The moment-SOS approach II: Some Applications

Jean B. Lasserre
LAAS-CNRS and Institute of Mathematics, Toulouse, France

Master 2 Optimization: Paris-Orsay, 2016
A preliminary result

- Let $B \subset \mathbb{R}^n$ be a simple set like e.g., a Box $[-1, 1]^n$ or an ellipsoid,
- $K := \{ (x, y) : x \in B; g_j(x, y) \geq 0, j = 1, \ldots, m \}$ and $h : \mathbb{R}^n \times \mathbb{R}^p : \rightarrow \mathbb{R}.$

Goal: Approximate the function $F : B \rightarrow \mathbb{R}$

$$x \mapsto F(x) := \sup_{y} \{ h(x, y) : (x, y) \in K \}, \quad x \in B.$$ 

... by POLYNOMIALS $J_k \in \mathbb{R}[x]$ of increasing degree ... and with some guarantee of convergence.
A preliminary result

Let \( B \subset \mathbb{R}^n \) be a simple set like e.g., a Box \([-1, 1]^n\) or an ellipsoid,

\[ K := \{ (x, y) : x \in B; g_j(x, y) \geq 0, j = 1, \ldots, m \} \] and \( h : \mathbb{R}^n \times \mathbb{R}^p :\rightarrow \mathbb{R} \).

Goal: Approximate the function \( F : B \rightarrow \mathbb{R} \)

\[ x \mapsto F(x) := \sup_y \{ h(x, y) : (x, y) \in K \}, \quad x \in B. \]

... by POLYNOMIALS \( J_k \in \mathbb{R}[x] \) of increasing degree ... and with some guarantee of convergence.
Fix \( k \in \mathbb{N} \) and consider the optimization problem:

\[
\inf_{J \in \mathbb{R}[x]_k} \left\{ \int_B (J(x) - F(x)) \, dx : J(x) \geq F(x), \quad \forall x \in B \right\}
\]

Or, equivalently:

\[
\inf_{J \in \mathbb{R}[x]_k} \left\{ \|J - F\|_1 : p(x) \geq F(x), \quad \forall x \in B \right\}
\]

where \( \| \cdot \|_1 \) is the \( L_1 \)-norm

\[
h \mapsto \|h\|_1 := \int_B |h(x)| \, dx, \quad h \in L_1(B).
\]
Fix $k \in \mathbb{N}$ and consider the optimization problem:

$$\inf_{J \in \mathbb{R}[x]_k} \left\{ \int_B (J(x) - F(x)) \, dx : J(x) \geq F(x), \quad \forall x \in B \right\}$$

Or, equivalently:

$$\inf_{J \in \mathbb{R}[x]_k} \left\{ \| J - F \|_1 : p(x) \geq F(x), \quad \forall x \in B \right\}$$

where $\| \cdot \|_1$ is the $L_1$-norm

$$h \mapsto \| h \|_1 := \int_B |h(x)| \, dx, \quad h \in L_1(B).$$
To fix ideas, let $\mathbf{B} := [-1, 1]^n = \{ x : x_i^2 \leq 1, i = 1, \ldots, n \}$ and set $g_0(x) = 1$ and $g_{m+i}(x) = 1 - x_i^2$, $i = 1, \ldots, n$.

A simple strategy:

Replace the difficult constraint $J(x) - F(x)$, $\forall x \in \mathbf{B}$ with the SOS-based positivity certificate

$$J(x) - h(x, y) = \sum_{j=0}^{m+n} \sigma_j(x, y) g_j(x, y),$$

for some SOS polynomials $\sigma_j$ such that $\deg(\sigma_j g_j) \leq 2k$.

Indeed

$$J(x) \geq h(x, y), \quad \forall (x, y) \in \mathbf{K}; x \in \mathbf{B} \Rightarrow J(x) \geq F(x), \quad \forall x \in \mathbf{B}.$$
To fix ideas, let $\mathbf{B} := [-1, 1]^n = \{ x : x_i^2 \leq 1, i = 1, \ldots, n \}$ and set $g_0(x) = 1$ and $g_{m+i}(x) = 1 - x_i^2$, $i = 1, \ldots, n$.

A simple strategy:

Replace the difficult constraint $J(x) - F(x)$, $\forall x \in \mathbf{B}$ with the SOS-based positivity certificate

$$J(x) - h(x, y) = \sum_{j=0}^{m+n} \sigma_j(x, y) g_j(x, y),$$

for some SOS polynomials ($\sigma_j$) such that $\deg(\sigma_j g_j) \leq 2k$.

Indeed

$$J(x) \geq h(x, y), \quad \forall (x, y) \in K; x \in \mathbf{B} \Rightarrow J(x) \geq F(x), \quad \forall x \in \mathbf{B}.$$
Then solve:

\[ \inf_{J, \sigma_j} \left\{ \int_B (J(x) - F(x)) \, dx : \right. \]

\[ \text{s.t. } J(x) - h(x, y) = \sum_{j=0}^{m+n} \sigma_j(x, y) g_j(x, y), \quad \forall (x, y) \in \mathbb{R}^{n+p} \]

\[ J \in \mathbb{R}[x]_k; \quad \sigma_j \text{ SOS; } \deg(\sigma_j g_j) \leq 2k \} . \]

**Theorem**

This is an SDP and it has an optimal solution \( J_k^* \in \mathbb{R}[x]_k \).
Moreover, \( \| J_k^* - f \|_1 \rightarrow 0 \) as \( k \rightarrow \infty \).
Then solve:

\[
\inf_{J, \sigma_j} \left\{ \int_B (J(x) - F(x)) \, dx : \right. \\
\text{s.t. } J(x) - h(x, y) = \sum_{j=0}^{m+n} \sigma_j(x, y) g_j(x, y), \quad \forall (x, y) \in \mathbb{R}^{n+p} \\
J \in \mathbb{R}[x]_k; \quad \sigma_j \text{ SOS; } \deg(\sigma_j g_j) \leq 2k \}.
\]

**Theorem**

This is an SDP and it has an optimal solution \( J_k^* \in \mathbb{R}[x]_k \).
Moreover, \( \|J_k^* - f\|_1 \to 0 \) as \( k \to \infty \).
Write $J(x) = \sum_{\alpha \in \mathbb{N}^n} J_\alpha x^\alpha$.

I. The criterion $\int_B (J(x) - f(x)) \, dx$ reads:

$$\sum_{\alpha \in \mathbb{N}^n} J_\alpha \int_B x^\alpha \, dx - \int_B F(x) \, dx = \sum_{\alpha \in \mathbb{N}^n} J_\alpha \gamma_\alpha - c,$$

where $\gamma_\alpha$ is known and $c$ is an (unknown) constant.

is LINEAR in the unknown coefficients $J_\alpha$ of $J \in \mathbb{R}[x]_k$. 
II. The constraint

\[ J(x) - h(x, y) = \sum_{j=0}^{m+n} \sigma_j(x, y) g_j(x, y), \]

reduces to

- \textbf{LINEAR} constraints on the coefficients \( J_\alpha \) and \( \sigma_{j_\alpha} \),
- + semidefinite constraints to state that the \( \sigma_j \)'s are SOS.
Let $f \in \mathbb{R}[x, y]$ and let $K \subset \mathbb{R}^n \times \mathbb{R}^p$ be the semi-algebraic set:

$$K := \{(x, y) : g_j(x, y) \geq 0, \quad j = 1, \ldots, m\},$$

and let $B \subset \mathbb{R}^n$ be the Euclidean unit ball or the box $[-1, 1]^n$.

