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:

$$\Omega := f_* = \min_{x} \{ f(x) : x \in \Omega \}$$
Background: A converging hierarchy of upper bounds

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

- Let $\lambda$ be a Borel measure whose support is EXACTLY $\Omega$, i.e., $\Omega$ 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

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

$\rho_t \geq f^*$ because $\sigma d\lambda$ is a prob. measure on $\Omega$, and so:

$$f \geq f^* \text{ on } \Omega \Rightarrow \int_{\Omega} f \sigma d\lambda \geq f^* \int_{\Omega} \sigma d\lambda = f^*.$$
Background: A converging hierarchy of upper bounds

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

- Let $\lambda$ be a Borel measure whose support is EXACTLY $\Omega$, i.e., $\Omega$ 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

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

$\rho_t \geq f^*$ because $\sigma \, d\lambda$ is a prob. measure on $\Omega$, and so:

$$f \geq f^* \text{ on } \Omega \Rightarrow \int_{\Omega} f \sigma \, d\lambda \geq f^* \int_{\Omega} \sigma \, d\lambda = f^*. \quad \text{[=1]}$$
Hence $\rho_{t+1} \geq \rho_t \geq f^*$ for all $t \in \mathbb{N}$.

The dual reads:

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

where

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

Computing $\rho_t^*$ 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:

$$\rho_t^* := \max_\theta \{ M_t(f \lambda) \succeq \theta M_t(\lambda) \}$$

where

- $M_t(\lambda)$ is the MOMENT matrix of order $t$, associated with the measure $\lambda$
- $M_t(f \lambda)$) is the LOCALIZING matrix of order $t$, associated with the measure $\lambda$ and the function $f$.

Computing $\rho^*_t$ is solving a Generalized Eigenvalue Problem for the pair of matrices $(M_t(\lambda), M_t(f \lambda))$. 
Hence $\rho_{t+1} \geq \rho_t \geq f^*$ for all $t \in \mathbb{N}$.

The dual reads:

$$\rho_t^* := \max_{\theta} \{ M_t(f \lambda) \succeq \theta M_t(\lambda) \}$$

where

- $M_t(\lambda)$ is the **MOMENT** matrix of order $t$, associated with the measure $\lambda$
- $M_t(f \lambda)$ is the **LOCALIZING** matrix of order $t$, associated with the measure $\lambda$ and the function $f$.

Computing $\rho_t^*$ is solving a **Generalized Eigenvalue Problem** for the pair of matrices $(M_t(\lambda), M_t(f \lambda))$. 
• The **Moment** matrix $M_t(\lambda)$ associated with $\lambda$ is real symmetric, with rows & columns indexed by $\alpha \in \mathbb{N}^t$, and with entries

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

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

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

If $\Omega$ is a “SIMPLE” set and $f$ is a “POLYNOMIAL” then $\rho_t$ can be computed easily.
• The Moment matrix $M_t(\lambda)$ associated with $\lambda$ is real symmetric, with rows & columns indexed by $\alpha \in \mathbb{N}^t$, and with entries

$$M_t(\lambda)(\alpha, \beta) = \int_{\Omega} x^{\alpha+\beta} d\lambda, \quad \alpha, \beta \in \mathbb{N}_n^t$$

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

$$M_t(f \lambda)(\alpha, \beta) = \int_{\Omega} f(x) x^{\alpha+\beta} d\lambda, \quad \alpha, \beta \in \mathbb{N}_n^t$$

✍️ If $\Omega$ is a "SIMPLE" set and $f$ is a "POLYNOMIAL" then $\rho_t$ can be computed easily
Illustrative Example

