Approximating Pareto Curves using Semidefinite Relaxations

Victor Magron[1,2]   Didier Henrion[1,2,3]   Jean-Bernard Lasserre[1,2]

June 16, 2014

Abstract

We consider the problem of constructing an approximation of the Pareto curve associated with the multiobjective optimization problem $\min_{\mathbf{x} \in \mathbf{S}}\{(f_1(\mathbf{x}), f_2(\mathbf{x}))\}$, where f_1 and f_2 are two conflicting polynomial criteria and $\mathbf{S} \subset \mathbb{R}^n$ is a compact basic semialgebraic set. We provide a systematic numerical scheme to approximate the Pareto curve. We start by reducing the initial problem into a scalarized polynomial optimization problem (POP). Three scalarization methods lead to consider different parametric POPs, namely (a) a weighted convex sum approximation, (b) a weighted Chebyshev approximation, and (c) a parametric sublevel set approximation. For each case, we have to solve a semidefinite programming (SDP) hierarchy parametrized by the number of moments or equivalently the degree of a polynomial sums of squares approximation of the Pareto curve. When the degree of the polynomial approximation tends to infinity, we provide guarantees of convergence to the Pareto curve in L^2-norm for methods (a) and (b), and L^1-norm for method (c).

Keywords   Parametric Polynomial Optimization Problems, Semidefinite Programming, Multicriteria Optimization, Sums of Squares Relaxations, Pareto Curve, Inverse Problem from Generalized Moments

1   Introduction

Let $\mathbf{P}$ be the bicriteria polynomial optimization problem $\min_{\mathbf{x} \in \mathbf{S}}\{(f_1(\mathbf{x}), f_2(\mathbf{x}))\}$, where $\mathbf{S} \subset \mathbb{R}^n$ is the basic semialgebraic set:

$$\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geq 0, \ldots, g_m(\mathbf{x}) \geq 0\}, \tag{1}$$

for some polynomials $f_1, f_2, g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$. Here, we assume the following:

[1]CNRS; LAAS; 7 avenue du colonel Roche, F-31400 Toulouse; France.

[2]Université de Toulouse; LAAS, F-31400 Toulouse, France.

[3]Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic

Assumption 1.1. The image space $\mathbb{R}^2$ is partially ordered with the positive orthant $\mathbb{R}^2_+$. That is, given $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{y} \in \mathbb{R}^2$, it holds $\mathbf{x} \geq \mathbf{y}$ whenever $\mathbf{x} - \mathbf{y} \in \mathbb{R}^2_+$.

For the multiobjective optimization problem $\mathbf{P}$, one is usually interested in computing, or at least approximating, the following optimality set, defined e.g. in [6, Definition 11.5].

Definition 1.2. Let Assumption 1.1 be satisfied. A point $\bar{\mathbf{x}} \in \mathbf{S}$ is called an *Edgeworth-Pareto (EP) optimal point* of Problem $\mathbf{P}$, when there is no $\mathbf{x} \in \mathbf{S}$ such that $f_j(\mathbf{x}) \leq f_j(\bar{\mathbf{x}})$, $j = 1, 2$ and $f(\mathbf{x}) \neq f(\bar{\mathbf{x}})$. A point $\bar{\mathbf{x}} \in \mathbf{S}$ is called a *weakly Edgeworth-Pareto optimal point* of Problem $\mathbf{P}$, when there is no $\mathbf{x} \in \mathbf{S}$ such that $f_j(\mathbf{x}) < f_j(\bar{\mathbf{x}})$, $j = 1, 2$.

In this paper, for conciseness, we will also use the following terminology:

Definition 1.3. The image set of weakly Edgeworth-Pareto optimal points is called the *Pareto curve.*

Given a positive integer p and $\lambda \in [0, 1]$ both fixed, a common workaround consists in solving the *scalarized* problem:

$$f^p(\lambda) := \min_{\mathbf{x} \in \mathbf{S}} \{[(\lambda f_1(\mathbf{x}))^p + ((1 - \lambda) f_2(\mathbf{x}))^p]^{1/p}\} \,, \tag{2}$$

which includes the weighted sum approximation ($p = 1$)

$$\mathbf{P}^1_\lambda : \ f^1(\lambda) := \min_{\mathbf{x} \in \mathbf{S}} (\lambda f_1(\mathbf{x}) + (1 - \lambda) f_2(\mathbf{x})) \,, \tag{3}$$

and the weighted Chebyshev approximation ($p = \infty$)

$$\mathbf{P}^\infty_\lambda : \ f^\infty(\lambda) := \min_{\mathbf{x} \in \mathbf{S}} \max \{\lambda f_1(\mathbf{x}), (1 - \lambda) f_2(\mathbf{x})\} \,. \tag{4}$$

Here, we assume that for almost all (a.a.) $\lambda \in [0, 1]$, the solution $\mathbf{x}^*(\lambda)$ of the scalarized problem (3) (resp. (4)) 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. Then, by computing a solution $\mathbf{x}^*(\lambda)$, one can approximate the set $\{(f_1^*(\lambda), f_2^*(\lambda)) : \lambda \in [0, 1]\}$, where

$$f_j^*(\lambda) := f_j(\mathbf{x}^*(\lambda)), \ j = 1, 2.$$

Other approaches include using a numerical scheme such as the modified Polak method [11]: first, one considers a finite discretization $(y_1^{(k)})$ of the interval $[a_1, b_1]$, where

$$a_1 := \min_{\mathbf{x} \in \mathbf{S}} f_1(\mathbf{x}), \quad b_1 := f_1(\bar{\mathbf{x}}) \,, \tag{5}$$

with $\bar{\mathbf{x}}$ being a solution of $\min_{\mathbf{x} \in \mathbf{S}} f_2(\mathbf{x})$. Then, for each k, one computes an optimal solution $\mathbf{x}_k$ of the constrained optimization problem $y_2^{(k)} := \min_{\mathbf{x} \in \mathbf{S}} \{f_2(\mathbf{x}) : f_1(\mathbf{x}) = y_1^{(k)}\}$ and select the Pareto front from the finite collection $\{(y_1^{(k)}, y_2^{(k)})\}$. This method can be improved with the iterative Eichfelder-Polak algorithm, see e.g. [3]. Assuming the smoothness of the Pareto curve, one can use the Lagrange multiplier of the equality constraint to select the next point $y_1^{(k+1)}$. It allows to combine the adaptive control of discretization points with the modified Polak method. In [2], Das and Dennis introduce

the Normal-boundary intersection method which can find a uniform spread of points on the Pareto curve with more than two conflicting criteria and without assuming that the Pareto curve is either connected or smooth. However, there is no guarantee that the NBI method succeeds in general and even in case it works well, the spread of points is only uniform under certain additional assumptions. Interactive methods such as STEM [1] rely on a *decision maker* to select at each iteration the weight λ (most often in the case $p = \infty$) and to make a trade-off between criteria after solving the resulting scalar optimization problem.

So discretization methods suffer from two major drawbacks. (i) They only provide a *finite subset* of the Pareto curve and (ii) for each discretization point one has to compute a *global* minimizer of the resulting optimization problem (e.g. (3) or (4)). Notice that when f and $\mathbf{S}$ are both convex then point (ii) is not an issue.

In a recent work [4], Gorissen and den Hertog avoid discretization schemes for convex problems with multiple linear criteria $f_1, f_2, \ldots, f_k$ and a convex polytope $\mathbf{S}$. They provide an inner approximation of $f(\mathbf{S}) + \mathbb{R}^k_+$ by combining robust optimization techniques with semidefinite programming; for more details the reader is referred to [4].

Contribution. We provide a numerical scheme with two characteristic features: It avoids a discretization scheme and approximates the Pareto curve in a relatively strong sense. More precisely, the idea is consider multiobjective optimization as a particular instance of *parametric polynomial optimization* for which some strong approximation results are available when the data are polynomials and semi-algebraic sets. In fact we will investigate this approach:

method (a) for the first formulation (3) when $p = 1$, this is a *weighted convex sum approximation*;

method (b) for the second formuation (4) when $p = \infty$, this is a *weighted Chebyshev approximation*;

method (c) for a third formulation inspired by [4], this is a *parametric sublevel set approximation*.

