On the hierarchical structure of Pareto critical sets

Bennet Gebken[1], Sebastian Peitz[1] and Michael Dellnitz[1]

[1]*Department of Mathematics, Paderborn University, Germany*

Abstract. In this talk we show that the boundary of the Pareto critical set of an unconstrained multiobjective optimization problem (MOP) consists of Pareto critical points of subproblems considering subsets of the objective functions. If the Pareto critical set is completely described by its boundary (e.g., if we have more objective functions than dimensions in the variable space), this can be used to solve the MOP by solving a number of MOPs with fewer objective functions. If this is not the case, the results can still give insight into the structure of the Pareto critical set. This technique is especially useful for efficiently solving many-objective optimization problems by breaking them down into MOPs with a reduced number of objective functions. For further details on this topic, we refer to [1].

Extended abstract

Given an unconstrained multiobjective optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \quad f = (f_1, ..., f_k) : \mathbb{R}^n \to \mathbb{R}^k \text{ differentiable,} \tag{MOP}$$

a necessary condition for (local) Pareto optimality at some point x is the non-existence of a common descent direction of all objective functions f_i. Such x are called *Pareto critical* and they form a set P called the *Pareto critical set*. If all objective functions are smooth, convex and $k \leq n$, then P is diffeomorphic to a $(k-1)$-dimensional simplex, so its structure is well understood [2]. To be precise, each of its facets is given by a Pareto critical set of an MOP that consists of $k-1$ objective functions of the original MOP. For example, Figure 1 shows P for the case of three objective functions in $\mathbb{R}^2$ that are paraboloids. There are also results about the structure for the general case showing

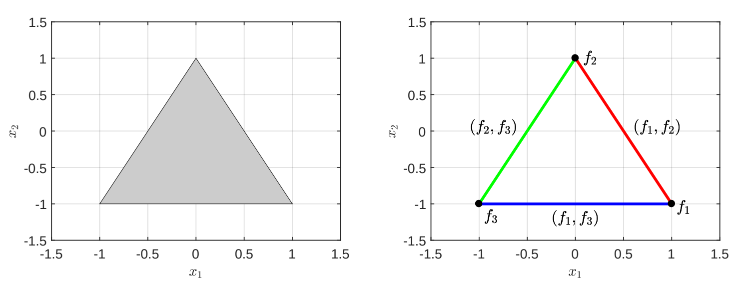

FIGURE 1. Pareto critical set P (left) and Pareto critical sets of all 2-objective subproblems (right) for three paraboloid objective functions in $\mathbb{R}^2$.

that generically, P is a stratified set [3] (i.e., roughly speaking, it is a manifold with boundary and corners). First results about its "boundary" indicate that it is still given by subproblems with fewer objective functions [4, 5]. The motivation for our work is to show how to (define and) calculate the boundary in the general case.

Our central approach is to use the fact that a point is Pareto critical if and only if it satisfies the KKT conditions [6], which means that $x \in P$ if and only if there is some *KKT multiplier* $\alpha \in (\mathbb{R}^{\geq 0})^k$ with

$$Df(x)^\top \alpha = 0 \quad \text{and} \quad \sum_{i=1}^{k} \alpha_i = 1. \tag{KKT}$$

This creates a relationship between P and the set of possible KKT multipliers $\Delta_k := \{\alpha \in (\mathbb{R}^{\geq 0})^k : \sum_{i=1}^k \alpha_i = 1\}$. Observe that a single Pareto critical point can have multiple different KKT multipliers and multiple Pareto critical points can have the same KKT multiplier, so in general, this induces no real mapping between P and Δ_k. Nevertheless, it is possible to use the KKT conditions to deduce properties of the boundary of P from the boundary of Δ_k.

Note that in general, the topological boundary ∂P of P (with respect to the subspace topology of $\mathbb{R}^n$) is not what we mean by "boundary". (For example, in Figure 1, the topological boundary of the Pareto critical set of the MOP with the objective function (f_1, f_2) is the set itself.) To define what we actually mean, we first have to classify the points in P via Δ_k:

$$\begin{aligned} P_0 &:= \{x \in P : \forall \alpha \in \Delta_k \text{ with } Df(x)^\top \alpha = 0 \text{ there is some } i \in \{1, ..., k\} \text{ with } \alpha_i = 0\}, \\ P_{int} &:= P \setminus P_0. \end{aligned}$$

In other words, P_0 contains all the Pareto critical points where for each KKT multiplier, not all objective functions are needed in (KKT), and P_{int} contains all the Pareto critical points where there is at least one KKT multiplier that uses all objective functions. Let $\mathrm{Tan}(A, x)$ be the tangent cone of $A \subseteq \mathbb{R}^n$ at $x \in \mathbb{R}^n$. Then we call

$$P_E := \{x \in P : \mathrm{Tan}(P_{int}, x) \neq -\mathrm{Tan}(P_{int}, x)\}$$

the *boundary* (or *edge*) *of P*. Our goal can now be rephrased as investigating the relationship between P_0, P_{int} and P_E.

