New merit functions and error bounds for non-convex multiobjective optimization

Hiroki Tanabe, Ellen H. Fukuda, and Nobuo Yamashita

Graduate School of Informatics, Kyoto University
{tanabehiroki@amp.i.kyoto-u.ac.jp},{ellen,nobuo}@i.kyoto-u.ac.jp

Abstract

Our aim is to propose merit functions and provide error bounds for non-convex multiobjective optimization problems. For such problems, the merit functions return zero at Pareto stationary points and strictly positive values otherwise. These functions are known to be important for establishing convergence rates in single-objective optimization, but the related studies for the multiobjective case are still recent. We then propose in this paper six merit functions for multiobjective optimization that differ, on whether they contain some regularization term and they linearize the objective functions. In particular, those with regularization terms necessarily have bounded values. We also compare the proposed merit functions and analyze the sufficient conditions for them to have error bounds. In addition, by considering the well-known Fenchel duality, we present efficient ways to compute these merit functions for particular problems, specifically for differentiable conic programming problems, differentiable constrained problems, and problems involving $\ell_1$ norms.

1 Introduction

Multiobjective optimization is an important field of research with many practical applications. It minimizes several objective functions at once, but usually, there does not exist a single point that minimizes all objective functions at the same time. Therefore, we use the concept of Pareto optimality. A point is called Pareto optimal, if there does not exist another point with the same or smaller objective function values, and with at least one objective function value being strictly smaller. However, for non-convex problems, it is difficult to get Pareto optimal solutions. Thus, we use the concept of Pareto stationarity. A point is called Pareto stationary if there does not exist a descent direction from this point.

Many algorithms for getting Pareto stationary solutions have been developed [6, 4, 6], but from a practical point of view (e.g. to estimate convergence
rates), it is important to know how far a feasible point is from the Pareto stationary set. In this paper, we meet such needs by developing new merit functions for multiobjective optimization, which return zero at the Pareto stationary solutions and strictly positive values otherwise. We also show the error bound properties associated to the proposed merit functions under the strong convexity assumption. This means that the distance between a point and the Pareto optimal set is bounded from above by the value of the new merit functions.

We point out that, as far as we know, there are few results about merit functions for multiobjective optimization and their error bound properties. In [8], the authors proposed merit functions for linear multiobjective optimization problems, and in [3], other functions were considered for convex problems. Both studies also presented some conditions to ensure the existence of error bounds, but we cannot apply their results directly to the non-convex context.

In this paper, we propose six types of merit functions for non-convex multiobjective optimization that differ in terms of computability. More precisely, each function may contain a regularization term, and a linearization of the objective function may also be considered. We will see, for instance, that the functions without the regularization can be unbounded, while the regularized ones are necessarily bounded. We also establish the relationship among all these functions, and we show sufficient conditions for them to have error bounds. Moreover, we present some efficient ways to compute merit functions for particular problems by considering the well-known Fenchel duality.

The outline of this paper is as follows. We present some notations and define Pareto optimality and level-boundedness in Section 2. In Section 3, we propose six types of merit functions for non-convex multiobjective optimization and we check how they are related to each other. We also present error bound properties for the proposed merit functions in Section 4. In Section 5, we observe that these merit functions can be evaluated efficiently at least when particular multiobjective optimization problems are considered. Finally, we conclude this paper in Section 6.

2 Preliminaries

In this paper, we consider the following constrained multiobjective optimization problem:

$$\begin{align*}
\min & \quad F(x) \\
\text{s.t.} & \quad x \in S,
\end{align*}$$

where $F: S \to \mathbb{R}^m$ is a vector-valued functions with $F := (F_1, \ldots, F_m)^\top$, $S \subseteq \mathbb{R}^n$ is a nonempty closed convex set and $\top$ denotes transpose. We assume that each $F_i$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$. 

2.1 Notations

We first present some notions that will be used in this paper. Let us denote by $\mathbb{R}$ the set of real numbers. We use the symbol $\| \cdot \|$ for the Euclidean norm in $\mathbb{R}^n$. The notation $u \leq v \ (u < v)$ means that $u_i \leq v_i \ (u_i < v_i)$ for all $i \in \{1, \ldots, m\}$. Moreover, we call

$$h'(x; d) := \lim_{t \to 0} \frac{h(x+td) - h(x)}{t}$$

the directional derivative of $h: S \to \mathbb{R} \cup \{\infty\}$ at $x$ in the direction $d$. Note that $h'(x; d) = \nabla h(x)^\top d$ when $h$ is differentiable at $x$, where $\nabla h(x)$ stands for the gradient of $h$ at $x$. Furthermore, we define the lower Dini derivative of $h$ at $x$ in the direction $d$ by

$$h'_-(x; d) := \liminf_{t \to 0} \frac{h(x+td) - h(x)}{t}.$$

2.2 Pareto optimality and stationarity

Now, we introduce the concept of optimality for the multiobjective optimization problem (1). Recall that $x^* \in S$ is Pareto optimal, if there is no $x \in S$ such that $F(x) \leq F(x^*)$ and $F(x) \neq F(x^*)$. Likewise, $x^* \in S$ is weakly Pareto optimal, if there does not exist $x \in S$ such that $F(x) < F(x^*)$. It is known that Pareto optimal points are always weakly Pareto optimal, and the converse is not always true. We also say that $\bar{x} \in S$ is Pareto stationary [2], if and only if,

$$\max_{i \in \{1, \ldots, m\}} F_i'(\bar{x}; z - \bar{x}) \geq 0 \text{ for all } z \in S.$$

We state below the relation among the three concepts of Pareto optimality.

Lemma 2.1. The following three statements hold.

1. If $x \in S$ is weakly Pareto optimal for (1), then $x$ is Pareto stationary.

2. Let every component $F_i$ of $F$ be convex. If $x \in S$ is Pareto stationary for (1), then $x$ is weakly Pareto optimal.

3. Let every component $F_i$ of $F$ be strictly convex. If $x \in S$ is Pareto stationary for (1), then $x$ is Pareto optimal.

Proof. It is clear from [12] Lemma 2.2.

