# Moments and positive polynomials for optimization IV: Another look at nonnegativity and optimization

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

Master 2 Optimization: Paris-Orsay, 2016

- Positivstellensatze for semi-algebraic sets **K** $\subset \mathbb{R}^n$ from the knowledge of defining polynomials
- $\rightarrow$ inner approximations of the cone of polynomials nonnegative on **K**
- Optimization: Semidefinite relaxations yield lower bounds
- Another look at nonnegativity from knowledge of a measure supported on **K**.
- $\rightarrow$ outer approximations of the cone of polynomials nonnegative on **K**
- Optimization: Semidefinite approximations yield upper bounds

- Positivstellensatze for semi-algebraic sets $\mathbf{K} \subset \mathbb{R}^n$ from the knowledge of defining polynomials
- $\rightarrow$ inner approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite relaxations yield lower bounds
- Another look at nonnegativity from knowledge of a measure supported on $\mathbf{K}$.
- $\rightarrow$ outer approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite approximations yield upper bounds

- Positivstellensatze for semi-algebraic sets $\mathbf{K} \subset \mathbb{R}^n$ from the knowledge of defining polynomials
- $\rightarrow$ inner approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite relaxations yield lower bounds
- Another look at nonnegativity from knowledge of a measure supported on $\mathbf{K}$.
- $\rightarrow$ outer approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite approximations yield upper bounds

- Positivstellensatze for semi-algebraic sets $\mathbf{K} \subset \mathbb{R}^n$ from the knowledge of defining polynomials
- $\rightarrow$ inner approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite relaxations yield lower bounds
- Another look at nonnegativity from knowledge of a measure supported on $\mathbf{K}$.
- $\rightarrow$ outer approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite approximations yield upper bounds

- Positivstellensatze for semi-algebraic sets $\mathbf{K} \subset \mathbb{R}^n$ from the knowledge of defining polynomials
- $\rightarrow$ inner approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite relaxations yield lower bounds
- Another look at nonnegativity from knowledge of a measure supported on $\mathbf{K}$.
- $\rightarrow$ outer approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite approximations yield upper bounds

- Positivstellensatze for semi-algebraic sets **K** $\subset \mathbb{R}^n$ from the knowledge of defining polynomials
- $\rightarrow$ inner approximations of the cone of polynomials nonnegative on **K**
- Optimization: Semidefinite relaxations yield lower bounds
- Another look at nonnegativity from knowledge of a measure supported on **K**.
- $\rightarrow$ outer approximations of the cone of polynomials nonnegative on **K**
- Optimization: Semidefinite approximations yield upper bounds

- Positivstellensatze for semi-algebraic sets $\mathbf{K} \subset \mathbb{R}^n$ from the knowledge of defining polynomials
- $\rightarrow$ inner approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite relaxations yield lower bounds
- Another look at nonnegativity from knowledge of a measure supported on $\mathbf{K}$.
- $\rightarrow$ outer approximations of the cone of polynomials nonnegative on $\mathbf{K}$
- Optimization: Semidefinite approximations yield upper bounds

# Let $\mathbf{K} \subseteq \mathbb{R}^n$ be closed

**A basic question is:**

Characterize the continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ that are nonnegative on $\mathbf{K}$

AND .....


**if one obtains ...**

a characterization amenable to practical computation!

# Positivstellensatze for basic semi-algebraic sets

Let $\mathbf{K} := \{\mathbf{x} \in \mathbb{R}^n \, : \, g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m\}$, for some polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$.

Here, knowledge on $\mathbf{K}$ is through its defining polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$.

Let $\mathcal{C}(\mathbf{K})_d$ be the CONVEX cone of polynomials of degree at most $d$, nonnegative on $\mathbf{K}$, and $\mathcal{C}_d$ the CONVEX cone of polynomials of degree at most $d$, nonnegative on $\mathbb{R}^n$.

Let $g_0(x) = 1$ for all $x$.

**The quadratic module associated with $(g_j)$ is the set**

$$Q(g) := \left\{ \sum_{j=0}^{m} \sigma_j \, g_j \; : \; \sigma_j \in \Sigma[\mathbf{x}] \right\}$$

Of course every element of $Q(g)$ is nonnegative on $\mathbf{K}$, and the $(\sigma_j)$ provide certificates of nonnegativity on $\mathbf{K}$.

# Truncated versions

The $k$-truncated quadratic module associated with the $(g_j)$ is the set

$$Q_k(g) := \left\{ \sum_{j=0}^{m} \sigma_j \, g_j \; : \; \sigma_j \in \Sigma[\mathbf{x}], \quad \deg \sigma_j \, g_j \leq 2k \right\}$$

And as one is interested in the cone of polynomials of degree at most $d$, nonnegative on $\mathbf{K}$,

. . . consider the $d$-truncated convex cone:

$$Q_k^d(g) := Q_k(g) \cap \mathbb{R}[\mathbf{x}]_d$$

## Observe that

$$Q_k^d(g) \subset \mathcal{C}(\mathbf{K})_d, \qquad \forall k,$$

and so, the convex cones $(Q_k^d(g))$, $k \in \mathbb{N}$, provide nested inner approximations of $\mathcal{C}(\mathbf{K})_d$.

## ... and ... TESTING whether $f \in Q_k^d(g)$

reduces to SOLVING a SEMIDEFINITE PROGRAM

(a convex optimization problem that can be solved efficiently)

## . . . which provides the basis of

moment-sos relaxations for polynomial programming!

## Observe that

$$Q_k^d(g) \subset \mathcal{C}(\mathbf{K})_d, \qquad \forall k,$$

and so, the convex cones $(Q_k^d(g))$, $k \in \mathbb{N}$, provide nested inner approximations of $\mathcal{C}(\mathbf{K})_d$.

## ... and ... TESTING whether $f \in Q_k^d(g)$

reduces to SOLVING a SEMIDEFINITE PROGRAM

(a convex optimization problem that can be solved efficiently)

## . . . which provides the basis of

moment-sos relaxations for polynomial programming!