Suppose that one wants to approximate the set:

$$R_f := \{x \in B : f(x, y) \leq 0 \text{ for all } y \text{ such that } (x, y) \in K\}$$

as closely as desired by a sequence of sets of the form:

$$\Theta_k := \{x \in B : J_k(x) \leq 0 \}$$

for some polynomials $J_k$. 
Recall that $\mathcal{B} = \{ x : 1 - x_i^2 \geq 0 \}$ and set $g_{m+i}(x) := 1 - x_i^2$, $i = 1, \ldots, n$. With $g_0 = 1$ and with $K \subset \mathbb{R}^n \times \mathbb{R}^p$ and $k \in \mathbb{N}$, let

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

Let $x \mapsto F(x) := \max \{ f(x, y) : (x, y) \in K \}$, and for every integer $k$ consider the optimization problem:

$$\rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_{\mathcal{B}} (J - F) \, dx : J(x) - f(x, y) \in Q_k(g); x \in \mathcal{B} \right\}$$

Remember that $J(x) - f(x, y) \in Q_k(g)$ implies:

$$J(x) \geq f(x, y), \quad \forall (x, y) \in K \quad \Rightarrow \quad J(x) \geq F(x) \quad \forall x \in \mathcal{B}. $$
Recall that $B = \{ x : 1 - x_i^2 \geq 0 \}$ and set $g_{m+i}(x) := 1 - x_i^2$, $i = 1, \ldots, n$. With $g_0 = 1$ and with $K \subset \mathbb{R}^n \times \mathbb{R}^p$ and $k \in \mathbb{N}$, let

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

Let $x \mapsto F(x) := \max \{ f(x, y) : (x, y) \in K \}$, and

for every integer $k$ consider the optimization problem:

$$\rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_B (J - F) \, dx : J(x) - f(x, y) \in Q_k(g); \ x \in B \right\}$$

Remember that $J(x) - f(x, y) \in Q_k(g)$ implies:

$$J(x) \geq f(x, y), \quad \forall (x, y) \in K \quad \Rightarrow \quad J(x) \geq F(x) \quad \forall x \in B.$$
Recall that $B = \{ x : 1 - x_i^2 \geq 0 \}$ and set $g_{m+i}(x) := 1 - x_i^2$, $i = 1, \ldots, n$. With $g_0 = 1$ and with $K \subset \mathbb{R}^n \times \mathbb{R}^p$ and $k \in \mathbb{N}$, let

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

Let $x \mapsto F(x) := \max \{ f(x, y) : (x, y) \in K \}$, and

for every integer $k$ consider the optimization problem:

$$ \rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_B (J - F) \, dx : J(x) - f(x, y) \in Q_k(g) ; x \in B \right\} $$

Remember that $J(x) - f(x, y) \in Q_k(g)$ implies:

$$ J(x) \geq f(x, y), \quad \forall (x, y) \in K \quad \Rightarrow \quad J(x) \geq F(x) \quad \forall x \in B. $$
1. The criterion

\[ \int_B (J - F) \, dx = \int_B -F \, dx + \sum_{\alpha} J_{\alpha} \int_B x^\alpha \, dx \]

unknown but constant\hspace{2in} easy to compute

is \textbf{LINEAR} in the coefficients \( J_{\alpha} \) of the unknown polynomial \( J \in \mathbb{R}[x]_k \)!

2. The constraint

\[ J(x) - f(x, y) = \sum_{j=0}^{m} \sigma_j(x, y) g_j(x, y) \]

is just \textbf{LINEAR CONSTRAINTS} + LMIs!
1. The criterion

\[ \int_B (J - F) \, dx = \int_B -F \, dx + \sum_{\alpha} J_\alpha \int_B x^\alpha \, dx \]

unknown but constant

is LINEAR in the coefficients \( J_\alpha \) of the unknown polynomial \( J \in \mathbb{R}[x]_k \)!

2. The constraint

\[ J(x) - f(x, y) = \sum_{j=0}^{m} \sigma_j(x, y) g_j(x, y) \]

is just LINEAR CONSTRAINTS + LMIs!
Hence, the optimization problem

\[ \rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_B (J - F) \, dx : J(x) - f(x, y) \in Q_k(g) \right\} \]

IS AN SDP! Moreover, it has an optimal solution \( J_k^* \in \mathbb{R}[x]_k \)!

- Alternatively, if one uses LP-based positivity certificates for \( J(x) - f(x, y) \), one ends up with solving an LP!

From the definition of \( J_k^* \), the sublevel sets

\[ \Theta_k := \{ x \in B : J_k^*(x) \leq 0 \} \subset R_f, \quad k \in \mathbb{N}, \]

provide a nested sequence of INNER approximations of \( R_f \).
Hence, the optimization problem

\[
\rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_{B} (J - F) \ dx : J(x) - f(x, y) \in Q_k(g) \right\}
\]

IS AN SDP! Moreover, it has an optimal solution \( J_k^* \in \mathbb{R}[x]_k \)!

• Alternatively, if one uses LP-based positivity certificates for \( J(x) - f(x, y) \), one ends up with solving an LP!

From the definition of \( J_k^* \), the sublevel sets

\[
\Theta_k := \{ x \in B : J_k^*(x) \leq 0 \} \subset R_f, \quad k \in \mathbb{N},
\]

provide a nested sequence of INNER approximations of \( R_f \).
Hence, the optimization problem

$$\rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_B (J - F) \, dx : J(x) - f(x, y) \in Q_k(g) \right\}$$

is an SDP! Moreover, it has an optimal solution $J^*_k \in \mathbb{R}[x]_k$!

- Alternatively, if one uses LP-based positivity certificates for $J(x) - f(x, y)$, one ends up with solving an LP!

From the definition of $J^*_k$, the sublevel sets

$$\Theta_k := \{ x \in B : J^*_k(x) \leq 0 \} \subset R_f, \quad k \in \mathbb{N}$$

provide a nested sequence of inner approximations of $R_f$. 
**Theorem (Lass)**

*(Strong) convergence in $L_1(B)$-norm takes place, that is:

$$\lim_{k \to \infty} \int_B |J_k^* - F| \, dx = 0$$

and, if in addition the set $\{x \in B : F(x) = 0\}$ has Lebesgue measure zero, then

$$\lim_{k \to \infty} \text{VOL}(R_f \setminus \Theta_k) = 0$$
Theorem (Lass)

(Strong) convergence in $L_1(\mathcal{B})$-norm takes place, that is:

$$\lim_{k \to \infty} \int_{\mathcal{B}} |J_k^* - F| \, dx = 0$$

and, if in addition the set \( \{x \in \mathcal{B} : F(x) = 0\} \) has Lebesgue measure zero, then

$$\lim_{k \to \infty} \text{VOL}(R_f \setminus \Theta_k) = 0$$
Let \( x \mapsto A(x) \in \mathbb{R}^{p \times p} \) where \( A(x) \) is the matrix-polynomial

\[
x \mapsto A(x) = \sum_{\alpha \in \mathbb{N}^n} A_\alpha \, x^\alpha = \sum_{\alpha \in \mathbb{N}^n} A_\alpha \, x_1^{\alpha_1} \cdots x_n^{\alpha_n}.
\]

for finitely many real symmetric matrices \((A_\alpha), \alpha \in \mathbb{N}^n\).

... and suppose one wants to approximate the set

\[
R_A := \{ x \in B : A(x) \succeq 0 \} = \{ x : \lambda_{\text{min}}(A(x)) \geq 0 \}.
\]

Then:

\[
R_A = \left\{ x \in B : \underbrace{y^T A(x) y \geq 0, \quad \forall y \text{ s.t. } \|y\|^2 = 1}_{f(x,y)} \right\}
\]
Let \( x \mapsto A(x) \in \mathbb{R}^{p \times p} \) where \( A(x) \) is the matrix-polynomial

\[
x \mapsto A(x) = \sum_{\alpha \in \mathbb{N}^n} A_\alpha \, x^\alpha = \sum_{\alpha \in \mathbb{N}^n} A_\alpha \, x_1^{\alpha_1} \cdots x_n^{\alpha_n}.
\]

for finitely many real symmetric matrices \((A_\alpha), \alpha \in \mathbb{N}^n\).

... and suppose one wants to approximate the set

\[
R_A := \{ x \in B : A(x) \succeq 0 \} = \{ x : \lambda_{\text{min}}(A(x)) \geq 0 \}.
\]

Then:

\[
R_A = \left\{ x \in B : \underbrace{y^T A(x) y}_{f(x,y)} \geq 0, \quad \forall \, y \text{ s.t. } \|y\|^2 = 1 \right\}
\]
Let $B$ be the unit disk $\{x : \|x\| \leq 1\}$ and let:

$$R_A := \left\{ x \in B : A(x) \begin{bmatrix} 1 - 16x_1x_2 & x_1 \\ x_1 & 1 - x_1^2 - x_2^2 \end{bmatrix} \succeq 0 \right\}$$

Then by solving relatively simple semidefinite programs, one may approximate $R_A$ with sublevel sets of the form:

$$\Theta_k := \{ x \in B : J_k^*(x) \geq 0 \}$$

for some polynomial $J_k^*$ of degree $k = 2, 4, \ldots$ and with

$$\text{VOL} (R_A \setminus \Theta_k) \to 0 \quad \text{as} \quad k \to \infty.$$
Let $\mathcal{B}$ be the unit disk $\{ x : \| x \| \leq 1 \}$ and let:

$$R_A := \left\{ x \in \mathcal{B} : A(x) \begin{bmatrix} 1 - 16x_1x_2 \\ x_1 \\ 1 - x_1^2 - x_2^2 \end{bmatrix} \succeq 0 \right\}$$

