\mathbf{x}]_2$, and $\mathbf{A} \in \mathbf{Z}^{m \times n}$, $\mathbf{b} \in \mathbf{Z}^m$.

is exactly equivalent to the MAXCUT problem

$$\min \{ \tilde{f}(\mathbf{x}, x_0) : \quad (\mathbf{x}, x_0) \in \{-1,1\}^{n+1} \}$$

where $\tilde{f} \in \mathbb{R}[\mathbf{x}, x_0]_2$ is explicit in terms of $\mathbf{A}$ and $\mathbf{b}$.

☞ Lass (2016) : A MAX-CUT formulation of 0/1 programs, *Oper. Res. Letters 44*, pp. 158–164.

☞ So here the set $\Omega = \{-1,1\}^{n+1}$ is very simple !

A typical example : quadratic 0/1 problems

The 0/1 problem

$$\min \{ f(\mathbf{x}) : \quad \mathbf{A}\,\mathbf{x} \,=\, \mathbf{b}; \quad \mathbf{x} \in \{0,1\}^n \}$$

with $f \in \mathbb{R}[\mathbf{x}]_2$, and $\mathbf{A} \in \mathbf{Z}^{m\times n}$, $\mathbf{b} \in \mathbf{Z}^m$.

is exactly equivalent to the MAXCUT problem

$$\min \{ \tilde{f}(\mathbf{x}, x_0) : \quad (\mathbf{x}, x_0) \in \{-1,1\}^{n+1} \}$$

where $\tilde{f} \in \mathbb{R}[\mathbf{x}, x_0]_2$ is explicit in terms of $\mathbf{A}$ and $\mathbf{b}$.

☞ Lass (2016) : A MAX-CUT formulation of 0/1 programs, *Oper. Res. Letters 44*, pp. 158–164.

☞ So here the set $\Omega = \{-1,1\}^{n+1}$ is very simple !

A typical example : quadratic 0/1 problems

The 0/1 problem

$$\min \{ f(\mathbf{x}) : \quad \mathbf{A}\,\mathbf{x} \,=\, \mathbf{b}; \quad \mathbf{x} \in \{0,1\}^n \}$$

with $f \in \mathbb{R}[\mathbf{x}]_2$, and $\mathbf{A} \in \mathbf{Z}^{m \times n}$, $\mathbf{b} \in \mathbf{Z}^m$.

is exactly equivalent to the MAXCUT problem

$$\min \{ \tilde{f}(\mathbf{x}, x_0) : \quad (\mathbf{x}, x_0) \in \{-1,1\}^{n+1} \}$$

where $\tilde{f} \in \mathbb{R}[\mathbf{x}, x_0]_2$ is explicit in terms of $\mathbf{A}$ and $\mathbf{b}$.

☞ Lass (2016) : A MAX-CUT formulation of 0/1 programs, *Oper. Res. Letters 44*, pp. 158–164.

☞ So here the set $\Omega = \{-1,1\}^{n+1}$ is very simple !

Recall : $f_* = \min\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$ and $f^* = \max\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$

Key observation :

f_* (resp f^*) is the left (resp. right) endpoint of the support of $\#\lambda$. Equivalently :

$$\begin{aligned} f^* &= \max\{\mathbf{x} : \mathbf{x} \in \mathrm{supp}(\#\lambda)\} \\ f_* &= \min\{\mathbf{x} : \mathbf{x} \in \mathrm{supp}(\#\lambda)\} \end{aligned}$$

☞ Lass (2011) : Bounding the support a measure from its marginal moments. Proc. Amer. Math. Soc. 139, pp. 3375–3382.

☞ Hence on may apply the preceding approach to obtain a hierarchy of upper bounds $(\tau_t^\ell)_{t\in\mathbb{N}}$ on f_* (and lower bounds $(\tau_t^u)_{t\in\mathbb{N}}$ on f^*) BUT NOW ON A UNIVARIATE PROBLEM !