Let \( n = 2, B = [-1, 1]^2 \), \( f(x) = x_1x_2 + x_2^2 \), and \( \lambda \) be the Lebesgue measure on \( B \). Then

\[
M_t(\lambda) = 4 \begin{bmatrix}
1 & 0 & 0 \\
0 & \frac{1}{3} & 0 \\
0 & 0 & \frac{1}{3}
\end{bmatrix};
M_1(f\lambda) = 4 \begin{bmatrix}
\frac{1}{3} & 0 & 0 \\
0 & \frac{1}{9} & \frac{1}{9} \\
0 & \frac{1}{9} & \frac{1}{5}
\end{bmatrix}.
\]

Hence

\[
\rho_1^* = \max \{ \theta : \begin{bmatrix}
\frac{1}{3} & 0 & 0 \\
0 & \frac{1}{9} & \frac{1}{9} \\
0 & \frac{1}{9} & \frac{1}{5}
\end{bmatrix} \leq \theta \begin{bmatrix}
1 & 0 & 0 \\
0 & \frac{1}{3} & 0 \\
0 & 0 & \frac{1}{3}
\end{bmatrix}\}.
\]

\[
f^* = 0 \leq \rho_1^* \approx 0.22
\]
Typical examples of such "simple sets" $\Omega$ are:
- Box $[a, b]^n$ and Simplex $\{x : e^Tx \leq 1\}$,
- ellipsoid $\{x : x^TQx \leq 1\}$ for $Q \succ 0$, and sphere,
- Hypercube $\{-1, 1\}^n$.
- $\mathbb{R}^n$ (with $\lambda$ the Gaussian measure), positive orthant $\mathbb{R}^n_+$ (with $\lambda$ the exponential measure)

as well as their affine transformations.
**Theorem (Lass (2011))**

Let $\Omega$ 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 $M_t(\lambda)$ and $M_t(f \lambda)$ are expressed in the basis of polynomials $(T_\alpha)_{\alpha \in \mathbb{N}^n}$ orthonormal w.r.t. $\lambda$, then:

$$\rho_t = \lambda_{\min}(M_t(f \lambda)).$$

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

Theorem (Lass (2011))

Let $\Omega$ 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 $M_t(\lambda)$ and $M_t(f \lambda)$ are expressed in the basis of polynomials $(T_\alpha)_{\alpha \in \mathbb{N}^n}$ orthonormal w.r.t. $\lambda$, then:

$$\rho_t = \lambda_{\text{min}}(M_t(f \lambda)).$$

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

As an illustrative example consider the bivariate Motzkin-like polynomial

\[ f(x) := x_1^3 4x_2^2 + x_1^2 x_2^4 - 3 x_1^2 x_2^2 + 1, \]

which has 4 global minimizers. Below is the optimal SOS density \( \sigma^* \) 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:

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

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

A new approach via a simple transformation

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

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

$$f_* = \min \{ f(x) : x \in \Omega \}$$

$$f^* = \max \{ f(x) : x \in \Omega \}$$
A new approach via a simple transformation

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

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

\[
f_* = \min \{ f(x) : x \in \Omega \}
\]

\[
f^* = \max \{ f(x) : x \in \Omega \}
\]
All moments of \( \#\lambda \) are obtained by:

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

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

For instance \( \Omega \) is a box, simplex, ellipsoid, ... and their affine transformations.
All moments of $\#\lambda$ are obtained by:

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

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

For instance $\Omega$ is a box, simplex, ellipsoid, ... and their affine transformations.
A typical example: quadratic 0/1 problems

The 0/1 problem

\[
\begin{align*}
\min \{ f(x) : \quad & Ax = b; \quad x \in \{0, 1\}^n \} \\
\text{with} \quad & f \in \mathbb{R}[x]_2, \text{ and } A \in \mathbb{Z}^{m \times n}, b \in \mathbb{Z}^m.
\end{align*}
\]

is exactly equivalent to the MAXCUT problem

\[
\begin{align*}
\min \{ \tilde{f}(x, x_0) : \quad & (x, x_0) \in \{-1, 1\}^{n+1} \} \\
\text{where} \quad & \tilde{f} \in \mathbb{R}[x, x_0]_2 \text{ is explicit in terms of } A \text{ and } b.
\end{align*}
\]