Then by solving relatively simple semidefinite programs, one may approximate $R_A$ with sublevel sets of the form:

$$\Theta_k := \left\{ x \in \mathcal{B} : J_k^*(x) \geq 0 \right\}$$

for some polynomial $J_k^*$ of degree $k = 2, 4, \ldots$ and with

$$\text{VOL} (R_A \setminus \Theta_k) \to 0 \quad \text{as} \quad k \to \infty.$$
$\Theta_2$ (left) and $\Theta_4$ (right) inner approximations (light gray) of (dark gray) embedded in unit disk $B$ (dashed).
$\Theta_6$ (left) and $\Theta_8$ (right) inner approximations (light gray) of (dark gray) embedded in unit disk $B$ (dashed).
A typical MINLP is of the form:

$$\mathbf{P} : \inf_x \{ f(x) : x \in \mathbf{K} \cap \mathbf{B}; x_i \in \{0, 1\}, \forall i \in I \},$$

where

- $\mathbf{K} := \{x : g_j(x) \leq 0, j = 1, \ldots, m\}$ for some $(g_j) \subset \mathbb{R}[x]$,
- $f$ is a (low) degree-$d$ non convex polynomial ($d \leq 4$).
- $\mathbf{B}$ is a box, e.g., $[0, 1]^n$

hence very difficult to solve!
A typical MINLP is of the form:

\[ P : \inf_x \{ f(x) : x \in K \cap B; x_i \in \{0, 1\}, \forall i \in I \}, \]

where

- **K** := \{x : g_j(x) \leq 0, j = 1, \ldots, m\} for some \((g_j) \subset \mathbb{R}[x],\)
- **f** is a (low) degree-d non convex polynomial \((d \leq 4).\)
- **B** is a box, e.g., \([0, 1]^n\)

hence very difficult to solve!
In the context of large scale MINLP the most efficient & popular strategy is to use BRANCH & BOUND combined with efficient LOWER BOUNDING techniques used at each node of the search tree.

Typically, $f$ is a sum $\sum_k f_k$ where each $f_k$ “sees” only very few variables (say 3, 4). The same observation is true for each $g_j$ in the constraints:

Hence a very appealing idea is to pre-compute CONVEX UNDERESTIMATORS $\hat{f}_k \leq f_k$ and $\hat{g}_j \leq g_j$ for each non convex $f_k$ and each non convex $g_j$, independently and separately!

→ hence potentially many BUT LOW-DIMENSIONAL problems.
• In the context of large scale MINLP the most efficient & popular strategy is to use BRANCH & BOUND combined with efficient LOWER BOUNDING techniques used at each node of the search tree.

• Typically, $f$ is a sum $\sum_k f_k$ where each $f_k$ “sees” only very few variables (say 3, 4). The same observation is true for each $g_j$ in the constraints:

Hence a very appealing idea is to pre-compute CONVEX UNDERESTIMATORS $\hat{f}_k \leq f_k$ and $\hat{g}_j \leq g_j$ for each non convex $f_k$ and each non convex $g_j$, independently and separately!

→ hence potentially many BUT LOW-DIMENSIONAL problems.
• In the context of large scale MINLP the most efficient & popular strategy is to use BRANCH & BOUND combined with efficient LOWER BOUNDING techniques used at each node of the search tree.

• Typically, $f$ is a sum $\sum_k f_k$ where each $f_k$ "sees" only very few variables (say 3, 4). The same observation is true for each $g_j$ in the constraints:

Hence a very appealing idea is to pre-compute **CONVEX UNDERESTIMATORS** $\hat{f}_k \leq f_k$ and $\hat{g}_j \leq g_j$ for each non convex $f_k$ and each non convex $g_j$, independently and separately!

→ hence potentially many **BUT LOW-DIMENSIONAL** problems.
Then at each node of the search tree of the B & B tree one computes a **LOWER BOUND** by solving convex optimization problems the form:

\[
\hat{P} : \inf_{x} \left\{ \sum_{k} \hat{f}_k(x) : \hat{g}_j(x) \leq 0, \; j = 1, \ldots, m; \; x \in B \right\}
\]

where some of the integer variables are fixed at 0 or 1.

and then one explores the search tree.
Hence one has to solve the generic problem

Compute a "tight" convex polynomial underestimator $p \leq f$ of a non convex polynomial $f$ on a box $B \subset \mathbb{R}^n$.

Message:

“Good" CONVEX POLYNOMIAL UNDERESTIMATORS can be computed efficiently!
Hence one has to solve the generic problem

Compute a "tight" convex polynomial underestimator $p \leq f$ of a non convex polynomial $f$ on a box $B \subset \mathbb{R}^n$.

Message:

“Good” CONVEX POLYNOMIAL UNDERESTIMATORS can be computed efficiently!
I: Characterizing convex polynomial underestimators

1. \( p(x) \leq f(x) \) for every \( x \in B \).

2. \( p \) convex on \( B \) \( \Rightarrow \nabla^2 p(x) \succeq 0 \) for all \( x \in B \),

\[ \iff \quad u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U, \]

where \( U := \{ u : \|u\|^2 \leq 1 \} \).

Hence we have the two "Positivity constraints"

\[ f(x) - p(x) \geq 0, \quad \forall x \in B \]
\[ u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U. \]
I: Characterizing convex polynomial underestimators

1. \( p(x) \leq f(x) \) for every \( x \in B \).

2. \( p \) convex on \( B \) \( \implies \nabla^2 p(x) \succeq 0 \) for all \( x \in B \).

\[ \iff \quad u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U, \]

where \( U := \{ u : \|u\|^2 \leq 1 \} \).

Hence we have the two "Positivity constraints"

\[ f(x) - p(x) \geq 0, \quad \forall x \in B \]
\[ u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U. \]
I: Characterizing convex polynomial underestimators

1. \( p(x) \leq f(x) \) for every \( x \in B \).

2. \( p \) convex on \( B \rightarrow \nabla^2 p(x) \succeq 0 \) for all \( x \in B \),

   \[ \iff u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U, \]

   where \( U := \{ u : \|u\|^2 \leq 1 \} \).

Hence we have the two "Positivity constraints"

\[
\begin{align*}
    f(x) - p(x) & \geq 0, \quad \forall x \in B \\
    u^T \nabla^2 p(x) u & \geq 0, \quad \forall (x, u) \in B \times U.
\end{align*}
\]
II: Characterizing "tightness"

One possibility is to evaluate the $L_1$-norm $\int_B |f(x) - p(x)| \, dx$

$$\rightarrow \int_B (f(x) - p(x)) \, dx = \int_B f(x) \, dx - \int_B p(x) \, dx$$

Indeed, writing $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_{\alpha} x^\alpha$, we have

$$\int_B p(x) \, dx = \sum_{\alpha \in \mathbb{N}^n} p_{\alpha} \int_B x^\alpha \, dx,$$

where $\gamma_\alpha$ is known (and easy to compute)!
II: Characterizing "tightness"

One possibility is to evaluate the $L_1$-norm $\int_B |f(x) - p(x)| \, dx$

$$\rightarrow \int_B (f(x) - p(x)) \, dx = \int_B f(x) \, dx - \int_B p(x) \, dx$$

Indeed, writing $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x^\alpha$,

$$\int_B p(x) \, dx = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \int_B x^\alpha \, dx,$$

where $\gamma_\alpha$ is known (and easy to compute)!
II: Characterizing "tightness"

One possibility is to evaluate the $L_1$-norm $\int_B |f(x) - p(x)| \, dx$

$$\rightarrow \int_B (f(x) - p(x)) \, dx = \int_B f(x) \, dx - \int_B p(x) \, dx$$

Indeed, writing $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x^\alpha$,

$$\int_B p(x) \, dx = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \int_B x^\alpha \, dx,$$

where $\gamma_\alpha$ is known (and easy to compute)!
Hence computing the best degree-$d$ convex polynomial underestimator of $f$ reduces to solve the CONVEX optimization problem:

\[
P: \quad \rho = \inf_{p \in \mathbb{R}[x]_d} \sum_{\alpha \in \mathbb{N}_d^n} p_\alpha \gamma_\alpha \\
\text{s.t.} \quad f(x) - p(x) \geq 0, \quad \forall x \in B \\
\quad \quad \quad u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U.
\]

which has an optimal solution $p^* \in \mathbb{R}[x]_d$
Hence we replace $P$ with the HIERARCHY of SEMIDEFINITE PROGRAMS

$$\rho_\ell = \inf_{p \in \mathbb{R}[x]_d} \sum_{\alpha \in \mathbb{N}_d^n} p_\alpha \gamma_\alpha$$

subject to

$$f - p = \sigma_0(x) + \sum_{j=1}^{m} \sigma_j(x)g_j(x) + \sigma_{m+1}(M - \|x\|^2)$$

$$u^T \nabla^2 p(x) u = \psi_0(x, u) + \sum_{j=1}^{m} \psi_j(x, u)g_j(x)$$

$$+ \psi_{m+1}(x, u)(M - \|x\|^2) + \psi_{m+2}(x, u)(1 - \|u\|^2);$$

$$\text{deg}(\sigma_j, \psi_j) \leq 2\ell,$$

parametrized by $\ell$.