2.3 Level-boundedness

Finally, we introduce the concept of level-boundedness of scalar-valued and vector-valued functions. For scalar-valued functions, a function $f: S \to \mathbb{R}$ is said to be level-bounded [10] if for all $\alpha \in \mathbb{R}$ the level set $\{x \in S \mid f(x) \leq \alpha\}$ is bounded. We also define the level-boundedness for vector-valued function as follows:
**Definition 2.1.** A vector-valued function $F: S \rightarrow \mathbb{R}^m$ is level-bounded if the level set $\{x \in S \mid F(x) \leq \zeta\}$ is bounded for all $\zeta \in \mathbb{R}^m$.

If $F_i$ is level-bounded for all $i \in \{1, \ldots, m\}$, then $F = (F_1, \ldots, F_m)^\top$ is also level-bounded. Note that even if $F$ is level-bounded, every $F_i$ is not necessarily level-bounded (e.g. $F(x) = (x - 1, -2x + 1)^\top$). Now, we show the existence of weakly Pareto optimal points under level-boundedness assumption.

**Theorem 2.1.** If $F$ is continuous and level-bounded, then (1) has a weakly Pareto optimal solution.

**Proof.** Let $F$ be continuous and level-bounded. Then the level set $\Omega_F(\alpha) := \{x \in S \mid F_i(x) \leq \alpha \text{ for all } i \in \{1, \ldots, m\}\}$ is bounded for all $\alpha \in \mathbb{R}$. Now, we have

$$\Omega_F(\alpha) = \left\{ x \in S \mid \max_{i \in \{1, \ldots, m\}} F_i(x) \leq \alpha \right\},$$

so $\max_i F_i$ is also level-bounded. Moreover, since $F$ is continuous, $\max_i F_i$ is also continuous. Thus, the problem

$$\min \max_{i \in \{1, \ldots, m\}} F_i(x) \quad \text{s.t. } x \in S$$

has a global optimal solution $x^\ast$. This gives

$$\max_{i \in \{1, \ldots, m\}} F_i(x^\ast) \leq \max_{i \in \{1, \ldots, m\}} F_i(x) \text{ for all } x \in S,$$

which means that $x^\ast$ is weakly Pareto optimal for (1). \qed

### 3 New merit functions for multiobjective optimization

A function is called merit function associated with an optimization problem if it returns zero at their solutions and strictly positive values otherwise. In this section, we propose new merit functions for nonlinear and non-convex multiobjective optimization problems. Let us now consider that each component $F_i$ of the objective function $F$ of (1) is defined by

$$F_i(x) := f_i(x) + g_i(x), \quad i \in \{1, \ldots, m\},$$

(2)

where $f_i: S \rightarrow \mathbb{R}$ has directional derivative $f_i'(x; z - x)$ for all $x, z \in S$ and $g_i$ is a closed, proper and convex function from $S$ to $\mathbb{R} \cup \{\infty\}$. This type of objective function has many applications, in particular, in machine learning.
For this problem, we can consider three types of merit functions as follows:

\[ u_\ell(x) := \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \left\{ F_i(x) - F_i(y) - \frac{\ell}{2} \|x - y\|^2 \right\}, \quad (3) \]

\[ v_\ell(x) := \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \left\{ -F_i'(x; y - x) - \frac{\ell}{2} \|x - y\|^2 \right\}, \quad (4) \]

\[ w_\ell(x) := \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \left\{ -f_i'(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|x - y\|^2 \right\}, \quad (5) \]

where \( \ell \geq 0 \) is a constant. The merit function \( u_\ell \) defined by (3) with \( \ell = 0 \) is proposed in \([8]\), and they prove that the solution of the equation \( u_0(x) = 0 \) is weakly Pareto optimal for the original problem, but they define the function only for linear multiobjective problems, while here we are considering the more general nonlinear cases. The function \( v_\ell \) defined by (4) are also proposed in \([3]\), where it was shown that the solution of \( v_\ell = 0 \) is weakly Pareto optimal for the original problem. However, once again, they define it only for convex problems, so we extend it for non-convex cases. The function \( w_\ell \) defined by (5) have not been studied as far as we know.

Moreover, by distinguishing \( \ell = 0 \) and \( \ell > 0 \), we have a total of six three types of merit functions, each one having different properties. For example, \( u_0, v_0 \) and \( w_0 \) may be unbounded, but for \( \ell > 0 \), \( v_\ell \) and \( w_\ell \) are necessarily bounded. Also, when \( f \) is convex (and thus the whole objective function is convex), \( u_\ell \) is bounded when \( \ell > 0 \). Furthermore, for \( \ell \geq 0 \), \( u_\ell \) is a merit function in the sense of weak Pareto optimality, while \( v_\ell \) and \( w_\ell \) are merit functions in the sense of Pareto stationarity. In the subsequent subsections, we discuss the six types of merit functions for nonlinear and non-convex multiobjective optimization problems, proving the mentioned results.

### 3.1 Relationship among the proposed merit functions

In this section, we show the connection among the proposed functions \( u_\ell, v_\ell \) and \( w_\ell \) defined by \([3],[4]\) and \([5]\), respectively. First, we show that some inequalities hold when considering these different proposed functions.

**Theorem 3.1.** For all \( \ell \geq 0 \), let \( u_\ell, v_\ell \) and \( w_\ell \) be defined by \([3],[4]\) and \([5]\), respectively. Define \( \nu_i \geq 0 \) as a convexity parameter of \( g_i \) for all \( i \in \{1, \ldots, m\} \) and set \( \nu := \min_{i \in \{1, \ldots, m\}} \nu_i \). Then, the following statements hold.

(i) For all \( x \in S \), we have

\[ v_{\nu + \ell}(x) \geq w_\ell(x). \]

