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

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

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

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

![Diagram of the set S and the function f(S)]
Some Examples: \( f(S) + \mathbb{R}^2_+ \) is convex

\[
\begin{align*}
g_1 &:= -x_1^2 + x_2 , \\
g_2 &:= -x_1 - 2x_2 + 3 , \\
S &:= \{ x \in \mathbb{R}^2 : g_1 \geq 0, g_2 \geq 0 \} .
\end{align*}
\]

\[
\begin{align*}
f_1 &:= -x_1 , \\
f_2 &:= x_1 + x_2^2 .
\end{align*}
\]
Some Examples: $f(S) + \mathbb{R}^2_+$ is not convex

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

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

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

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

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

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

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

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

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

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

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

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

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

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

When $p = \infty$, weighted Chebyshev approximation $P^\infty_{\lambda}$

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

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

Suppose that $\forall x \in S, \mu_j < f_j(x), j = 1, 2$. A point $\bar{x} \in S$ is a weakly EP optimal point of Problem $P$ $\iff \exists \lambda$ such that $\bar{x}$ is a solution of Problem $P^\infty_{\lambda}$. 
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
     \[ \int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda, \quad j = 1, 2. \]
  2. The second provides certified underestimators of the Pareto curve by computing a hierarchy of outer approximations of $f(S)$. 
Outline

1. Parametric POP
2. Outer Approximations of $f(S)$
3. Perspectives
Preliminaries

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

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

\[
(P) \left\{ \begin{array}{l}
\rho := \min_{\mu \in \mathcal{M}(\hat{S})} \int_{\hat{S}} f_\lambda(x) d\mu(\lambda, x) \\
\text{s.t. } \int_{\hat{S}} \lambda^k d\mu(\lambda, x) = 1/(1 + k), \ k \in \mathbb{N}
\end{array} \right.
\]
Preliminaries

Lemma (Corollary of [1, Theorem 2.2])

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

\[
\rho = \int_{\hat{S}} f_{\lambda}(x) d\mu^* = \int_{0}^{1} f^*(\lambda) d\lambda.
\]

Moreover, suppose that \((P)\) has a unique global minimizer \(x^*(\lambda) \in S\) and let \(f^*_j(\lambda) := f_j(x^*(\lambda))\), \(j = 1, 2\). 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 $g \in \mathbb{R}[\lambda, x]$ with $g(\lambda, x) := \sum_{k, \alpha} g_{k\alpha} \lambda^k x^\alpha$.

- Consider the real sequence $z = (z_{k\alpha}), (k, \alpha) \in \mathbb{N}_d^{n+1}$

- The entries $(i, \alpha), (j, \beta) \in \mathbb{N}_d^{n+1}$ of the localizing matrix $M_d(g z)$ are:
  $$M_d(g z)((i, \alpha), (j, \beta)) := \sum_{r, \gamma} g_{r\gamma} z_{(i+j+r)(\alpha+\beta+\gamma)}$$

- 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 $k_{\text{max}} := \max\{d, v_1, \ldots, v_m\}$.

Consider the semidefinite relaxations of $(P)$:

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

The dual SDP of $(P_s)$ reads:

$$
(D_s) \begin{cases} 
\min_{p,(\sigma_l)} \sum_k p_k / (1 + k) \\
\text{s.t. } f_\lambda(x) - p(\lambda) = \sum_{l=0}^m \sigma_l(\lambda, x) g_l(x) \\
\quad p \in \mathbb{R}_{2s-k_{\max}}[\lambda], \sigma_l \in \Sigma[\lambda, x], l = 0, \ldots, m, \\
\quad \deg(\sigma_l g_l) \leq 2s - k_{\max}, l = 0, \ldots, m.
\end{cases}
$$

- The hierarchy $(D_s)$ provides a sequence $(p_s)$ of polynomial underestimators of $f^*(\lambda)$.
- $\lim_{s \to \infty} \int_0^1 (f^*(\lambda) - p_s(\lambda)) d\lambda = 0$
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 \( x^*(\lambda) \) and let \( z^s \) be an optimal solution of \((P_s)\). Then,

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

In particular,

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

\(^1\)J.B. Lasserre. A “joint + marginal” approach to parametric polynomial optimization (2010)
An inverse problem from generalized moments

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

- Approximation of the vector $m_j := (m_j^k)$
- 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\( p_{sj} \in \mathbb{R}_{2s'}[\lambda] \), whose vector of coefficients is \( p_{sj} = H_s^{-1}m_j \), \( j = 1, 2 \), where \( H_s \in S^{2s'+1} \) is the Hankel matrix, whose entries are defined by:

\[
H_s(a, b) := 1 / (1 + a + b), \quad a, b = 0, \ldots, 2s',
\]
An inverse problem from generalized moments

Proof.

\[ \int_0^1 (f^*_j(\lambda) - h(\lambda))^2 d\lambda = \int_0^1 f^*_j(\lambda)^2 d\lambda - 2 \int_0^1 f^*_j(\lambda) h(\lambda) d\lambda + \int_0^1 h(\lambda)^2 d\lambda, \]

where

\[ A = \int_0^1 f^*_j(\lambda)^2 d\lambda, \]

\[ B = \int_0^1 f^*_j(\lambda) h(\lambda) d\lambda, \]

\[ C = \int_0^1 h(\lambda)^2 d\lambda. \]

