CHARACTERIZATION OF THE EQUALITY OF WEAK EFFICIENCY AND EFFICIENCY ON CONVEX FREE DISPOSAL HULLS

NAOKI HAMADA AND SHUNSUKE ICHIKI

Abstract. In solving a multi-objective optimization problem by scalarization techniques, solutions to a scalarized problem are, in general, weakly efficient rather than efficient to the original problem. Thus, it is crucial to understand what problem ensures that all weakly efficient solutions are efficient. In this paper, we give a characterization of the equality of the weakly efficient set and the efficient set, provided that the free disposal hull of the domain is convex. Using this characterization, we see that the set of weakly efficient solutions is equal to the set of efficient solutions (and also the set of strictly efficient solutions) in strongly convex problems. As practical applications, we consider the structure of the set of efficient solutions to a location problem under Mahalanobis distances and a multi-objective reformulation of the elastic net. Multi-objective optimization and Free disposal hull and Weak efficiency and Efficiency and Location problem and Sparse modeling

1. Introduction

The aim of multi-objective optimization is to find efficient solutions to a given problem. In order to do so, a lot of scalarization techniques have been developed so far (see for example [3, 5, 18, 19, 21, 25, 33]). Nevertheless, there is no scalarization method that ensures for a wide variety of problems that all solutions optimal to scalarized problems are efficient to the original problem. In general, scalarization methods only ensure that their solutions are weakly efficient to the original problem, which means users may waste computation resources for finding inefficient, undesirable solutions. Thus, it is crucial to understand conditions that the weak efficiency coincides with the efficiency.

In the literature, the relationship between the weak efficiency and the efficiency has been investigated. In some cases, the set of weakly efficient solutions to a given problem can be described as the union of the sets of efficient solutions to its subproblems [2, 14, 16, 31]. This property was named the Pareto reducibility [22] and further investigated [13, 23, 24]. Some relationships of the weak efficiency and the efficiency on quasi-convex problems are collected in [4, 15]. However, the equality between the weak efficiency and the efficiency, both of which are of the original problem (rather than subproblems), is still unclear.

In this paper, we give a characterization of the equality of the set of weakly efficient solutions and the set of efficient solutions, provided that the free disposal hull [1] of the image of an objective mapping is convex (see Proposition 2.3 in Section 2). This claim is derived from our main theorem (Theorem 2.1 in Section 2), which gives a similar characterization of the equality of the weakly efficient set
and the efficient set on partially ordered Euclidean space without objective functions. Moreover, Proposition 2.3 yields a lot of mathematical applications (see Corollary 2.4 in Section 2 and Corollaries 6.1 to 6.3 in Section 6). In particular, it follows from Proposition 2.3 that the set of weakly efficient solutions is equal to the set of efficient solutions for any strongly convex problem (Corollary 2.4). Corollary 2.4 is the most essential application of Proposition 2.3 in this paper. Using this corollary, we discuss two practical problems: the location problem under Mahalanobis distances, which appears in the modeling of phenotypic divergence of species in evolutionary biology [26]; and the hyper-parameter tuning of the elastic net [34], which is a widely-used regression method to derive a sparse and stable model from high-dimensional data.

This paper is organized as follows. Firstly, in Section 2, we present the main results (Theorem 2.1 and Proposition 2.3) and the primary application (Corollary 2.4). Implications of Theorem 2.1 are discussed with illustrative examples in Section 3. Section 4 is devoted to the proof of Theorem 2.1. In Section 5, some lemmas for the proofs of Corollaries 2.4 and 6.1 to 6.3 are prepared. In Section 6, Corollaries 6.1 to 6.3 and the proofs of Corollaries 2.4 and 6.1 to 6.3 are given. In Section 7, we investigate a multi-objective version of the location problem and the elastic net and discuss the structure of their sets of efficient solutions as applications of our result. Section 8 provides concluding remarks.

2. PRELIMINARIES AND THE STATEMENTS OF THE MAIN RESULTS

Throughout this paper, \( \mathbb{R}^m \) will be the Euclidean space of dimension \( m \geq 1 \). Unless otherwise stated, it is not necessary to assume that mappings are continuous.

Let \( I \) be a nonempty subset of \( M = \{1, \ldots, m\} \), where \( m \) is a positive integer. Let \( y = (y_1, \ldots, y_m) \) and \( y' = (y'_1, \ldots, y'_m) \) be two elements of \( \mathbb{R}^m \). The inequality \( y \leq_I y' \) (resp., \( y <_I y' \)) means that \( y_i \leq y'_i \) (resp., \( y_i < y'_i \)) for all \( i \in I \). The inequality \( y \preceq_I y' \) means that \( y_i \leq y'_i \) for all \( i \in I \) and there exists \( j \in I \) such that \( y_j < y'_j \).

Let \( Y \) be a subset of \( \mathbb{R}^m \). Let \( \text{Min}_I Y \) (resp., \( \text{WMin}_I Y \)) be the set consisting of all elements \( y' \in Y \) such that there does not exist any element \( y \in Y \) satisfying \( y \preceq_I y' \) (resp., \( y <_I y' \)). For simplicity, set \( \text{Min} Y = \text{Min}_M Y \) and \( \text{WMin} Y = \text{WMin}_M Y \). Then, the set \( \text{Min} Y \) (resp., \( \text{WMin} Y \)) is called the efficient set (resp., the weakly efficient set) of \( Y \).

For a subset \( Z \) of \( \mathbb{R}^m \), the set \( \mathbb{R}^m_{\geq 0} + Z \) is called the free disposal hull of \( Z \) (denoted by \( \text{FDH} Z \)), where

\[
\mathbb{R}^m_{\geq 0} = \{ (y_1, \ldots, y_m) \in \mathbb{R}^m \mid y_1 \geq 0, \ldots, y_m \geq 0 \}.
\]

For details on free disposal hulls, see [1]. A subset \( Z \) of \( \mathbb{R}^m \) is said to be convex if \( tx + (1-t)y \in Z \) for all \( x, y \in Z \) and all \( t \in [0, 1] \).

The main theorem of this paper is the following.

**Theorem 2.1.** Let \( Y \) be a subset of \( \mathbb{R}^m \). If the free disposal hull of \( Y \) is convex, then the following \((\alpha)\) and \((\beta)\) are equivalent:

\[(\alpha) \text{ WMin} Y = \text{Min} Y.\]

\[(\beta) \bigcup_{\emptyset \neq I \subseteq M} \text{Min}_I Y \subseteq \text{Min} Y, \text{ where } M = \{1, \ldots, m\}.\]
Remark 2.2. As in the proof of Theorem 2.1, the hypothesis that the free disposal hull of \( Y \) is convex is used only in the proof of (\( \beta \)) (see Section 4.2). In the proof of (\( \alpha \)) \( \Rightarrow \) (\( \beta \)) of Theorem 2.1, it is not necessary to assume that the free disposal hull of \( Y \) is convex (see Section 4.1).