Recall : $f_* = \min\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$ and $f^* = \max\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$

Key observation :

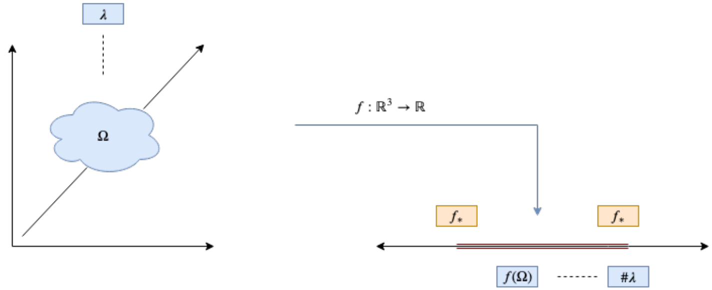
f_* (resp f^*) is the left (resp. right) endpoint of the support of $\#\lambda$.
Equivalently :

$$f^* = \max\{\mathbf{x} : \mathbf{x} \in \mathrm{supp}(\#\lambda)\}$$
$$f_* = \min\{\mathbf{x} : \mathbf{x} \in \mathrm{supp}(\#\lambda)\}$$

☞ Lass (2011) : Bounding the support a measure from its marginal moments. Proc. Amer. Math. Soc. 139, pp. 3375–3382.

☞ Hence on may apply the preceding approach to obtain a hierarchy of upper bounds $(\tau_t^\ell)_{t\in\mathbb{N}}$ on f_* (and lower bounds $(\tau_t^u)_{t\in\mathbb{N}}$ on f^*) BUT NOW ON A UNIVARIATE PROBLEM !

Illustration



Multivariate pb with λ $\rightarrow$ Univariate pb with $\#\lambda$

The sequence

$$\tau_t^\ell \;:=\; \max\{\,\theta:\; \mathbf{M}_t(x;\,\#\lambda)\,\succeq\,\theta\,\mathbf{M}_t(\#\lambda)\,\},\qquad t\in\mathbb{N}$$

is monotone decreasing and converges to f_* as $t\to\infty$.

The sequence

$$\tau_t^u \;:=\; \min\{\,\theta:\; \theta\,\mathbf{M}_t(\#\lambda)\,\succeq\,\mathbf{M}_t(x;\,\#\lambda)\,\},\qquad t\in\mathbb{N}$$

is monotone increasing and converges to f^* as $t\to\infty$.

Link with tri-diagonal Hankel matrices

Let $(T_j)_{j\in\mathbb{N}}$ be a basis of ORTHONORMAL (univariate) POLYNOMIALS w.r.t. the measure $\#\lambda$, that is :

$$\int T_i\,T_j\,d\#\lambda\;=\;\delta_{i=j},\qquad \forall i,j\in\mathbb{N}.$$

In this new basis, the moment matrix $\widehat{\mathbf{M}}_t(\#\lambda)$ is the $(t+1)\times(t+1)$ identity matrix I_t and therefore

$$\tau_t^{\ell}\;:=\;\max\left\{\theta:\;\widehat{\mathbf{M}}_t(x;\,\#\lambda)\;\succeq\;\theta\,\mathrm{I}_t,\right\}\;=\;\lambda_{\min}(\widehat{\mathbf{M}}_t(x;\,\#\lambda))$$

$$\tau_t^{u}\;:=\;\min\left\{\theta:\;\theta\,\mathrm{I}_t\;\succeq\;\widehat{\mathbf{M}}_t(x;\,\#\lambda)\right\}\;=\;\lambda_{\max}(\widehat{\mathbf{M}}_t(x;\,\#\lambda))$$

The polynomials $(T_j)_{j\in\mathbb{N}}$ obey the three-term recurrence

$$x\,T_j(x) \;=\; a_j\,T_{j+1}(x) + b_j\,T_j(x) + a_{j-1}\,T_{j-1}(x),$$

for all $x \in \mathbb{R}$ and $j \in \mathbb{N}$.