## Observe that

$$Q_k^d(g) \subset \mathcal{C}(\mathbf{K})_d, \qquad \forall k,$$

and so, the convex cones $(Q_k^d(g))$, $k \in \mathbb{N}$, provide nested inner approximations of $\mathcal{C}(\mathbf{K})_d$.

## ... and ... TESTING whether $f \in Q_k^d(g)$

reduces to SOLVING a SEMIDEFINITE PROGRAM

(a convex optimization problem that can be solved efficiently)

## . . . which provides the basis of

moment-sos relaxations for polynomial programming!

Recall the fundamental and powerful representation result:

**Putinar-Prestel-Jacobi Positivstellensatz**

Assume that for some $M > 0$, the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ is in $Q(g)$ and let $f \in \mathbb{R}[X]_d$. Then:

$$[\,\mathbf{K} \text{ compact and } f > 0 \text{ on } \mathbf{K}\,] \quad \Rightarrow \quad f \in Q_k^d(g)$$

for some integer $k$.

Recall the fundamental and powerful representation result:

**Putinar-Prestel-Jacobi Positivstellensatz**

Assume that for some $M > 0$, the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ is in $Q(g)$ and let $f \in \mathbb{R}[X]_d$. Then:

$$[\,\mathbf{K} \text{ compact and } f > 0 \text{ on } \mathbf{K}\,] \quad \Rightarrow \quad f \in Q_k^d(g)$$

for some integer $k$.

Recall the fundamental and powerful representation result:

**Putinar-Prestel-Jacobi Positivstellensatz**

Assume that for some $M > 0$, the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ is in $Q(g)$ and let $f \in \mathbb{R}[X]_d$. Then:

$$[\,\mathbf{K} \text{ compact and } f > 0 \text{ on } \mathbf{K}\,] \quad \Rightarrow \quad f \in Q_k^d(g)$$

for some integer $k$.

In fact, Putinar's Positivstellensatz can be re-stated as:

$$\overline{\left(\bigcup_{k=0}^{\infty} Q_k^d(g)\right)} = \mathcal{C}(\mathbf{K})_d$$

( if $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ is in $Q(g)$ )

# Optimization: Hierarchy of semidefinite relaxations

**Consider the global optimization problem**

$$f^* = \min\{f(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K}\}$$

and with $2k_0 \geq \deg f$, consider the semidefinite programs:

$$\rho_k := \max_{\lambda}\{\, \lambda \; : \; f - \lambda \in Q_k(g) \,\}, \quad k \geq k_0$$

We have already seen:

**Theorem**

*Let $\mathbf{K}$ be compact and assume that the polynomial $M - \|\mathbf{x}\|^2$ belongs to $Q(g)$. Then $\rho_k \uparrow f^* := \min\{f(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K}\}$.*

**Consider the global optimization problem**

$$f^* = \min\{f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K}\}$$

and with $2k_0 \geq \deg f$, consider the semidefinite programs:

$$\rho_k := \max_{\lambda}\{\, \lambda \,:\, f - \lambda \in Q_k(g)\,\}, \quad k \geq k_0$$

We have already seen:

**Theorem**

*Let* $\mathbf{K}$ *be compact and assume that the polynomial* $M - \|\mathbf{x}\|^2$ *belongs to* $Q(g)$*. Then* $\rho_k \uparrow f^* := \min\{f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K}\}$.

# Optimization: Hierarchy of semidefinite relaxations

**Consider the global optimization problem**

$$f^* = \min\{f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K}\}$$

and with $2k_0 \geq \deg f$, consider the semidefinite programs:

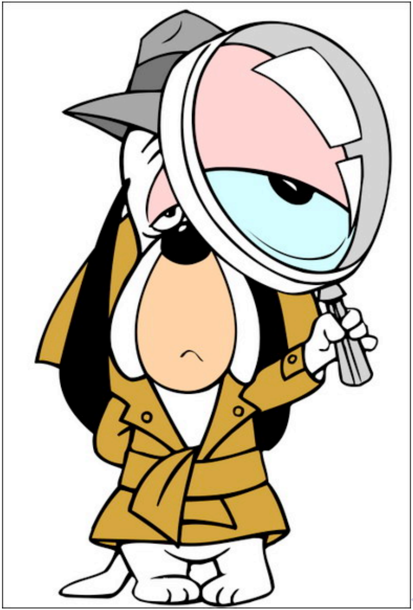
$$\rho_k := \max_{\lambda}\{\, \lambda \,:\, f - \lambda \in Q_k(g) \,\}, \quad k \geq k_0$$

We have already seen:

**Theorem**

*Let* $\mathbf{K}$ *be compact and assume that the polynomial* $M - \|\mathbf{x}\|^2$ *belongs to* $Q(g)$*. Then* $\rho_k \uparrow f^* := \min\{f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K}\}$.

# Another look at of nonnegativity



Let $\mathbf{K} \subseteq \mathbb{R}^n$ be an arbitrary closed set, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function.

### Support of a measure

On a separable metric space $X$, the support $\mathrm{supp}\,\mu$ of a Borel measure $\mu$ is the (unique) smallest closed set such that $\mu(X \setminus \mathbf{K}) = 0$.

Here the knowledge on $\mathbf{K}$ is through a measure $\mu$ with $\mathrm{supp}\,\mu = \mathbf{K}$, and is independent of the representation of $\mathbf{K}$.

### Lemma (Let $\mu$ be such that $\mathrm{supp}\,\mu = \mathbf{K}$)

*A continuous function $f : X \to \mathbb{R}$ is nonnegative on $\mathbf{K}$ if and only if the signed Borel measure $\nu(B) = \int_{\mathbf{K} \cap B} f\, d\mu$, $B \in \mathcal{B}$, is a positive measure.*

Let $\mathbf{K} \subseteq \mathbb{R}^n$ be an arbitrary closed set, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function.

### Support of a measure

On a separable metric space $X$, the support $\mathrm{supp}\,\mu$ of a Borel measure $\mu$ is the (unique) smallest closed set such that $\mu(X \setminus \mathbf{K}) = 0$.

