The moment-LP and moment-SOS approaches

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

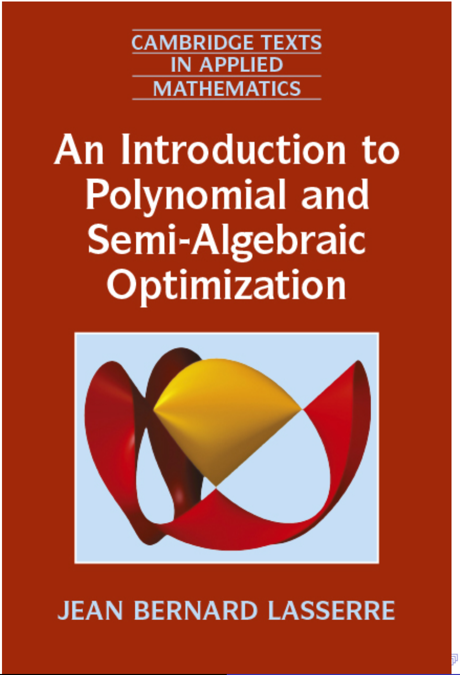
NIPS-2014, Optimization workshop, Montreal

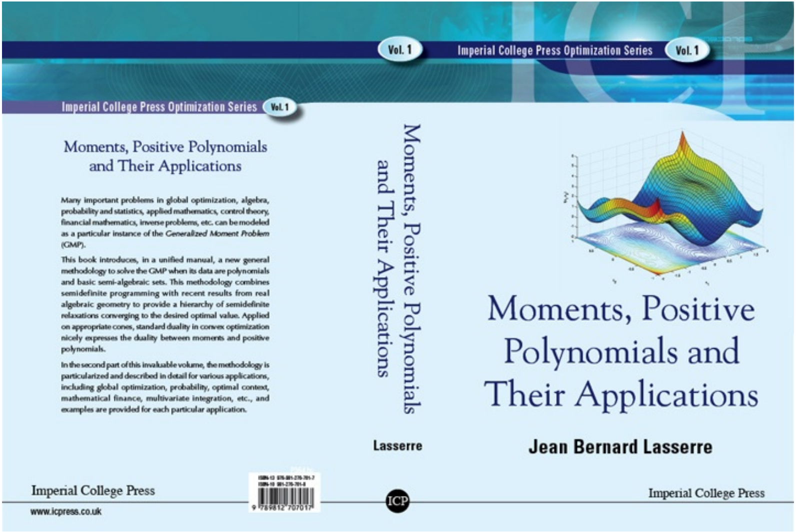
The moment-LP and moment-SOS approaches

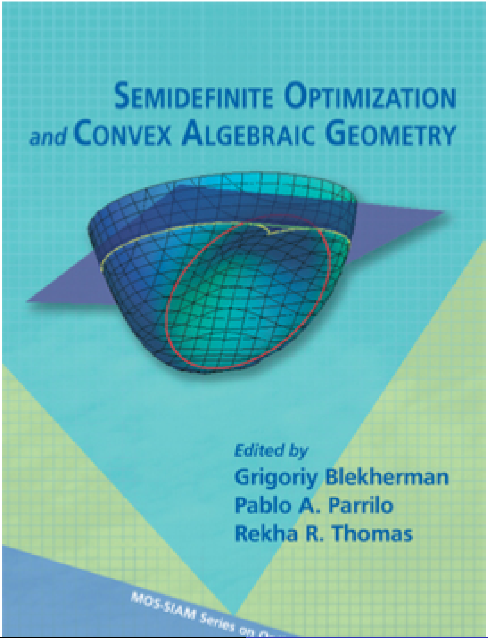
Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

NIPS-2014, Optimization workshop, Montreal


CAMBRIDGE TEXTS IN APPLIED MATHEMATICS
An Introduction to Polynomial and Semi-Algebraic Optimization
JEAN BERNARD LASSERRE


Vol. 1
Imperial College Press Optimization Series
Vol. 1
Imperial College Press Optimization Series
Vol. 1
Moments, Positive Polynomials and Their Applications
Many important problems in global optimization, algebra, probability and statistics, applied mathematics, control theory, financial mathematics, inverse problems, etc. can be modeled as a particular instance of the Generalized Moment Problem (GMP).
This book introduces, in a unified manual, a new general methodology to solve the GMP when its data are polynomials and basic semi-algebraic sets. This methodology combines semidefinite programming with recent results from real algebraic geometry to provide a hierarchy of semidefinite relaxations converging to the desired optimal value. Applied on appropriate cones, standard duality in convex optimization nicely expresses the duality between moments and positive polynomials.
In the second part of this invaluable volume, the methodology is particularized and described in detail for various applications, including global optimization, probability, optimal control, mathematical finance, multivariate integration, etc., and examples are provided for each particular application.
Imperial College Press
www.icpress.co.uk
Moments, Positive Polynomials and Their Applications
Lasserre
ICP
Moments, Positive Polynomials and Their Applications
Jean Bernard Lasserre
Imperial College Press


SEMIDEFINITE OPTIMIZATION
and CONVEX ALGEBRAIC GEOMETRY
Edited by
Grigoriy Blekherman
Pablo A. Parrilo
Rekha R. Thomas
MOS-SIAM Series on

- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- An alternative characterization of nonnegativity

- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- An alternative characterization of nonnegativity

- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- An alternative characterization of nonnegativity

- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- An alternative characterization of nonnegativity

Why Polynomial Optimization?

After all ...

the polynomial optimization problem:

$$f^* = \min\{f(\mathbf{x}) : \quad g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m\}$$

is just a particular case of Non Linear Programming (NLP)!

True!

... if one is interested with a LOCAL optimum only!!

Why Polynomial Optimization?

After all ...

the polynomial optimization problem:

$$f^* = \min\{f(\mathbf{x}) : \quad g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m\}$$

is just a particular case of Non Linear Programming (NLP)!

True!

... if one is interested with a LOCAL optimum only!!

When searching for a local minimum ...

Optimality conditions and descent algorithms use basic tools from REAL and CONVEX analysis and linear algebra

The focus is on how to improve f by looking at a NEIGHBORHOOD of a nominal point $\mathbf{x} \in \mathbf{K}$, i.e., LOCALLY AROUND $\mathbf{x} \in \mathbf{K}$, and in general, no GLOBAL property of $\mathbf{x} \in \mathbf{K}$ can be inferred.

The fact that f and g_j are POLYNOMIALS does not help much!

When searching for a local minimum ...

Optimality conditions and descent algorithms use basic tools from REAL and CONVEX analysis and linear algebra

The focus is on how to improve f by looking at a NEIGHBORHOOD of a nominal point $\mathbf{x} \in \mathbf{K}$, i.e., LOCALLY AROUND $\mathbf{x} \in \mathbf{K}$, and in general, no GLOBAL property of $\mathbf{x} \in \mathbf{K}$ can be inferred.

The fact that f and g_j are POLYNOMIALS does not help much!

When searching for a local minimum ...

Optimality conditions and descent algorithms use basic tools from REAL and CONVEX analysis and linear algebra

The focus is on how to improve f by looking at a NEIGHBORHOOD of a nominal point $\mathbf{x} \in \mathbf{K}$, i.e., LOCALLY AROUND $\mathbf{x} \in \mathbf{K}$, and in general, no GLOBAL property of $\mathbf{x} \in \mathbf{K}$ can be inferred.

The fact that f and g_j are POLYNOMIALS does not help much!

BUT for GLOBAL Optimization

... the picture is different!

Remember that for the GLOBAL minimum f^*:

$$f^* = \sup \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}.$$

... and so to compute f^* one needs

TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$!

BUT for GLOBAL Optimization

... the picture is different!

Remember that for the GLOBAL minimum f^*:

$$f^* = \sup \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}.$$

... and so to compute f^* one needs

TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$!

BUT for GLOBAL Optimization

... the picture is different!

Remember that for the GLOBAL minimum f^*:

$$f^* = \sup \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}.$$

... and so to compute f^* one needs

TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$!

REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

($\star$ Stronger Positivstellensatzë exist for analytic functions but are useless from a computational viewpoint.)

REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

($\star$ Stronger Positivstellensatzë exist for analytic functions but are useless from a computational viewpoint.)

REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

($\star$ Stronger Positivstellensatzë exist for analytic functions but are useless from a computational viewpoint.)