$$J = \begin{bmatrix} b_0 & a_0 & 0 & \cdots & \cdots & 0 \\ a_0 & b_1 & a_1 & 0 & \cdots & 0 \\ 0 & a_1 & b_2 & a_2 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

is called the *Jacobi matrix* associated with the orthonormal polynomials $(T_j)_{j\in\mathbb{N}}$;

Hence using the three-term recurrence relation :

$$\widehat{\mathbf{M}}_t(x;\, \#\lambda) = \begin{bmatrix} b_0 & a_0 & 0 & \cdots & \cdots & 0 \\ a_0 & b_1 & a_1 & 0 & \cdots & 0 \\ 0 & a_1 & b_2 & a_2 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & a_{t-1} & b_t \end{bmatrix}$$

is the t-truncation of the Jacobi matrix J.

and therefore :

☞ $\lambda_{\min}(\widehat{\mathbf{M}}_t(x;\, \#\lambda))$ is the smallest root of polynomial T_{t+1}.

☞ $\lambda_{\max}(\widehat{\mathbf{M}}_t(x;\, \#\lambda))$ is the largest root of polynomial T_{t+1}.

Take home message

The global minimum f_* (resp. maximum f^*) of a polynomial on $\Omega \subset \mathbb{R}^n$ can be approximated from above (resp. from below) and as closely as desired, by a sequence $(\tau_t^\ell)_{t\in\mathbb{N}} \downarrow f_*$ (resp. $(\tau_t^u)_{t\in\mathbb{N}} \uparrow f^*$)

- τ_t^ℓ is the smallest root of the univariate orthonormal polynomial T_{t+1}.
- τ_t^u is the largest root of the univariate orthonormal polynomial T_{t+1}.

However

Computing the polynomials $(T_j)_{j\in\mathbb{N}}$ requires computing moments $(\#\lambda_j)_{j\in\mathbb{N}}$ of the measure $\#\lambda$

☞ Ω needs to be simple enough (e.g., sphere, unit ball, unit box, simplex, etc.)

☞ can still be very tedious for large t

Another application of the pushforward

Let f be a nonnegative homogeneous polynomial, and let

$$\Omega = \{\, \mathbf{x} : f(\mathbf{x}) \,\leq\, 1 \,\} \,\subset\, \mathbf{B}, \text{ be compac.t}$$

Compute the Lebesgue volume

$$\rho \;=\; \mathrm{vol}(\Omega) \;=\; \int_{\Omega} d\mathbf{x}$$

... and possibly the moments

$$\rho_\alpha \;=\; \int_{\Omega} \mathbf{x}^\alpha \, d\mathbf{x}, \qquad \alpha \in \mathbb{N}^n,$$

of the Lebesgue measure on Ω

Another application of the pushforward

Let f be a nonnegative homogeneous polynomial, and let

$$\Omega = \{\, \mathbf{x} : f(\mathbf{x}) \,\leq\, 1 \,\} \,\subset\, \mathbf{B}, \text{ be compac.t}$$

Compute the Lebesgue volume

$$\rho \;=\; \mathrm{vol}(\Omega) \;=\; \int_{\Omega} d\mathbf{x}$$

... and possibly the moments

$$\rho_\alpha \;=\; \int_{\Omega} \mathbf{x}^\alpha \, d\mathbf{x}, \quad \alpha \in \mathbb{N}^n,$$

of the Lebesgue measure on Ω

Motivation

It turns out that :

$$\mathrm{vol}(\Omega) = \int_\Omega d\mathbf{x} = \frac{1}{\Gamma(1+n/d)} \int_\Omega \exp(-f(\mathbf{x}))\, d\mathbf{x}.$$

see e.g. Morozov & Shakirov, *Introduction to integral discriminants*, J. High Energy physics

I. ☞ $\int_\Omega \exp(-f(\mathbf{x}))\, d\mathbf{x}$, called an integral discriminant, is ubiquitous in statistical and quantum Physics.