Now, in order to state Proposition 2.3, we will prepare some definitions. Let \( f = (f_1, \ldots, f_m) : X \rightarrow \mathbb{R}^m \) be a mapping, and \( I = \{ i_1, \ldots, i_k \} \) (\( i_1 < \cdots < i_k \)) be a nonempty subset of \( M = \{ 1, \ldots, m \} \), where \( X \) is a given set and \( k \) is the number of the elements of \( I \). Let \( f_I : X \rightarrow \mathbb{R}^k \) be the mapping defined by \( f_I = (f_{i_1}, \ldots, f_{i_k}) \). A point \( x^* \in X \) is called an efficient solution (resp., a weakly efficient solution) to the following multi-objective optimization problem:

\[
\text{minimize } f_I(x) = (f_{i_1}(x), \ldots, f_{i_k}(x)),
\]

if \( f(x^*) \in \text{Min}_I f(X) \) (resp., \( f(x^*) \in \text{WMin}_I f(X) \)). By \( S(f_I, X) \) (resp., \( \text{WS}(f_I, X) \)), we denote the set consisting of all efficient solutions (resp., all weakly efficient solutions). Namely,

\[
S(f_I, X) = f^{-1}(\text{Min}_I f(X)),
\]

\[
\text{WS}(f_I, X) = f^{-1}(\text{WMin}_I f(X)).
\]

It is well known that a solution to a weighting problem is a weakly efficient solution (for example, see [19, Theorem 3.1.1 (p. 78)]). On the other hand, a solution to a weighting problem is not necessarily an efficient solution. For a given mapping \( f : X \rightarrow \mathbb{R}^m \), if \( \text{WS}(f, X) = S(f, X) \), then a solution to weighting problem is always an efficient solution. Therefore, characterizations of \( \text{WS}(f, X) = S(f, X) \) are useful and significant.

Proposition 2.3 is an application of Theorem 2.1 to multi-objective optimization problems.

Proposition 2.3. Let \( f = (f_1, \ldots, f_m) : X \rightarrow \mathbb{R}^m \) be a mapping, where \( X \) is a given set. If the free disposal hull of \( f(X) \) is convex, then the following (\( \alpha \)) and (\( \beta \)) are equivalent:

\[
(\alpha) \quad \text{WS}(f, X) = S(f, X),
\]

\[
(\beta) \quad \bigcup_{\emptyset \neq I \subseteq M} S(f_I, X) \subseteq S(f, X), \text{ where } M = \{ 1, \ldots, m \}.
\]

Notice that Proposition 2.3 is easily shown by setting \( Y = f(X) \) in Theorem 2.1.

From this proposition, we will drive four applications in this paper. The most important one is stated in Corollary 2.4 in this section (for the other applications, see Corollaries 6.1 to 6.3 in Section 6.2). In order to state Corollary 2.4, we will prepare some definitions.

Let \( X \) be a convex subset of \( \mathbb{R}^n \). A function \( f : X \rightarrow \mathbb{R} \) is said to be convex if

\[
f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)
\]

for all \( x, y \in X \) and all \( t \in [0, 1] \). A function \( f : X \rightarrow \mathbb{R} \) is said to be strongly convex if there exists \( \alpha > 0 \) satisfying

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

for all \( x, y \in X \) and all \( t \in [0, 1] \), where \( \|x-y\| \) denotes the Euclidean norm of \( x-y \). For details on convex functions and strongly convex functions, see [20]. A
mapping \( f = (f_1, \ldots, f_m) : X \to \mathbb{R}^m \) is said to be convex (resp., strongly convex) if every \( f_i \) is convex (resp., strongly convex).

Let \( f = (f_1, \ldots, f_m) : X \to \mathbb{R}^m \) be a mapping, where \( X \) is a set. Then, \( x^* \in X \) is called a strictly efficient solution if there does not exist \( x \in X \) (\( x \neq x^* \)) such that \( f_i(x) \leq f_i(x^*) \) for all \( i = 1, \ldots, m \). We denote the set of all strictly efficient solutions to the problem minimizing \( f \) by \( \text{SS}(f, X) \).

**Corollary 2.4.** Let \( X \) be a convex subset of \( \mathbb{R}^n \), and \( f : X \to \mathbb{R}^m \) be a strongly convex mapping. Then, we have
\[
\text{WS}(f, X) = S(f, X) = \text{SS}(f, X).
\]

3. **Illustration of Theorem 2.1**

In this section, we denote by \( \overline{xy} \) the line segment with end points \( x, y \in \mathbb{R}^2 \). First, we see an example that \((\beta)\) implies \((\alpha)\) in Theorem 2.1.

**Example 3.1.** Let us consider the situation shown in Figure 1. The domain \( Y \) in this case is defined by the convex hull of four points \( p_1 = (0, 1) \), \( p_2 = (1, 0) \), \( p_3 = (2, 1) \), \( p_4 = (1, 2) \), as shown in dark gray in the figure.

![Figure 1](image_url)

**Figure 1.** The condition \((\beta)\) implies \((\alpha)\) on convex FDH \( Y \).

It is easy to check that
\[
\min\{1\} Y = \{ p_1 \}, \quad \min\{2\} Y = \{ p_2 \}, \quad \text{WMin} Y = \text{Min} Y = p_1p_2.
\]

Since \( \min\{1\} Y \subseteq \min Y \) and \( \min\{2\} Y \subseteq \min Y \), we can see the condition \((\beta)\) in Theorem 2.1 holds. The free disposal hull of \( Y \) is the region shown in light gray in
the figure, which is a convex set. Thus, we can apply Theorem 2.1 and obtain \((\alpha)\). Actually, the condition \(W\text{Min} Y = \text{Min} Y\) holds in this example.

On the other hand, Example 3.2 shows that \(\text{FDH} Y\) is convex, but both \((\alpha)\) and \((\beta)\) do not hold.

Example 3.2. Let us consider the situation shown in Figure 2. The domain \(Y\) is the convex hull of four points \(p_1 = (0, 2), p_2 = (0, 1), p_3 = (1, 0), p_4 = (2, 0),\) as shown in dark gray in the figure. We have the same free disposal hull as in Example 3.1 which is a convex set, and thus we can apply Theorem 2.1 to this case.

\[
\begin{align*}
\text{Min}_{\{1\}} Y &= p_1p_2, \\
\text{Min}_{\{2\}} Y &= p_3p_4, \\
\text{Min} Y &= p_2p_3, \\
W\text{Min} Y &= p_1p_2 \cup p_2p_3 \cup p_3p_4.
\end{align*}
\]

Since \(\text{Min}_{\{1\}} Y \not\subseteq \text{Min} Y\), the condition \((\beta)\) in Theorem 2.1 does not hold. By Theorem 2.1 the condition \((\alpha)\) \(W\text{Min} Y = \text{Min} Y\) does not hold, as shown in the above equations.

The following example shows why the assumption of Theorem 2.1 is required.

Example 3.3. Let us consider the situation shown in Figure 3 where the domain \(Y\) is the nonconvex polygon with five vertices \(p_1 = (0, 3), p_2 = (1, 2), p_3 = (1, 1), p_4 = (2, 0), p_5 = (2, 3),\) shown in dark gray.

\begin{figure}[h]
\centering
\includegraphics[width=0.6\textwidth]{figure2.png}
\caption{The condition \(\neg(\beta)\) implies \(\neg(\alpha)\) on convex \(\text{FDH} Y\).}
\end{figure}
We can easily check:
\[
\begin{align*}
\text{Min}_{\{1\}} Y &= \{ p_1 \}, \\
\text{Min}_{\{2\}} Y &= \{ p_4 \}, \\
\text{Min} Y &= \{ p_1, p_2 \} \cup \{ p_3 \} \setminus \{ p_2 \}, \\
\text{WMin} Y &= \{ p_1, p_2 \} \cup \{ p_2, p_3 \} \cup \{ p_3, p_4 \}.
\end{align*}
\]
Since \(\text{Min}_{\{1\}} Y \subseteq \text{Min} Y\) and \(\text{Min}_{\{2\}} Y \subseteq \text{Min} Y\), the condition (\(\beta\)) holds. However, the free disposal hull of \(Y\) is a nonconvex set, as shown in light gray in the figure. Hence, we cannot apply Theorem 2.1 to this case. In such a case, (\(\alpha\)) can be false even if (\(\beta\)) is true. Actually, in this example, the condition (\(\alpha\)) does not hold as seen in the above equations.

