A "joint+marginal" approach to parametric optimlization

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

RTRA, Toulouse, February 2010

This work is based on:

The "joint+marginal" approach to parametric polynomial optimization

To appear in **SIAM J. Optimization**

- Parametric Optimization
- The "joint+marginal" approach
- A hierarchy of semidefinite relaxations
- -> in optimization yields a "joint+marginal" algorithm

This work is based on:

The "joint+marginal" approach to parametric polynomial optimization

To appear in **SIAM J. Optimization**

- Parametric Optimization
- The "joint+marginal" approach
- A hierarchy of semidefinite relaxations
- -> in optimization yields a "joint+marginal" algorithm

This work is based on:

The "joint+marginal" approach to parametric polynomial optimization

To appear in **SIAM J. Optimization**

- Parametric Optimization
- The "joint+marginal" approach
- A hierarchy of semidefinite relaxations
- -> in optimization yields a "joint+marginal" algorithm

This work is based on:

The "joint+marginal" approach to parametric polynomial optimization

To appear in **SIAM J. Optimization**

- Parametric Optimization
- The "joint+marginal" approach
- A hierarchy of semidefinite relaxations
- -> in optimization yields a "joint+marginal" algorithm

This work is based on:

The "joint+marginal" approach to parametric polynomial optimization

To appear in **SIAM J. Optimization**

- Parametric Optimization
- The "joint+marginal" approach
- A hierarchy of semidefinite relaxations
- -> in optimization yields a "joint+marginal" algorithm

Parametric Optimization

Let $\mathbf{Y} \subset \mathbb{R}^p$ be a compact set, called the parameter set.

Let $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$ be the set:

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

for some continuous functions $h_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

Consider the following optimization problem:

$$J(\mathbf{y}) := \inf_{\mathbf{x}} \{ f(\mathbf{x}, \mathbf{y}) \,:\, x \in \mathbf{K}_{\mathbf{y}} \},$$

where for each $\mathbf{y} \in \mathbf{Y}$, the $\mathbf{K}_{\mathbf{y}} \subset \mathbb{R}^n$ is defined by:

$$\mathbf{K}_{\mathbf{y}} := \{ \mathbf{x} \in \mathbb{R}^n \,:\, (\mathbf{x}, \mathbf{y}) \in \mathbf{K} \}$$

Parametric optimization is concerned with:

- the global optimal value function $\mathbf{y} \mapsto J(\mathbf{y})$, and
- the global minimizer set function $\mathbf{y} \mapsto \mathbf{x}^*_i(\mathbf{y})$
- the optimal dual multiplier set function $\mathbf{y} \mapsto \lambda^*_j(\mathbf{y})$ associated with the constraint $h_j(\mathbf{x}, \mathbf{y}) \geq 0$.

In general, getting full information is impossible, and one is satisfied with local information (e.g. sensitivity analysis) around some (even local) minimizer $\mathbf{x}^*(\mathbf{y}) \in \mathbf{K}_{\mathbf{y}}$, $\mathbf{y} \in \mathbf{Y}$. (See e.g. the book by Bonnans and Shapiro.)

This talk

For polynomial optimization much more is possible!

Parametric optimization is concerned with:

- the global optimal value function $\mathbf{y} \mapsto J(\mathbf{y})$, and
- the global minimizer set function $\mathbf{y} \mapsto \mathbf{x}_i^*(\mathbf{y})$
- the optimal dual multiplier set function $\mathbf{y} \mapsto \lambda_j^*(\mathbf{y})$ associated with the constraint $h_j(\mathbf{x}, \mathbf{y}) \geq 0$.

In general, getting full information is impossible, and one is satisfied with local information (e.g. sensitivity analysis) around some (even local) minimizer $\mathbf{x}^*(\mathbf{y}) \in \mathbf{K}_{\mathbf{y}}$, $\mathbf{y} \in \mathbf{Y}$. (See e.g. the book by Bonnans and Shapiro.)

This talk

For polynomial optimization much more is possible!

Parametric optimization is concerned with:

- the global optimal value function $\mathbf{y} \mapsto J(\mathbf{y})$, and
- the global minimizer set function $\mathbf{y} \mapsto \mathbf{x}^*_i(\mathbf{y})$
- the optimal dual multiplier set function $\mathbf{y} \mapsto \lambda^*_j(\mathbf{y})$ associated with the constraint $h_j(\mathbf{x}, \mathbf{y}) \geq 0$.

In general, getting full information is impossible, and one is satisfied with local information (e.g. sensitivity analysis) around some (even local) minimizer $\mathbf{x}^*(\mathbf{y}) \in \mathbf{K}_{\mathbf{y}}$, $\mathbf{y} \in \mathbf{Y}$. (See e.g. the book by Bonnans and Shapiro.)

This talk

For polynomial optimization much more is possible!

The "joint+marginal" approach

Let φ be a Borel probability measure on $\mathbf{Y}$, with a positive density with respect to the Lebesgue measure on the smallest affine variety that contains $\mathbf{Y}$. For instance,

$$\varphi(B) := \left(\int_{\mathbf{Y}} d\mathbf{y}\right)^{-1} \int_B d\mathbf{y}, \qquad \forall B \in \mathcal{B}(\mathbf{Y}),$$

is uniformly distributed on $\mathbf{Y}$.

For a discrete set of parameters $\mathbf{Y}$ (finite or countable) take for φ a discrete probability measure on $\mathbf{Y}$ with strictly positive weight at each point of the support.

Sometimes, e.g. in the context of optimization with data uncertainty, φ is already specified.

