SIAM J. OPTIM.
Vol. 30, No. 3, pp. 2659–2686

TOPOLOGY OF PARETO SETS OF STRONGLY CONVEX PROBLEMS*

NAOKI HAMADA†, KENTA HAYANO‡, SHUNSUKE ICHIKI§, YUTARO KABATA¶, AND HIROSHI TERAMOTO‖

Dedicated to Professor Takashi Nishimura on the occasion of his 60th birthday

Abstract. A multiobjective optimization problem is simplicial if the Pareto set and front are homeomorphic to a simplex and, under the homeomorphisms, each face of the simplex corresponds to the Pareto set and front of a subproblem that treats a subset of objective functions. In this paper, we show that strongly convex problems are simplicial under a mild assumption on the ranks of the differentials of the objective mappings. We further prove that one can make any strongly convex problem satisfy the assumption by a generic linear perturbation, provided that the dimension of the source is sufficiently larger than that of the target. We demonstrate that the location problems, a biological modeling, and the ridge regression can be reduced to multiobjective strongly convex problems via appropriate transformations preserving the Pareto ordering and the topology.

Key words. multiobjective optimization, strongly convex mapping, simplicial problem, topology of differentiable mapping

AMS subject classifications. 90C25, 90C29, 57R35, 57R45

DOI. 10.1137/19M1271439

1. Introduction.

Multiobjective optimization arises in various fields of science and engineering (e.g., data mining [26, 27], finance [29], car design [36], and more [46]). A common scenario in the classical decision making is that users first specify their preference, then scalarize objective functions according to the preference, and finally solve a scalarized problem to find the preferred Pareto solution [4, 21]. In contrast to this a priori approach, the recent computational power enables us to take an a posteriori approach: users first solve a multiobjective problem to obtain the

*Received by the editors July 1, 2019; accepted for publication (in revised form) June 28, 2020; published electronically September 28, 2020.
https://doi.org/10.1137/19M1271439

Funding: The second author was supported by JSPS KAKENHI grant JP17K14194. The third author was supported by JSPS KAKENHI grants JP16J06911, JP19J00650, and JP17H06128. The fourth author was supported by JSPS KAKENHI grant JP16J02200. The fifth author was supported by JSPS KAKENHI grant JP19K03484 and by JST PRESTO grant JPMJPR16E8. The authors were supported by the RIKEN Center for Advanced Intelligence Project. This work was also supported by the Research Institute for Mathematical Sciences, a Joint Usage/Research Center located in Kyoto University.

†Engineering Division, KLab Inc., Tokyo, 106-6122, Japan, and RIKEN AIP Fujitsu Collaboration Center, RIKEN Center for Advanced Intelligence Project, Tokyo, 103-0027, Japan (hamanaoki@gmail.com).

‡Department of Mathematics, Faculty of Science and Technology, Keio University, Yokohama, 223-8522, Japan, and Mathematical Science Team, RIKEN Center for Advanced Intelligence Project, Tokyo, 103-0027, Japan (k-hayano@math.keio.ac.jp).

§Department of Mathematical and Computing Science, School of Computing, Tokyo Institute of Technology, Tokyo 152-8552, Japan (ichiki@c.titech.ac.jp).

¶School of Information and Data Sciences, Nagasaki University, Nagasaki 852-8131, Japan (kabata@imi.kyushu-u.ac.jp).

‖Molecule & Life Nonlinear Sciences Laboratory, Research Institute for Electronic Science, Hokkaido University, Sapporo 001-0020, Japan, JST, PRESTO, Department of Research Promotion, Tokyo, 102-0076, Japan, and Institute for Chemical Reaction Design and Discovery, Hokkaido University, Sapporo, Japan (teramoto@es.hokudai.ac.jp).

whole Pareto set (or an approximation of it), then consider the trade-off between objective functions, and finally make a choice of the preferred solution. In the a posteriori approach, some multiobjective solvers utilize scalarization for guiding search directions in order to obtain the entire Pareto set (see, for example, [3, 5, 10, 45]).

In general it becomes easier to obtain the whole Pareto set when assuming topological conditions on (the Pareto sets/fronts of) problems. Lovison introduced a piecewise linear approximation of the Pareto sets under some assumptions on the *genericity* of problems and configurations of sample points [17]. Applying a generalized homotopy method [10] one can find the entire *connected* Pareto set from a single Pareto solution. Furthermore, we can efficiently compute a parametric-surface approximation of the entire Pareto set, provided that the problem is *simplicial* [15]. Figure 1.1 depicts an example of a simplicial problem with three objective functions f_1, f_2, f_3. As this figure indicates, there exists a homeomorphism Φ from a simplex Δ^2 to the Pareto set, sending each face of Δ^2 to the Pareto set of a subproblem that treats a subset of $\{ f_1, f_2, f_3 \}$ (for the precise definition of simplicial problems, see subsection 2.2). The property that Φ maps the faces to the Pareto sets of the subproblems has a great utility: it is known that using this correspondence, one can reduce the number of solutions to compute an approximation of the mapping Φ (see [15, 35]).

Simpliciality has been widely observed in practical problems for a long time but has not been rigorously studied until very recently. In 1967, Kuhn dealt with a location problem under the Euclidean norm and showed that its Pareto set is the convex hull of demand points [16]. Thus it is easy to see that for m demand points in general position, the Pareto set becomes an $(m-1)$-simplex whose $(k-1)$-faces are the Pareto sets of subproblems with k demand points $(1 \le k \le m)$. In 1973, Smale claimed (without giving proofs) that the Pareto set of a pure exchange economy with m agents is homeomorphic to the $(m-1)$-simplex under some assumptions [32]. In 2012, Shoval et al. rediscovered Kuhn's result in the biological context and further pointed out (without mathematical rigor) that the Pareto set will be curved when different objectives are measured by different norms [30]. It was 2019 when Kobayashi et al. first defined simpliciality [15]. So far, simpliciality has been observed only in specific problems and has not been proven in any general class of optimization problems such as linear problems, convex problems, etc.

Another important aspect is the genericity of properties, which means that such a property holds true for "almost all" problems. Smale conjectured in [32] that in the space of C^∞-mappings endowed with the Whitney C^∞-topology, there is a dense

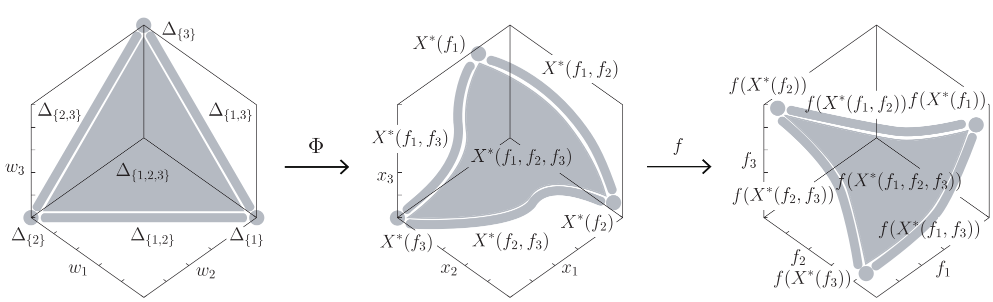


FIG. 1.1. *The Pareto set $X^*(f)$ (middle) and front $f(X^*(f))$ (right) of a simplicial problem $f = (f_1, f_2, f_3)$. There exists a homeomorphism Φ from a simplex Δ^2 to $X^*(f)$, sending each face to the Pareto set of a subproblem. The composite $f \circ \Phi : \Delta^2 \to f(X^*(f))$ is also a homeomorphism.*

subset[1] of mappings such that the Pareto critical set (a superset of the local Pareto set) is a stratified set of dimension $m-1$, which is a disjoint union of k-manifolds (called *strata*) with some gluing conditions ($0 \leq k \leq m-1$). De Melo gave a proof of it under different assumptions than those Smale put forth [1]. There remained a question of whether the stratification given in his proof is always "fine" or not in the sense that the local Pareto set is a union of strata. Wan showed under another generic condition that the local Pareto set admits a Whitney stratification for unconstrained problems [41] and constrained problems [42]. Independently of [41], Vershik and Chernyakov also proved a similar result [40] and extended it to the case where the Pareto ordering is replaced by an arbitrary ordering defined by a convex cone [39]. Recently, Lovison and Pecci provided a comprehensive survey around Wan's results [18]. Gebken, Peitz, and Dellnitz studied a situation that the Pareto critical set can be determined by a reduced number of objectives when the dimension of the objective space is greater than that of the decision space under some regularity assumptions (though the relationship to genericity was not investigated) [6]. Compared to the above studies, simpliciality gives a concrete stratification of the Pareto set and front with its computationally efficient approximation method. Although simpliciality does not seem to be a generic property in the mapping space (i.e., the class of all problems), we still expect that there exists a restricted problem class where simpliciality is ensured via certain generic perturbations.

In this paper, we will show that unconstrained strongly convex problems are such a problem class.[2]

THEOREM 1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex C^r-mapping* ($2 \leq r \leq \infty$). *The multiobjective optimization problem of minimizing f is C^{r-1}-weakly simplicial. Furthermore, this problem is C^{r-1}-simplicial if the corank of the differential df_x is equal to* 1 *for any $x \in X^*(f)$.*

For the definitions of strong convexity and weak simpliciality, see section 2. The assumption on the corank of df_x in the latter part of Theorem 1.1 is called the *rank assumption* in [32].

Under the assumptions of Theorem 1.1, the weighted-sum scalarization gives an instance of the required homeomorphism Φ for the problem to be simplicial, that is, the mapping $x^* : \Delta^{m-1} \to X^*(f)$ defined by

$$x^*(w) = \arg\min_{x \in \mathbb{R}^n} \left(\sum_{i=1}^{m} w_i f_i(x) \right) \tag{1.1}$$

is a diffeomorphism. Since the scalarized function $\sum_{i=1}^{m} w_i f_i : \mathbb{R}^n \to \mathbb{R}$ is strongly convex (see subsection 2.3), the mapping x^* is well-defined and can be efficiently computed by, e.g., gradient methods. By this property, scalarization-based subdivision methods [5, 8, 25, 31] are able to provide a sample of solutions for parametric-surface approximation methods developed in [15, 35] to build an approximation of Φ.

Note that (strict) convexity of a mapping does not necessarily imply that the corresponding problem is simplicial. For example, the single-objective problem of minimizing the function $\exp(x)$ (defined on $\mathbb{R}$) does not have a Pareto solution (i.e., a minimizer); in particular it is not simplicial, although $\exp(x)$ is strictly convex.

[1]Smale also characterized such a subset. See [32, Proposition, p. 540] for precise conditions.

[2]While it has been stated that *any* convex problem is simplicial in the literature, it is too optimistic to expect so, as one can easily find a counterexample. See, e.g., Example 3.4 and Remark 4.4.

Remark 1.2. The strong convexity guarantees not only simpliciality but also that the mapping x^* is a valid homeomorphism Φ between Δ^{m-1} and the Pareto set $X^*(f)$. Indeed, as illustrated with a counterexample in Appendix B, it may happen that for a convex and C^∞-simplicial problem the above formula (1.1) for x^* is not even an univocally defined mapping.

As Example 3.5 indicates, we cannot drop the rank assumption in Theorem 1.1. Nevertheless, one can make any strongly convex mapping satisfy the rank assumption by a generic (i.e., almost all, including arbitrarily small) linear perturbation, under the assumption that the dimension of the source is sufficiently larger than that of the target (we call this assumption the *dimension assumption*). Note that Smale [32] also showed the genericity of the rank assumption, in which the dimension assumption was introduced. However, the topology Smale considered is somewhat complicated, in particular adding small linear (even polynomial) terms to a problem is not a perturbation in the sense of Smale. Thus one cannot deduce our result from that due to Smale. See section 4 for details. We will further observe that Theorem 4.1 would not hold without the dimension assumption (cf. Remark 4.4). Note that a small linear perturbation of a strongly convex problem does not cause substantial changes of the Pareto set (see Remark 4.5 for details).

Remark 1.3. That is to say, the Pareto set of a strongly convex problem is "stable" with respect to linear perturbations of objective functions under the Euclidean topology. To the best of the authors' knowledge, this type of stability is novel and logically independent of existing results on stability. Smale [32] and de Melo [2] studied the stability of the Pareto critical set of a generic problem with respect to perturbations of objective functions under the Whitney C^∞-topology. Miglierina and Molho studied the stability of the Pareto set of a convex problem with respect to perturbations of the feasible region under the topology of Hausdorff set-convergence [24]. They also showed that the strict global optimum of a scalarized problem, if it is a proper Pareto point of the original problem, is stable with respect to perturbations of the feasible region under the topology of Kuratowski–Painleve set-convergence [23]. Since they use different perturbations of different objects from ours, any of their stabilities do not imply our stability and vice versa. Furthermore, Smale [33] and Miglierina [22] discussed a quite different type of stability concerning behavior of "admissible curves." Since they did not consider perturbations, this type of stability is not related to ours.