In Theorem 2.1 the assumption (the free disposal hull of \(Y\) is convex) is not a necessary condition. In the following example, we will give a case where the free disposal hull is nonconvex but the condition (\(\alpha\)) holds (thus, (\(\beta\)) also holds).

**Example 3.4.** Let us consider the situation shown in Figure 4 where the domain \(Y\) is a nonconvex polygon with four vertices \(p_1 = (0, 3), p_2 = (2, 2), p_3 = (3, 0), p_4 = (3, 3)\), as shown in dark gray in the figure.

We can easily check:
\[
\begin{align*}
\text{Min}_{\{1\}} Y &= \{ p_1 \}, \\
\text{Min}_{\{2\}} Y &= \{ p_3 \}, \\
\text{WMin} Y &= \text{Min} Y = \{ p_1, p_2 \} \cup \{ p_2, p_3 \}.
\end{align*}
\]
4. Proof of Theorem 2.1

In the case $Y = \emptyset$, it is trivially seen that both $(\alpha)$ and $(\beta)$ hold. Hence, in what follows, we will consider the case $Y \neq \emptyset$.

4.1. Proof of $(\alpha) \Rightarrow (\beta)$. Let $I$ be a nonempty subset of $M$. Then, it is clearly seen that

\[ \text{Min}_I Y \subseteq \text{WMin}_I Y \subseteq \text{WMin} Y. \]

By $(\alpha)$, we get $\text{Min}_I Y \subseteq \text{Min} Y$. Thus, we have $(\beta)$. \hfill $\Box$

4.2. Proof of $(\beta) \Rightarrow (\alpha)$. It is sufficient to show that $\text{WMin} Y \subseteq \text{Min} Y$. Let $y^* = (y^*_1, \ldots, y^*_m) \in \text{WMin} Y$ be an arbitrary element. Set

\[
\begin{align*}
A &= \{ (y_1 - y^*_1, \ldots, y_m - y^*_m) \in \mathbb{R}^m \mid (y_1, \ldots, y_m) \in \text{FDH} Y \}, \\
B &= \{ y \in \mathbb{R}^m \mid y <_M 0 \}.
\end{align*}
\]

Here, note that $0 = (0, \ldots, 0) \in \mathbb{R}^m$ in the above description of $B$. Then, we will have $A \cap B = \emptyset$ by contradiction. Suppose that $A \cap B \neq \emptyset$. Then, there exist $y' \in Y$ and $z \in \mathbb{R}_{\geq 0}^m$ satisfying $y' + z - y^* <_M 0$. Since $z \in \mathbb{R}_{\geq 0}^m$, we get $y' - y^* <_M 0$. This contradicts $y^* \in \text{WMin} Y$. Hence, we have $A \cap B = \emptyset$. 

In the following lemma, $(,)\text{ stands for the inner product in } \mathbb{R}^m.$

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{The condition ($\beta$) holds, and ($\alpha$) does on nonconvex FDH $Y$.}
\end{figure}
Lemma 4.1 (Separation theorem [17]). Let $D_1$ and $D_2$ be nonempty convex subsets of $\mathbb{R}^m$ satisfying $D_1 \cap D_2 = \emptyset$. Then, there exist $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$ ($a \neq 0$) and $b \in \mathbb{R}$ such that the following both assertions hold.

(1) For any $y \in D_1$, we have $\langle a, y \rangle \geq b$.

(2) For any $y \in D_2$, we have $\langle a, y \rangle \leq b$.

Note that $A$ and $B$ are nonempty convex subsets of $\mathbb{R}^m$. Hence, by Lemma 4.1, there exist $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$ ($a \neq 0$) and $b \in \mathbb{R}$ such that the following both assertions hold.

(1') For any $y \in A$, we have $\langle a, y \rangle \geq b$.

(2') For any $y \in B$, we have $\langle a, y \rangle \leq b$.

Then, we will show that $b = 0$. Since $0 = (0, \ldots, 0, 0) \in A$, we have $\langle a, 0 \rangle \geq b$ by (1'). Namely, we get $b \leq 0$. Since $(-\varepsilon, \ldots, -\varepsilon) \in B$ for any sufficiently small $\varepsilon > 0$, it is clearly seen that $b = 0$ by (2').

We will show that $a_i \geq 0$ for any $i \in M$ by contradiction. Suppose that there exists an element $i' \in M$ satisfying $a_{i'} < 0$. Let $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$ be the element given by

$$y_i = \begin{cases} -1 & \text{if } i \neq i', \\ \frac{\sum_{j=1, j \neq i'}^m |a_j| + 1}{a_{i'}} & \text{if } i = i'. \end{cases}$$

Then, we get $y \in B$ and $\langle a, y \rangle > 0$. This contradicts (2'). Hence, it follows that $a_i \geq 0$ for any $i \in M$.

Now, set

$$I = \{ i \in M \mid a_i > 0 \}.$$

Notice that $I \neq \emptyset$. Set $I = \{ i_1, \ldots, i_k \}$, where $k$ is an integer $(1 \leq k \leq m)$ and $i_1 < \cdots < i_k$.

We will show that $y^* \in \text{Min}_I Y$. Let $y = (y_1, \ldots, y_m) \in Y$ be any element. Since $(y_1 - y_1^*, \ldots, y_m - y_m^*) \in A$ and $b = 0$, by (1'), we have

$$a_{i_1} (y_{i_1} - y_{i_1}^*) + \cdots + a_{i_k} (y_{i_k} - y_{i_k}^*) \geq 0.$$ 

Since $a_{i_1} > 0, \ldots, a_{i_k} > 0$, the element $y \in Y$ does not satisfy $y \leq_I y^*$. Therefore, we obtain $y^* \in \text{Min}_I Y$. By the assumption (\beta), it follows that $y^* \in \text{Min}_I Y$.

5. Lemmas for the applications of Proposition 2.3

Firstly, we will give the following well-known result. For the sake of the readers’ convenience, we also give the proof.

Lemma 5.1. Let $X$ be a convex subset of $\mathbb{R}^n$, and $f : X \to \mathbb{R}^m$ be a convex mapping. Then, the free disposal hull of $f(X)$ is convex.

