Approximating Pareto Curves using Semidefinite Relaxations

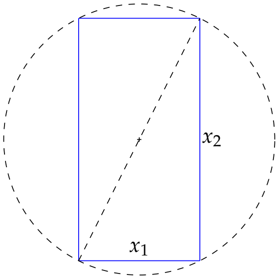
Victor MAGRON

Postdoc LAAS-CNRS
(Joint work with Didier Henrion and Jean-Bernard Lasserre)

Applications of Real Algebraic Geometry
2014 March 1

Multiobjective Polynomial Optimization

- Optimization Problems with several criteria in engineering, economics, applied mathematics.
- Design of a beam of length l, heigth x_1 and width x_2:
  1. **light** construction: minimize the volume lx_1x_2
  2. **cheap** construction: minimize the sectional area $\pi/4(x_1^2 + x_2^2)$
  3. under stress and nonnegativity constraints



Multiobjective Polynomial Optimization

- Let $f_1, f_2 \in \mathbb{R}_d[\mathbf{x}]$ two conflicting criteria
- Let $\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geqslant 0, \dots, g_m(\mathbf{x}) \geqslant 0\}$ a semialgebraic set

$$(\mathbf{P}) \left\{ \min_{\mathbf{x} \in \mathbf{S}} (f_1(\mathbf{x}) \; f_2(\mathbf{x}))^\top \right\}$$

Assumption

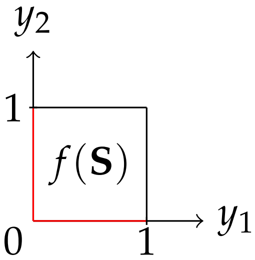
The image space $\mathbb{R}^2$ is partially ordered in a natural way ($\mathbb{R}^2_+$ is the ordering cone).

Multiobjective Polynomial Optimization

Definition

Let the previous assumption be satisfied. A point $\bar{\mathbf{x}} \in \mathbf{S}$ is called a *weakly Edgeworth-Pareto (EP) optimal point* of Problem **P**, when there is no $\mathbf{x} \in \mathbf{S}$ such that $f_j(\mathbf{x}) < f_j(\bar{\mathbf{x}}),\ \ j = 1, 2$.

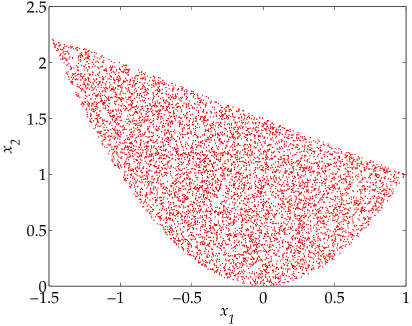
$$f_1(\mathbf{x}) := x_1 \ ,$$
$$f_2(\mathbf{x}) := x_2 \ ,$$
$$\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^2 : 0 \leqslant x_1 \leqslant 1,\ \ 0 \leqslant x_2 \leqslant 1\} \ .$$

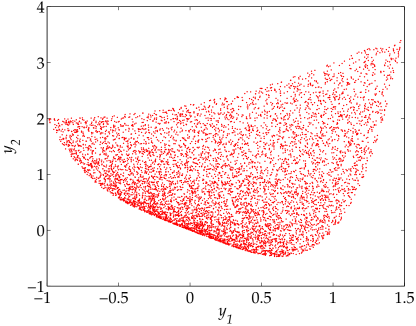


Some Examples: $f(\mathbf{S}) + \mathbb{R}^2_+$ is convex

$$g_1 := -x_1^2 + x_2 \ ,$$
$$g_2 := -x_1 - 2x_2 + 3 \ ,$$
$$\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^2 : g_1 \geqslant 0, g_2 \geqslant 0\} \ .$$

$$f_1 := -x_1 \ ,$$
$$f_2 := x_1 + x_2^2 \ .$$

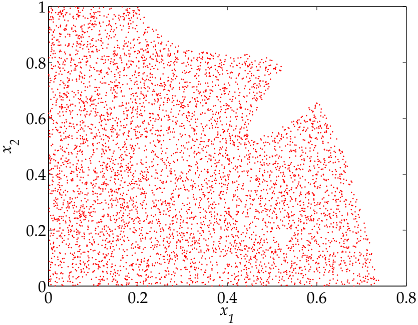


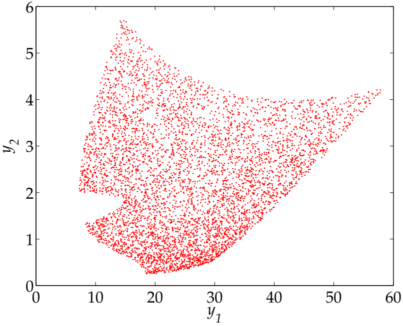


Some Examples: $f(\mathbf{S}) + \mathbb{R}^2_+$ is not convex

$$g_1 := -(x_1 - 2)^3/2 - x_2 + 2.5 \ ,$$
$$g_2 := -x_1 - x_2 + 8(-x_1 + x_2 + 0.65)^2 + 3.85 \ ,$$
$$\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^2 : g_1 \geqslant 0, g_2 \geqslant 0\} \ .$$

$$f_1 := (x_1 + x_2 - 7.5)^2/4 + (-x_1 + x_2 + 3)^2 \ ,$$
$$f_2 := (x_1 - 1)^2/4 + (x_2 - 4)^2/4 \ .$$





Scalarization Techniques

- Common workaround by reducing **P** to a scalar POP :

$$(\mathbf{P}^p_\lambda) \left\{ \min_{\mathbf{x} \in \mathbf{S}} f_\lambda(\mathbf{x}) := \left(\lambda |f_1(\mathbf{x}) - \mu_1|^p + (1-\lambda)|f_2(\mathbf{x}) - \mu_2|^p\right)^{\frac{1}{p}} \right\} ,$$

with the weight $\lambda \in [0,1]$ and the goals $\mu_1, \mu_2 \in \mathbb{R}$.