When using some weighted combination of criteria ($p = 1$, method (a) or $p = \infty$, method (b)) we treat each function $\lambda \mapsto f_j(\lambda)$, $j = 1, 2$, as the signed density of the signed Borel measure $d\mu_j := f_j(\lambda)d\lambda$ with respect to the Lebesgue measure $d\lambda$ on $[0, 1]$. Then the procedure consists of two distinct steps:

1. In a first step, we solve a hierarchy of semidefinite programs (called SDP hierarchy) which permits to approximate any finite number $s + 1$ of moments $\mathbf{m}_j := (m_j^k)$, $k = 0, \ldots, s$ where :

$$m_j^k := \int_0^1 \lambda^k f_j^*(\lambda)d\lambda\,, \quad k = 0, \ldots, s\,, j = 1, 2\,.$$

More precisely, for any fixed integer s, step d of the SDP hierarchy provides an approximation $\mathbf{m}_j^d$ of $\mathbf{m}_j$ which converges to $\mathbf{m}_j$ as $d \to \infty$.

2. The second step consists of two *density estimation* problems: namely, for each $j = 1, 2$, and given the moments $\mathbf{m}_j$ of the measure $f_j^* d\lambda$ with unknown density f_j^* on $[0, 1]$, one computes a univariate polynomial $h_{s,j} \in \mathbb{R}_s[\lambda]$ which solves the optimization problem $\min_{h \in \mathbb{R}_s[\lambda]} \int_0^1 (f_j^*(\lambda) - h)^2 d\lambda$ if the moments $\mathbf{m}_j$ are known exactly. The corresponding vector of coefficients $\mathbf{h}_j^s \in \mathbb{R}^{s+1}$ is given by $\mathbf{h}_j^s = \mathbf{H}_s(\lambda)^{-1} \mathbf{m}_j$, $j = 1, 2$, where $\mathbf{H}_s(\lambda)$ is the s-moment matrix of the Lebesgue measure $d\lambda$ on $[0, 1]$; therefore in the expression for $\mathbf{h}_j^s$ we replace $\mathbf{m}_j$ with its approximation.

Hence for both methods (a) and (b), we have L^2-*norm convergence guarantees.*

Alternatively, in our method (c), one can estimate the Pareto curve by solving for each $\lambda \in [a_1, b_1]$ the following parametric POP:

$$\mathbf{P}_\lambda^u: \quad f^u(\lambda) := \min_{\mathbf{x} \in \mathbf{S}} \left\{ f_2(\mathbf{x}) : f_1(\mathbf{x}) \leq \lambda \right\}, \tag{6}$$

with a_1 and b_1 as in (5). Notice that by definition $f^u(\lambda) = f_2^*(\lambda)$. Then, we derive an SDP hierarchy parametrized by d, so that the optimal solution $q_{2d} \in \mathbb{R}[\lambda]_{2d}$ of the d-th relaxation underestimates f_2^* over $[a_1, b_1]$. In addition, q_{2d} converges to f_2^* with respect to the L_1-norm, as $d \to \infty$. In this way, one can approximate from below the set of Pareto points, as closely as desired. Hence for method (c), we have L^1-*norm convergence guarantees.*

It is important to observe that even though $\mathbf{P}_\lambda^1$, $\mathbf{P}_\lambda^\infty$ and $\mathbf{P}_\lambda^u$ are all global optimization problems we do *not* need to solve them exactly. In all cases the information provided at step d of the SDP hierarchy (i.e. $\mathbf{m}_j^d$ for $\mathbf{P}_\lambda^1$ and $\mathbf{P}_\lambda^\infty$ and the polynomial q_{2d} for $\mathbf{P}_\lambda^u$) permits to define an approximation of the Pareto front. In other words even in the absence of convexity the SDP hierarchy allows to approximate the Pareto front and of course the higher in the hierarchy the better is the approximation.

The paper is organized as follows. Section 2 is dedicated to recalling some background about moment and localizing matrices. Section 3 describes our framework to approximate the set of Pareto points using SDP relaxations of parametric optimization programs. These programs are presented in Section 3.1 while we describe how to reconstruct the Pareto curve in Section 3.2. Section 4 presents some numerical experiments which illustrate the different approximation schemes.

2 Preliminaries

Let $\mathbb{R}[\lambda, \mathbf{x}]$ (resp. $\mathbb{R}[\lambda, \mathbf{x}]_{2d}$) denote the ring of real polynomials (resp. of degree at most $2d$) in the variables λ and $\mathbf{x} = (x_1, \ldots, x_n)$, whereas $\Sigma[\lambda, \mathbf{x}]$ (resp. $\Sigma[\lambda, \mathbf{x}]_d$) denotes its subset of sums of squares (SOS) of polynomials (resp. of degree at most $2d$). For every $\alpha \in \mathbb{N}^n$ the notation $\mathbf{x}^\alpha$ stands for the monomial $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ and for every $d \in \mathbb{N}$, let $\mathbb{N}_d^{n+1} := \{\beta \in \mathbb{N}^{n+1} : \sum_{j=1}^{n+1} \beta_j \leq d\}$, whose cardinal is $s_n(d) = \binom{n+1+d}{d}$. A polynomial $f \in \mathbb{R}[\lambda, \mathbf{x}]$ is written

$$(\lambda, \mathbf{x}) \mapsto f(\lambda, \mathbf{x}) = \sum_{(k,\alpha) \in \mathbb{N}^{n+1}} f_{k\alpha} \lambda^k \mathbf{x}^\alpha,$$

and f can be identified with its vector of coefficients $\mathbf{f} = (f_{k\alpha})$ in the canonical basis $(\mathbf{x}^\alpha)$, $\alpha \in \mathbb{N}^n$. For any symmetric matrix $\mathbf{A}$ the notation $\mathbf{A} \succeq 0$ stands for $\mathbf{A}$ being semidefinite positive. A real sequence $\mathbf{z} = (z_{k\alpha})$, $(k, \alpha) \in \mathbb{N}^{n+1}$, has a *representing measure* if there exists some finite Borel measure μ on $\mathbb{R}^{n+1}$ such that

$$z_{k\alpha} = \int_{\mathbb{R}^{n+1}} \lambda^k \mathbf{x}^\alpha \, d\mu(\lambda, \mathbf{x}), \quad \forall (k, \alpha) \in \mathbb{N}^{n+1} .$$

Given a real sequence $\mathbf{z} = (z_{k\alpha})$ define the linear functional $L_\mathbf{z} : \mathbb{R}[\lambda, \mathbf{x}] \to \mathbb{R}$ by:

$$f \left(= \sum_{(k,\alpha)} f_{k\alpha} \lambda^k \mathbf{x}^\alpha \right) \quad \mapsto L_\mathbf{z}(f) = \sum_{(k,\alpha)} f_{k\alpha} z_{k\alpha}, \quad f \in \mathbb{R}[\lambda, \mathbf{x}] .$$

Moment matrix The *moment* matrix associated with a sequence $\mathbf{z} = (z_{k\alpha})$, $(k, \alpha) \in \mathbb{N}^{n+1}$, is the real symmetric matrix $\mathbf{M}_d(\mathbf{z})$ with rows and columns indexed by $\mathbb{N}_d^{n+1}$, and whose entry $(i, \alpha), (j, \beta)$ is just $z_{(i+j)(\alpha+\beta)}$, for every $(i, \alpha), (j, \beta) \in \mathbb{N}_d^{n+1}$.

If $\mathbf{z}$ has a representing measure μ then $\mathbf{M}_d(\mathbf{z}) \succeq 0$ because

$$\langle \mathbf{f}, \mathbf{M}_d(\mathbf{z}) \mathbf{f} \rangle = \int f^2 \, d\mu \geq 0, \quad \forall \mathbf{f} \in \mathbb{R}^{s_n(d)}.$$

Localizing matrix With $\mathbf{z}$ as above and $g \in \mathbb{R}[\lambda, \mathbf{x}]$ (with $g(\lambda, \mathbf{x}) = \sum_{\ell, \gamma} g_{\ell\gamma} \lambda^\ell \mathbf{x}^\gamma$), the *localizing* matrix associated with $\mathbf{z}$ and g is the real symmetric matrix $\mathbf{M}_d(g\,\mathbf{z})$ with rows and columns indexed by $\mathbb{N}_d^n$, and whose entry $((i, \alpha), (j, \beta))$ is just $\sum_{\ell, \gamma} g_{\ell\gamma} z_{(i+j+\ell)(\alpha+\beta+\gamma)}$, for every $(i, \alpha), (j, \beta) \in \mathbb{N}_d^{n+1}$.