To this end, first note that the KKT conditions can be written as the nonlinear system of equations $\tilde{F}(x, \alpha) = 0$ with

$$\tilde{F} : \mathbb{R}^n \times \overline{\Delta^{k-1}} \to \mathbb{R}^n, \quad (x, \alpha) \mapsto \sum_{i=1}^{k-1} \alpha_i \nabla f_i(x) + (1 - \sum_{i=1}^{k-1} \alpha_i) \nabla f_k(x)$$

and $\Delta^{k-1} := \{\alpha \in (\mathbb{R}^{>0})^{k-1} : \sum_{i=1}^{k-1} \alpha_i < 1\}$. Let $\mathcal{M} := (\tilde{F}|_{\mathbb{R}^n \times \Delta^{k-1}})^{-1}(0)$ be the set of points in P_{int} paired with their corresponding (positive) KKT multipliers. In particular, $pr_x(\mathcal{M}) = P_{int}$, where $pr_x : \mathbb{R}^{n+k-1} \to \mathbb{R}^n$ is the projection onto the first n components. We will assume from now on that f is C^2 and that $D_x\tilde{F}(x, \alpha)$ is invertible for all $(x, \alpha) \in \mathcal{M}$. Similar to [7], Theorem 5.1, this allows us to show that $\mathcal{M}$ is a manifold. In particular, its tangent space is given by $T_{(x,\alpha)}\mathcal{M} = \ker(D\tilde{F}(x, \alpha))$.

Since we have defined the boundary of P via tangent cones of P_{int}, we now look at the relationship between $\mathrm{Tan}(P_{int}, x) = \mathrm{Tan}(pr_x(\mathcal{M}), x)$ and $pr_x(T_{(x,\alpha)}\mathcal{M})$ for $(x, \alpha) \in \mathcal{M}$. We can immediately show the following lemma by taking curves in $T_{(x,\alpha)}\mathcal{M}$ and projecting them onto $\mathbb{R}^n$ via pr_x.

Lemma 1 *$pr_x(T_{(x,\alpha)}\mathcal{M}) \subseteq Tan(P_{int}, x)$.*

To show the converse of Lemma 1, we first need a result about the uniqueness of KKT multipliers, i.e., about the set $A(x) := \{\alpha \in \overline{\Delta^{k-1}} : \tilde{F}(x, \alpha) = 0\}$.

Lemma 2
1. *Let $x \in P$. If $rk(Df(x)) < k - 1$ then $A(x) \cap \partial\Delta^{k-1} \neq \emptyset$.*
2. *Let $x \in P_{int}$. Then $rk(Df(x)) = k - 1 \Leftrightarrow |A(x)| = 1$.*

As a converse of Lemma 1 we get:

Lemma 3 *Let $x \in P_{int}$ with $rk(Df(x)) = k - 1$. Then there exists $\alpha \in A(x) \cap \Delta^{k-1}$ with*

$$Tan(P_{int}, x) \subseteq pr_x(T_{(x,\alpha)}\mathcal{M}).$$

Combining Lemma 1 and Lemma 3, we get $\mathrm{Tan}(P_{int}, x) = pr_x(T_{(x,\alpha)}\mathcal{M})$ for $x \in P_{int}$ with $rk(Df(x)) = k - 1$. In particular, $\mathrm{Tan}(P_{int}, x)$ is a vector space, so such x can not lie in P_E. This leads us to our first main result.

Theorem 1 *If $x \in P_E$ then $A(x) \cap \partial\Delta^{k-1} \neq \emptyset$.*

For $I \subseteq \{1, ..., k\}$ let P^I be the Pareto critical set of the problem (MOP_I) with the objective function $f^I := (f_i)_{i \in I}$. By Theorem 1, we know that

$$P_E \subseteq \bigcup_{I \subseteq \{1,...,k\}, |I| = k-1} P^I.$$

What is left to investigate is the question of how tight this covering is. Lemma 2 shows that if the rank of the Jacobian of f is too small, then every Pareto critical point has a KKT multiplier with a zero component somewhere. In this case, taking $k - 1$ objective functions is too much.