- Possible choice: $\mu_j < \min_{\mathbf{x} \in \mathbf{S}} f_j(\mathbf{x}), \;\; j = 1, 2.$

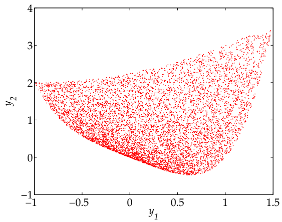
Scalarization Techniques

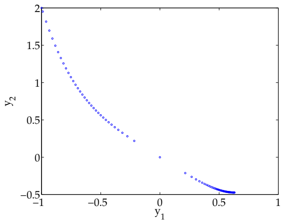
When $p = 1$, weighted sum formulation $\mathbf{P}^1_\lambda$:

$$f_\lambda(\mathbf{x}) := \lambda f_1(\mathbf{x}) + (1-\lambda) f_2(\mathbf{x})$$

Theorem ([Borwein 77], [Arrow-Barankin-Blackwell 53])

Assume that $f(\mathbf{S}) + \mathbb{R}^2_+$ is convex. A point $\bar{\mathbf{x}} \in \mathbf{S}$ is a weakly EP optimal point of Problem $\mathbf{P} \iff \exists \lambda$ such that $\bar{\mathbf{x}}$ is a solution of Problem $\mathbf{P}^1_\lambda$.





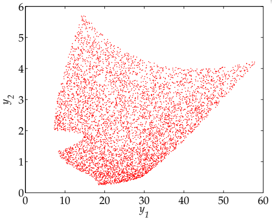
Scalarization Techniques

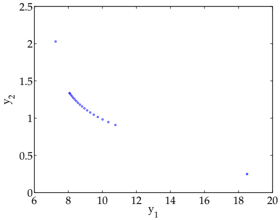
When $p = 1$, weighted sum formulation $\mathbf{P}_\lambda^1$:

$$f_\lambda(\mathbf{x}) := \lambda f_1(\mathbf{x}) + (1 - \lambda) f_2(\mathbf{x})$$

Theorem ([Borwein 77], [Arrow-Barankin-Blackwell 53])

Assume that $f(\mathbf{S}) + \mathbb{R}_+^2$ is convex. A point $\bar{\mathbf{x}} \in \mathbf{S}$ is a weakly EP optimal point of Problem $\mathbf{P} \iff \exists \lambda$ such that $\bar{\mathbf{x}}$ is a solution of Problem $\mathbf{P}_\lambda^1$.





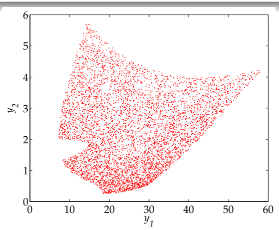
Scalarization Techniques

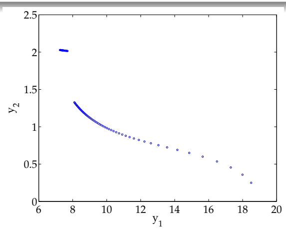
When $p = \infty$, weigthed Chebyshev approximation $\mathbf{P}_\lambda^\infty$

$$f_\lambda(\mathbf{x}) := \max\{\lambda(f_1(\mathbf{x}) - \mu_1), (1-\lambda)(f_2(\mathbf{x}) - \mu_2)\}$$

Theorem ([Jahn 87], [Bowman 76], [Steuer-Choo 83])

Suppose that $\forall \mathbf{x} \in \mathbf{S}, \mu_j < f_j(\mathbf{x}),\ j = 1, 2$. A point $\bar{\mathbf{x}} \in \mathbf{S}$ is a weakly EP optimal point of Problem $\mathbf{P} \iff \exists \lambda$ such that $\bar{\mathbf{x}}$ is a solution of Problem $\mathbf{P}_\lambda^\infty$.





Questions

- Is it mandatory to use discretization schemes?
- Can we approximate the Pareto curve in a relatively strong sense?

Contributions

Yes!

- We provide two numerical schemes that **avoid computing finitely many points**.
  1. The first **approximates the Pareto curve** $(f_1^*(\lambda), f_2^*(\lambda)), \lambda \in [0,1]$, with polynomials that minimize the L_2-norm $\int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda,\ j = 1, 2.$
  2. The second provides certified underestimators of the Pareto curve by computing **a hierarchy of outer approximations** of $f(\mathbf{S})$.