LP-based certificate

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

Theorem (Krivine-Vasilescu-Handelman's Positivstellensatz)

Let $\mathbf{K}$ *be compact and the family* $\{g_j, (1 - g_j)\}$ *generate* $\mathbb{R}[\mathbf{x}]$*. If* $f > 0$ *on* $\mathbf{K}$ *then:*

$$\star \quad f(\mathbf{x}) = \sum_{\alpha,\beta} c_{\alpha\beta} \prod_{j=1}^{m} g_j(\mathbf{x})^{\alpha_j} (1 - g_j(\mathbf{x}))^{\beta_j}, \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

for some NONNEGATIVE scalars $(c_{\alpha\beta})$*.*

Testing whether $\star$ holds for some

NONNEGATIVE $(c_{\alpha\beta})$ with $|\alpha + \beta| \leq M$, is SOLVING an LP!

LP-based certificate

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

Theorem (Krivine-Vasilescu-Handelman's Positivstellensatz)

Let $\mathbf{K}$ *be compact and the family* $\{g_j, (1 - g_j)\}$ *generate* $\mathbb{R}[\mathbf{x}]$. *If* $f > 0$ *on* $\mathbf{K}$ *then:*

$$\star \quad f(\mathbf{x}) = \sum_{\alpha,\beta} c_{\alpha\beta} \prod_{j=1}^{m} g_j(\mathbf{x})^{\alpha_j} \, (1 - g_j(\mathbf{x}))^{\beta_j}, \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

for some NONNEGATIVE scalars $(c_{\alpha\beta})$.

Testing whether $\star$ holds for some

NONNEGATIVE $(c_{\alpha\beta})$ with $|\alpha + \beta| \leq M$, is SOLVING an LP!

Indeed with $d = \max_j[\deg g_1, \ldots, \deg g_m]$ and $\sum_i \alpha_i + \beta_i \leq M$,

$$\sum_{\alpha,\beta} c_{\alpha\beta} \prod_{j=1}^{m} g_j(X)^{\alpha_j} (1 - g_j(X))^{\beta_j},$$

is a polynomial of degree at most $M + d$, which can be written

$$\sum_{\alpha,\beta} c_{\alpha\beta} \prod_{j=1}^{m} g_j(X)^{\alpha_j} (1 - g_j(X))^{\beta_j} = \sum_{\gamma \in \mathbb{N}^n_{M+d}} X^\gamma \quad \underbrace{\theta_\gamma(c)}_{\text{linear in } c}$$

And so the identity

$$f(\mathbf{x}) = \sum_\gamma f_\gamma X^\gamma = \sum_{\alpha,\beta} c_{\alpha\beta} \prod_{j=1}^{m} g_j(X)^{\alpha_j} (1 - g_j(X))^{\beta_j},$$

for all $X \in \mathbb{R}^n$, holds if and only if

$$\{ f_\gamma = \theta_\gamma(c), \quad \forall \gamma \in \mathbb{N}^n_{M+d};\ c \geq 0 \}. \quad \to c \in \text{polyhedron!}.$$

SOS-based certificate

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

Theorem (Putinar's Positivstellensatz)

If $\mathbf{K}$ *is compact (+ a technical Archimedean assumption) and* $f > 0$ *on* $\mathbf{K}$ *then:*

$$\dagger \quad f(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sum_{j=1}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

for some SOS polynomials $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.

Testing whether $\dagger$ holds for some

SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound, is SOLVING an SDP!

SOS-based certificate

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

Theorem (Putinar's Positivstellensatz)

If $\mathbf{K}$ *is compact (+ a technical Archimedean assumption) and* $f > 0$ *on* $\mathbf{K}$ *then:*

$$\dagger \quad f(\mathbf{x}) \;=\; \sigma_0(\mathbf{x}) + \sum_{j=1}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

for some SOS polynomials $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$*.*

Testing whether † holds for some

SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound, is SOLVING an SDP!

Checking whether a given polynomial is SOS

reduces to solving an SDP ... that one may solve efficiently to arbitrary precision, in time polynomial in the input size!

Indeed, let $v_d(X) = (X^\alpha\ (= X_1^{\alpha_1} \cdots X_n^{\alpha_n})),\ |\alpha| := \sum_i \alpha_i \leq d$, be a basis of $\mathbb{R}[X]_d$ (polynomials of degree at most d) Example

with $n = 2$ and $d = 3$,

$$v_3(X) = (1,\ X_1,\ X_2,\ X_1^2,\ X_1X_2,\ X_2^2,\ X_1^3,\ X_1^2X_2,\ X_1X_2^2,\ X_2^3\ )$$

Let $f \in \mathbb{R}[X]_{2d}$ be an SOS polynomial, that is,

$$f(X) = \sum_{k=1}^{s} q_k(X)^2,$$

for some polynomials $\{q_k\}_{k=1}^{s} \subset \mathbb{R}[X]_d$.

Checking whether a given polynomial is SOS

reduces to solving an SDP ... that one may solve efficiently to arbitrary precision, in time polynomial in the input size!

Indeed, let $v_d(X) = (X^\alpha \, (= X_1^{\alpha_1} \cdots X_n^{\alpha_n})), \, |\alpha| := \sum_i \alpha_i \leq d$, be a basis of $\mathbb{R}[X]_d$ (polynomials of degree at most d) Example

with $n = 2$ and $d = 3$,

$$v_3(X) \, = \, (1, \, X_1, \, X_2, \, X_1^2, \, X_1 X_2, \, X_2^2, \, X_1^3, \, X_1^2 X_2, \, X_1 X_2^2, \, X_2^3 \,)$$

Let $f \in \mathbb{R}[X]_{2d}$ be an SOS polynomial, that is,

$$f(X) \, = \, \sum_{k=1}^{s} q_k(X)^2,$$

for some polynomials $\{q_k\}_{k=1}^{s} \subset \mathbb{R}[X]_d$.

Denote also $q_k = \{q_{k\alpha}\}_{\alpha\in\mathbb{N}^n}$, the vector of coefficients of the polynomial q_k, in the basis $v_d(X)$, that is,

$$q_k(X) = \langle q_k, v_d(X)\rangle = \sum_{|\alpha|\leq r} q_{k\alpha} X^\alpha$$

and define the real symmetric matrix $Q := \sum_{k=1}^s q_k q_k^T \succeq 0$.

$$\langle v_d(X), Q\, v_d(X)\rangle = \sum_{k=1}^s \langle q_k, v_d(X)\rangle^2 = \sum_{k=1}^s q_k(X)^2 = f(X)$$

Conversely, let $Q \succeq 0$ be a real $s(d)\times s(d)$ positive semidefinite symmetric matrix ($s(d)$ is the dimension of the vector space $\mathbb{R}[X]_d$). As $Q \succeq 0$, write $Q = \sum_{k=1}^s q_k q_k^T$, so that

$$f(X) := \langle v_d(X), Q\, v_d(X)\rangle \left(= \sum_{k=1}^s \langle q_k, v_d(X)\rangle^2 = \sum_{k=1}^s q_k(X)^2 \right)$$

is SOS.

Denote also $q_k = \{q_{k\alpha}\}_{\alpha\in\mathbb{N}^n}$, the vector of coefficients of the polynomial q_k, in the basis $v_d(X)$, that is,

$$q_k(X) = \langle q_k, v_d(X)\rangle = \sum_{|\alpha|\leq r} q_{k\alpha}X^\alpha$$

and define the real symmetric matrix $Q := \sum_{k=1}^s q_k q_k^T \succeq 0$.

$$\langle v_d(X), Q\, v_d(X)\rangle = \sum_{k=1}^s \langle q_k, v_d(X)\rangle^2 = \sum_{k=1}^s q_k(X)^2 = f(X)$$

Conversely, let $Q \succeq 0$ be a real $s(d)\times s(d)$ positive semidefinite symmetric matrix ($s(d)$ is the dimension of the vector space $\mathbb{R}[X]_d$). As $Q \succeq 0$, write $Q = \sum_{k=1}^s q_k q_k^T$, so that

$$f(X) := \langle v_d(X), Q\, v_d(X)\rangle \left(= \sum_{k=1}^s \langle q_k, v_d(X)\rangle^2 = \sum_{k=1}^s q_k(X)^2\right)$$

is SOS.