The following lemma gives a first upper bound for the size of the subproblems that we have to consider.

Lemma 4 *Let $x \in P$. Then there is some $I \subseteq \{1, ..., k\}$ with $|I| = rk(Df(x)) + 1$ such that $rk(Df^I(x)) = rk(Df(x))$ and $x \in P^I$. Moreover, if $x \in P_0$ then $x \in P_0^I$.*

This shows that the union of the Pareto critical sets of all subproblems of size $\max_{x \in P} rk(Df(x)) + 1$ is the complete Pareto critical set. In particular, for each $x \in P_{int}$ there is some I so that we can apply Lemma 2 to (MOP_I). This allows us to show the following lemma, where $A^I(x)$ denotes the set of KKT multipliers of x with respect to (MOP_I).

Lemma 5 *Let $x \in P_E$ and $m := \max_{x \in P} rk(Df(x))$. Then there has to be some $I \subseteq \{1, ..., k\}$ with $|I| \leq m + 1$ such that either $I = \{i\}$ and $\nabla f_i(x) = 0$ or $A^I(x) \cap \partial \Delta^{|I|-1} \neq \emptyset$.*

As a corollary of this we get our second main result.

Corollary 1 *Let $m := \max_{x \in P} rk(Df(x)) > 0$. Then*

$$P_E \subseteq \bigcup_{I \subseteq \{1,...,k\}, |I| = m} P^I.$$

This is a tight covering in the sense that we can not take subproblems with fewer objective functions and still cover P_E. This can for example be seen in the three paraboloid example (Figure 1). For MOPs with $k > n$ this tells us that it is sufficient to consider all n-objective subproblems.

Obviously, each subproblem (MOP_I) for $I \subseteq \{1, ..., k\}$ is an MOP itself, so we can apply the theory above to calculate the boundary of its Pareto critical set P^I. If we do this iteratively, starting with the original MOP, we get a hierarchy of boundaries of P. Although information is lost when calculating these "lower-dimensional" boundaries, they are easier to compute and can still give important information about the structure of the full critical set.

The following example shows an application of our main results.

Example 1

$$\min_{x \in \mathbb{R}^2} f(x) \quad \text{with} \quad f(x) := \begin{pmatrix} 0.5(x_1 - 1)^2 + x_2^2 \\ 2x_1^2 + 2(x_2 - 1)^2 \\ 2(x_1 + 1)^2 + x_2^5 \\ -2x_1^3 + 2(x_2 + 1)^2 \end{pmatrix}.$$

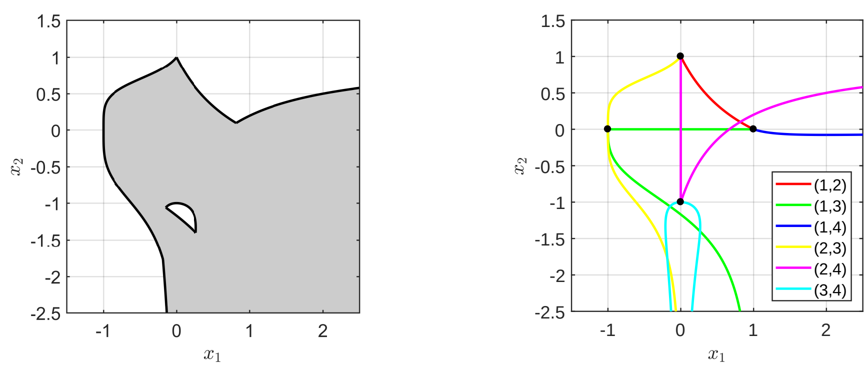

FIGURE 2. Pareto critical set P (left) and Pareto critical sets of all 2-objective subproblems (right) for Example 1.

Since the Pareto critical set in Example 1 is unbounded, Figure 2 (left) only shows the interesting part of it. By Corollary 1, it is sufficient to only consider the 2-objective subproblems, whose Pareto critical sets are depicted in Figure 2 (right). Since $rk(Df) = n$ in this example, they divide the variable space into different components. There are three notable features of P on the left side that can be explained by looking at the boundary structure on the right side: First of all, the hole in P arises as one of the components on the right side, given by the area between $P^{\{1,3\}}$ (green) and $P^{\{3,4\}}$ (teal). In particular, this means that this hole will also occur if we consider the MOP with the objective function (f_1, f_3, f_4). So in a way, the hole is caused by those three objective functions. Secondly, there are two types of corners in P. The first type are corners that arise as Pareto critical points of 1-objective subproblems. These could also be characterized as boundaries of the 2-objective subproblems. For example, $(0, 1)$ is a critical point of f_2 and a boundary point of the subproblem with the objective function (f_1, f_2). The second type are corners that arise as (transverse) intersections of two boundaries, e.g., the intersection of the red and the purple boundary. Lastly, one might expect a corner of P at $(-1, 0)$, as there is a critical point of f_2. The fact that there is no corner is caused by f_3, because $P^{\{2,3\}}$ goes "smoothly" through $(-1, 0)$.