Outline

1. Parametric POP
2. Outer Approximations of $f(\mathbf{S})$
3. Perspectives

Preliminaries

Parametric POP $(\mathbf{P}_\lambda^1)$: $f^*(\lambda) := \min_{\mathbf{x} \in \mathbf{S}} f_\lambda(\mathbf{x})$

- Let $\hat{\mathbf{S}} := [0,1] \times \mathbf{S}$
- Let $\mathcal{M}(\hat{\mathbf{S}})$ the set of probability measures supported on $\hat{\mathbf{S}}$

$$(P) \begin{cases} \rho := \min_{\mu \in \mathcal{M}(\hat{\mathbf{S}})} & \int_{\hat{\mathbf{S}}} f_\lambda(\mathbf{x}) d\mu(\lambda, \mathbf{x}) \\ \text{s.t.} & \int_{\hat{\mathbf{S}}} \lambda^k d\mu(\lambda, \mathbf{x}) = 1/(1+k),\ k \in \mathbb{N} \end{cases} .$$

Preliminaries

Lemma (Corollary of [1, Theorem 2.2])

Problem (P) has an optimal solution $\mu^* \in \mathcal{M}(\hat{\mathbf{S}})$. Then,

$$\rho = \int_{\hat{\mathbf{S}}} f_\lambda(\mathbf{x}) d\mu^* = \int_0^1 f^*(\lambda) d\lambda \ .$$

Moreover, suppose that (P) has a unique global minimizer $\mathbf{x}^*(\lambda) \in \mathbf{S}$ and let $f_j^*(\lambda) := f_j(\mathbf{x}^*(\lambda)), \ j = 1, 2$. Then,

$$\rho = \int_0^1 [\lambda f_1^*(\lambda) + (1 - \lambda) f_2^*(\lambda)] d\lambda \ .$$

[1] J.B. Lasserre. A “joint + marginal” approach to parametric polynomial optimization (2010)

A hierarchy of semidefinite relaxations

- Let $g \in \mathbb{R}[\lambda, \mathbf{x}]$ with $g(\lambda, \mathbf{x}) := \sum_{k,\alpha} g_{k\alpha} \lambda^k \mathbf{x}^\alpha$.
- Consider the real sequence $\mathbf{z} = (z_{k\alpha}),\ (k, \alpha) \in \mathbb{N}_d^{n+1}$
- The entries $(i, \alpha), (j, \beta) \in \mathbb{N}_d^{n+1}$ of the localizing matrix $\mathbf{M}_d(g\,\mathbf{z})$ are:

$$\mathbf{M}_d(g\,\mathbf{z})((i, \alpha), (j, \beta)) := \sum_{r,\gamma} g_{r\gamma} z_{(i+j+r)(\alpha+\beta+\gamma)}$$

- Consider the linear functional $L_z(g) := \sum_{k,\alpha} g_{k\alpha} z_{k\alpha}$

A hierarchy of semidefinite relaxations

- Let $v_l := \deg g_l$, $l = 1, \ldots, m$.
- Let $k_{\max} := \max\{d, v_1, \ldots, v_m\}$.

Consider the semidefinite relaxations of (P):

$$(P_s) \begin{cases} \min_{\mathbf{z}} & L_{\mathbf{z}}(f_\lambda) \\ \text{s.t.} & \mathbf{M}_{s-v_l}(g_l \, \mathbf{z}) \succcurlyeq 0, \; l = 0, \ldots, m \;, \\ & L_{\mathbf{z}}(\lambda^k) = 1/(1+k), \; k = 0, \ldots, 2s - k_{\max} \;. \end{cases}$$

Polynomial underestimators of $f^*(\lambda)$

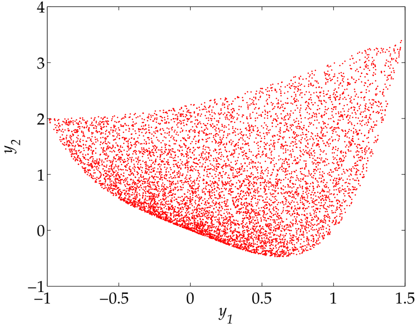
The dual SDP of (P_s) reads:

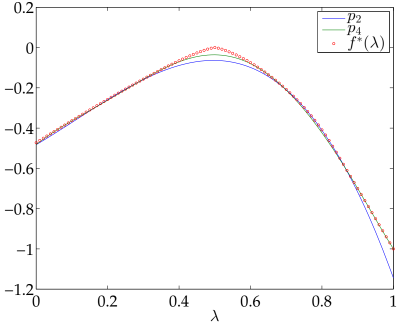
$$(D_s) \begin{cases} \min\limits_{p,(\sigma_l)} & \sum\limits_k p_k/(1+k) \\ \text{s.t.} & f_\lambda(\mathbf{x}) - p(\lambda) = \sum\limits_{l=0}^{m} \sigma_l(\lambda, \mathbf{x}) g_l(\mathbf{x}) \\ & p \in \mathbb{R}_{2s-k_{\max}}[\lambda],\ \sigma_l \in \Sigma[\lambda, \mathbf{x}],\ l = 0, \ldots, m\ , \\ & \deg(\sigma_l g_l) \leqslant 2s - k_{\max},\ l = 0, \ldots, m\ . \end{cases}$$

- The hierarchy (D_s) provides a sequence (p_s) of **polynomial underestimators** of $f^*(\lambda)$.
- $\lim\limits_{s \to \infty} \int_0^1 (f^*(\lambda) - p_s(\lambda)) d\lambda = 0$