Next, write the matrix $v_d(X)\, v_d(X)^T$ as:

$$v_d(X)\, v_d(X)^T \;=\; \sum_{\alpha\in\mathbb{N}^n_{2d}} B_\alpha\, x^\alpha,$$

for some real symmetric matrices (B_α). Checking whether

$$\begin{aligned}\underbrace{f(X)}_{\sum_\alpha f_\alpha X^\alpha} \;:=\langle v_d(X), Q\, v_d(X)\rangle \;&=\; \langle Q, v_d(X)\, v_d(X)^T\rangle\\ &=\; \sum_{\alpha\in\mathbb{N}^n_{2d}} \langle Q, B_\alpha\rangle\, X^\alpha\end{aligned}$$

for some $Q \succeq 0$ reduces to checking the LMI

$$\left\{\begin{array}{rcll}\langle B_\alpha, Q\rangle & = & f_\alpha, & \alpha\in\mathbb{N}^n,\ |\alpha|\leq 2d\\ Q & \succeq & 0 & \end{array}\right. .$$

has a solution!

Example

Let $t \mapsto f(t) = 6 + 4t + 9t^2 - 4t^3 + 6t^4$. Is f an SOS? Do we have

$$f(t) = \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}^T \underbrace{\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}}_{Q \succeq 0} \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix} ?$$

for some $Q \succeq 0$? We must have:

$$a = 6\,;\ 2\,b = 4;\ d + 2\,c = 9;\ 2\,e = -4;\ f = 6.$$

And so we must find a scalar c such that

$$Q = \begin{bmatrix} 6 & 2 & c \\ 2 & 9-2c & -2 \\ c & -2 & 6 \end{bmatrix} \succeq 0.$$

Example

Let $t \mapsto f(t) = 6 + 4t + 9t^2 - 4t^3 + 6t^4$. Is f an SOS? Do we have

$$f(t) = \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}^T \underbrace{\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}}_{Q \succeq 0} \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix} ?$$

for some $Q \succeq 0$? We must have:

$$a = 6\,;\, 2\,b = 4;\ d + 2\,c = 9;\ 2\,e = -4;\ f = 6.$$

And so we must find a scalar c such that

$$Q = \begin{bmatrix} 6 & 2 & c \\ 2 & 9 - 2c & -2 \\ c & -2 & 6 \end{bmatrix} \succeq 0.$$

With $c = -4$ we have

$$Q = \begin{bmatrix} 6 & 2 & -4 \\ 2 & 17 & -2 \\ -4 & -2 & 6 \end{bmatrix} \succeq 0.$$

et

$$Q = 2\begin{bmatrix} \sqrt(2/2) \\ 0 \\ \sqrt(2)/2 \end{bmatrix}\begin{bmatrix} \sqrt(2/2) \\ 0 \\ \sqrt(2)/2 \end{bmatrix}' + 9\begin{bmatrix} 2/3 \\ -1/3 \\ -2/3 \end{bmatrix}\begin{bmatrix} 2/3 \\ 1/3 \\ -2/3 \end{bmatrix}'$$

$$+18\begin{bmatrix} 1/\sqrt(18) \\ 4/\sqrt(18) \\ -1/\sqrt(18) \end{bmatrix}\begin{bmatrix} 1/\sqrt(18) \\ 4/\sqrt(18) \\ -1/\sqrt(18) \end{bmatrix}'$$

and so

$$f(t) = (1 + t^2)^2 + (2 - t - 2t^2)^2 + (1 + 4t - t^2)^2$$

which is an SOS polynomial.

SUCH POSITIVITY CERTIFICATES

allow to infer GLOBAL Properties of
FEASIBILITY and OPTIMALITY,

... the analogue of (well-known) previous ones

valid in the CONVEX CASE ONLY!

- Farkas Lemma $\rightarrow$ Krivine-Stengle
- KKT-Optimality conditions $\rightarrow$ Schmüdgen-Putinar

SUCH POSITIVITY CERTIFICATES

allow to infer GLOBAL Properties of
FEASIBILITY and OPTIMALITY,

... the analogue of (well-known) previous ones

valid in the CONVEX CASE ONLY!

- Farkas Lemma $\to$ Krivine-Stengle
- KKT-Optimality conditions $\to$ Schmüdgen-Putinar

SUCH POSITIVITY CERTIFICATES

allow to infer GLOBAL Properties of
FEASIBILITY and OPTIMALITY,

... the analogue of (well-known) previous ones

valid in the CONVEX CASE ONLY!

- Farkas Lemma $\rightarrow$ Krivine-Stengle
- KKT-Optimality conditions $\rightarrow$ Schmüdgen-Putinar

SUCH POSITIVITY CERTIFICATES

allow to infer GLOBAL Properties of
FEASIBILITY and OPTIMALITY,

... the analogue of (well-known) previous ones

valid in the CONVEX CASE ONLY!

- Farkas Lemma $\rightarrow$ Krivine-Stengle
- KKT-Optimality conditions $\rightarrow$ Schmüdgen-Putinar

In addition, polynomials NONNEGATIVE ON A SET $\mathbf{K} \subset \mathbb{R}^n$ are ubiquitous. They also appear in many important applications (outside optimization),

. . . modeled as

particular instances of the so called Generalized Moment Problem, among which: Probability, Optimal and Robust Control, Game theory, Signal processing, multivariate integration, etc.

The Generalized Moment Problem

$$(GMP): \quad \inf_{\mu_i \in M(\mathbf{K}_i)} \{ \sum_{i=1}^{s} \int_{\mathbf{K}_i} f_i \, d\mu_i : \sum_{i=1}^{s} \int_{\mathbf{K}_i} h_{ij} \, d\mu_i \overset{\leq}{=} b_j, \quad j \in J \}$$

with $M(\mathbf{K}_i)$ space of Borel measures on $\mathbf{K}_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, s$.

$$\text{Global OPTIM} \quad \rightarrow \inf_{\mu \in M(\mathbf{K})} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

is the simplest instance of the GMP!

The Generalized Moment Problem

$$(GMP): \quad \inf_{\mu_i \in M(\mathbf{K}_i)} \{ \sum_{i=1}^{s} \int_{\mathbf{K}_i} f_i \, d\mu_i : \sum_{i=1}^{s} \int_{\mathbf{K}_i} h_{ij} \, d\mu_i \overset{\leq}{\equiv} b_j, \quad j \in J \}$$

with $M(\mathbf{K}_i)$ space of Borel measures on $\mathbf{K}_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, s$.

$$\text{Global OPTIM} \quad \rightarrow \inf_{\mu \in M(\mathbf{K})} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

is the simplest instance of the GMP!

For instance, one may also want:

• To approximate sets defined with QUANTIFIERS, like .e.g.,

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

$$D_f := \{x \in \mathbf{B} \quad : \quad f(x,y) \leq 0 \text{ for some } y \text{ such that } (x,y) \in \mathbf{K}\}$$

where $f \in \mathbb{R}[x,y]$, $\mathbf{B}$ is a simple set (box, ellipsoid).

• To compute convex polynomial underestimators $p \leq f$ of a polynomial f on a box $\mathbf{B} \subset \mathbb{R}^n$. (Very useful in MINLP.)

For instance, one may also want:

• To approximate sets defined with QUANTIFIERS, like .e.g.,

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

$$D_f := \{x \in \mathbf{B} \quad : \quad f(x,y) \leq 0 \text{ for some } y \text{ such that } (x,y) \in \mathbf{K}\}$$

where $f \in \mathbb{R}[x,y]$, **B** is a simple set (box, ellipsoid).

• To compute convex polynomial underestimators $p \leq f$ of a polynomial f on a box $\mathbf{B} \subset \mathbb{R}^n$. (Very useful in MINLP.)