Applying Theorem 1.1, we will show in section 5 that several practical problems are (weakly) simplicial. We will observe in subsection 5.1 that Kuhn's location problem [16], which is convex but not strongly convex, can be made strongly convex by a Pareto-order-preserving transformation on the target space. In particular, our theorem gives an alternative proof of Kuhn's result on simpliciality. A multiobjective problem for modeling phenotypic divergence of species proposed by Shoval et al. [30] also becomes strongly convex after the same transformation (see subsection 5.2). We can thus give a rigorous proof for their observations of simpliciality. The ridge regression [11] used in data analysis is a single-objective (strongly convex) problem in its original form. In subsection 5.3, we will reformulate it into a multiobjective strongly convex one and show that the Pareto set of the reformulated problem contains the solutions to the original problem with different values of the hyper-parameter (i.e., the coefficient of the regularization term). With few solutions (i.e., regression models trained with some specific hyper-parameters), we can efficiently compute a parametric-surface approximation of the whole Pareto set (i.e., the set of regression models trained with all possible hyper-parameters), which will reduce the effort of hyper-parameter tuning.

Remark 1.4. As we have stated, Smale's pure exchange economy [32] seems to be simplicial. However, this problem cannot be a strongly convex problem, even after any Pareto-order-preserving transformation. Thus our results imply nothing about Smale's discussion and vice versa; see Appendix C for a detailed comparison.

The paper is organized as follows. After preliminaries in section 2, the proof of the main theorem is given in section 3. In section 4, it is shown that, under some assumption on the dimensions of the source and the target, any strongly convex problem becomes simplicial after a generic linear perturbation. Section 5 is devoted to discussing practical problems.

2. Preliminaries.

We introduce the definition of strongly convex problems and their properties and define C^r-(weakly) simplicial problems. Throughout the paper, we denote the index set $\{ 1, \dots, m \}$ by M.

2.1. Multiobjective optimization.

A multiobjective optimization problem is a problem of minimizing objective functions $f_1, \dots, f_m : X \to \mathbb{R}$ over a subset $X \subseteq \mathbb{R}^n$:

$$\begin{aligned} &\text{minimize } f(x) = (f_1(x), \dots, f_m(x)) \\ &\text{subject to } x \in X (\subseteq \mathbb{R}^n). \end{aligned}$$

According to the *Pareto ordering*, i.e.,

$$f(x) \prec f(y) \overset{\text{def}}{\Longleftrightarrow} f_i(x) \le f_i(y) \ \forall i \in M \text{ and } f_j(x) < f_j(y) \ \exists j \in M,$$

we basically would like to obtain the *Pareto set*

$$X^*(f) = \{ x \in X \mid f(y) \not\prec f(x) \ \forall y \in X \}$$

and the *Pareto front*

$$f(X^*(f)) = \{ f(x) \in \mathbb{R}^m \mid x \in X^*(f) \} .$$

2.2. Simplicial problems.

Here, we explain the definition of C^r-(weakly) simplicial problems for $0 \le r \le \infty$. For $\varepsilon \ge 0$, we define the subset $\Delta_\varepsilon^{m-1} \subsetneq \mathbb{R}^m$ as follows:

$$\Delta_\varepsilon^{m-1} = \left\{ (w_1, \dots, w_m) \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^m w_i = 1,\ w_i > -\varepsilon \right\} \subsetneq \mathbb{R}^m.$$

Note that the closure $\overline{\Delta_0^{m-1}}$ is the standard simplex, which we will denote by

$$\Delta^{m-1} = \left\{ (w_1, \dots, w_m) \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^m w_i = 1,\ w_i \ge 0 \right\} \subsetneq \mathbb{R}^m.$$

We also denote a face of Δ^{m-1} for $I \subseteq M$ by

$$\Delta_I = \left\{ (w_1, \dots, w_m) \in \Delta^{m-1} \;\middle|\; w_i = 0 \ (i \notin I) \right\} \subsetneq \mathbb{R}^m.$$

Figure 2.1 describes these sets for $m = 3$.

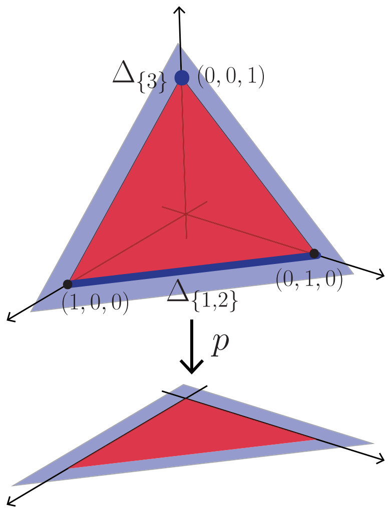


FIG. 2.1. *The upper figure contains Δ^2_ε (blue), Δ^2 (red), and its faces (dark blue), and the lower one contains D^2_ε (blue) and D^2 (red). Color available online.*

For a subset $U \subseteq \mathbb{R}^m$, a continuous mapping $f : \Delta^{m-1} \to U$ is a C^r*-mapping* if there exist $\varepsilon > 0$ and a C^r-mapping $\tilde{f} : \Delta^{m-1}_\varepsilon \to \mathbb{R}^m$ satisfying $\tilde{f}|_{\Delta^{m-1}} = f$.[3] A subspace $X \subseteq \mathbb{R}^n$ is C^r*-diffeomorphic* to the simplex Δ^{m-1} if there exist $\varepsilon > 0$ and a C^r-immersion $\phi : \Delta^{m-1}_\varepsilon \to \mathbb{R}^n$ such that $\phi|_{\Delta^{m-1}} : \Delta^{m-1} \to X$ is a homeomorphism. The reader can refer to [20, section 2] for more general definition of diffeomorphisms between manifolds with corners.

DEFINITION 2.1. *Let X be a subset of $\mathbb{R}^n$ and $f = (f_1, \dots, f_m)$ be a mapping from X to $\mathbb{R}^m$. For $I = \{ i_1, \dots, i_k \} \subseteq M$ such that $i_1 < \cdots < i_k$, we put $f_I = (f_{i_1}, \dots, f_{i_k})$. The problem of minimizing f is C^r-*simplicial *if there exists a C^r-mapping $\Phi : \Delta^{m-1} \to X^*(f)$ such that both of the restrictions $\Phi|_{\Delta_I} : \Delta_I \to X^*(f_I)$ and $f|_{X^*(f_I)}$ are C^r-diffeomorphisms for any $I \subseteq M$. The problem of minimizing f is C^r-*weakly simplicial *if there exists a C^r-mapping $\phi : \Delta^{m-1} \to f(X^*(f))$ satisfying $\phi(\Delta_I) = f(X^*(f_I))$ for any $I \subseteq M$.*

We will show later an example of weakly simplicial but not simplicial problems in Example 3.4 (see also Figure 3.1).

2.3. Pareto solutions of strongly convex mappings. In this subsection, a characterization of Pareto solutions of strongly convex C^1-mappings is given (see Proposition 2.5). We begin this subsection by quickly reviewing the definition of (strong) convexity. A subset X of $\mathbb{R}^n$ is *convex* if $tx + (1-t)y \in X$ for all $x, y \in X$ and all $t \in [0, 1]$. Let X be a convex set in $\mathbb{R}^n$. A function $f : X \to \mathbb{R}$ is *convex* if

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$$

for all $x, y \in X$ and all $t \in [0, 1]$. A function $f : X \to \mathbb{R}$ is *strongly convex* if there exists $\alpha > 0$ such that

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) - \frac{1}{2}\alpha t(1-t)\|x-y\|^2$$

[3]The usual definition of a C^r-mapping on a manifold with corners is slightly different from that given here: the latter is stronger (as a condition) than the former.

for all $x, y \in X$ and all $t \in [0,1]$, where $\|x-y\|$ denotes the Euclidean norm of $x-y$. A mapping $f = (f_1, \ldots, f_m) : X \to \mathbb{R}^m$ is *(strongly) convex* if every f_i is (strongly) convex. A problem of minimizing a strongly convex mapping is called a *strongly convex problem.* The following are basic properties of (strongly) convex mappings which will be needed later on.

LEMMA 2.2 (see [28, Theorem 2.1.2, p. 54]). *Let $X \subseteq \mathbb{R}^n$ be a convex open subset. A C^1-function $f : X \to \mathbb{R}$ is convex if and only if $f(x) + df_x \cdot (y-x) \le f(y)$ for any $x, y \in X$.*

LEMMA 2.3 (see [28, Theorem 2.1.11, p. 65]). *Let $X \subseteq \mathbb{R}^n$ be a convex open subset. A C^2-function $f : X \to \mathbb{R}$ is strongly convex if and only if there exists $\beta > 0$ such that $m(f)_x \ge \beta$ for any $x \in X$, where $m(f)_x$ is the minimal eigenvalue of the Hessian matrix of f at x.*

LEMMA 2.4 (see [28, Theorem 2.2.6, p. 85]). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strongly convex C^1-function. Then, there exists a unique point such that the function f is minimized.*

In the rest of this subsection we will prove the following proposition.

PROPOSITION 2.5. *Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex C^1-mapping. Then, $x \in X^*(f)$ if and only if there exists $(w_1, \ldots, w_m) \in \Delta^{m-1}$ such that f satisfies the following equivalent conditions:*

1. $\sum_{i=1}^m w_i (df_i)_x = 0.$
2. *The point $x \in \mathbb{R}^n$ is the unique minimizer of the function $\sum_{i=1}^m w_i f_i$.*

For the proof, we prepare some lemmas.

LEMMA 2.6 (see [21, Theorem 3.1.3, p. 79]). *Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a mapping and let $(w_1, \ldots, w_m) \in \Delta^{m-1}$. If $x \in \mathbb{R}^n$ is the unique minimizer of the function $\sum_{i=1}^m w_i f_i$, then $x \in X^*(f)$.*

The following is a special case of the Karush–Kuhn–Tucker necessary condition for Pareto optimality.

LEMMA 2.7 (see [21, Theorem 3.1.5, p. 39]). *Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a C^1-mapping. If $x \in X^*(f)$, then there exists $(w_1, \ldots, w_m) \in \Delta^{m-1}$ satisfying $\sum_{i=1}^m w_i (df_i)_x = 0$.*

LEMMA 2.8. *Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a convex C^1-mapping. Let $(w_1, \ldots, w_m) \in \Delta^{m-1}$. Then, the following conditions for $x \in \mathbb{R}^n$ are equivalent:*

1. $\sum_{i=1}^m w_i (df_i)_x = 0.$
2. *The function $\sum_{i=1}^m w_i f_i : \mathbb{R}^n \to \mathbb{R}$ attains its minimum at x.*

Proof of Lemma 2.8. Set $g = \sum_{i=1}^m w_i f_i$. Then, for any $x \in \mathbb{R}^n$, we have

$$\sum_{i=1}^m w_i (df_i)_x = dg_x. \tag{2.1}$$

Since g is convex, we can deduce from Lemma 2.2 that the following inequality holds for any $y \in \mathbb{R}^n$:

$$g(x) + dg_x \cdot (y-x) \le g(y). \tag{2.2}$$

Suppose that $\sum_{i=1}^m w_i (df_i)_x = 0$. We can easily deduce assertion 2 from (2.1) and (2.2). Suppose that $\sum_{i=1}^m w_i f_i : \mathbb{R}^n \to \mathbb{R}$ attains its minimum at x. Since dg_x is equal to 0 and the equality (2.1) holds, we have assertion 1. ∎

Proof of Proposition 2.5. Suppose that $x \in X^*(f)$. Using Lemma 2.7, we can verify that there exists $(w_1, \dots, w_m) \in \Delta^{m-1}$ satisfying $\sum_{i=1}^m w_i (df_i)_x = 0$. From Lemma 2.8, the point $x \in \mathbb{R}^n$ is a minimizer of $\sum_{i=1}^m w_i f_i$. Since $\sum_{i=1}^m w_i f_i$ is a strongly convex C^1-function, by Lemma 2.4, we have the assertion 2. Finally, suppose assertion 2. Then, from Lemma 2.6, we get $x \in X^*(f)$. ◻

2.4. Fold singularities. In this subsection we will briefly review the definition and basic properties of fold singularities (for details, see [7]). For $0 \le k \le \min\{n, m\}$, we define a subset $S_k \subsetneq J^1(\mathbb{R}^n, \mathbb{R}^m)$ as follows:

$$S_k = \left\{ j^1 g(x) \in J^1(\mathbb{R}^n, \mathbb{R}^m) \;\middle|\; \begin{array}{l} x \in \mathbb{R}^n,\ g : \mathbb{R}^n \to \mathbb{R}^m : C^2\text{-mapping}, \\ \min\{n, m\} - \operatorname{rank}(dg_x) = k \end{array} \right\},$$

where $j^1 g : \mathbb{R}^n \to J^1(\mathbb{R}^n, \mathbb{R}^m)$ is the 1-jet extension of g. Let $g : \mathbb{R}^n \to \mathbb{R}^m$ be a C^2-mapping, $S \subseteq J^1(\mathbb{R}^n, \mathbb{R}^m)$ be a submanifold, and $x \in \mathbb{R}^n$. The mapping $j^1 g$ is *transverse* to S at x if either of the following conditions holds:

- $j^1 g(x)$ is not contained in S,
- $j^1 g(x) \in S$ and $d(j^1 g)_x(T_x\mathbb{R}^n) + T_{j^1 g(x)} S = T_{j^1 g(x)} J^1(\mathbb{R}^n, \mathbb{R}^m)$.

The mapping $j^1 g$ is *transverse* to S if it is transverse to S at any point in $\mathbb{R}^n$.