Polynomial underestimators of $f^*(\lambda)$

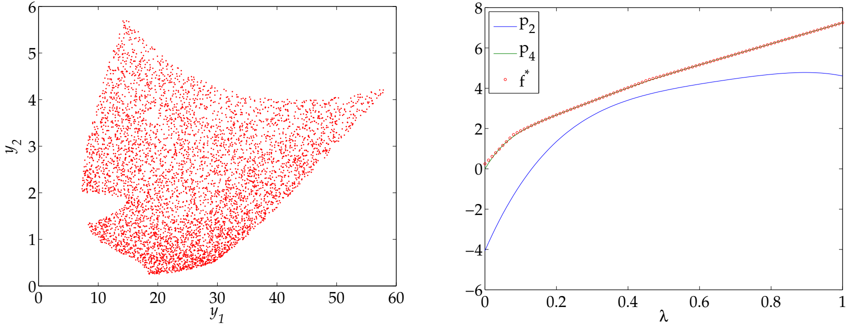
On the convex example:





Polynomial underestimators of $f^*(\lambda)$

On the non-convex example:



An inverse problem from generalized moments

Lemma (Corollary of [1, Theorem 3.3])

Assume that for a.a. $\lambda \in [0,1]$, Problem (P) has a unique global optimizer $\mathbf{x}^*(\lambda)$ and let $\mathbf{z}^s$ be an optimal solution of (P_s). Then,

$$\lim_{s \to \infty} z^s_{k\alpha} = \int_0^1 \lambda^k (\mathbf{x}^*(\lambda))^\alpha d\lambda, \ \ k \in \mathbb{N} \ .$$

In particular,

$$m_j^k := \lim_{s \to \infty} \sum_\alpha f_{j\alpha} z^s_{k\alpha} = \int_0^1 \lambda^k f_j^*(\lambda) d\lambda, \ \ j = 1, 2, \ \ k \in \mathbb{N} \ .$$

[1] J.B. Lasserre. A “joint + marginal” approach to parametric polynomial optimization (2010)

An inverse problem from generalized moments

Let $2s' := 2s - k_{\max}$. One can compute:

- Approximation of the vector $\mathbf{m}_j := (m_j^k)$
- Approximations of $f_j^*(\lambda)$, $j = 1, 2$, by solving:

$$\min_{h \in \mathbb{R}_{2s'}[\lambda]} \left\{ \int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda \right\}, \; j = 1, 2 \; .$$

An inverse problem from generalized moments

Theorem

The Problem $\min_{h \in \mathbb{R}_{2s'}[\lambda]} \left\{ \int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda \right\}$ has an optimal solution $\mathbf{p}_{sj} \in \mathbb{R}_{2s'}[\lambda]$, whose vector of coefficients is $\mathbf{p}_{sj} = \mathbf{H}_s^{-1} \mathbf{m}_j$, $j = 1, 2$, where $\mathbf{H}_s \in \mathcal{S}^{2s'+1}$ is the Hankel matrix, whose entries are defined by:

$$\mathbf{H}_s(a, b) := 1/(1 + a + b), \; a, b = 0, \ldots, 2s', \; .$$

An inverse problem from generalized moments

Proof.

$$\int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda = \underbrace{\int_0^1 f_j^*(\lambda)^2 d\lambda}_{A} - 2 \underbrace{\int_0^1 f_j^*(\lambda) h(\lambda) d\lambda}_{B} + \underbrace{\int_0^1 h(\lambda)^2 d\lambda}_{C},$$

$$B = \mathbf{h}'\mathbf{m}_j, \quad C = \mathbf{h}'\mathbf{H}_s\mathbf{h} \quad ,$$

thus the problem can be reformulated as:

$$\min_{\mathbf{h}} \{\mathbf{h}'\mathbf{H}_s\mathbf{h} - 2\mathbf{h}'\mathbf{m}_j\}, \quad j = 1, 2 \quad .$$

An inverse problem from generalized moments

Proof.

$$\int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda = \underbrace{\int_0^1 f_j^*(\lambda)^2 d\lambda}_{A} - 2\underbrace{\int_0^1 f_j^*(\lambda) h(\lambda) d\lambda}_{B} + \underbrace{\int_0^1 h(\lambda)^2 d\lambda}_{C},$$

$$B = \mathbf{h}'\mathbf{m}_j, \;\; C = \mathbf{h}'\mathbf{H}_s\mathbf{h} \;\; ,$$

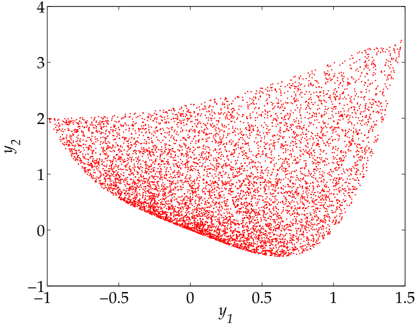
thus the problem can be reformulated as:

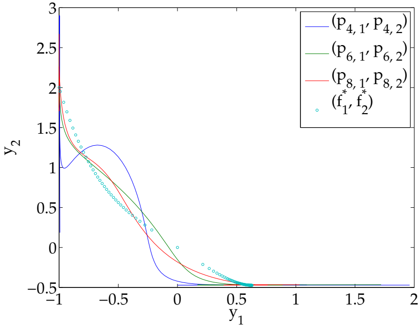
$$\min_{\mathbf{h}} \{\mathbf{h}'\mathbf{H}_s\mathbf{h} - 2\mathbf{h}'\mathbf{m}_j\}, \;\; j = 1, 2 \;\; .$$