When it comes to numerical algorithms, our results can in principle be used in any algorithm that can calculate the Pareto critical set, such as continuation methods [7, 8] or (globalized) descent methods [9, 10], by simply applying them to each possible subproblem (of proper size). As a result, one gets a boundary structure like in Figure 2, (right). Since there are $\binom{k}{m} = \frac{k!}{m!(k-m)!}$ subproblems of size m, the number of subproblems that have to be solved can become very large. This is partly due to the fact that Pareto critical sets (and in particular their boundaries) can become complicated for a larger number of objectives. But in addition to the relevant subproblems, there can also be subproblems whose Pareto critical set is redundant in the sense that it does not actually lie on the boundary P_E of the original MOP. An example for this is the subproblem with the objective function (f_1, f_4) in Figure 2. The reason for this is that Corollary 1 only states that we get a superset of P_E. As there is no definitive way to tell a priori which subproblems are necessary for an arbitrary MOP, in theory, one has to solve them all. But in practice there are heuristics that can be applied. For example, one can start with rough approximations of all P^I and then only refine the ones that are relevant.

If the topological boundary ∂P coincides with the boundary P_E (like in Example 1), the critical set P is completely characterized by P_E and the variable space is divided into different (connected) components (like in Figure 2, right). This typically happens when $k \geq n + 1$, i.e., when the dimension of the image space is larger than the dimension of the variable space. This means that after calculating the boundary structure, one merely has to test one point of each component for Pareto criticality, and then fill out those components that contain a critical point. Since the "dimension" of the boundary is smaller than the "dimension" of the full critical set, this is a highly efficient way to solve such MOPs. If P_E does not coincide with ∂P, which is typically the case when $k \leq n$, it can still give important information about the structure of P and could, for example, be used to generate starting points for an algorithm that calculates the complete Pareto (critical) set.

We conclude that the boundary of the Pareto critical set (defined via tangent cones) can be covered by Pareto critical sets of subproblems with fewer objective functions (Theorem 1). The number of objective functions that have to be considered in the subproblems depends on the rank of the Jacobian of the full objective function (Corollary 1). This can be used to analyze the structure of Pareto critical sets, which we demonstrated in an example. Additionally, this can be used to efficiently solve many-objective optimization problems by breaking them down into MOPs with fewer objective functions.

REFERENCES

[1] B. Gebken, S. Peitz, and M. Dellnitz, "On the hierarchical structure of Pareto critical sets", 2018, arXiv:1803.06864 .

[2] S. Smale, "Global Analysis and Economics I: Pareto Optimum and a Generalization of Morse Theory," in *Dynamical Systems* (Academic Press, 1973), pp. 531 – 544.

[3] W. De Melo, "On the structure of the Pareto set of generic mappings," in *Bulletin of the Brazilian Mathematical Society*, Vol. 7 (Springer, 1976), pp. 121–126.

[4] S. Peitz, "Exploiting structure in multiobjective optimization and optimal control," Ph.D. thesis, Paderborn University 2017.

[5] A. Lovison and F. Pecci, "Hierarchical stratification of Pareto sets", 2014, arXiv:1407.1755 .

[6] M. Ehrgott, *Multicriteria Optimization* (Springer-Verlag Berlin Heidelberg, 2005).

[7] C. Hillermeier, *Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach* (Birkhäuser Basel, 2001).

[8] O. Schütze, *Set Oriented Methods for Global Optimization* (PhD thesis, Paderborn University, Paderborn, Germany, 2004).

[9] J. Fliege and B. F. Svaiter, "Steepest Descent Methods for Multicriteria Optimization," in *Mathematical Methods of Operations Research. Vol 51 (2000), No 3, 479-494* (2000).

[10] M. Dellnitz, O. Schütze, and T. Hestermeyer, "Covering Pareto Sets by Multilevel Subdivision Techniques," in *Journal of Optimization Theory and Applications*, Vol. 124 (2005), pp. 113–136.