Suppose that n is greater than or equal to m. For a C^2-mapping $f : \mathbb{R}^n \to \mathbb{R}^m$, we denote the critical point set of f by $\operatorname{Crit}(f) \subseteq \mathbb{R}^n$. A point $x \in \operatorname{Crit}(f)$ is called a *fold* if the following conditions hold:

1. $j^1 f$ is transverse to S_1 at x_0.
2. $T_{x_0} S_1(f) \oplus \ker df_{x_0} = T_{x_0}\mathbb{R}^n$, where $S_1(f) = (j^1 f)^{-1}(S_1)$.

Note that we can easily deduce from condition 2 that the restriction $f|_{\operatorname{Crit}(f)}$ is an immersion around a fold.

Remark 2.9. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a C^∞-mapping and $x \in \operatorname{Crit}(f)$ be a fold. One can take coordinate neighborhoods (U, φ) and (V, ψ) at x and $f(x)$, respectively, so that they satisfy

$$\psi \circ f \circ \varphi^{-1}(x_1, \dots, x_n) = \left(x_1, \dots, x_{m-1}, \sum_{k=m}^{n} \pm x_k^2 \right).$$

In what follows we will give a useful criterion for detecting fold singularities. Let $f = (f_1, \dots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a C^2-mapping and $x_0 \in \operatorname{Crit}(f)$. Suppose that the corank[4] of df_{x_0} is 1 and the matrix $\left(\frac{\partial f_i}{\partial x_j}(x_0)\right)_{1 \le i,j \le m-1}$ is invertible. We define the function $\Lambda_f : \mathbb{R}^n \to \mathbb{R}^{n-m+1}$ as follows:

$$\Lambda_f(x) = (\lambda_1(x), \dots, \lambda_{n-m+1}(x)),$$

where

$$\lambda_i(x) = \det \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(x) & \cdots & \dfrac{\partial f_m}{\partial x_1}(x) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_1}{\partial x_{m-1}}(x) & \cdots & \dfrac{\partial f_m}{\partial x_{m-1}}(x) \\ \dfrac{\partial f_1}{\partial x_{m-1+i}}(x) & \cdots & \dfrac{\partial f_m}{\partial x_{m-1+i}}(x) \end{pmatrix}.$$

[4]For a linear mapping $\varphi : V \to W$ the nonnegative number $\dim W - \operatorname{rank}(\varphi)$ is called the *corank* of φ.

LEMMA 2.10. *Under the situation above, x_0 is a fold if and only if the following conditions hold:*

1. *the differential $(d\Lambda_f)_{x_0}$ has rank $n-m+1$,*
2. $\ker(d\Lambda_f)_{x_0} \oplus \ker df_{x_0} = T_{x_0}\mathbb{R}^n$.

Remark 2.11. The two conditions in Lemma 2.10 are equivalent to those in the original definition above. Indeed, the first condition is equivalent to the condition that $j^1 f$ is transverse to S_1 at x_0, and $T_{x_0}S_1(f)$ is equal to $\ker(d\Lambda_f)_{x_0}$.

3. Proof of the main result. In this section we will show that strongly convex problems are simplicial. Let $f = (f_1, \dots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex C^r-mapping ($2 \le r \le \infty$). Since $\sum_{i=1}^m w_i f_i$ is strongly convex for any $(w_1, \dots, w_m) \in \Delta^{m-1}$, there exists a unique point $x \in \mathbb{R}^n$ such that $\sum_{i=1}^m w_i f_i$ is minimized (see Lemma 2.4). We denote this minimizer by $\arg\min_{x\in\mathbb{R}^n}(\sum_{i=1}^m w_i f_i(x)) \in \mathbb{R}^n$, which is contained in $X^*(f)$ by Lemma 2.6. We can thus define a mapping $x^* : \Delta^{m-1} \to X^*(f)$ as follows:

$$x^*(w) = \arg\min_{x\in\mathbb{R}^n}\left(\sum_{i=1}^m w_i f_i(x)\right).$$

THEOREM 3.1. *Let $f = (f_1, \dots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex C^r-mapping $(2 \le r \le \infty)$. Then the following assertions hold:*

1. *The mapping $x^* : \Delta^{m-1} \to X^*(f)$ (and thus $f \circ x^* : \Delta^{m-1} \to f(X^*(f))$) is a surjective mapping of class C^{r-1}.*
2. *Assume that the corank of df_x is equal to 1 for any $x \in X^*(f)$.*
   A. *The mapping $x^* : \Delta^{m-1} \to X^*(f)$ is a C^{r-1}-diffeomorphism.*
   B. *The restriction $f|_{X^*(f)} : X^*(f) \to \mathbb{R}^m$ is a C^{r-1}-embedding.*

Note that this theorem obviously holds for $m = 1$. For this reason, in the rest of this section we will assume $m \ge 2$.

Theorem 1.1 is an easy consequence of Theorem 3.1, because any subproblem of a strongly convex problem is trivially strongly convex. In particular, by applying Theorem 3.1 to each subproblem, we can show that the image of the restriction x^* on Δ_I is equal to $X^*(f_I)$ for any $I \subseteq M$. Thus a strongly convex problem is weakly simplicial. We can further deduce from assertion 2 of Theorem 3.1 that a strongly convex problem is simplicial under a suitable assumption on the coranks of differentials.

Remark 3.2. The corank assumption in 2 implies that n is greater than or equal to $m-1$. As we will show in the proof, under this assumption any point in $X^*(f)$ for a mapping f is a fold[5] if $n \ge m$.

Remark 3.3. In general, the mapping x^* for a strongly convex problem (without the corank assumption) is not necessarily a diffeomorphism. We will give an explicit example of such a problem with a noninjective x^* in Example 3.4.

Proof of 1 *in Theorem* 3.1. First of all, we can immediately deduce from Proposition 2.5 that x^* is surjective. Let $\delta' > 0$ be a positive number and $g : \Delta^{m-1}_{\delta'} \times \mathbb{R}^n \to \mathbb{R}^n$ be a C^{r-1}-mapping defined by $g(w,x) = \sum_{k=1}^m w_k (df_k)_x$. We can easily deduce from Lemma 2.8 that x^* is an implicit function of the equation $g(w,x) = 0$ defined on Δ^{m-1}. Differentiating g, we have $\frac{\partial g_i}{\partial x_j} = \sum_{k=1}^m w_k \frac{\partial f_k}{\partial x_i \partial x_j}$. Thus the matrix $\left(\frac{\partial g_i}{\partial x_j}\right)_{1\le i,j\le n}$ is equal to $\sum_{k=1}^m w_k H(f_k)_x$, where $H(f_k)_x$ is the Hessian matrix of f_k at x.[6] Since f_k

[5]It is indeed a "definite" fold (for details see [14]).

[6]It is called a generalized Hessian defined in [32] (its explicit expression appears in [34]).

is strongly convex, the Hessian matrix $H(f_k)_x$ is positive definite (see Lemma 2.3). Thus, the matrix $(\frac{\partial g_i}{\partial x_j})_{1\leq i,j\leq n}$ is invertible on Δ^{m-1}. By the implicit function theorem, for any $y \in \Delta^{m-1}$ there exists an open neighborhood $U_y \subseteq \Delta^{m-1}_{\delta'}$ and a (unique) C^{r-1}-mapping $x^*_y : U_y \to \mathbb{R}^n$ such that $x^*_y(y) = x^*(y)$ and $g(w, x^*_y(w)) = 0$ for any $w \in U_y$. We can further deduce from uniqueness of an implicit function that x^*_y coincides with $x^*_{y'}$ on $U_y \cap U_{y'}$ for distinct $y, y' \in \Delta^{m-1}$. Since $U = \bigcup_{y\in\Delta^{m-1}} U_y$ is an open neighborhood of Δ^{m-1}, one can take $\delta < \delta'$ so that Δ^{m-1}_δ is contained in U.[7] We can define $\widetilde{x^*} : \Delta^{m-1}_\delta \to \mathbb{R}^n$ by $\widetilde{x^*}(w) = x^*_y(w)$ for $w \in U_y$. It is easy to see that $\widetilde{x^*}$ is a C^{r-1}-mapping and is an extension of x^*. Thus x^* is a C^{r-1}-mapping. □

Proof of 2.A *in Theorem* 3.1. For $\varepsilon \geq 0$, we define a subset $D^{m-1}_\varepsilon \subsetneq \mathbb{R}^{m-1}$ by

$$D^{m-1}_\varepsilon = \left\{ (z_1, \ldots, z_{m-1}) \in \mathbb{R}^{m-1} \;\middle|\; \sum_i z_i < 1 + \varepsilon,\ z_i > -\varepsilon \right\}.$$

We denote the closure $\overline{D^{m-1}_0}$ by D^{m-1}. It is easy to check that the projection $p : \Delta^{m-1}_\varepsilon \to D^{m-1}_\varepsilon$ defined by $p(w_1, \ldots, w_m) = (w_1, \ldots, w_{m-1})$ is a diffeomorphism. In what follows we will identify Δ^{m-1}_ε with D^{m-1}_ε by p.

Since $\widetilde{x^*}$ constructed in the proof above is an implicit function of the equation $g(w, x) = 0$, the following equality holds:

$$0 = \sum_{k=1}^{m} w_k (df_k)_{\widetilde{x^*}(w)} = \sum_{k=1}^{m-1} z_k (df_k)_{\widetilde{x^*}(z)} + \left(1 - \sum_{k=1}^{m-1} z_k\right)(df_m)_{\widetilde{x^*}(z)},$$

where $(z_1, \ldots, z_{m-1}) \in D^{m-1}_\delta$ is a point corresponding to $w \in \Delta^{m-1}_\delta$. Differentiating both sides of the equation above by z_j ($j = 1, \ldots, m-1$), we obtain the following equality:

$$\begin{aligned} 0 = &\sum_{k=1}^{m-1} z_k \left(H(f_k)_{\widetilde{x^*}(z)} \frac{\partial \widetilde{x^*}}{\partial z_j} \right) + (df_j)_{\widetilde{x^*}(z)} \\ &+ \left(1 - \sum_{k=1}^{m-1} z_k\right)\left(H(f_m)_{\widetilde{x^*}(z)} \frac{\partial \widetilde{x^*}}{\partial z_j} \right) - (df_m)_{\widetilde{x^*}(z)}. \end{aligned}$$

We thus obtain

$$\begin{aligned} \frac{\partial \widetilde{x^*}}{\partial z_j} &= -\left(\sum_{k=1}^{m-1} z_k H(f_k)_{\widetilde{x^*}(z)} + \left(1 - \sum_{k=1}^{m-1} z_k\right) H(f_m)_{\widetilde{x^*}(z)} \right)^{-1} \Big((df_j)_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)} \Big) \\ &= -A(z)\Big((df_j)_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)} \Big), \end{aligned}$$

where we denote the matrix $(\sum_{k=1}^{m-1} z_k H(f_k)_{\widetilde{x^*}(z)} + (1 - \sum_{k=1}^{m-1} z_k) H(f_m)_{\widetilde{x^*}(z)})^{-1}$ by $A(z)$, which is positive definite for $z \in D^{m-1}$. Since D^{m-1} is compact and the corank of $df_{\widetilde{x^*}(z)}$ is equal to 1 for any $z \in D^{m-1}$, by retaking δ if necessary, we can assume that the corank of $df_{\widetilde{x^*}(z)}$ is equal to 1 and $A(z)$ is positive definite, and thus invertible,

[7]If not, we can take $x_n \in U^c \cap \Delta^{m-1}_{1/n}$ for any $n \in \mathbb{N}$. Since $\overline{\Delta^{m-1}_1}$ is compact, $\{ x_n \}_{n\in\mathbb{N}}$ has a cluster point x, which is contained in Δ^{m-1}. However, x is also contained in U^c since it is closed, contradicting the fact that $\Delta^{m-1} \subsetneq U$.

for any $z \in D_\delta^{m-1}$. (Note that the condition $\operatorname{corank}(df_x) = 1$ is an open condition in $\operatorname{Crit}(f)$.)

We will show that the matrix

$$\text{(3.1)} \qquad \Big((df_1)_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)} \quad \cdots \quad (df_{m-1})_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)} \Big)$$

has rank $m-1$ for any $z \in D^{m-1}$. If not, there exists $a_i \in \mathbb{R}$ with $(a_1, \ldots, a_{m-1}) \neq 0$ such that

$$\sum_{i=1}^{m-1} a_i \Big((df_i)_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)} \Big) = 0.$$

On the other hand, by the definition of $\widetilde{x^*}$, we obtain

$$\sum_{j=1}^{m-1} z_j (df_j)_{\widetilde{x^*}(z)} + \left(1 - \sum_{i=1}^{m-1} z_i \right) (df_m)_{\widetilde{x^*}(z)} = 0.$$

Thus, two vectors $(z_1, \ldots, z_{m-1}, 1 - \sum_{i=1}^{m-1} z_i)$ and $(a_1, \ldots, a_{m-1}, -\sum_{i=1}^{m-1} a_i)$ are contained in $\ker df_{\widetilde{x^*}(z)}$. However, these are linearly independent and contradict the assumption $\operatorname{corank}(df_{\widetilde{x^*}(z)}) = 1$. Therefore, the matrix in (3.1) has rank $m-1$ for any $z \in D^{m-1}$. Since the condition that the matrix in (3.1) has rank $m-1$ is an open condition for z, we can assume that this condition holds for any D_δ^{m-1} by making δ sufficiently small. The differential

$$d\widetilde{x^*}_z = -A(z) \Big((df_1)_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)}, \ldots, (df_{m-1})_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)} \Big)$$

also has rank $m-1$ since $A(z)$ is invertible for any $z \in D_\delta^{m-1}$.

