Approximating Pareto Curves using Semidefinite Relaxations

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

Optimisation Non Linéaire en Variables Continues et Discrètes
18 June 2014
Multiobjective Polynomial Optimization

- Optimization Problems with several criteria in engineering, economics, applied mathematics.
- Design of a beam of length $l$, height $x_1$ 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

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

**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 an *Edgeworth-Pareto (EP) optimal point* of Problem \( \mathbf{P} \), when there is no \( x \in S \) such that \( f_j(x) \leq f_j(\bar{x}), \ j = 1, 2 \) and \( f(x) \neq f(\bar{x}) \).

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

\[
\begin{align*}
f_1(x) & := x_1 , \\
f_2(x) & := x_2 , \\
S & := \{ x \in \mathbb{R}^2 : 0 \leq x_1 \leq 1, \ 0 \leq x_2 \leq 1 \} .
\end{align*}
\]
Pareto Curve

**Definition**

The image set of weakly Edgeworth-Pareto optimal points is called the \textit{Pareto curve}.
Some Examples: $f(S) + \mathbb{R}^2_+$ is convex

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

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

\begin{align*}
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 .
\end{align*}
Scalarization Techniques

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

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

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

- Possible choice: $\mu_j < \min_{x \in S} f_j(x), \ j = 1, 2$. 
Weighted convex sum approximation: method (a)

\[ (P^1_\lambda) : \quad f^1(\lambda) := \min_{x \in S} f^1(\lambda, x) \]

\[ f^1(\lambda, x) := \lambda f_1(x) + (1 - \lambda) f_2(x) \]

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

Assume that \( f(S) + \mathbb{R}^2_+ \) is convex. A point \( \bar{x} \in S \) is an EP optimal point of Problem \( P \iff \exists \lambda \) such that \( \bar{x} \) is an image unique solution of Problem \( P^1_\lambda \).
Weighted convex sum approximation: method (a)

\[(P_{\lambda}^1) : \quad f^1(\lambda) := \min_{x \in S} f^1(\lambda, x)\]

\[f^1(\lambda, x) := \lambda f_1(x) + (1 - \lambda) f_2(x)\]

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

Assume that \(f(S) + \mathbb{R}_+^2\) is convex. A point \(\bar{x} \in S\) is an EP optimal point of Problem \(P \iff \exists \lambda\) such that \(\bar{x}\) is an image unique solution of Problem \(P_{\lambda}^1\).
Weighthed Chebyshev approximation: method (b)

\[(P^\infty_\lambda) : \quad f^\infty(\lambda) := \min_{x \in S} f^\infty(\lambda, x)\]

\[f^\infty(\lambda, x) := \max\{\lambda (f_1(x) - \mu_1), (1 - \lambda) (f_2(x) - \mu_2)\}\]

Theorem ([Jahn 10, Corollary 11.21 (a)], [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 an EP optimal point of Problem \(P \iff \exists \lambda \in (0, 1)\) such that \(\bar{x}\) is an image unique solution of Problem \(P^\infty_\lambda\).
Parametric sublevel set approximation: method (c)

Inspired by previous research on multiobjective linear optimization [1]

For each $\lambda \in [a_1, b_1]$, consider the following parametric POP

$$(P^u_{\lambda}) : \quad f^u(\lambda) := \min_{x \in S} \{ f_2(x) : f_1(x) \leq \lambda \} ,$$

with $a_1 := \min_{x \in S} f_1(x)$, $b_1 := f_1(\bar{x})$ and $\bar{x}$ a solution of $\min_{x \in S} f_2(x)$.

Lemma

Suppose that $\bar{x} \in S$ is an optimal solution of Problem $P^u_{\lambda}$, with $\lambda \in [a_1, b_1]$. Then $\bar{x}$ belongs to the set of weakly EP points of Problem $P$.

---

1B. Gorissen, D. den Hertog. *Approximating the pareto set of multiobjective linear programs via robust optimization.* (2012)
Questions

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

Yes!

We provide two approaches together with numerical schemes that avoid computing finitely many points.