The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

LP- and SDP-hierarchies for optimization

Replace $f^* = \sup_{\lambda,\sigma_j} \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K}\}$ with:

The SDP-hierarchy indexed by $d \in \mathbb{N}$:

$$f_d^* = \sup \{ \lambda : f - \lambda = \underbrace{\sigma_0}_{SOS} + \sum_{j=1}^{m} \underbrace{\sigma_j}_{SOS} g_j; \quad \deg(\sigma_j\, g_j) \leq 2d \}$$

or, the LP-hierarchy indexed by $d \in \mathbb{N}$:

$$\theta_d = \sup \{ \lambda : f - \lambda = \sum_{\alpha,\beta} \underbrace{c_{\alpha\beta}}_{\geq 0} \prod_{j=1}^{m} g_j^{\alpha_j} (1 - g_j)^{\beta_j}; \quad |\alpha + \beta| \leq 2d \}$$

LP- and SDP-hierarchies for optimization

Replace $f^* = \sup_{\lambda,\sigma_j} \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}$ with:

The SDP-hierarchy indexed by $d \in \mathbb{N}$:

$$f_d^* = \sup \{ \lambda : f - \lambda = \underbrace{\sigma_0}_{SOS} + \sum_{j=1}^m \underbrace{\sigma_j}_{SOS} g_j; \quad \deg(\sigma_j g_j) \leq 2d \}$$

or, the LP-hierarchy indexed by $d \in \mathbb{N}$:

$$\theta_d = \sup \{ \lambda : f - \lambda = \sum_{\alpha,\beta} \underbrace{c_{\alpha\beta}}_{\geq 0} \prod_{j=1}^m g_j^{\alpha_j} (1 - g_j)^{\beta_j}; \quad |\alpha + \beta| \leq 2d \}$$

Theorem

Both sequence (f_d^*), and (θ_d), $d \in \mathbb{N}$, are MONOTONE NON DECREASING and when $\mathbf{K}$ is compact (and satisfies a technical Archimedean assumption) then:

$$f^* = \lim_{d\to\infty} f_d^* = \lim_{d\to\infty} \theta_d.$$

- What makes this approach exciting is that it is at the crossroads of several disciplines/applications:
  - Commutative, Non-commutative, and Non-linear ALGEBRA
  - Real algebraic geometry, and Functional Analysis
  - Optimization, Convex Analysis
  - Computational Complexity in Computer Science,

which BENEFIT from interactions!

- As mentioned ... potential applications are ENDLESS!

• What makes this approach exciting is that it is at the crossroads of several disciplines/applications:

- Commutative, Non-commutative, and Non-linear ALGEBRA
- Real algebraic geometry, and Functional Analysis
- Optimization, Convex Analysis
- Computational Complexity in Computer Science,

which BENEFIT from interactions!

• As mentioned ... potential applications are ENDLESS!

• Has already been proved useful and successful in applications with modest problem size, notably in optimization, control, robust control, optimal control, estimation, computer vision, etc. (If sparsity then problems of larger size can be addressed)

• HAS initiated and stimulated new research issues:

- in Convex Algebraic Geometry (e.g. semidefinite representation of convex sets, algebraic degree of semidefinite programming and polynomial optimization)
- in Computational algebra (e.g., for solving polynomial equations via SDP and Border bases)
- Computational Complexity where LP- and SDP-HIERARCHIES have become an important tool to analyze Hardness of Approximation for 0/1 combinatorial problems ($\rightarrow$ links with quantum computing)

• Has already been proved useful and successful in applications with modest problem size, notably in optimization, control, robust control, optimal control, estimation, computer vision, etc. (If sparsity then problems of larger size can be addressed)

• HAS initiated and stimulated new research issues:

- in Convex Algebraic Geometry (e.g. semidefinite representation of convex sets, algebraic degree of semidefinite programming and polynomial optimization)
- in Computational algebra (e.g., for solving polynomial equations via SDP and Border bases)
- Computational Complexity where LP- and SDP-HIERARCHIES have become an important tool to analyze Hardness of Approximation for 0/1 combinatorial problems ($\rightarrow$ links with quantum computing)

Recall that both LP- and SDP- hierarchies are

GENERAL PURPOSE METHODS ....

NOT TAILORED to solving specific hard problems!!

Recall that both LP- and SDP- hierarchies are

GENERAL PURPOSE METHODS ....

NOT TAILORED to solving specific hard problems!!

When solving the optimization problem

$$\mathbf{P}: \quad f^* = \min \{ f(\mathbf{x}) : \ g_j(\mathbf{x}) \geq 0, \ j = 1, \ldots, m \}$$

one does NOT distinguish between CONVEX, CONTINUOUS NON CONVEX, and 0/1 (and DISCRETE) problems! A boolean variable x_i is modelled via the equality constraint "$x_i^2 - x_i = 0$".

In Non Linear Programming (NLP),

modeling a 0/1 variable with the polynomial equality constraint "$x_i^2 - x_i = 0$" and applying a standard descent algorithm would be considered "stupid"!

Each class of problems has its own *ad hoc* tailored algorithms.

A remarkable property of the SOS hierarchy: I

When solving the optimization problem

$$\mathbf{P}: \quad f^* = \min \{ f(\mathbf{x}) : \; g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m \}$$

one does NOT distinguish between CONVEX, CONTINUOUS NON CONVEX, and 0/1 (and DISCRETE) problems! A boolean variable x_i is modelled via the equality constraint "$x_i^2 - x_i = 0$".

In Non Linear Programming (NLP),

modeling a 0/1 variable with the polynomial equality constraint
"$x_i^2 - x_i = 0$"
and applying a standard descent algorithm would be considered "stupid"!

Each class of problems has its own *ad hoc* tailored algorithms.

A remarkable property of the SOS hierarchy: I

When solving the optimization problem

$$\mathbf{P}: \quad f^* = \min \{ f(\mathbf{x}) : \; g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m \}$$

one does NOT distinguish between CONVEX, CONTINUOUS NON CONVEX, and 0/1 (and DISCRETE) problems! A boolean variable x_i is modelled via the equality constraint "$x_i^2 - x_i = 0$".

In Non Linear Programming (NLP),

modeling a 0/1 variable with the polynomial equality constraint "$x_i^2 - x_i = 0$" and applying a standard descent algorithm would be considered "stupid"!

Each class of problems has its own *ad hoc* tailored algorithms.

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems!

$\rightarrow$ The Theoretical Computer Science community speaks of a META-algorithm ...

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems!

$\rightarrow$ The Theoretical Computer Science community speaks of a META-algorithm ...

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems!

$\rightarrow$ The Theoretical Computer Science community speaks of a META-algorithm ...

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems!

$\rightarrow$ The Theoretical Computer Science community speaks of a META-algorithm ...

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems!

$\rightarrow$ The Theoretical Computer Science community speaks of a META-algorithm ...

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems!

$\rightarrow$ The Theoretical Computer Science community speaks of a META-algorithm ...

A remarkable property: II

FINITE CONVERGENCE of the SOS-hierarchy is GENERIC!

... and provides a GLOBAL OPTIMALITY CERTIFICATE,

the analogue for the NON CONVEX CASE of the KKT-OPTIMALITY conditions in the CONVEX CASE!

Theorem (Marshall, Nie)

Let $\mathbf{x}^* \in \mathbf{K}$ *be a global minimizer of*

$$\mathbf{P}: \quad f^* = \min\{f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, \, j = 1, \ldots, m\}.$$

and assume that:

(i) *The gradients* $\{\nabla g_j(\mathbf{x}^*)\}$ *are linearly independent,*

(ii) *Strict complementarity holds* ($\lambda_j^* \, g_j(\mathbf{x}^*) = 0$ *for all* j.)

(iii) *Second-order sufficiency conditions hold at* $(\mathbf{x}^*, \lambda^*) \in \mathbf{K} \times \mathbb{R}_+^m$.

Then $f(\mathbf{x}) - f^* = \sigma_0^*(\mathbf{x}) + \sum_{j=1}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$, *for some SOS polynomials* $\{\sigma_j^*\}$.

Moreover, the conditions (i)-(ii)-(iii) HOLD GENERICALLY!

Theorem (Marshall, Nie)

Let $\mathbf{x}^* \in \mathbf{K}$ *be a global minimizer of*

$$\mathbf{P}: \quad f^* = \min \{f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m\}.$$

and assume that:

(i) *The gradients* $\{\nabla g_j(\mathbf{x}^*)\}$ *are linearly independent,*

(ii) *Strict complementarity holds* ($\lambda_j^* g_j(\mathbf{x}^*) = 0$ *for all* j*.)*

(iii) *Second-order sufficiency conditions hold at* $(\mathbf{x}^*, \lambda^*) \in \mathbf{K} \times \mathbb{R}_+^m$.

Then $f(\mathbf{x}) - f^* = \sigma_0^*(\mathbf{x}) + \sum_{j=1}^{m} \sigma_j^*(\mathbf{x}) g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$, *for some SOS polynomials* $\{\sigma_j^*\}$.

