A proximal gradient method with Bregman distance in multi-objective optimization

Kangming Chen[1], Ellen H. Fukuda[1], and Nobuo Yamashita [1]

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

Abstract

In the last two decades, many descent algorithms were extended to solve multi-objective optimization problems. Recently, the multi-objective proximal gradient descent method was also proposed for problems where each objective function is written as the sum of a differentiable function and a proper convex but not necessarily differentiable one. However, it requires the differentiable part of each objective to have a Lipschitz continuous gradient, which limits its application. Moreover, the method solves subproblems using Euclidean distances only.

The so-called Bregman scheme is common in single-objective proximal gradient-type methods. In this case, the Euclidean distance is replaced by the more general and flexible Bregman distance. Combined with the notion of relative smoothness, we have an assumption less demanding than the Lipschitz continuity of the gradients. Thus in this work, we propose a proximal gradient method with Bregman distance for multi-objective optimization. At each iteration of our method, we compute the search direction by solving a subproblem that contains the Bregman distance. This subproblem can be solved easily depending on the choice of the Bregman scheme. We also consider two stepsize strategies: the constant stepsize and the backtracking procedure. In both cases, we prove convergence of the generated sequence to a Pareto stationary point, and analyze the convergence rate through some merit functions. Specifically we get the convergence rates for non-convex ($O(\sqrt{1/k})$), convex ($O(1/k)$), and strongly convex $\left(O\left(r^k\right)\right.$ for some $r \in (0,1)$) problems.

Keywords: multi-objective, proximal gradient method, Bregman distance, relative smooth.

This manuscript is the first author's master thesis which was submitted to Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University on Feb. 14, 2022.

1 Introduction

In this paper, we consider the following unconstrained multi-objective optimization problem:

$$\begin{aligned} \min \quad & F(x) \\ \text{s.t} \quad & x \in \mathbf{R}^n, \end{aligned} \tag{1}$$

where $F : \mathbf{R}^n \to (\mathbf{R} \cup \{\infty\})^m$ is a vector-valued function with $F := (F_1, \ldots, F_m)^\top$ and $\top$ denotes transpose. We assume that each component $F_i : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ is defined by

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

where $f_i : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ is continuously differentiable and $g_i : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ is proper convex and closed but not necessarily differentiable.

In multi-objective optimization we usually can not get a single point that minimize all given objective functions at once. So instead the Pareto optimality concept is used in multi-objective optimization.

A common solution strategy for multi-objective optimization is the scalarization approach [1], where the problem is converted to a single objective optimization using some parameters. The disadvantage of this method is that the parameters are not known in advance, requiring the users to decide them. Some adaptive scalarization techniques were also proposed in [2, 3] to choose parameters automatically during the course of the algorithm, but they require convexity of the objectives.

Recently, many descent methods for multi-objective optimization algorithms have been proposed. For instance, the steepest descent method has been discussed in [4], and Newton's method has been shown in [5]. The projected gradient method is proposed in [6, 7], the subgradient method [8] and the proximal point method [9] were also proposed. More recently, the proximal gradient method was also extended to multi-objective problem in [10]. As the notion of Bregman distance was proposed in [11] and began to be used in optimization [12], many researchers started to consider optimization methods based on Bregman distances in particular for proximal point methods [13] and proximal gradient methods [14, 15]. But in multi-objective optimization, most methods are only discussed under the Euclidean distance, and did not discuss the case of Bregman distance in detail. In this paper we will propose the proximal gradient method with Bregman distance in multi-objective optimization.

We also use the concept of relative smoothness here. A similar notion was proposed in [16] as a new descent lemma without Lipschitz gradient continuity, where the reference function is required to a Legendre function. Recently Lu et al. gave the definition of relative smoothness and relative strong convexity in [17] showing that it is less strict than the usual Lipschitz continuity of the gradients assumption. Similarly, in this work, we will assume the less restrictive relative smoothness for the differentiable part of the objective function, making adaptations to deal with multi-objectives. We observe that the use of Bregman distance is interesting for the computational point of view, because the subproblems in our algorithm may become easier to solve. Furthermore, we will consider two types of stepsizes, and for both of them, we will show convergence to Pareto stationary points and convergence rates.

The outline of this paper is as follows. In Section 2, we present proximal gradient method for multi-objective problems, definition of Bregman function and some preliminary materials. In Section 3, we propose a proximal gradient method with Bregman function for multi-objective optimization considering both constant stepsizes and backtracking procedure. Section 4 contains the proof of convergence to Pareto stationary points. And we prove the convergence rates for non-convex ($O(\sqrt{1/k})$), convex ($O(1/k)$), and strongly convex $\left(O\left(r^k\right)\right.$ for some $r \in (0,1)$) problems.

2 Preliminaries

First we present some notations used in this paper. $\mathbf{R}$ denotes the set of real numbers and $\mathbf{N}$ denotes the set of positive integers. The symbol $\|\cdot\|$ stands for the Euclidean norm in $\mathbf{R}^n$. We define Jf is the Jacobian matrix of function f. We also define the relation $\leq$ ($<$) in $\mathbf{R}^m$ as $u \leq v (u < v)$ if and only if $u_i \leq v_i$ $(u_i < v_i)$ for all $i = 1, \ldots, m$.