1. Parametric POP: for methods (a) and (b) (resp. method (c)), we approximate the Pareto curve with polynomials so that convergence in $L_2$-norm (resp. $L_1$-norm) holds.

2. Hierarchy of outer approximation: we provide certified underestimators of the Pareto curve with strong convergence to $f(S)$ in $L_1$-norm.
Contributions

Yes!

**We provide two approaches together with numerical schemes that avoid computing finitely many points.**

1. Parametric POP: for methods (a) and (b) (resp. method (c)), we approximate the Pareto curve with polynomials so that convergence in $L_2$-norm (resp. $L_1$-norm) holds

2. Hierarchy of outer approximation: we provide certified underestimators of the Pareto curve with strong convergence to $f(S)$ in $L_1$-norm
Contributions

Yes!

We provide two approaches together with numerical schemes that avoid computing finitely many points.

1. Parametric POP: for methods (a) and (b) (resp. method (c)), we approximate the Pareto curve with polynomials so that convergence in $L_2$-norm (resp. $L_1$-norm) holds

2. Hierarchy of outer approximation: we provide certified underestimators of the Pareto curve with strong convergence to $f(S)$ in $L_1$-norm
Outline

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

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

**Assumption**

For almost all $\lambda \in [0, 1]$, the solution $x^*(\lambda)$ of the scalarized problem ($P^1_\lambda$) is unique.

Non-uniqueness may be tolerated on a Borel set $B \subset [0, 1]$, in which case one assumes image uniqueness of the solution.
Preliminaries: method (a)

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

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

\[
(P) \begin{cases} 
    \rho := \min_{\mu \in \mathcal{M}(K)} \int_{K} f(\lambda, x) d\mu(\lambda, x) \\
    \text{s.t. } \int_{K} \lambda^k d\mu(\lambda, x) = 1/(1 + k), \ k \in \mathbb{N}.
\end{cases}
\]
Preliminaries: method (a)

Lemma (Corollary of [2, Theorem 2.2])

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

\[
\rho = \int_K f(\lambda, x) d\mu^* = \int_0^1 f^*(\lambda) d\lambda.
\]

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

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

\[\text{2J.B. Lasserre. A “joint + marginal” approach to parametric polynomial optimization (2010)}\]
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 $z = (z_{k\alpha})$, $(k, \alpha) \in \mathbb{N}_d^{n+1}$.

- Consider the linear functional $L_z(g) := \sum_{k, \alpha} g_{k\alpha} z_{k\alpha}$. 
A hierarchy of semidefinite relaxations

- Let $g_0 := 1$.
- Let $d_0 := \max\{\deg f_1, \deg f_2, \deg g_1, \ldots, \deg g_m\}$.

Consider the semidefinite relaxations of $(P)$ for $d \geq d_0$:

$$
(P_d) \begin{cases}
\min_z L_z(f) \\
\text{s.t. } M_{d-v_l}(g_l z) \succeq 0, \ l = 0, \ldots, m, \\
L_z(\lambda^k) = 1/(1 + k), \ k = 0, \ldots, 2d
\end{cases}
$$

- $M_d(z)$ is the moment matrix associated with $z$
- $M_{d-v_l}(g_l z)$ is the localizing matrix associated with $z$ and $g_l$
Polynomial underestimators of $f^*(\lambda)$

The dual SDP of $(P_d)$ reads:

$$
(D_d) \begin{cases} 
\max_{q, (\sigma_l)} \sum_{k=0}^{2d} q_k / (1 + k) \\
\text{s.t. } f(\lambda, x) - q(\lambda) = \sum_{l=0}^{m} \sigma_l(\lambda, x) g_l(x) \\
q \in \mathbb{R}_{2d}[\lambda], \ \sigma_l \in \Sigma[\lambda, x], \ l = 0, \ldots, m, \\
\deg(\sigma_l g_l) \leq 2d, \ l = 0, \ldots, m.
\end{cases}
$$

- The hierarchy $(D_d)$ provides a sequence $(q_d)$ of polynomial underestimators of $f^*(\lambda)$.
- $\lim_{d \to \infty} \int_0^1 (f^*(\lambda) - q_d(\lambda)) d\lambda = 0$
Polynomial underestimators of $f^*(\lambda)$