Proof of Lemma 5.1. Let $y = (y_1, \ldots, y_m)$, $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_m) \in \text{FDH} f(X)$ be arbitrary points and $t \in [0, 1]$ be an arbitrary element. Then, there exist $x \in X$ (resp., $\bar{x} \in X$) and $z = (z_1, \ldots, z_m) \in \mathbb{R}^m_+ \ (\text{resp., } \bar{z} = (\bar{z}_1, \ldots, \bar{z}_m) \in \mathbb{R}^m_+)$ such that $y = f(x) + z$ (resp., $\bar{y} = f(\bar{x}) + \bar{z}$). Let $i$ be an arbitrary integer satisfying $1 \leq i \leq m$. Since $f_i$ is convex, we have

$$f_i(tx + (1-t)\bar{x}) \leq tf_i(x) + (1-t)f_i(\bar{x}),$$
where \( f = (f_1, \ldots, f_m) \). Since \( z_i \geq 0 \) and \( \tilde{z}_i \geq 0 \) for any \( i = 1, \ldots, m \), we also get
\[
tf_i(x) + (1 - t)f_i(\tilde{x}) \leq tf_i(x + z_i) + (1 - t)(f_i(\tilde{x} + \tilde{z}_i) = ty_i + (1 - t)\tilde{y}_i.
\]
Hence, we obtain
\[
f_i(tx + (1 - t)x) \leq ty_i + (1 - t)\tilde{y}_i.
\]
Since we have \[(5.1)\]
for any integer \( i \) satisfying \( 1 \leq i \leq m \), it follows that \( ty + (1 - t)\tilde{y} \in \text{FDH } f(X) \). \[\square\]

In the following, for two sets \( U, V \), and a subset \( W \) of \( U \), the restriction of a given mapping \( g : U \to V \) to \( W \) is denoted by \( g|_W : W \to V \).

**Lemma 5.2.** Let \( f : X \to \mathbb{R}^m \) be a mapping, where \( X \) is a given set. Let \( I = \{i_1, \ldots, i_k\} \) (\( i_1 < \cdots < i_k \)) be a nonempty subset of \( M = \{1, \ldots, m\} \), where \( k \) is the number of the elements of \( I \). If \( f|_{S(f_i, X)} : S(f_i, X) \to \mathbb{R}^k \) is injective, then we have \( S(f, X) \subseteq S(f, X) \).

**Proof of Lemma 5.2.** Suppose that there exists an element \( x \in S(f_i, X) \) such that \( x \not\in S(f, X) \). Then, there exists an element \( y \in X \) (\( y \not= x \)) satisfying \( f_i(y) \leq f_i(x) \) for any \( i \in I \). Since \( I \subseteq M \), it follows that \( f_i(y) \leq f_i(x) \) for any \( i \in I \). Since \( x \in S(f_i, X) \), we get \( f_i(x) = f_i(y) \). Therefore, we have \( y \in S(f_i, X) \). This contradicts the assumption that \( f|_{S(f_i, X)} : S(f_i, X) \to \mathbb{R}^k \) is injective. \[\square\]

**Lemma 5.3.** Let \( X \) be a convex subset of \( \mathbb{R}^n \), and \( f : X \to \mathbb{R}^m \) be a strongly convex mapping. Then, \( f|_{S(f_i, X)} : S(f_i, X) \to \mathbb{R}^m \) is injective.

**Proof of Lemma 5.3.** Suppose that \( f|_{S(f_i, X)} : S(f_i, X) \to \mathbb{R}^m \) is not injective. Then, there exist \( x, y \in S(f_i, X) \) such that \( x \not= y \) and \( f|_{S(f_i, X)}(x) = f|_{S(f_i, X)}(y) \). Let \( i \) be an arbitrary integer satisfying \( 1 \leq i \leq m \). Since \( f = (f_1, \ldots, f_m) \) is strongly convex, there exists \( \alpha_i > 0 \) satisfying
\[
f_i(tx + (1 - t)y) \leq tf_i(x) + (1 - t)f_i(y) - \frac{1}{2}\alpha_i t(1 - t)\|x - y\|^2
\]
for the points \( x, y \in S(f_i, X) \) and all \( t \in [0, 1] \). Set \( t = \frac{1}{2} \). Then, we get
\[
f_i\left(\frac{x + y}{2}\right) \leq \frac{f_i(x) + f_i(y)}{2} - \frac{\alpha_i}{8}\|x - y\|^2.
\]
Since \( f|_{S(f_i, X)}(x) = f|_{S(f_i, X)}(y) \), we have
\[
f_i\left(\frac{x + y}{2}\right) \leq f_i(x) - \frac{\alpha_i}{8}\|x - y\|^2.
\]
Since \( x \not= y \) and \( \alpha_i > 0 \), it follows that
\[
f_i\left(\frac{x + y}{2}\right) < f_i(x).
\]
This contradicts \( x \in S(f_i, X) \). \[\square\]

6. **Proof of Corollary 2.4 and Other Applications of Proposition 2.3**

We will show Corollary 2.4 in Section 6.1. Other applications of Proposition 2.3 are given in Section 6.2.
6.1. Proof of Corollary 2.4. Since \( f|_{S(f,X)} \) is injective by Lemma 5.3, it is not hard to see that \( S(f,X) = SS(f,X) \).

Now, we will show that \( WS(f,X) = S(f,X) \). Since \( f \) is strongly convex, by Lemma 5.1, the set \( \text{FDH}(f) \) is convex. Thus, by Proposition 2.3 in order to show that \( WS(f,X) = S(f,X) \), it is sufficient to show

\[
\bigcup_{\emptyset \neq I \subseteq M} S(f_I,X) \subseteq S(f,X),
\]

where \( M = \{1, \ldots, m\} \). Let \( I \) be any nonempty subset of \( M \). Since \( f|_{S(f_I,X)} : S(f_I,X) \rightarrow \mathbb{R}^k \) is strongly convex, by Lemma 5.3, the mapping \( f|_{S(f_I,X)} \) is injective, where \( k \) is the number of the elements of \( I \). By Lemma 5.2, we have \( S(f_I,X) \subseteq S(f,X) \). Thus, we obtain (6.1).

6.2. Other applications of Proposition 2.3. Proposition 2.3 gives a characterization of the equality of the weak efficiency and the efficiency for possibly nonconvex problems having a convex image \( f(X) \) as follows:

**Corollary 6.1.** Let \( f = (f_1, \ldots, f_m) : X \rightarrow \mathbb{R}^m \) be a mapping, where \( X \) is a given set. If \( f(X) \) is convex, then the following (\( \alpha \)) and (\( \beta \)) are equivalent:

(\( \alpha \)) \( WS(f,X) = S(f,X) \).

(\( \beta \)) \[
\bigcup_{\emptyset \neq I \subseteq M} S(f_I,X) \subseteq S(f,X), \text{ where } M = \{1, \ldots, m\}.
\]

**Proof of Corollary 6.1.** Since \( f(X) \) is convex, it is clearly seen that the free disposal hull of \( f(X) \) is also convex. Thus, by Proposition 2.3, we have Corollary 6.1. \( \square \)

Unfortunately, it is not easy to check the convexity of the image of a given mapping which is possibly nonconvex. A more workable condition ensuring this characterization is the convexity of a given mapping itself.

**Corollary 6.2.** Let \( X \) be a convex subset of \( \mathbb{R}^n \) and \( f = (f_1, \ldots, f_m) : X \rightarrow \mathbb{R}^m \) be a convex mapping. Then, the following (\( \alpha \)) and (\( \beta \)) are equivalent:

(\( \alpha \)) \( WS(f,X) = S(f,X) \).

(\( \beta \)) \[
\bigcup_{\emptyset \neq I \subseteq M} S(f_I,X) \subseteq S(f,X), \text{ where } M = \{1, \ldots, m\}.
\]

**Proof of Corollary 6.2.** Since \( f \) is convex, by Lemma 5.1, the free disposal full of \( f(X) \) is convex. Therefore, by Proposition 2.3, we get Corollary 6.2. \( \square \)

As a direct consequence, the characterization is valid for convex programming problems.