→ and optimal solution $p^*_\ell \in \mathbb{R}[x]_d$. 
Hence we replace $P$ with the **HIERARCHY of SEMIDEFINITE PROGRAMS**

$$\rho_\ell = \inf_{p \in \mathbb{R}[x]_d} \sum_{\alpha \in \mathbb{N}^n_d} p_\alpha \gamma_\alpha$$

s.t.  
$$f - p = \sigma_0(x) + \sum_{j=1}^{m} \sigma_j(x)g_j(x) + \sigma_{m+1}(M - \|x\|^2)$$

$$u^T \nabla^2 \rho(x) u = \psi_0(x, u) + \sum_{j=1}^{m} \psi_j(x, u)g_j(x)$$

$$+ \psi_{m+1}(x, u)(M - \|x\|^2) + \psi_{m+2}(x, u)(1 - \|u\|^2);$$

$$\deg(\sigma_j, \psi_j) \leq 2\ell,$$

parametrized by $\ell$.

→ and optimal solution $p_\ell^* \in \mathbb{R}[x]_d$. 

Jean B. Lasserre  semidefinite characterization
Theorem (Lass & T. Phan Thanh (JOGO 2013))

\[ \rho_\ell \rightarrow \rho \text{ and } p^*_\ell \rightarrow p^* \in \mathbb{R}[x]_d, \text{ as } \ell \rightarrow \infty \]

→ Provides the best results in the comparison:

III. Applications in probability

Let $K \subseteq \mathbb{R}^n$, $S \subset K$ be Borel subsets, and $\Gamma \subset \mathbb{N}^n$.

Finding an upper bound (if possible optimal) on $\mathbb{P}(X \in S)$, the probability that a $K$-valued random variable $X \in S$, given some of its moments $\gamma = \{\gamma_\alpha\}$, $\alpha \in \Gamma \subset \mathbb{N}^n$.

.... is equivalent to solving:

\[
\sup_{\mu \in M(K)} \{ \mu(S) \mid \int_K X^\alpha \, d\mu = \gamma_\alpha, \quad \alpha \in \Gamma \}
\]

- $M(K)$ is the (convex) set of probability measures on $K \subseteq \mathbb{R}^n$.
- $f_\alpha \equiv X^\alpha$, $\alpha \in \Gamma$ (polynomial); $f_0 = I_S$ (piecewise polynomial)
Let $K \subseteq \mathbb{R}^n$, $S \subset K$ be Borel subsets, and $\Gamma \subset \mathbb{N}^n$.

Finding an upper bound (if possible optimal) on $\text{Prob}(X \in S)$, the probability that a $K$-valued random variable $X \in S$, given some of its moments $\gamma = \{\gamma_\alpha\}$, $\alpha \in \Gamma \subset \mathbb{N}^n$ ....

.... is equivalent to solving:

$$\sup_{\mu \in M(K)} \{ \mu(S) \mid \int_K X^\alpha \, d\mu = \gamma_\alpha, \quad \alpha \in \Gamma \}$$

- $M(K)$ is the (convex) set of probability measures on $K \subseteq \mathbb{R}^n$.
- $f_\alpha \equiv X^\alpha$, $\alpha \in \Gamma$ (polynomial); $f_0 = I_S$ (piecewise polynomial).
Let $K \subseteq \mathbb{R}^n$, $S \subset K$ be Borel subsets, and $\Gamma \subset \mathbb{N}^n$.

Finding an upper bound (if possible optimal) on $\text{Prob}(X \in S)$, the probability that a $K$-valued random variable $X \in S$, given some of its moments $\gamma = \{\gamma_\alpha\}$, $\alpha \in \Gamma \subset \mathbb{N}^n$....

.... is equivalent to solving:

$$\sup_{\mu \in M(K)} \left\{ \mu(S) \mid \int_K f_\alpha \, d\mu = \gamma_\alpha, \; \alpha \in \Gamma \right\}$$

- $M(K)$ is the (convex) set of probability measures on $K \subseteq \mathbb{R}^n$.
- $f_\alpha \equiv X^\alpha$, $\alpha \in \Gamma$ (polynomial); $f_0 = I_S$ (piecewise polynomial)
Notice that writing

\[ \mu = \nu + \varphi, \quad \text{with} \]

- \( \varphi \) supported on \( S \), and
- \( \nu \) supported on \( K \setminus S \),

equivalently, one has to solve

\[
P : \sup_{\nu, \varphi} \left\{ \int_S 1 \ d\varphi \mid \int_{K \setminus S} X^\alpha \ d\nu + \int_S X^\alpha \ d\varphi = \gamma_\alpha, \quad \alpha \in \Gamma \right\}
\]

(since we maximize \( \varphi(S) \), one may take \( \nu \) supported on \( K \).)
Assume that $\Gamma \subset \mathbb{N}^n_d$. Then the dual of $P$ reads:

$$
P^*: \inf_{p_\alpha} \left\{ \sum_{\alpha \in \Gamma} p_{\alpha} \gamma_{\alpha} : \ p \geq 1 \text{ on } S; \quad p \geq 0 \text{ on } K \right\}
$$

where $p \in \mathbb{R}[x]_d$ is a polynomial

$$
x \mapsto p(x) = \sum_{\alpha \in \mathbb{N}^n_d} p_{\alpha} x^\alpha; \quad p_{\alpha} = 0 \quad \forall \alpha \in \mathbb{N}^n_d \setminus \Gamma.
$$

Let $K$ and $S \subset K$ be compact semi-algebraic sets:

$$
K = \{x \in \mathbb{R}^n : h_k(x) \geq 0, \quad k = 1, \ldots, p\}
$$

$$
S = \{x \in \mathbb{R}^n : g_j(x) \geq 0, \quad j = 1, \ldots, m\}
$$
Assume that $\Gamma \subset \mathbb{N}^n_d$. Then the dual of $P$ reads:

\[
P^* = \inf_{p_\alpha} \left\{ \sum_{\alpha \in \Gamma} p_\alpha \gamma_\alpha : \quad p \geq 1 \text{ on } S; \quad p \geq 0 \text{ on } K \right\}
\]

where $p \in \mathbb{R}[x]_d$ is a polynomial

\[
x \mapsto p(x) = \sum_{\alpha \in \mathbb{N}^n_d} p_\alpha x^\alpha; \quad p_\alpha = 0 \quad \forall \alpha \in \mathbb{N}^n_d \setminus \Gamma.
\]

Let $K$ and $S \subset K$ be compact semi-algebraic sets:

\[
K = \{ x \in \mathbb{R}^n : h_k(x) \geq 0, \quad k = 1, \ldots, p \}
\]

\[
S = \{ x \in \mathbb{R}^n : g_j(x) \geq 0, \quad j = 1, \ldots, m \}
\]
Again the basic idea of the moment-SOS approach is to replace the positivity constraints

\[ p - 1 \geq 0 \text{ on } S; \quad p \geq 0 \text{ on } K \]

with the SOS-positivity certificates

\[
\begin{align*}
p(x) - 1 &= \sigma_0(x) + \sum_{k=1}^{p} \sigma_k(x) h_k(x) \\
p(x) &= \psi_0(x) + \sum_{j=1}^{m} \psi_j(x) g_j(x)
\end{align*}
\]

for some SOS polynomial \((\sigma_k) \subset \mathbb{R}[x]_t\) and \((\psi_j) \subset \mathbb{R}[x]_t\),

and let \(t\) increase.
Again the basic idea of the moment-SOS approach is to replace the positivity constraints

\[ p - 1 \geq 0 \text{ on } S; \quad p \geq 0 \text{ on } K \]

with the SOS-positivity certificates

\[ p(x) - 1 = \sigma_0(x) + \sum_{k=1}^{p} \sigma_k(x) h_k(x) \]