If $\mathbf{z}$ has a representing measure μ whose support is contained in the set $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ then $\mathbf{M}_d(g\,\mathbf{z}) \succeq 0$ because

$$\langle \mathbf{f}, \mathbf{M}_d(g\,\mathbf{z}) \mathbf{f} \rangle = \int f^2 \, g \, d\mu \geq 0, \quad \forall \mathbf{f} \in \mathbb{R}^{s_n(d)} .$$

In the sequel, we assume that $\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geq 0, \ldots, g_m(\mathbf{x}) \geq 0\}$ is contained in a box. It ensures that there is some integer $M > 0$ such that the quadratic polynomial $g_{m+1}(\mathbf{x}) := M - \sum_{i=1}^n x_i^2$ is nonnegative over $\mathbf{S}$. Then, we add the redundant polynomial constraint $g_{m+1}(\mathbf{x}) \geq 0$ to the definition of $\mathbf{S}$.

3 Approximating the Pareto Curve

3.1 Reduction to Scalar Parametric POP

Here, we show that computing the set of Pareto points associated with Problem $\mathbf{P}$ can be achieved with three different parametric polynomial problems. Recall that the feasible set of Problem $\mathbf{P}$ is $\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geq 0, \ldots, g_{m+1}(\mathbf{x}) \geq 0\}$.

Method (a): convex sum approximation Consider the scalar objective function $f(\lambda, \mathbf{x}) := \lambda f_1(\mathbf{x}) + (1 - \lambda) f_2(\mathbf{x})$, $\lambda \in [0, 1]$. Let $\mathbf{K}^1 := [0, 1] \times \mathbf{S}$. Recall from (3) that function $f^1 : [0, 1] \to \mathbb{R}$ is the optimal value of Problem $\mathbf{P}_\lambda^1$, i.e. $f^1(\lambda) = \min_{\mathbf{x}}\{f(\lambda, \mathbf{x}) : (\lambda, \mathbf{x}) \in \mathbf{K}^1\}$. If the set $f(\mathbf{S}) + \mathbb{R}_+^2$ is convex, then one can recover the Pareto curve by computing $f^1(\lambda)$, for all $\lambda \in [0, 1]$, see [6, Table 11.5].

Lemma 3.1. *Assume that $f(\mathbf{S}) + \mathbb{R}_+^2$ is convex. Then, a point $\overline{\mathbf{x}} \in \mathbf{S}$ belongs to the set of EP points of Problem $\mathbf{P}$ if and only if there exists some weight $\lambda \in [0, 1]$ such that $\overline{\mathbf{x}}$ is an image unique optimal solution of Problem $\mathbf{P}_\lambda^1$.*

Method (b): weighted Chebyshev approximation Reformulating Problem $\mathbf{P}$ using the Chebyshev norm approach is more suitable when the set $f(\mathbf{S}) + \mathbb{R}_+^2$ is not convex. We optimize the scalar criterion $f(\lambda, \mathbf{x}) := \max[\lambda f_1(\mathbf{x}), (1 - \lambda) f_2(\mathbf{x})]$, $\lambda \in [0, 1]$. In this case, we assume without loss of generality that both f_1 and f_2 are positive. Indeed, for each $j = 1, 2$, one can always consider the criterion $\tilde{f}_j := f_j - a_j$, where a_j is any lower bound of the global minimum of f_j over $\mathbf{S}$. Such bounds can be computed efficiently by solving polynomial optimization problems using an SDP hierarchy, see e.g. [8]. In practice, we introduce a lifting variable ω to represent the max of the objective function. For scaling purpose, we introduce the constant $C := \max(M_1, M_2)$, with $M_j := \max_{\mathbf{x} \in \mathbf{S}} f_j$, $j = 1, 2$. Then, one defines the constraint set $\mathbf{K}^\infty := \{(\lambda, \mathbf{x}, \omega) \in \mathbb{R}^{n+2} : \mathbf{x} \in \mathbf{S}, \lambda \in [0, 1], \lambda f_1(\mathbf{x})/C \leq \omega, (1 - \lambda) f_2(\mathbf{x})/C \leq \omega\}$, which leads to the reformulation of $\mathbf{P}_\lambda^\infty : f^\infty(\lambda) = \min_{\mathbf{x}, \omega}\{\omega : (\lambda, \mathbf{x}, \omega) \in \mathbf{K}^\infty\}$ consistent with (4). The following lemma is a consequence of [6, Corollary 11.21 (a)].

Lemma 3.2. *Suppose that f_1 and f_2 are both positive. Then, a point $\overline{\mathbf{x}} \in \mathbf{S}$ belongs to the set of EP points of Problem $\mathbf{P}$ if and only if there exists some weight $\lambda \in (0, 1)$ such that $\overline{\mathbf{x}}$ is an image unique optimal solution of Problem $\mathbf{P}_\lambda^\infty$.*

Method (c): parametric sublevel set approximation Here, we use an alternative method inspired by [4]. Problem $\mathbf{P}$ can be approximated using the criterion f_2 as the objective function and the constraint set

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

which leads to the parametric POP $\mathbf{P}_\lambda^u : f^u(\lambda) = \min_{\mathbf{x}}\{f_2(\mathbf{x}) : (\lambda, \mathbf{x}) \in \mathbf{K}^u\}$ which is consistent with (6), and such that $f^u(\lambda) = f_2^*(\lambda)$ for all $\lambda \in [0, 1]$, with a_1 and b_1 as in (5).

Lemma 3.3. *Suppose that $\overline{\mathbf{x}} \in \mathbf{S}$ is an optimal solution of Problem $\mathbf{P}_\lambda^u$, with $\lambda \in [0, 1]$. Then $\overline{\mathbf{x}}$ belongs to the set of weakly EP points of Problem $\mathbf{P}$.*

Proof. Suppose that there exists $\mathbf{x} \in \mathbf{S}$ such that $f_1(\mathbf{x}) < f_1(\overline{\mathbf{x}})$ and $f_2(\mathbf{x}) < f_2(\overline{\mathbf{x}})$. Then $\mathbf{x}$ is feasible for Problem $\mathbf{P}_\lambda^u$ (since $(f_1(\mathbf{x}) - a_1)/(b_1 - a_1) \leq \lambda$) and $f_2(\overline{\mathbf{x}}) \leq f_2(\mathbf{x})$, which leads to a contradiction. □

Note that if a solution $\mathbf{x}^*(\lambda)$ is unique then it is EP optimal. Moreover, if a solution $\mathbf{x}^*(\lambda)$ of Problem $\mathbf{P}^u(\lambda)$ solves also the optimization problem $\min_{\mathbf{x} \in \mathbf{S}}\{f_1(\mathbf{x}) : f_2(\mathbf{x}) \leq \lambda\}$, then it is an EP optimal point (see [10] for more details).

3.2 A Hierarchy of Semidefinite Relaxations

Notice that the three problems $\mathbf{P}_\lambda^1$, $\mathbf{P}_\lambda^\infty$ and $\mathbf{P}_\lambda^u$ are particular instances of the generic parametric optimization problem $f^*(y) := \min_{(y,\mathbf{x})\in\mathbf{K}} f(y,\mathbf{x})$. The feasible set $\mathbf{K}$ (resp. the objective function f^*) corresponds to $\mathbf{K}^1$ (resp. f^1) for Problem $\mathbf{P}_\lambda^1$, $\mathbf{K}^\infty$ (resp. f^∞) for Problem $\mathbf{P}_\lambda^\infty$ and $\mathbf{K}^u$ (resp. f^u) for Problem $\mathbf{P}_\lambda^u$. We write $\mathbf{K} := \{(y,\mathbf{x}) \in \mathbb{R}^{n'+1} : p_1(y,\mathbf{x}) \geq 0, \ldots, p_{m'}(y,\mathbf{x}) \geq 0\}$. Note also that $n' = n$ (resp. $n' = n+1$) when considering Problem $\mathbf{P}_\lambda^1$ and Problem $\mathbf{P}_\lambda^u$ (resp. Problem $\mathbf{P}_\lambda^\infty$).