So here the set \( \Omega = \{-1, 1\}^{n+1} \) is very simple!
A typical example: quadratic 0/1 problems

### The 0/1 problem

\[
\min \left\{ f(x) : \quad Ax = b; \quad x \in \{0, 1\}^n \right\}
\]

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

### is exactly equivalent to the MAXCUT problem

\[
\min \left\{ \tilde{f}(x, x_0) : \quad (x, x_0) \in \{-1, 1\}^{n+1} \right\}
\]

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

---


So here the set \( \Omega = \{-1, 1\}^{n+1} \) is very simple!
A typical example: quadratic 0/1 problems

The 0/1 problem

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

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

is exactly equivalent to the MAXCUT problem

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

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


So here the set $\Omega = \{-1, 1\}^{n+1}$ is very simple!
Recall: $f_* = \min \{ f(x) : x \in \Omega \}$ and $f^* = \max \{ f(x) : x \in \Omega \}$

Key observation:

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

Equivalently:

$$f^* = \max \{ x : x \in \text{supp}(\#\lambda) \}$$

$$f_* = \min \{ x : x \in \text{supp}(\#\lambda) \}$$


Hence on may apply the preceding approach to obtain a hierarchy of upper bounds $(\tau^\ell_t)_{t \in \mathbb{N}}$ on $f_*$ (and lower bounds $(\tau^u_t)_{t \in \mathbb{N}}$ on $f^*$) BUT NOW ON A UNIVARIATE PROBLEM!
Recall: $f_\star = \min \{ f(x) : x \in \Omega \}$ and $f^\star = \max \{ f(x) : x \in \Omega \}$

**Key observation:**

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

$$f^\star = \max \{ x : x \in \text{supp}(\# \lambda) \}$$

$$f_\star = \min \{ x : x \in \text{supp}(\# \lambda) \}$$


Hence on may apply the preceding approach to obtain a hierarchy of upper bounds $(\tau^\ell_t)_{t \in \mathbb{N}}$ on $f_\star$ (and lower bounds $(\tau^u_t)_{t \in \mathbb{N}}$ on $f^\star$) BUT NOW ON A UNIVARIATE PROBLEM!
Illustration

Multivariate pb with $\lambda$ $\rightarrow$ Univariate pb with $\#\lambda$
The sequence

$$\tau^l_t := \max \{ \theta : M_t(x; \#\lambda) \succeq \theta M_t(\#\lambda) \}, \quad t \in \mathbb{N}$$

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

The sequence

$$\tau^u_t := \min \{ \theta : \theta M_t(\#\lambda) \succeq M_t(x; \#\lambda) \}, \quad t \in \mathbb{N}$$

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:

\[
\int T_i T_j \, d\#\lambda = \delta_{i=j}, \quad \forall i, j \in \mathbb{N}.
\]

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

\[
\tau^\ell_t := \max \{ \theta : \hat{M}_t(x; \#\lambda) \succeq \theta I_t, \} = \lambda_{\min}(\hat{M}_t(x; \#\lambda))
\]

\[
\tau^\mu_t := \min \{ \theta : \theta I_t \succeq \hat{M}_t(x; \#\lambda) \} = \lambda_{\max}(\hat{M}_t(x; \#\lambda))
\]
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 \textit{Jacobi matrix} associated with the orthonormal polynomials \((T_j)_{j \in \mathbb{N}}\);
Hence using the three-term recurrence relation:

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

is the \( t \)-truncation of the Jacobi matrix \( J \).

and therefore:

- \( \lambda_{\text{min}}(\hat{M}_t(x; \# \lambda)) \) is the smallest root of polynomial \( T_{t+1} \).
- \( \lambda_{\text{max}}(\hat{M}_t(x; \# \lambda)) \) 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 $(\tau^\ell_t)_{t \in \mathbb{N}} \downarrow f_*$ (resp. $(\tau^u_t)_{t \in \mathbb{N}} \uparrow f^*$)

- $\tau^\ell_t$ is the smallest root of the univariate orthonormal polynomial $T_{t+1}$.
- $\tau^u_t$ is the largest root of the univariate orthonormal polynomial $T_{t+1}$. 