On the convex example:

Degree 4 underestimator
Polynomial underestimators of $f^*(\lambda)$

On the convex example:

Degree 6 underestimator
An inverse problem from generalized moments

Lemma (Corollary of [2, Theorem 3.3])
Assume that for a.a. \( \lambda \in [0,1] \), Problem \((P)\) has a unique global optimizer \( x^*(\lambda) \) and let \( z^d = (z^d_{k\alpha}) \) be an optimal solution of \((P_d)\). Then,

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

In particular, for \( s \in \mathbb{N} \),

\[
m^k_j := \lim_{d \to \infty} \sum_{\alpha} f_{j\alpha} z^d_{k\alpha} = \int_0^1 \lambda^k f_j^*(\lambda) d\lambda, \quad j = 1, 2, \quad k = 0, \ldots, s.
\]

---

An inverse problem from generalized moments

For a fixed $s \in \mathbb{N}$, one can compute:

- Approximation $m_{j}^{sd}$ of the vector $m_{j}^{s} := (m_{j}^{k})$
- Approximations of $f_{j}^{*}(\lambda)$, $j = 1, 2$, by solving:

$$\min_{h \in \mathbb{R}_{s}[\lambda]} \left\{ \int_{0}^{1} (f_{j}^{*}(\lambda) - h(\lambda))^{2}d\lambda \right\}, \ j = 1, 2 .$$
An inverse problem from generalized moments

Theorem

The Problem
\[
\min_{h \in \mathbb{R}_s[\lambda]} \left\{ \int_0^1 (f_j^*(\lambda) - h(\lambda))^2 d\lambda \right\}
\]
has an optimal solution
\[
h_{sj} \in \mathbb{R}_s[\lambda],
\]
whose vector of coefficients is
\[
h_{sj} = H_s^{-1}m_j, \quad j = 1, 2,
\]
where
\[
H_s \in \mathcal{S}^{2s+1}
\]
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

\[
A = \int_0^1 f_j^*(\lambda)^2 d\lambda, \quad B = \int_0^1 f_j^*(\lambda) h(\lambda) d\lambda, \quad C = \int_0^1 h(\lambda)^2 d\lambda,
\]

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.

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

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

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

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

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

thus the problem can be reformulated as:

\[
\min_{h} \{ h' H_s h - 2 h' m_j \}, \ j = 1, 2.
\]
Weighted convex sum approximation: method (a)

On the convex example:
Weighted convex sum approximation: method (a)

On the convex example:

![Graph showing Pareto curves for a convex example with a degree of 6.]
Weighted convex sum approximation: method (a)

On the convex example:
Weighted Chebyshev approximation: method (b)

On the non-convex example:
Weighted Chebyshev approximation: method (b)

On the non-convex example:

Degree 6
Weighted Chebyshev approximation: method (b)

On the non-convex example:

Degree 8
Parametric sublevel set approximation: method (c)

Scaling the problem:

\[ K^u := \{ (\lambda, x) \in [0, 1] \times S : (f_1(x) - a_1) / (b_1 - a_1) \leq \lambda \} , \]

Parametric POP:

\[ (P^u_\lambda) : f^u(\lambda) = \min_{x \in S} \{ f_2(x) : (\lambda, x) \in K^u \} \]

Solving the dual SDP \( D_d \) yields underestimators for \( \lambda \mapsto f^u(\lambda) \) over \([a_1, b_1]\). One can directly approximate the Pareto curve from below!
Parametric sublevel set approximation: method (c)

On the non-convex example:
Parametric sublevel set approximation: method (c)

On the non-convex example:

Degree 6
Parametric sublevel set approximation: method (c)

On the non-convex example:

Degree 8
Parametric sublevel set approximation: method (c)

Medium size random bicriteria problem:

- $Q_1, Q_2 \in \mathbb{R}^{15 \times 15}, q_1, q_2 \in \mathbb{R}^{15}$
- $\min_{x \in [-1,1]^{15}} \{f_1(x), f_2(x)\}$
- $f_j(x) := x^\top Q_j x / n^2 - q_j^\top x / n$
Parametric sublevel set approximation: method (c)