Let $\mathcal{M}(\mathbf{K})$ be the space of probability measures supported on $\mathbf{K}$. The function f^* is well-defined because f is a polynomial and $\mathbf{K}$ is compact. Let $\mathbf{a} = (a_k)_{k\in\mathbb{N}}$, with $a_k = 1/(k+1)$, $\forall k \in \mathbb{N}$ and consider the optimization problem:

$$\mathcal{P}: \quad \begin{cases} \rho := \displaystyle\min_{\mu\in\mathcal{M}(\mathbf{K})} \ \int_\mathbf{K} f(y,\mathbf{x})\,d\mu(y,\mathbf{x}) \\ \quad\text{s.t.} \ \displaystyle\int_\mathbf{K} y^k d\mu(y,\mathbf{x}) = a_k, \ k\in\mathbb{N}. \end{cases} \tag{7}$$

Lemma 3.4. *The optimization problem $\mathcal{P}$ has an optimal solution $\mu^* \in \mathcal{M}(\mathbf{K})$ and if ρ is as in (7) then*

$$\rho = \int_\mathbf{K} f(y,\mathbf{x})\,d\mu^* = \int_0^1 f^*(y)\,dy. \tag{8}$$

Suppose that for almost all (a.a.) $y \in [0,1]$, the parametric optimization problem $f^*(y) = \min_{(y,\mathbf{x})\in\mathbf{K}} f(y,\mathbf{x})$ has a unique global minimizer $\mathbf{x}^*(y)$ and let $f_j^* : [0,1] \to \mathbb{R}$ be the function $y \mapsto f_j^*(y) := f_j(\mathbf{x}^*(y))$, $j = 1,2$. Then for Problem $\mathbf{P}_\lambda^1$, $\rho = \int_0^1 \lambda f_1^*(\lambda) + (1-\lambda)f_2^*(\lambda)\,d\lambda$, for Problem $\mathbf{P}_\lambda^\infty$, $\rho = \int_0^1 \max\{\lambda f_1^*(\lambda), (1-\lambda)f_2^*(\lambda)\}\,d\lambda$ and for Problem $\mathbf{P}_\lambda^u$, $\rho = \int_0^1 f_2^*(\lambda)\,d\lambda$.

Proof. The proof of (8) follows from [9, Theorem 2.2] with y in lieu of $\mathbf{y}$. Now, consider the particular case of Problem $\mathbf{P}_\lambda^1$. If $\mathbf{P}_\lambda^1$ has a unique optimal solution $\mathbf{x}^*(\lambda) \in \mathbf{S}$ for a.a. $\lambda \in [0,1]$ then $f^*(\lambda) = \lambda f_1^*(\lambda) + (1-\lambda)f_2^*(\lambda)$ for a.a. $\lambda \in [0,1]$. The proofs for $\mathbf{P}_\lambda^\infty$ and $\mathbf{P}_\lambda^u$ are similar. $\square$

We set $p_0 := 1$, $v_l := \lceil \deg p_l/2 \rceil$, $l = 0, \ldots, m'$ and $d_0 := \max(\lceil d_1/2\rceil, \lceil d_2/2\rceil, v_1, \ldots, v_{m'})$. Then, consider the following semidefinite relaxations for $d \geq d_0$:

$$\begin{cases} \min_\mathbf{z} \ L_\mathbf{z}(f) \\ \quad\text{s.t.} \ \mathbf{M}_d(\mathbf{z}) \succeq 0, \\ \qquad \mathbf{M}_{d-v_l}(p_l\,\mathbf{z}) \succeq 0, \ l = 1, \ldots, m', \\ \qquad L_\mathbf{z}(y^k) = a_k, \quad k = 0, \ldots, 2d. \end{cases} \tag{9}$$

Lemma 3.5. *Assume that for a.a. $y \in [0,1]$, the parametric optimization problem $f^*(y) = \min_{(y,\mathbf{x})\in\mathbf{K}} f(y,\mathbf{x})$ has a unique global minimizer $\mathbf{x}^*(y)$, and let $\mathbf{z}^d = (z_{k\alpha}^d)$, $(k,\alpha) \in \mathbb{N}_{2d}^{n+1}$, be an optimal solution of (9). Then*

$$\lim_{d\to\infty} z_{k\alpha}^d = \int_0^1 y^k\,(\mathbf{x}^*(y))^\alpha dy. \tag{10}$$

In particular, for $s \in \mathbb{N}$, for all $k = 0, \ldots, s$, $j = 1,2$,

$$m_j^k := \lim_{d\to\infty} \sum_\alpha f_{j\alpha} z_{k\alpha}^d = \int_0^1 y^k\,f_j^*(y)\,dy. \tag{11}$$

Proof. Let $\mu^* \in \mathcal{M}(\mathbf{K})$ be an optimal solution of problem $\mathcal{P}$. From [9, Theorem 3.3],

$$\lim_{d\to\infty} z_{k\alpha}^d = \int_{\mathbf{K}} y^k \mathbf{x}^\alpha \, d\mu^*(y, \mathbf{x}) = \int_0^1 y^k \, (\mathbf{x}^*(y))^\alpha dy \,,$$

which is (10). Next, from (10), one has for $s \in \mathbb{N}$:

$$\lim_{d\to\infty} \sum_\alpha f_{j\alpha} z_{k\alpha}^d = \int_0^1 y^k \, f_j(\mathbf{x}^*(y)) \, dy = \int_0^1 y^k \, f_j^*(y) \, dy \,,$$

for all $k = 0, \ldots, s$, $j = 1, 2$. Thus (11) holds. $\qquad\square$

The dual of the SDP (9) reads:

$$
\begin{cases}
\rho_d^* := & \displaystyle\max_{q,(\sigma_l)} \int_0^1 q(y) \, dy \; \left(= \sum_{k=0}^{2d} q_k \, a_k\right) \\
\text{s.t.} & f(y, \mathbf{x}) - q(y) = \sum_{k=0}^{m'} \sigma_l(y, \mathbf{x}) \, p_l(y, \mathbf{x}) \,, \forall y \,, \forall \mathbf{x} \,, \\
& q \in \mathbb{R}[y]_{2d}, \sigma_l \in \Sigma[y, \mathbf{x}]_{d - v_l}, \; l = 0, \ldots, m' \,.
\end{cases}
\tag{12}
$$

Lemma 3.6. *Consider the dual semidefinite relaxations defined in* (12). *Then, one has:*

(i) $\rho_d \uparrow \rho$ *as* $d \to \infty$.
(ii) *Let* q_{2d} *be a nearly optimal solution of* (12), *i.e., such that* $\int_0^1 q_{2d}(y) dy \geq \rho_d^* - 1/d$. *Then* q_{2d} *underestimates* f^* *over* $\mathbf{S}$ *and* $\lim_{d\to\infty} \int_0^1 |f^*(y) - q_{2d}(y)| d\mu = 0$.

Proof. It follows from [9, Theorem 3.5]. $\qquad\square$

Note that one can directly approximate the Pareto curve from below when considering Problem $\mathbf{P}_\lambda^u$. Indeed, solving the dual SDP (12) yields polynomials that underestimate the function $\lambda \mapsto f_2^*(\lambda)$ over $[0, 1]$.

Remark. In [4, Appendix A], the authors derive the following relaxation from Problem $\mathbf{P}_\lambda^u$:

$$
\begin{cases}
\displaystyle\max_{q\in\mathbb{R}[y]_d} & \displaystyle\int_0^1 q(\lambda) \, d\lambda \,, \\
\text{s.t.} & f_2(\mathbf{x}) \geq q\left(\frac{f_1(\mathbf{x}) - a_1}{b_1 - a_1}\right) \,, \forall \mathbf{x} \in \mathbf{S} \,.
\end{cases}
\tag{13}
$$

Since one wishes to approximate the Pareto curve, suppose that in (13) one also imposes that q is nonincreasing over $[0, 1]$. For even degree approximations, the formulation (13) is equivalent to

$$
\begin{cases}
\displaystyle\max_{q\in\mathbb{R}[y]_{2d}} & \displaystyle\int_0^1 q(\lambda) \, d\lambda \,, \\
\text{s.t.} & f_2(\lambda) \geq q(\lambda) \,, \forall \lambda \in [0, 1] \,, \\
& \frac{f_1(\mathbf{x}) - a_1}{b_1 - a_1} \leq \lambda \,, \forall \lambda \in [0, 1] \,, \forall \mathbf{x} \in \mathbf{S} \,.
\end{cases}
\tag{14}
$$

Thus, our framework is related to [4] by observing that (12) is a strengthening of (14).

When using the reformulations $\mathbf{P}_\lambda^1$ and $\mathbf{P}_\lambda^\infty$, computing the Pareto curve is computing (or at least providing good approximations) of the functions $f_j^* : [0, 1] \to \mathbb{R}$ defined above, and we consider this problem as an *inverse problem from generalized moments*.

