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. 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.

1. INTRODUCTION

Multiobjective optimization arises in various fields of science and engineering (e.g., data mining [18, 19], finance [21], car design [25], and more [32]). 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 [2,16]. 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 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 [1, 3, 8, 31]).

In general it becomes easier to obtain the whole Pareto set if it has a simple topological structure. For example, applying a generalized homotopy method [8] 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* [12]. Figure 1 depicts an example of a simplicial problem with the 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 (for the precise definition of simplicial problems, see section 2.2). The property that Φ sends the faces to the Pareto sets has a great utility: it is known that using this correspondence, we can reduce the number of solutions to obtain an approximation of the mapping Φ , i.e., a parametric-surface approximation of the Pareto set (cf. [12,24]).

In the literature, some researchers also studied hierarchical structures of solution sets of various types. Shoval et al. pointed out that when minimizing distances from



FIGURE 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.

given points, one can decompose the boundary of the Pareto set into the Pareto sets of subproblems [22]. Mainly based on Wan's results [28,29], Lovison and Pacci showed a generic property of mappings that the local Pareto set admits a Whitney stratification [14]. Gebken et al. investigated a situation that the Pareto critical set can be determined by a reduced number of objectives when the number of objectives are greater than that of variables [4]. Compared to the above studies, the theory of simplicial problems uniquely provides efficient computation of the whole Pareto set of some class of mappings from high-dimensional decision spaces to low-dimensional objective spaces, which are frequently seen in practical problems.

It is therefore important to investigate what kinds of known problem classes, such as linear problems, convex problems¹, etc., are simplicial. In this paper, we will show that problems minimizing strongly convex mappings are simplicial under the assumption on the differentials of the mappings.

Theorem 1.1. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex C^r -mapping $(2 \le r \le \infty)$. The multiobjective optimization problem 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. Under the assumptions of theorem 1.1, the weighted-sum scalarization gives an instance of the aforementioned homeomorphism Φ , that is, the mapping $x^* : \Delta^{m-1} \to X^*(f)$ defined by

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

is a homeomorphism. Since the scalarized function $\sum_{i=1}^{m} w_i f_i : \mathbb{R}^n \to \mathbb{R}$ is strongly convex (see section 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 [3, 6, 17, 23] are able to provide solutions for parametric-surface approximation methods in [12, 24] to build an approximation of Φ .

¹While 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.

Note that (strict) convexity of a mapping does not necessarily imply that the corresponding problem is simplicial. For example, the single objective problem 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.

As the example given in example 3.5 indicates, we cannot drop the assumption on the corank in theorem 1.1. Nevertheless, one can make any strongly convex mapping satisfy the corank assumption by a generic linear perturbation, provided that the dimension of the source is sufficiently larger than that of the target. See theorem 4.1 for details. We will further observe that theorem 4.1 would not hold without the assumption on the dimensions (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).

Applying theorem 1.1, we can show that several problems appearing in practical situations are simplicial. In 1967, Kuhn [13] showed that the Pareto set of a location problem under the Euclidean norm is the convex hull of demand points, which becomes a simplex when the demand points are in general position. We will observe in section 5 that this problem can be made strongly convex by suitable transformations on the target space preserving the Pareto ordering. In particular, our theorem gives an alternative proof of Kuhn's result. Shoval et al. [22] proposed a multiobjective model for phenotypic divergence of species in evolutionary biology and indicated that its Pareto set is a curved simplex. Again, this problem becomes strongly convex after suitable transformations (see section 5). We can thus give a rigorous proof for the observation in [22].

The paper is organized as follows: After preliminaries in section 2, the proof of the main theorem will be given in section 3. In section 4, we will show 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 will be 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 $\{1, \ldots, m\}$ by M.

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

minimize
$$f(x) = (f_1(x), \dots, f_m(x))$$

subject to $x \in X (\subseteq \mathbb{R}^n)$.

According to the Pareto ordering, i.e.,

 $f(x) \prec f(y) \iff f_i(x) \leq f_i(y)$ for all $i \in M$ and $f_j(x) < f_j(y)$ for some $j \in M$, we basically would like to obtain the *Pareto set*

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

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:

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

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

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

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

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

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 $\tilde{f}|_{\Delta^{m-1}} = f^{2}$. 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 [15, §.2] for more general definition of diffeomorphisms between manifolds with corners.