Moreover, the conditions (i)-(ii)-(iii) HOLD GENERICALLY!

Certificates of positivity already exist in convex optimization

$$f^* = f(\mathbf{x}^*) = \min \{ f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \}$$

when f and $-g_j$ are CONVEX. Indeed if Slater's condition holds there exist nonnegative KKT-multipliers $\lambda_j^* \in \mathbb{R}_+^m$ such that:

$$\nabla f(\mathbf{x}^*) - \sum_{j=1}^{m} \lambda_j^* \, g_j(\mathbf{x}^*) = 0; \quad \lambda_j^* \, g_j(\mathbf{x}^*) = 0, \; j = 1, \ldots, m.$$

... and so ... the Lagrangian

$$L_{\lambda^*}(\mathbf{x}) := f(\mathbf{x}) - f^* - \sum_{j=1} \lambda_j^* \, g_j(\mathbf{x}),$$

satisfies

$L_{\lambda^*}(\mathbf{x}^*) = 0$ and $L_{\lambda^*}(\mathbf{x}) \geq 0$ for all $\mathbf{x}$. Therefore:

$$L_{\lambda^*}(\mathbf{x}) \geq 0 \Rightarrow f(\mathbf{x}) \geq f^* \quad \forall \mathbf{x} \in \mathbf{K}!$$

Certificates of positivity already exist in convex optimization

$$f^* = f(\mathbf{x}^*) = \min \{ f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \}$$

when f and $-g_j$ are CONVEX. Indeed if Slater's condition holds there exist nonnegative KKT-multipliers $\lambda_j^* \in \mathbb{R}_+^m$ such that:

$$\nabla f(\mathbf{x}^*) - \sum_{j=1}^{m} \lambda_j{}^* g_j(\mathbf{x}^*) = 0; \quad \lambda_j{}^* g_j(\mathbf{x}^*) = 0, \; j = 1, \ldots, m.$$

... and so ... the Lagrangian

$$L_{\lambda^*}(\mathbf{x}) := f(\mathbf{x}) - f^* - \sum_{j=1} \lambda_j{}^* g_j(\mathbf{x}),$$

satisfies

$L_{\lambda^*}(\mathbf{x}^*) = 0$ and $L_{\lambda^*}(\mathbf{x}) \geq 0$ for all $\mathbf{x}$. Therefore:

$$L_{\lambda^*}(\mathbf{x}) \geq 0 \Rightarrow f(\mathbf{x}) \geq f^* \quad \forall \mathbf{x} \in \mathbf{K}!$$

In summary:

KKT-OPTIMALITY when f and $-g_j$ are CONVEX	PUTINAR's CERTIFICATE in the non CONVEX CASE
$\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*) = 0$	$\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \sigma_j(\mathbf{x}^*) \nabla g_j(\mathbf{x}^*) = 0$
$f(\mathbf{x}) - f^* - \sum_{j=1}^m \lambda_j^* g_j(\mathbf{x})$	$f(\mathbf{x}) - f^* - \sum_{j=1}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x})$
≥ 0 for all $\mathbf{x} \in \mathbb{R}^n$	$(= \sigma_0^*(\mathbf{x})) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.
	for some SOS $\{\sigma_j^*\}$, and $\sigma_j^*(\mathbf{x}^*) = \lambda_j^*$.

In summary:

KKT-OPTIMALITY when f and $-g_j$ are CONVEX	PUTINAR's CERTIFICATE in the non CONVEX CASE
$\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*) = 0$	$\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \sigma_j(\mathbf{x}^*) \nabla g_j(\mathbf{x}^*) = 0$
$f(\mathbf{x}) - f^* - \sum_{j=1}^m \lambda_j^* g_j(\mathbf{x})$ ≥ 0 for all $\mathbf{x} \in \mathbb{R}^n$	$f(\mathbf{x}) - f^* - \sum_{j=1}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x})$ $(= \sigma_0^*(\mathbf{x})) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. for some SOS $\{\sigma_j^*\}$, and $\sigma_j^*(\mathbf{x}^*) = \lambda_j^*$.

So even though both LP- and SDP-relaxations were not designed for solving specific hard problems ...

The SDP-relaxations behave reasonably well ("efficiently"?) as they provide the BEST LOWER BOUNDS in very different contexts (in contrast to LP-relaxations).

$\rightarrow$ The Theoretical Computer Science (TCS) community even speaks of a META-ALGORITHM

$\rightarrow$ ... considered as the most promising tool to prove/disprove the Unique Games Conjecture (UGC)

So even though both LP- and SDP-relaxations were not designed for solving specific hard problems ...

The SDP-relaxations behave reasonably well ("efficiently"?) as they provide the BEST LOWER BOUNDS in very different contexts (in contrast to LP-relaxations).

$\rightarrow$ The Theoretical Computer Science (TCS) community even speaks of a META-ALGORITHM

$\rightarrow$ ... considered as the most promising tool to prove/disprove the Unique Games Conjecture (UGC)

So even though both LP- and SDP-relaxations were not designed for solving specific hard problems ...

The SDP-relaxations behave reasonably well ("efficiently"?) as they provide the BEST LOWER BOUNDS in very different contexts (in contrast to LP-relaxations).

$\rightarrow$ The Theoretical Computer Science (TCS) community even speaks of a META-ALGORITHM

$\rightarrow$ ... considered as the most promising tool to prove/disprove the Unique Games Conjecture (UGC)

A Lagrangian interpretation of LP-relaxations

Consider the optimization problem

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

where $\mathbf{K}$ is the compact basic semi-algebraic set:

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

Assume that:

• For every $j = 1, \ldots, m$ (and possibly after scaling), $g_j(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathbf{K}$.

• The family $\{g_j, 1 - g_j\}$ generate $\mathbb{R}[\mathbf{x}]$.

Lagrangian relaxation

The dual method of multipliers, or Lagrangian relaxation consists of solving: $\rho := \max_{\mathbf{u}}\{\, G(\mathbf{u}) : \mathbf{u} \geq 0 \,\}$,

$$\text{with} \quad \mathbf{u} \mapsto \quad G(\mathbf{u}) := \min_{\mathbf{x}} \left\{ f(x) - \sum_{j=1}^{m} u_j \, g_j(x) \right\}.$$

Equivalently:

$$\rho = \max_{\mathbf{u},\lambda} \{\lambda : \; f(x) - \sum_{j=1}^{m} u_j \, g_j(x) \geq \lambda, \quad \forall x.\}$$

In general, there is a DUALITY GAP, i.e., $\rho < f^*$,

except in the CONVEX case where f and $-g_j$ are all convex (and under some conditions).

With $d \in \mathbb{N}$ fixed, consider the new optimization problem $\mathbf{P}_d$:

$$f_d^* = \min_x \left\{ f(x) : \quad \begin{array}{l} \displaystyle\prod_{j=1}^{m} g_j(x)^{\alpha_j}(1 - g_j(x))^{\beta_j} \;\geq\; 0 \\ \forall\, \alpha, \beta : \; |\alpha + \beta| = \sum_j \alpha_j + \beta_j \leq 2d \end{array} \right\}$$

Of course

$\mathbf{P}$ and $\mathbf{P}_d$ are equivalent and so $f_d^* = f^*$.

... because $\mathbf{P}_d$ is just $\mathbf{P}$ with additional redundant constraints!

The Lagrangian relaxation of $\mathbf{P}_d$ consists of solving:

$$\rho_d = \max_{\mathbf{u}\geq 0,\lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1 - g_j(x))^{\beta_j} \geq \lambda, \quad \forall x.$$

$$|\alpha + \beta| \leq 2d\}$$

Theorem

$\rho_d \leq f^*$ for all $d \in \mathbb{N}$, and if $\mathbf{K}$ is compact and the family of polynomials $\{g_j, 1 - g_j\}$ generates $R[\mathbf{x}]$, then:

$$\lim_{d\to\infty} \rho_d = f^*.$$

The Lagrangian relaxation of $\mathbf{P}_d$ consists of solving:

$$\rho_d = \max_{\mathbf{u}\geq 0,\lambda} \left\{\lambda : \quad f(x) - \sum_{\substack{\alpha,\beta \\ |\alpha+\beta|\leq 2d}} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} \geq \lambda, \quad \forall x. \right\}$$

Theorem

$\rho_d \leq f^*$ for all $d \in \mathbb{N}$, and if $\mathbf{K}$ is compact and the family of polynomials $\{g_j, 1-g_j\}$ generates $R[\mathbf{x}]$, then:

$$\lim_{d\to\infty} \rho_d = f^*.$$

The previous theorem provides a rationale

for the well-known fact that :

adding redundant constraints to **P** helps when doing relaxations!