A related infinite-dimensional linear program:

Consider the infinite-dimensional LP:

$$\mathbf{P}: \quad \rho := \inf_{\mu \in \mathbf{M}(\mathbf{K})} \left\{ \int_{\mathbf{K}} f \, d\mu \; : \; \pi\mu = \varphi \right\}$$

where: $\mathbf{M}(\mathbf{K})$ is the of Borel probability measures on $\mathbf{K}$, and

$\pi : \mathbf{M}(\mathbf{K}) \to \mathbf{M}(\mathbf{Y})$ is the projection (or, marginal) on $\mathbf{Y}$.

Whence the name "joint+marginal"-approach since:

- μ is a joint distribution on the variables $\mathbf{x}$ AND the parameters $\mathbf{y}$.
- φ is the marginal of μ on $\mathbf{Y}$ (fixed, as a constraint on μ).

A related infinite-dimensional linear program:

Consider the infinite-dimensional LP:

$$\mathbf{P}: \quad \rho := \inf_{\mu \in \mathbf{M}(\mathbf{K})} \left\{ \int_{\mathbf{K}} f \, d\mu \; : \; \pi\mu = \varphi \right\}$$

where: $\mathbf{M}(\mathbf{K})$ is the of Borel probability measures on $\mathbf{K}$, and

$\pi : \mathbf{M}(\mathbf{K}) \rightarrow \mathbf{M}(\mathbf{Y})$ is the projection (or, marginal) on $\mathbf{Y}$.

Whence the name "joint+marginal"-approach since:

- μ is a joint distribution on the variables $\mathbf{x}$ AND the parameters $\mathbf{y}$.
- φ is the marginal of μ on $\mathbf{Y}$ (fixed, as a constraint on μ).

The dual $\mathbf{P}^*$ is the infinite-dimensional LP:

$$\mathbf{P}^*: \quad \rho^* := \sup_{g \in C(\mathbf{Y})} \int_{\mathbf{Y}} g(\mathbf{y})\, d\varphi(\mathbf{y})$$
$$f(\mathbf{x},\mathbf{y}) - g(\mathbf{y}) \geq 0 \quad \forall (\mathbf{x},\mathbf{y}) \in \mathbf{K}.$$

where $C(\mathbf{Y})$ is the set of continuous functions on $\mathbf{Y}$.

In other words, among the continuous functions g on $\mathbf{Y}$ such that:

$$f(\mathbf{x},\mathbf{y}) \geq g(\mathbf{y}) \qquad \forall \mathbf{x} \in \mathbf{K}_{\mathbf{y}},$$

one searches for the one that maximizes $\int_{\mathbf{Y}} g d\varphi$.

The dual $\mathbf{P}^*$ is the infinite-dimensional LP:

$$\mathbf{P}^*: \quad \rho^* := \sup_{g \in C(\mathbf{Y})} \int_{\mathbf{Y}} g(\mathbf{y})\, d\varphi(\mathbf{y})$$
$$f(\mathbf{x},\mathbf{y}) - g(\mathbf{y}) \geq 0 \quad \forall (\mathbf{x},\mathbf{y}) \in \mathbf{K}.$$

where $C(\mathbf{Y})$ is the set of continuous functions on $\mathbf{Y}$.

In other words, among the continuous functions g on $\mathbf{Y}$ such that:

$$f(\mathbf{x},\mathbf{y}) \geq g(\mathbf{y}) \qquad \forall \mathbf{x} \in \mathbf{K}_{\mathbf{y}},$$

one searches for the one that maximizes $\int_{\mathbf{Y}} g d\varphi$.

Why those LPs?

We assume that $\mathbf{K}$ is compact.

As we shall see ....

Any optimal solution μ^* of the primal $\mathbf{P}$ encodes *all* information on the optimal solutions $\mathbf{x}^*(\mathbf{y})$ of $\mathbf{P}_\mathbf{y}$.

Similarly ....

Ther is no duality gap $\rho = \rho^*$ and so, in particular, the optimal value function $\mathbf{y} \mapsto J(\mathbf{y})$ of $\mathbf{P}_\mathbf{y}$ can be nicely approximated by polynomials.

Why those LPs?

We assume that $\mathbf{K}$ is compact.

As we shall see ....

Any optimal solution μ^* of the primal **P** encodes *all* information on the optimal solutions $\mathbf{x}^*(\mathbf{y})$ of $\mathbf{P}_\mathbf{y}$.

Similarly ....

Ther is no duality gap $\rho = \rho^*$ and so, in particular, the optimal value function $\mathbf{y} \mapsto J(\mathbf{y})$ of $\mathbf{P}_\mathbf{y}$ can be nicely approximated by polynomials.

Theorem (The primal side ...)

Assume that $\mathbf{K}$ *is compact and* $\mathbf{K}_{\mathbf{y}} \neq \emptyset$ *for every* $\mathbf{y} \in \mathbf{Y}$. *Let*

$\mathbf{X}^*_{\mathbf{y}} := \{\mathbf{x} \in \mathbb{R}^n \,:\, f(\mathbf{x},\mathbf{y}) = J(\mathbf{y})\}$, $\mathbf{y} \in \mathbf{Y}$. *Then:*

(a) $\rho \,=\, \int_{\mathbf{Y}} J(\mathbf{y})\, d\varphi(\mathbf{y})$ *and* $\mathbf{P}$ *has an optimal solution.*

(b) *For every optimal solution* μ^* *of* $\mathbf{P}$, *and for* φ*-almost all* $\mathbf{y} \in \mathbf{Y}$, *there is a probability measure* $\psi^*(d\mathbf{x} \,|\, \mathbf{y})$ *on* $\mathbb{R}^n$, *concentrated on* $\mathbf{X}^*_{\mathbf{y}}$, *such that:*

$$\mu^*(C \times B) \,=\, \int_B \psi^*(C \,|\, \mathbf{y})\, d\varphi(\mathbf{y}), \qquad \forall B \in \mathcal{B}(\mathbf{Y}),\; C \in \mathcal{B}(\mathbb{R}^n).$$

continued ...