Here the knowledge on $\mathbf{K}$ is through a measure $\mu$ with $\mathrm{supp}\,\mu = \mathbf{K}$, and is independent of the representation of $\mathbf{K}$.

### Lemma (Let $\mu$ be such that $\mathrm{supp}\,\mu = \mathbf{K}$)

*A continuous function $f : X \to \mathbb{R}$ is nonnegative on $\mathbf{K}$ if and only if the signed Borel measure $\nu(B) = \int_{\mathbf{K}\cap B} f\, d\mu$, $B \in \mathcal{B}$, is a positive measure.*

# proof

The *only if part* is straightforward. For the *if part*, if $\nu$ is a positive measure then $f(\mathbf{x}) \geq 0$ for $\mu$-almost all $\mathbf{x} \in \mathbf{K}$. That is, there is a Borel set $G \subset \mathbf{K}$ such that $\mu(G) = 0$ and $f(\mathbf{x}) \geq 0$ on $\mathbf{K} \setminus G$.

Next, observe that $\overline{\mathbf{K} \setminus G} \subset \mathbf{K}$ and $\mu(\overline{\mathbf{K} \setminus G}) = \mu(\mathbf{K})$. Therefore $\overline{\mathbf{K} \setminus G} = \mathbf{K}$ by minimality of $\mathbf{K}$.

Hence, let $\mathbf{x} \in \mathbf{K}$ be fixed, arbitrary. As $\mathbf{K} = \overline{\mathbf{K} \setminus G}$, there is a sequence $(\mathbf{x}_k) \subset \mathbf{K} \setminus G$, $k \in \mathbb{N}$, with $\mathbf{x}_k \to \mathbf{x}$ as $k \to \infty$. But since $f$ is continuous and $f(\mathbf{x}_k) \geq 0$ for every $k \in \mathbb{N}$, we obtain the desired result $f(\mathbf{x}) \geq 0$. $\square$

# proof

The *only if part* is straightforward. For the *if part*, if $\nu$ is a positive measure then $f(\mathbf{x}) \geq 0$ for $\mu$-almost all $\mathbf{x} \in \mathbf{K}$. That is, there is a Borel set $G \subset \mathbf{K}$ such that $\mu(G) = 0$ and $f(\mathbf{x}) \geq 0$ on $\mathbf{K} \setminus G$.

Next, observe that $\overline{\mathbf{K} \setminus G} \subset \mathbf{K}$ and $\mu(\overline{\mathbf{K} \setminus G}) = \mu(\mathbf{K})$. Therefore $\overline{\mathbf{K} \setminus G} = \mathbf{K}$ by minimality of $\mathbf{K}$.

Hence, let $\mathbf{x} \in \mathbf{K}$ be fixed, arbitrary. As $\mathbf{K} = \overline{\mathbf{K} \setminus G}$, there is a sequence $(\mathbf{x}_k) \subset \mathbf{K} \setminus G$, $k \in \mathbb{N}$, with $\mathbf{x}_k \to \mathbf{x}$ as $k \to \infty$. But since $f$ is continuous and $f(\mathbf{x}_k) \geq 0$ for every $k \in \mathbb{N}$, we obtain the desired result $f(\mathbf{x}) \geq 0$. $\square$

# proof

The *only if part* is straightforward. For the *if part*, if $\nu$ is a positive measure then $f(\mathbf{x}) \geq 0$ for $\mu$-almost all $\mathbf{x} \in \mathbf{K}$. That is, there is a Borel set $G \subset \mathbf{K}$ such that $\mu(G) = 0$ and $f(\mathbf{x}) \geq 0$ on $\mathbf{K} \setminus G$.

Next, observe that $\overline{\mathbf{K} \setminus G} \subset \mathbf{K}$ and $\mu(\overline{\mathbf{K} \setminus G}) = \mu(\mathbf{K})$. Therefore $\overline{\mathbf{K} \setminus G} = \mathbf{K}$ by minimality of $\mathbf{K}$.

Hence, let $\mathbf{x} \in \mathbf{K}$ be fixed, arbitrary. As $\mathbf{K} = \overline{\mathbf{K} \setminus G}$, there is a sequence $(\mathbf{x}_k) \subset \mathbf{K} \setminus G$, $k \in \mathbb{N}$, with $\mathbf{x}_k \to \mathbf{x}$ as $k \to \infty$. But since $f$ is continuous and $f(\mathbf{x}_k) \geq 0$ for every $k \in \mathbb{N}$, we obtain the desired result $f(\mathbf{x}) \geq 0$. $\square$

# Moment and localizing matrix

Let $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, be the moment of a finite Borel measure $\mu$ on $\mathbb{R}^n$, i.e.,

$$y_\alpha = \int_{\mathbb{R}^n} x^\alpha \, d\mu \quad \left( = \int_{\mathbb{R}^n} x_1^{\alpha_1} \cdots x_n^{\alpha_n} \, d\mu \right), \quad \forall \alpha \in \mathbb{N}^n.$$

**The "Moment matrix" $M_d(y)$ has its rows and columns**

indexed in the basis $\{X^\alpha\}$ of $\mathbb{R}[X]_d$, and with entries:

$$\begin{aligned} M_d(y)(\alpha, \beta) &= \int_{\mathbb{R}^n} X^{\alpha+\beta} \, d\mu \\ &= y_{\alpha+\beta} \qquad \forall\, \alpha, \beta \in \mathbb{N}^n, \quad |\alpha|, |\beta| \leq d. \end{aligned}$$

For instance in $\mathbb{R}^2$ : $M_1(y) = \overbrace{\left[\begin{array}{c|cc} y_{00} & y_{10} & y_{01} \\ - & - & - \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{array}\right]}^{1 \qquad X_1 \quad X_2}$

**Importantly . . .**

$$M_d(y) \succeq 0 \iff \int_{\mathbb{R}^n} h^2 \, d\mu \geq 0, \qquad \forall h \in \mathbb{R}[X]_d$$