- For any fixed $s \in \mathbb{N}$, we first compute approximations $\mathbf{m}_j^{sd} = (m_j^{kd})$, $k = 0, \ldots, s$, $d \in \mathbb{N}$, of the generalized moments $m_j^k = \int_0^1 \lambda^k f_j^*(\lambda) \, d\lambda$, $k = 0, \ldots, s$, $j = 1, 2$, with the convergence property $(\mathbf{m}_j^{sd}) \to \mathbf{m}_j^s$ as $d \to \infty$, for each $j = 1, 2$.

- Then we solve the inverse problem: given a (good) approximation $(\mathbf{m}_j^{sd})$ of $\mathbf{m}_j^s$, find a polynomial $h_{s,j}$ of degree at most s such that $m_j^{kd} = \int_0^1 \lambda^k h_{s,j}(\lambda) \, d\lambda$, $k = 0, \ldots, s$, $j = 1, 2$. Importantly, if $(\mathbf{m}_j^{sd}) = (\mathbf{m}_j^s)$ then $h_{s,j}$ minimizes the L_2-norm $\int_0^1 (h(\lambda) - f_j^*(\lambda))^2 d\lambda$ (see A for more details).

Computational considerations The presented parametric optimization methodology has a high computational cost mainly due to the size of SDP relaxations (9) and the state-of-the-art for SDP solvers. Indeed, when the relaxation order d is fixed, the size of the SDP matrices involved in (9) grows like $O((n+1)^d)$ for Problem $\mathbf{P}_\lambda^1$ and like $O((n+2)^d)$ for problems $\mathbf{P}_\lambda^\infty$ and $\mathbf{P}_\lambda^u$. By comparison, when using a discretization scheme, one has to solve N polynomial optimization problems, each one being solved by programs whose SDP matrix size grows like $O(n^d)$. Section 4 compares both methods.

Therefore these techniques are of course limited to problems of modest size involving a small or medium number of variables n. We have been able to handle non convex problems with about 15 variables. However when a correlative sparsity pattern is present then one may benefit from a sparse variant of the SDP relaxations for parametric POP which permits to handle problems of much larger size (e.g. with more than 500 variables); see e.g. [12, 7] for more details.

4   Numerical Experiments

The semidefinite relaxations of problems $\mathbf{P}_\lambda^1$, $\mathbf{P}_\lambda^\infty$ and $\mathbf{P}_\lambda^u$ have been implemented in MATLAB, using the Gloptipoly software package [5], on an Intel Core i5 CPU (2.40 GHz).

4.1   Case 1: $f(\mathbf{S}) + \mathbb{R}_+^2$ is convex

We have considered the following test problem mentioned in [6, Example 11.8]:

Example 1. Let

$$g_1 := -x_1^2 + x_2 , \qquad\qquad\qquad f_1 := -x_1 ,$$
$$g_2 := -x_1 - 2x_2 + 3 , \qquad\qquad f_2 := x_1 + x_2^2 .$$
$$\mathbf{S} := \{\mathbf{x} \in \mathbb{R}^2 : g_1(\mathbf{x}) \geq 0, g_2(\mathbf{x}) \geq 0\} .$$

Figure 1 displays the discretization of the feasible set $\mathbf{S}$ as well as the image set $f(\mathbf{S})$. The weighted sum approximation method (a) being suitable when the set $f(\mathbf{S}) + \mathbb{R}_+^2$ is convex, one reformulates the problem as a particular instance of Problem $\mathbf{P}_\lambda^1$.

For comparison, we fix discretization points $\lambda_1, \ldots, \lambda_N$ uniformly distributed on the interval $[0, 1]$ (in our experiments, we set $N = 100$). Then for each $\lambda_i, i = 1, \ldots, N$, we compute
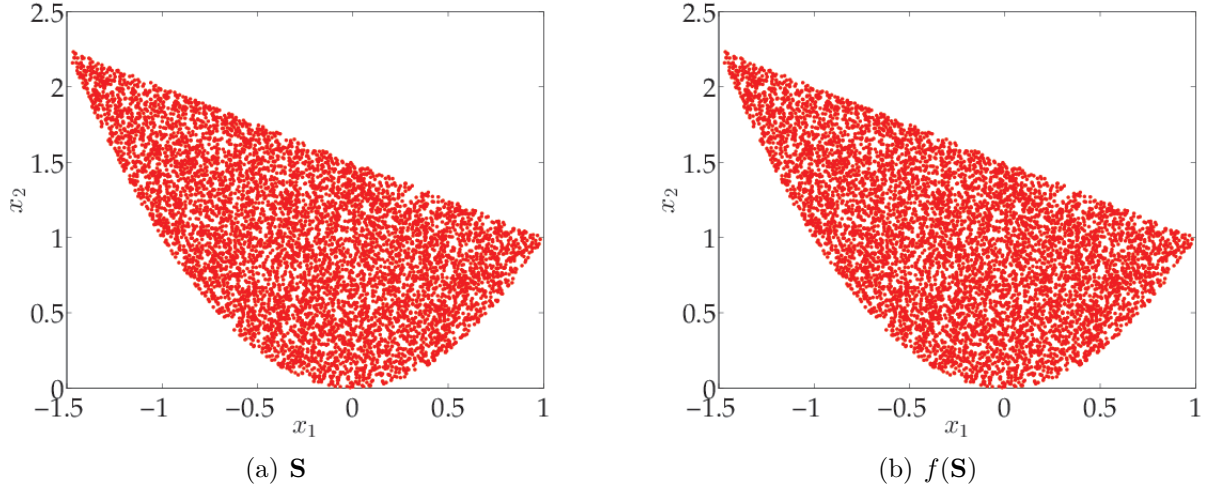
(a) **S**   (b) $f(\mathbf{S})$

Figure 1: Preimage and image set of f for Example 1

the optimal value $f^*(\lambda_i)$ of the polynomial optimization problem $\mathbf{P}^1_{\lambda_i}$. The dotted curves from Figure 2 display the results of this discretization scheme. From the optimal solution of the dual SDP (12) corresponding to our method (a), namely weighted convex sum approximation, one obtains the degree 4 polynomial q_4 (resp. degree 6 polynomial q_6) with moments up to order 8 (resp. 12), displayed on Figure 2 (a) (resp. (b)). One observes that $q_4 \leq f^*$ and $q_6 \leq f^*$, which illustrates Lemma 3.6 (ii). The higher relaxation order also provides a tighter underestimator, as expected.

(a) Degree 4 underestimator   (b) Degree 6 underestimator
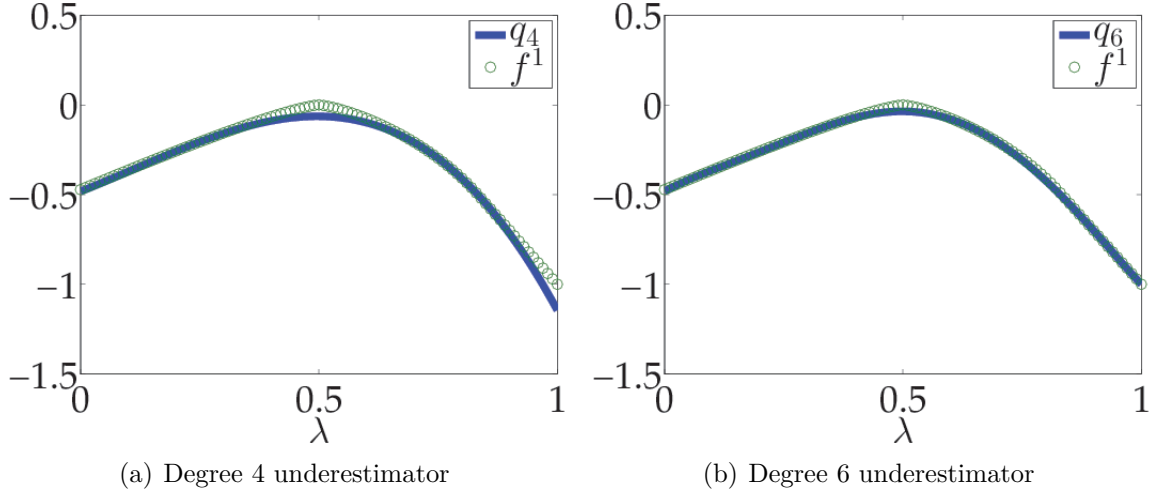
Figure 2: A hierarchy of polynomial underestimators of the Pareto curve for Example 1 obtained by weighted convex sum approximation (method (a))

Then, for each $\lambda_i, i = 1, \ldots, 100$, we compute an optimal solution $\mathbf{x}^*(\lambda_i)$ of Problem $\mathbf{P}^1_{\lambda_i}$ and we set $f^*_{1i} := f_1(\mathbf{x}^*(\lambda_i))$, $f^*_{2i} := f_2(\mathbf{x}^*(\lambda_i))$. Hence, we obtain a discretization (f^*_1, f^*_2) of the Pareto curve, represented by the dotted curve on Figure 3. The required CPU running time for the corresponding SDP relaxations is 26sec.