II. From the above formula it follows that

☞ vol(Ω) is a strictly CONVEX function of the coefficients of the polynomial f.

☞ very useful for solving the following Problem $\mathbf{P}$:

$\mathbf{P}$: Compute nonnegative homogeneous polynomial f of degree $2d$ such that $\mathbf{K} \subset \Omega$ and Ω has minimum volume.

where $\mathbf{K} \subset \mathbb{R}^n$ is a given compact (not necessarily convex) set.

Theorem

☞ *Problem $\mathbf{P}$ is a CONVEX problem with a unique optimal solution f^**

II. From the above formula it follows that

☞ vol(Ω) is a strictly CONVEX function of the coefficients of the polynomial f.

☞ very useful for solving the following Problem **P** :

P : Compute nonnegative homogeneous polynomial f of degree $2d$ such that $\mathbf{K} \subset \Omega$ and Ω has minimum volume.

where $\mathbf{K} \subset \mathbb{R}^n$ is a given compact (not necessarily convex) set.

Theorem

☞ *Problem* **P** *is a CONVEX problem with a unique optimal solution* f^*

☞ $d = 2$ (quadratic case) : Ω_{f^*} is the celebrated Löwner-John ellipsoid

☞ However, given f, computing $\mathrm{vol}(\Omega_f)$ is difficult !

☞ $d = 2$ (quadratic case) : Ω_{f^*} is the celebrated Löwner-John ellipsoid

☞ However, given f, computing $\mathrm{vol}(\Omega_f)$ is difficult !

Computing $\mathrm{vol}(\Omega)$

Let λ be the Lebesgue probability measure on a box $\mathbf{B} \supset \Omega$.

General approach

(i) Either approximate $\mathrm{vol}(\Omega)$ by Monte-Carlo : λ-sample on $\mathbf{B}$ and COUNT points that fall into Ω. This provides a (random) estimate of $\mathrm{vol}(\Omega)$.

(ii) Or SOLVE† (or approximate)

$$\mathrm{vol}(\Omega) \;=\; \max_{\phi} \{\, \phi(\Omega) : \phi \,\leq\, \lambda \,\}$$

where the “max" is over measures ϕ supported on Ω.

☞ † Henrion D., Lasserre J.B., Savorgnan C. (2009) Approximate volume and integration for basic semi-algebraic sets. SIAM Review 51, pp. 722–743

(i) ☞ simple method that can handle potentially relatively large dimensions. On the other hand, it only provides a (random) estimate of $\mathrm{vol}(\Omega)$.

(ii) ☞ $\phi^* := \lambda_\Omega$ is the unique optimal solution and applying the Moment-SOS hierarchy provides a monotone sequence of upper bounds $(\rho_d)_{d\in\mathbb{N}} \downarrow \mathrm{vol}(\Omega)$ as $d \to \infty$.

• Additional linear constraints coming from Stokes' theorem applied to ϕ^* significantly accelerate the (otherwise slow) convergence.

• However, in view of the present status of SDP-solvers, this method is limited to problems of modest size.

(i) ☞ simple method that can handle potentially relatively large dimensions. On the other hand, it only provides a (random) estimate of vol(Ω).

(ii) ☞ $\phi^* := \lambda_\Omega$ is the unique optimal solution and applying the Moment-SOS hierarchy provides a monotone sequence of upper bounds $(\rho_d)_{d\in\mathbb{N}} \downarrow \mathrm{vol}(\Omega)$ as $d \to \infty$.

• Additional linear constraints coming from Stokes' theorem applied to ϕ^* significantly accelerate the (otherwise slow) convergence.

• However, in view of the present status of SDP-solvers, this method is limited to problems of modest size.

(i) ☞ simple method that can handle potentially relatively large dimensions. On the other hand, it only provides a (random) estimate of $\mathrm{vol}(\Omega)$.