Definition 2.1. Let X be a subset of \mathbb{R}^n and $f = (f_1, \ldots, f_m)$ be a mapping from X to \mathbb{R}^m . For $I = \{i_1, \ldots, i_k\} \subseteq M$ such that $i_1 < \cdots < i_k$, we put $f_I = (f_{i_1}, \ldots, f_{i_k})$. The problem 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 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 $I \subseteq M$.

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 with 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

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

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 minimizing a strongly convex mapping is called a *strongly convex problem*. The followings are basic properties of (strongly) convex mappings which will be needed later on:

Lemma 2.2 ([20, 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 $f(x) + df_x \cdot (y - x) \leq f(y)$ for any $x, y \in X$.

²The 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.

Lemma 2.3 ([20, 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 ([20, 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 an element $(w_1, \ldots, w_m) \in$ Δ^{m-1} such that f satisfies one (and hence both) of the following equivalent conditions:

- (1) $\sum_{i=1}^{m} w_i (df_i)_x = 0.$
- (2) The point $x \in \mathbb{R}^n$ is a unique element such that the function $\sum_{i=1}^m w_i f_i$ is minimized.

For the proof, we prepare some lemmas.

Lemma 2.6 ([16, 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 $(w_1, \ldots, w_m) \in \Delta^{m-1}$ be an element. If $x \in \mathbb{R}^n$ is a unique element such that the function $\sum_{i=1}^m w_i f_i$ is minimized, then $x \in X^*(f)$.

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

Lemma 2.7 ([16, 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 an element $(w_1, \ldots, w_m) \in \Delta^{m-1}$ 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 $(w_1,\ldots,w_m) \in \Delta^{m-1}$ be an element. 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

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

Since q is convex, we can deduce from lemma 2.2 that the following inequality holds $\frac{1}{2}$ for any $y \in \mathbb{R}^n$:

(2)
$$g(x) + dg_x \cdot (y - x) \le g(y).$$

Suppose that $\sum_{i=1}^{m} w_i (df_i)_x = 0$. We can easily deduce the assertion 2 from eq. (1) and eq. (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 eq. (1) holds, we have the assertion 1.

Proof of proposition 2.5. Suppose that $x \in X^*(f)$. Using lemma 2.7, we can verify that there exists an element $(w_1, \ldots, w_m) \in \Delta^{m-1}$ satisfying $\sum_{i=1}^m w_i (df_i)_x = 0$. From lemma 2.8, the point $x \in \mathbb{R}^n$ is an element such that $\sum_{i=1}^m w_i f_i$ is minimized. Since $\sum_{i=1}^m w_i f_i$ is strongly convex C^1 -function, by lemma 2.4, we have the assertion 2. Finally, suppose the 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 [5]). 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:

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

where $j^1g: \mathbb{R}^n \to J^1(\mathbb{R}^n, \mathbb{R}^m)$ 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^1g is transverse to S at x if either of the following conditions holds:

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

The mapping j^1g 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 $S_1(f) = (j^1f)^{-1}(S_1)$.

Note that we can easily deduce from the 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 $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a C^2 -mapping and $x_0 \in \operatorname{Crit}(f)$. Suppose that the corank ³ 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 regular. We define the function $\lambda_f : \mathbb{R}^n \to \mathbb{R}^{n-m+1}$ as follows:

$$\lambda_f(x) = (J_1(x), \dots, J_{n-m+1}(x)),$$

where

$$J_{i}(x) = \det \begin{pmatrix} \frac{\partial f_{1}}{\partial x_{1}}(x) & \cdots & \frac{\partial f_{m}}{\partial x_{1}}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{1}}{\partial x_{m-1}}(x) & \cdots & \frac{\partial f_{m}}{\partial x_{m-1}}(x) \\ \frac{\partial f_{1}}{\partial x_{m-1+i}}(x) & \cdots & \frac{\partial f_{m}}{\partial x_{m-1+i}}(x) \end{pmatrix}$$

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) $\ker(d\lambda_f)_{x_0} \oplus \ker df_{x_0} = T_{x_0}\mathbb{R}^n$.