10

We compute an optimal solution of the primal SDP (9) at order $d = 5$, in order to provide a good approximation of $s + 1$ moments with $s = 4, 6, 8$. Then, we approximate each function f_j^*, $j = 1, 2$ with a polynomial h_{sj} of degree s by solving the inverse problem from generalized moments (see Appendix A). The resulting Pareto curve approximation using degree 4 estimators h_{41} and h_{42} is displayed on Figure 3 (a). For comparison purpose, higher degree approximations are also represented on Figure 3 (b) (degree 6 polynomials) and Figure 3 (c) (degree 8 polynomials). It consumes only 0.4sec to compute the two degree 4 polynomials h_{41} and h_{42}, 0.5sec for the degree 6 polynomials and 1.4sec for the degree 8 polynomials.

(a) Degree 4 estimators
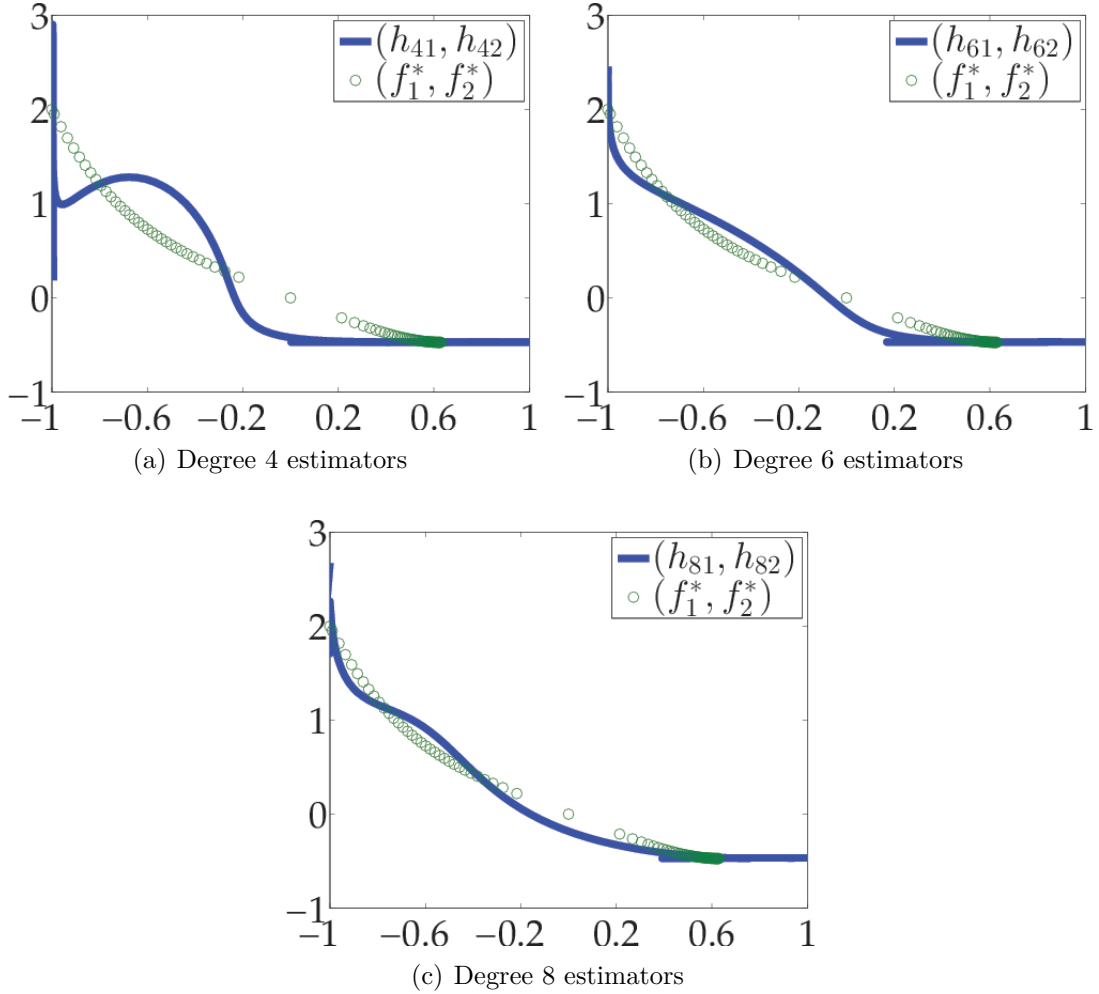
(b) Degree 6 estimators

(c) Degree 8 estimators

Figure 3: A hierarchy of polynomial approximations of the Pareto curve for Example 1 obtained by the weighted convex sum approximation (method (a))

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

We have also solved the following two-dimensional nonlinear problem proposed in [13]:

Example 2. Let

$$g_1 := -(x_1 - 2)^3/2 - x_2 + 2.5 \, , \qquad\qquad f_1 := \frac{(x_1 + x_2 - 7.5)^2}{4} + (x_2 - x_1 + 3)^2 \, ,$$
$$g_2 := -x_1 - x_2 + 8(x_2 - x_1 + 0.65)^2 + 3.85 \, , \qquad f_2 := 0.4(x_1 - 1)^2 + 0.4(x_2 - 4)^2 \, .$$
$$\mathbf{S} := \{\mathbf{x} \in [0,5] \times [0,3] : g_1(\mathbf{x}) \geq 0, g_2(\mathbf{x}) \geq 0\} \, .$$

Figure 4 depicts the discretization of the feasible set $\mathbf{S}$ as well as the image set $f(\mathbf{S})$ for this problem. Note that the Pareto curve is non-connected and non-convex.

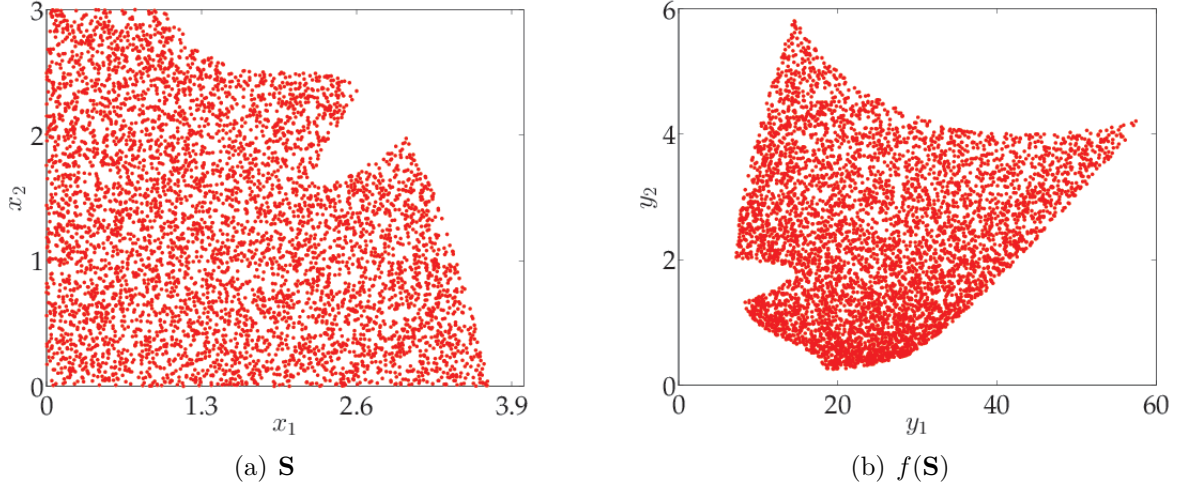(a) $\mathbf{S}$ (b) $f(\mathbf{S})$

Figure 4: Preimage and image set of f for Example 2

In this case, the weighted convex sum approximation of method (a) would not allow to properly reconstruct the Pareto curve, due to the apparent nonconvex geometry of the set $f(\mathbf{S}) + \mathbb{R}_+^2$. Hence we have considered methods (b) and (c).

Method (b): weighted Chebyshev approximation As for Example 1, one solves the SDP (9) at order $d = 5$ and approximate each function f_j^*, $j = 1, 2$ using polynomials of degree 4, 6 and 8. The approximation results are displayed on Figure 5. Degree 8 polynomials give a closer approximation of the Pareto curve than degree 4 or 6 polynomials. The solution time range is similar to the benchmarks of Example 1. The SDP running time for the discretization is about 3min. The degree 4 polynomials are obtained after 1.3sec, the degree 6 polynomials h_{61}, h_{62} after 9.7sec and the degree 8 polynomials after 1min.