Jean B. Lasserre

Optimization & 3-diagonal Hankel matrices
However

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

\(\Omega\) 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

$$\Omega = \{ \mathbf{x} : f(\mathbf{x}) \leq 1 \} \subset \mathbb{B},$$

be compact.

Compute the Lebesgue volume

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

... 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 $\Omega$. 
Another application of the pushforward

Let $f$ be a nonnegative homogeneous polynomial, and let

$$\Omega = \{ x : f(x) \leq 1 \} \subset \mathbb{B},$$

be compact.

Compute the Lebesgue volume

$$\rho = \text{vol}(\Omega) = \int_{\Omega} dx$$

... and possibly the moments

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

of the Lebesgue measure on $\Omega$. 
It turns out that:

\[
\text{vol}(\Omega) = \int_{\Omega} dx = \frac{1}{\Gamma(1 + n/d)} \int_{\Omega} \exp(-f(x)) \, dx.
\]

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

I. \( \int_{\Omega} \exp(-f(x)) \, dx \), called an *integral discriminant*, is ubiquitous in *statistical and quantum Physics*. 
II. From the above formula it follows that

\[ \text{vol}(\Omega) \text{ is a strictly CONVEX function of the coefficients of the polynomial } f. \]

very useful for solving the following Problem \( P \):

\( P \) : Compute nonnegative homogeneous polynomial \( f \) of degree \( 2d \) such that \( K \subset \Omega \) and \( \Omega \) has minimum volume.

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

Theorem

\( \text{Problem } P \text{ is a CONVEX problem with a unique optimal solution } f^* \)
II. From the above formula it follows that

\[ \text{vol}(\Omega) \] 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 \( K \subset \Omega \) and \( \Omega \) has minimum volume.

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

Theorem

\[ \text{Problem } P \text{ is a CONVEX problem with a unique optimal solution } f^* \]
\( d = 2 \) (quadratic case): \( \Omega_{f^*} \) is the celebrated Löwner-John ellipsoid

However, given \( f \), computing vol(\( \Omega_f \)) is difficult!
\[ d = 2 \text{ (quadratic case)} \]: \( \Omega_{f^*} \) is the celebrated Löwner-John ellipsoid

However, given \( f \), computing \( \text{vol}(\Omega_f) \) is difficult!
Let $\lambda$ be the Lebesgue probability measure on a box $\mathbf{B} \supset \Omega$.

**General approach**

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

(ii) Or **SOLVE**† (or approximate)

$$\text{vol}(\Omega) = \max_{\phi} \{ \phi(\Omega) : \phi \leq \lambda \}$$

where the “max” is over measures $\phi$ supported on $\Omega$.

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

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

- Additional linear constraints coming from Stokes’ theorem applied to $\phi^*$ 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) A simple method that can handle potentially relatively large dimensions. On the other hand, it only provides a (random) estimate of $\text{vol}(\Omega)$.

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

- Additional linear constraints coming from Stokes’ theorem applied to $\phi^*$ 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) A simple method that can handle potentially relatively large dimensions. On the other hand, it only provides a (random) estimate of $\text{vol}(\Omega)$.