Thus, the problem can be reformulated as:

\[ \min_h \{ h'H_s h - 2h'm_j \}, \quad j = 1, 2. \]
An inverse problem from generalized moments

Proof.

\[
\int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda = \int_0^1 f_j^*(\lambda)^2 d\lambda - 2 \int_0^1 f_j^*(\lambda)h(\lambda) d\lambda + \int_0^1 h(\lambda)^2 d\lambda,
\]

\[
B = h'm_j, \quad C = h'H_s h,
\]

thus the problem can be reformulated as:

\[
\min_h \{h'H_s h - 2h'm_j\}, \quad j = 1, 2.
\]
With the weighted sum approach

On the convex example:
With the weighted sum approach

On the non-convex example:
With the weighted Chebyshev approximation

On the non-convex example:
Approximation of sets defined with "∃"

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

- Another point of view:
  
  $$f(S) = \{ y \in B : \exists x \in S \text{ s.t. } h(x,y) \leq 0 \} ,$$

  with
  
  $$h(x,y) := (y_1 - f_1(x))^2 + (y_2 - f_2(x))^2 .$$

- Approximate $f(S)$ as closely as desired by a sequence of sets of the form:
  
  $$\Theta_k := \{ y \in B : J_k(y) \leq 0 \} ,$$

  for some polynomials $J_k \in \mathbb{R}_{2k}[y]_{2d}$.
Approximation of sets defined with "∃"

Let $g_0 := 1$ and $Q_k(S)$ be the $k$-truncated quadratic module generated by $g_0, \ldots, g_m$:

$$Q_k(S) = \left\{ \sum_{l=0}^{m} \sigma_l(x, y)g_l(x), \text{ with } \sigma_l \in \Sigma_{k-v_l}[x, y] \right\}$$

Define $H(y) := \min_{x \in S} h(x, y)$

Hierarchy of Semidefinite programs:

$$\rho_k := \min_{J \in \mathbb{R}_{2k}[y], \sigma_l} \left\{ \int_B (H - J)dy : h - J \in Q_k(S) \right\}.$$ 

Yet another SOS program with an optimal solution $J_k \in \mathbb{R}_{2k}[y]$!
A hierarchy of outer approximations of $f(S)$

From the definition of $J_k$, the sublevel sets

$$\Theta_k := \{y \in \mathcal{B} : J_k(y) \leq 0\} \supset f(S), \quad k \geq k_{\text{max}},$$

provide a sequence of certified outer approximations of $f(S)$.

It comes from the following:

$$\forall (x, y) \in S \times \mathcal{B}, J(y) \leq h(x, y) \iff \forall y, J(y) \leq H(y).$$
Strong convergence property

Theorem [Lasserre]

1. The sequence of underestimators \((J_k)_{k \geq k_{\text{max}}}\) converges to \(H\) w.r.t. the \(L_1(\mathcal{B})\)-norm:
   \[
   \lim_{k \to \infty} \int_{\mathcal{B}} |H - J_k| \, dy = 0 .
   \]

2. When the contour set \(\{y \in \mathcal{B} : H(y) = 0\}\) has a null Lebesgue measure, then
   \[
   \lim_{k \to \infty} V(\Theta_k \setminus f(S)) = 0 .
   \]
Strong convergence property

Theorem [Lasserre]

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

\[
\lim_{k \to \infty} \int_B |H - J_k| \, dy = 0 .
\]

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

\[
\lim_{k \to \infty} V(\Theta_k \backslash f(S)) = 0 .
\]
Back to the non-convex example
Back to the non-convex example
Back to the non-convex example

\[ f(S) \]

\[ \Theta_4 \]
Back to the non-convex example

\[ f(S) \]

\[ \Theta_5 \]
Back to the non-convex example

\[ f(S) \]
Branch and Bound: Zoom on the left

$f(S)$

$\Theta_2$
Branch and Bound: Zoom on the left

\[ f(S) \]

\[ \Theta_3 \]
Branch and Bound: Zoom on the left

\[ f(S) \]

\[ \Theta_4 \]
Branch and Bound: Zoom on the left

\[ f(S) \]
Branch and Bound: Zoom on the left

\[
f(S)
\]

\[\Theta_6\]
Transcendental conflicting criteria

Now, consider the following Problem:

\[
(P) \left\{ \min_{x \in S} (f_1(x) f_2(x))^\top \right. \}
\]

with transcendental criteria \( f_1, f_2 \).

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

Definition: Semiconvex function

Let $\gamma \geq 0$. A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be $\gamma$-semiconvex if the function $x \mapsto \phi(x) + \frac{\gamma}{2} \|x\|^2_2$ 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 $f = \sup_{w \in B} (a(w) + w)$ for some function $a : B \to \mathbb{R} \cup \{-\infty\}$ is precisely the set of lower semicontinuous $\gamma$-semiconvex functions.
Transcendental conflicting criteria
Transcendental conflicting criteria
Transcendental conflicting criteria

\[
y = \arctan(m \cdot a_1 a_2 a_3) - a_1 - a_2 - a_3
\]
Sublevel sets of semialgebraic underestimators

The sublevel sets

$$\Theta_k := \{ y \in B : J_k(y) \leq 0 \} \supset f(S), \ k \geq k_{\text{max}},$$

provide a sequence of certified outer approximations of $f(S)$.

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

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

with $F \in \mathbb{R}_{2k}[y], \sigma \in \Sigma_{k_0}[y]$. 
Thanks for your attention!

Kiitos Alexander, Cordian!!!