□

With the weighted sum approach

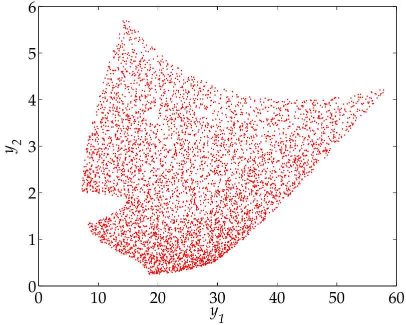
On the convex example:

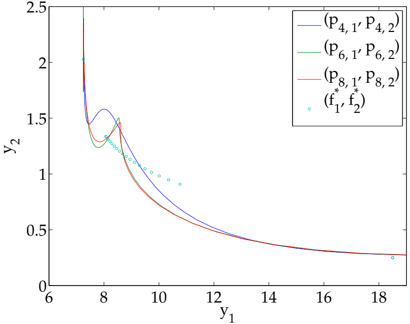




With the weighted sum approach

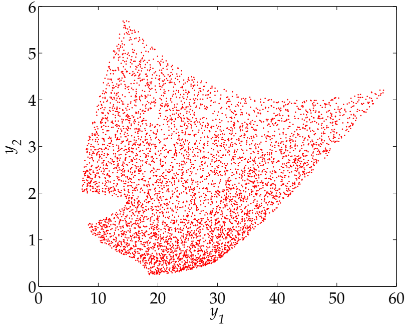
On the non-convex example:

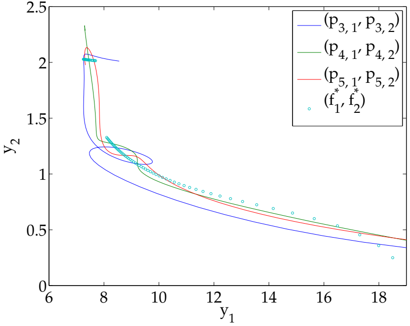




With the weigthed Chebyshev approximation

On the non-convex example:





Approximation of sets defined with "∃"

Let $\mathbf{B} \subset \mathbb{R}^2$ be the unit ball and assume that $f(\mathbf{S}) \subset \mathbf{B}$.

- Another point of view:

$$f(\mathbf{S}) = \{\mathbf{y} \in \mathbf{B} : \exists \mathbf{x} \in \mathbf{S} \text{ s.t. } h(\mathbf{x}, \mathbf{y}) \leqslant 0\} \enspace ,$$

  with

$$h(\mathbf{x}, \mathbf{y}) := (y_1 - f_1(\mathbf{x}))^2 + (y_2 - f_2(\mathbf{x}))^2 \enspace .$$

- Approximate $f(\mathbf{S})$ as closely as desired by a sequence of sets of the form :

$$\Theta_k := \{\mathbf{y} \in \mathbf{B} : J_k(\mathbf{y}) \leqslant 0\} \enspace ,$$

  for some polynomials $J_k \in \mathbb{R}_{2k}[\mathbf{y}]_{2d}$.

Approximation of sets defined with "∃"

- Let $g_0 := 1$ and $\mathbf{Q}_k(\mathbf{S})$ be the k-truncated quadratic module generated by $g_0, \dots, g_m$:

$$\mathbf{Q}_k(\mathbf{S}) = \Big\{ \sum_{l=0}^{m} \sigma_l(\mathbf{x}, \mathbf{y}) g_l(\mathbf{x}), \ \text{with } \sigma_l \in \Sigma_{k-v_l}[\mathbf{x}, \mathbf{y}] \Big\}$$

- Define $H(\mathbf{y}) := \min_{\mathbf{x} \in \mathbf{S}} h(\mathbf{x}, \mathbf{y})$
- Hierarchy of Semidefinite programs:

$$\rho_k := \min_{J \in \mathbb{R}_{2k}[\mathbf{y}], \sigma_l} \Big\{ \int_{\mathbf{B}} (H - J) d\mathbf{y} : h - J \in \mathbf{Q}_k(\mathbf{S}) \Big\} \ .$$

Yet another SOS program with an optimal solution $J_k \in \mathbb{R}_{2k}[\mathbf{y}]$!

A hierarchy of outer approximations of $f(\mathbf{S})$

From the definition of J_k, the sublevel sets

$$\Theta_k := \{\mathbf{y} \in \mathbf{B} : J_k(\mathbf{y}) \leqslant 0\} \supset f(\mathbf{S}), \;\; k \geqslant k_{\max} \;\; ,$$

provide a sequence of certified outer approximations of $f(\mathbf{S})$.