Medium size random bicriteria problem:

- $Q_1, Q_2 \in \mathbb{R}^{15 \times 15}$, $q_1, q_2 \in \mathbb{R}^{15}$
- $\min_{x \in [-1,1]^{15}} \{ f_1(x), f_2(x) \}$
- $f_j(x) := x^\top Q_j x / n^2 - q_j^\top x / n$
Outline

1. Parametric POP
2. Outer Approximations of $f(S)$
3. Perspectives
Approximation of sets defined with "∃"

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

- Another point of view:

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

  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:

  $$ \Theta_d := \{ y \in \mathcal{B} : J_d(y) \leq 0 \} , $$

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

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

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

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

- Hierarchy of Semidefinite programs:

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

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

From the definition of \( J_d \), the sublevel sets

\[
\Theta_d := \{ y \in B : J_d(y) \leq 0 \} \supseteq f(S), \quad d \geq d_0
\]

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

It comes from the following:

\[
\forall (x, y) \in S \times B, J(y) \leq h(x, y) \iff \forall y, J(y) \leq H(y).
\]
Strong convergence property

**Theorem**

1. The sequence of underestimators \((J_d)_{d \geq d_0}\) converges to \(H\) w.r.t the \(L_1(B)\)-norm:
   \[
   \lim_{d \to \infty} \int_B |H - J_d|\,dy = 0 .
   \]

2. \[
   \lim_{d \to \infty} V(\Theta_d \setminus f(S)) = 0 .
   \]
Strong convergence property

Theorem

1. The sequence of underestimators \((J_d)_{d \geq d_0}\) converges to \(H\) w.r.t the \(L_1(B)\)-norm:
   \[
   \lim_{d \to \infty} \int_B |H - J_d| \, dy = 0.
   \]

2. \[
   \lim_{d \to \infty} V(\Theta_d \setminus f(S)) = 0.
   \]
Back to the non-convex example
Back to the non-convex example
Back to the non-convex example
Back to the non-convex example

\[ f(S) \]

\[ \Theta_5 \]
Back to the non-convex example
Branch and Bound: Zoom on the left

\[ f(S) \]

\[ \Theta_2 \]
Branch and Bound: Zoom on the left
Branch and Bound: Zoom on the left

\[ f(S) \]

\[ Θ_4 \]
Branch and Bound: Zoom on the left

\[ f(S) \]

\[ y_1 \]

\[ y_2 \]

\[ 7.4 \quad 7.6 \quad 7.8 \quad 8 \quad 8.2 \quad 8.4 \]

\[ 0 \quad 1 \quad 2 \quad 3 \quad 4 \]

\[ \Theta_5 \]
Branch and Bound: Zoom on the left

\[ f(S) \]
Outline

1. Parametric POP

2. Outer Approximations of $f(S)$

3. Perspectives
Transcendental conflicting criteria

Now, consider the following Problem:

$$(P) \left\{ \min_{x \in S} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}^T \right. \}$$

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 [2]

---

Transcendental conflicting criteria

**Definition: Semiconvex function**

Let $\gamma \geq 0$. A function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be $\gamma$-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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ which can be written as the max-plus linear combination $f = \sup_{w \in B} (a(w) + w)$ for some function $a : B \rightarrow \mathbb{R} \cup \{-\infty\}$ is precisely the set of lower semicontinuous $\gamma$-semiconvex functions.
Transcendental conflicting criteria
Transcendental conflicting criteria
Transcendental conflicting criteria

\[ y = \arctan(m a_1 a_2 a_3) \]

\[ \text{par}_{a_1} \]

\[ \text{par}_{a_2} \]

\[ \text{par}_{a_3} \]
Sublevel sets of semialgebraic underestimators

The sublevel sets

$$\Theta_d := \{ y \in B : J_d(y) \leq 0 \} \supseteq f(S), \ d \geq d_0,$$

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 \( F \in \mathbb{R}_{2d}[y], \sigma \in \Sigma_{d_0}[y]. \)
Thank you for your attention!


http://homepages.laas.fr/vmagron/