Moreover, suppose that \( f_i \) has a convexity parameter \( \mu_i \in \mathbb{R} \) and let \( \mu := \min_{i \in \{1, \ldots, m\}} \mu_i \). Then, if \( \mu + \ell \geq 0 \), we obtain

\[ w_{\mu + \ell}(x) \geq u_\ell(x) \quad \text{for all } x \in S, \]

and if \( \mu + \nu + \ell \geq 0 \), it follows that

\[ v_{\mu + \nu + \ell}(x) \geq u_\ell(x) \quad \text{for all } x \in S. \]
(ii) Assume that all $i \in \{1, \ldots, m\}$ $f_i$ is continuously differentiable, $\nabla f_i$ is Lipschitz continuous with Lipschitz constant $L_i > 0$, and let $L := \max_{i \in \{1, \ldots, m\}} L_i$. Then, we get

$$u_\ell(x) \geq w_{L+\ell}(x) \text{ for all } x \in S.$$ 

Moreover, assume that $F_i$ is continuously differentiable, $\nabla F_i$ is Lipschitz continuous with Lipschitz constant $\hat{L}_i > 0$ for all $i \in \{1, \ldots, m\}$, and define $\hat{L} := \max_{i \in \{1, \ldots, m\}} \hat{L}_i$. Then, we have

$$u_\ell(x) \geq v_{L+\ell}(x) \text{ for all } x \in S.$$ 

Proof. Suppose that $f_i$ is Lipschitz continuous with Lipschitz constant $L_i$. Then, it follows from [1, Proposition A.24] that

$$f_i(y) - f_i(x) \leq \nabla f_i(x)^\top (y - x) + \frac{L_i}{2} \|y - x\|^2.$$ 

By the definition of $L$, we have

$$f_i(x) - f_i(y) - \frac{\ell}{2} \|x - y\|^2 \geq \nabla f_i(x)^\top (x - y) - \frac{L_i + \ell}{2} \|x - y\|^2.$$ 

Therefore, we immediately get $u_\ell(x) \geq w_{L+\ell}(x)$ for all $x \in S$ by the definitions (3) and (5) of $u_\ell$ and $w_{L+\ell}$, respectively. Furthermore, we can prove the second inequality in the same manner.

Next, we prove the relationships between coefficients and function values in the same proposed functions.

**Theorem 3.2.** For all $\ell \geq 0$, let $u_\ell$, $v_\ell$ and $w_\ell$ be defined by (3), (4) and (5), respectively. Let $r$ be an arbitrary scalar such that $r \geq \ell$. Then, for all $x \in S$, we have

$$u_r(x) \leq u_\ell(x), \quad v_r(x) \leq v_\ell(x) \quad \text{and} \quad w_r(x) \leq w_\ell(x).$$
Moreover, if $\ell > 0$, then we get
\[ \frac{\ell}{r} v_\ell(x) \leq v_r(x) \quad \text{and} \quad \frac{\ell}{r} w_\ell(x) \leq w_r(x). \]
Furthermore, if $F_i$ is convex for all $i \in \{1, \ldots, m\}$, then we obtain
\[ \frac{\ell}{r} u_\ell(x) \leq u_r(x). \]

Proof. From the definitions (3), (4) and (5) of $u_\ell$, $v_\ell$ and $w_\ell$, we can easily show that
\[ u_r(x) \leq u_\ell(x), \quad v_r(x) \leq v_\ell(x) \quad \text{and} \quad w_r(x) \leq w_\ell(x). \]

Now, let $\ell > 0$. Then, from the definition (5) of $w_\ell$, we have
\[ w_\ell(x) = \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|x - y\|^2 \right\} \]
\[ \leq \sup_{\ell \in S} \min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; \frac{\ell}{r} (y - x)) + g_i(x) - g_i(y) - \frac{\ell}{2} \left\| \frac{\ell}{r} (x - y) \right\|^2 \right\} \]
\[ \leq \sup_{\ell \in S} \min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; \frac{\ell}{r} (y - x)) + g_i(x) - g_i \left( x - \frac{\ell}{r} (x - y) \right) - \frac{\ell}{2} \left\| \frac{\ell}{r} (x - y) \right\|^2 \right\} \]
where the first inequality follows from the convexity of $g_i$. Since $S$ is convex, we have $x - (\ell/r)(x - y) \in S$. Therefore, from the definition (5) of $w_r$, we get
\[ \frac{\ell}{r} u_\ell(x) \leq \frac{\ell}{r} w_\ell(x). \]

From (6) with $g_i = 0$, we can easily show that
\[ \frac{\ell}{r} v_\ell(x) \leq \frac{\ell}{r} v_r(x). \]

When $F_i$ is convex, we can prove that
\[ u_\ell(x) \leq \frac{\ell}{r} u_r(x) \]
by setting $f_i = 0$ in (6).

Remark 3.1. For unconstrained problems, we can consider the following inequality.
\[ w_L(x) \geq \tau u_0(x) \quad \text{for all} \quad x \in \mathbb{R}^n \quad \text{for some} \quad \tau > 0, \]
which is an extension of the proximal-PL condition for scalar optimization [7]. Under (7), we can prove that multiobjective proximal gradient methods [12] have linear convergence rate [13]. Note that if each $f_i$ is strongly convex, then (7) holds from Theorem 3.1 (i) and Theorem 3.2.
3.2 Conditions for the proposed functions to be merit functions

In this section, we show the conditions for the proposed functions to be merit functions in the Pareto sense. We start by proving that those functions are necessarily nonnegative.

**Theorem 3.3.** Let \( \ell \geq 0 \). Then, functions \( u_\ell \), \( v_\ell \) and \( w_\ell \) defined by (3), (4) and (5), respectively, are nonnegative on \( S \).

**Proof.** By the definition (3), we get

\[
   u_\ell(x) \geq \min_{i \in \{1, \ldots, m\}} \left\{ F_i(x) - F_i(y) - \frac{\ell}{2} \|x - y\|^2 \right\} = 0 \quad \text{for all } x \in S.
\]

We can also show that \( v_\ell(x) \geq 0 \) and \( w_\ell(x) \geq 0 \) in the same manner. \( \square \)

Next, we show that \( u_\ell \) defined by (3) is in fact a merit function in the sense of weak Pareto optimality when \( \ell = 0 \) or \( F_i \) is convex for all \( i \in \{1, \ldots, m\} \).

**Theorem 3.4.** For all \( \ell \geq 0 \), let \( u_\ell \) be defined by (3). If \( x \in S \) is weakly Pareto optimal for (1), then we have \( u_\ell(x) = 0 \). Conversely, if \( u_\ell(x) = 0 \), then \( x \) is weakly Pareto optimal for (1), if one of the following conditions holds.

(i) \( \ell = 0 \).
(ii) \( F_i \) is convex for all \( i \in \{1, \ldots, m\} \).