**Corollary 6.3.** Let \( X \) be a convex subset of \( \mathbb{R}^n \) and \( f = (f_1, \ldots, f_m) : X \rightarrow \mathbb{R}^m \) be a convex mapping. Let \( g_1, \ldots, g_\ell \) be convex functions of \( X \) into \( \mathbb{R} \), where \( \ell \) is a positive integer. Set

\[
\Omega = \{ x \in X \mid g_1(x) \leq 0, \ldots, g_\ell(x) \leq 0 \}.
\]

Then, the following (\( \alpha \)) and (\( \beta \)) are equivalent:

(\( \alpha \)) \( WS(f|_{\Omega},\Omega) = S(f|_{\Omega},\Omega) \).

(\( \beta \)) \[
\bigcup_{\emptyset \neq I \subseteq M} S(f_{I|_{\Omega}},\Omega) \subseteq S(f|_{\Omega},\Omega), \text{ where } M = \{1, \ldots, m\}.
\]
(\beta) \bigcup_{\emptyset \neq I \subseteq M} S((f|_I) \Omega) \subseteq S(f|_\Omega, \Omega), \text{ where } M = \{1, \ldots, m\}.

Proof of Corollary 6.3 Since \(g_1, \ldots, g_\ell\) are convex functions, it is clearly seen that \(\Omega\) is convex. Since the mapping \(f|_\Omega : \Omega \rightarrow \mathbb{R}^m\) is convex, by Corollary 6.2, we get Corollary 6.3. \(\square\)

7. Practical applications of Corollary 2.4

In this section, we apply Corollary 2.4 to two practical problems and demonstrate its usefulness. Throughout this section, the standard \((m-1)\)-simplex and its \(I\)-face (\(\emptyset \neq I \subseteq M\)) are denoted by

\[
\Delta^{m-1} = \left\{ w = (w_1, \ldots, w_m) \in \mathbb{R}^m \left| \sum_{i=1}^{m} w_i = 1, w_i \geq 0 \right. \right\},
\]

\[
\Delta_I^{m-1} = \left\{ w = (w_1, \ldots, w_m) \in \Delta^{m-1} \left| w_i = 0 \ (i \notin I) \right. \right\},
\]

respectively.

7.1. Location problems. Let \(p_1, \ldots, p_m\) be points in \(\mathbb{R}^n\) and \(A_1, \ldots, A_m\) be symmetric positive definite \(n \times n\) real matrices. The multi-objective location problem under squared Mahalanobis distances is defined by

\[
\begin{align*}
\text{minimize} & \quad f(x) := (f_1(x), \ldots, f_m(x)) \\
\text{where} & \quad f_i(x) = (x - p_i)^T A_i (x - p_i) \quad (i = 1, \ldots, m).
\end{align*}
\]

(7.1)

The mapping \(f : \mathbb{R}^n \rightarrow \mathbb{R}^m\) in (7.1) is strongly convex since each component function \(f_i : \mathbb{R}^n \rightarrow \mathbb{R}\) is an affine transformation of the squared Euclidean norm \(\|x\|^2 = \sum_{j=1}^{n} x_j^2\), which is strongly convex by definition. It is known [14, Corollary 1] that any convex problem—thus the problem (7.1)—satisfies the following equation:

\[
\text{WS}(f, \mathbb{R}^n) = \bigcup_{\emptyset \neq I \subseteq M} S(f|_I, \mathbb{R}^n),
\]

where \(M = \{1, \ldots, m\}\). The property (7.2) is called the Pareto reducibility [22, Definition 1]. Furthermore, it suffices to solve at most \((n+1)\)-objective subproblems [31, Theorem 2]:

\[
\text{WS}(f, \mathbb{R}^n) = \bigcup_{\emptyset \neq I \subseteq M} S(f|_I, \mathbb{R}^n),
\]

where \(|I|\) is the number of the elements of \(I\). Applying Corollary 2.4 to (7.3), we have

\[
S(f, \mathbb{R}^n) = \bigcup_{\emptyset \neq I \subseteq M} S(f|_I, \mathbb{R}^n),
\]

(7.4)

This stronger version of Pareto reducibility is more useful in the sense that efficient solutions to subproblems never become inefficient to the original problem. Each

\footnote{We consider squared distances for differentiability. Since the Mahalanobis distance is nonnegative, its square has the same ordering and thus preserves the Pareto set of each subproblem.}
S(f₁, \mathbb{R}ⁿ) (\emptyset \neq I \subseteq M) can be easily obtained in a way described below. Since f in (7.1) is a strongly convex \( C^∞ \)-mapping, its weighted-sum scalarization

\[ h_w(x) := \sum_{i=1}^{m} w_i f_i(x) \quad (\text{given } w \in \Delta^{m-1}) \]

is a strongly convex \( C^∞ \)-function. The problem minimizing \( h_w \) has a unique solution for every \( w \in \Delta^{m-1} \) (see [20] Theorem 2.2.6 (p. 85)). We denote this minimizing solution by \( \arg\min_{x \in \mathbb{R}ⁿ} h_w(x) \). Thus we can define the mapping \( x^* : \Delta^{m-1} \rightarrow S(f, \mathbb{R}ⁿ) \) given by

\[ x^*(w) := \arg\min_{x \in \mathbb{R}ⁿ} h_w(x), \]

which is a \( C^∞ \)-surjection satisfying

\[ x^*(\Delta^{m-1}) = S(f₁, \mathbb{R}ⁿ) \]

for each \( I \) satisfying \( \emptyset \neq I \subseteq M \) (see [7] Theorems 1.1 and 3.1). While we cannot express (7.5) in a closed form, the mapping \( x^* : \Delta^{m-1} \rightarrow S(f, \mathbb{R}ⁿ) \) can be approximated by a Bézier simplex [11] Theorem 3.1. For numerical computation, see [27].

Let us compare our result to related ones in the literature. The Pareto reducibility bounded with at most \( (n + 1) \)-objective subproblems (7.3) is shown in various settings: for convex problems [14, 31] as mentioned above, strictly quasiconvex problems with upper semi-continuity along line segments [16] Theorem 3.2, and lexicographic quasiconvex problems with upper semi-continuity along line segments [24] Theorem 4.4. However, the condition WS(f, \mathbb{R}ⁿ) = S(f, \mathbb{R}ⁿ) is not guaranteed in any of those cases, and thus some S(f₁, \mathbb{R}ⁿ) in (7.3) may contain inefficient solutions.

If \( n = 2 \) and \( A_i \) is the identity matrix for all \( i = 1, \ldots, m \), then (7.1) becomes the classical location problem for which Kuhn [12] Theorem 4.4 showed that S(f, \mathbb{R}²) is the convex hull of \( p₁, \ldots, p_m \). His result can be easily extended to the case \( n > 2 \) [31]. However, for non-identity \( A_i \), the set S(f, \mathbb{R}ⁿ) is not necessarily the convex hull of \( p₁, \ldots, p_m \) since S(f, \mathbb{R}ⁿ) in this case can be nonconvex. Let us examine the following mapping \( f = (f₁, f₂, f₃) : \mathbb{R}³ \rightarrow \mathbb{R}³ \), which is adopted from [7] Example 3.5:

\[ f₁(x, y, z) = x^2 + (-y + x)^2 + z^2, \]
\[ f₂(x, y, z) = 2(x - 1)^2 + (-y + x - 1)^2 + z^2, \]
\[ f₃(x, y, z) = (x - 2)^2 + (y + x - 2)^2 + z^2. \]