## The “Localizing matrix" $M_d(\theta y)$ w.r.t. a polynomial $\theta \in \mathbb{R}[X]$

with $X \mapsto \theta(X) = \sum_\gamma \theta_\gamma X^\gamma$, has its rows and columns also indexed in the basis $\{X^\alpha\}$ of $\mathbb{R}[X]_d$, and with entries:

$$
\begin{aligned}
M_d(\theta\, y)(\alpha,\beta) &= \int_{\mathbb{R}^n} \theta(X)\, X^{\alpha+\beta}\, d\mu \\
&= \sum_{\gamma\in\mathbb{N}^n} \theta_\gamma\, y_{\alpha+\beta+\gamma}, \qquad \left\{ \begin{array}{l} \alpha,\beta \in \mathbb{N}^n \\ |\alpha|, |\beta| \leq d. \end{array} \right.
\end{aligned}
$$

For instance, in $\mathbb{R}^2$, and with $X \mapsto \theta(X) := 1 - X_1^2 - X_2^2$,

$$M_1(\theta\, y) = \overbrace{\begin{bmatrix} y_{00} - y_{20} - y_{02}, & y_{10} - y_{30} - y_{12}, & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12}, & y_{20} - y_{40} - y_{22}, & y_{11} - y_{21} - y_{12} \\ y_{01} - y_{21} - y_{03}, & y_{11} - y_{21} - y_{12}, & y_{02} - y_{22} - y_{04} \end{bmatrix}}^{1 \qquad\qquad X_1 \qquad\qquad X_2}.$$

**Importantly ...**

$$M_d(\theta\, y) \succeq 0 \quad \Longleftrightarrow \quad \int_{\mathbb{R}^n} h^2\, \theta\, d\mu \geq 0, \qquad \forall h \in \mathbb{R}[X]_d$$

## Theorem

*Let* $\mathbf{K} \subseteq [-1,1]^n$ *be compact and let* $\mu$ *be an arbitrary, fixed, finite Borel measure on* $\mathbf{K}$ *with* $\mathrm{supp}\,\mu = \mathbf{K}$ *and with moments* $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$.

(a) $f \in \mathbb{R}[\mathbf{x}]$ *is nonnegative on* $\mathbf{K}$ *if and only if*

$$M_d(f\,\mathbf{y}) \succeq 0, \qquad d = 0, 1, \ldots$$

(b) *If in addition,* $f$ *is also concave on* $\mathbf{K}$*, then one may replace* $\mathbf{K}$ *with* $\mathrm{co}\,(\mathbf{K})$.

## Theorem

*Let* $\mathbf{K} \subseteq [-1,1]^n$ *be compact and let* $\mu$ *be an arbitrary, fixed, finite Borel measure on* $\mathbf{K}$ *with* $\mathrm{supp}\,\mu = \mathbf{K}$ *and with moments* $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$.

(a) $f \in \mathbb{R}[\mathbf{x}]$ *is nonnegative on* $\mathbf{K}$ *if and only if*

$$M_d(f\,\mathbf{y}) \succeq 0, \qquad d = 0, 1, \ldots$$

(b) *If in addition,* $f$ *is also concave on* $\mathbf{K}$*, then one may replace* $\mathbf{K}$ *with* $\mathrm{co}\,(\mathbf{K})$.

# Sketch of proof

Consider the signed measure $d\nu = f\,d\mu$. As $\mathbf{K} \subseteq [-1,1]^n$,

$$|z_\alpha| = \left|\int_{\mathbf{K}} \mathbf{x}^\alpha f d\mu\right| \leq \int_{\mathbf{K}} |f| d\mu = \|f\|_1, \qquad \forall \alpha \in \mathbb{N}^n.$$

and so $z$ is the moment sequence of a finite (positive) Borel measure $\psi$ on $[-1,1]^n$.

As $\mathbf{K}$ is compact this implies $\nu = \psi$, and so, $\nu$ is a positive Borel measure, and with support equal to $\mathbf{K}$.

By the Lemma that we have seen, $f \geq 0$ on $\mathbf{K}$. $\square$

# Sketch of proof

Consider the signed measure $d\nu = f\,d\mu$. As $\mathbf{K} \subseteq [-1,1]^n$,

$$|z_\alpha| = \left|\int_{\mathbf{K}} \mathbf{x}^\alpha f d\mu\right| \leq \int_{\mathbf{K}} |f| d\mu = \|f\|_1, \qquad \forall \alpha \in \mathbb{N}^n.$$

and so $z$ is the moment sequence of a finite (positive) Borel measure $\psi$ on $[-1,1]^n$.

As $\mathbf{K}$ is compact this implies $\nu = \psi$, and so, $\nu$ is a positive Borel measure, and with support equal to $\mathbf{K}$.

By the Lemma that we have seen, $f \geq 0$ on $\mathbf{K}$. □

# Sketch of proof

Consider the signed measure $d\nu = f\, d\mu$. As $\mathbf{K} \subseteq [-1,1]^n$,

$$|z_\alpha| = \left| \int_{\mathbf{K}} \mathbf{x}^\alpha f d\mu \right| \leq \int_{\mathbf{K}} |f| d\mu = \|f\|_1, \qquad \forall \alpha \in \mathbb{N}^n.$$

and so $z$ is the moment sequence of a finite (positive) Borel measure $\psi$ on $[-1,1]^n$.

As $\mathbf{K}$ is compact this implies $\nu = \psi$, and so, $\nu$ is a positive Borel measure, and with support equal to $\mathbf{K}$.

By the Lemma that we have seen, $f \geq 0$ on $\mathbf{K}$. $\square$

Let identify $f \in \mathbb{R}[\mathbf{x}]_d$ with its vector of coefficient $f \in \mathbb{R}^{s(d)}$, with $s(d) = \binom{n+d}{n}$.

**Observe that, for every $k = 1, \ldots,$ the set**

$$\Delta_k := \{ f \in \mathbb{R}^{s(d)} \; : \; M_k(f\,\mathbf{y}) \succeq 0 \},$$

is the feasible set associated with a Linear Matrix Inequality, and so a CONVEX SET (and in fact, here, a CONVEX CONE).