On the other hand ...

we don't know HOW TO COMPUTE ρ_d!

The previous theorem provides a rationale

for the well-known fact that :

adding redundant constraints to **P** helps when doing relaxations!

On the other hand ...

we don't know HOW TO COMPUTE ρ_d!

The LP-hierarchy may be viewed as

the BRUTE FORCE SIMPLIFICATION of

$$\rho_d = \max_{\mathbf{u}\geq 0,\lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} \geq \lambda, \quad \forall x.$$

$$|\alpha+\beta| \leq 2d\}$$

to ...

$$\theta_d = \max_{\mathbf{u}\geq 0,\lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} - \lambda = 0, \quad \forall x$$

$$|\alpha+\beta| \leq 2d\}$$

The LP-hierarchy may be viewed as

the BRUTE FORCE SIMPLIFICATION of

$$\rho_d = \max_{\mathbf{u}\geq 0,\lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} \geq \lambda, \quad \forall x.$$

$$|\alpha+\beta| \leq 2d\}$$

to ...

$$\theta_d = \max_{\mathbf{u}\geq 0,\lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} - \lambda = 0, \quad \forall x$$

$$|\alpha+\beta| \leq 2d\}$$

and indeed, ... with $|\alpha+\beta| \leq 2d$,

the set of $(\mathbf{u}, \lambda)$ such that $\mathbf{u} \geq 0$ and

$$f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} - \lambda = 0, \quad \forall x.$$

is a CONVEX POLYTOPE!

and so, computing θ_d is solving a Linear Program!

and one has $f^* \geq \rho_d \geq \theta_d$ for all d.

and indeed, ... with $|\alpha+\beta| \leq 2d$,

the set of $(\mathbf{u}, \lambda)$ such that $\mathbf{u} \geq 0$ and

$$f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} - \lambda = 0, \quad \forall x.$$

is a CONVEX POLYTOPE!

and so, computing θ_d is solving a Linear Program!

and one has $f^* \geq \rho_d \geq \theta_d$ for all d.

However as already mentioned

- For most easy convex problems (except LP) finite convergence is impossible!
- Other obstructions to exactness occur

Typically, if $\mathbf{K}$ is the polytope $\{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \ j = 1, \ldots, m\}$ and $f^* = f(\mathbf{x}^*)$ with $g_j(\mathbf{x})^* = 0, \ j \in J(\mathbf{x}^*)$, then finite convergence is impossible as soon as the exists $\mathbf{x} \neq \mathbf{x}^*$ with $J(\mathbf{x}) = J(\mathbf{x}^*)$ ($\mathbf{x}$ not necessarily in $\mathbf{K}$)

However as already mentioned

- For most easy convex problems (except LP) finite convergence is impossible!
- Other obstructions to exactness occur

Typically, if $\mathbf{K}$ is the polytope $\{\mathbf{x} : g_j(\mathbf{x}) \geq 0,\, j = 1, \ldots, m\}$ and $f^* = f(\mathbf{x}^*)$ with $g_j(\mathbf{x})^* = 0$, $j \in J(\mathbf{x}^*)$, then finite convergence is impossible as soon as the exists $\mathbf{x} \neq \mathbf{x}^*$ with $J(\mathbf{x}) = J(\mathbf{x}^*)$ ($\mathbf{x}$ not necessarily in $\mathbf{K}$)

A less brutal simplification

With $k \geq 1$ FIXED, consider the LESS BRUTAL SIMPLIFICATION of

$$\rho_d = \max_{\mathbf{u} \geq 0, \lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1 - g_j(x))^{\beta_j} \geq \lambda, \quad \forall x.$$

$$|\alpha + \beta| \leq 2d\}$$

to ...

$$\rho_d^k = \max_{\mathbf{u} \geq 0, \lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1 - g_j(x))^{\beta_j} - \lambda = \sigma(x),$$

$$|\alpha + \beta| \leq 2d; \quad \sigma \text{ SOS of degree at most } 2k\}$$

A less brutal simplification

With $k \geq 1$ FIXED, consider the LESS BRUTAL SIMPLIFICATION of

$$\rho_d = \max_{\mathbf{u}\geq 0,\lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} \geq \lambda, \quad \forall x.$$

$$|\alpha+\beta| \leq 2d\}$$

to ...

$$\rho_d^k = \max_{\mathbf{u}\geq 0,\lambda} \{\lambda : \quad f(x) - \sum_{\alpha,\beta} u_{\alpha\beta} \prod_{j=1}^{m} g_j(x)^{\alpha_j}(1-g_j(x))^{\beta_j} - \lambda = \sigma(x),$$

$$|\alpha+\beta| \leq 2d; \quad \sigma \text{ SOS of degree at most } 2k\}$$

Why such a simplification?

- With k fixed, $\rho_d^k = f^*$ as $d \to \infty$.
- Computing ρ_d^k is now solving an SDP (and not an LP any more!)
- However, the size of the LMI constraint of this SDP is $\binom{n+k}{n}$ (fixed) and does not depend on d!
- For convex problems where f and $-g_j$ are SOS-CONVEX polynomials, the first relaxation in the hierarchy is exact, that is, $\rho_1^k = f^*$ (never the case for the LP-hierarchy)

• A polynomial f is SOS-CONVEX if its Hessian $\nabla^2 f(x)$ factors as $L(x)\,L(x)^T$ for some polynomial matrix $L(x)$. For instance, separable polynomials $f(x) = \sum_{i=1}^n f_i(x_i)$, with convex f_i's are SOS-CONVEX.

Why such a simplification?

- With k fixed, $\rho_d^k = f^*$ as $d \to \infty$.
- Computing ρ_d^k is now solving an SDP (and not an LP any more!)
- However, the size of the LMI constraint of this SDP is $\binom{n+k}{n}$ (fixed) and does not depend on d!
- For convex problems where f and $-g_j$ are SOS-CONVEX polynomials, the first relaxation in the hierarchy is exact, that is, $\rho_1^k = f^*$ (never the case for the LP-hierarchy)

• A polynomial f is SOS-CONVEX if its Hessian $\nabla^2 f(x)$ factors as $L(x)\, L(x)^T$ for some polynomial matrix $L(x)$. For instance, separable polynomials $f(x) = \sum_{i=1}^n f_i(x_i)$, with convex f_i's are SOS-CONVEX.

Why such a simplification?

- With k fixed, $\rho_d^k = f^*$ as $d \to \infty$.
- Computing ρ_d^k is now solving an SDP (and not an LP any more!)
- However, the size of the LMI constraint of this SDP is $\binom{n+k}{n}$ (fixed) and does not depend on d!
- For convex problems where f and $-g_j$ are SOS-CONVEX polynomials, the first relaxation in the hierarchy is exact, that is, $\rho_1^k = f^*$ (never the case for the LP-hierarchy)

• A polynomial f is SOS-CONVEX if its Hessian $\nabla^2 f(x)$ factors as $L(x)\,L(x)^T$ for some polynomial matrix $L(x)$. For instance, separable polynomials $f(x) = \sum_{i=1}^n f_i(x_i)$, with convex f_i's are SOS-CONVEX.

Why such a simplification?

- With k fixed, $\rho_d^k = f^*$ as $d \to \infty$.
- Computing ρ_d^k is now solving an SDP (and not an LP any more!)
- However, the size of the LMI constraint of this SDP is $\binom{n+k}{n}$ (fixed) and does not depend on d!
- For convex problems where f and $-g_j$ are SOS-CONVEX polynomials, the first relaxation in the hierarchy is exact, that is, $\rho_1^k = f^*$ (never the case for the LP-hierarchy)

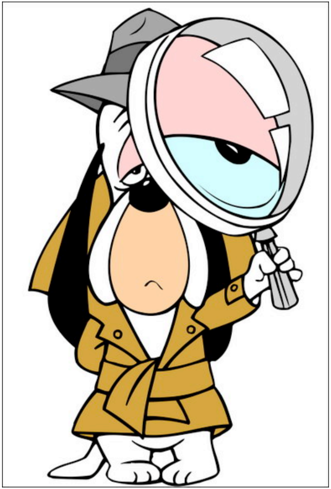
• A polynomial f is SOS-CONVEX if its Hessian $\nabla^2 f(x)$ factors as $L(x)\,L(x)^T$ for some polynomial matrix $L(x)$. For instance, separable polynomials $f(x) = \sum_{i=1}^n f_i(x_i)$, with convex f_i's are SOS-CONVEX.