It can be rewritten in the form of the location problem (7.1):

\[ A₁ = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A₂ = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A₃ = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \]
\[ p₁ = (0 \ 0 \ 0)^T, \quad p₂ = (1 \ 0 \ 0)^T, \quad p₃ = (2 \ 0 \ 0)^T. \]

The set of efficient solutions is nonconvex as shown in Figure 5.
Another result that partially overlaps with our result is \cite[Theorem 4.2]{4}. This theorem implies \eqref{7.4} for any strongly convex mapping \( f : \mathbb{R}^2 \to \mathbb{R}^m \) since \( S(f_I, \mathbb{R}^n) = SS(f_I, \mathbb{R}^n) \) \((\emptyset \neq I \subseteq M)\) holds, including a spacial case of \eqref{7.1} where we set \( n = 2 \) and choose \( m \) and \( A_1, \ldots, A_m \) arbitrarily. However, the authors of \cite{4} gave a counter example to show that their result cannot be extended to the case \( n > 2 \). Therefore, the unique strength of Corollary 2.4 among existing results is to derive \eqref{7.4} in the case where \( n > 2 \) and \( A_i \) is not identity for some \( i = 1, \ldots, m \). In practice, such a case appears in the modeling of phenotypic divergence of species in evolutionary biology \cite{26} (see also \cite[Section 5.2]{7}).

7.2. Elastic net. The elastic net \cite{34} is a sparse modeling method that is originally a single-objective problem but can be reformulated as a multi-objective one. Let us consider a linear regression model:

\[
y = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n + \delta,
\]

where \( x_i \) and \( \theta_i \) \((i = 1, \ldots, n)\) are a predictor and its coefficient, \( y \) is a response to be predicted, and \( \delta \) is a Gaussian noise. Given a matrix \( X \) with \( m \) rows of observations and \( n \) columns of predictors, a row vector \( y \) of \( m \) responses, the (original) elastic net regressor is the solution to the following problem:

\[
\min_{\theta \in \mathbb{R}^n} g_{\mu, \lambda}(\theta) := \frac{1}{2m} \|X\theta - y\|^2 + \mu |\theta| + \frac{\lambda}{2} \|\theta\|^2,
\]

where \( \|\cdot\| \) is the \( \ell_2 \)-norm, \( |\cdot| \) is the \( \ell_1 \)-norm, and \( \mu, \lambda \) are fixed non-negative numbers for regularization. Note that with \( \mu = \lambda = 0 \), the problem \eqref{7.6} reduces to the ordinary least squares (OLS) regression; with \( \mu > 0 \) and \( \lambda = 0 \), it turns into the lasso regression \cite{28}, which find a sparse solution that suppresses ineffective predictors; with \( \mu = 0 \) and \( \lambda > 0 \), it becomes the ridge regression \cite{9}, which finds a stable solution against multicollinear predictors. Thus the elastic net regression, with \( \mu > 0 \) and \( \lambda > 0 \), inherits both of the lasso and ridge properties. Choosing
appropriate values for \( \mu \) and \( \lambda \) involves a 2-D black-box search on an unbounded domain, which often requires a great deal of computational effort.

In order to avoid such a costful hyper-parameter search, we reformulate the problem into a multi-objective strongly convex one and take the same approach as in Section 7.1: first, consider its weighting problem; next, make an approximation of the weight-solution mapping (which requires fewer models to train than the original hyper-parameter search does); then, compare possible models on the approximation and find the best weight (which is computationally cheap and does not require additional training); and finally, send the best weight back to a hyper-parameter in the original problem.

For this purpose, we separate the OLS term and the regularization terms into individual objective functions:

\[
 f_1(\theta) = \frac{1}{2m} \|X\theta - y\|^2, \quad f_2(\theta) = |\theta|, \quad f_3(\theta) = \frac{1}{2} \|\theta\|^2.
\]

The functions \( f_1 \) and \( f_2 \) are convex but may not be strongly convex. We add a small amount of \( f_3 \) values to each function, making them strongly convex:

\[
 \minimize_{\theta \in \mathbb{R}^n} \tilde{f}(\theta) := (\tilde{f}_1(\theta), \tilde{f}_2(\theta), \tilde{f}_3(\theta))
\]

where \( \tilde{f}_i(\theta) = f_i(\theta) + \epsilon f_3(\theta) \) \((i = 1, 2, 3)\).

In (7.7), we assume that \( \epsilon \) is a positive real number. Hence, the mapping \( \tilde{f} \) in (7.7) is a strongly convex \( C^0 \)-mapping.

Now let us consider how to obtain the whole set of weakly efficient solutions to (7.7). As is the case of the location problem discussed in Section 7.1, the mapping \( \tilde{f} : \mathbb{R}^n \to \mathbb{R}^3 \) in (7.7) is a strongly convex \( C^0 \)-mapping, and thus the weighting problem

\[
 \minimize_{\theta \in \mathbb{R}^n} h_w(\theta) := w_1 \tilde{f}_1(\theta) + w_2 \tilde{f}_2(\theta) + w_3 \tilde{f}_3(\theta)
\]

has a unique solution for every weight \( w = (w_1, w_2, w_3) \in \Delta^2 \). We denote this solution by \( \arg \min_{\theta \in \mathbb{R}^n} h_w(\theta) \). Thus, one can again define a mapping \( \theta^* : \Delta^2 \to \mathbb{R}^n \) by

\[
 \theta^*(w) := \arg \min_{\theta \in \mathbb{R}^n} h_w(\theta).
\]

Unlike the location problem, the mapping \( \tilde{f} \) in this case is of class \( C^0 \). Thus [7, Theorems 1.1 and 3.1], which are valid for \( C^2 \)-mappings, are no longer available for deriving properties of \( \theta^* \). Instead, we apply [19, Theorem 3.1.1 (p. 78)] and have

\[
 \theta^*(\Delta^2_I) \subseteq \text{WS}(\tilde{f}_I, \mathbb{R}^n) \text{ for all } I \text{ satisfying } \emptyset \neq I \subseteq \{1, 2, 3\}.
\]

It is made stronger by Corollary 2.4

\[
 \theta^*(\Delta^2_I) \subseteq S(\tilde{f}_I, \mathbb{R}^n) \text{ for all } I \text{ satisfying } \emptyset \neq I \subseteq \{1, 2, 3\}.
\]

By [19, Theorem 3.1.4 (p. 79)], it follows that

\[
 \theta^*(\Delta^2_I) = S(\tilde{f}_I, \mathbb{R}^n) \text{ for all } I \text{ satisfying } \emptyset \neq I \subseteq \{1, 2, 3\}.
\]
Note that for any \( w = (w_1, w_2, w_3) \in \Delta^2 \setminus \Delta^2_{12,3} \), the point \( \theta^*(w) \) is the minimizer of the function \( g_{\mu(w), \lambda(w)} \) in (7.6), where
\[
\mu(w) = \frac{w_2}{w_1}, \\
\lambda(w) = \frac{w_3 + \varepsilon}{w_1},
\]
and \( \varepsilon \) is given in (7.7). Here, the equations in (7.8) are easily obtained by comparing (7.6) and (7.7).