³For a linear mapping $\varphi: V \to W$ the non-negative number dim $W - \operatorname{rank}(\varphi)$ is called the corank of φ .

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 $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ 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 $(w_1, \ldots, w_m) \in \Delta^{m-1}$, 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 minimizing point 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 $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be a strongly convex C^r -mapping $(2 \le r \le \infty)$.

- (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) Suppose 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 follows from this theorem as follows: It is easy to see that any subproblem of a strongly convex problem is again 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 the assertion 2 of theorem 3.1 that a strongly convex problem is simplicial under the 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 fold⁴ 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 non-injective 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 $g: \Delta_{\delta'}^{m-1} \times \mathbb{R}^n \to \mathbb{R}^n$ 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 $\left(\frac{\partial g_i}{\partial x_j}\right)_{1 \leq i,j \leq n}$ is equal to $\sum_{k=1}^m w_k H(f_k)_x$, where $H(f_k)_x$ is the Hessian matrix of

⁴It is indeed a "definite" fold.

 f_k at x. Since f_k is strongly convex, the Hessian matrix $H(f_k)_x$ is positive definite (see lemma 2.3). Thus, the matrix $\left(\frac{\partial g_i}{\partial x_j}\right)_{1\leq i,j\leq n}$ is regular on Δ^{m-1} . By the implicit function theorem, for any $y \in \Delta^{m-1}$ there exists an open neighborhood $U_y \subseteq \Delta_{\delta'}^{m-1}$ 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_{δ}^{m-1} is contained in U.⁵ We can define $\widetilde{x^*}: \Delta_{\delta}^{m-1} \to \mathbb{R}^n$ 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 $D_{\varepsilon}^{m-1} \subsetneq \mathbb{R}^{m-1}$ by

$$D_{\varepsilon}^{m-1} = \left\{ \left. (z_1, \dots, z_{m-1}) \in \mathbb{R}^{m-1} \right| \sum_i z_i < 1+\varepsilon, \ z_i > -\varepsilon \right\}.$$

We denote the closure D_0^{m-1} by D^{m-1} . It is easy to check that the projection $p: \Delta_{\varepsilon}^{m-1} \to D_{\varepsilon}^{m-1}$ defined by $p(w_1, \ldots, w_m) = (w_1, \ldots, w_{m-1})$ is a diffeomorphism. In what follows we will identify $\Delta_{\varepsilon}^{m-1}$ with D_{ε}^{m-1} by p.

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

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

We thus obtain:

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

where we denote the matrix $\left(\sum_{k=1}^{m-1} z_k H(f_k)_{\widetilde{x^*}(z)} + \left(1 - \sum_{k=1}^{m-1} z_k\right) H(f_m)_{\widetilde{x^*}(z)}\right)^{-1}$ 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

⁵ If not, we can take $x_n \in U^c \cap \Delta_{1/n}^{m-1}$ for any $n \in \mathbb{N}$. Since $\overline{\Delta_1^{m-1}}$ 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 $\Delta^{m-1} \subsetneq U$.

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

We will show that the matrix

(3)
$$((df_1)_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)} \cdots (df_{m-1})_{\widetilde{x^*}(z)} - (df_m)_{\widetilde{x^*}(z)})$$

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

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

On the other hand, by the definition of 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 corank $\left(df_{\widetilde{x}(z)} \right) = 1$. Therefore, the matrix in eq. (3) has rank m-1for any $z \in D^{m-1}$. Since the condition that the matrix in eq. (3) has rank m-1is 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

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

also has rank m-1 since A(z) is regular 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 $w, z \in$ Δ^{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 $n \ge m$.