Why such a simplification?

- With k fixed, $\rho_d^k = f^*$ as $d \to \infty$.
- Computing ρ_d^k is now solving an SDP (and not an LP any more!)
- However, the size of the LMI constraint of this SDP is $\binom{n+k}{n}$ (fixed) and does not depend on d!
- For convex problems where f and $-g_j$ are SOS-CONVEX polynomials, the first relaxation in the hierarchy is exact, that is, $\rho_1^k = f^*$ (never the case for the LP-hierarchy)

• A polynomial f is SOS-CONVEX if its Hessian $\nabla^2 f(x)$ factors as $L(x)\,L(x)^T$ for some polynomial matrix $L(x)$. For instance, separable polynomials $f(x) = \sum_{i=1}^n f_i(x_i)$, with convex f_i's are SOS-CONVEX.

An alternative moment-approach



So far we have considered

LP- and SDP-moment approaches based on
CERTIFICATES of POSITIVITY on **K**

That is:

One approximates FROM INSIDE the (convex cone) $C_d(\mathbf{K})$ of polynomials nonnegative on **K**: For instance if $\mathbf{K} = \{\mathbf{x} : \; g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m\}$, by the convex cones:

$$C_d^k(\mathbf{K}) \;=\; \{ \underbrace{\sigma_0}_{SOS} + \sum_{j=1}^{m} \underbrace{\sigma_j}_{SOS} \; g_j \; : \; \deg(\sigma_j g_j) \leq 2k \,\} \cap \mathbb{R}[\mathbf{x}]_d$$

$$\Gamma_d^k(\mathbf{K}) \;=\; \{ \sum_{(\alpha,\beta)\in\mathbb{N}^{2m}_{2k}} \underbrace{c_{\alpha\beta}}_{\geq 0} \prod_{j=1}^{m} g_j^{\alpha_j} (1-g_j)^{\beta_j} \,\} \cap \mathbb{R}[\mathbf{x}]_d$$

So far we have considered

LP- and SDP-moment approaches based on CERTIFICATES of POSITIVITY on $\mathbf{K}$

That is:

One approximates FROM INSIDE the (convex cone) $C_d(\mathbf{K})$ of polynomials nonnegative on $\mathbf{K}$: For instance if $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j = 1, \ldots, m\}$, by the convex cones:

$$C_d^k(\mathbf{K}) = \{ \underbrace{\sigma_0}_{SOS} + \sum_{j=1}^m \underbrace{\sigma_j}_{SOS} g_j \,:\, \deg(\sigma_j g_j) \leq 2k \} \cap \mathbb{R}[\mathbf{x}]_d$$

$$\Gamma_d^k(\mathbf{K}) = \{ \sum_{(\alpha,\beta)\in\mathbb{N}^{2m}_{2k}} \underbrace{c_{\alpha\beta}}_{\geq 0} \prod_{j=1}^m g_j^{\alpha_j}(1-g_j)^{\beta_j} \} \cap \mathbb{R}[\mathbf{x}]_d$$

An alternative is to try

to approximate $C_d(\mathbf{K})$ FROM OUTSIDE!

Given a sequence $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$:

• Let $L_y : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be the Riesz linear functional:

$$g \,(= \sum_\beta g_\beta \, \mathbf{x}^\beta) \;\mapsto\; L_y(g) \;:=\; \sum_\beta g_\beta \, y_\beta$$

• Define the localizing matrix $\mathbf{M}_k(g\,y)$ with respect to y and $g \in \mathbb{R}[\mathbf{x}]$ is the real symmetric matrix with rows and columns indexed by $\alpha \in \mathbb{N}^n$ and with entries

$$\mathbf{M}_k(g\,y)[\alpha,\beta] \;=\; L_y(\mathbf{x}^{\alpha+\beta}\, g_j), \qquad \alpha,\beta \in \mathbb{N}^n_k.$$

⋆ If y comes from a measure μ then

$$L_y(\mathbf{x}^{\alpha+\beta}\, g_j) \;=\; \int \mathbf{x}^{\alpha+\beta}\, g_j(x)\, d\mu.$$

An alternative is to try

to approximate $C_d(\mathbf{K})$ FROM OUTSIDE!

Given a sequence $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$:

• Let $L_y : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be the Riesz linear functional:

$$g \,(= \sum_\beta g_\beta \, \mathbf{x}^\beta) \;\mapsto\; L_y(g) \;:=\; \sum_\beta g_\beta \, y_\beta$$

• Define the localizing matrix $\mathbf{M}_k(g\,y)$ with respect to y and $g \in \mathbb{R}[\mathbf{x}]$ is the real symmetric matrix with rows and columns indexed by $\alpha \in \mathbb{N}^n$ and with entries

$$\mathbf{M}_k(g\,y)[\alpha, \beta] \;=\; L_y(\mathbf{x}^{\alpha+\beta}\, g_j), \qquad \alpha, \beta \in \mathbb{N}^n_k.$$

$\star$ If y comes from a measure μ then

$$L_y(\mathbf{x}^{\alpha+\beta}\, g_j) \;=\; \int \mathbf{x}^{\alpha+\beta}\, g_j(x)\, d\mu.$$

An alternative is to try
to approximate $C_d(\mathbf{K})$ FROM OUTSIDE!

Given a sequence $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$:

• Let $L_y : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be the Riesz linear functional:

$$g \; (= \sum_\beta g_\beta \, \mathbf{x}^\beta) \; \mapsto \; L_y(g) \; := \; \sum_\beta g_\beta \, y_\beta$$

• Define the localizing matrix $\mathbf{M}_k(g\,y)$ with respect to y and $g \in \mathbb{R}[\mathbf{x}]$ is the real symmetric matrix with rows and columns indexed by $\alpha \in \mathbb{N}^n$ and with entries

$$\mathbf{M}_k(g\,y)[\alpha, \beta] \; = \; L_y(\mathbf{x}^{\alpha+\beta} \, g_j), \qquad \alpha, \beta \in \mathbb{N}^n_k.$$

$\star$ If y comes from a measure μ then

$$L_y(\mathbf{x}^{\alpha+\beta} \, g_j) \; = \; \int \mathbf{x}^{\alpha+\beta} \, g_j(x) \, d\mu.$$

Theorem

Let $\mathbf{K} \subset \mathbb{R}^n$ *be compact and let* $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, *be the moments of a Borel measure whose support is* $\mathbf{K}$. *Then a polynomial* g_j *is nonnegative on* $\mathbf{K}$ *if and only if:*

$$\mathbf{M}_k(g_j\, y) \succeq 0, \quad k = 0, 1, \ldots$$

- So if y is known checking whether $\mathbf{M}_k(g_j\, y) \succeq 0$ is just computing the smallest eigenvalue of the matrix $\mathbf{M}_k(g_j\, y)$!
- The set $\Delta_k \subset \mathbb{R}[x]_d$ defined by:

$$\Delta_k := \{g \in \mathbb{R}[\mathbf{x}]_d : \mathbf{M}_k(g_j\, y) \succeq 0\}, \qquad k = 0, 1, \ldots$$

is a convex cone described by a LINEAR MATRIX INEQUALITY (LMI) on its coefficients (g_α), $\alpha \in \mathbb{N}^n_d$!

Theorem

Let $\mathbf{K} \subset \mathbb{R}^n$ *be compact and let* $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, *be the moments of a Borel measure whose support is* $\mathbf{K}$. *Then a polynomial* g_j *is nonnegative on* $\mathbf{K}$ *if and only if:*

$$\mathbf{M}_k(g_j\, y) \succeq 0, \quad k = 0, 1, \ldots$$

- So if y is known checking whether $\mathbf{M}_k(g_j\, y) \succeq 0$ is just computing the smallest eigenvalue of the matrix $\mathbf{M}_k(g_j\, y)$!
- The set $\Delta_k \subset \mathbb{R}[x]_d$ defined by:

$$\Delta_k := \{g \in \mathbb{R}[\mathbf{x}]_d : \mathbf{M}_k(g_j\, y) \succeq 0\}, \qquad k = 0, 1, \ldots$$

is a convex cone described by a LINEAR MATRIX INEQUALITY (LMI) on its coefficients (g_α), $\alpha \in \mathbb{N}^n_d$!