**Remark 7.1.** If \( \theta^* \) is continuous, then it can be approximated by a Bézier simplex as discussed in Section 7.1. However, the continuity of \( \theta^* \) for a strongly convex \( C^0 \)-mapping \( \tilde{f} \) is currently an open problem, although it is true for a strongly convex \( C^2 \)-mapping (for the details on this result for a strongly convex \( C^2 \)-mapping, see [7]). To investigate the truthiness of the above assumption, the authors conducted

![Figure 6](image-url)

**Figure 6.** Numerical computation of the multi-objective elastic net (with 5151 points).

a numerical computation of \( \theta^* \) with some dataset (see Appendix A for detailed settings). As shown in Figure 6, the authors made a set \( W \) of 5151 sampling points taken from \( \Delta^2 \) (Figure 6a) and computed its image under \( \theta^* \) (Figures 6c and 6d).
and $\tilde{f} \circ \theta^*$ (Figure 6b), where the color of points indicates the correspondence in those mappings. Since the color changes continuously in Figures 6c and 6d (resp. Figure 6b), the mapping $\theta^*$ (resp. $\tilde{f} \circ \theta^*$) for this dataset seems to be continuous.

Motivated by the above observation, the authors applied a Bézier simplex fitting method [27] to this sample and obtained an approximation of $\theta^*$ and $\tilde{f} \circ \theta^*$ (Figure 7). As shown in the figure, the mappings $\theta^*$ and $\tilde{f} \circ \theta^*$ are accurately approximated. See also Appendix A for detailed settings. Once we have obtained an approximation of $\theta^*$ and $\tilde{f} \circ \theta^*$, we can explore the best weight on the approximation with little computational cost. By (7.8), such a weight is converted to a hyper-parameter for the original elastic net (7.6).

**Remark 7.2.** To find the best weight on the approximation, weights corresponding to typical hyper-parameters will help for comparison. If we have some hyper-parameters in the original problem, then transformation from a hyper-parameter $(\mu, \lambda)$ to a weight $(w_1, w_2, w_3)$ gives such “guiding” weights. For any $\mu, \lambda$ such that

\begin{equation}
0 \leq \mu \leq \frac{\lambda - \varepsilon}{\varepsilon},
\end{equation}
the minimizer of the function $g_{\mu,\lambda}$ in (7.6) is the point $\theta^*(w(\mu, \lambda))$, where $w(\mu, \lambda) = (w_1(\mu, \lambda), w_2(\mu, \lambda), w_3(\mu, \lambda))$ is defined by

$$
\begin{align*}
w_1(\mu, \lambda) &= \frac{1 + \varepsilon}{\lambda + \mu + 1}, \\
w_2(\mu, \lambda) &= \frac{(1 + \varepsilon)\mu}{\lambda + \mu + 1}, \\
w_3(\mu, \lambda) &= \frac{\lambda - \varepsilon(\mu + 1)}{\lambda + \mu + 1}.
\end{align*}
$$

(7.10)

Here, by (7.9), note that $\lambda + \mu + 1 \neq 0$ and $w(\mu, \lambda) \in \Delta^2 \setminus \Delta^2_{\{2, 3\}}$ (see Appendix B which also contains the derivation of (7.9) and (7.10)).

8. Conclusion

In this paper, we have given a characterization of the equality of weak efficiency and efficiency when the free disposal hull of the domain is convex. An important problem class in which this equality holds is the strongly convex problems. By this fact, we have seen that the location problem under Mahalanobis distances has a stronger version of the Pareto reducibility, which makes the problem easier to solve. It has also been seen that the hyper-parameter tuning of the elastic net can be reformulated as a multi-objective optimization problem, and this reformulation allows us to obtain an approximation of the whole set of efficient solutions, which contains the all possible models trained with varying hyper-parameters.

The scope of the multi-objective reformulation discussed in Section 7.2 is not limited to the elastic net. The same idea can be applied to a wide range of sparse modeling methods, including the group lasso [32], the fused lasso [29], the graphical lasso [6], the smooth lasso [8], and their elastic net counterparts.

Acknowledgements

The authors are grateful to Kenta Hayano, Yutaro Kabata, and Hiroshi Teramoto for their kind comments. S. Ichiki was supported by JSPS KAKENHI Grant Number JP19J00650. This work is based on the discussions at 2018 IMI Joint Use Research Program, Short-term Joint Research “multi-objective optimization and singularity theory: Classification of Pareto point singularities”.

References

[1] Saleh Abdullah R. Al-Mezel, Falleh Rajallah M. Al-Solamy, and Qamrul Hasan Ansari. Fixed Point Theory, Variational Analysis, and Optimization. CRC Press, Boca Raton, FL, 2014.
[2] J. Benoist and N. Popovici. The structure of the efficient frontier of finite-dimensional completely-shaded sets. Journal of Mathematical Analysis and Applications, 250(1):98–117, 2000.
[3] Indraneel Das and J. E. Dennis. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. SIAM Journal on Optimization, 8(3):631–657, 1998.
[4] Matthias Ehrgott and Stefan Nickel. On the number of criteria needed to decide Pareto optimality. Mathematical Methods of Operations Research, 55(3):329–345, Jun 2002.
[5] G. Eichfelder. Adaptive Scalarization Methods in Multiobjective Optimization. Springer-Verlag, Berlin, Heidelberg, 2008.
[6] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3):432–441, 12 2007.
[7] Naoki Hamada, Kenta Hayano, Shunsuke Ichiki, Yutaro Kabata, and Hiroshi Teramoto. Topology of Pareto sets of strongly convex problems. ArXiv e-prints, 2019. [http://arxiv.org/abs/1904.03615]

[8] Mohamed Hebiri and Sara van de Geer. The smooth-lasso and other $\ell_1 + \ell_2$-penalized methods. Electron. J. Statist., 5:1184–1226, 2011.

[9] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55–67, 1970.

[10] David W Hosmer and Stanley Lemeshow. Applied Logistic Regression. Wiley, New York, 1989.

[11] 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, volume 33 of AAAI-19, pages 2304–2313. AAAI Press, 2019.

[12] Harold W Kuhn. On a pair of dual nonlinear programs. Nonlinear Programming, 1:38–45, 1967.

[13] D. La Torre and N. Popovici. Arcwise cone-quasiconvex multicriteria optimization. Operations Research Letters, 38(2):143–146, 2010.

[14] T.J. Lowe, J.-F. Thisse, J.E. Ward, and R.E. Wendell. On efficient solutions to multiple objective mathematical programs. Management Science, 30(11):1346–1349, 1984.

[15] D.T. Luc. Theory of Vector Optimization, volume 319 of Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 1989.

[16] C. Malivert and N. Boissard. Structure of efficient sets for strictly quasi convex objectives. Journal of Convex Analysis, 1:143–150, 1994.

[17] Jiří Matoušek. Lectures on discrete geometry, volume 212 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2002.

[18] Achille Messac and Christopher A. Mattson. Normal constraint method with guarantee of even representation of complete Pareto frontier. AIAA Journal, 42(10):2101–2111, 2004.

[19] Kaisa Miettinen. Nonlinear Multiobjective Optimization, volume 12 of International Series in Operations Research & Management Science. Springer-Verlag, GmbH, 1999.

[20] Yurii Nesterov. Introductory Lectures on Convex Optimization: A Basic Course. Kluwer Academic Publishers, 2004.

[21] A. Pascoletti and P. Serafini. Scalarizing vector optimization problems. Journal of Optimization Theory and Applications, 42:499–524, 1984.

[22] N. Popovici. Pareto reducible multicriteria optimization problems. Optimization, 54(3):253–263, 2005.

[23] N. Popovici. Structure of efficient sets in lexicographic quasiconvex multicriteria optimization. Operations Research Letters, 34(2):142–148, 2006.