We next show that the mapping x^* is injective. Assume that $x^*(w)$ is equal to $x^*(w')$ for $w, w' \in \Delta^{m-1}$. Since the corank of $df_{x^*(w)}$ is 1 and $\sum_{j=1}^m w_j (df_j)_{x^*(w)} = 0$, we can obtain $\operatorname{Im}(df_{x^*(w)}) = \langle w \rangle^\perp$. In the same way, we can also prove that $\operatorname{Im}(df_{x^*(w')})$ is equal to $\langle w' \rangle^\perp$. From the assumption, we can deduce that $\langle w \rangle^\perp$ is equal to $\langle w' \rangle^\perp$ and thus $w = w'$.

We have shown that x^* is an injective immersion. Since Δ^{m-1} is compact, x^* is a homeomorphism and thus a diffeomorphism to its image, which is equal to $X^*(f)$. ∎

Proof of 2.B *in Theorem* 3.1. We first prove that $f|_{X^*(f)}$ is injective. Let $w, z \in \Delta^{m-1}$, $x = x^*(w)$, and $y = x^*(z)$. Suppose that $f(x)$ is equal to $f(y)$. Then $(\sum_{i=1}^m w_i f_i)(x)$ is also equal to $(\sum_{i=1}^m w_i f_i)(y)$. Since the function $\sum_{i=1}^m w_i f_i$ is strongly convex, the point minimizing $\sum_{i=1}^m w_i f_i$ is unique (see Lemma 2.4). Thus, x is equal to y.

As we noted in Remark 3.2, the corank assumption implies that n is greater than or equal to $m-1$. If $n = m-1$, this assumption further implies that f is an immersion at any point in $X^*(f)$. The restriction $f|_{X^*(f)}$ is thus an embedding since any injective immersion on a compact manifold is an embedding. In what follows we will assume $n \geq m$.

We next show that any point $x \in X^*(f)$ is a fold of f. The following transformations preserve strong convexity of f:

- $(f_1, \ldots, f_m) \mapsto (f_{\sigma(1)}, \ldots, f_{\sigma(m)}) \quad (\sigma \in \mathfrak{S}_m)$,
- $(f_1, \ldots, f_m) \mapsto (f_1, \ldots, f_m + \alpha f_i) \quad (\alpha > 0, i = 1, \ldots, m-1)$, and
- linear transformations of the source of f,

where $\mathfrak{S}_m$ is the symmetric group of degree m. By applying these transformations if necessary, we can assume the following:

$$df_x = \begin{array}{c} \overbrace{\hphantom{I_{m-1} \quad 0 \cdots 0 \quad 0}}^{m} \\ \left(\begin{array}{ccc|c} & & & 0 \\ & I_{m-1} & & \vdots \\ & & & 0 \\ \hline 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{array}\right) \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\ \vdots\\0\end{array}}\right\} m-1 \\ \left.\vphantom{\begin{array}{c}0\\ \vdots\\0\end{array}}\right\} n-m+1 \end{array}$$

Let $\Lambda_f : \mathbb{R}^n \to \mathbb{R}^{n-m+1}$ be the mapping defined in subsection 2.4. By Lemma 2.10, it suffices to show the following:
(A) the rank of $(d\Lambda_f)_x$ is $n - m + 1$,
(B) $\ker(d\Lambda_f)_x \oplus \ker df_x = T_x\mathbb{R}^n$.
By the assumptions above, we can calculate $(d\Lambda_f)_x$ as follows:

$$\begin{aligned}(d\Lambda_f)_x &= \pm\left(\frac{\partial^2 f_m}{\partial x_j \partial x_{m-1+i}}(x)\right)_{1\le i\le n-m+1, 1\le j\le n} \\ &= \pm\left(\; 0_{(n-m+1)\times(m-1)} \; \middle| \; I_{n-m+1} \; \right) H(f_m)_x.\end{aligned}$$

Since f_m is strongly convex, the Hessian matrix $H(f_m)_x$ is positive definite, in particular invertible by Lemma 2.3. Thus, condition (A) holds. The above calculation also implies the following equality:

$$\ker(d\Lambda_f)_x = \left\langle H(f_m)_x^{-1} e_1, \ldots, H(f_m)_x^{-1} e_{m-1} \right\rangle.$$

Let $v \in \ker(d\Lambda_f)_x \cap \ker(df)_x$. From the equality above, we can find $w \in \mathbb{R}^{m-1} \times \{ 0 \} \subsetneq \mathbb{R}^m$ such that $v = H(f_m)_x^{-1} w$. Thus the following holds:

$$0 = df_x(v) = \begin{pmatrix} I_{m-1} & 0 \\ 0 & 0 \end{pmatrix} H(f_m)_x^{-1} w.$$

We can deduce the following from this equality:

$${}^t w H(f_m)_x^{-1} w = 0.$$

Since $H(f_m)$ is positive definite by Lemma 2.3, w is equal to 0. Since the corank of df_x is equal to 1, we can deduce the following from condition (A):

$$\dim \ker(d\Lambda_f)_x + \dim \ker(df_x) = n.$$

Thus condition (B) also holds. We can eventually conclude that $f|_{X^*(f)}$ is an immersion. ∎

3.1. Examples. One of the simplest and most representative instances of strongly convex problem is the multiobjective location problem under the Euclidean norm. It is well-known that the Pareto set (resp., the Pareto front) of this problem is the convex hull of minimizers (resp., their values) of individual objective functions [16]. Thus, if these minimizers are in general position, then the convex hull becomes a simplex and this problem is a C^0-simplicial problem.

In this section we will show that in the strongly convex case, the condition that minimizers are in general position is no longer necessary nor sufficient to ensure C^r-simpliciality, and the corank assumption is still essential to determine the topology of the Pareto set and the Pareto front. To this end, we will give two examples of strongly convex mappings from $\mathbb{R}^3$ to $\mathbb{R}^3$ and discuss the configurations of Pareto sets of them. As we mentioned in the beginning of section 3, for any strongly convex mapping $f : \mathbb{R}^3 \to \mathbb{R}^3$ we can define a mapping $x^* : \Delta^2 \to X^*(f)$. Example 3.4 has a corank 2 differential at a point in the Pareto set, and the corresponding x^* is not a diffeomorphism (despite the fact that the minimizers of the three component functions are in general position). This example implies that we cannot drop the corank assumption in assertion 2 of Theorem 3.1. In Example 3.5, the minimizers of the three component functions are not in general position. Nevertheless, the corank assumption is fulfilled and (1.1) defines a diffeomorphism between Δ^2 and the Pareto set.

Example 3.4 (general position with corank 2). Define a mapping $f = (f_1, f_2, f_3) : \mathbb{R}^3 \to \mathbb{R}^3$ as follows:

$$\begin{aligned} f_1(x, y, z) &= x^2 + y^2 + z^2, \\ f_2(x, y, z) &= x + y + x^2 + y^2 + z^2, \\ f_3(x, y, z) &= -(x + y) + x^2 + 2y^2 + z^2. \end{aligned}$$

The mapping f is strongly convex. We will check that $X^*(f)$ contains a singularity of corank 2 and is not diffeomorphic to Δ^2. The differentials at $p = (x, y, z) \in \mathbb{R}^3$ are

$$\begin{aligned} df_{1,p} &= (2x, 2y, 2z), \\ df_{2,p} &= (1 + 2x, 1 + 2y, 2z), \\ df_{3,p} &= (-1 + 2x, -1 + 4y, 2z), \end{aligned}$$

and thus the corank of $df_{\mathbf{0}}$ is 2. Since f is strongly convex, the mapping $x^* : \Delta^2 \to X^*(f)$ is surjective by assertion 1 of Theorem 3.1. Regarding Δ^2 as $D^2 = \{ (w_2, w_3) \mid w_2, w_3 \geq 0, w_2 + w_3 \leq 1 \}$, we obtain

$$x^*(w_2, w_3) = \left(-\frac{w_2 - w_3}{2}, -\frac{w_2 - w_3}{2(1 + w_3)}, 0 \right).$$

Obviously x^* maps the line defined by $w_2 - w_3 = 0$ in $\Delta^2 = D^2$ into a single point (the origin), while it is injective at points outside the above line. Thus $X^*(f)(= x^*(\Delta^2))$ is not diffeomorphic to Δ^2. Figure 3.1 describes the Pareto set of f, together with the contours of the functions f_1 (red), f_2 (blue), and f_3 (green).

For $\varepsilon \in \mathbb{R} - \{ 0 \}$, we define another mapping $h_\varepsilon : \mathbb{R}^3 \to \mathbb{R}^3$ as follows:

$$h_\varepsilon(x, y, z) = (f_1(x, y, z) + \varepsilon z, f_2(x, y, z), f_3(x, y, z)).$$

Note that the mapping h_ε is a linear perturbation of f. It is easy to verify that the mapping h_ε is strongly convex and never has corank 2 critical points. Thus the problem of minimizing h_ε is simplicial. The Pareto sets of the problems for various ε's are shown in Figure 3.2. They are indeed homeomorphic to the 2-simplex. We will see in section 4 that in general any strongly convex problem becomes simplicial after a generic linear perturbation (see Theorem 4.1).

Example 3.5 (nongeneral position without corank 2). We define a mapping $f = (f_1, f_2, f_3) : \mathbb{R}^3 \to \mathbb{R}^3$ as follows:

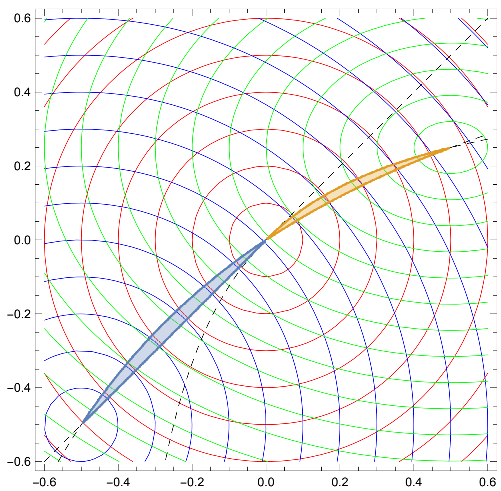
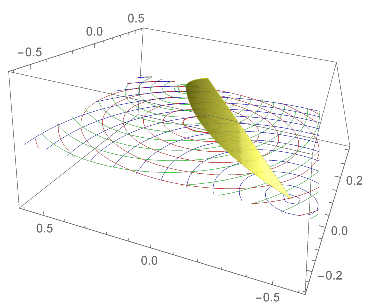
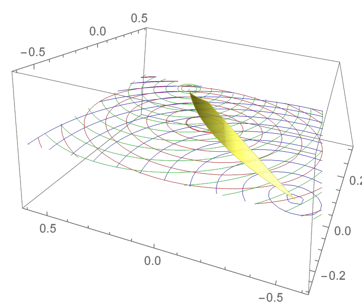


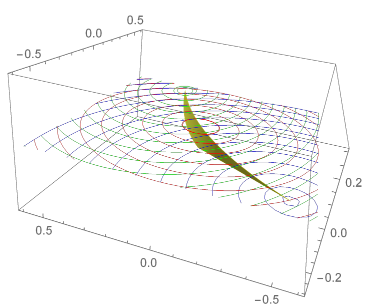
Fig. 3.1. *The Pareto set of f. The union of two domains colored orange and blue is the Pareto set. Color available online.*



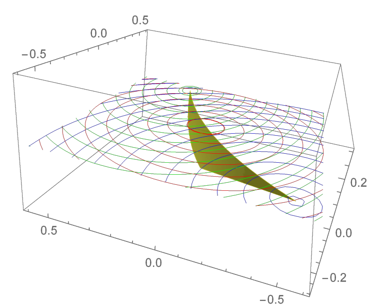
(a) $\varepsilon = -0.5$.



(b) $\varepsilon = -0.25$.



(c) $\varepsilon = 0.25$.



(d) $\varepsilon = 0.5$.

Fig. 3.2. *The Pareto sets of the problems minimizing h_ε for several ε's. Color available online.*

$$
\begin{aligned}
f_1(x,y,z) &= x^2 + (-y+x)^2 + z^2,\\
f_2(x,y,z) &= 2(x-1)^2 + (-y+x-1)^2 + z^2,\\
f_3(x,y,z) &= (x-2)^2 + (y+x-2)^2 + z^2.
\end{aligned}
$$

The mapping f is strongly convex. The minimizers of the functions f_1, f_2, and f_3 are respectively $(0,0,0), (1,0,0)$, and $(2,0,0)$, which are not in general position. Nevertheless, the corank of the differential df_p is equal to 1 for any $p \in X^*(f)$, and thus the Pareto set of the problem minimizing f is diffeomorphic to the 2-simplex by Theorem 3.1.