\[ p(x) = \psi_0(x) + \sum_{j=1}^{m} \psi_j(x) g_j(x) \]

for some SOS polynomial \( (\sigma_k) \subset \mathbb{R}[x]_t \) and \( (\psi_j) \subset \mathbb{R}[x]_t \), and let \( t \) increase.
Again the basic idea of the moment-SOS approach is to replace the positivity constraints

\[ p - 1 \geq 0 \text{ on } S; \quad p \geq 0 \text{ on } K \]

with the SOS-positivity certificates

\[
\begin{align*}
p(x) - 1 &= \sigma_0(x) + \sum_{k=1}^{p} \sigma_k(x) h_k(x) \\
p(x) &= \psi_0(x) + \sum_{j=1}^{m} \psi_j(x) g_j(x)
\end{align*}
\]

for some SOS polynomial \((\sigma_k) \subset \mathbb{R}[x]_t\) and \((\psi_j) \subset \mathbb{R}[x]_t\),

and let \(t\) increase.
One ends up in solving the hierarchy of semidefinite programs indexed by \( t \in \mathbb{N} \):

\[
\rho_t = \min_{p \in \mathbb{R}[x]_d} \sum_{\alpha \in \Gamma} p_{\alpha} \gamma_{\alpha} : \quad \text{subject to:}
\]

\[
p_{\alpha} = 0, \quad \forall \alpha \in \mathbb{N}^n_d \setminus \Gamma
\]

\[
p(x) - 1 = \sigma_0(x) + \sum_{k=1}^{p} \sigma_k(x) h_k(x)
\]

\[
p(x) = \psi_0(x) + \sum_{j=1}^{m} \psi_j(x) g_j(x)
\]

\( \sigma_k, \psi_j \) SOS; \quad \text{deg}(\sigma_k h_k) \leq 2t; \quad \text{deg}(\psi_j g_j) \leq 2t;\]
Let $S \subset \mathbb{R}^n$ be a compact basic semi-algebraic set. Let $K$ be a BOX $[0, a]^n$ containing $S$ and let:

$$
\gamma_\alpha = \int_K X^\alpha \, dx = \frac{a^{n+|\alpha|}}{\prod_{k=1}^n (1 + \alpha_k)!}, \quad \forall \alpha \in \mathbb{N}^n
$$

**Theorem**

The (Lebesgue) volume of the set $S$ is obtained as:

$$
\sup_{\nu, \varphi} \left\{ \int_S 1 \, d\varphi \mid \int_S X^\alpha \, d\varphi + \int_K X^\alpha \, d\nu = \gamma_\alpha, \; \alpha \in \mathbb{N}^n \right\}
$$
Let $S \subset \mathbb{R}^n$ be a compact basic semi-algebraic set. Let $K$ be a BOX $[0, a]^n$ containing $S$ and let:

$$
\gamma_\alpha = \int_K X^\alpha \, dx = \frac{a^{n+|\alpha|}}{\prod_{k=1}^n (1 + \alpha_k)!}, \quad \forall \alpha \in \mathbb{N}^n
$$

**Theorem**

The (Lebesgue) volume of the set $S$ is obtained as:

$$
\sup_{\nu, \varphi} \left\{ \int_S 1 \, d\varphi \mid \int_S X^\alpha \, d\varphi + \int_K X^\alpha \, d\nu = \gamma_\alpha, \quad \forall \alpha \in \mathbb{N}^n \right\}
$$
Let $K_j \subset \mathbb{R}^{n_j}, j = 1, \ldots, p$, and $K := K_1 \times K_2 \cdots \times K_p \subset \mathbb{R}^n$, and with natural projections $\pi_j : K \rightarrow K_j, j = 1, \ldots, p$.

Let $\nu_j$ be a given Borel measure on $K_j, j = 1, \ldots, p$.

For a measure $\mu$ on $K$, denote $\pi_j \mu$ its marginal on $K_j$, i.e.

$$\pi_j \mu(B) := \mu(\pi_j^{-1}(B)) = \mu(\{x \in K : \pi_j x \in B\}), \quad B \in \mathcal{B}(K_j)$$

Consider the optimization problem:

$$f^* = \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu \mid \pi_j \mu = \nu_j, \quad j = 1, \ldots, p \right\}$$

with $M(K)$ being the set of finite Borel measures on $K$. 
Let $K_j \subset \mathbb{R}^{n_j}, j = 1, \ldots, p$, and $K := K_1 \times K_2 \cdots \times K_p \subset \mathbb{R}^n$, and with natural projections $\pi_j : K \rightarrow K_j, j = 1, \ldots, p$.

Let $\nu_j$ be a given Borel measure on $K_j, j = 1, \ldots, p$.

For a measure $\mu$ on $K$, denote $\pi_j \mu$ its marginal on $K_j$, i.e.

$$\pi_j \mu(B) := \mu(\pi_j^{-1}(B)) = \mu(\{x \in K : \pi_j x \in B\}), \quad B \in \mathcal{B}(K_j)$$

Consider the optimization problem:

$$f^* = \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu \mid \pi_j \mu = \nu_j, \quad j = 1, \ldots, p \right\}$$

with $M(K)$ being the set of finite Borel measures on $K$. 
Generalization of the famous Monge-Kantorovich transportation problem, with many other interesting applications, particularly in Probability.

If $K_j$ is compact then the constraint on marginals

$$\pi_j \mu = \nu_j$$

is equivalent to the countably many linear equalities

$$\int_K X^\alpha \, d\mu = \int_{K_j} X^\alpha \, d\nu_j, \quad \forall \alpha \in \mathbb{N}^{n_j}$$

between moments of $\mu$ and $\nu_j$ ...

because the space of polynomials is dense (for the sup-norm) in the space $C(K_j)$ of continuous functions on $K_j$. 

Jean B. Lasserre

semidefinite characterization
Hence for each $d \in \mathbb{N}$, we may consider the truncated version

$$\rho_d = \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu : \int_K X^\alpha \, d\mu = \int_{K_j} X^\alpha \, d\nu_j, \quad \alpha \in \mathbb{N}_d^n; j = 1, \ldots, p \right\}$$

with dual

$$\rho_d^* = \sup_{p_j \in \mathbb{R}[X_j_d]} \left\{ \sum_{j=1}^p \int_{K_j} p_j \, d\nu_j : f(X) - \sum_{j=1}^p p_j(X_j) \geq 0 \quad \text{on } K \right\}$$

$$= \sum_{\alpha \in \mathbb{N}^n_d} p_{j\alpha} \gamma_{j\alpha}$$
Hence for each \( d \in \mathbb{N} \), we may consider the truncated version

\[
\rho_d = \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu : \int_K X^\alpha \, d\mu = \int_{K_j} X^\alpha \, d\nu_j, \quad \alpha \in \mathbb{N}^n_d; j = 1, \ldots, p \right\}
\]

with dual

\[
\rho_d^* = \sup_{p_j \in \mathbb{R}[X_j]_d} \left\{ \sum_{j=1}^{p} \int_{K_j} p_j \, d\nu_j : f(X) - \sum_{j=1}^{p} p_j(X_j) \geq 0 \quad \text{on } K \right\}
\]

\[= \sum_{\alpha \in \mathbb{N}^n_d} p_{j\alpha} \gamma_{j\alpha} \]
Then if $K \subset \mathbb{R}^n$ is compact and in the form

$$K = \{ x \in \mathbb{R}^n : g_j(X) \geq 0, \quad j = 1, \ldots, m \},$$

one replaces the positivity-on-$K$ constraint

$$f(X) - \sum_{j=1}^{p} p_j(X_j) \geq 0 \quad \text{on } K$$

with the SOS-based Putinar's certificate

$$f(X) - \sum_{j=1}^{p} p_j(X_j) = \sigma_0(X) + \sum_{j=1}^{m} \sigma_j(X) g_j(X),$$

with SOS-weights $(\sigma_j) \subset \mathbb{R}[X]$. 

Jean B. Lasserre  
semidefinite characterization
... and solve the hierarchy of semidefinite programs

\[
\theta_d = \sup_{p_j \in \mathbb{R}[X_j]_d} \left\{ \sum_{j=1}^p \sum_{\alpha \in \mathbb{N}_d^{n_j}} p_{j\alpha} \gamma_{j\alpha} : \right. \\
\left. \text{s.t.} \quad f(X) - \sum_{j=1}^p p_j(X_j) = \sigma_0(X) + \sum_{j=1}^m \sigma_j(X) g_j(X) \right\},
\]

with SOS-weights \((\sigma_j)\) of degree at most 2\(d\).