(c) Assume that for φ-almost all $\mathbf{y} \in \mathbf{Y}$, the set of minimizers $\mathbf{X}^*_{\mathbf{y}}$ is the singleton $\{\mathbf{x}^*(\mathbf{y})\}$ for some $\mathbf{x}^*(\mathbf{y}) \in \mathbf{K}_{\mathbf{y}}$. Then there is a measurable mapping $g : \mathbf{Y} \to \mathbf{K}_{\mathbf{y}}$ such that

$$g(\mathbf{y}) = \mathbf{x}^*(\mathbf{y}) \text{ for every } \mathbf{y} \in \mathbf{Y}; \quad \rho = \int_{\mathbf{Y}} f(g(\mathbf{y}), \mathbf{y})\, d\varphi(\mathbf{y}),$$

and for every $\alpha \in \mathbb{N}^n$, and $\beta \in \mathbb{N}^p$:

$$\int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{y}^\beta \, d\mu^*(\mathbf{x}, \mathbf{y}) = \int_{\mathbf{Y}} \mathbf{y}^\beta \, g(\mathbf{y})^\alpha \, d\varphi(\mathbf{y}).$$

Theorem (The dual side ...)

(a) There is no duality gap, i.e.,

$$\rho \,=\, \rho^* \,=\, \int_{\mathbf{Y}} J(\mathbf{y})\, d\varphi(\mathbf{y}),$$

(b) One may use polynomials of $\mathbb{R}[\mathbf{y}]$ to approximate ρ^*.

(c) Let $(p_i) \subset \mathbb{R}[\mathbf{y}]$ be any maximizing sequence. Then:

L_1-norm convergence:

$$\text{as } i \to \infty, \qquad \int_{\mathbf{Y}} |J(\mathbf{y}) - p_i(\mathbf{y})|\, d\varphi(\mathbf{y}) \,\to\, 0$$

φ-almost sure convergence: *Let $\tilde{p}_i := \max_{k=0,..,i} p_i$. Then*

$$\text{as } i \to \infty, \qquad \tilde{p}_i \,\to\, J \qquad \varphi\text{-almost surely in } \mathbf{Y}$$

Polynomial Parametric Optimization

In general, P and P* are intractable!

However .... when:

- $\mathbf{Y}$ and $\mathbf{K}$, are basic semi-algebraic sets, and:
- either one already knows the moments of φ, or $\mathbf{Y}$ is simple enough (e.g. a box, a simplex, a hyper-sphere) so that they can be computed.

.... then one can approximate the optimal value ρ of P, and:

- The optimal value mapping $\mathbf{y} \mapsto J(\mathbf{y})$
- The global minimizer mapping $\mathbf{y} \mapsto \mathbf{x}^*(\mathbf{y})$,

... via the hierarchy of semidefinite relaxations

of the moment-s.o.s. approach in polynomial optimization.

Polynomial Parametric Optimization

In general, P and P* are intractable!

However .... when:

- **Y** and **K**, are basic semi-algebraic sets, and:
- either one already knows the moments of φ, or **Y** is simple enough (e.g. a box, a simplex, a hyper-sphere) so that they can be computed.

.... then one can approximate the optimal value ρ of **P**, and:

- The optimal value mapping $\mathbf{y} \mapsto J(\mathbf{y})$
- The global minimizer mapping $\mathbf{y} \mapsto \mathbf{x}^*(\mathbf{y})$,

... via the hierarchy of semidefinite relaxations

of the moment-s.o.s. approach in polynomial optimization.

In general, P and P^* are intractable!

However .... when:

- **Y** and **K**, are basic semi-algebraic sets, and:
- either one already knows the moments of φ, or **Y** is simple enough (e.g. a box, a simplex, a hyper-sphere) so that they can be computed.

.... then one can approximate the optimal value ρ of **P**, and:

- The optimal value mapping $\mathbf{y} \mapsto J(\mathbf{y})$
- The global minimizer mapping $\mathbf{y} \mapsto \mathbf{x}^*(\mathbf{y})$,

... via the hierarchy of semidefinite relaxations

of the moment-s.o.s. approach in polynomial optimization.

The moment-s.o.s. approach

Let $\mathbb{N}^n_i := \{\alpha \in \mathbb{N}^n : \sum_j \alpha_j| \leq i\}$.

With a sequence $\mathbf{z} = (z_{\alpha\beta})$, indexed in the canonical basis $(\mathbf{x}^\alpha\, \mathbf{y}^\beta)$ of $\mathbb{R}[\mathbf{x}, \mathbf{y}]$, let $L_\mathbf{z} : \mathbb{R}[\mathbf{x}, \mathbf{y}] \to \mathbb{R}$ be the linear mapping:

$$f\,(= \sum_{\alpha\beta} f_{\alpha\beta}\mathbf{x}^\alpha \mathbf{y}^\beta) \mapsto L_\mathbf{z}(f) := \sum_{\alpha\beta} f_{\alpha\beta}\, z_{\alpha\beta}, \qquad f \in \mathbb{R}[\mathbf{x}, \mathbf{y}].$$

The moment matrix $\mathbf{M}_i(\mathbf{z})$

associated with a sequence $\mathbf{z} = (z_{\alpha\beta})$, has its rows and columns indexed in the canonical basis $(\mathbf{x}^\alpha\, \mathbf{y}^\beta)$, and with entries.

$$\mathbf{M}_i(\mathbf{z})(\alpha, \beta), (\delta, \gamma)) = L_\mathbf{z}(\mathbf{x}^\alpha \mathbf{y}^\beta\, \mathbf{x}^\delta \mathbf{y}^\gamma) = z_{(\alpha+\delta)(\beta+\gamma)},$$

for every $\alpha, \delta \in \mathbb{N}^n_i$ and every $\beta, \gamma \in \mathbb{N}^p_i$

.