Indeed the entry $(\alpha, \beta)$ of $M_k(f\,\mathbf{y})$ is just

$$\sum_{\gamma \in \mathbb{N}^n} f_\gamma \, y_{\alpha+\beta+\gamma}$$

and so $M_k(f\,\mathbf{y}) \succeq 0$ is a Linear Matrix Inequality (LMI) on the vector of coefficients of $f$.

Let identify $f \in \mathbb{R}[\mathbf{x}]_d$ with its vector of coefficient $f \in \mathbb{R}^{s(d)}$, with $s(d) = \binom{n+d}{n}$.

**Observe that, for every $k = 1, \ldots,$ the set**

$$\Delta_k := \{ f \in \mathbb{R}^{s(d)} \,:\, M_k(f\,\mathbf{y}) \succeq 0 \},$$

is the feasible set associated with a Linear Matrix Inequality, and so a CONVEX SET (and in fact, here, a CONVEX CONE).

Indeed the entry $(\alpha, \beta)$ of $M_k(f\,\mathbf{y})$ is just

$$\sum_{\gamma \in \mathbb{N}^n} f_\gamma \, y_{\alpha+\beta+\gamma}$$

and so $M_k(f\,\mathbf{y}) \succeq 0$ is a Linear Matrix Inequality (LMI) on the vector of coefficients of $f$.

Example: Let $f \in \mathbb{R}[\mathbf{x}]$ be the polynomial:

$$\mathbf{x} \mapsto f(\mathbf{x}) := a + b\,x_1 x_2.$$

$$M_1(f\,\mathbf{y}) = \begin{bmatrix} a\,y_{00} + b\,y_{11}, & a\,y_{10} + b\,y_{21}, & a\,y_{01} + b\,y_{12} \\ a\,y_{10} + b\,y_{21}, & a\,y_{20} + by_{31}, & a\,y_{11} + b\,y_{22} \\ a\,y_{01} + b\,y_{12}, & a\,y_{11} + b\,y_{22}, & a\,y_{02} + b\,y_{13} \end{bmatrix} \succeq 0.$$

Equivalently,

$$a \begin{bmatrix} y_{00}, & y_{10}, & y_{01} \\ y_{10}, & y_{20}, & y_{11} \\ y_{01}, & y_{11}, & y_{02} \end{bmatrix} + b \begin{bmatrix} y_{11}, & y_{21}, & y_{12} \\ y_{21}, & y_{31}, & y_{22} \\ y_{12}, & y_{22}, & y_{13} \end{bmatrix} \succeq 0.$$

which defines a CONVEX CONE in $\mathbb{R}^2$ for the coefficients $(a, b)$ of polynomials of the form $a + bx_1x_2$.

## and so . . .

one obtains a nested hierarchy of spectrahedra

$$\Delta_0 \supset \Delta_1 \cdots \supset \Delta_k \cdots \supset \mathcal{C}(\mathbf{K})_d,$$

with no lifting, which provide

tighter and tighter outer approximations of $\mathcal{C}(\mathbf{K})_d$.

## So we get the sandwich $Q^d_k(g) \subset \mathcal{C}(\mathbf{K})_d \subset \Delta_k$ for all $k$, and

$$\overline{\left(\bigcup_{k=0}^{\infty} Q^d_k(g)\right)} \;=\; \mathcal{C}(\mathbf{K})_d \;=\; \left(\bigcap_{k=0}^{\infty} \Delta_k\right)$$

$\downarrow$ Inner approximations — representation dependent

$\downarrow$ Outer approximations — independent of representation

# Application to optimization

## Theorem (A hierarchy of upper bounds)

*Let* $f \in \mathbb{R}[\mathbf{x}]_d$ *be fixed and* $\mathbf{K} \subset \mathbb{R}^n$ *be closed. Let* $\mu$ *be such that* $\mathrm{supp}\,\mu = \mathbf{K}$ *and with moment sequence* $\mathbf{y} = (\mathbf{y}_\alpha)$, $\alpha \in \mathbb{N}^n$*.*
*Consider the hierarchy of semidefinite programs:*

$$u_k = \min_{\sigma} \left\{ \int_{\mathbf{K}} f \underbrace{\sigma\, d\mu}_{d\nu} \;:\; \int_{\mathbf{K}} \underbrace{\sigma\, d\mu}_{d\nu} = 1;\; \sigma \in \Sigma[\mathbf{x}]_d \right\},$$

*with dual:*

$$\begin{aligned} u_k^* &= \max_{\lambda} \{ \lambda \;:\; M_k(f-\lambda, \mathbf{y}) \succeq 0 \} \\ &= \max_{\lambda} \{ \lambda \;:\; \lambda\, M_k(\mathbf{y}) \preceq M_k(f, \mathbf{y}) \} \end{aligned}$$

*Then* $u_k^*, u_k \downarrow f^* = \min_{\mathbf{x}} \{ f(\mathbf{x}) \;:\; \mathbf{x} \in \mathbf{K} \}$.

# Interpetation of $u_k$ and $u_k^*$

- Computing $u_k^*$ is a generalized eigenvalue problem!
- Next, recall that

$$f^* = \min_{\psi} \{ \int_{\mathbf{K}} f\, d\psi \,:\, \psi(\mathbf{K}) = 1,\ \psi(\mathbb{R}^n \setminus \mathbf{K}) = 0 \}$$

whereas

$$u_k = \min_{\nu} \{ \int_{\mathbf{K}} f\, \underbrace{\sigma d\mu}_{d\nu} \,:\, \nu(\mathbf{K}) = 1,\ \nu(\mathbb{R}^n \setminus \mathbf{K}) = 0;\ \sigma \in \Sigma[\mathbf{x}]_k \}$$

that is, one optimizes over the subspace of Borel probability measures absolutely continuous with respect to $\mu$, and with density $\sigma \in \Sigma[\mathbf{x}]_k$.