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

- $(f_1, \ldots, f_m) \mapsto (f_{\sigma(1)}, \ldots, f_{\sigma(m)})$ $(\sigma \in \mathfrak{S}_m),$ $(f_1, \ldots, f_m) \mapsto (f_1, \ldots, f_m + \alpha f_i)$ $(\alpha > 0, i = 1, \ldots, m 1),$ 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 followings:

$$(df_m)_x = 0, \quad \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{1 \le i,j \le m-1} = I_{m-1}, \quad \frac{\partial f_i}{\partial x_j}(x) = 0 \quad \begin{pmatrix} i = 1, \dots, m \\ j = m, \dots, n \end{pmatrix}.$$

Let $\lambda_f : \mathbb{R}^n \to \mathbb{R}^{n-m+1}$ be the mapping defined in section 2.4. By lemma 2.10, it suffices to show the followings:

- (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:

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

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

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

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:

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

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 the condition (A):

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

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

3.1. **Examples.** One of the most simple and 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 a convex hull of minimizing points (resp. their values) of individual objective functions [13]. Thus, if these minimizing points 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 minimizing points 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 $x^* : \Delta^2 \to X^*(f)$. The first example (given in 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 minimizing points of the three component functions are in general position). This example implies that we cannot drop the corank assumption in the assertion 2 of theorem 3.1. The second example (given in example 3.5) satisfies the corank assumption, and thus the corresponding x^* is a diffeomorphism (although the minimizing points of the three component functions are *not* in general position).

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:

$$f_1(x, y, z) = x^2 + y^2 + z^2,$$

$$f_2(x, y, z) = x + y + x^2 + y^2 + z^2,$$

$$f_3(x, y, z) = -(x + y) + x^2 + 2y^2 + z^2$$

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

$$df_{1,p} = (2x, 2y, 2z)$$

$$df_{2,p} = (1 + 2x, 1 + 2y, 2z)$$

$$df_{3,p} = (-1 + 2x, -1 + 4y, 2z),$$

and thus the corank of df_0 is 2. Since f is strongly convex, the mapping x^* : $\Delta^2 \to X^*(f)$ is surjective by the assertion 1 of theorem 3.1. Regarding Δ^2 as $D^2 = \{ (w_2, w_3) \mid w_2, w_3 \ge 0, w_2 + w_3 \le 1 \}$, 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 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 2 describes the Pareto set of f, together with the contours of the functions f_1 (red), f_2 (blue) and f_3 (green).



FIGURE 2. The Pareto set of f. The union of two domains colored with orange and blue is the Pareto set.

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 minimizing h_{ε} is simplicial. 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 (non-general position without corank 2). We define a mapping $f = (f_1, f_2, f_3) : \mathbb{R}^3 \to \mathbb{R}^3$ as follows:

$$f_1(x, y, z) = x^2 + (-y + x)^2 + z^2,$$

$$f_2(x, y, z) = 2(x - 1)^2 + (-y + x - 1)^2 + z^2,$$

$$f_3(x, y, z) = (x - 2)^2 + (y + x - 2)^2 + z^2.$$

The mapping f is strongly convex. We will check that f satisfies the assumption in the assertion 2 of theorem 3.1 (although the minimizing points of f_1, f_2, f_3 are not in general position). Let p = (x, y, z) be a point in $X^*(f)$. It is easy to see that z is equal to 0. Thus the differentials at p are calculated as follows:

$$(df_1)_p = (4x - 2y, -2x + 2y, 0), (df_2)_p = (6x - 2y - 6, -2x + 2y + 2, 0), (df_3)_p = (4x + 2y - 8, 2x + 2y - 4, 0).$$

Suppose that the corank of df_p is greater than or equal to 2. Then the following equalities hold:

(4)
$$0 = \det \begin{pmatrix} 4x - 2y & -2x + 2y \\ 6x - 2y - 6 & -2x + 2y + 2 \end{pmatrix} = 4\left(x - \frac{y}{2} - \frac{1}{2}\right)^2 - (y - 3)^2 + 8,$$

(5)
$$0 = \det \begin{pmatrix} 4x - 2y & -2x + 2y \\ 4x + 2y - 8 & 2x + 2y - 4 \end{pmatrix} = 2 \left(8(x-1)^2 - 4\left(y - \frac{3}{2}\right)^2 + 1 \right).$$

These equalities give rise to two hyperbolas given in fig. 3(a) (the red hyperbola is defined by eq. (4), while the blue one is defined by eq. (5)). As shown in fig. 3(a), the two hyperbolas intersect at two points. One is the origin **0** and let q = (x', y') be the other. Since the rank of $((df_1)_0, (df_2)_0, (df_3)_0)$ is 2, (x, y) is not equal to (0, 0). Thus we obtain (x, y) = (x', y'). However, since y' - x' > 1 and x', y' > 1, all of the three values -2x' + 2y', -2x' + 2y' + 2 and 2x' + 2y' - 4 are greater than 0, contradicting proposition 2.5. Hence we can conclude that there is no point $p \in X^*(f)$ with corank $(df_p) \ge 2$. Figure 3(b) describes the Pareto set of f, together with the contours of the functions f_1 (red), f_2 (blue) and f_3 (green).