Let q be the polynomial $(\mathbf{x},\mathbf{y}) \mapsto q(\mathbf{x},\mathbf{y}) := \sum_{u,v} q_{uv}\mathbf{x}^u\mathbf{y}^v$.

The localizing matrix $\mathbf{M}_i(q,\mathbf{z})$

associated with $q \in \mathbb{R}[\mathbf{x},\mathbf{y}]$ and a sequence $\mathbf{z} = (z_{\alpha\beta})$, has its rows and columns indexed in the canonical basis $(\mathbf{x}^\alpha\,\mathbf{y}^\beta)$, and with entries.

$$\begin{aligned}\mathbf{M}_i(q\,\mathbf{z})(\alpha,\beta),(\delta,\gamma)) &= L_{\mathbf{z}}(q(\mathbf{x},\mathbf{y})\mathbf{x}^\alpha\mathbf{y}^\beta\,\mathbf{x}^\delta\mathbf{y}^\gamma) \\ &= \sum_{u\in\mathbb{N}^n,v\in\mathbb{N}^p} q_{uv}z_{(\alpha+\delta+u)(\beta+\gamma+v)},\end{aligned}$$

for every $\alpha,\delta \in \mathbb{N}^n_i$ and every $\beta,\gamma \in \mathbb{N}^p_i$.

Let $h_j \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$ for every $j = 1, \ldots, m$, and recall that

$$\mathbf{K} := \{ (\mathbf{x}, \mathbf{y}) \; : \; \mathbf{y} \in \mathbf{Y} \, ; \quad h_j(\mathbf{x}, \mathbf{y}) \geq 0, \quad j = 1, \ldots, m \}.$$

The parameter set $\mathbf{Y}$ is the basic semi-algebraic set

$$\mathbf{Y} := \{ \mathbf{y} \in \mathbb{R}^p \; : \; h_k(\mathbf{y}) \geq 0, \quad k = m+1, \ldots, t \}$$

Let φ be a probability measure on $\mathbf{Y}$,

absoliutely continuous with respect to the Lebesgue measure, and whose moments $\gamma = (\gamma_\beta)$ with

$$\gamma_\beta = \int_{\mathbf{Y}} \mathbf{y}^\beta \, d\varphi(\mathbf{y}), \qquad \forall \beta \in \mathbb{N}^p,$$

are all known.

Primal semidefinite relaxations:

Let $v_k := \lceil (\deg h_k)/2 \rceil$ for every $k = 1, \ldots, t$ and let $i_0 := \max[\lceil (\deg f)/2 \rceil, \max_k v_k]$.

For $\mathbf{i} \geq i_0$, consider the semidefinite program:

$$\begin{array}{rll} \rho_{\mathbf{i}} = & \inf_{\mathbf{z}} & L_{\mathbf{z}}(f) \\ & \text{s.t.} & \mathbf{M}_{\mathbf{i}}(\mathbf{z}) \succeq 0 \\ & & \mathbf{M}_{\mathbf{i}-v_j}(h_j\,\mathbf{z}) \succeq 0, \quad j = 1, \ldots, t \\ & & L_{\mathbf{z}}(\mathbf{y}^{\beta}) = \gamma_{\beta}, \quad \forall\, \beta \in \mathbb{N}^p_{\mathbf{i}}. \end{array}$$

which is a relaxation of **P**, and

$$\rho_{i_0} \leq \cdots \leq \rho_{\mathbf{i}-1} \leq \rho_{\mathbf{i}} \leq \cdots \leq \rho.$$

Dual semidefinite relaxation

The dual reads:

$$\rho_{\mathbf{i}}^* = \sup_{g,(\sigma_i)} \int_{\mathbf{Y}} g(\mathbf{y})\, d\varphi(\mathbf{y}) \quad (= \sum_\beta g_\beta \gamma_\beta)$$

$$\text{s.t.} \quad f(\mathbf{x},\mathbf{y}) - g(\mathbf{y}) = \sigma_0(\mathbf{x},\mathbf{y}) + \sum_{j=1}^t \sigma_j(\mathbf{x},\mathbf{y})\, h_j(\mathbf{x},\mathbf{y})$$

$$p \in \mathbb{R}[\mathbf{y}];\ \sigma_j \in \Sigma[\mathbf{x},\mathbf{y}], \quad j = 1,\ldots,t$$
$$\deg p \leq 2i,\ \deg \sigma_j h_j \leq 2\mathbf{i}, \quad j = 1,\ldots,t$$

(Compare with $\mathbf{P}^*$ where $f(\mathbf{x},\mathbf{y}) - g(\mathbf{y}) \geq 0$ for all $\mathbf{x} \in \mathbf{K_y}$).

Theorem (Results for the Primal **P**)

(a) $\rho_{\mathbf{i}} \uparrow \rho$ *as* $\mathbf{i} \to \infty$.

(b) *Let* $\mathbf{z}^{\mathbf{i}}$ *be a nearly optimal solution, e.g. such that* $L_{\mathbf{z}^{\mathbf{i}}}(f) \leq \rho_{\mathbf{i}} + 1/\mathbf{i}$. *If for* φ*-almost all* $\mathbf{y} \in \mathbf{Y}$, $J(\mathbf{y})$ *is attained at a unique optimal solution* $\mathbf{x}^*(\mathbf{y})$*, then:*

$$\lim_{\mathbf{i}\to\infty} {z^{\mathbf{i}}}_{\alpha\beta} = \int_{\mathbf{Y}} \mathbf{y}^{\beta}\, \mathbf{x}^*(\mathbf{y})^{\alpha}\, d\varphi(\mathbf{y}), \qquad \forall\, \alpha \in \mathbb{N}^n,\ \beta \in \mathbb{N}^p.$$

In particular, for every $k = 1, \ldots, n$,

$$\lim_{i\to\infty} {z^{\mathbf{i}}}_{e(k)\beta} = \int_{\mathbf{Y}} \mathbf{y}^{\beta}\, \mathbf{x}^*_k(\mathbf{y})\, d\varphi(\mathbf{y}), \qquad \forall\, \beta \in \mathbb{N}^p,$$

where $e(k)_j = \delta_{j=k}$, $j = 1, \ldots, n$ *(with* $\delta_{\cdot}$ *being the Kronecker symbol).*

Theorem (Results for the Primal **P**)

(a) $\rho_{\mathbf{i}} \uparrow \rho$ *as* $\mathbf{i} \to \infty$.