[24] N. Popovici. Involving the Helly number in Pareto reducibility. Operations Research Letters, 36(2):173–176, 2008.

[25] H. Sato. Inverted PBI in MOEA/D and its impact on the search performance on multi and many-objective optimization. In Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation, GECCO ’14, pages 645–652, New York, NY, USA, 2014. ACM.

[26] 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.

[27] Akinori Tanaka, Akiyoshi Sannai, Ken Kobayashi, and Naoki Hamada. Asymptotic risk of Bézier simplex fitting. ArXiv e-prints, 2019. [http://arxiv.org/abs/1904.03615]

[28] Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996.

[29] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparcity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(1):91–108, 2005.

[30] W. N. Venables and B. D. Ripley. Modern Applied Statistics with S. Springer, fourth edition, 2002.

[31] J. Ward. Structure of efficient sets for convex objectives. Mathematics of Operations Research, 14(2):249–257, 1989.
Appendix A. Experimental settings in Remark 7.1

In order to numerically examine the mapping $\theta^*$ in Section 7.2, the authors applied the elastic net to a real dataset. The authors adapted the Birthwt dataset in the R package MASS, which contains 189 births at the Baystate Medical Centre, Springfield, Massachusetts during 1986 [10,30]. The dataset was modified as follows: continuous variables related to a mother’s age and weight ($age_1$, $age_2$, $age_3$, $lwt_1$, $lwt_2$, $lwt_3$) and a baby’s weight ($x$) were treated as predictors and a response, respectively. The other variables, which are “one-hot” flags, were discarded. The predictors and response were normalized and fed to the elastic net. The mappings the authors were numerically computing are

$$\theta^* : \Delta^2 \rightarrow S(\tilde{f}, R^6),$$

$$\tilde{f} \circ \theta^* : \Delta^2 \rightarrow \operatorname{Min} \tilde{f}(R^6)(\subset R^3).$$

The authors sampled 5151 grid points on $\Delta^2$ as follows:

$$w = \frac{1}{100}(n_1, n_2, n_3) \text{ such that } n_1, n_2, n_3 \in \{0, \ldots, 100\}, \ n_1 + n_2 + n_3 = 100.$$ 

For each point $w$, we computed the value of $\theta^*(w)$ and $\tilde{f} \circ \theta^*(w)$. To do so, the weight $w = (w_1, w_2, w_3)$ was converted to the regularization coefficient $(\mu, \lambda)$ according to (7.8), where the magnitude of perturbation $\varepsilon$ was set to $1e-16$. Then, the original elastic net problem (7.6) was solved by the coordinate descent method. For the implementation of the elastic net and the coordinate descent method, the Python package scikit-learn was used. Resulting points, $\theta^*(w)$ and $\tilde{f} \circ \theta^*(w)$, were plotted as Figure 6 where each point is colored by its corresponding weight $w$ whose $(w_1, w_2, w_3)$-coordinates are converted to RGB values.

The authors computed a simultaneous approximation of both $\theta^*$ and $\tilde{f} \circ \theta^*$ with a single Bézier simplex. More specifically, the Bézier simplex approximates to the following mapping:

$$(\theta^*, \tilde{f} \circ \theta^*) : \Delta^2 \rightarrow S(\tilde{f}, R^6) \times \operatorname{Min} \tilde{f}(R^6)(\subset R^3).$$

The degree of the Bézier simplex was set to 30. The authors trained the Bézier simplex by the all-at-once fitting [27] using the grid points $w$ together with the resulting points $(\theta^*(w), \tilde{f} \circ \theta^*(w))$ computed as above. The authors plotted the values of the trained Bézier simplex for the grid points as Figure 7. Points were colored in the same way as Figure 6.

3The degree the authors used in this experiment is much higher than that used in [27] (which was two or three). This is required to capture sharp changes around non-smooth points of $\theta^*$. We expect that the required degree would be decreased if one can subdivide the Bézier simplex at non-smooth points. The development of such a subdivision algorithm is future work.
Appendix B. Derivation of (7.9) and (7.10) in Remark 7.2

Let us first derive the equations in (7.10) that converts \((\mu, \lambda)\) to \((w_1, w_2, w_3)\), which is true for some superset of \(\Delta^2 \setminus \Delta^2_{(2,3)}\). By \(\mu = w_2/w_1\) in (7.8) and \(w_1 + w_2 + w_3 = 1\), we have

\[
\begin{align*}
\text{(B.1)} & \\
& w_3 = 1 - w_1 - w_2 \\
& = 1 - w_1 - \mu w_1 \\
& = 1 - (\mu + 1)w_1.
\end{align*}
\]

By \(\lambda = (w_3 + \epsilon)/w_1\) in (7.8), we have

\[
\text{(B.2)} \quad w_3 = \lambda w_1 - \epsilon.
\]

Combining (B.1) and (B.2), we have

\[
1 - (\mu + 1)w_1 = \lambda w_1 - \epsilon \iff (\lambda + \mu + 1)w_1 = 1 + \epsilon.
\]

Since \(\mu, \lambda \geq 0\), we have \(\lambda + \mu + 1 > 0\) and

\[
\text{(B.3)} \quad w_1 = \frac{1 + \epsilon}{\lambda + \mu + 1}.
\]

Then, we substitute (B.3) to (B.2) and obtain

\[
\begin{align*}
\text{(B.4)} & \\
& w_3 = \lambda w_1 - \epsilon \\
& = \lambda \times \frac{1 + \epsilon}{\lambda + \mu + 1} - \epsilon \\
& = \frac{\lambda(1 + \epsilon) - \epsilon(\lambda + \mu + 1)}{\lambda + \mu + 1} \\
& = \frac{\lambda - \epsilon \mu - \epsilon}{\lambda + \mu + 1}.
\end{align*}
\]

Finally, we have

\[
\begin{align*}
\text{(B.5)} & \\
& w_2 = 1 - w_1 - w_3 \\
& = \frac{(\lambda + \mu + 1) - (1 + \epsilon) - (\lambda - \epsilon \mu - \epsilon)}{\lambda + \mu + 1} \\
& = \frac{(1 + \epsilon)\mu}{\lambda + \mu + 1}.
\end{align*}
\]

Next, let us restrict the possible range of \(\mu, \lambda\) to satisfy \(w(\mu, \lambda) \in \Delta^2 \setminus \Delta^2_{(2,3)}\), which derives (7.9). By \(0 < w_1 \leq 1\), it follows from (B.3) that

\[
1 + \epsilon \leq \lambda + \mu + 1,
\]

which can be simplified to

\[
\text{(B.6)} \quad \epsilon - \lambda \leq \mu.
\]

By \(0 \leq w_2 \leq 1\), it follows from (B.5) that

\[
(1 + \epsilon)\mu \leq \lambda + \mu + 1,
\]

which can be simplified to

\[
\epsilon \mu \leq \lambda + 1.
\]
Since $\varepsilon > 0$, we have

(B.7) $\mu \leq \frac{\lambda + 1}{\varepsilon}$.

By $0 \leq w_3 \leq 1$, it follows from (B.4) that

$$0 \leq \lambda - \varepsilon \mu - \varepsilon \leq \lambda + \mu + 1.$$ 

It can be simplified to

(B.8) $-1 \leq \mu \leq \frac{\lambda - \varepsilon}{\varepsilon}$.

By (B.6) to (B.8) and $\mu \geq 0$, we have

$$0 \leq \mu \leq \frac{\lambda - \varepsilon}{\varepsilon}.$$