Ideally, when $k$ is large, $\sigma(\mathbf{x}) > 0$ in a neighborhood of a global minimizer $\mathbf{x}^* \in \mathbf{K}$.

# Interpetation of $u_k$ and $u_k^*$

- Computing $u_k^*$ is a generalized eigenvalue problem!
- Next, recall that

$$f^* = \min_{\psi} \{ \int_{\mathbf{K}} f \, d\psi \; : \; \psi(\mathbf{K}) = 1, \; \psi(\mathbb{R}^n \setminus \mathbf{K}) = 0 \}$$

whereas

$$u_k = \min_{\nu} \{ \int_{\mathbf{K}} f \underbrace{\sigma d\mu}_{d\nu} \; : \; \nu(\mathbf{K}) = 1, \; \nu(\mathbb{R}^n \setminus \mathbf{K}) = 0; \; \sigma \in \Sigma[\mathbf{x}]_k \}$$

that is, one optimizes over the subspace of Borel probability measures absolutely continuous with respect to $\mu$, and with density $\sigma \in \Sigma[\mathbf{x}]_k$.

Ideally, when $k$ is large, $\sigma(\mathbf{x}) > 0$ in a neighborhood of a global minimizer $\mathbf{x}^* \in \mathbf{K}$.

# Interpetation of $u_k$ and $u_k^*$

- Computing $u_k^*$ is a generalized eigenvalue problem!
- Next, recall that

$$f^* = \min_{\psi} \{ \int_{\mathbf{K}} f\, d\psi \; : \; \psi(\mathbf{K}) = 1, \; \psi(\mathbb{R}^n \setminus \mathbf{K}) = 0 \}$$

whereas

$$u_k = \min_{\nu} \{ \int_{\mathbf{K}} f \underbrace{\sigma d\mu}_{d\nu} \; : \; \nu(\mathbf{K}) = 1, \; \nu(\mathbb{R}^n \setminus \mathbf{K}) = 0; \; \sigma \in \Sigma[\mathbf{x}]_k \}$$

that is, one optimizes over the subspace of Borel probability measures absolutely continuous with respect to $\mu$, and with density $\sigma \in \Sigma[\mathbf{x}]_k$.

Ideally, when $k$ is large, $\sigma(\mathbf{x}) > 0$ in a neighborhood of a global minimizer $\mathbf{x}^* \in \mathbf{K}$.

- Computing $u_k^*$ is a generalized eigenvalue problem!
- Next, recall that

$$f^* = \min_{\psi}\{\int_{\mathbf{K}} f\,d\psi \;:\; \psi(\mathbf{K}) = 1,\; \psi(\mathbb{R}^n \setminus \mathbf{K}) = 0\}$$

whereas

$$u_k = \min_{\nu}\{\int_{\mathbf{K}} f\,\underbrace{\sigma\,d\mu}_{d\nu} \;:\; \nu(\mathbf{K}) = 1,\; \nu(\mathbb{R}^n \setminus \mathbf{K}) = 0;\; \sigma \in \Sigma[\mathbf{x}]_k\}$$

that is, one optimizes over the subspace of Borel probability measures absolutely continuous with respect to $\mu$, and with density $\sigma \in \Sigma[\mathbf{x}]_k$.

Ideally, when $k$ is large, $\sigma(\mathbf{x}) > 0$ in a neighborhood of a global minimizer $\mathbf{x}^* \in \mathbf{K}$.

# Interpetation of $u_k$ and $u_k^*$

- Computing $u_k^*$ is a generalized eigenvalue problem!
- Next, recall that

$$f^* = \min_{\psi} \{ \int_{\mathbf{K}} f \, d\psi \; : \; \psi(\mathbf{K}) = 1, \; \psi(\mathbb{R}^n \setminus \mathbf{K}) = 0 \}$$

whereas

$$u_k = \min_{\nu} \{ \int_{\mathbf{K}} f \, \underbrace{\sigma \, d\mu}_{d\nu} \; : \; \nu(\mathbf{K}) = 1, \; \nu(\mathbb{R}^n \setminus \mathbf{K}) = 0; \; \sigma \in \Sigma[\mathbf{x}]_k \}$$

that is, one optimizes over the subspace of Borel probability measures absolutely continuous with respect to $\mu$, and with density $\sigma \in \Sigma[\mathbf{x}]_k$.

Ideally, when $k$ is large, $\sigma(\mathbf{x}) > 0$ in a neighborhood of a global minimizer $\mathbf{x}^* \in \mathbf{K}$.

• Also works for non-compact closed sets but then $\mu$ has to satisfy a Carleman-type sufficient condition which limits the growth of the moments. For example, take

$$d\mu = \mathrm{e}^{-\|\mathbf{x}\|^2/2}\, d\nu$$

where $\nu$ is an arbitrary finite Borel measure with support $\mathbf{K}$.

• The sequences of upper bounds $(u_k, u_k^*)$ complement the sequences of lower bounds $(\rho_k, \rho_k^*)$ obtained from SDP-relaxations.

• Of course, for practical computation, the previous semidefinite relaxations require knowledge of the moment sequence $\mathbf{y} = (\mathbf{y}_\alpha)$, $\alpha \in \mathbb{N}^n$.

• Also works for non-compact closed sets but then $\mu$ has to satisfy a Carleman-type sufficient condition which limits the growth of the moments. For example, take

$$d\mu = \mathrm{e}^{-\|\mathbf{x}\|^2/2}\, d\nu$$

where $\nu$ is an arbitrary finite Borel measure with support $\mathbf{K}$.
• The sequences of upper bounds $(u_k, u_k^*)$ complement the sequences of lower bounds $(\rho_k, \rho_k^*)$ obtained from SDP-relaxations.

• Of course, for practical computation, the previous semidefinite relaxations require knowledge of the moment sequence $\mathbf{y} = (\mathbf{y}_\alpha)$, $\alpha \in \mathbb{N}^n$.