It is easy to see that the z-coordinate of the point in $X^*(f)$ is equal to 0. We will first show that the corank of df_p is 1 on the xy-plane except for a single point, and then check that this exceptional point is not in $X^*(f)$. The differentials at $p = (x, y, 0)$ are calculated as follows:

$$\begin{aligned}
(df_1)_p &= (4x - 2y, -2x + 2y, 0),\\
(df_2)_p &= (6x - 2y - 6, -2x + 2y + 2, 0),\\
(df_3)_p &= (4x + 2y - 8, 2x + 2y - 4, 0).
\end{aligned}$$

The minors of the Jacobian matrix Jf_p are then calculated as follows:

$$\det\begin{pmatrix} 4x - 2y & -2x + 2y \\ 6x - 2y - 6 & -2x + 2y + 2 \end{pmatrix} = 4\left(x - \frac{y}{2} - \frac{1}{2}\right)^2 - (y-3)^2 + 8, \tag{3.2}$$

$$\det\begin{pmatrix} 4x - 2y & -2x + 2y \\ 4x + 2y - 8 & 2x + 2y - 4 \end{pmatrix} = 2\left(8(x-1)^2 - 4\left(y - \frac{3}{2}\right)^2 + 1\right), \tag{3.3}$$

$$\det\begin{pmatrix} 6x - 2y - 6 & -2x + 2y + 2 \\ 4x + 2y - 8 & 2x + 2y - 4 \end{pmatrix} = 5\left(2x + \frac{y}{5} - 3\right)^2 - \frac{41}{5}\left(y - \frac{35}{41}\right)^2 + \frac{40}{41}. \tag{3.4}$$

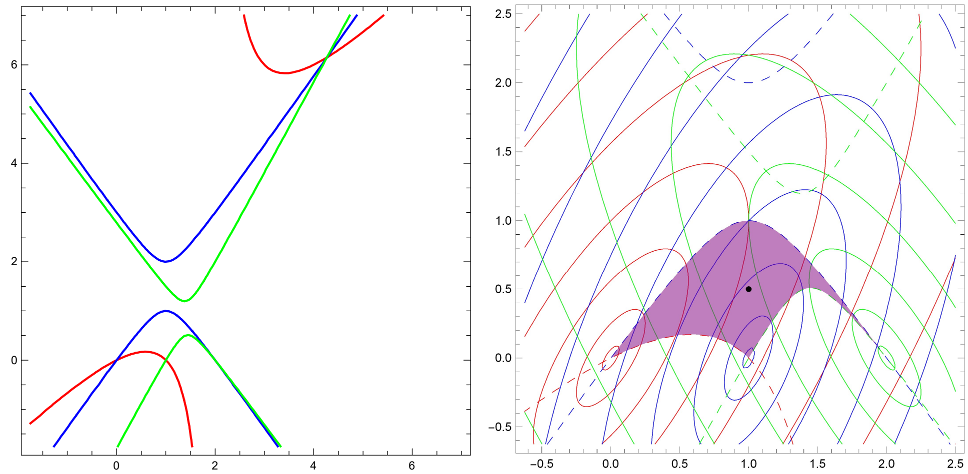
The zero-sets of these polynomials are hyperbolas shown in Figure 3.3(a): the red, blue, and green curves are the zero-sets of the polynomials (3.2), (3.3), and (3.4), respectively. As this figure indicates, there is a unique intersection point of the three hyperbolas, which we denote by (x', y'). This observation implies that the corank of df_p is equal to 1 for any $p = (x, y, 0) \neq (x', y', 0)$. Since $y' - x' > 1$ and $x', y' > 1$, all three values $-2x' + 2y', -2x' + 2y' + 2$, and $2x' + 2y' - 4$ are greater than 0. We can deduce from Proposition 2.5 that $(x', y', 0)$ is not in $X^*(f)$. Hence there is no point $p \in X^*(f)$ with $\operatorname{corank}(df_p) \geq 2$.

We next determine the Pareto set $X^*(f)$. Since the corank of df_p is 1 for any $p = (x, y, 0) \neq (x', y', 0)$, there exist unique numbers $a(p), b(p), c(p) \in \mathbb{R}$ satisfying the following conditions:

- $a(p)(df_1)_p + b(p)(df_2)_p + c(p)(df_3)_p = 0$,
- $a(p) + b(p) + c(p) = 1$.

By Proposition 2.5, the Pareto set $X^*(f)$ is the set of points p with $a(p), b(p), c(p) \geq 0$. Since one of the numbers $a(p), b(p), c(p)$ is equal to 0 on the unions of three hyperbolas in Figure 3.3(a) and $a(p), b(p), c(p)$ depend continuously on p, $X^*(f)$ is one of the connected components of the complement of the three hyperbolas. One can easily check that $(a(p_0), b(p_0), c(p_0))$ is equal to $(1/3, 1/2, 1/6)$, where $p_0 = (1, 1/2, 0)$. Thus, the Pareto set $X^*(f)$ is the component containing p_0, which is shown in Figure 3.3(b): the purple region is $X^*(f)$, while the black point is p_0.

4. Generic linear perturbations of strongly convex mappings. In this section, we will investigate the multiobjective optimization problem of minimizing a generically linearly perturbed strongly convex mapping. Let $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ be the space consisting of all linear mappings from $\mathbb{R}^n$ into $\mathbb{R}^m$. In what follows we will regard $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ as the Euclidean space $(\mathbb{R}^n)^m$ in the obvious way. The purpose of this section is to show the following theorem.



(a) The three hyperbolas defined by (3.2), (3.3) and (3.4).

(b) The Pareto set of f (projected on the xy-plane).

Fig. 3.3. *The dashed curves in* (b) *are parts of the hyperbolas in* (a)*. Color available online.*

Theorem 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ $(n \geq m)$ be a strongly convex C^r-mapping $(2 \leq r \leq \infty)$. If $n - 2m + 4 > 0$, then there exists a subset $\Sigma \subsetneq \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with Lebesgue measure zero such that for any $\pi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) - \Sigma$ the mapping $f + \pi$ never has differential with corank greater than 1 on its Pareto set. In particular, the multiobjective optimization problem of minimizing $f + \pi : \mathbb{R}^n \to \mathbb{R}^m$ is C^{r-1}-simplicial.*

As we mentioned in section 1, we would like to emphasize that one *cannot* deduce Theorem 4.1 immediately from the preceding results on genericity of problems. Smale's genericity result and other results on genericities ([1, 2], for example) are in the mapping space with the Whitney C^r-topology $(2 \leq r \leq \infty)$, based on Thom's transversality theorem. In [19], Lucchetti and Miglierina consider some stabilities on the topology of continuous convergence. On the other hand, Ichiki discussed in [12] some transverality theorems on generic linear perturbations, which are different from Thom's. Based on this result, we consider a different version of genericity, which is considered in the linear mapping space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \simeq (\mathbb{R}^n)^m$ with the Euclidean topology. Note that the following inclusions hold:

- $\boxed{\begin{array}{c}\text{Euclidean}\\ \text{topology}\end{array}} \subsetneq \boxed{\begin{array}{c}\text{Topology of}\\ \text{continuous convergence}\\ \text{relative to } C^r(\mathbb{R}^n, \mathbb{R}^m)\end{array}} = \boxed{\begin{array}{c}\text{Discrete}\\ \text{topology}\end{array}}$ in $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$,
- $\boxed{\begin{array}{c}\text{Topology of}\\ \text{continuous convergence}\end{array}} \subsetneq \boxed{\begin{array}{c}\text{Whitney } C^r\text{-topology}\\ (0 \leq r \leq \infty)\end{array}}$ in $C^r(\mathbb{R}^n, \mathbb{R}^m)$.

Thus, a sequence of linearly perturbed mappings $\{ f + \pi_i : \mathbb{R}^n \to \mathbb{R}^m \}_{i=1}^{\infty}$ in general does not converge under the topology of continuous convergence. (Indeed, as shown above, if we induce the relative topology of continuous convergence into the space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, it becomes the discrete topology.)

In order to prove Theorem 4.1, we first observe that strong convexity is preserved under linear perturbations.

LEMMA 4.2. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex mapping. Then, for any $\pi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, the mapping $f + \pi : \mathbb{R}^n \to \mathbb{R}^m$ is also a strongly convex mapping.*

Proof of Lemma 4.2. Obviously, it is sufficient to show the statement under the assumption that f is a function (i.e., $m = 1$). For $x, y \in \mathbb{R}^n$ and $t \in [0,1]$, the following holds:

$$\begin{aligned}
&t((f+\pi)(x)) + (1-t)((f+\pi)(y)) - (f+\pi)(tx+(1-t)y)\\
&= t(f(x)+\pi(x)) + (1-t)(f(y)+\pi(y)) - f(tx+(1-t)y) - \pi(tx+(1-t)y)\\
&= tf(x) + (1-t)f(y) - f(tx+(1-t)y),
\end{aligned}$$

where the last equality holds since π is linear. Since f is strongly convex, there exists $\alpha > 0$ satisfying the following inequality for any $x, y \in \mathbb{R}^n$ and $t \in [0,1]$:

$$tf(x) + (1-t)f(y) - f(tx+(1-t)y) \geq \frac{1}{2}\alpha t(1-t)\|x-y\|^2.$$

Hence, the mapping $f + \pi$ is also strongly convex. $\square$

Before proving Theorem 4.1, we will briefly review the result in [12] needed here. Let $S_k \subsetneq J^1(\mathbb{R}^n, \mathbb{R}^m)$ be the subset defined in subsection 2.4. It is known that S_k is a submanifold of $J^1(\mathbb{R}^n, \mathbb{R}^m)$ satisfying the following (see [7, 43]):

$$\operatorname{codim} S_k = (n - v + k)(m - v + k),$$

where $\operatorname{codim} S_k = \dim J^1(\mathbb{R}^n, \mathbb{R}^m) - \dim S_k$ and $v = \min\{\, n, m \,\}$. The following lemma is merely a special case of [12, Theorem 1].

LEMMA 4.3 (cf. [12]). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a C^r-mapping. Let k be an integer satisfying $1 \leq k \leq \min\{\, n, m \,\}$. If $r > \max\{\, n - \operatorname{codim} S_k, 0 \,\} + 1$, then there exists a subset $\Sigma \subsetneq \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with Lebesgue measure zero such that for any $\pi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) - \Sigma$, the mapping $j^1(f + \pi) : \mathbb{R}^n \to J^1(\mathbb{R}^n, \mathbb{R}^m)$ is transverse to S_k.*

Proof of Theorem 4.1. In the case $m = 1$, it is clearly seen that Theorem 4.1 holds. Hence, we will consider the case $m \geq 2$. Since $n \geq m$, the codimension of S_2 is equal to $2(n - m + 2)$. By the assumption $n - 2m + 4 > 0$, we can obtain the inequality $\operatorname{codim} S_2 > n$. Let k be an integer with $2 \leq k \leq m$. It follows that

$$n - \operatorname{codim} S_k \leq n - \operatorname{codim} S_2 < 0.$$

In particular, for a mapping $g : \mathbb{R}^n \to \mathbb{R}^m$, transversality of $j^1 g$ to S_k is equivalent to the condition that $j^1 g(\mathbb{R}^n) \cap S_k = \emptyset$, that is, g has no corank k critical points (see [7, Chapter II, Proposition 4.2]). Furthermore, the following inequality holds:

$$r \geq 2 > \max\{\, n - \operatorname{codim} S_k, 0 \,\} + 1. \qquad \square$$

We can deduce from Lemma 4.3, together with the observations above, that there exists $\Sigma_k \subsetneq \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with Lebesgue measure zero such that the mapping $f + \pi$ has no corank k critical points for any $\pi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) - \Sigma_k$. Set $\Sigma = \bigcup_{l=2}^m \Sigma_l \subsetneq \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, which also has Lebesgue measure zero. Last, we can easily verify that Σ satisfies the conditions in Theorem 4.1.

Remark 4.4. We cannot drop the assumption $n - 2m + 4 > 0$ in Theorem 4.1. Let $G(x) = (g_1(x) - x_3, g_2(x) - x_4, g_1(x) + x_3, g_2(x) + x_4)$ be a map from $\mathbb{R}^4$ to $\mathbb{R}^4$, where

$$g_1(x) = x_1^2 + x_3x_2 + \frac{1}{2}\left(x_2^2 + x_4x_1\right) + x_3^2 + x_4^2,$$
$$g_2(x) = x_2^2 + x_4x_1 + \frac{1}{2}\left(x_1^2 + x_3x_2\right) + x_3^2 + x_4^2.$$

The mapping G is strongly convex and has a corank 2 critical point at the origin. In what follows, we will see that the Pareto set of any small linear perturbation of G contains a corank 2 critical point. (Note that the inequality $n - 2m + 4 > 0$ does not hold when $n = m = 4$.)

The Jacobi matrix of G at the origin is $dG_0 = \left(\begin{smallmatrix} O & O \\ -I & I \end{smallmatrix}\right)$, where I and O are 2×2 unit matrix and zero matrix, respectively. Define the cokernel of dG_0 as

$$\operatorname{coker} dG_0 = \left\{ (v_1, v_2, v_3, v_4) \in \mathbb{R}^4 \;\middle|\; \sum_{i=1}^{4} v_i (dG_i)_0 = 0 \right\};$$

then, the subspace $\operatorname{coker} dG_0$ is equal to $\langle (1,0,1,0),\ (0,1,0,1) \rangle$. In particular, the subspace $\operatorname{coker} dG_0$ intersects with the interior of Δ^3.