One may prove that \(\rho_d \to f^*\) as \(d \to \infty\).
... and solve the hierarchy of semidefinite programs

\[ \theta_d = \sup_{p_j \in \mathbb{R}[X_j]} \left\{ \sum_{j=1}^{p} \sum_{\alpha \in \mathbb{N}_d} p_{j\alpha} \gamma_{j\alpha} : \right\} \]

s.t. \[ f(X) - \sum_{j=1}^{p} p_j(X_j) = \sigma_0(X) + \sum_{j=1}^{m} \sigma_j(X) g_j(X) \}

with SOS-weights \((\sigma_j)\) of degree at most 2d.

One may prove that \(\rho_d \to f^*\) as \(d \to \infty\).
Let $\mu$ be the Gaussian measure on $\mathbb{R}^n$ with density $x \mapsto \exp(-\|x\|^2)$ and let $K \subset \mathbb{R}^n$ be the non necessarily compact basic semi-algebraic set

$$K = \{ x \in \mathbb{R}^n : g_j(X) \geq 0, \quad j = 1, \ldots, m \}.$$

**Goal:** Approximate $\mu(K)$ as closely as desired

Can be difficult even in small dimension $n = 2, 3$. 
Let $\mu$ be the Gaussian measure on $\mathbb{R}^n$ with density $x \mapsto \exp(-\|x\|^2)$ and let $K \subset \mathbb{R}^n$ be the non necessarily compact basic semi-algebraic set

$$K = \{ x \in \mathbb{R}^n : g_j(X) \geq 0, \quad j = 1, \ldots, m \}.$$

Goal:

Approximate $\mu(K)$ as closely as desired

Can be difficult even in small dimension $n = 2, 3$. 

Jean B. Lasserre

semidefinite characterization
Theorem (Lass 2015)

Let \( f \in \mathbb{R}[x] \) be strictly positive \( \mu \)-a.e. on \( K \), and let \( M(K) \) (resp. \( M(\mathbb{R}^n) \)) be the space of finite Borel measures on \( K \) (resp. \( \mathbb{R}^n \)).

Then the optimization problem:

\[
f_1^* = \sup_{\nu, \phi} \left\{ \int_K f \, d\phi : \phi + \nu = \mu; \, \phi \in M(K), \, \nu \in M(\mathbb{R}^n) \right\},
\]

has a unique optimal solution \((\phi^*, \nu^*) = (\mu_K, \mu - \mu_K)\) where \( \mu_K \) is the restriction of \( \mu \) to \( K \), that is:

\[
\phi^*(B) = \mu_K(B) = \mu(K \cap B), \quad \forall B \in \mathcal{B}(\mathbb{R}^n).
\]

In particular, \( \phi^*(K) = \mu(K) \), and \( f^* = \mu(K) \) if \( f = 1 \).
Proof

From $\phi + \nu = \mu$ one deduces $\phi \leq \mu$ and therefore

$$f^* \leq \int_{\mathbf{K}} f \, d\mu = \int_{\mathbf{K}} f \, d\mu_{\mathbf{K}}.$$ 

On the other hand the pair $(\phi^*, \nu^*) = (\mu_{\mathbf{K}}, \mu - \mu_{\mathbf{K}})$ is a feasible solution with associated cost

$$\int_{\mathbf{K}} f \, d\phi^* = \int_{\mathbf{K}} f \, d\mu_{\mathbf{K}},$$

which proves the optimality of $(\phi^*, \nu^*)$.

Uniqueness is more delicate. Assume there is another optimal solution $\phi, \nu)$. From $\phi \leq \mu$ one deduces $\phi \ll \mu$ and so by Radon-Nykodim

$$\phi(B \cap \mathbf{K}) = \int_{B \cap \mathbf{K}} g \, d\mu \leq \int_{B \cap \mathbf{K}} d\mu, \quad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

for some nonnegative measurable function $g$. Hence $g \leq 1$, $\mu$-a.e. on $\mathbf{K}$. 

Jean B. Lasserre  
semidefinite characterization
Proof

From $\phi + \nu = \mu$ one deduces $\phi \leq \mu$ and therefore

$$f^* \leq \int_K f \, d\mu = \int f \, d\mu_K.$$ 

On the other hand the pair $(\phi^*, \nu^*) = (\mu_K, \mu - \mu_K)$ is a feasible solution with associated cost

$$\int_K f \, d\phi^* = \int f \, d\mu_K,$$

which proves the optimality of $(\phi^*, \nu^*)$.

**Uniqueness** is more delicate. Assume there is another optimal solution $\phi, \nu$). From $\phi \leq \mu$ one deduces $\phi \ll \mu$ and so by Radon-Nykodim

$$\phi(B \cap K) = \int_{B \cap K} g \, d\mu \leq \int_{B \cap K} d\mu, \quad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

for some nonnegative measurable function $g$. Hence $g \leq 1$, $\mu$-a.e. on $K$. 

Jean B. Lasserre  

semidefinite characterization
On the other hand, by optimality of $\phi^*$ and $\phi$,

$$f^* = \int_K f \, d\mu = \int f \, d\phi^* = \int f \, d\phi = \int f \, g \, d\mu$$

which implies

$$0 = \int_K f(1 - g) \, d\mu,$$

Combining this with $f > 0$ and $g \leq 1$ $\mu$-a.e. on $K$, yields $g = 1$, $\mu$-a.e. on $K$.

This yields the desired result that $\phi = \phi^*$. □
A possible dual for the above LP is the LP:

\[
\rho^* = \inf_{p \in \mathbb{R}[x]} \left\{ \int_K p \, d\mu : p \geq f \text{ on } K; \ p \geq 0 \text{ on } \mathbb{R}^n \right\},
\]

Indeed it trivially holds that \( \rho^* \geq f^* \).

A tractable version is obtained by replacing:

- the "hard" positivity constraint \( p - f \geq 0 \text{ on } K \), with the positivity-on-K certificate

\[
p - f = \sigma_0 + \sum_{j=1}^m \sigma_j g_j; \quad \sigma_j \text{ is SOS for all } j\}
\]

- the "hard" positivity constraint \( p \geq 0 \text{ on } \mathbb{R}^n \) with \( p \) is SOS.
A possible dual for the above LP is the LP:

\[
\rho^* = \inf_{p \in \mathbb{R}[x]} \left\{ \int_K p \, d\mu : p \geq f \text{ on } K; \, p \geq 0 \text{ on } \mathbb{R}^n \right\},
\]

Indeed it trivially holds that \( \rho^* \geq f^* \).

A tractable version is obtained by replacing:

- the "hard" positivity constraint \( p - f \geq 0 \text{ on } K \), with the positivity-on-\( K \) certificate

\[
p - f = \sigma_0 + \sum_{j=1}^{m} \sigma_j g_j; \quad \sigma_j \text{ is SOS for all } j\}
\]

- the "hard" positivity constraint \( p \geq 0 \text{ on } \mathbb{R}^n \) with \( p \text{ is SOS} \).
so as to obtain the hierarchy of semidefinite approximations indexed by $d \in \mathbb{N}$:

$$
\rho_d^* = \inf_{p \in \mathbb{R}[x]_d} \left\{ \int_{\mathbb{R}^n} p \, d\mu : p - f = \sigma_0 + \sum_{j=1}^m \sigma_j g_j; \quad p, \sigma_j \text{ all SOS} \right\}
$$

where the degree of the SOS $p, \sigma_j$ is bounded by $2d$.

**Theorem (Lass 2015)**

For every $d \in \mathbb{N}$, $\rho_d^* \geq f^*$ and $\rho_d^* \to f^*$ as $d \to \infty$. 
so as to obtain the hierarchy of semidefinite approximations
indexed by \( d \in \mathbb{N} \):

\[
\rho_d^* = \inf_{p \in \mathbb{R}[x]_d} \{ \int_{\mathbb{R}^n} p \, d\mu : p - f = \sigma_0 + \sum_{j=1}^{m} \sigma_j g_j; \quad p, \sigma_j \text{ all SOS} \}
\]

where the degree of the SOS \( p, \sigma_j \) is bounded by \( 2d \).