**Proof.** First, we set \( \ell = 0 \). It follows from (3) that

\[
   u_0(x) = 0 \iff \min_{i \in \{1, \ldots, m\}} \{ F_i(x) - F_i(y) \} \leq 0 \quad \text{for all } y \in S,
\]

which is equivalent to the existence of \( i \in \{1, \ldots, m\} \) such that

\[
   F_i(x) - F_i(y) \leq 0 \quad \text{for all } y \in S.
\]

In other words, there does not exist \( y \in S \) such that

\[
   F_i(x) - F_i(y) > 0 \quad \text{for all } i \in \{1, \ldots, m\}.
\]

Therefore, \( x \in S \) is weakly Pareto optimal for (1) if and only if \( u_0(x) = 0 \). Now, let \( \ell > 0 \). Let us suppose that \( x \in S \) is weakly Pareto optimal for (1). Then, with the same argument as before, it follows that \( u_\ell(x) \leq 0 \). This, together with Theorem 3.3 yields \( u_\ell(x) = 0 \). Finally, if \( u_\ell(x) = 0 \) and \( F_i \) is convex for all \( i \in \{1, \ldots, m\} \), we can show that \( x \) is weakly Pareto optimal from Lemma 2.1(ii) and Theorem 3.3 with \( f_i = 0 \). \( \square \)

For functions \( v_\ell \) and \( w_\ell \) defined by (4) and (5), respectively, we can show that they are merit functions in the sense of Pareto stationarity.
Theorem 3.5. Assume that $\ell \geq 0$. Define $v_\ell$ and $w_\ell$ by (4) and (5), respectively, and let $x \in S$. Then, the following three statements are equivalent.

(i) $x$ is Pareto stationary for (1).

(ii) $v_\ell(x) = 0$.

(iii) $w_\ell(x) = 0$.

Proof. $(i) \iff (iii)$ Let $x \in S$. Assume that $w_\ell(x) = 0$. It follows from (5) that

$$
\min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|x - y\|^2 \right\} \leq 0 \quad \text{for all } y \in S. \tag{8}
$$

Since $S$ is convex, if $z \in S$, then $x + \alpha(z - x) \in S$ for all $\alpha \in (0, 1)$. Therefore, by substituting $y = x + \alpha(z - x)$ into (8) we obtain

$$
\min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; \alpha(z - x)) + g_i(x) - g_i(x + \alpha(z - x)) - \frac{\ell \alpha}{2} \|z - x\|^2 \right\}
$$

for all $z \in S$.

Dividing both sides by $\alpha$ yields

$$
\min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; \alpha(z - x)) / \alpha + g_i(x) / \alpha - g_i(x + \alpha(z - x)) / \alpha - \frac{\ell \alpha}{2} \|z - x\|^2 \right\}
$$

for all $z \in S$.

By taking $\alpha \downarrow 0$ and from the continuity of the involved functions, we get

$$
\min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; z - x) - g'_i(x; z - x) \right\} \leq 0 \quad \text{for all } z \in S,
$$

which shows that $x$ is Pareto stationary for (1).

Let us now prove the converse by contrapositive. Assume that $w_\ell(x) > 0$. Then, for some $y \in S$ we have

$$
\min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|x - y\|^2 \right\} > 0.
$$

Since $g_i$ is convex, we obtain

$$
\min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; y - x) - g'_i(x; y - x) - \frac{\ell}{2} \|x - y\|^2 \right\} > 0.
$$

By taking $\alpha \downarrow 0$ and from the continuity of the involved functions, we get

$$
\max_{i \in \{1, \ldots, m\}} F'_i(x; y - x) \leq -\frac{\ell}{2} \|x - y\|^2 < 0,
$$

which shows that $x$ is not Pareto stationary for (1). Since $w_\ell(x) \geq 0$ from Theorem 3.3 the claim holds.

$(i) \iff (ii)$ We can show this statement simply by using $(i) \iff (iii)$ with $g_i = 0$. \qed

9
From Theorems 3.3–3.5 we conclude that function $u_0$ is a merit function in the sense of weak Pareto optimality, and functions $v_\ell$ and $w_\ell$ are merit functions only in the sense of Pareto stationarity for all $\ell \geq 0$. As indicated by the following example, even if $v_\ell(x) = 0$, $x$ is not necessarily weakly Pareto optimal for (1).

**Example 3.1.** Consider the single-objective function $F: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$F(x) = -x^2,$$

and let $S = [-2, 1]$. Then, we have

$$v_\ell(0) = \sup_{y \in S} \left\{ -F'(0; y - 0) - \frac{\ell}{2}(y - 0)^2 \right\} = \sup_{y \in S} \left\{ -\frac{\ell}{2}y^2 \right\} = 0.$$

However, $x = 0$ is not the global optimal point (i.e. weakly Pareto optimal point) for $F$.

### 3.3 Minimization of $u_0$

In this subsection, let us analyze the following single-objective optimization problem:

$$\begin{align*}
\min & \quad u_0(x) \\
\text{s.t.} & \quad x \in S.
\end{align*}$$

(9)

If the global optimal solution $x^*$ of (9) exists and satisfies $u_0(x^*) > 0$, then $x^*$ is not weakly Pareto optimal from Theorem 3.4. However, as shown in the next theorem, the global solutions of (9) are always weakly Pareto optimal for (1). Before proving this, we need the following basic result.

**Lemma 3.1.** Let $G_i: S \rightarrow \mathbb{R}$ and $H_i: S \rightarrow \mathbb{R}$ be upper and lower semicontinuous, respectively, for all $i \in \{1, \ldots, m\}$. Then, we have

$$\sup_{x \in S} \min_{i \in \{1, \ldots, m\}} G_i(x) - \sup_{x \in S} \min_{i \in \{1, \ldots, m\}} H_i(x) \leq \sup_{x \in S} \max_{i \in \{1, \ldots, m\}} [G_i(x) - H_i(x)].$$

**Proof.** Let $f: S \rightarrow \mathbb{R}$ and $g: S \rightarrow \mathbb{R}$ be upper semicontinuous. Then, it follows that

$$\sup_{x \in S} (f(x) + g(x)) \leq \sup_{x \in S} f(x) + \sup_{x \in S} g(x).$$

Now, define $h: S \rightarrow \mathbb{R}$ as $h := f + g$. Then, we obtain

$$\sup_{x \in S} h(x) - \sup_{x \in S} f(x) \leq \sup_{x \in S} (h(x) - f(x)).$$
Substituting $h(x) = \min_{i \in \{1, \ldots, m\}} G_i(x)$ and $f(x) = \min_{i \in \{1, \ldots, m\}} H_i(x)$ into the above inequality, we get

$$\sup_{x \in S} \min_{i \in \{1, \ldots, m\}} G_i(x) - \sup_{x \in S} \min_{i \in \{1, \ldots, m\}} H_i(x) \leq \sup_{x \in S} \left[ \min_{i \in \{1, \ldots, m\}} G_i(x) - \min_{i \in \{1, \ldots, m\}} H_i(x) \right].$$