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

- Additional linear constraints coming from Stokes’ theorem applied to $\phi^*$ 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:

\[
0 = \int_{\Omega} \text{Div}(X \cdot \mathbf{x}^\alpha (1 - f)) \, dx = \int_{\Omega} \text{Div}(X \cdot \mathbf{x}^\alpha (1 - f)) \, d\phi^*
\]

\[
= \int \mathbf{x}^\alpha \left[ (n + |\alpha|) (1 - f) - \langle \mathbf{x}, \nabla f \rangle \right] d\phi^*
\]

\[
= \int p_\alpha(\mathbf{x}) \, d\phi^* \quad \text{a moment constraint on } \phi^*
\]

Hence one may equivalently solve:

\[
\text{vol}(\Omega) = \max_{\phi \in \mathcal{M}(\Omega)} \{ \phi(\Omega) : \phi \leq \lambda; \int p_\alpha \, d\phi = 0, \quad \alpha \in \mathbb{N}^n \}
\]

⚠️ The associated relaxations of the Moment-SOS hierarchy converge much faster!
Stokes’ theorem

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

$$
0 = \int_{\Omega} \text{Div}(X \cdot x^\alpha (1 - f)) \, dx = \int \text{Div}(X \cdot x^\alpha (1 - f)) \, d\phi^*
$$

$$
= \int x^\alpha [(n + |\alpha|)(1 - f) - \langle x, \nabla f \rangle] \, d\phi^*
$$

$$
= \int p_\alpha(x) \, d\phi^* \quad \text{a moment constraint on } \phi^*
$$

Hence one may equivalently solve:

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

The associated relaxations of the Moment-SOS hierarchy converge much faster!
Let the measure $\#\lambda$ on $\mathbb{R}$ be the pushforward of $\lambda$ by the mapping $f : \mathcal{B} \to \mathbb{R}$.

That is:

$$\#\lambda(B) = \lambda(f^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$
Introduction
Let $I := f(B) \subset \mathbb{R}$. Notice that:

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

$$\gamma_k := \int_I z^k d\#\lambda(z) = \int_B f(x)^k \lambda(dx), \quad k = 0, 1, \ldots$$

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 \(\rho\) is computing the one-dimensional measure of the interval \([0, 1]\) for the measure \(\#\lambda\) on \(\mathbb{R}\) …

Therefore Jasour et al.\(^\dagger\) et al. suggest to solve:

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

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

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 \( \rho \) 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:

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

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

Hence

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

The value $\rho$ 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 \textit{n-dimensional Lebesgue-volume} of $\Omega$ by computation of the \textit{1-dimensional $\#\lambda$-volume} of the interval $[0, 1]$.

The value $\rho$ can be approximated as closely as desired by solving \textit{appropriate SDP relaxations} associated with the \textit{Moment-SOS hierarchy}.

However ...

- Convergence $(\rho_d)_{d \in \mathbb{N}} \downarrow \rho$ is typically \textit{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 $\Omega$ by computation of the 1-dimensional $\#\lambda$-volume of the interval $[0, 1]$.

The value $\rho$ 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!

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

so that \( \rho = \lambda(\Omega) = \phi_0^* \).
The homogeneous case

Take home message:
When \( f \) is \textbf{homogeneous} then one can do much better!

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

so that \( \rho = \lambda(\Omega) = \phi_0^* \).
Suppose that $f$ is **NONNEGATIVE** and **HOMOGENEOUS** of degree $t$. Then by Stokes’ Theorem with vector field $X = x$:

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

Let \((\phi_j^*)_{j \in \mathbb{N}}\) be the moments of \(\phi^*\). 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 \(\phi^*\), is just \(\phi_0^* H_d^*\) with:

\[
H_d^* = \begin{bmatrix}
1 & \frac{n}{n+t} & \cdots & \frac{n}{n+d t} \\
\frac{n}{n+t} & \cdots & \cdots & \frac{n}{n+(d+1) t} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{n}{n+d t} & \cdots & \cdots & \frac{n}{n+2d t}
\end{bmatrix}
\]