Theorem

Let $\mathbf{K} \subset \mathbb{R}^n$ *be compact and let* $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, *be the moments of a Borel measure whose support is* $\mathbf{K}$. *Then a polynomial* g_j *is nonnegative on* $\mathbf{K}$ *if and only if:*

$$\mathbf{M}_k(g_j\, y) \succeq 0, \quad k = 0, 1, \ldots$$

- So if y is known checking whether $\mathbf{M}_k(g_j\, y) \succeq 0$ is just computing the smallest eigenvalue of the matrix $\mathbf{M}_k(g_j\, y)$!
- The set $\Delta_k \subset \mathbb{R}[x]_d$ defined by:

$$\Delta_k := \{g \in \mathbb{R}[\mathbf{x}]_d : \mathbf{M}_k(g_j\, y) \succeq 0\}, \qquad k = 0, 1, \ldots$$

is a convex cone described by a LINEAR MATRIX INEQUALITY (LMI) on its coefficients (g_α), $\alpha \in \mathbb{N}^n_d$!

- Of course $C_d(\mathbf{K}) \subset \Delta_k \subset \Delta_{k-1}$, for all $k = 0, 1, \ldots$, and so
- $C_d(\mathbf{K}) = \cap_{k=0}^{\infty} \Delta_k$, i.e.,

The convex cones Δ_k form a nested sequence of INNER APPROXIMATIONS of $C_d(\mathbf{K})$.

Examples of sets for which the moments of a measure μ can be computed easily include:

- In the compact case: hyper-Rectangle $[a, b]^n$, Ellipsoid $\{\mathbf{x} : (\mathbf{x} - m)^T Q(\mathbf{x} - m) \leq 1\}$, simplex $\{\mathbf{x} \geq 0 : \sum_i a_i x_i \leq b\}$, hypercube $\{-1, 1\}^n$ with μ being uniformly distributed, and
- in the non-compact case: $\mathbb{R}^n$ with $d\mu = \exp(-\|\mathbf{x}\|^2)d\mathbf{x}$, and $\mathbb{R}^n_+$ with $d\mu = \exp(-\sum_i |x_i|)d\mathbf{x}$.

- Of course $C_d(\mathbf{K}) \subset \Delta_k \subset \Delta_{k-1}$, for all $k = 0, 1, \ldots$, and so
- $C_d(\mathbf{K}) = \cap_{k=0}^{\infty} \Delta_k$, i.e.,

The convex cones Δ_k form a nested sequence of INNER APPROXIMATIONS of $C_d(\mathbf{K})$.

Examples of sets for which the moments of a measure μ can be computed easily include:

- In the compact case: hyper-Rectangle $[a, b]^n$, Ellipsoid $\{\mathbf{x} : (\mathbf{x} - m)^T Q(\mathbf{x} - m) \leq 1\}$, simplex $\{\mathbf{x} \geq 0 : \sum_i a_i x_i \leq b\}$, hypercube $\{-1, 1\}^n$ with μ being uniformly distributed, and
- in the non-compact case: $\mathbb{R}^n$ with $d\mu = \exp(-\|\mathbf{x}\|^2)d\mathbf{x}$, and $\mathbb{R}^n_+$ with $d\mu = \exp(-\sum_i |x_i|)d\mathbf{x}$.

Application to optimization

Let $f^* = \min_{\mathbf{x}}\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$ and let $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, be the moments of a measure μ whose support is $\mathbf{K}$.

For each $d \in \mathbb{N}$ consider the optimization problem:

$$\rho_d = \max_{\lambda}\{\lambda : \mathbf{M}_d(f\,y) \succeq \lambda\,\mathbf{M}_d(y)\}.$$

with the single unknown λ.

- Computing ρ_d is solving a generalized eigenvalue problem associated with $\mathbf{M}_d(f\,y)$ and $\mathbf{M}_d(y)$.
- $\rho_d \geq f^*$ for all d and $\rho_d \to f^*$ as $d \to \infty$

Application to optimization

Let $f^* = \min_{\mathbf{x}}\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$ and let $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, be the moments of a measure μ whose support is $\mathbf{K}$.

For each $d \in \mathbb{N}$ consider the optimization problem:

$$\rho_d = \max_{\lambda}\{\lambda : \mathbf{M}_d(f\,y) \succeq \lambda\,\mathbf{M}_d(y)\}.$$

with the single unknown λ.

- Computing ρ_d is solving a generalized eigenvalue problem associated with $\mathbf{M}_d(f\,y)$ and $\mathbf{M}_d(y)$.
- $\rho_d \geq f^*$ for all d and $\rho_d \to f^*$ as $d \to \infty$

In other words: the sequence (ρ_d), $d \in \mathbb{N}$, provides a converging sequence of upper bounds on f^*!

Example: MAX-CUT problem: $f(x) = \mathbf{x}^T Q\, \mathbf{x}$ and $\mathbf{K} = \{-1, 1\}^n$. Take for μ the measure uniformly distributed on $\mathbf{K}$ with weights $1/2$, and so with moments:

$$y_\alpha = \int_{\{-1,1\}^n} \mathbf{x}^\alpha \, d\mathbf{x} = \begin{cases} 0 \text{ if } \alpha_i \text{ is odd for some } i \\ 1 \text{ otherwise} \end{cases}$$

Then build up the localizing matrix $\mathbf{M}_d(f\, y)$ and solve

$$\rho_d = \max_\lambda \{\lambda : \mathbf{M}_d(f\, y) \succeq \lambda\, \mathbf{M}_d(y)\}.$$

In fact, same as computing

the smallest eigenvalue of $\widehat{\mathbf{M}}_d(f\, y)$ (keeping only the rows and columns of $\mathbf{M}_d(f\, y)$ indexed by square-free monomials $(\mathbf{x}^\alpha)$.

In other words: the sequence (ρ_d), $d \in \mathbb{N}$, provides a converging sequence of upper bounds on f^*!

Example: MAX-CUT problem: $f(x) = \mathbf{x}^T Q\,\mathbf{x}$ and $\mathbf{K} = \{-1,1\}^n$. Take for μ the measure uniformly distributed on $\mathbf{K}$ with weights $1/2$, and so with moments:

$$y_\alpha = \int_{\{-1,1\}^n} \mathbf{x}^\alpha \, d\mathbf{x} = \left\{ \begin{array}{l} 0 \text{ if } \alpha_i \text{ is odd for some } i \\ 1 \text{ otherwise} \end{array} \right.$$

Then build up the localizing matrix $\mathbf{M}_d(f\,y)$ and solve

$$\rho_d = \max_\lambda \{\lambda : \mathbf{M}_d(f\,y) \succeq \lambda\,\mathbf{M}_d(y)\}.$$

In fact, same as computing

the smallest eigenvalue of $\widehat{\mathbf{M}}_d(f\,y)$ (keeping only the rows and columns of $\mathbf{M}_d(f\,y)$ indexed by square-free monomials $(\mathbf{x}^\alpha)$.

In other words: the sequence (ρ_d), $d \in \mathbb{N}$, provides a converging sequence of upper bounds on f^*!

Example: MAX-CUT problem: $f(x) = \mathbf{x}^T Q\,\mathbf{x}$ and $\mathbf{K} = \{-1, 1\}^n$. Take for μ the measure uniformly distributed on $\mathbf{K}$ with weights $1/2$, and so with moments:

$$y_\alpha = \int_{\{-1,1\}^n} \mathbf{x}^\alpha \, d\mathbf{x} = \begin{cases} 0 \text{ if } \alpha_i \text{ is odd for some } i \\ 1 \text{ otherwise} \end{cases}$$

Then build up the localizing matrix $\mathbf{M}_d(f\,y)$ and solve

$$\rho_d = \max_\lambda \{\lambda : \mathbf{M}_d(f\,y) \succeq \lambda\,\mathbf{M}_d(y)\}.$$

In fact, same as computing

the smallest eigenvalue of $\widehat{\mathbf{M}}_d(f\,y)$ (keeping only the rows and columns of $\mathbf{M}_d(f\,y)$ indexed by square-free monomials ($\mathbf{x}^\alpha$).

THANK YOU!!