Now, let $j_x \in \arg\min_{i \in \{1, \ldots, m\}} H_i(x)$. Since

$$\min_{i \in \{1, \ldots, m\}} H_i(x) = H_{j_x}(x),$$

we have

$$\sup_{x \in S} \min_{i \in \{1, \ldots, m\}} G_i(x) - \sup_{x \in S} \min_{i \in \{1, \ldots, m\}} H_i(x) \leq \sup_{x \in S} \left[ G_{j_x}(x) - H_{j_x}(x) \right] \leq \sup_{x \in S} \max_{i \in \{1, \ldots, m\}} \left[ G_i(x) - H_i(x) \right],$$

where the second and third inequalities come from the definition of minimum and maximum.

**Theorem 3.6.** Let $u_0$ be defined by (3). If $x^* \in S$ is global optimal for (9), then $x^*$ is weakly Pareto optimal for (1).

**Proof.** Let $x^* \in S$ be a global optimal solution of (9). Then, for all $z \in S$, we have $u_0(x^*) \leq u_0(z)$. This gives

$$0 \leq u_0(z) - u_0(x^*) = \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \{F_i(z) - F_i(y)\} - \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \{F_i(x^*) - F_i(y)\} \leq \sup_{y \in S} \max_{i \in \{1, \ldots, m\}} \{F_i(z) - F_i(y) - (F_i(x^*) - F_i(y))\} = \max_{i \in \{1, \ldots, m\}} \{F_i(z) - F_i(x^*)\},$$

where the first equality comes from the definition (3) of $u_0$, and Lemma 3.1 yields the second inequality. Therefore, $x$ is weakly Pareto optimal for (1) by definition.

Since the objective function of (3) is generally non-convex, the problem (9) does not necessarily have global optimal solutions. However, we can prove that every stationary point of (9) is Pareto stationary for (1).

**Theorem 3.7.** Let $u_0$ be defined by (3) and assume that there exists a directional derivative $F_i'(x; z - x)$ for all $i \in \{1, \ldots, m\}$ and $x, z \in S$. If $u_0$ also
has an lower Dini derivative \((u_0)'_-(x; z - x)\) for all \(x, z \in S\) and is stationary for (9), that is,
\[
(u_0)'_-(x; z - x) \geq 0 \quad \text{for all } z \in S,
\]
then \(x\) is Pareto stationary for (11).

Proof. Let \(x \in S\) be stationary for (9). By the definition (3) of \(u_0\), we see that for all \(z \in S\),
\[
(u_0)'_-(x; z - x) = \liminf_{t \searrow 0} \frac{1}{t} \left[ \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \{F_i(x + t(z - x)) - F_i(y)\} \right. \\
\left. - \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \{F_i(x) - F_i(y)\} \right] \\
\leq \liminf_{t \searrow 0} \frac{1}{t} \sup_{y \in S} \max_{i \in \{1, \ldots, m\}} \{F_i(x + t(z - x)) - F_i(y)\} - \{F_i(x) - F_i(y)\} \\
= \liminf_{t \searrow 0} \max_{i \in \{1, \ldots, m\}} \frac{F_i(x + t(z - x)) - F_i(x)}{t},
\]
where the definition of lower Dini derivative yields the first equality, and the inequality follows from Lemma 3.1. Since
\[
\max_{i \in \{1, \ldots, m\}} \frac{F_i(x + t(z - x)) - F_i(x)}{t}
\]
is continuous, we have
\[
(u_0)'_-(x; z - x) = \max_{i \in \{1, \ldots, m\}} \liminf_{t \searrow 0} \frac{F_i(x + t(z - x)) - F_i(x)}{t} \\
= \max_{i \in \{1, \ldots, m\}} F_i'(x; z - x),
\]
where the second equality comes from the existence of \(F_i'(x; z - x)\). Therefore, by (10) we get
\[
\max_{i \in \{1, \ldots, m\}} F_i'(x; z - x) \geq 0 \quad \text{for all } z \in S,
\]
which shows that \(x\) is Pareto stationary for (11). \qed

Now, we state below a sufficient condition for the level-boundedness of the merit function \(u_0\).

**Theorem 3.8.** Let \(u_0\) be defined by (3). If \(F_i\) is level-bounded for all \(i \in \{1, \ldots, m\}\), then \(u_0\) is also level-bounded.

**Proof.** Suppose, contrary to our claim, that \(u_0\) is not level-bounded. Then, there exists \(\alpha \in \mathbb{R}\) such that \(\{x \in S \mid u_0(x) \leq \alpha\}\) is not bounded. By the definition (3) of \(u_0\), the inequality \(u_0(x) \leq \alpha\) can be written as
\[
\sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \{F_i(x) - F_i(y)\} \leq \alpha.
\]
This implies for some fixed \( z \in S \) that there exists \( j \in \{1, \ldots, m\} \) such that
\[
F_j(x) \leq F_j(z) + \alpha.
\]

Therefore, it follows that
\[
\{x \in S \mid u_0(x) \leq \alpha\} \subseteq \bigcup_{j=1}^{m} \{x \in S \mid F_j(x) \leq F_j(z) + \alpha\}.
\]

Since \( F_i \) is level-bounded for all \( i \in \{1, \ldots, m\} \), the right-hand side must be bounded, which contradicts the unboundedness of the left-hand side.

As indicated by the following example, even if \( F \) is level-bounded, \( u_0 \) is not necessarily level-bounded.

**Example 3.2.** Consider the bi-objective function \( F: \mathbb{R} \to \mathbb{R}^2 \) with each component given by
\[
F_1(x) := x^2, \quad F_2(x) := 0.
\]
Then, the merit function \( u_0 \) defined by (3) is written as
\[
u_0(x) = \sup_{y \in \mathbb{R}} \min \{F_1(x) - F_1(y), F_2(x) - F_2(y)\} = \sup_{y \in \mathbb{R}} \min \{(x^2 - y^2), 0\} = 0.
\]

On the other hand, \( F \) is level-bounded because \( \lim_{\|z\| \to \infty} F_1(x) = +\infty \).