Method (c): parametric sublevel set approximation Better approximations can be directly obtained by reformulating Example 2 as an instance of Problem $\mathbf{P}_\lambda^u$ and compute the degree d optimal solutions q_{2d} of the dual SDP (12). Figure 6 reveals that with degree 4 polynomials one can already capture the change of sign of the Pareto front curvature (arising when the values of f_1 lie over $[10, 18]$). Observe also that higher-degree polynomials yield tighter underestimators of the left part of the Pareto front. The CPU

(a) Degree 4 estimators

(b) Degree 6 estimators
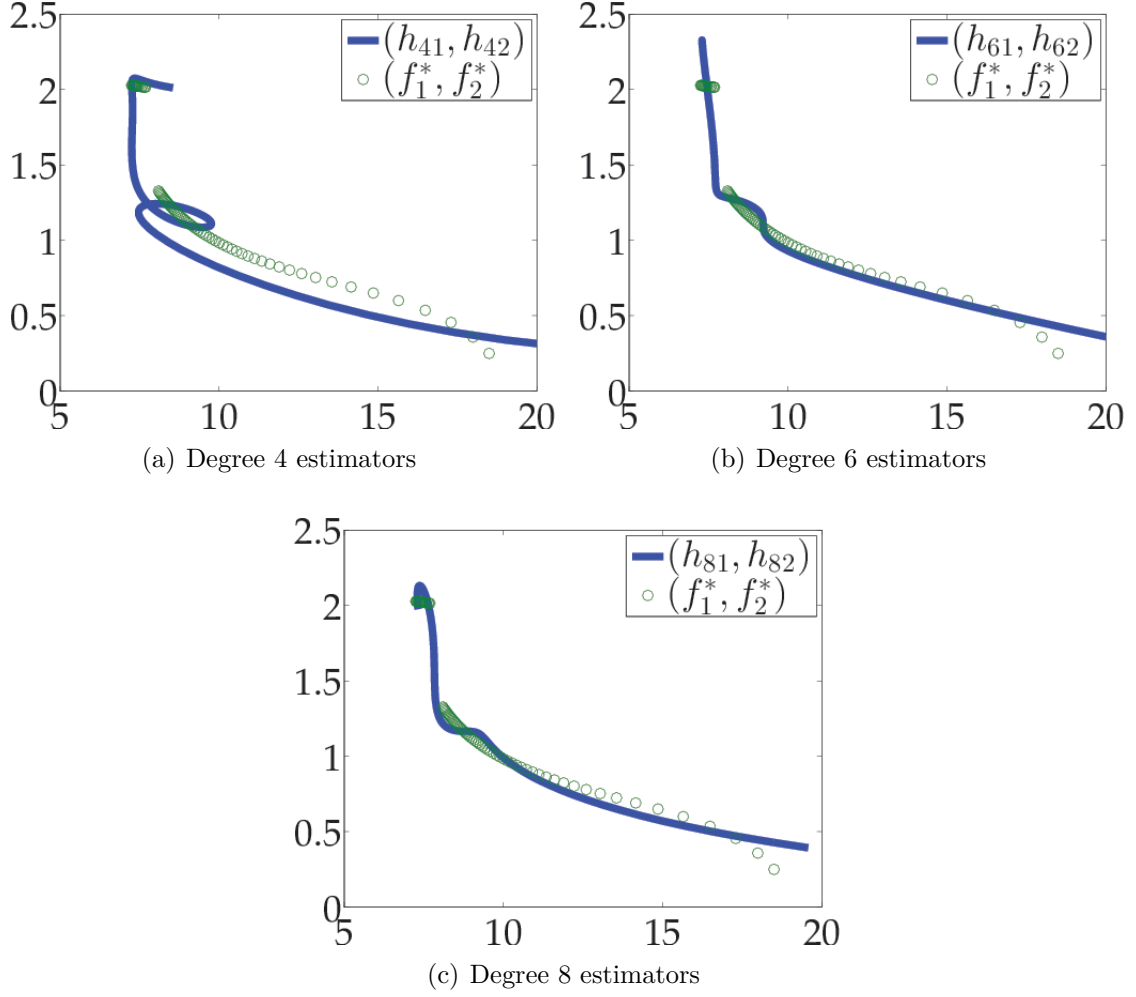
(c) Degree 8 estimators

Figure 5: A hierarchy of polynomial approximations of the Pareto curve for Example 2 obtained by the Chebyshev norm approximation (method (b))

time ranges from 0.5sec to compute the degree 4 polynomial q_4, to 1sec for the degree 6 computation and 1.7sec for the degree 8 computation. The discretization of the Pareto front is obtained by solving the polynomial optimization problems $\mathbf{P}_{\lambda_i}^u, i = 1, \ldots, N$. The corresponding running time of SDP programs is 51sec.

The same approach is used to solve the random bicriteria problem of Example 3.

Example 3. Here, we generate two random symmetric real matrices $\mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{R}^{15 \times 15}$ as well as two random vectors $\mathbf{q}_1, \mathbf{q}_2 \in \mathbb{R}^{15}$. Then we solve the quadratic bicriteria problem $\min_{\mathbf{x} \in [-1,1]^{15}} \{f_1(\mathbf{x}), f_2(\mathbf{x})\}$, with $f_j(\mathbf{x}) := \mathbf{x}^\top \mathbf{Q}_j \mathbf{x}/n^2 - \mathbf{q}_j^\top \mathbf{x}/n$, for each $j = 1, 2$.

Experimental results are displayed in Figure 7. For a 15 variable random instance, it consumes 18min of CPU time to compute q_4 against only 0.5sec for q_2 but the degree 4 underestimator yields a better point-wise approximation of the Pareto curve. The running time of SDP programs is more than 8 hours to compute the discretization of the front.

(a) Degree 4 estimators

(b) Degree 6 estimators
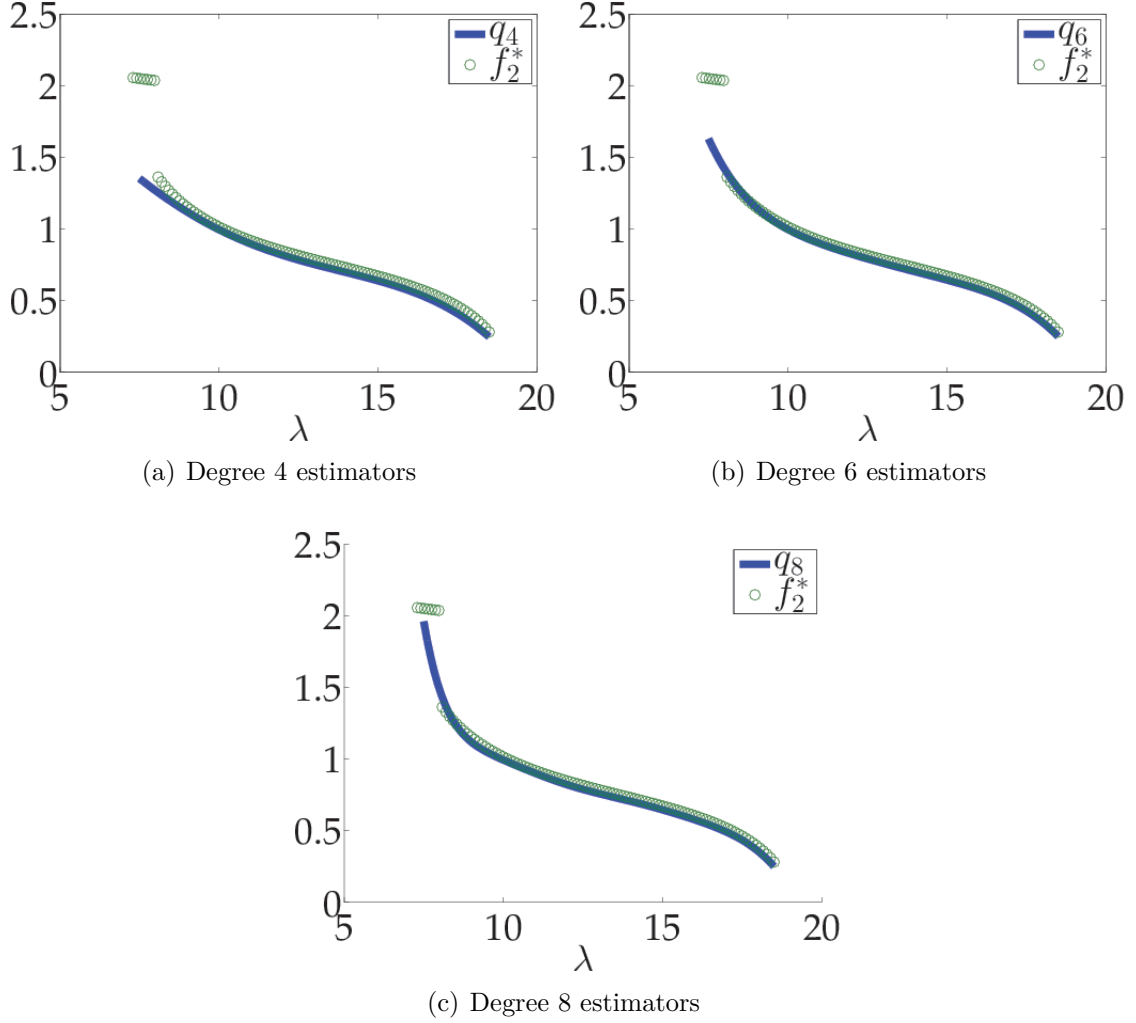
(c) Degree 8 estimators