4. Generic linear perturbations of strongly convex mappings

In this section, we will investigate the multiobjective optimization problem 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:



FIGURE 3. The dotted lines in fig. 3(a) are asymptotes of the hyperbolas.

Theorem 4.1. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ $(n \ge m)$ be a strongly convex C^r -mapping $(2 \le r \le \infty)$. 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 minimizing $f + \pi : \mathbb{R}^n \to \mathbb{R}^m$ is C^{r-1} -simplicial.

We begin with observing 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 $t \in [0, 1]$:

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

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

Before proving theorem 4.1, we will briefly review the result in [10] needed here. Let $S_k \subsetneq J^1(\mathbb{R}^n, \mathbb{R}^m)$ be the subset defined in section 2.4. It is known that S_k is a submanifold of $J^1(\mathbb{R}^n, \mathbb{R}^m)$ satisfying the following (see [5]):

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

where 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 [10, Theorem 1]:

Lemma 4.3 (cf. [10]). 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 $m \ge 2$. Since $n \ge m$, the codimension of S_2 is equal to 2(n - m + 2). By the assumption n - 2m + 4 > 0, we can obtain the inequality codim $S_2 > n$. Let k be an integer with $2 \le k \le m$. It follows that

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

In particular, for a mapping $g : \mathbb{R}^n \to \mathbb{R}^m$, transversality of j^1g to S_k is equivalent to the condition that $j^1g(\mathbb{R}^n) \cap S_k = \emptyset$, that is, g has no corank k critical points (see [5, Ch. II, Proposition 4.2]). Furthermore, the following inequality holds:

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

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. Lastly, 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

$$g_1(x) = x_1^2 + x_3 x_2 + \frac{1}{2} (x_2^2 + x_4 x_1) + x_3^2 + x_4^2,$$

$$g_2(x) = x_2^2 + x_4 x_1 + \frac{1}{2} (x_1^2 + x_3 x_2) + x_3^2 + x_4^2.$$

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 $dG_0 = \begin{pmatrix} O & O \\ -I & I \end{pmatrix}$, where I and O are 2×2 unit matrix and zero matrix, respectively. Define the cokernel of dG_0 as

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

then, the subspace coker dG_0 is equal to $\langle (1,0,1,0), (0,1,0,1) \rangle$. In particular, the subspace 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 $d(G+\pi)_x = \begin{pmatrix} A(x,\pi) & B(x,\pi) \\ C(x,\pi) & D(x,\pi) \end{pmatrix}$ 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 $\begin{pmatrix} I \\ -D^{-1}(x,\pi)C(x,\pi) & I \end{pmatrix}$ to $d(G+\pi)_x$ from the right, we obtain the matrix $\begin{pmatrix} A(x,\pi)-B(x,\pi)D^{-1}(x,\pi)C(x,\pi) & B(x,\pi) \\ O & D(x,\pi) \end{pmatrix}$. 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

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

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 coker $d(G + \cdot)_{\hat{x}(\cdot)} : V \to Gr(2, \mathbb{R}^4)$ where $Gr(2, \mathbb{R}^4)$ is all 2-dimensional linear subspaces of \mathbb{R}^4 . Since coker dG_0 intersects with the interior of Δ^3 , so does 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 $(2 \leq r \leq \infty)$. 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 the Pareto set of a linearly perturbed mapping is close to that of the original one. 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 lots of practical problems are *not* necessarily strongly convex, we can apply suitable "structurepreserving" 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 to find 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 [13] is the following

problem:⁶

(6) minimize
$$f(x) = (f_1(x), \dots, f_m(x))$$
 subject to $x \in \mathbb{R}^n$
where $f_i(x) = ||x - p_i||$ $(i = 1, \dots, m)$.