• Also works for non-compact closed sets but then $\mu$ has to satisfy a Carleman-type sufficient condition which limits the growth of the moments. For example, take

$$d\mu = \mathrm{e}^{-\|\mathbf{x}\|^2/2}\, d\nu$$

where $\nu$ is an arbitrary finite Borel measure with support $\mathbf{K}$.
• The sequences of upper bounds $(u_k, u_k^*)$ complement the sequences of lower bounds $(\rho_k, \rho_k^*)$ obtained from SDP-relaxations.

• Of course, for practical computation, the previous semidefinite relaxations require knowledge of the moment sequence $\mathbf{y} = (\mathbf{y}_\alpha)$, $\alpha \in \mathbb{N}^n$.

This is possible for relatively simple sets $\mathbf{K}$ like a box, a simplex, the discrete set, an ellipsoid, etc., where one can compute all moments of a measure $\mu$ whose support is $\mathbf{K}$. For instance take $\mu$ to be uniformly distributed, or $\mathbf{K} = \mathbb{R}^n$ (or $\mathbf{K} = \mathbb{R}^n_+$) with

$$d\mu = \mathrm{e}^{-\|\mathbf{x}\|^2/2}\, d\mathbf{x}, \qquad \mathbf{K} = \mathbb{R}^n$$

$$d\mu = \mathrm{e}^{-\sum_i x_i}\, d\mathbf{x}, \qquad \mathbf{K} = \mathbb{R}^n_+$$

$$d\mu = \, d\mathbf{x}, \left\{ \begin{array}{lcl} \mathbf{K} & = & [a_1, b_1] \times \cdots \times [a_n, b_n] \\ \mathbf{K} & = & \{\mathbf{x} \in \mathbb{R}^n_+ \,:\, \sum_{i=1}^n x_i \,\leq\, 1\} \end{array} \right.$$

For $\mathbf{K} = \{-1, 1\}^n$ or $\mathbf{K} = \{0, 1\}^n$ take $\mu$ to be uniformly distributed.

# A sequence of eigenvalue problems

## Practical calculation

If instead of the usual canonical basis of monomials $(X^\alpha)$, $\alpha \in \mathbb{N}^n$, one now uses the basis of polynomials $(P_\alpha)$, $\alpha \in \mathbb{N}^n$, that are ORTHONORMAL with respect to the known measure $\mu$, then the moments matrix $M_k(\mathbf{y})$ expressed in that basis is the IDENTITY matrix! Indeed,

$$M_k(\mathbf{y})(\alpha, \beta) = \int_{\mathbb{R}^n} P_\alpha \, P_\beta \, d\mu = \delta_{\alpha=\beta}.$$

## Then ...

$$u_k^* = \max_\lambda \{ \lambda : \lambda \, M_k(\mathbf{y}) \preceq M_k(f, \mathbf{y}) \},$$

i.e., $u_k^*$ is the smallest eigenvalue of the matrix $M_k(f, \mathbf{y})$!

# A sequence of eigenvalue problems

## Practical calculation

If instead of the usual canonical basis of monomials $(X^\alpha)$, $\alpha \in \mathbb{N}^n$, one now uses the basis of polynomials $(P_\alpha)$, $\alpha \in \mathbb{N}^n$, that are ORTHONORMAL with respect to the known measure $\mu$, then the moments matrix $M_k(\mathbf{y})$ expressed in that basis is the IDENTITY matrix! Indeed,

$$M_k(\mathbf{y})(\alpha, \beta) \,=\, \int_{\mathbb{R}^n} P_\alpha \, P_\beta \, d\mu \,=\, \delta_{\alpha=\beta}.$$

## Then ...

$$u_k^* \,=\, \max_\lambda \{ \, \lambda \,:\, \lambda \, M_k(\mathbf{y}) \preceq M_k(f, \, \mathbf{y}),$$

i.e., $u_k^*$ is the smallest eigenvalue of the matrix $M_k(f, \, \mathbf{y})$!

## Computing a basis of polynomials

$(P_\alpha)$, $\alpha \in \mathbb{N}^n$, orthonormal with respect to $\mu$ is easy if one knows the moments of $\mu$!

For instance: $P_0 = 1$, and

$$P_{10} = \det\left(\begin{bmatrix} y_0 & y_{10} \\ 1 & X_1 \end{bmatrix}\right); \qquad P_{01} = \det\left(\begin{bmatrix} y_0 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ 1 & X_1 & X_2 \end{bmatrix}\right),$$

etc., plus scaling so as to have $\int {P_\alpha}^2 \, d\mu = 1$.

## Computing a basis of polynomials

$(P_\alpha)$, $\alpha \in \mathbb{N}^n$, orthonormal with respect to $\mu$ is easy if one knows the moments of $\mu$!

For instance: $P_0 = 1$, and

$$P_{10} = \det\left(\begin{bmatrix} y_0 & y_{10} \\ 1 & X_1 \end{bmatrix}\right); \qquad P_{01} = \det\left(\begin{bmatrix} y_0 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ 1 & X_1 & X_2 \end{bmatrix}\right),$$

etc., plus scaling so as to have $\int {P_\alpha}^2\, d\mu = 1$.

**Ex 1: With $\mathbf{K} = \mathbb{R}^n_+$ and $\mathbf{x} \mapsto f_{\mathbf{A}}(\mathbf{x}) := \mathbf{x}^T \mathbf{A}\, \mathbf{x}$**

for real symmetric matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, one may thus provide outer approximations of the convex cone of COPOSITIVE matrices, that is, matrices $\mathbf{A}$ such that

$$\mathbf{x}^T \mathbf{A}\, \mathbf{x} \geq 0, \qquad \forall \mathbf{x} \in \mathbb{R}^n_+,$$

an important tool for 0/1 combinatorial optimization problems. These outer approximations complement the inner approximations already obtained by Parrilo, and DeKlerk and Pasechnik.

**Ex 2: With $\mathbf{K} = \{-1, 1\}^n$ and $\mathbf{x} \mapsto f_{\mathbf{A}}(\mathbf{x}) := \mathbf{x}^T \mathbf{A}\, \mathbf{x}$**

for real symmetric matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, one may thus provide a hierarchy of upper bounds for MAXCUT problem with matrix $\mathbf{A}$.