which is the moment matrix of the probability measure

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

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

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

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

\[
H_d^* = \begin{pmatrix}
1 & \frac{n}{n+t} & \cdots & \frac{n}{n+d} \\
\frac{n}{n+t} & \ddots & \cdots & \frac{n}{n+(d+1)} \\
\vdots & \ddots & \ddots & \vdots \\
\frac{n}{n+d} & \cdots & \frac{n}{n+(d+1)} & \frac{n}{n+2d}
\end{pmatrix}
\]

which is the moment matrix of the probability measure

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

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

can be approximated as closely as desired by

\[ \tau_d = \max_{\theta} \{ \theta : \theta H_d^* \preceq H_d(\#\lambda) \} \]
\[ = \lambda_{\min}(H_d(\#\lambda), H_d^*) \]

a **GENERALIZED EIGENVALUE PROBLEM** associated with two **HANKEL** moment matrices.

**Theorem**

\[ \tau_d \downarrow \rho \text{ as } d \to \infty. \]
But then:

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

can be approximated as closely as desired by

\[ \tau_d = \max_{\theta} \{ \theta : \theta H_d^* \preceq H_d(\#\lambda) \} \]

\[ = \lambda_{\min}(H_d(\#\lambda), H_d^*) \]

a generalized eigenvalue problem associated with two Hankel moment matrices.

**Theorem**

\[ \tau_d \downarrow \rho \text{ as } d \to \infty. \]
To visualize & appreciate the simplicity of the approach, let 
\( n = 2 \) and 
\[ f(x) = \|x\|^2 = x_1^2 + x_2^2, \]
and 
\[ B = [-1, 1]^2, \]
so that 
\[ \text{vol}(\Omega) = \pi. \]
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 \( \pi \) whereas 
\[ 4 \cdot \rho_1 = 4. \]
\[
H_2^* = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}; \quad H_2(\#\lambda) = \begin{bmatrix} 1 & 2/3 & 28/45 \\ 2/3 & 28/45 & 24/35 \\ 28/45 & 24/35 & 2/9 + 8/21 + 6/25 \end{bmatrix}
\]
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 \( \pi \) with only moments of order 4.
To visualize & appreciate the simplicity of the approach, let $n = 2$ and $f(x) = \|x\|^2 = x_1^2 + x_2^2$, and $B = [-1, 1]^2$, so that $\text{vol}(\Omega) = \pi$. 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 $\pi$ whereas $4 \cdot \rho_1 = 4$.

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

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 $\pi$ with only moments of order 4.
<table>
<thead>
<tr>
<th>$d$</th>
<th>$d = 1$</th>
<th>$d = 2$</th>
<th>$d = 3$</th>
<th>$d = 4$</th>
<th>$d = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_d$</td>
<td>12.19</td>
<td>11.075</td>
<td>9.163</td>
<td>8.878</td>
<td>8.499</td>
</tr>
<tr>
<td>$\tau_d$</td>
<td>6.839</td>
<td>5.309</td>
<td>5.001</td>
<td>4.945</td>
<td>4.936</td>
</tr>
</tbody>
</table>

**Table** – $n = 4$, $\rho = 4.9348$; $\rho_d$ versus $\tau_d$

<table>
<thead>
<tr>
<th>$d$</th>
<th>$d = 3$</th>
<th>$d = 4$</th>
<th>$d = 5$</th>
<th>$d = 6$</th>
<th>$d = 7$</th>
<th>$d = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^n \tau_d$</td>
<td>7.97</td>
<td>5.569</td>
<td>4.639</td>
<td>4.272</td>
<td>4.133</td>
<td>4.083</td>
</tr>
<tr>
<td>$\frac{(2^n \tau_d - \rho^<em>)}{\rho^</em>}$</td>
<td>96%</td>
<td>37%</td>
<td>14%</td>
<td>5.26%</td>
<td>1.83%</td>
<td>0.60%</td>
</tr>
</tbody>
</table>

**Table** – $n = 8$, $\rho = 4.0587$; $\tau_d$ and relative error
THANK YOU!