The mapping f is called a *distance mapping* [11], which is not differentiable. Each f_i is convex but not strongly convex, and thus so is the problem eq. (6).

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

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

The mapping $T \circ f$ (called a *distance-squared mapping* [11]) is differentiable and strongly convex, in particular the problem eq. (7) is strongly convex. Since T preserves the Pareto ordering, the Pareto sets of eqs. (6) and (7) are identical and the Pareto fronts are homeomorphic.

For the original problem eq. (6), there is a weight with which the weighted sum scalarization has non-unique solutions (e.g. the problem 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 $t \in [0, 1]$), in particular we cannot define a mapping x^* given in section 3. On the other hand, for the transformed problem eq. (7), every scalarized problem has a unique solution and the entire Pareto set consists of such elements by the 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 $n \ge m-1$ and p_1, \ldots, p_m are in general position. Thus, the assertion 2 of theorem 3.1 guarantees that the problem eq. (7) is (C^{∞}) -simplicial. Since T preserves the Pareto ordering, one can easily see that the problem eq. (6) is also (C^0) -simplicial.

5.2. **Phenotypic divergence model.** Another example minimizing distances from points arises in evolutionary biology. Let A_i be a symmetric, positive definite matrix of size n and $p_i \in \mathbb{R}^n$ (i = 1, ..., m). Shoval et al. [22] provided a model for describing phenotypic divergence of species, which is an extension of the location problem:

(8) minimize
$$f(x) = (f_1(x), \dots, f_m(x))$$
 subject to $x \in \mathbb{R}^n$
where $f_i(x) = ||A_i(x - p_i)||$ $(i = 1, \dots, m)$

As before, the problem minimizing eq. (8) is convex but not strongly convex. We can again apply the transformation T used in the previous subsection and obtain

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

Since affine transformations of the source space preserve strong convexity of a problem, each component of $T \circ f$ (and thus the problem eq. (9)) is strongly convex. Applying the assertion 1 of theorem 3.1 we can conclude that both of the problems eqs. (8) and (9) are weakly simplicial. In order to further show that these problems are simplicial, we have to check the corank condition in the assertion 2 of theorem 3.1, which would be a hard task, even if the demand points p_1, \ldots, p_m are in general position. Indeed, problems appearing in section 3.1 are special cases of the

16

⁶While Kuhn [13] originally considered a planar case (n = 2), we consider the problem in general dimension.

problems we are dealing with here. As discussed in section 3.1, generality of the configuration of demand points does not necessarily imply the corank condition.

5.3. **Ridge regression.** The ridge regression [9] can be reformulated as a multiobjective strongly convex problem. Let us consider a linear regression model

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

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

(10) minimize
$$g^{\lambda}(\theta) = \|\overline{X}\theta - \overline{y}\|^2 + \lambda \|\theta\|^2$$
 subject to $\theta \in \mathbb{R}^p$

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 the problem eq. (10) with various candidates of λ .

We consider the following multiobjective reformulation of the eq. (10):

(11) minimize
$$f^{\mu}(\theta) = (f_{1}^{\mu}(\theta), f_{2}(\theta))$$
 subject to $\theta \in \mathbb{R}^{p}$
where $f_{1}^{\mu}(\theta) = \left\|\overline{X}\theta - \overline{y}\right\|^{2} + \mu \|\theta\|^{2} \quad (\mu > 0),$
 $f_{2}(\theta) = \|\theta\|^{2}.$

Notice that $\|\overline{X}\theta - \overline{y}\|^2$ 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

(12)
$$\theta^*(w) = \arg\min_{\theta} (w_1 f_1^{\mu}(\theta) + w_2 f_2(\theta))$$

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 eq. (10) for any hyper-parameter $\lambda \geq \mu$ by solving the multiobjective problem eq. (11). Since the problem eq. (11) is weakly simplicial, the mapping $\theta^* : \Delta^1 \to \mathbb{R}^p$ in eq. (12) can be approximated by a Bézier simplex with small samples (see [24, 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 [26], the group lasso [30], the fused lasso [27], the smooth lasso [7], and the elastic net [33]. For example, the group lasso [30]

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

is reformulated as

minimize
$$f(\theta) = (f_0(\theta), \dots, f_m(\theta))$$
 subject to $\theta \in \mathbb{R}^p$
where $f_0(\theta) = \|\overline{X}\theta - \overline{y}\|^2$,
 $f_i(\theta) = \|\theta_{(i)}\|^2$ $(i = 1, \dots, m)$.