(b) *Let* $\mathbf{z}^{\mathbf{i}}$ *be a nearly optimal solution, e.g. such that* $L_{\mathbf{z}^{\mathbf{i}}}(f) \leq \rho_{\mathbf{i}} + 1/\mathbf{i}$. *If for* φ*-almost all* $\mathbf{y} \in \mathbf{Y}$, $J(\mathbf{y})$ *is attained at a unique optimal solution* $\mathbf{x}^*(\mathbf{y})$, *then:*

$$\lim_{\mathbf{i}\to\infty} z^{\mathbf{i}}{}_{\alpha\beta} = \int_{\mathbf{Y}} \mathbf{y}^{\beta}\, \mathbf{x}^*(\mathbf{y})^{\alpha}\, d\varphi(\mathbf{y}), \qquad \forall\, \alpha \in \mathbb{N}^n,\ \beta \in \mathbb{N}^p.$$

In particular, for every $k = 1, \ldots, n$,

$$\lim_{i\to\infty} z^{\mathbf{i}}{}_{e(k)\beta} = \int_{\mathbf{Y}} \mathbf{y}^{\beta}\, \mathbf{x}_k^*(\mathbf{y})\, d\varphi(\mathbf{y}), \qquad \forall\, \beta \in \mathbb{N}^p,$$

where $e(k)_j = \delta_{j=k}$, $j = 1, \ldots, n$ *(with* $\delta_{\cdot}$ *being the Kronecker symbol).*

Theorem (Results for the Primal **P**)

(a) $\rho_{\mathbf{i}} \uparrow \rho$ *as* $\mathbf{i} \to \infty$.

(b) *Let* $\mathbf{z}^{\mathbf{i}}$ *be a nearly optimal solution, e.g. such that* $L_{\mathbf{z}^{\mathbf{i}}}(f) \leq \rho_{\mathbf{i}} + 1/\mathbf{i}$. *If for* φ*-almost all* $\mathbf{y} \in \mathbf{Y}$, $J(\mathbf{y})$ *is attained at a unique optimal solution* $\mathbf{x}^*(\mathbf{y})$, *then:*

$$\lim_{\mathbf{i}\to\infty} z^{\mathbf{i}}{}_{\alpha\beta} = \int_{\mathbf{Y}} \mathbf{y}^{\beta}\, \mathbf{x}^*(\mathbf{y})^{\alpha}\, d\varphi(\mathbf{y}), \qquad \forall\, \alpha \in \mathbb{N}^n,\ \beta \in \mathbb{N}^p.$$

In particular, for every $k = 1, \ldots, n$,

$$\lim_{i\to\infty} z^{\mathbf{i}}{}_{e(k)\beta} = \int_{\mathbf{Y}} \mathbf{y}^{\beta}\, \mathbf{x}_k^*(\mathbf{y})\, d\varphi(\mathbf{y}), \qquad \forall\, \beta \in \mathbb{N}^p,$$

where $e(k)_j = \delta_{j=k}$, $j = 1, \ldots, n$ *(with* $\delta_{\cdot}$ *being the Kronecker symbol).*

Assume $\mathbf{x}_k^*(\mathbf{y}) \geq 0$ on $\mathbf{Y}$.

Then for sufficiently large i,

$$\mathbf{z}^{\mathbf{i}}{}_{e(k)\beta} \approx \int_{\mathbf{Y}} \mathbf{y}^{\beta}\, \mathbf{x}_k^*(\mathbf{y})\, d\varphi(\mathbf{y}), \qquad \forall\, \beta \in \mathbb{N}^p.$$

That is, one has a good approximation of all moments of the measure $\mathbf{x}_k^*(\mathbf{y})\, d\varphi(\mathbf{y})$ with density $\mathbf{x}_k^*(\mathbf{y})$ on $\mathbf{Y}$.

Hence one may approximate the optimal k-th coordinate function $\mathbf{y} \mapsto \mathbf{x}_k^*(\mathbf{y})$ by e.g. maximum-entropy estimation methods.

Assume $\mathbf{x}_k^*(\mathbf{y}) \geq 0$ on $\mathbf{Y}$.

Then for sufficiently large i,

$$\mathbf{z^i}_{e(k)\beta} \approx \int_{\mathbf{Y}} \mathbf{y}^\beta \, \mathbf{x}_k^*(\mathbf{y}) \, d\varphi(\mathbf{y}), \qquad \forall \, \beta \in \mathbb{N}^p.$$

That is, one has a good approximation of all moments of the measure $\mathbf{x}_k^*(\mathbf{y}) \, d\varphi(\mathbf{y})$ with density $\mathbf{x}_k^*(\mathbf{y})$ on $\mathbf{Y}$.

Hence one may approximate the optimal k-th coordinate function $\mathbf{y} \mapsto \mathbf{x}_k^*(\mathbf{y})$ by e.g. maximum-entropy estimation methods.

Theorem (Results for the dual $\mathbf{P}^*$)

Consider the dual semidefinite relaxations. Then:

(a) $\rho^*_{\mathbf{i}} \uparrow \rho$ *as* $\mathbf{i} \to \infty$.