(ii) ☞ $\phi^* := \lambda_\Omega$ is the unique optimal solution and applying the Moment-SOS hierarchy provides a monotone sequence of upper bounds $(\rho_d)_{d\in\mathbb{N}} \downarrow \mathrm{vol}(\Omega)$ as $d \to \infty$.

• Additional linear constraints coming from Stokes' theorem applied to ϕ^* significantly accelerate the (otherwise slow) convergence.

• However, in view of the present status of SDP-solvers, this method is limited to problems of modest size.

Stokes' theorem

With vector field $X = \mathbf{x}$, and $\alpha \in \mathbb{N}^n$ arbitrary :

$$\begin{aligned} 0 &= \int_\Omega \mathrm{Div}(X \cdot \mathbf{x}^\alpha (1 - f))\, dx = \int \mathrm{Div}(X \cdot \mathbf{x}^\alpha (1 - f))\, d\phi^* \\ &= \int \underbrace{\mathbf{x}^\alpha \; [(n + |\alpha|)\,(1 - f) - \langle \mathbf{x}, \nabla f \rangle]}_{p_\alpha(\mathbf{x})} \, d\phi^* \\ &= \int p_\alpha(\mathbf{x})\, d\phi^* \quad \text{a moment constraint on } \phi^* \end{aligned}$$

Hence one may equivalently solve :

$$\mathrm{vol}(\Omega) = \max_{\phi \in \mathscr{M}(\Omega)} \{ \phi(\Omega) : \phi \le \lambda; \quad \int p_\alpha \, d\phi = 0, \quad \alpha \in \mathbb{N}^n \}$$

☞ The associated relaxations of the Moment-SOS hierarchy converge much faster !

Stokes' theorem

With vector field $X = \mathbf{x}$, and $\alpha \in \mathbb{N}^n$ arbitrary :

$$\begin{aligned} 0 &= \int_{\Omega} \mathrm{Div}(X \cdot \mathbf{x}^{\alpha}(1-f))\,dx = \int \mathrm{Div}(X \cdot \mathbf{x}^{\alpha}(1-f))\,d\phi^{*} \\ &= \int \underbrace{\mathbf{x}^{\alpha}\;[(n+|\alpha|)\,(1-f) - \langle \mathbf{x}, \nabla f\rangle]}_{p_{\alpha}(\mathbf{x})}\,d\phi^{*} \\ &= \int p_{\alpha}(\mathbf{x})\,d\phi^{*} \qquad \text{a moment constraint on } \phi^{*} \end{aligned}$$

Hence one may equivalently solve :

$$\mathrm{vol}(\Omega) = \max_{\phi \in \mathscr{M}(\Omega)} \{\, \phi(\Omega) : \phi \leq \lambda; \quad \int p_{\alpha}\,d\phi = 0, \quad \alpha \in \mathbb{N}^n \}$$

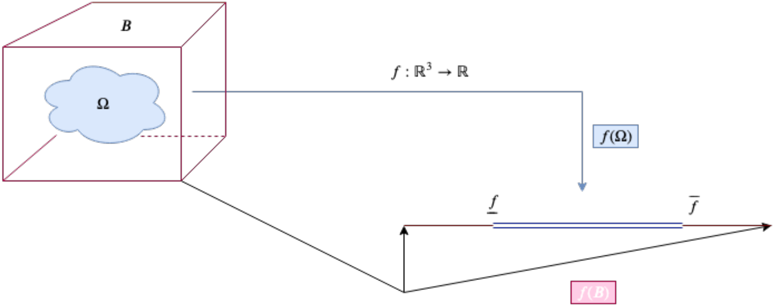
☞ The associated relaxations of the Moment-SOS hierarchy converge much faster !

Another approach via the pushforward

Let the measure $\#\lambda$ on $\mathbb{R}$ be the pushforward of λ by the mapping $f : \mathbf{B} \to \mathbb{R}$.