Here, $\theta_{(i)}$ is a coefficient vector corresponding to the *i*-th group of variables. Each group-wise 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 us possible to use different regularization constants for different groups and investigate their consequences. Unfortunately, the functions f_0, \ldots, 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 $\tilde{f}_0, \ldots, \tilde{f}_m$ 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 has 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 lots of 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.

Acknowledgements

K. H. was supported by JSPS KAKENHI Grant Number JP17K14194. S. I. was supported by JSPS KAKENHI Grant Number JP19J00650. Y. K. was supported by JSPS KAKENHI Grant Number JP16J02200. H. T. was supported by JSPS KAKENHI Grant Number JP19K03484 and JST PRESTO Grant Number

JPMJPR16E8. The authors were supported by RIKEN Center for Advanced Intelligence Project. This work is based on the discussions at 2017 IMI Joint Use Research Program, Short-term Joint Research "Visualization of solution sets in manyobjective optimization via vector-valued stratified Morse theory" and was supported by the Research Institute for Mathematical Sciences, a Joint Usage/Research Center located in Kyoto University.

References

- K. Deb and H. Jain. An evolutionary many-objective optimization algorithm using referencepoint-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, Aug 2014.
- [2] Matthias Ehrgott. Multicriteria Optimization. Springer-Verlag, Berlin, Heidelberg, 2nd edition, 2005.
- [3] G. Eichfelder. Adaptive Scalarization Methods in Multiobjective Optimization. Springer-Verlag, Berlin, Heidelberg, 2008.
- [4] Bennet Gebken, Sebastian Peitz, and Michael Dellnitz. On the hierarchical structure of pareto critical sets. Journal of Global Optimization, 73(4):891–913, Apr 2019.
- [5] M. Golubitsky and V. Guillemin. Stable Mappings and Their Singularities, volume 14 of Graduate Texts in Mathematics. Springer New York, 1974.
- [6] Naoki Hamada, Yuichi Nagata, Shigenobu Kobayashi, and Isao Ono. Adaptive weighted aggregation: A multiobjective function optimization framework taking account of spread and evenness of approximate solutions. In *Proceedings of the 2010 IEEE Congress on Evolutionary Computation (CEC 2010)*, pages 787–794, 2010.
- [7] Mohamed Hebiri and Sara van de Geer. The smooth-lasso and other ℓ₁+ℓ₂-penalized methods. Electron. J. Statist., 5:1184–1226, 2011.
- [8] C. Hillermeier. Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach, volume 25 of International Series of Numerical Mathematics. Birkhäuser Verlag, Basel, Boston, Berlin, 2001.
- [9] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [10] Shunsuke Ichiki. Transversality theorems on generic linearly perturbed mappings. Methods and Applications of Analysis special volume in memory of John Mather, to appear.
- [11] Shunsuke Ichiki and Takashi Nishimura. Distance-squared mappings. *Topology Appl.*, 160(8):1005–1016, 2013.
- [12] Ken Kobayashi, Naoki Hamada, Akiyoshi Sannai, Akinori Tanaka, Kenichi Bannai, and Masashi Sugiyama. Bézier simplex fitting: Describing Pareto fronts of simplicial problems with small samples in multi-objective optimization. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, to appear.
- [13] Harold W Kuhn. On a pair of dual nonlinear programs. Nonlinear Programming, 1:38–45, 1967.
- [14] A. Lovison and F. Pecci. Hierarchical stratification of Pareto sets. ArXiv e-prints, 2014.
- [15] P. W. Michor. Manifolds of Differentiable Mappings, volume 3 of Shiva mathematics series. Birkhauser, 1980.
- [16] Kaisa M. Miettinen. Nonlinear Multiobjective Optimization, volume 12 of International Series in Operations Research & Management Science. Springer-Verlag, GmbH, 1999.
- [17] D. Mueller-Gritschneder, H. Graeb, and U. Schlichtmann. A successive approach to compute the bounded Pareto front of practical multiobjective optimization problems. *SIAM Journal* on Optimization, 20(2):915–934, 2009.
- [18] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello. A survey of multiobjective evolutionary algorithms for data mining: Part I. *IEEE Transactions on Evolutionary Computation*, 18(1):4–19, Feb 2014.
- [19] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello. Survey of multiobjective evolutionary algorithms for data mining: Part II. *IEEE Transactions on Evolutionary Computation*, 18(1):20–35, Feb 2014.
- [20] Yurii Nesterov. Introductory Lectures on Convex Optimization: A Basic Course. Kluwer Academic Publishers, 2004.