It comes from the following:

$$\forall (\mathbf{x}, \mathbf{y}) \in \mathbf{S} \times \mathbf{B}, J(\mathbf{y}) \leqslant h(\mathbf{x}, \mathbf{y}) \Longleftrightarrow \forall \mathbf{y}, J(\mathbf{y}) \leqslant H(\mathbf{y}) \;\; .$$

Strong convergence property

Theorem [Lasserre]

1. The sequence of underestimators $(J_k)_{k \geqslant k_{\max}}$ converges to H w.r.t the $L_1(\mathbf{B})$-norm:

$$\lim_{k \to \infty} \int_{\mathbf{B}} |H - J_k| d\mathbf{y} = 0 \enspace .$$

2. When the contour set $\{\mathbf{y} \in \mathbf{B} : H(\mathbf{y}) = 0\}$ has a null Lebesgue measure, then

$$\lim_{k \to \infty} V(\Theta_k \backslash f(\mathbf{S})) = 0 \enspace .$$

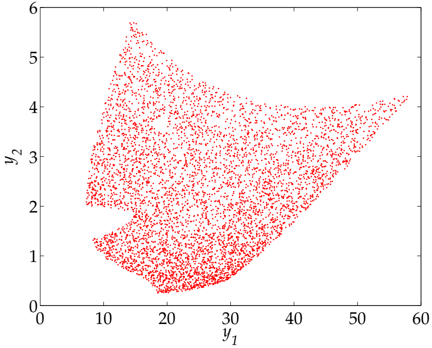
Strong convergence property

Theorem [Lasserre]

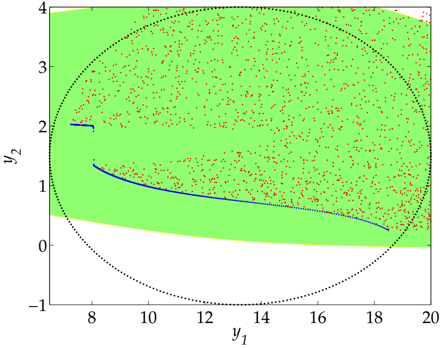
1. The sequence of underestimators $(J_k)_{k \geqslant k_{\max}}$ converges to H w.r.t the $L_1(\mathbf{B})$-norm:
$$\lim_{k \to \infty} \int_{\mathbf{B}} |H - J_k| d\mathbf{y} = 0 \ .$$

2. When the contour set $\{\mathbf{y} \in \mathbf{B} : H(\mathbf{y}) = 0\}$ has a null Lebesgue measure, then
$$\lim_{k \to \infty} V(\Theta_k \backslash f(\mathbf{S})) = 0 \ .$$