### 4 Error bound properties of the proposed merit functions

In this section, we show the error bound properties of the merit functions \( u_\ell, v_\ell \) and \( w_\ell \) defined by (3), (4) and (5), respectively, under strong convexity assumption. This means that we can estimate the distance between a point and the Pareto optimal set. Now, let \( X^* \subseteq S \) be a set of Pareto optimal points for (1) and \( \text{dist}\{x, X^*\} := \min\{\|x - y\| \mid y \in X^*\} \).

**Theorem 4.1.** For arbitrary \( \ell \geq 0 \), let \( w_\ell \) be defined by (5). Assume that \( f_i \) and \( g_i \) have convexity parameters \( \mu_i \in \mathbb{R} \) and \( \nu_i \geq 0 \), respectively, and \( 2\mu_i + \nu_i - \ell > 0 \) for all \( i \in \{1, \ldots, m\} \). Then, it follows that
\[
w_\ell(x) \geq \frac{\rho}{2} \text{dist}\{x, X^*\}^2 \quad \text{for all } x \in S,
\]
where \( \rho := \min_{i \in \{1, \ldots, m\}} \{2\mu_i + \nu_i - \ell\} \).
Proof. By the definition (5), we have

\[
w_\ell(x) = \sup_{y \in S} \min_{i \in \{1, \ldots, m\}} \left\{ -f'_i(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|y - x\|^2 \right\}
\]

\[
= \sup_{y \in S} \min_{\lambda \in C} \sum_{i=1}^{m} \lambda_i \left\{ -f'_i(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|y - x\|^2 \right\},
\]

where \( C := \{\lambda \in \mathbb{R}^m \mid \lambda_i \geq 0 \text{ and } \sum_{i=1}^{m} \lambda_i = 1 \text{ for all } i \in \{1, \ldots, m\}\}. \) By using Sion’s minimax theorem [11], we get

\[
w_\ell(x) = \min_{\lambda \in C} \sup_{y \in S} \sum_{i=1}^{m} \lambda_i \left\{ -f'_i(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|y - x\|^2 \right\}.
\]

Since \( C \) is compact, there exists \( \lambda^* \in C \) such that

\[
w_\ell(x) = \sup_{y \in S} \sum_{i=1}^{m} \lambda_i^* \left\{ -f'_i(x; y - x) + g_i(x) - g_i(y) - \frac{\ell}{2} \|y - x\|^2 \right\}.
\]

Now, let \( x^* \in S \) be the solution of

\[
\min_{x \in S} \sum_{i=1}^{m} \lambda_i^* F_i(x)
\]

The problem (11) is a weighting scalarization of (1), so \( x^* \) is Pareto optimal for (11) [9]. Hence, we obtain

\[
w_\ell(x) \geq \sum_{i=1}^{m} \lambda_i^* \left\{ -f'_i(x^*; x - x) + g_i(x^*) - g_i(-x^*) - \frac{\ell}{2} \|x - x^*\|^2 \right\}
\]

\[
\geq \sum_{i=1}^{m} \lambda_i^* \left\{ f_i(x) - f_i(x^*) + g_i(x) - g_i(x^*) + \frac{\mu_i - \ell}{2} \|x - x^*\|^2 \right\}
\]

\[
\geq \sum_{i=1}^{m} \lambda_i^* \left\{ -f'_i(x^*; x - x^*) + g'_i(x^*; x) - g'_i(x^*; x) + \frac{2\mu_i + \nu_i - \ell}{2} \|x - x^*\|^2 \right\}
\]

\[
\geq \frac{\rho}{2} \|x - x^*\|^2
\]

\[
\geq \frac{\rho}{2} \text{dist}\{x, X^*\}^2
\]

where the second and third inequality follows from the convexity of \( f_i \) and \( g_i \), and the fourth one comes from the optimality condition of (11) and the definition of \( \rho \).

We can easily introduce error bound property of \( u_\ell \) and \( v_\ell \) by setting \( f_i = 0 \) and \( g_i = 0 \), respectively, in Theorem 4.1.
**Corollary 4.1.** For arbitrary \( \ell \geq 0 \), let \( u_\ell \) and \( v_\ell \) be defined by (3) and (4), respectively. Assume that \( F_i \) is strongly convex with modulus \( \mu_i > 0 \) for all \( i \in \{1, \ldots, m\} \). Then, if \( \mu_i > \ell \), we have

\[
u_\ell(x) \geq \rho \frac{\ell}{2} \text{dist}\{x, X^*\}^2 \quad \text{for all } x \in S,
\]

where \( \rho := \min_{i \in \{1, \ldots, m\}} \{\mu_i - \ell\} \). Furthermore, if \( 2\mu_i > \ell \), it follows that

\[
u_\ell(x) \geq \hat{\rho} \frac{\ell}{2} \text{dist}\{x, X^*\}^2 \quad \text{for all } x \in S,
\]

where \( \hat{\rho} := \min_{i \in \{1, \ldots, m\}} \{2\mu_i - \ell\} \).

By using Corollary 4.1, we can immediately show the boundedness of Pareto optimal sets under the assumption of strong convexity of the objective functions.

**Theorem 4.2.** If \( F_i \) is strongly convex for all \( i \in \{1, \ldots, m\} \), then the Pareto optimal set of (1) is bounded.

## 5 Evaluation of the proposed merit functions

The proposed merit functions are usually difficult to evaluate. However, for particular problems, we can calculate them easily. In this section, we propose the algorithm of computing them by considering Fenchel duality.