- [21] 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 Transactions on Evolutionary Computation*, 17(3):321–344, June 2013.
- [22] 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(6085):1157–1160, 2012.
- [23] H. K. Singh, A. Isaacs, and T. Ray. A Pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems. *IEEE Transactions on Evolutionary Computation*, 15(4):539–556, 2011.
- [24] Akinori Tanaka, Akiyoshi Sannai, Ken Kobayashi, and Naoki Hamada. Asymptotic risk of Bézier simplex fitting. arXiv e-prints, Jun 2019.
- [25] M. Tayarani-N., X. Yao, and H. Xu. Meta-heuristic algorithms in car engine design: A literature survey. *IEEE Transactions on Evolutionary Computation*, 19(5):609–629, Oct 2015.
- [26] Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996.
- [27] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(1):91–108, 2005.
- [28] Yieh-Hei Wan. On local Pareto optima. Journal of Mathematical Economics, 2(1):35–42, 1975.
- [29] Yieh-Hei Wan. On the algebraic criteria for local Pareto optima-I. Topology, 16(1):113-117, 1977.
- [30] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1):49–67, 2006.
- [31] Q. Zhang and H. Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. IEEE Transactions on Evolutionary Computation, 11(6):712-731, December 2007.
- [32] Aimin Zhou, Bo-Yang Qu, Hui Li, Shi-Zheng Zhao, Ponnuthurai Nagaratnam Suganthan, and Qingfu Zhang. Multiobjective evolutionary algorithms: A survey of the state of the art. Swarm and Evolutionary Computation, 1(1):32–49, 2011.
- [33] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67(2):301–320, 2005.

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

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

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

Then, γ is a C^{r-1} -mapping. Set $\tilde{x} = \Gamma(\tilde{w}, \tilde{\pi})$. 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 $\widetilde{w} = (\widetilde{w}_1, \ldots, \widetilde{w}_m)$ and $\widetilde{\pi} = (\widetilde{\pi}_1, \ldots, \widetilde{\pi}_m)$. 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

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

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, \ldots, U_l of $0 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ satisfying

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

for any $(w,\pi) \in V_i \times U_i$ $(1 \leq i \leq l)$. 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. \Box

ARTIFICIAL INTELLIGENCE LABORATORY, FUJITSU LABORATORIES LTD., KAWASAKI, 211-8588, JAPAN; RIKEN AIP FUJITSU COLLABORATION CENTER, RIKEN CENTER FOR ADVANCED INTELLIGENCE PROJECT, TOKYO, 103-0027, JAPAN.

E-mail address: hamada-naoki@fujitsu.com

DEPARTMENT OF MATHEMATICS FACULTY OF SCIENCE AND TECHNOLOGY, KEIO UNIVERSITY, YOKOHAMA, 223-8522, JAPAN; MATHEMATICAL SCIENCE TEAM, RIKEN CENTER FOR ADVANCED INTELLIGENCE PROJECT, TOKYO, 103-0027, JAPAN.

E-mail address: k-hayano@math.keio.ac.jp

INSTITUTE OF MATHEMATICS FOR INDUSTRY, KYUSHU UNIVERSITY, FUKUOKA, 819-0395, JAPAN. *E-mail address*: s-ichiki@imi.kyushu-u.ac.jp

INSTITUTE OF MATHEMATICS FOR INDUSTRY, KYUSHU UNIVERSITY, FUKUOKA, 819-0395, JAPAN. *E-mail address*: 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.

E-mail address: teramoto@es.hokudai.ac.jp