Figure 6: A hierarchy of polynomial underestimators of $\lambda \mapsto f_2^*(\lambda)$ for Example 2 obtained by the parametric sublevel set approximation (method (c))

5   Conclusion

The present framework can tackle multicriteria polynomial problems by solving semidefinite relaxations of parametric optimization programs. The reformulations based on the weighted sum approach and the Chebyshev approximation allow to recover the Pareto curve, defined here as the set of weakly Edgeworth-Pareto points, by solving an inverse problem from generalized moments. An alternative method builds directly a hierarchy of polynomial underestimators of the Pareto curve. The numerical experiments illustrate the fact that the Pareto curve can be estimated as closely as desired using semidefinite programming within a reasonable amount of time for problem still of modest size. Finally our approach could be extended to higher-dimensional problems by exploiting the system properties such as sparsity patterns or symmetries.

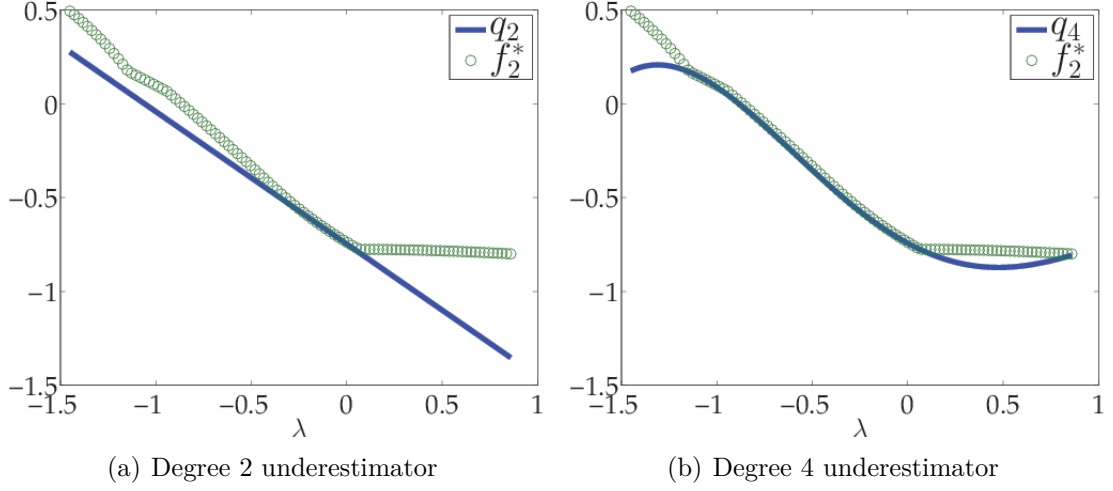(a) Degree 2 underestimator　　　　(b) Degree 4 underestimator

Figure 7: A hierarchy of polynomial underestimators of $\lambda \mapsto f_2^*(\lambda)$ for Example 3 obtained by the parametric sublevel set approximation (method (c))

Acknowledgments

A Appendix. An Inverse Problem from Generalized Moments

Suppose that one wishes to approximate each function f_j^*, $j = 1, 2$, with a polynomial of degree s. One way to do this is to search for $h_j \in \mathbb{R}_s[\lambda]$, $j = 1, 2$, optimal solution of

$$\min_{h \in \mathbb{R}_s[\lambda]} \int_0^1 (h(\lambda) - f_j^*(\lambda))^2 d\lambda, \quad j = 1, 2. \tag{15}$$

Let $\mathbf{H}_s \in \mathbb{R}^{(s+1) \times (s+1)}$ be the Hankel matrix associated with the moments of the Lebesgue measure on $[0, 1]$, i.e. $\mathbf{H}_s(i, j) = 1/(i + j + 1)$, $i, j = 0, \dots, s$.

Theorem A.1. *For each $j = 1, 2$, let $\mathbf{m}_j^s = (m_j^k) \in \mathbb{R}^{s+1}$ be as in (11). Then (15) has an optimal solution $h_{s,j} \in \mathbb{R}_s[\lambda]$ whose vector of coefficient $\mathbf{h}_{s,j} \in \mathbb{R}^{s+1}$ is given by:*

$$\mathbf{h}_{s,j} = \mathbf{H}_s^{-1} \mathbf{m}_j^s, \quad j = 1, 2. \tag{16}$$

Proof. Write

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

15

and observe that

$$A = \mathbf{h}^\top \mathbf{H}_s \mathbf{h} , \quad B = \sum_{k=0}^{s} h_k \int_0^1 \lambda^k f_j^*(\lambda) \, d\lambda = \sum_{k=0}^{s} h_k \, m_j^k = \mathbf{h}^\top \mathbf{m}_j ,$$

and so, as C is a constant, (15) reduces to

$$\min_{\mathbf{h} \in \mathbb{R}^{s+1}} \mathbf{h}^\top \mathbf{H}_s \mathbf{h} - 2\mathbf{h}^\top \mathbf{m}_j , \quad j = 1, 2 ,$$

from which (16) follows. $\qquad\qquad\square$

References

[1] R. Benayoun, J. Montgolfier, J. Tergny, and O. Laritchev. Linear programming with multiple objective functions: Step method (stem). *Mathematical Programming*, 1(1):366–375, 1971.

[2] Indraneel Das and J. E. Dennis. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM J. on Optimization*, 8(3):631–657, March 1998.

[3] Gabriele Eichfelder. Scalarizations for adaptively solving multi-objective optimization problems. *Comput. Optim. Appl.*, 44(2):249–273, November 2009.

[4] Bram L. Gorissen and Dick den Hertog. Approximating the pareto set of multiobjective linear programs via robust optimization. *Operations Research Letters*, 40(5):319 – 324, 2012.

[5] Didier Henrion, Jean-Bernard Lasserre, and Johan Lofberg. GloptiPoly 3: moments, optimization and semidefinite programming. *Optimization Methods and Software*, 24(4-5):pp. 761–779, August 2009.

[6] J. Jahn. *Vector Optimization: Theory, Applications, and Extensions*. Springer, 2010.

[7] Jean B. Lasserre. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM Journal on Optimization*, 17(3):822–843.

[8] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[9] Jean B. Lasserre. A "joint+marginal" approach to parametric polynomial optimization. *SIAM Journal on Optimization*, 20(4):1995–2022, 2010.

[10] K. Miettinen. *Nonlinear Multiobjective Optimization*, volume 12 of *International Series in Operations Research and Management Science*. Kluwer Academic Publishers, Dordrecht, 1999.

[11] Elijah Polak. On the approximation of solutions to multiple criteria decision making problems. In Milan Zeleny, editor, *Multiple Criteria Decision Making Kyoto 1975*, volume 123 of *Lecture Notes in Economics and Mathematical Systems*, pages 271–282. Springer Berlin Heidelberg, 1976.

[12] Hayato Waki, Sunyoung Kim, Masakazu Kojima, and Masakazu Muramatsu. Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity. *SIAM Journal on Optimization*, 17(1):218–242, 2006.

[13] Benjamin Wilson, David Cappelleri, Timothy W. Simpson, and Mary Frecker. Efficient Pareto Frontier Exploration using Surrogate Approximations. *Optimization and Engineering*, 2(1):31–50, 2001.