### 5.1 Differentiable conic problems

Here, we assume that the constraint set \( S \subseteq \mathbb{R}^n \) of (1) is a cone, and the objective function \( F_i \) is differentiable. Then, the merit function \( v_\ell \) defined by (4) with \( \ell > 0 \) can be written as

\[
v_\ell(x) = \max_{y \in S} \min_{i \in \{1, \ldots, m\}} \left\{ -\nabla F_i(x)^\top (y - x) - \frac{\ell}{2} \|y - x\|^2 \right\} = -\min_{d \in S - x} \max_{i \in \{1, \ldots, m\}} \left\{ \nabla F_i(x)^\top d + \frac{\ell}{2} \|d\|^2 \right\}.
\]

(12)

Now, we define functions \( h_1 : \mathbb{R}^m \to \mathbb{R} \) and \( h_2 : \mathbb{R}^n \to \mathbb{R} \) as

\[
h_1(d) := \max_{i \in \{1, \ldots, m\}} y_i, \quad h_2^x(y) := \begin{cases} \frac{\ell}{2} \|d\|^2, & d \in S - x, \\ \infty, & \text{otherwise}. \end{cases}
\]

Then, we get the conjugate function \( h_1^* : \mathbb{R}^m \to \mathbb{R} \) of \( h_1 \) as follows:

\[
h_1^*(v) = \begin{cases} 0, & \text{if } v \geq 0 \text{ and } \sum_{i=1}^m v_i = 1, \\ \infty, & \text{otherwise}. \end{cases}
\]
Moreover, we have the conjugate function \((h^2_x)^*: \mathbb{R}^n \to \mathbb{R}\) of \(h^2_x\) as follows:

\[
(h^2_x)^*(u) = \sup_{d \in S - x} \left\{ u^\top d - \frac{\ell}{2} \|d\|^2 \right\}
\]

\[
= \sup_{z \in S} \left\{ u^\top (z - x) - \frac{\ell}{2} \|z - x\|^2 \right\}
\]

\[
= -u^\top x - \frac{\ell}{2} \|x\|^2 + \sup_{z \in S} \left\{ (u + \ell x)^\top z - \frac{\ell}{2} \|z\|^2 \right\}
\]

\[
= -u^\top x - \frac{\ell}{2} \|x\|^2 + p^*(u + \ell x),
\]

where \(p^*: \mathbb{R}^n \to \mathbb{R}\) is a conjugate function of \(p: \mathbb{R}^n \to \mathbb{R}\) defined by

\[
p(d) := \begin{cases} 
\frac{\ell}{2} \|d\|^2, & d \in S, \\
\infty, & \text{otherwise}.
\end{cases}
\]

Let \(P_S: \mathbb{R}^n \to \mathbb{R}\) be a projection onto \(S\). Then, we have

\[
p^*(u) = \sup_{z \in S} \left\{ u^\top z - \frac{\ell}{2} \|z\|^2 \right\}
\]

\[
= \frac{1}{2\ell} \|u\|^2 + \sup_{z \in S} \left\{ -\frac{\ell}{2} \left\| z - \frac{1}{\ell} u \right\|^2 \right\}
\]

\[
= \frac{1}{2\ell} \|u\|^2 - \frac{1}{2\ell} \min_{z \in S} \|z - u\|^2
\]

\[
= \frac{1}{2\ell} \|u\|^2 - \frac{1}{2\ell} \|P_S(u) - u\|^2
\]

\[
= \frac{1}{2\ell} \|P_S(u)\|^2,
\]

where the third equality follows since \(S\) is a cone, the fourth equality holds from the definition of projection, and the last one follows from the Pythagorean theorem. Therefore, by considering Fenchel duality of (12), we obtain

\[
v_\ell(x) = \frac{\ell}{2} \|x\|^2 - \beta(x),
\]

where \(\beta(x)\) is an optimal value of the following optimization problem:

\[
\min_{\eta \in \mathbb{R}^m} \sum_{i=1}^{m} \eta_i \nabla f_i(x)^\top x + \frac{1}{2\ell} \left\| P_S \left( -\sum_{i=1}^{m} \eta_i \nabla f_i(x) + \ell x \right) \right\|^2
\]

s.t. \(\eta \ge 0, \sum_{i=1}^{m} \eta_i = 1\).

(13)

Observe that the dimension of the above problem’s variable is \(m\). Therefore, if the projection onto \(S\) can be computed easily and \(m\) is quite smaller than \(n\),
it would be easier to evaluate the merit function $v_\ell$ with (13) instead of (12). In particular, for two objective functions and variables restricted to the non-negative orthant (i.e. $m = 2$ and $S = \{x \in \mathbb{R}^n \mid x \geq 0\}$), we get a piecewise quadratic programming problem with a unique variable, by considering the fact that one variable can be deduced from the constraint $\eta_1 + \eta_2 = 1$ of (13). Thus, we can evaluate (13) analytically.

5.2 Constrained and differentiable problems

In this section, we assume that the problem is unconstrained, that is, $S = \mathbb{R}^n$, each component $F_i$ of the objective function $F$ of (1) is defined by (2), $f_i$ is continuously differentiable, and

$$g_i(x) = \begin{cases} 0 & x \in C, \\ \infty & \text{otherwise,} \end{cases}$$

where $C \subseteq \mathbb{R}^n$ is convex. Then, the partially merit function $u_0$ defined by (??) returns $+\infty$ when $x /\in C$. Moreover, for all $x \in C$, we have

$$u_0(x) = \sup_{y \in \mathbb{R}^n} \min_{i \in \{1, \ldots, m\}} \{\nabla f_i(x)\top (x - y) - g_i(y)\} = \max_{d \in \mathbb{R}^n} \left\{ -\max_{i \in \{1, \ldots, m\}} \{\nabla f_i(x)\top d\} - g_i(-(-d) + x) \right\} \quad (14)$$

Now, we define functions $h_1: \mathbb{R}^m \to \mathbb{R}$ and $h_2: \mathbb{R}^n \to \mathbb{R}$ as

$$h_1(y) := \max_{i \in \{1, \ldots, m\}} y_i, \quad h_2(y) := \begin{cases} 0 & x - y \in C, \\ \infty & \text{otherwise.} \end{cases}$$

Immediately, we get the conjugate function $h_1^*: \mathbb{R}^m \to \mathbb{R}$ and $h_2^*: \mathbb{R}^n \to \mathbb{R}$ of them as follows

$$h_1^*(v) = \begin{cases} 0 & \text{if } v \geq 0 \text{ and } \sum_{i=1}^m v_i = 1 \\ \infty & \text{otherwise,} \end{cases}$$

$$h_2^*(u) = \sup_{w \in C} u\top(x - w).$$

Therefore, by considering Fenchel duality of (14), we obtain

$$u_0(x) = \min_{\eta \in \mathbb{R}^m} \left\{ \sup_{w \in C} \left[ \sum_{i=1}^m \eta_i \nabla f_i(x)\top (x - w) \right] \mid \sum_{i=1}^m \eta_i = 1, \eta \geq 0 \right\}. \quad (15)$$