That is :

$$\#\lambda(B) \;=\; \lambda(f^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$


B
Ω
$f:\mathbb{R}^3\to\mathbb{R}$
$f(\Omega)$
$\underline{f}$
$\overline{f}$
$f(B)$

Let $I := f(\mathbf{B}) \subset \mathbb{R}$. Notice that :

All moments γ_k of $\#\lambda$ are obtained in closed form. That is :

$$\gamma_k \;:=\; \int_I z^k \, d\#\lambda(z) \;=\; \int_{\mathbf{B}} f(\mathbf{x})^k \, \lambda(d\mathbf{x}), \quad k = 0, 1, \ldots$$

Next, observe that

$$f(\Omega) \;=\; \{z \in I : \; 0 \leq \; z \leq 1\,\}.$$

Then :

$$\#\lambda([0,1]) = \int_{0\leq z\leq 1} \#\lambda(dz) = \lambda(f^{-1}([0,1])) = \lambda(\Omega)$$

That is, computing the n-dimensional volume ρ is computing the one-dimensional measure of the interval $[0,1]$ for the measure $\#\lambda$ on $\mathbb{R}$ ...

☞ Therefore Jasour et al.[†] et al. suggest to solve :

$$\rho = \max_{\phi} \{ \phi([0,1]) : \phi \leq \#\lambda; \ \mathrm{supp}(\phi) = [0,1] \}$$

Indeed $\phi^* = 1_{[0,1]}(z)\, d\#\lambda(z)$ is the unique optimal solution.

† A. Jasour, A. Hofmann, and B.C. Williams. *Moment-Sum-Of-Squares Approach For Fast Risk Estimation In Uncertain Environments*, arXiv:1810.01577, 2018.

Then :

$$\#\lambda([0,1]) = \int_{0\leq z\leq 1} \#\lambda(dz) = \lambda(f^{-1}([0,1])) = \lambda(\Omega)$$

That is, computing the n-dimensional volume ρ is computing the one-dimensional measure of the interval $[0,1]$ for the measure $\#\lambda$ on $\mathbb{R}$ . . .

☞ Therefore Jasour et al.[†] et al. suggest to solve :

$$\rho = \max_{\phi} \{\phi([0,1]) : \phi \leq \#\lambda;\ \mathrm{supp}(\phi) = [0,1]\}$$

Indeed $\phi^* = 1_{[0,1]}(z)\, d\#\lambda(z)$ is the unique optimal solution.

† A. Jasour, A. Hofmann, and B.C. Williams. *Moment-Sum-Of-Squares Approach For Fast Risk Estimation In Uncertain Environments*, arXiv:1810.01577, 2018.

Hence

☞ One has replaced computation of the n-dimensional Lebesgue-volume of Ω by computation of the 1-dimensional $\#\lambda$-volume of the interval $[0, 1]$

The value ρ can be approximated as closely as desired by solving appropriate SDP relaxations associated with the Moment-SOS hierarchy.

However ...

☞ Convergence $(\rho_d)_{d\in\mathbb{N}} \downarrow \rho$ is typically VERY SLOW !

☞ One cannot use Stokes constraints because one does not know the density of $\#\lambda$.

Hence

☞ One has replaced computation of the n-dimensional Lebesgue-volume of Ω by computation of the 1-dimensional $\#\lambda$-volume of the interval $[0, 1]$

The value ρ can be approximated as closely as desired by solving appropriate SDP relaxations associated with the Moment-SOS hierarchy.

However ...

☞ Convergence $(\rho_d)_{d\in\mathbb{N}} \downarrow \rho$ is typically VERY SLOW !

☞ One cannot use Stokes constraints because one does not know the density of $\#\lambda$.

Hence

☞ One has replaced computation of the n-dimensional Lebesgue-volume of Ω by computation of the 1-dimensional $\#\lambda$-volume of the interval $[0, 1]$

The value ρ can be approximated as closely as desired by solving appropriate SDP relaxations associated with the Moment-SOS hierarchy.