(b) *Let* $(p_{\mathbf{i}}, (\sigma^{\mathbf{i}}_j))$ *be a nearly optimal solution e.g.. such that* $\int_{\mathbf{Y}} p_{\mathbf{i}} d\varphi \geq \rho^*_{\mathbf{i}} - 1/\mathbf{i}$. *Then* $p_{\mathbf{i}} \leq J(\cdot)$ *and*

$$\lim_{\mathbf{i}\to\infty} \int_{\mathbf{Y}} |J(\mathbf{y}) - p_{\mathbf{i}}(\mathbf{y})| \, d\varphi(\mathbf{y}) = 0$$

Moreover if one defines

$$\tilde{p}_0 := p_0, \quad \mathbf{y} \mapsto \tilde{p}_{\mathbf{i}}(\mathbf{y}) \; := \; \max\left[\, \tilde{p}_{\mathbf{i}-1}(\mathbf{y}), p_{\mathbf{i}}(\mathbf{y}) \,\right], \quad \mathbf{i} = 1, 2, \ldots,$$

then $\tilde{p}_{\mathbf{i}} \to J(\cdot)$ φ*-almost uniformly on* $\mathbf{Y}$.

Theorem (Results for the dual $\mathbf{P}^*$)

Consider the dual semidefinite relaxations. Then:

(a) $\rho_{\mathbf{i}}^* \uparrow \rho$ *as* $\mathbf{i} \to \infty$.

(b) *Let* $(p_{\mathbf{i}}, (\sigma_j^{\mathbf{i}}))$ *be a nearly optimal solution e.g.. such that* $\int_{\mathbf{Y}} p_{\mathbf{i}} d\varphi \geq \rho_{\mathbf{i}}^* - 1/\mathbf{i}$. *Then* $p_{\mathbf{i}} \leq J(\cdot)$ *and*

$$\lim_{\mathbf{i}\to\infty} \int_{\mathbf{Y}} |J(\mathbf{y}) - p_{\mathbf{i}}(\mathbf{y})| \, d\varphi(\mathbf{y}) = 0$$

Moreover if one defines

$$\tilde{p}_0 := p_0, \quad \mathbf{y} \mapsto \tilde{p}_{\mathbf{i}}(\mathbf{y}) := \max\left[\tilde{p}_{\mathbf{i}-1}(\mathbf{y}), p_{\mathbf{i}}(\mathbf{y})\right], \quad \mathbf{i} = 1, 2, \ldots,$$

then $\tilde{p}_{\mathbf{i}} \to J(\cdot)$ φ*-almost uniformly on* $\mathbf{Y}$.

Theorem (Results for the dual $\mathbf{P}^*$)

Consider the dual semidefinite relaxations. Then:

(a) $\rho^*_{\mathbf{i}} \uparrow \rho$ *as* $\mathbf{i} \to \infty$.

(b) *Let* $(p_{\mathbf{i}}, (\sigma^{\mathbf{i}}_j))$ *be a nearly optimal solution e.g.. such that* $\int_{\mathbf{Y}} p_{\mathbf{i}} d\varphi \geq \rho^*_{\mathbf{i}} - 1/\mathbf{i}$. *Then* $p_{\mathbf{i}} \leq J(\cdot)$ *and*

$$\lim_{\mathbf{i}\to\infty} \int_{\mathbf{Y}} |J(\mathbf{y}) - p_{\mathbf{i}}(\mathbf{y})| \, d\varphi(\mathbf{y}) = 0$$

Moreover if one defines

$$\tilde{p}_0 := p_0, \quad \mathbf{y} \mapsto \tilde{p}_{\mathbf{i}}(\mathbf{y}) \;:=\; \max\left[\,\tilde{p}_{\mathbf{i}-1}(\mathbf{y}), p_{\mathbf{i}}(\mathbf{y})\,\right], \quad \mathbf{i} = 1, 2, \ldots,$$

then $\tilde{p}_{\mathbf{i}} \to J(\cdot)$ φ*-almost uniformly on* $\mathbf{Y}$.

Illustrative Examples

With $\mathbf{Y} := [0,1]$, let $\mathbf{K} := \{(\mathbf{x},\mathbf{y}) \,:\, 1 - {x_1}^2 - {x_2}^2 \geq 0\} \subset \mathbb{R}^2$, and $f(\mathbf{x},\mathbf{y}) := \mathbf{y}x_1 + (1-\mathbf{y})x_2$. One easily obtains:

$$J(\mathbf{y}) = -\sqrt{\mathbf{y}^2 + (1-\mathbf{y})^2},$$

and

$$\mathbf{x}_1^*(\mathbf{y}) \;=\; \frac{-\mathbf{y}}{\sqrt{\mathbf{y}^2+(1-\mathbf{y})^2}}; \quad \mathbf{x}_2^*(\mathbf{y}) \;=\; \frac{\mathbf{y}-1}{\sqrt{\mathbf{y}^2+(1-\mathbf{y})^2}}.$$

With 8 moments of the uniform distribution of $[0,1]$, one obtains:

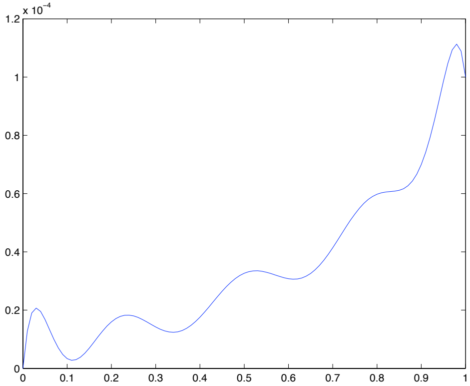

Figure: $J(\mathbf{y}) - p_4(\mathbf{y})$ on $[0, 1]$. (Scale is 10^{-4})

And the Boltzmann-Shannon maximum-entropy estimation of $\mathbf{x}^*(\mathbf{y})$ with 8 moments gives:

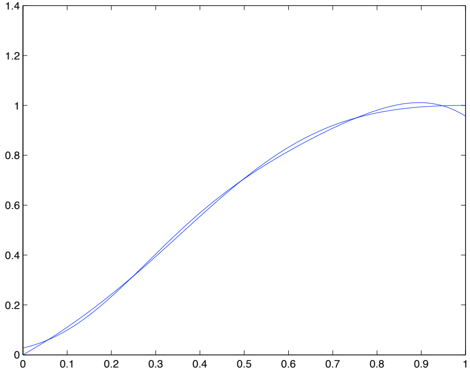

Figure: max-entropy estimate $g_1(\mathbf{y})$ vs $-\mathbf{x}^*(\mathbf{y}) = \mathbf{y}/\sqrt{\mathbf{y}^2 + (1-\mathbf{y})^2}$

Example 2:

$\mathbf{Y} = [0,1]$, $f(\mathbf{x},\mathbf{y}) := (1-2\mathbf{y})(x_1+x_2)$, and

$$\mathbf{K} := \{(\mathbf{x},\mathbf{y}) \,:\, \mathbf{y}{x_1}^2 + {x_2}^2 - \mathbf{y} <= 0;\; x_1^2 + \mathbf{y}x^2 - \mathbf{y} <= 0\}.$$

That is, for each $\mathbf{y} \in \mathbf{Y}$ the set $\mathbf{K}_\mathbf{y}$ is the intersection of two ellipsoids. $J(\mathbf{y}) = -2|1-2\mathbf{y}|\sqrt{\frac{\mathbf{y}}{1+\mathbf{y}}}$.

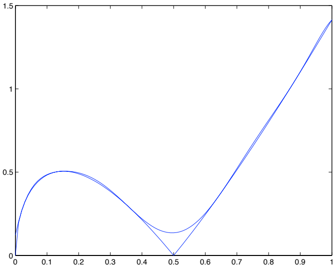

Figure: $-p_2(\mathbf{y})$ and $-J(\mathbf{y})$ on $[0,1]$ (4 moments)

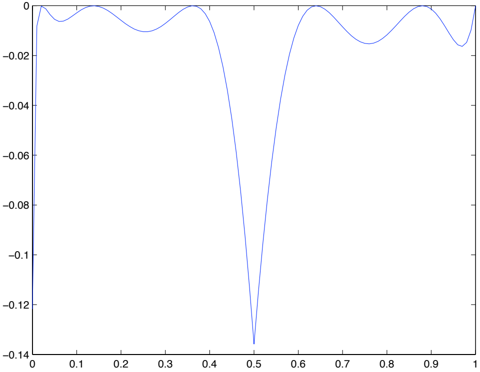

Figure: The curve $p_2(\mathbf{y}) - J(\mathbf{y})$ on $[0, 1]$ (4 moments)

Example 3:

Consider the following system of 4 quadratic equations in 4 variables and one parameter $y \in \mathbf{Y} = [0, 1]$:

$$x_1x_2 - x_1x_3 - x_4 = \mathbf{y} \qquad x_2x_3 - x_2x_4 - x_1 = \mathbf{y}$$
$$-x_1x_3 + x_3x_4 - x_2 = \mathbf{y} \qquad x_1x_4 - x_2x_4 - x_3 = \mathbf{y},$$

for which one wishes to compute the minimum norm $J(\mathbf{y})$ of real solutions as a function of $\mathbf{y} \in \mathbf{Y}$.

	$\mathbf{y}$=0.1	$\mathbf{y}$=0.5	$\mathbf{y}$=1
$J(\mathbf{y})$	0.0400	1.0000	2.0000
$p_6(\mathbf{y})$	0.0384	0.9264	1.9887
$p_8(\mathbf{y})$	0.0390	0.9395	2.0000

Table: $J(\mathbf{y})$ versus $p_k(\mathbf{y})$

A "joint+marginal" algorithm for Optimization

Given $\mathbf{K} \subset \mathbb{R}^n$ and $f \in \mathbb{R}[\mathbf{x}]$, consider the polynomial optimization problem $\mathbf{P}: \min_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K} \}$.

IDEA: Consider $\mathbf{x}_1$ as a parameter $\mathbf{y}$ in some interval $\mathbf{Y} \subset \mathbb{R}$

(to be determined, e.g., easily when $\mathbf{K}$ is convex), so that:

$$J(\mathbf{y}) = \min_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}; \mathbf{x}_1 = \mathbf{y} \}, \qquad \mathbf{y} \in \mathbf{Y}.$$

- Compute a polynomial approximation $p_k(\mathbf{y})$ of $J(\mathbf{y})$ via the k-th semidefinite relaxation of the "joint+marginal" approach.
- Minimize the univariate polynomial $p_k(\mathbf{y})$ on $\mathbf{Y}$ (easy! reduces to solving a single semidefinite program), and get $z_1 \in \mathbf{Y}$.
- In $\mathbf{P}$ fix $\mathbf{x}_1 := z_1$, and repeat for a $(n-1)$ optimization problem to obtain $\mathbf{x}_2 = z_2$, etc. until we get $(z_1, z_2, \ldots, z_n)$

A "joint+marginal" algorithm for Optimization

Given $\mathbf{K} \subset \mathbb{R}^n$ and $f \in \mathbb{R}[\mathbf{x}]$, consider the polynomial optimization problem $\mathbf{P}: \min_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$.

IDEA: Consider $\mathbf{x}_1$ as a parameter $\mathbf{y}$ in some interval $\mathbf{Y} \subset \mathbb{R}$

(to be determined, e.g., easily when $\mathbf{K}$ is convex), so that:

$$J(\mathbf{y}) = \min_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K};\ \mathbf{x}_1 = \mathbf{y}\}, \qquad \mathbf{y} \in \mathbf{Y}.$$

• Compute a polynomial approximation $p_k(\mathbf{y})$ of $J(\mathbf{y})$ via the k-th semidefinite relaxation of the "joint+marginal" approach.