Now, we assume that there exist $\bar{C} \subseteq \mathbb{R}^n$ and $\hat{f}_{\eta,x}: C \to \mathbb{R}$ such that $\min_{w \in \bar{C}} \hat{f}_{\eta,x}(\hat{w}) = \sup_{w \in C} \left[ \sum_{i=1}^m \eta_i \nabla f_i(x)\top (x - w) \right]$. Note that if $C$ is a convex polyhedron and the objective function attains a maximum in some $w \in C$, we can write $\bar{C}$ and $\hat{f}$ explicitly. Then, the optimization problem in (15) is an $m + \hat{n}$ dimensional problem, and if $m$ and $\hat{n}$ is quite smaller than $n$, we can evaluate the merit function $u_0$ by (15) more easily than by (14).
5.3 Problems that include $\ell_1$ norm

In this subsection, we suppose that the problem is unconstrained, that is, $S = \mathbb{R}^n$. We also assume that each component $F_i$ of the objective function $F$ of (1) is defined by (2), $f_i$ is continuously differentiable and $g_i(x) = \tau \|x\|_1$ for all $i \in \{1, \ldots, m\}$, where $\| \cdot \|_1$ denotes the $\ell_1$-norm of a vector. Then, the merit function $w_0$ defined by (5) can be written as

$$w_0(x) = \sup_{y \in \mathbb{R}^n} \min_{i \in \{1, \ldots, m\}} \{ \nabla f_i(x)^\top (x - y) + \tau \|x\|_1 - \tau \|y\|_1 \} = \tau \|x\|_1 + \max_{d \in \mathbb{R}^n} \left[ - \max_{i \in \{1, \ldots, m\}} \{ \nabla f_i(x)^\top d \} - \tau \|x + d\|_1 \right]. \quad (16)$$

Now, we define functions $h_1 : \mathbb{R}^m \to \mathbb{R}$ and $h_3^\tau : \mathbb{R}^n \to \mathbb{R}$ as

$$h_1(y) := \max_{i \in \{1, \ldots, m\}} y_i, \quad h_3^\tau(y) := \tau \|x - y\|_1.$$

Then, we get the conjugate functions $h_1^* : \mathbb{R}^m \to \mathbb{R}$ and $h_3^{\tau*} : \mathbb{R}^n \to \mathbb{R}$ of them as follows:

$$h_1^*(v) = \begin{cases} 0, & \text{if } v \geq 0 \text{ and } \sum_{i=1}^m v_i = 1, \\ +\infty, & \text{otherwise}, \end{cases}$$

$$h_3^{\tau*}(u) = \begin{cases} u^\top x, & \text{if } \|u_i\|_\infty \leq \tau, \\ +\infty, & \text{otherwise}. \end{cases}$$

Therefore, by considering Fenchel duality of (16), we obtain

$$w_0(x) = \tau \|x\|_1 + \min_{\eta \in \mathbb{R}^m} \left\{ \sum_{i=1}^m \eta_i \nabla f_i(x)^\top x \left| \left| \sum_{i=1}^m \eta_i \nabla f_i(x) \right| \right|_\infty \leq \tau, \right. \quad (17)$$

$$\left. \sum_{i=1}^m \eta_i = 1, \eta \geq 0 \right\}.$$

The above characterization shows that we need to solve a linear programming problem with $m$-dimensional variable. As in the previous example, if $m$ is relatively smaller then $n$, it would be easier to evaluate the merit function $w_0$ with (17) instead of (16). Once again, if two objectives are considered, we get a linear programming problem with a unique variable, and thus, we can evaluate (17) analytically.

6 Conclusion

In this paper, we proposed six types of merit functions for problem (1). We verified how these functions are related to each other, and we showed the conditions for them to be merit functions in the Pareto sense. We also presented
error bound properties of them by assuming strong convexity, and we finally suggested some efficient ways to evaluate the proposed merit functions for three types of particular problems.

For future works, we should present efficient ways of computing other merit functions. This also have connection with another work, which would be the proposal of algorithms for getting Pareto stationary solutions by using the merit functions studied here.

Acknowledgements

This work was supported by the Grant-in-Aid for Scientific Research (C) (17K00032 and 19K11840) and Grant-in-Aid for JSPS Fellows (20J21961) from Japan Society for the Promotion of Science.

References

[1] Bertsekas, D. P.: *Nonlinear Programming*, Vol. 9, Athena Scientific, Belmont, Mass., second edition, 1976.

[2] Custódio, A. L., Madeira, J. F., Vaz, A. I. and Vicente, L. N.: Direct multisearch for multiobjective optimization, *SIAM Journal on Optimization*, Vol. 21 (2011), 1109–1140.

[3] Dutta, J., Kesarwani, P. and Gupta, S.: Gap functions and error bounds for nonsmooth convex vector optimization problem, *Optimization*, Vol. 66 (2017), 1807–1836.

[4] Fliege, J., Drummond, L. M. and Svaiter, B. F.: Newton’s method for multiobjective optimization, *SIAM Journal on Optimization*, Vol. 20 (2009), 602–626.

[5] Fliege, J. and Svaiter, B. F.: Steepest descent methods for multicriteria optimization, *Mathematical Methods of Operations Research*, Vol. 51 (2000), 479–494.

[6] Fukuda, E. H. and Graña Drummond, L. M.: A survey on multiobjective descent methods, *Pesquisa Operacional*, Vol. 34 (2014), 585–620.

[7] Karimi, H., Nutini, J. and Schmidt, M.: Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition, Technical report, 2016.

[8] Liu, C. G., Ng, K. F. and Yang, W. H.: Merit functions in vector optimization, *Mathematical Programming*, Vol. 119 (2009), 215–237.

[9] Miettinen, K.: *Nonlinear Multiobjective Optimization*, Springer US, 1998.
[10] Rockafellar, R. T. and Wets, R. J. B.: Variational Analysis, Vol. 317 of Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[11] Sion, M.: On general minimax theorems, Pacific Journal of Mathematics, Vol. 8 (1958), 171–176.

[12] Tanabe, H., Fukuda, E. H. and Yamashita, N.: Proximal gradient methods for multiobjective optimization and their applications, Computational Optimization and Applications, Vol. 72 (2019), 339–361.

[13] Tanabe, H., Fukuda, E. H. and Yamashita, N.: Convergence rates analysis of multiobjective proximal gradient methods, (2020).