However ...

☞ Convergence $(\rho_d)_{d\in\mathbb{N}} \downarrow \rho$ is typically VERY SLOW !

☞ One cannot use Stokes constraints because one does not know the density of $\#\lambda$.

The homogeneous case

Take home message :

When f is homogeneous then one can do much better !

$$\text{Let} \quad \phi_j^* = \int_{[0,1]} z^j \, d\#\lambda(z), \quad j = 0, 1, \ldots$$

so that $\rho = \lambda(\Omega) = \phi_0^*$.

The homogeneous case

Take home message :

When f is homogeneous then one can do much better !

$$\text{Let} \quad \phi_j^* = \int_{[0,1]} z^j \, d\#\lambda(z), \quad j = 0, 1, \ldots$$

so that $\rho = \lambda(\Omega) = \phi_0^*$.

Suppose that f is NONNEGATIVE and HOMOGENEOUS of degree t. Then by Stokes' Theorem with vector field $X = \mathbf{x}$:

$$
\begin{aligned}
0 &= \int_{\Omega} \left[n\,(1 - f(\mathbf{x})^j) + \langle \mathbf{x}, \nabla (1 - f(\mathbf{x})^j) \rangle \right] \, d\lambda(\mathbf{x}) \\
&= n\,\lambda(\Omega) - (n + jt) \, \int_{\Omega} f(\mathbf{x})^j \, d\lambda(\mathbf{x}) \\
&= n\,\lambda(\Omega) - (n + jt) \, \int_{f(\Omega)} z^j \, d\#\lambda(z) \\
&= n\,\phi_0^* - (n + jt)\,\phi_j^*, \quad j = 1, 2, \ldots
\end{aligned}
$$

Theorem

Let $(\phi_j^*)_{j\in\mathbb{N}}$ be the moments of ϕ^*. Then :

$$\phi_j^* \,=\, \frac{n}{n+j\,t}\,\phi_0^*,\quad j=1,2,\ldots$$

As a consequence the moment matrix $H_d(\phi^*)$ of ϕ^*, is just $\phi_0^*\,H_d^*$ with :

$$H_d^* \,=\, \left[\begin{array}{cccc} 1 & \frac{n}{n+t} & \cdots & \frac{n}{n+d\,t} \\ \frac{n}{n+t} & \cdots & \cdots & \frac{n}{n+(d+1)\,t} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{n}{n+d\,t} & \cdots & \cdots & \frac{n}{n+2d\,t} \end{array}\right]$$

☞ which is the moment matrix of the probability measure

$$d\gamma(x) \,=\, \frac{n}{t}\,x^{\frac{n}{t}-1}\,dx \quad \text{on } [0,1]$$

Theorem

Let $(\phi_j^*)_{j\in\mathbb{N}}$ be the moments of ϕ^*. Then :

$$\phi_j^* \;=\; \frac{n}{n+j\,t}\,\phi_0^*,\quad j=1,2,\ldots$$

As a consequence the moment matrix $H_d(\phi^*)$ of ϕ^*, is just $\phi_0^*\,H_d^*$ with :

$$H_d^* \;=\; \left[\begin{array}{cccc} 1 & \frac{n}{n+t} & \cdots & \frac{n}{n+d\,t} \\ \frac{n}{n+t} & \cdots & \cdots & \frac{n}{n+(d+1)\,t} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{n}{n+d\,t} & \cdots & \cdots & \frac{n}{n+2d\,t} \end{array}\right]$$

☞ which is the moment matrix of the probability measure

$$d\gamma(x) \;=\; \frac{n}{t}\,x^{\frac{n}{t}-1}\,dx \quad \text{on } [0,1]$$

But then :

$$\rho = \max_{\phi} \{ \phi(\mathbb{R}) : \phi \leq \#\lambda; \; \mathrm{supp}(\phi) = [0,1] \}$$

can be approximated as closely as desired by

$$\begin{aligned} \tau_d &= \max_{\theta} \{ \theta : \theta\, H_d^* \preceq H_d(\#\lambda) \} \\ &= \lambda_{\min}(H_d(\#\lambda), H_d^*) \end{aligned}$$

a GENERALIZED EIGENVALUE PROBLEM associated with two HANKEL moment matrices.

