Operations Research Letters 42 (2014) 432–437

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Approximating Pareto curves using semidefinite relaxations

Victor Magron [a,b,*], Didier Henrion [a,b,c], Jean-Bernard Lasserre [a,b]

[a] LAAS-CNRS, 7 avenue du colonel Roche, F-31400 Toulouse, France
[b] Université de Toulouse, LAAS, F-31400 Toulouse, France
[c] Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic

ARTICLE INFO

Article history:
Received 18 April 2014
Received in revised form
24 July 2014
Accepted 28 July 2014
Available online 4 August 2014

Keywords:
Parametric polynomial optimization problems
Semidefinite programming
Multicriteria optimization
Sums of squares relaxations
Pareto curve
Inverse problem from generalized moments

ABSTRACT

We approximate as closely as desired the Pareto curve associated with bicriteria polynomial optimization problems. We use three formulations (including the weighted sum approach and the Chebyshev approximation) and each of them is viewed as a parametric polynomial optimization problem. For each case is associated a hierarchy of semidefinite relaxations and from an optimal solution of each relaxation one approximates the Pareto curve by solving an inverse problem (first two cases) or by building a polynomial underestimator (third case).

1. Introduction

Let $\mathbf{P}$ be the bicriteria polynomial optimization problem (POP) $\min_{\mathbf{x}\in\mathbf{S}}\{(f_1(\mathbf{x}), f_2(\mathbf{x}))\}$, 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 assumption:

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 set of Edgeworth–Pareto (EP) optimal points, defined e.g. in [6, Definition 11.3].

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

* Corresponding author at: LAAS-CNRS, 7 avenue du colonel Roche, F-31400 Toulouse, France. Tel.: +33 0561336945.
E-mail address: magron@laas.fr (V. Magron).

http://dx.doi.org/10.1016/j.orl.2014.07.007
0167-6377/

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

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

with $\overline{\mathbf{x}}$ 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 curve 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 us to combine the adaptive control of discretization points with the modified Polak method. In [2], Das and Dennis introduce the Normal-boundary intersection (NBI) 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 (SDP); 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 with three methods: *method* (a) for the first formulation (3) when $p = 1$, this is a *weighted convex sum approximation*; *method* (b) for the second formulation (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 (SDP) which permits us 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:

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

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 us to define an approximation of the Pareto curve. In other words even in the absence of convexity the SDP hierarchy allows us to approximate the Pareto curve 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 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 $\mathbb{N}_d^{n+1} := \{\beta \in \mathbb{N}^{n+1} : \sum_{j=1}^{n+1} \beta_j \le d\}$, whose cardinal is $s_n(d) = \binom{n+1+d}{d}$. A polynomial $f \in \mathbb{R}[\lambda, \mathbf{x}]$ is written $f(\lambda, \mathbf{x}) = \sum_{(k,\alpha)\in\mathbb{N}^{n+1}} f_{k\alpha}\lambda^k\mathbf{x}^\alpha$ 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:

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

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 $(i, \alpha), (j, \beta) \in \mathbb{N}_d^{n+1}$.