Back to the non-convex example

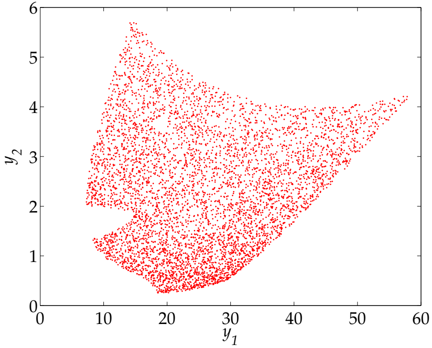


$f(\mathbf{S})$

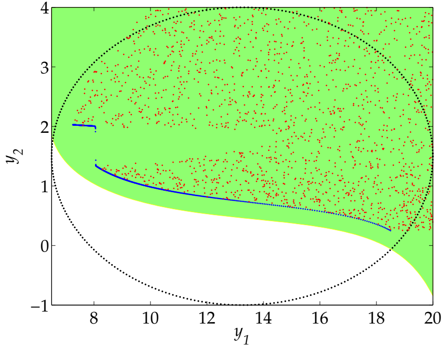


Θ_2

Back to the non-convex example

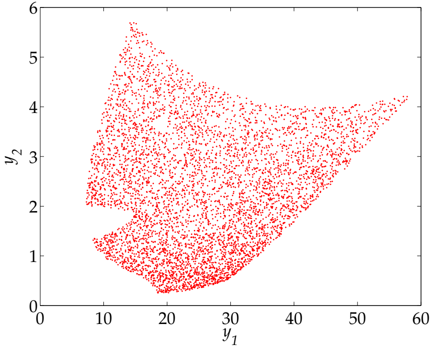


$f(\mathbf{S})$

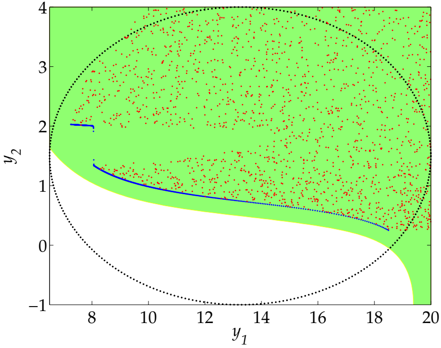


Θ_3

Back to the non-convex example

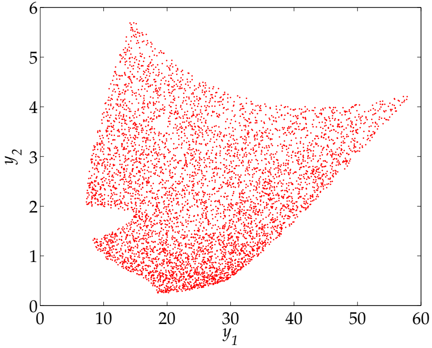


$f(\mathbf{S})$

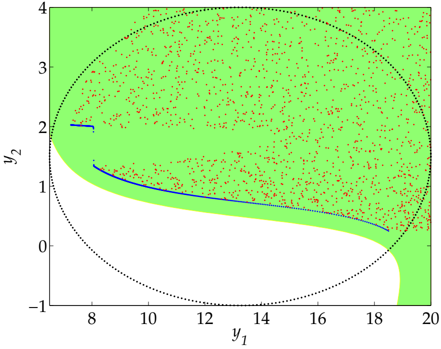


Θ_4

Back to the non-convex example

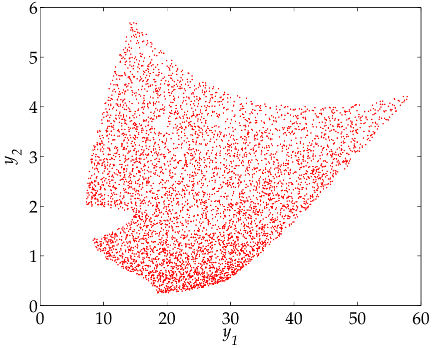


$f(\mathbf{S})$

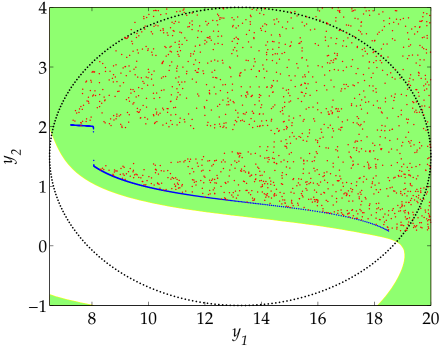


Θ_5

Back to the non-convex example

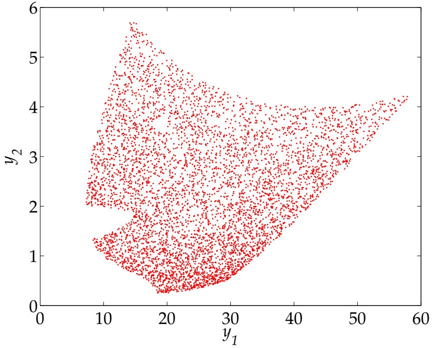


$f(\mathbf{S})$

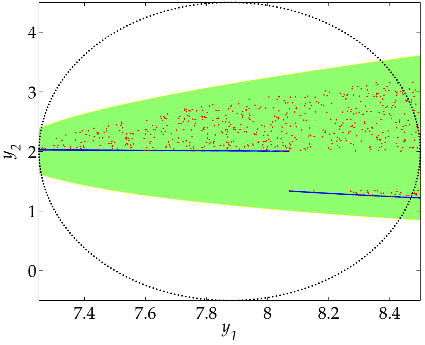


Θ_6

Branch and Bound: Zoom on the left

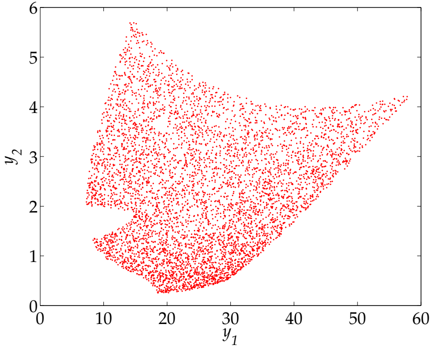


$f(\mathbf{S})$

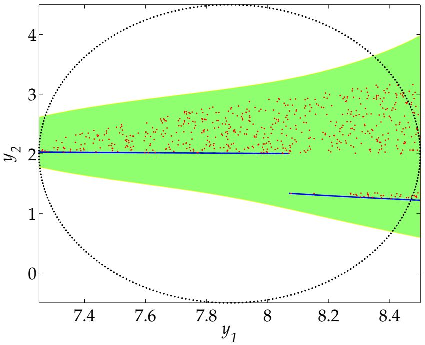


Θ_2

Branch and Bound: Zoom on the left

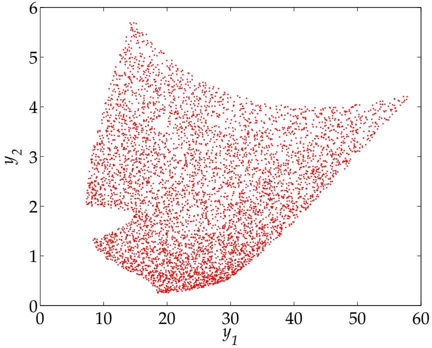


$f(\mathbf{S})$

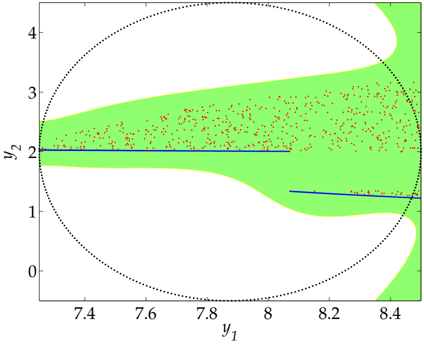


Θ_3

Branch and Bound: Zoom on the left

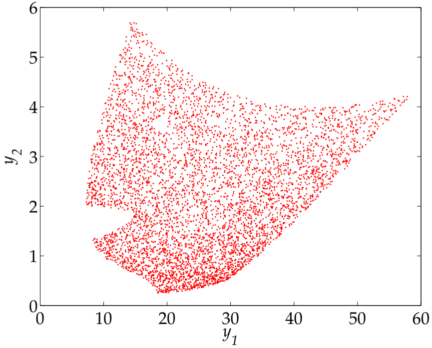


$f(\mathbf{S})$

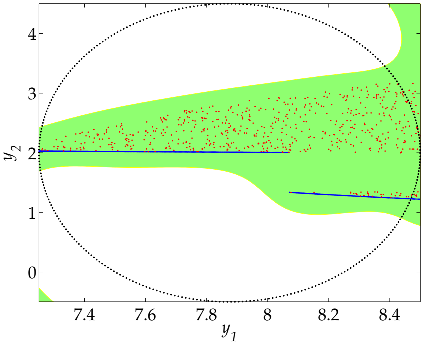


Θ_4

Branch and Bound: Zoom on the left

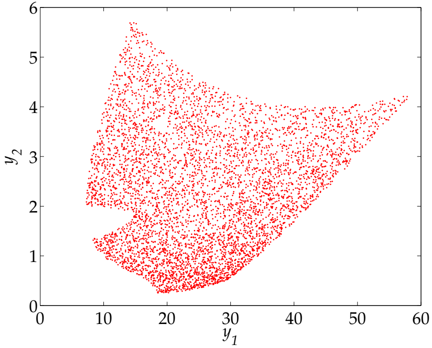


$f(\mathbf{S})$

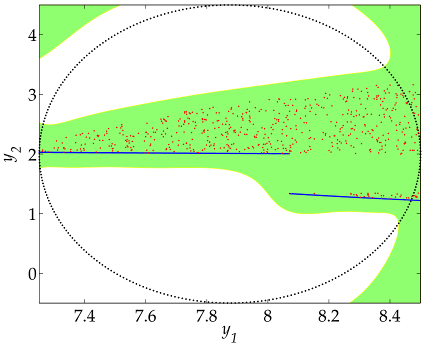


Θ_5

Branch and Bound: Zoom on the left



$f(\mathbf{S})$



Θ_6

Transcendental conflicting criteria

Now, consider the following Problem:

$$(\mathbf{P}) \left\{ \min_{\mathbf{x} \in \mathbf{S}} \quad (f_1(\mathbf{x})\, f_2(\mathbf{x}))^\top \ . \right\}$$

with transcendental criteria f_1, f_2.