# Some experiments

• $\mathbf{K} = \mathbb{R}^2_+$ with $d\mu = \mathrm{e}^{-\sum_i x_i} d\mathbf{x}$ so that

$$\mathbf{y}_{ij} \; = \; i!\,j!, \qquad \forall i,j = 0,1,\ldots$$

$$\mathbf{x} \mapsto f(\mathbf{x}) := x_1^2 x_2^2\,(x_1^2 + x_2^2 - 1) \text{ with } f^* = -1/27 \approx -0.037.$$

$$u_0 = 92; \quad u_1 = 1.51; \quad u_{14} = -0.011.$$

• The same problem on the box $\mathbf{K} = [0,1]$ now yields

$$u_0 = 0.222; \quad u_1 = -0.055; \quad u_{14} = -0.0311,$$

and some numerical problems occur.

# Some experiments

- $\mathbf{K} = \mathbb{R}^2_+$ with $d\mu = \mathrm{e}^{-\sum_i x_i} d\mathbf{x}$ so that

$$\mathbf{y}_{ij} \;=\; i!\,j!, \qquad \forall i,j = 0,1,\ldots$$

$$\mathbf{x} \mapsto f(\mathbf{x}) := x_1^2 x_2^2\,(x_1^2 + x_2^2 - 1) \text{ with } f^* = -1/27 \approx -0.037.$$

$$u_0 = 92; \quad u_1 = 1.51; \quad u_{14} = -0.011.$$

- The same problem on the box $\mathbf{K} = [0,1]$ now yields

$$u_0 = 0.222; \quad u_1 = -0.055; \quad u_{14} = -0.0311,$$

and some numerical problems occur.

# Some experiments

- $\mathbf{K} = \mathbb{R}^2_+$ with $d\mu = \mathrm{e}^{-\sum_i x_i} d\mathbf{x}$ so that

$$\mathbf{y}_{ij} \;=\; i!\,j!, \qquad \forall i,j = 0,1,\ldots$$

$$\mathbf{x} \mapsto f(\mathbf{x}) := x_1^2 x_2^2\,(x_1^2 + x_2^2 - 1) \text{ with } f^* = -1/27 \approx -0.037.$$

$$u_0 = 92; \quad u_1 = 1.51; \quad u_{14} = -0.011.$$

- The same problem on the box $\mathbf{K} = [0,1]$ now yields

$$u_0 = 0.222; \quad u_1 = -0.055; \quad u_{14} = -0.0311,$$

and some numerical problems occur.

- Some randomly generated MAXCUT problems

$$f^* = \min_x \{\mathbf{x}Q\mathbf{x} \;:\; x \in \{-1,1\}^n\}$$

with $n = 11$ variables.

| $d$ | $u_0$ | $u_1$ | $u_2$ | $u_3$ | $u_4$ | $f^*$ |
|---|---|---|---|---|---|---|
| Ex1 | 0 | -1.928 | -3.748 | -5.22 | -6.37 | -7.946 |
| Ex2 | 0 | -1.56 | -3.103 | -4.314 | -5.282 | -6.863 |
| Ex3 | 0 | -1.910 | -3.694 | -5.078 | -6.161 | -8.032 |
| Ex4 | 0 | -2.164 | -4.1664 | -5.7971 | -7.06 | -9.198 |
| Ex5 | 0 | -1.825 | -3.560 | -4.945 | -5.924 | -7.467 |

Table: MAXCUT: $n = 11$; $Q$ random.

# Illustrating duality

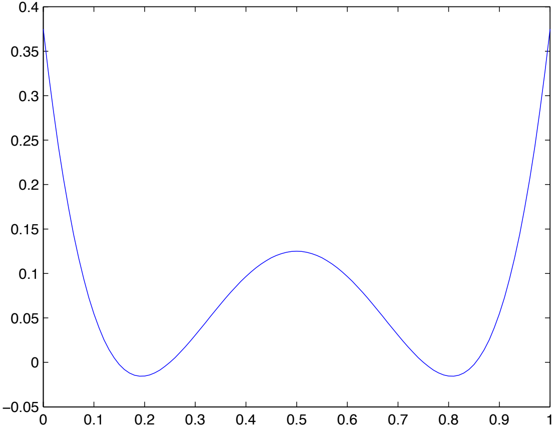



Figure: $f(x) = 0.375 - 5x + 21x^2 - 32x^3 + 16x^4$ on $[0, 1]$

Solving the dual yields the SOS polynomial density $\sigma_k$ with

$$u_k = \int f(\mathbf{x}) \underbrace{\sigma_k(\mathbf{x})\, d\mathbf{x}}_{d\nu_k(\mathbf{x})}$$

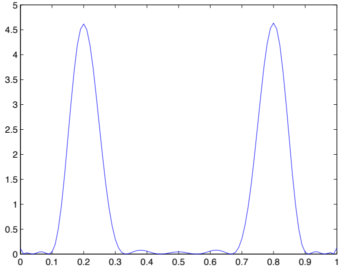



Figure: The probability density $\sigma_{10}(x)dx$ on $[0, 1]$

# Preliminary conclusions

- Rapid decrease in first steps, but convergence is slow
- Numerical stability problems to be expected.
- Use bases different from the monomial basis.
- Rather see this technique as a complement to lower bounds obtained from semidefinite relaxations

# Preliminary conclusions

- Rapid decrease in first steps, but convergence is slow
- Numerical stability problems to be expected.
- Use bases different from the monomial basis.
- Rather see this technique as a complement to lower bounds obtained from semidefinite relaxations

# Preliminary conclusions

- Rapid decrease in first steps, but convergence is slow
- Numerical stability problems to be expected.
- Use bases different from the monomial basis.
- Rather see this technique as a complement to lower bounds obtained from semidefinite relaxations

# Preliminary conclusions

- Rapid decrease in first steps, but convergence is slow
- Numerical stability problems to be expected.
- Use bases different from the monomial basis.
- Rather see this technique as a complement to lower bounds obtained from semidefinite relaxations