For $(x, \pi) \in \mathbb{R}^4 \times \mathcal{L}(\mathbb{R}^4, \mathbb{R}^4)$, put $d(G + \pi)_x = \left(\begin{smallmatrix} A(x,\pi) & B(x,\pi) \\ C(x,\pi) & D(x,\pi) \end{smallmatrix}\right)$, where $A(x, \pi)$, $B(x, \pi)$, $C(x, \pi)$, and $D(x, \pi)$ are 2×2 matrices. We can take an open neighborhood U of $(0, 0) \in \mathbb{R}^4 \times \mathcal{L}(\mathbb{R}^4, \mathbb{R}^4)$ so that the matrix $D(x, \pi)$ is invertible for any $(x, \pi) \in U$. By multiplying the matrix $\left(\begin{smallmatrix} I & O \\ -D^{-1}(x,\pi)C(x,\pi) & I \end{smallmatrix}\right)$ to $d(G + \pi)_x$ from the right, we obtain the matrix $\left(\begin{smallmatrix} A(x,\pi) - B(x,\pi)D^{-1}(x,\pi)C(x,\pi) & B(x,\pi) \\ O & D(x,\pi) \end{smallmatrix}\right)$. Put $E(x, \pi) = A(x, \pi) - B(x, \pi)D^{-1}(x, \pi)C(x, \pi)$. The matrix $E(0, 0)$ is the zero matrix, and

$$\left.\frac{\partial E(x,\pi)}{\partial x_1}\right|_{(x,\pi)=(0,0)} = \begin{pmatrix} 4 & 2 \\ 0 & 0 \end{pmatrix}, \quad \left.\frac{\partial E(x,\pi)}{\partial x_2}\right|_{(x,\pi)=(0,0)} = \begin{pmatrix} 0 & 0 \\ 2 & 4 \end{pmatrix},$$
$$\left.\frac{\partial E(x,\pi)}{\partial x_3}\right|_{(x,\pi)=(0,0)} = \begin{pmatrix} 0 & 0 \\ 2 & 1 \end{pmatrix}, \quad \left.\frac{\partial E(x,\pi)}{\partial x_4}\right|_{(x,\pi)=(0,0)} = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$$

hold. These equalities imply that the rank of the Jacobi matrix of E with respect to x is 4. Therefore, applying the implicit function theorem we can obtain $\hat{x} : V \to \mathbb{R}^4$, where V is an open neighborhood of $0 \in \mathcal{L}(\mathbb{R}^4, \mathbb{R}^4)$, such that the equality $E(\hat{x}(\pi), \pi) = O$ holds for any $\pi \in V$. Note that $\hat{x}(\pi)$ is continuous with respect to π. Then, we can define a continuous mapping $\operatorname{coker} d(G + \cdot)_{\hat{x}(\cdot)} : V \to Gr(2, \mathbb{R}^4)$ where $Gr(2, \mathbb{R}^4)$ is all two-dimensional linear subspaces of $\mathbb{R}^4$. Since $\operatorname{coker} dG_0$ intersects with the interior of Δ^3, so does $\operatorname{coker} d(G + \pi)_{\hat{x}(\pi)}$ if π is sufficiently small. This proves that $\hat{x}(\pi)$ is a Pareto solution to $G + \pi$ with corank 2 differential for a sufficiently small π.

Remark 4.5. Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex C^r-mapping ($2 \le r \le \infty$). We can deduce from Lemma 4.2 that the mapping $\sum_{i=1}^{m} w_i(f_i + \pi_i)$ is strongly convex for any $(w, \pi) \in \Delta^{m-1} \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, where $\pi = (\pi_1, \ldots, \pi_m)$. By the same argument as in section 3, we can define a mapping $\Gamma : \Delta^{m-1} \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}^n$ as follows:

$$\Gamma(w, \pi) = \arg\min_{x \in \mathbb{R}^n} \left(\sum_{i=1}^{m} w_i (f_i + \pi_i)(x) \right).$$

Then the mapping Γ is continuous, and thus for any $\varepsilon > 0$ there exists an open neighborhood U of $0 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ satisfying the following inequality for any $(w, \pi) \in \Delta^{m-1} \times U$:

$$\|\Gamma(w, \pi) - \Gamma(w, 0)\| < \varepsilon.$$

This inequality implies that a small linear perturbation of a strongly convex problem does not cause substantial changes of the Pareto set. For the proof of the statement here, see Appendix A.

5. Applications.

As we have seen, strongly convex problems have a variety of desirable properties which make their Pareto sets easy to understand. Although many practical problems are *not* necessarily strongly convex, we can apply suitable "structure-preserving" transformations to some of these problems so that they become strongly convex. In this section, we will give several examples of such problems.

5.1. Location problems.

One of the most traditional examples is the location problem, which requires finding the best place $x \in \mathbb{R}^n$ for a facility so that the weighted-sum $\sum_{i=1}^m w_i \|x - p_i\|$ (for given $w \in \Delta^{m-1}$) of distances from demand points $p_1, \ldots, p_m \in \mathbb{R}^n$ is minimized. Its multiobjective version [16] is the following problem:[8]

$$\begin{aligned} &\text{minimize } f(x) = (f_1(x), \ldots, f_m(x)) \text{ subject to } x \in \mathbb{R}^n, \\ &\quad \text{where } f_i(x) = \|x - p_i\| \quad (i = 1, \ldots, m). \end{aligned} \tag{5.1}$$

The mapping f is called a *distance mapping* [13], which is not differentiable. Each f_i is convex but not strongly convex, and thus so is the problem (5.1).

Let us consider the transformation of the target $T : [0, \infty)^m \to [0, \infty)^m$ defined by $T(y_1, \ldots, y_m) = (y_1^2, \ldots, y_m^2)$, which preserves the Pareto ordering of $[0, \infty)^m$. We have a transformed problem

$$\text{minimize } T \circ f(x) \text{ subject to } x \in \mathbb{R}^n. \tag{5.2}$$

The mapping $T \circ f$ (called a *distance-squared mapping* [13]) is differentiable and strongly convex; in particular the problem (5.2) is strongly convex. Since T preserves the Pareto ordering, the Pareto sets of (5.1) and (5.2) are identical and the Pareto fronts are homeomorphic.

For the original problem (5.1), the weighted-sum scalarization may have non-unique solutions (e.g., the problem of minimizing $\sum_{i=1}^m w_i \|x - p_i\|$ with $w_1 = w_2 = 1/2$, $w_3 = \cdots = w_m = 0$ has solutions $tp_1 + (1-t)p_2$ for $t \in [0, 1]$). Thus we cannot define a mapping x^* given in section 3. On the other hand, for the transformed problem (5.2), every scalarized problem has a unique solution. Since x^* is well-defined in this case, the set of such solutions coincides with the entire Pareto set by assertion 1 of Theorem 3.1. It is further easy to verify that the corank of $d(T \circ f)_x$ is 1 for any $x \in X^*(T \circ f)$, provided that $p_1, \ldots, p_m$ are in general position (which requires $n \geq m - 1$). Thus, the assertion 2 of Theorem 3.1 guarantees that the problem (5.2) is (C^∞-)simplicial. Since T preserves the Pareto ordering, one can easily see that problem (5.1) is also (C^0-)simplicial.

5.2. Phenotypic divergence modeling.

Another example of minimizing distances from points arises in evolutionary biology. Let A_i be a symmetric, positive definite matrix of size n and let $p_i \in \mathbb{R}^n$ $(i = 1, \ldots, m)$. Shoval et al. [30] provided

[8] While Kuhn [16] originally considered a planar case ($n = 2$), we consider the problem in general dimension.

a model for describing phenotypic divergence of species, which is an extension of the location problem:

$$\begin{aligned} &\text{minimize } f(x) = (f_1(x), \dots, f_m(x)) \text{ subject to } x \in \mathbb{R}^n, \\ &\quad \text{where } f_i(x) = \|A_i(x - p_i)\| \quad (i = 1, \dots, m). \end{aligned} \tag{5.3}$$

As before, the problem of minimizing (5.3) is convex but not strongly convex. We can again apply the transformation T used in the previous subsection and obtain

$$\text{minimize } T \circ f(x) \text{ subject to } x \in \mathbb{R}^n. \tag{5.4}$$

Since affine transformations of the source space preserve the strong convexity of a problem, each component of $T \circ f$ (and thus problem (5.4)) is strongly convex. Applying assertion 1 of Theorem 3.1, we can conclude that both problems (5.3) and (5.4) are weakly simplicial. In order to further show that these problems are simplicial, we have to check the corank condition in assertion 2 of Theorem 3.1, which would be a hard task, even if the demand points $p_1, \dots, p_m$ are in general position. Indeed, problems appearing in subsection 3.1 are special cases of the problems we are dealing with here. As discussed in subsection 3.1, generality of the configuration of demand points does not necessarily imply the corank condition.

5.3. Ridge regression. The ridge regression [11] can be reformulated as a multiobjective strongly convex problem. Let us consider a linear regression model

$$y = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p + \varepsilon,$$

where $x_1, \dots, x_p$ are predictor variables, y is a response variable, ε is a Gaussian random variable expressing noise, and $\theta_1, \dots, \theta_p$ are the predictors' coefficients to be estimated from observations. Given n observations of $x_1, \dots, x_p$ and y, which are denoted by $\overline{x}_{i1}, \dots, \overline{x}_{ip}$ and $\overline{y}_i$ $(i = 1, \dots, n)$, an observation matrix and a response vector are formed as

$$\overline{X} = \begin{pmatrix} \overline{x}_{11} & \dots & \overline{x}_{1p} \\ \vdots & \ddots & \vdots \\ \overline{x}_{n1} & \dots & \overline{x}_{np} \end{pmatrix}, \quad \overline{y} = \begin{pmatrix} \overline{y}_1 \\ \vdots \\ \overline{y}_n \end{pmatrix}.$$

The ridge regressor is the solution to the following problem:

$$\text{minimize } g^\lambda(\theta) = \left\|\overline{X}\theta - \overline{y}\right\|^2 + \lambda\|\theta\|^2 \text{ subject to } \theta \in \mathbb{R}^p, \tag{5.5}$$

where $\|\cdot\|$ is the Euclidean norm and λ is a positive number predetermined by users. To obtain a good regressor, users have to find an appropriate value of λ by repeatedly solving problem (5.5) with various candidates of λ.

We consider the following multiobjective reformulation of the (5.5):

$$\begin{aligned} &\text{minimize } f^\mu(\theta) = (f_1^\mu(\theta), f_2(\theta)) \text{ subject to } \theta \in \mathbb{R}^p, \\ &\quad \text{where } f_1^\mu(\theta) = \left\|\overline{X}\theta - \overline{y}\right\|^2 + \mu\|\theta\|^2 \quad (\mu > 0), \\ &\quad\quad f_2(\theta) = \|\theta\|^2. \end{aligned} \tag{5.6}$$

Notice that $\left\|\overline{X}\theta - \overline{y}\right\|^2$ is convex but not ensured to be strongly convex. We thus add $\mu\|\theta\|^2$ to guarantee strong convexity of f_1^μ. By Theorems 1.1 and 3.1, this problem is weakly simplicial and the mapping

$$\theta^*(w) = \arg\min_\theta (w_1 f_1^\mu(\theta) + w_2 f_2(\theta)) \tag{5.7}$$

is well-defined and continuous on Δ^1, satisfying $\theta^*(\Delta_I) = X^*(f_I^\mu)$ for all $I \subseteq \{1, 2\}$.

Note that for $w = (w_1, w_2) \in \Delta^1 - \{ (0,1) \}$, the point $\theta^*(w)$ is the minimizer of the function $g^{\lambda(w)}$, where $\lambda(w) = \mu + \frac{w_2}{w_1}$. In particular, we can obtain the solutions to the original problems (5.5) for any hyper-parameter $\lambda \geq \mu$ by solving the multiobjective problem (5.6). Since the problem (5.6) is weakly simplicial, the mapping $\theta^* : \Delta^1 \to \mathbb{R}^p$ in (5.7) can be approximated by a Bézier simplex with small samples (see [35, section 4]).

Remark 5.1. As the ridge regression problem is a special case of ℓ_p-regularized regression problems, it is quite natural to expect that one can apply the same idea to solve various sparse modeling problems, including the lasso [37], the group lasso [44], the fused lasso [38], the smooth lasso [9], and the elastic net [47]. For example, the group lasso [44]

$$\text{minimize } \left\|\overline{X}\theta - \overline{y}\right\|^2 + \lambda \sum_{i=1}^{m} \left\|\theta_{(i)}\right\| \text{ subject to } \theta \in \mathbb{R}^p$$

is reformulated as

$$\begin{aligned} &\text{minimize } f(\theta) = (f_0(\theta), \dots, f_m(\theta)) \text{ subject to } \theta \in \mathbb{R}^p, \\ &\quad \text{where } f_0(\theta) = \left\|\overline{X}\theta - \overline{y}\right\|^2, \\ &\qquad\qquad f_i(\theta) = \left\|\theta_{(i)}\right\|^2 \quad (i = 1, \dots, m). \end{aligned}$$

Here, $\theta_{(i)}$ is a coefficient vector corresponding to the ith group of variables. Each groupwise regularization term is squared to be a C^2-function without changing the Pareto ordering. Note that the original formulation uses the single regularization constant λ for all the groups but our reformulation makes it possible to use different regularization constants for different groups and investigate their consequences. Unfortunately, the functions $f_0, \dots, f_m$ appearing in the reformulated problem are not always strongly convex. In order to apply our technique, we have to take other strongly convex functions $\tilde{f}_0, \dots, \tilde{f}_m$ approximating to the original ones (say, $\tilde{f}_i = f_i + \mu\|\theta\|^2$ for small $\mu > 0$) so that the resulting Pareto set and front are close to those for the original problem. The issue arising here will be addressed in a forthcoming project.

6. Conclusions. In this paper, we have shown that C^r-strongly convex problems are C^{r-1}-weakly simplicial. We have further proved that they are C^{r-1}-simplicial under some mild assumption on the coranks of the differentials of the objective mappings. The example given after the proof illustrated the necessity of the corank assumption.