**Theorem (Lass 2015)**

*For every \( d \in \mathbb{N} \), \( \rho_d^* \geq f^* \) and \( \rho_d^* \to f^* \) as \( d \to \infty \).*
One may do the same for the complement $\mathbf{K}^c := \mathbb{R}^n \setminus \mathbf{K}$ as soon as one can write

$$
\mathbf{K}^c = \bigcup_{i=1}^{p} \Omega_i; \quad \mu(\Omega_i \cap \Omega_j) = 0 \quad \forall (i, j)
$$

so that $\mu(\mathbf{K}^c) = \sum_{i=1}^{p} \mu(\Omega_i)$. In doing so one obtains for each $i = 1, \ldots, p$ a sequence $(\theta_{id})_{d \in \mathbb{N}}$ such that

$$
\sum_{i=1}^{p} \theta_{id} \geq \mu(\mathbf{K}^c) \quad \text{and} \quad \lim_{d \to \infty} \sum_{i=1}^{p} \theta_{id} = \mu(\mathbf{K}^c) = \mu(\mathbb{R}^n) - \mu(\mathbf{K}).
$$

**Theorem (Lass 2015)**

With $f = 1$ one obtains $\mu(\mathbb{R}^n) - \sum_{j=1}^{p} \theta_{jd} \leq \mu(\mathbf{K}) \leq \rho^*_d$ for all $d$, and

$$
\lim_{d \to \infty} \omega^*_d = \mu(\mathbf{K}) = \lim_{d \to \infty} \rho^*_d.
$$
One may do the same for the complement $K^c := \mathbb{R}^n \setminus K$ as soon as one can write

$$K^c = \bigcup_{i=1}^{p} \Omega_i; \quad \mu(\Omega_i \cap \Omega_j) = 0 \quad \forall (i, j)$$

so that $\mu(K^c) = \sum_{i=1}^{p} \mu(\Omega_i)$. In doing so one obtains for each $i = 1, \ldots, p$ a sequence $(\theta_{id})_{d \in \mathbb{N}}$ such that

$$\sum_{i=1}^{p} \theta_{id} \geq \mu(K^c) \quad \text{and} \quad \lim_{d \to \infty} \sum_{i=1}^{p} \theta_{id} = \mu(K^c) = \mu(\mathbb{R}^n) - \mu(K).$$

**Theorem (Lass 2015)**

*With $f = 1$ one obtains $\mu(\mathbb{R}^n) - \sum_{j=1}^{p} \theta_{jd} \leq \mu(K) \leq \rho_d^*$ for all $d$, for all $d$.*

and

$$\lim_{d \to \infty} \omega_d^* = \mu(K) = \lim_{d \to \infty} \rho_d^*.$$
Examples

Let $n = 2$, and $d\mu = \exp(-\|x\|^2/\sigma) \, dx$ and let $K$ be the non-convex quadratic

$$
x \mapsto x^T Ax = 0.56 x_1^2 + 0.96 x_1 x_2 - 1.24 x_2^2.
$$

$$
K = \{(x, y) : (x - u)^T A (x - u) \leq 1\} \quad \text{(non-compact)},
$$

with $u = (0.1, 0.5)$ and $(0.5, 0.1)$.

<table>
<thead>
<tr>
<th>$u = (0.5, 0.1)$</th>
<th>$u = (0.1, 0.5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>$\rho^*_9$</td>
</tr>
<tr>
<td>1</td>
<td>2.829605</td>
</tr>
<tr>
<td>0.8</td>
<td>1.876731</td>
</tr>
<tr>
<td>1</td>
<td>2.989832</td>
</tr>
<tr>
<td>0.8</td>
<td>1.969188</td>
</tr>
</tbody>
</table>
Examples

Let $n=2$, and $d\mu = \exp(-\|x\|^2/\sigma)\,dx$ and let $K$ be the non-convex quadratic

$$x \mapsto x^T Ax = 0.56 x_1^2 + 0.96 x_1 x_2 - 1.24 x_2^2.$$ 

$$K = \{(x, y) : (x - u)^T A (x - u) \leq 1\} \quad \text{(non-compact)},$$

with $u = (0.1, 0.5)$ and $(0.5, 0.1)$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$\rho^*_9$</th>
<th>$\omega^*_9$</th>
<th>$100 (\rho^<em>_9 - \omega^</em>_9)/\omega^*_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.829605</td>
<td>2.824718</td>
<td>0.17%</td>
</tr>
<tr>
<td>0.8</td>
<td>1.876731</td>
<td>1.876609</td>
<td>0.006%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$u = (0.1, 0.5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0.8</td>
</tr>
</tbody>
</table>
More details and (non-compact) examples in arXiv:1508.06132.

Conclusion

- Provides a sequence of converging upper and lower bounds on $\mu(K)$ for non necessarily compact basic semi-algebraic sets $K$.
- A general methodology not set-$K$-dependent.
- Also works for the exponential measure on the positive orthant $\mathbb{R}_+^n$, and in fact any measure $\mu$ provided that it satisfies Carleman’s condition and one knows all its moments.
With rough basic implementation and present state-of-the-art SDP solvers, one can obtain a few upper and lower bounds only and for dimension $n = 2$ or $n = 3$. For $d \geq 15$ numerical accuracy problems show up.

- Some non-trivial tricks (based on Stokes’ formula) permit to improve the quality of bounds.
- Much remains to be done for a better implementation.
With rough basic implementation and present state-of-the-art SDP solvers, one can obtain a few upper and lower bounds only and for dimension $n = 2$ or $n = 3$. For $d \geq 15$ numerical accuracy problems show up.

Some non-trivial tricks (based on Stokes’ formula) permit to improve the quality of bounds.

Much remains to be done for a better implementation.
With rough basic implementation and present state-of-the-art SDP solvers, one can obtain a few upper and lower bounds only and for dimension $n = 2$ or $n = 3$. For $d \geq 15$ numerical accuracy problems show up.

Some non-trivial tricks (based on Stokes’ formula) permit to improve the quality of bounds.

Much remains to be done for a better implementation.
V. Lebesgue decomposition in action

Given two measures $\mu$ and $\nu$ on $\mathbb{R}^n$,

one would like to approximate the Lebesgue decomposition

$$\phi + \psi = \mu; \quad \phi \ll \nu; \quad \psi \perp \nu,$$

of $\mu$ with respect to $\nu$.

... based on the sole knowledge of the moments

$$y_\alpha = \int_{\mathbb{R}^n} x^\alpha \, d\mu, \quad z_\alpha = \int_{\mathbb{R}^n} x^\alpha \, d\nu, \quad \alpha \in \mathbb{N}^n.$$

of $\mu$ and $\nu$.
Given two measures $\mu$ and $\nu$ on $\mathbb{R}^n$, one would like to approximate the Lebesgue decomposition

$$\phi + \psi = \mu; \quad \phi \ll \nu; \quad \psi \perp \nu,$$

of $\mu$ with respect to $\nu$.

... based on the sole knowledge of the moments

$$y_\alpha = \int_{\mathbb{R}^n} x^\alpha \, d\mu, \quad z_\alpha = \int_{\mathbb{R}^n} x^\alpha \, d\nu, \quad \alpha \in \mathbb{N}^n.$$
By definition of $\phi$ and $\psi$:

$\phi$ has a **DENSITY** w.r.t. $\nu$ in $L_1(\nu)$ (called the **Radon-Nikodym** derivative of $\mu$ w.r.t. $\nu$). That is, there exists a nonnegative measurable function $f \in L_1(\nu)$ such that:

$$
\phi(A) = \int_A f(x) \, d\nu(x), \quad \forall A \in \mathcal{B}(\mathbb{R}^n).
$$
CLAIM: If one assumes that:

- $f$ is in $L_\infty(\nu)$ (instead of $L_1(\nu)$), and $\|f\|_\infty < M$ for some $M$,
- Both moment sequences $(y_\alpha)$ and $(z_\alpha), \alpha \in \mathbb{N}^n$ satisfy Carleman’s condition:

\[
+\infty = \sum_{k=1}^{\infty} \left( \int x_i^{2k} \, d\mu \right)^{-1/2k} = \sum_{k=1}^{\infty} \left( \int x_i^{2k} \, d\nu \right)^{-1/2k}
\]

for all $i = 1, \ldots, n$.

THEN ... one may approximate as closely as desired any fixed set of moments of $\phi$ and $\psi$. 
CLAIM: If one assumes that:

- \( f \) is in \( L_\infty(\nu) \) (instead of \( L_1(\nu) \)), and \( \|f\|_\infty < M \) for some \( M \),

- Both moment sequences \( (y_\alpha) \) and \( (z_\alpha) \), \( \alpha \in \mathbb{N}^n \) satisfy Carleman’s condition:

\[
+\infty = \sum_{k=1}^{\infty} \left( \int X_i^{2k} \ d\mu \right)^{-1/2k} = \sum_{k=1}^{\infty} \left( \int X_i^{2k} \ d\nu \right)^{-1/2k}
\]

for all \( i = 1, \ldots, n \).