- Generalization of the single criterion problem $\min_{\mathbf{x} \in \mathbf{S}} f(\mathbf{x})$
- Hard to combine SOS hierarchies with Taylor/Chebyshev approximations

Transcendental conflicting criteria

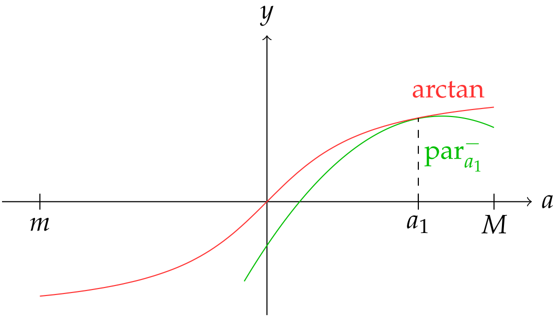
Definition: Semiconvex function

Let $\gamma \geqslant 0$. A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be γ-*semiconvex* if the function $\mathbf{x} \mapsto \phi(\mathbf{x}) + \frac{\gamma}{2}\|\mathbf{x}\|_2^2$ is convex.

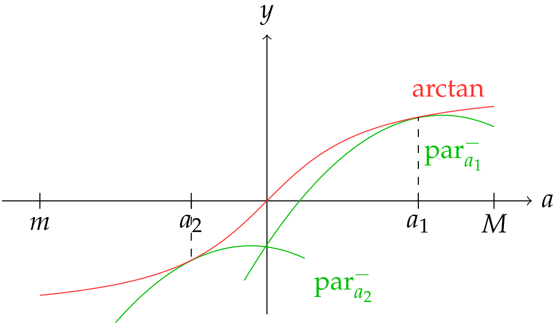
Proposition (by Legendre-Fenchel duality)

The set of functions $f : \mathbb{R}^n \to \mathbb{R}$ which can be written as the max-plus linear combination $f = \sup_{w \in \mathcal{B}}(a(w) + w)$ for some function $a : \mathcal{B} \to \mathbb{R} \cup \{-\infty\}$ is precisely the set of lower semicontinuous γ-semiconvex functions.

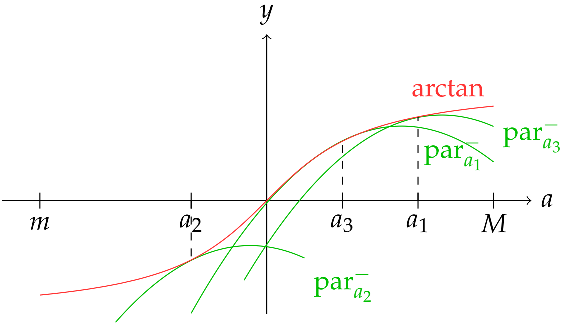
Transcendental conflicting criteria



Transcendental conflicting criteria



Transcendental conflicting criteria



Sublevel sets of semialgebraic underestimators

The sublevel sets

$$\Theta_k := \{\mathbf{y} \in \mathbf{B} : J_k(\mathbf{y}) \leqslant 0\} \supset f(\mathbf{S}), \;\; k \geqslant k_{\max} \;\; ,$$

provide a sequence of certified outer approximations of $f(\mathbf{S})$.

To avoid Branch and bound iterations (“Zooms”), one could underestimate H with a rational function

$$J := F/1 + \sigma \;\; ,$$

with $F \in \mathbb{R}_{2k}[\mathbf{y}], \sigma \in \Sigma_{k_0}[\mathbf{y}]$.

Thanks for your attention!

Kiitos Alexander, Cordian!!!