We have also shown that one can always make any strongly convex problem satisfy the corank assumption by a generic linear perturbation, provided that the dimension of the decision space is sufficiently larger than that of the objective space. Note that this theorem would not hold without the assumption on the dimension pair. We have proved that a small linear perturbation does not change the Pareto set considerably.

While many multiobjective optimization problems appearing in practice are not strongly convex, we have demonstrated that several examples of such problems can be reduced to strongly convex problems via transformations preserving the Pareto ordering and the topology. The location problems, a phenotypic divergence model, and the ridge regression have been reformulated into multiobjective strongly convex problems and shown to be (weakly) simplicial.

We plan to extend the theorems to those for C^1-mappings. To do this, we will require different techniques since one cannot define the Hessian matrices for

C^1-mappings. Another interesting research project is to find a transformation that makes problems strongly convex without causing substantial changes of the Pareto set.

Appendix A. The effect of linear perturbations on the Pareto sets. In this appendix we will show the following statement mentioned in Remark 4.5.

PROPOSITION A.1. *The mapping* Γ *defined in Remark* 4.5 *is continuous. Moreover, for any* $\varepsilon > 0$*, there exists an open neighborhood* U *of* $0 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ *satisfying the following inequality for any* $(w, \pi) \in \Delta^{m-1} \times U$*:*

$$\|\Gamma(w, \pi) - \Gamma(w, 0)\| < \varepsilon.$$

Proof of Proposition A.1. Let $(\widetilde{w}, \widetilde{\pi}) \in \Delta^{m-1} \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ be an arbitrary element. We will show that Γ is continuous at the point $(\widetilde{w}, \widetilde{\pi})$.

Now, let $\gamma : \mathbb{R}^{m-1} \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \times \mathbb{R}^n \to \mathbb{R}^n$ be the mapping defined by

$$\gamma(w_1, \dots, w_{m-1}, \pi, x) = d\left(\sum_{i=1}^{m-1} w_i(f_i + \pi_i) + \left(1 - \sum_{i=1}^{m-1} w_i\right)(f_m + \pi_m)\right)_x .$$

Then, γ is a C^{r-1}-mapping. Set $\widetilde{x} = \Gamma(\widetilde{w}, \widetilde{\pi})$. By the definition of Γ, we have

$$d\left(\sum_{i=1}^{m} \widetilde{w}_i(f_i + \widetilde{\pi}_i)\right)_{\widetilde{x}} = 0,$$

where $\widetilde{w} = (\widetilde{w}_1, \dots, \widetilde{w}_m)$ and $\widetilde{\pi} = (\widetilde{\pi}_1, \dots, \widetilde{\pi}_m)$. Since $\widetilde{w}_m = 1 - \sum_{i=1}^{m-1} \widetilde{w}_i$, we get $\gamma(p(\widetilde{w}), \widetilde{\pi}, \widetilde{x}) = 0$, where $p : \Delta^{m-1} \to \mathbb{R}^{m-1}$ is the mapping defined by

$$p(w_1, \dots, w_m) = (w_1, \dots, w_{m-1}).$$

Let $\gamma_{(p(\widetilde{w}), \widetilde{\pi})} : \mathbb{R}^n \to \mathbb{R}^n$ be the mapping defined by $\gamma_{(p(\widetilde{w}), \widetilde{\pi})}(x) = \gamma(p(\widetilde{w}), \widetilde{\pi}, x)$. It is easy to verify that the following equality holds:

$$d(\gamma_{(p(\widetilde{w}), \widetilde{\pi})})_{\widetilde{x}} = H\left(\sum_{i=1}^{m} \widetilde{w}_i(f_i + \widetilde{\pi}_i)\right)_{\widetilde{x}},$$

where $H(\sum_{i=1}^{m} \widetilde{w}_i(f_i + \widetilde{\pi}_i))_{\widetilde{x}}$ is the Hessian matrix of $\sum_{i=1}^{m} \widetilde{w}_i(f_i + \widetilde{\pi}_i)$ at $\widetilde{x}$. Since the function $\sum_{i=1}^{m} \widetilde{w}_i(f_i + \widetilde{\pi}_i)$ is strongly convex, we can deduce from Lemma 2.3 that the determinant $\det d(\gamma_{(p(\widetilde{w}), \widetilde{\pi})})_{\widetilde{x}}$ is not equal to 0. Therefore, by the implicit function theorem, there exist an open neighborhood W of $(p(\widetilde{w}), \widetilde{\pi}) \in \mathbb{R}^{m-1} \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and a mapping $\varphi : W \to \mathbb{R}^n$ satisfying

$$\gamma(z, \pi, \varphi(z, \pi)) = 0$$

for any $(z, \pi) \in W \subseteq \mathbb{R}^{m-1} \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Thus the mapping Γ is continuous at $(\widetilde{w}, \widetilde{\pi})$ since $\Gamma(w, \pi)$ is equal to $\varphi(p(w), \pi)$ for any $(w, \pi) \in (p \times \mathrm{id})^{-1}(W)$.

Let $\varepsilon > 0$ be an arbitrary real number. Since Γ is continuous and Δ^{m-1} is compact, there exist an open covering $\{ V_i \}_{i=1}^{l}$ of Δ^{m-1} and open neighborhoods $U_1, \dots, U_l$ of $0 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ satisfying

$$\|\Gamma(w, \pi) - \Gamma(w, 0)\| < \varepsilon$$

for any $(w, \pi) \in V_i \times U_i$ $(1 \le i \le l)$. The intersection $U = \bigcap_{i=1}^{l} U_i$ is an open neighborhood of $0 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with the desired property. ∎

Appendix B. Counterexample in Remark 1.2. The following example illustrates that in absence of strong convexity, the relation (1.1) between Pareto optimal points and weights w_i of a scalarization of a convex C^∞-simplicial problem does not define a diffeomorphism. Actually, for some w, the function $w \mapsto x^*$ is not even univocally defined. Let us first check that the following problem is convex and C^∞-simplicial:

$$
\begin{aligned}
&\text{minimize } f(x) := (f_1(x), f_2(x)) \text{ subject to } x \in \mathbb{R},\\
&\text{where } f_1(x) = \int_0^x 2g(y) - 1\mathrm{d}y,\\
&f_2(x) = f_1(3 - x),\\
&g(x) = \frac{h(x+1)}{h(x+1) + h(1-x)},\\
&h(x) = \begin{cases} \exp(-1/x) & (x > 0),\\ 0 & (x \le 0). \end{cases}
\end{aligned}
\tag{B.1}
$$

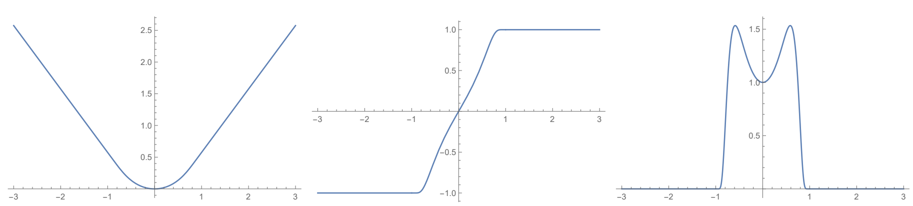
The function $h : \mathbb{R} \to \mathbb{R}$ is known as a mollifier, which is a C^∞-function, and thus $f_1, f_2, g : \mathbb{R} \to \mathbb{R}$ are C^∞-functions as well. The first and second order derivatives of f_1 are

$$
\begin{aligned}
f_1'(x) &= 2g(x) - 1,\\
f_1''(x) &= 2g'(x).
\end{aligned}
$$

The graphs of f_1, f_1', f_1'' are shown in Figure B.1. Since $f_1''(x) \ge 0$ for all $x \in \mathbb{R}$, the functions f_1, f_2 are convex (and thus so is the mapping f). Note that f_1, f_2 are not strictly convex, more specifically, not strongly convex (and thus neither is the mapping f). This is because $f_1'(x)$ is constant for $x \le -1$ or $x \ge 1$, i.e., the function $f_1(x)$ (resp., $f_2(x)$) is linear for $x \le -1$ or $x \ge 1$ (resp., $x \le 2$ or $x \ge 4$).

From the convexity of f and the first order optimality condition, we obtain the Pareto set of each subproblem as follows:

$$
\begin{aligned}
X^*(f) &= \{\, x \in \mathbb{R} \mid 0 \le x \le 3 \,\},\\
X^*(f_1) &= \{\, 0 \,\},\\
X^*(f_2) &= \{\, 3 \,\}.
\end{aligned}
$$



(a) $f_1(x) = \int_0^x 2g(y) - 1\mathrm{d}y$ (b) $f_1'(x) = 2g(x) - 1$ (c) $f_1''(x) = 2g'(x)$

FIG. B.1. *The objective function and its first and second order derivatives.*

Let $\Delta^1 = \{ (w_1, w_2) \in \mathbb{R}^2 \mid 0 \le w_1 \le 1,\ w_2 = 1 - w_1 \}$. We define a mapping $\Phi : \Delta^1 \to X^*(f)$ by

$$\Phi(w_1, w_2) = 3w_2.$$

Then, it is a C^∞-diffeomorphism satisfying simpliciality conditions:

$$\Phi(\Delta_{\{1\}}) = \Phi(1, 0) = 0 = X^*(f_1),$$
$$\Phi(\Delta_{\{2\}}) = \Phi(0, 1) = 3 = X^*(f_2).$$

Since $f|_{X^*(f)} : X^*(f) \to f(X^*(f))$ is a C^∞-diffeomorphism, so is $f \circ \Phi : \Delta^1 \to f(X^*(f))$, which implies that the problem of minimizing f is C^∞-simplicial.

Now, let us see that this convex C^∞-simplicial problem has an ill-posed weighted-sum scalarization that corresponds one weight to multiple solutions and thus not a required diffeomorphism (even not a mapping). The Pareto set $X^*(f)$ has a subset

$$A = \{ x \in \mathbb{R} \mid 1 \le x \le 2 \},$$

on which the differential of f becomes constant:

$$df|_A = \begin{pmatrix} f_1' \\ f_2' \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Let us consider a family of weighted-sum functions

$$f_w(x) = w_1 f_1(x) + w_2 f_2(x) \quad \left(w \in \Delta^1\right).$$

On the subset A, the derivative $df_w|_A = w_1 - w_2$ vanishes when $w_1 = w_2 = 1/2$. This means that the weighted-sum function with such a weight

$$f_{(\frac{1}{2},\frac{1}{2})}(x) = \frac{1}{2} f_1(x) + \frac{1}{2} f_2(x)$$

has multiple minimizers. Thus, the correspondence $x^* : \Delta^1 \rightrightarrows \mathbb{R}^2$ defined by

$$x^*(w) = \arg\min_{x \in \mathbb{R}} f_w(x)$$

is a set-valued mapping and not a diffeomorphism required for simplicial problems. However, if the value of each objective function is squared, then the resulting function

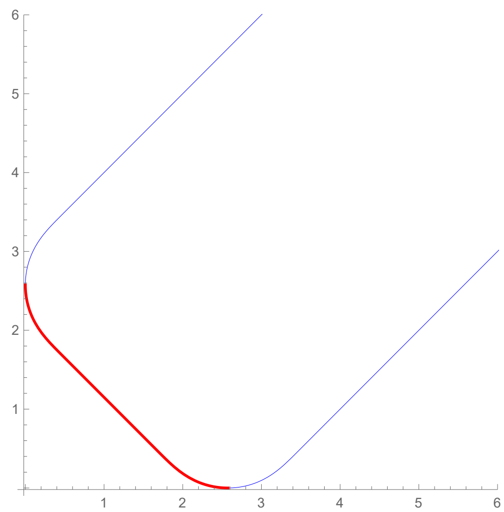
$$f_w^2(x) = w_1 \big(f_1(x)\big)^2 + w_2 \big(f_2(x)\big)^2 \quad \left(w \in \Delta^1\right)$$

has one and only one minimizer for each $w \in \Delta^1$, i.e., the weighted-sum scalarization is well-posed. In this case, the correspondence $x^{**} : \Delta^1 \rightrightarrows \mathbb{R}^2$ defined by

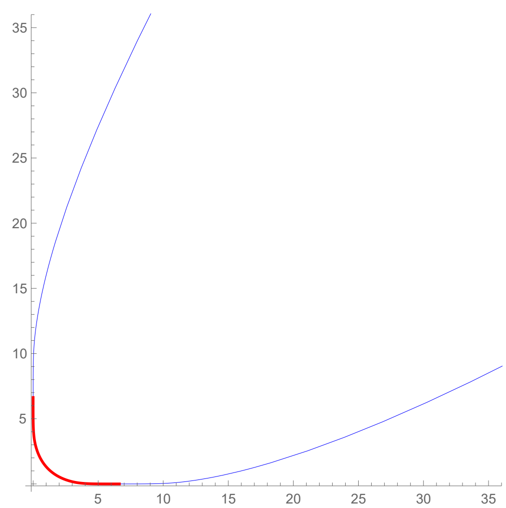
$$x^{**}(w) = \arg\min_{x \in \mathbb{R}} f_w^2(x)$$

is a point-valued mapping and a C^∞-diffeomorphism required for simplicial problems. Geometrically, this is due to the fact that the Pareto front $f(X^*(f))$ has a flat region $f(A)$ (Figure B.2(a)) while $f^2(X^*(f))$ has no such region (Figure B.2(b)).