Theorem

$\tau_d \downarrow \rho$ *as* $d \to \infty$.

But then :

$$\rho = \max_{\phi} \{ \phi(\mathbb{R}) : \phi \leq \#\lambda; \; \mathrm{supp}(\phi) = [0,1] \}$$

can be approximated as closely as desired by

$$\begin{aligned} \tau_d &= \max_{\theta} \{ \theta : \theta H_d^* \preceq H_d(\#\lambda) \} \\ &= \lambda_{\min}(H_d(\#\lambda), H_d^*) \end{aligned}$$

a GENERALIZED EIGENVALUE PROBLEM associated with two HANKEL moment matrices.

Theorem

$\tau_d \downarrow \rho$ as $d \to \infty$.

To visualize & appreciate the simplicity of the approach, let $n=2$ and $f(x)=\|\mathbf{x}\|^2=x_1^2+x_2^2$, and $\mathbf{B}=[-1,1]^2$, so that $\mathrm{vol}(\Omega)=\pi$. Then :

$$H_1^* = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}; \quad H_1(\#\lambda) = \begin{bmatrix} 1 & 2/3 \\ 2/3 & 28/45 \end{bmatrix}$$

This yields $4\cdot\tau_1 \approx 3.20$ which is already a good upper bound on π whereas $4\cdot\rho_1=4$.

$$H_2^* = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}; H_2(\#\lambda) = \begin{bmatrix} 1 & \frac{2}{3} & \frac{28}{45} \\ \frac{2}{3} & \frac{28}{45} & \frac{24}{35} \\ \frac{28}{45} & \frac{24}{35} & \frac{2}{9}+\frac{8}{21}+\frac{6}{25} \end{bmatrix}$$

This yields $4\cdot\tau_2 \approx 3.1440$ while $4\cdot\rho_2 = 3.8928$. Hence $4\cdot\tau_2$ already provides a very good upper bound on π with only moments of order 4.

To visualize & appreciate the simplicity of the approach, let $n = 2$ and $f(x) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$, and $\mathbf{B} = [-1, 1]^2$, so that $\mathrm{vol}(\Omega) = \pi$. Then :

$$H_1^* = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}; \quad H_1(\#\lambda) = \begin{bmatrix} 1 & 2/3 \\ 2/3 & 28/45 \end{bmatrix}$$

This yields $4 \cdot \tau_1 \approx 3.20$ which is already a good upper bound on π whereas $4 \cdot \rho_1 = 4$.

$$H_2^* = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}; H_2(\#\lambda) = \begin{bmatrix} 1 & \frac{2}{3} & \frac{28}{45} \\ \frac{2}{3} & \frac{28}{45} & \frac{24}{35} \\ \frac{28}{45} & \frac{24}{35} & \frac{2}{9} + \frac{8}{21} + \frac{6}{25} \end{bmatrix}$$

This yields $4 \cdot \tau_2 \approx 3.1440$ while $4 \cdot \rho_2 = 3.8928$. Hence $4 \cdot \tau_2$ already provides a very good upper bound on π with only moments of order 4.

d	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
ρ_d	12.19	11.075	9.163	8.878	8.499
τ_d	6.839	5.309	5.001	4.945	4.936

TABLE – $n = 4$, $\rho = 4.9348$; ρ_d versus τ_d

d	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$
$2^n\tau_d$	7.97	5.569	4.639	4.272	4.133	4.083
$\frac{(2^n\tau_d-\rho^*)}{\rho^*}$	96%	37%	14%	5.26%	1.83%	0.60%

TABLE – $n = 8$, $\rho = 4.0587$; τ_d and relative error

THANK YOU !