THEN ... one may approximate as closely as desired any fixed set of moments of \( \phi \) and \( \psi \).
CLAIM: If one assumes that:

- $f$ is in $L_{\infty}(\nu)$ (instead of $L_1(\nu)$), and $\|f\|_{\infty} < M$ for some $M$,
- Both moment sequences $(y_{\alpha})$ and $(z_{\alpha})$, $\alpha \in \mathbb{N}^n$ satisfy Carleman’s condition:

$$+\infty = \sum_{k=1}^{\infty} \left( \int X_i^{2k} \, d\mu \right)^{-1/2k} = \sum_{k=1}^{\infty} \left( \int X_i^{2k} \, d\nu \right)^{-1/2k}$$

for all $i = 1, \ldots, n$.

THEN ... one may approximate as closely as desired any fixed set of moments of $\phi$ and $\psi$. 

Jean B. Lasserre

semidefinite characterization
Denote the moments of $\mu$ and $\nu$ by:
\[
\mu_\alpha = \int x^\alpha \, d\mu, \quad \nu_\alpha = \int x^\alpha \, d\nu, \quad \alpha \in \mathbb{N}^n.
\]

Let $\gamma > 0$ be fixed, and consider the hierarchy of semidefinite programs $P_d$ indexed by $d \in \mathbb{N}$:
\[
P_d : \quad \rho_d = \sup_{y, u, v} \ y_0
\]
subject to:
\[
\begin{align*}
y_\alpha + u_\alpha &= \mu_\alpha, & \forall \alpha \in \mathbb{N}^n_d \\
y_\alpha + v_\alpha &= \gamma \nu_\alpha, & \forall \alpha \in \mathbb{N}^n_d
\end{align*}
\]
\[
M_d(y), \ M_d(u), \ M_d(v) \succeq 0
\]
Let $\phi^*$ and $\psi^*$ be the Lebesgue decomposition of $\mu$ w.r.t. $\nu$, and let $f^* \in L_1(\nu)$ be the density of $\phi^*$ w.r.t. $\nu$.

**Theorem (Lass 2015)**

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution $(y^d, u^d, v^d)$.

(ii) Moreover as $d \to \infty$, the triplet of sequences $(y^d, u^d, v^d)$ converges to some triplet of sequences $(y^*, u^*, v^*)$ where

$$y^*_\alpha = \int x^\alpha (\gamma \wedge f^*) \, d\nu = \int x^\alpha f^*_\gamma \, d\nu, \quad \forall \alpha \in \mathbb{N}^n.$$ 

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$y^*_\alpha = \int x^\alpha \, d\phi, \quad \forall \alpha \in \mathbb{N}^n.$$
Let $\phi^*$ and $\psi^*$ be the Lebesgue decomposition of $\mu$ w.r.t. $\nu$, and let $f^* \in L_1(\nu)$ be the density of $\phi^*$ w.r.t. $\nu$.

**Theorem (Lass 2015)**

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution $(y^d, u^d, v^d)$.

(ii) Moreover as $d \to \infty$, the triplet of sequences $(y^d, u^d, v^d)$ converges to some triplet of sequences $(y^*, u^*, v^*)$ where

$$y^*_\alpha = \int x^\alpha (\gamma \wedge f^*) \, d\nu = \int x^\alpha f^*_\gamma \, d\nu, \quad \forall \alpha \in \mathbb{N}^n.$$

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$y^*_\alpha = \int x^\alpha \, d\phi, \quad \forall \alpha \in \mathbb{N}^n.$$
Let $\phi^*$ and $\psi^*$ be the Lebesgue decomposition of $\mu$ w.r.t. $\nu$, and let $f^* \in L_1(\nu)$ be the density of $\phi^*$ w.r.t. $\nu$.

**Theorem (Lass 2015)**

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution $(y^d, u^d, v^d)$.

(ii) Moreover as $d \to \infty$, the triplet of sequences $(y^d, u^d, v^d)$ converges to some triplet of sequences $(y^*, u^*, v^*)$ where

$$
y^*_\alpha = \int x^\alpha (\gamma \wedge f^*) d\nu = \int x^\alpha f^*_\gamma d\nu, \quad \forall \alpha \in \mathbb{N}^n.
$$

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$
y^*_\alpha = \int x^\alpha d\phi, \quad \forall \alpha \in \mathbb{N}^n.
$$
Let $\phi^*$ and $\psi^*$ be the Lebesgue decomposition of $\mu$ w.r.t. $\nu$, and let $f^* \in L_1(\nu)$ be the density of $\phi^*$ w.r.t. $\nu$.

**Theorem (Lass 2015)**

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution $(y^d, u^d, v^d)$.

(ii) Moreover as $d \to \infty$, the triplet of sequences $(y^d, u^d, v^d)$ converges to some triplet of sequences $(y^*, u^*, v^*)$ where

$$y^*_\alpha = \int x^\alpha (\gamma \land f^*) \, d\nu = \int x^\alpha f^*_\gamma \, d\nu, \quad \forall \alpha \in \mathbb{N}^n.$$  

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$y^*_\alpha = \int x^\alpha \, d\phi, \quad \forall \alpha \in \mathbb{N}^n.$$
Examples

Let $n = 2$, $p \in (0, 1)$ and

- $\nu$ is the Gaussian with density $x \mapsto \exp(-||x||^2)$,
- $\theta$ is the measure uniformly distributed on the circle $\{x : x_1^2 + x_2^2 = 1\}$.

Define the measure $\mu$ to be

$$\mu = p \nu + (1 - p) \theta,$$

so that the Lebesgue decomposition of $\mu$ w.r.t. $\nu$ is $(\phi, \psi) = (p \nu, (1 - p) \theta)$.
The table below show relative error between the approximate moments \( u = (u_\alpha) \) of degree 2 and 4, of the singular part \( \psi \) and those of \( p \theta \) computed with moments up to order \( 2d = 14 \).

<table>
<thead>
<tr>
<th>approx. moments</th>
<th>( x_1^2 )</th>
<th>( x_1^4 )</th>
<th>( x_1^2 x_2^2 )</th>
<th>( L_u^d((x_1^2 + x_2^2 - 1)^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p=0.1 )</td>
<td>0.19%</td>
<td>0.52%</td>
<td>0.53%</td>
<td>0.001</td>
</tr>
<tr>
<td>( p=0.2 )</td>
<td>3.7%</td>
<td>8.12%</td>
<td>12.14%</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Same thing but now with \( \nu \) being uniformly supported on the unit box \([-1, 1]^n\).

<table>
<thead>
<tr>
<th>approx. moments</th>
<th>( x_1^2 )</th>
<th>( x_1^4 )</th>
<th>( x_1^2 x_2^2 )</th>
<th>( L_u^d((x_1^2 + x_2^2 - 1)^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p=0.1 )</td>
<td>0.26%</td>
<td>0.93%</td>
<td>0.61%</td>
<td>0.0001</td>
</tr>
<tr>
<td>( p=0.2 )</td>
<td>8.8%</td>
<td>10.2%</td>
<td>7.5%</td>
<td>0.08</td>
</tr>
</tbody>
</table>
The table below show relative error between the approximate moments \( u = (u_\alpha) \) of degree 2 and 4, of the singular part \( \psi \) and those of \( p \theta \) computed with moments up to order \( 2d = 14 \).

<table>
<thead>
<tr>
<th>approx. moments</th>
<th>( x_1^2 )</th>
<th>( x_1^4 )</th>
<th>( x_1^2 x_2^2 )</th>
<th>( L_u^d((x_1^2 + x_2^2 - 1)^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>p=0.1</td>
<td>0.19%</td>
<td>0.52%</td>
<td>0.53%</td>
<td>0.001</td>
</tr>
<tr>
<td>p=0.2</td>
<td>3.7%</td>
<td>8.12%</td>
<td>12.14%</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Same thing but now with \( \nu \) being uniformly supported on the unit box \([-1, 1]^n\).

<table>
<thead>
<tr>
<th>approx. moments</th>
<th>( x_1^2 )</th>
<th>( x_1^4 )</th>
<th>( x_1^2 x_2^2 )</th>
<th>( L_u^d((x_1^2 + x_2^2 - 1)^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>p=0.1</td>
<td>0.26%</td>
<td>0.93%</td>
<td>0.61%</td>
<td>0.0001</td>
</tr>
<tr>
<td>p=0.2</td>
<td>8.8%</td>
<td>10.2%</td>
<td>7.5%</td>
<td>0.08</td>
</tr>
</tbody>
</table>