• Minimize the univariate polynomial $p_k(\mathbf{y})$ on $\mathbf{Y}$ (easy! reduces to solving a single semidefinite program), and get $z_1 \in \mathbf{Y}$.

• In $\mathbf{P}$ fix $\mathbf{x}_1 := z_1$, and repeat for a $(n-1)$ optimization problem to obtain $\mathbf{x}_2 = z_2$, etc. until we get $(z_1, z_2, \ldots, z_n)$

A "joint+marginal" algorithm for Optimization

Given $\mathbf{K} \subset \mathbb{R}^n$ and $f \in \mathbb{R}[\mathbf{x}]$, consider the polynomial optimization problem $\mathbf{P}: \min_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$.

IDEA: Consider $\mathbf{x}_1$ as a parameter $\mathbf{y}$ in some interval $\mathbf{Y} \subset \mathbb{R}$

(to be determined, e.g., easily when $\mathbf{K}$ is convex), so that:

$$J(\mathbf{y}) = \min_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K};\ \mathbf{x}_1 = \mathbf{y}\}, \qquad \mathbf{y} \in \mathbf{Y}.$$

- Compute a polynomial approximation $p_k(\mathbf{y})$ of $J(\mathbf{y})$ via the k-th semidefinite relaxation of the "joint+marginal" approach.
- Minimize the univariate polynomial $p_k(\mathbf{y})$ on $\mathbf{Y}$ (easy! reduces to solving a single semidefinite program), and get $z_1 \in \mathbf{Y}$.
- In $\mathbf{P}$ fix $\mathbf{x}_1 := z_1$, and repeat for a $(n-1)$ optimization problem to obtain $\mathbf{x}_2 = z_2$, etc. until we get $(z_1, z_2, \ldots, z_n)$

A "joint+marginal" algorithm for Optimization

Given $\mathbf{K} \subset \mathbb{R}^n$ and $f \in \mathbb{R}[\mathbf{x}]$, consider the polynomial optimization problem $\mathbf{P}: \min_{\mathbf{x}} \{f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K}\}$.

IDEA: Consider $\mathbf{x}_1$ as a parameter $\mathbf{y}$ in some interval $\mathbf{Y} \subset \mathbb{R}$

(to be determined, e.g., easily when $\mathbf{K}$ is convex), so that:

$$J(\mathbf{y}) = \min_{\mathbf{x}} \{ f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K};\ \mathbf{x}_1 = \mathbf{y}\}, \qquad \mathbf{y} \in \mathbf{Y}.$$

- Compute a polynomial approximation $p_k(\mathbf{y})$ of $J(\mathbf{y})$ via the k-th semidefinite relaxation of the "joint+marginal" approach.
- Minimize the univariate polynomial $p_k(\mathbf{y})$ on $\mathbf{Y}$ (easy! reduces to solving a single semidefinite program), and get $z_1 \in \mathbf{Y}$.
- In $\mathbf{P}$ fix $\mathbf{x}_1 := z_1$, and repeat for a $(n-1)$ optimization problem to obtain $\mathbf{x}_2 = z_2$, etc. until we get $(z_1, z_2, \ldots, z_n)$

The rationale behind the "joint+marginal" algorithm:

- The larger k, the better the approximation of $J(\mathbf{y})$ by the univariate polynomial $p_k(\mathbf{y})$. And so in minimizing $p_k(\mathbf{y})$ over $\mathbf{Y}$ one has a good chance to obtain $z_1 \approx \mathbf{x}_1^*$, where $\mathbf{x}^*$ is a global minimizer of $\mathbf{P}$. And so at the end one may expect $z \approx \mathbf{x}^*$.
- The interest is to precisely have k not too large so as to handle relatively large size problems, and use the output point z as the initial point of a local minimization algorithm to next obtain a local minimizer $\tilde{\mathbf{x}} \in \mathbf{K}$ with reasonable hope that $\tilde{\mathbf{x}}$ is not far away of $\mathbf{x}^*$.
- Preliminary results are encouraging!

The rationale behind the "joint+marginal" algorithm:

- The larger k, the better the approximation of $J(\mathbf{y})$ by the univariate polynomial $p_k(\mathbf{y})$. And so in minimizing $p_k(\mathbf{y})$ over $\mathbf{Y}$ one has a good chance to obtain $z_1 \approx \mathbf{x}_1^*$, where $\mathbf{x}^*$ is a global minimizer of $\mathbf{P}$. And so at the end one may expect $z \approx \mathbf{x}^*$.
- The interest is to precisely have k not too large so as to handle relatively large size problems, and use the output point z as the initial point of a local minimization algorithm to next obtain a local minimizer $\tilde{\mathbf{x}} \in \mathbf{K}$ with reasonable hope that $\tilde{\mathbf{x}}$ is not far away of $\mathbf{x}^*$.
- Preliminary results are encouraging!

The rationale behind the "joint+marginal" algorithm:

- The larger k, the better the approximation of $J(\mathbf{y})$ by the univariate polynomial $p_k(\mathbf{y})$. And so in minimizing $p_k(\mathbf{y})$ over $\mathbf{Y}$ one has a good chance to obtain $z_1 \approx \mathbf{x}_1^*$, where $\mathbf{x}^*$ is a global minimizer of $\mathbf{P}$. And so at the end one may expect $z \approx \mathbf{x}^*$.
- The interest is to precisely have k not too large so as to handle relatively large size problems, and use the output point z as the initial point of a local minimization algorithm to next obtain a local minimizer $\tilde{\mathbf{x}} \in \mathbf{K}$ with reasonable hope that $\tilde{\mathbf{x}}$ is not far away of $\mathbf{x}^*$.
- Preliminary results are encouraging!

THANK YOU !!