Note that a higher dimensional version of this example is constructed by using f_1 in the one-dimensional case:



(a) f



(b) f^2

FIG. B.2. *The image (blue) and the Pareto front (red). Color available online.*

$$
\begin{aligned}
&\text{minimize } F(x) := (F_1(x), F_2(x)) \text{ subject to } x \in \mathbb{R}^n,\\
&\text{where } F_1(x) = f_1(x_1) + \sum_{i=2}^{n} x_i^2,\\
&F_2(x) = F_1(3 - x_1, x_2, \dots, x_n).
\end{aligned}
$$

Appendix C. Simpliciality of pure exchange economy. In [32], Smale dealt with the structure of the Pareto set of a pure exchange economy. In such an economy, there are l commodities, and the total amount of the ith commodity w_i is fixed ($i = 1, \dots, l$); there are m agents, and the jth agent has l commodities whose amounts are represented by an $l$vector x_j ($j = 1, \dots, m$); and agents continue to exchange commodities each other until their utility functions $u_1, \dots, u_m$ cannot be maximized simultaneously. The possible final assignments of commodities are represented as the Pareto set of the following problem:

$$\begin{aligned}
&\text{maximize } u(x_1, \dots, x_m) = (u_1(x_1), \dots, u_m(x_m)) \\
&\text{subject to } (x_1, \dots, x_m) \in W := \{ (x_1, \dots, x_m) \in \overline{P}^m \mid \textstyle\sum_{i=1}^m x_i = w \} \left(\subsetneq (\mathbb{R}^l)^m\right), \\
&\quad \text{where } \overline{P} = \{ (t_1, \dots, t_l) \in \mathbb{R}^l \mid t_i \geq 0 \text{ for all } i = 1, \dots, l \}, \ w \in \overline{P}, \\
&\qquad u_i : \overline{P} \to \mathbb{R} : \text{function which is as differentiable as necessary.}
\end{aligned}$$

Without giving proofs, Smale claimed the following proposition.[9]

Proposition C.1 (Smale [32, Proposition, pp. 533–534]). *Assume that the utility functions $u_1, \dots, u_m$ satisfy the following properties:*
(a) *"convexity": $u_i^{-1}[c, \infty)$ is strictly convex for each i, c;*
(b) *"monotonicity": Define for all $x', x \in \overline{P}$, $x' > x$ if $x'_i > x_i$ for all i and similarly for $y', y \in \mathbb{R}^m$. Then $x' > x$ implies $u(x') > u(x)$.*
Then the Pareto set of the above problem is homeomorphic to the $(m-1)$-simplex.

Compared to Smale's proposition, our results are different in the following sense:

- The ith objective function u_i in Smale's problem depends only on the ith parameter $x_i \in \overline{P}$. In particular the mapping $u_i : W \to \mathbb{R}$ will never be strongly convex, even after any Pareto-order-preserving transformation (cf. section 5).
- The dimensions of the source and the target manifolds in Smale's problems are respectively $l(m-1)$ and m, while we consider the Euclidean spaces with arbitrary dimensions.
- Smale did not analyze the topology of the Pareto front and the Pareto sets/fronts of subproblems and thus did not show the problem is (weakly) simplicial in the sense of Definition 2.1. As we mentioned in section 1, the correspondence between Pareto sets of subproblems and faces of the simplex is important for the approximation method proposed in [15, 35].
- Smale observed that the Pareto set of the problem above is homeomorphic to the simplex, while we show that the Pareto set of a strongly convex problem is diffeomorphic to the simplex.

In any case, the result observed by Smale does not imply our result, and vice versa.

Acknowledgment. This work is based on the discussions at the 2017 IMI Joint Use Research Program, Short-term Joint Research "Visualization of solution sets in many-objective optimization via vector-valued stratified Morse theory," at Kyushu University.

REFERENCES

[1] W. de Melo, *On the structure of the Pareto set of generic mappings*, Bol. Soc. Brasil. Mat., 7 (1976), pp. 121–126, https://doi.org/10.1007/BF02584786.

[2] W. de Melo, *Stability and optimization of several functions*, Topology, 15 (1976), pp. 1–12, https://doi.org/10.1016/0040-9383(76)90045-8.

[3] K. Deb and H. Jain, *An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part* I*: Solving problems with box constraints*, IEEE Trans. Evol. Comput., 18 (2014), pp. 577–601, https://doi.org/10.1109/TEVC.2013.2281535.

[4] M. Ehrgott, *Multicriteria Optimization*, 2nd ed., Springer-Verlag, Berlin, 2005, https://doi.org/10.1007/3-540-27659-9.

[9] He wrote that this proposition is due to communication with Truman Bewley.

[5] G. Eichfelder, *Adaptive Scalarization Methods in Multiobjective Optimization*, Springer-Verlag, Berlin, 2008, https://doi.org/10.1007/978-3-540-79159-1.

[6] B. Gebken, S. Peitz, and M. Dellnitz, *On the hierarchical structure of Pareto critical sets*, J. Global Optim., 73 (2019), pp. 891–913, https://doi.org/10.1007/s10898-019-00737-6.

[7] M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities*, Grad. Texts in Math. 14, Springer, New York, 1974, https://doi.org/10.1007/978-1-4615-7904-5.

[8] N. Hamada, Y. Nagata, S. Kobayashi, and I. Ono, *Adaptive weighted aggregation: A multiobjective function optimization framework taking account of spread and evenness of approximate solutions*, in Proceedings of the IEEE Congress on Evolutionary Computation, 2010, pp. 787–794, https://doi.org/10.1109/CEC.2010.5586368.

[9] M. Hebiri and S. van de Geer, *The smooth-lasso and other $\ell_1 + \ell_2$-penalized methods*, Electron. J. Stat., 5 (2011), pp. 1184–1226, https://doi.org/10.1214/11-EJS638.

[10] C. Hillermeier, *Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach*, Internat. Ser. Numer. Math. 25, Birkhäuser Verlag, Basel, 2001, https://doi.org/10.1007/978-3-0348-8280-4.

[11] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, 12 (1970), pp. 55–67, https://doi.org/10.1080/00401706.1970.10488634.

[12] S. Ichiki, *Transversality theorems on generic linearly perturbed mappings*, Methods Appl. Anal., 25 (2018), pp. 323–336, https://doi.org/10.4310/MAA.2018.v25.n4.a4.

[13] S. Ichiki and T. Nishimura, *Distance-squared mappings*, Topology Appl., 160 (2013), pp. 1005–1016, https://doi.org/10.1016/j.topol.2013.03.011.

[14] B. Kalmár, *Cobordism invariants of fold maps*, in Real and Complex Singularities, Contemp. Math. 459, AMS, Providence, RI, 2008, pp. 103–115, https://doi.org/10.1090/conm/459/08965.

[15] K. Kobayashi, N. Hamada, A. Sannai, A. Tanaka, K. Bannai, and M. Sugiyama, *Bézier simplex fitting: Describing Pareto fronts of simplicial problems with small samples in multi-objective optimization*, in Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 2304–2313, https://doi.org/10.1609/aaai.v33i01.33012304.

[16] H. W. Kuhn, *On a pair of dual nonlinear programs*, in Nonlinear Programming, J. Abadie, ed., Wiley, New York, 1967, pp. 38–54.

[17] A. Lovison, *Singular continuation: Generating piecewise linear approximations to Pareto sets via global analysis*, SIAM J. Optim., 21 (2011), pp. 463–490, https://doi.org/10.1137/100784746.

[18] A. Lovison and F. Pecci, *Hierarchical Stratification of Pareto Sets*, https://arxiv.org/abs/1407.1755, 2014.

[19] R. E. Lucchetti and E. Miglierina, *Stability for convex vector optimization problems*, Optimization, 53 (2004), pp. 517–528, https://doi.org/10.1080/02331930412331327166.

[20] P. W. Michor, *Manifolds of Differentiable Mappings*, Shiva Math. Ser. 3, Birkhauser, Basek, 1980.

[21] K. M. Miettinen, *Nonlinear Multiobjective Optimization*, Internat. Ser. Oper. Res. Management Sci. 12, Springer-Verlag, Berlin, 1999, https://doi.org/10.1007/978-1-4615-5563-6.

[22] E. Miglierina, *Stability of critical points for vector valued functions and Pareto efficiency*, J. Inform. Optim. Sci., 24 (2003), pp. 413–422, https://doi.org/10.1080/02522667.2003.10699575.

[23] E. Miglierina and E. Molho, *Scalarization and stability in vector optimization*, J. Optim. Theory Appl., 114 (2002), pp. 657–670, https://doi.org/10.1023/A:1016031214488.

[24] E. Miglierina and E. Molho, *Well-posedness and convexity in vector optimization*, Math. Methods Oper. Res., 58 (2003), pp. 375–385, https://doi.org/10.1007/s001860300310.

[25] D. Mueller-Gritschneder, H. Graeb, and U. Schlichtmann, *A successive approach to compute the bounded Pareto front of practical multiobjective optimization problems*, SIAM J. Optim., 20 (2009), pp. 915–934, https://doi.org/10.1137/080729013.

[26] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello, *A survey of multiobjective evolutionary algorithms for data mining: Part* I, IEEE Trans. Evol. Comput., 18 (2014), pp. 4–19, https://doi.org/10.1109/TEVC.2013.2290086.

[27] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello, *Survey of multiobjective evolutionary algorithms for data mining: Part* II, IEEE Trans. Evol. Comput., 18 (2014), pp. 20–35, https://doi.org/10.1109/TEVC.2013.2290082.

[28] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer, Dordrecht, 2004, https://doi.org/10.1007/978-1-4419-8853-9.

[29] A. Ponsich, A. L. Jaimes, and C. A. C. Coello, *A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and*

economics applications, IEEE Trans. Evol. Comput., 17 (2013), pp. 321–344, https://doi.org/10.1109/TEVC.2012.2196800.

[30] O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, and U. Alon, *Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space*, Science, 336 (2012), pp. 1157–1160, https://doi.org/10.1126/science.1217405.

[31] H. K. Singh, A. Isaacs, and T. Ray, *A Pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems*, IEEE Trans. Evol. Comput., 15 (2011), pp. 539–556, https://doi.org/10.1109/TEVC.2010.2093579.

[32] S. Smale, *Global analysis and economics* I*: Pareto optimum and a generalization of Morse theory*, in Dynamical Systems, M. M. Peixoto, ed., Academic Press, New York, 1973, pp. 531–544, https://doi.org/10.1016/B978-0-12-550350-1.50044-8.

[33] S. Smale, *Global analysis and economics.* III*. Pareto optima and price equilibria*, J. Math. Econom., 1 (1974), pp. 107–117, https://doi.org/10.1016/0304-4068(74)90002-0.

[34] S. Smale, *Optimizing Several Functions*, in Proceedings of the International Conference on Manifolds and Related Topics in Topology, Manifolds, University of Tokyo Press, 1975, pp. 69–75.

[35] A. Tanaka, A. Sannai, K. Kobayashi, and N. Hamada, *Asymptotic risk of Bézier simplex fitting*, in Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 2416–2424, https://doi.org/10.1609/aaai.v34i03.5622.

[36] M. Tayarani-N., X. Yao, and H. Xu, *Meta-heuristic algorithms in car engine design: A literature survey*, IEEE Trans. Evol. Comput., 19 (2015), pp. 609–629, https://doi.org/10.1109/TEVC.2014.2355174.

[37] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 58 (1996), pp. 267–288, http://www.jstor.org/stable/2346178.

[38] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, *Sparsity and smoothness via the fused lasso*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 67 (2005), pp. 91–108, https://doi.org/10.1111/j.1467-9868.2005.00490.x.

[39] A. M. Vershik and A. G. Chernyakov, *Critical points of fields of convex polytopes and the Pareto-Smale optimum with respect to a convex cone*, Sov. Math. Dokl., 26 (1982), pp. 353–356.

[40] A. M. Vershik and A. G. Chernyakov, *Fields of the convex polyhedra and the Pareto-Smale optimum*, in Functional Analysis, Optimization, and Mathematical Economics, Oxford University Press, New York, 1990, pp. 290–313.

[41] Y.-H. Wan, *On the algebraic criteria for local Pareto optima.* I, Topology, 16 (1977), pp. 113–117, https://doi.org/10.1016/0040-9383(77)90035-0.

[42] Y.-H. Wan, *On the algebraic criteria for local Pareto optima.* II, Trans. Amer. Math. Soc., 245 (1978), pp. 385–397, http://www.jstor.org/stable/1998874.

[43] H. Whitney, *Singularities of mappings of Euclidean spaces*, in Symposium internacional de topología algebraica International symposium on algebraic topology, Universidad Nacional Autónoma de México and UNESCO, Mexico City, 1958, pp. 285–301.

[44] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 68 (2006), pp. 49–67, https://doi.org/10.1111/j.1467-9868.2005.00532.x.

[45] Q. Zhang and H. Li, *MOEA/D: A multiobjective evolutionary algorithm based on decomposition*, IEEE Trans. Evol. Comput., 11 (2007), pp. 712–731, https://doi.org/10.1109/TEVC.2007.892759.

[46] A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, and Q. Zhang, *Multiobjective evolutionary algorithms: A survey of the state of the art*, Swarm Evol. Comput., 1 (2011), pp. 32–49, https://doi.org/10.1016/j.swevo.2011.03.001.

[47] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 67 (2005), pp. 301–320, http://www.jstor.org/stable/3647580.