One way to solve problem (1) is by using the multi-objective proximal gradient method, proposed in [10]. Define $\tilde{\psi}_x : \mathbf{R}^n \to \mathbf{R}$ by

$$\tilde{\psi}_x(d) := \max_{i \in \{1,\ldots,m\}} \left\{ \nabla f_i(x)^\top d + g_i(x+d) - g_i(x) \right\}.$$

The multi-objective proximal gradient method generates a sequence $\left\{x^k\right\}$ iteratively with the following procedure:

$$x^{k+1} := x^k + t_k d^k,$$

where d^k is a search direction and t_k is a stepsize. At every iteration k, we define this d^k by solving

$$d^k := \operatorname*{argmin}_{d \in \mathbf{R}^n} \left\{ \tilde{\psi}_{x^k}(d) + \frac{\ell}{2}\|d\|^2 \right\}.$$

Note that this is equal to the following problem when $t_k = 1$:

$$x^{k+1} = \operatorname*{argmin}_{x \in \mathbf{R}^n} \psi_{x^k}(x) + \frac{\ell}{2}\left\|x - x^k\right\|^2, \tag{2}$$

where

$$\psi_{x^k}(x) = \max_{i=1,\ldots,m} \nabla f_i\left(x^k\right)^\top \left(x - x^k\right) + g_i(x) - g_i\left(x^k\right).$$

It is also proved that each accumulation point of the sequence generated by this method with and without line searches, if it exists, is Pareto stationary [10].

In this paper, we will replace the norm distance used in subproblem (2) with a distance-like function, called Bregman function, which we define below.

Definition 2.1 (Bregman distance)**.** *[18] Let* $\omega : \mathbf{R}^n \to (-\infty, \infty]$ *be a proper closed and convex function that is differentiable over* $\operatorname{dom}(\partial\omega)$. *The Bregman distance associated with* ω *is the function* $B_\omega : \operatorname{dom}(\omega) \times \operatorname{dom}(\partial\omega) \to \mathbf{R}$ *given by*

$$B_\omega(x,y) = \omega(x) - \omega(y) - \langle \nabla\omega(y), x - y \rangle.$$

The assumptions on ω (given a set C) are gathered in the following.

Assumption 2.1. *Let $\omega : \mathbf{R}^n \to (-\infty, \infty]$ be defined as in Definition 1. We assume the following:*

- *ω is proper closed and convex.*
- *ω is differentiable over $\operatorname{dom}(\partial\omega)$.*
- *$C \subseteq \operatorname{dom}(\omega)$.*
- *$\omega + \delta_C$ is σ-strongly convex ($\sigma > 0$).*

Now we give the basic properties for general Bregman distances satisfying the above assumption. We also present some examples below.

Lemma 2.1. *Suppose that $C \subseteq \mathbb{R}^n$ is nonempty closed and convex and that ω satisfies the properties in Assumption 2.1. Let B_ω be the Bregman distance associated with ω. Then*
(a) $B_\omega(x, y) \geq \frac{\sigma}{2}\|x - y\|^2$ for all $x \in C, y \in C \cap \operatorname{dom}(\partial\omega)$.
(b) Let $x \in C$ and $y \in C \cap \operatorname{dom}(\partial\omega)$. Then

- *$B_\omega(x, y) \geq 0$,*
- *$B_\omega(x, y) = 0$ if and only if $x = y$.*

Example 2.1 (Euclidean Distance)**.** *If $\omega : \mathbb{R}^n \to \mathbb{R}$ with $\omega(x) = \frac{1}{2}\|x\|^2$, then ω satisfies Assumption 2.1 with $\sigma = 1$ and $B_\omega(x, y) = \frac{1}{2}\|x - y\|^2$.*

Example 2.2 (KL Relative Entropy)**.** *If $\omega : \mathbb{R}^n_+ \to \mathbb{R}$ with $\omega(x) = \sum_{j=1}^n x_j \log x_j$ (with the convention $0 \log 0 = 0$), then ω satisfies Assumption 2.1 with $C = \left\{x \in \mathbf{R}^n \mid x \geqslant 0, \sum_{i=1}^n x_i = 1\right\}$, $\sigma = 1$ and $B_\omega(x, y) = \sum_{j=1}^n x_j \log \frac{x_j}{y_j} - \sum_{i=j}^n \left(x_j - y_j\right)$.*

Lemma 2.2 (Three-points lemma)**.** *[13] Suppose that $\omega : \mathbf{R}^n \to (-\infty, \infty]$ satisfies the Assumption 2.1. Assume that $a, b \in \operatorname{dom}(\partial\omega)$ and $c \in \operatorname{dom}(\omega)$. Then the following equality holds:*

$$\langle \nabla\omega(b) - \nabla\omega(a), c - a\rangle = B_\omega(c, a) + B_\omega(a, b) - B_\omega(c, b).$$

The next lemma is essential in the analysis of convergence of the proximal gradient method with Bregman distance.

Lemma 2.3. *[13] For any proper closed convex function $\theta : \mathbf{R}^n \to (-\infty, \infty]$ and any $z \in \mathbf{R}^n$, if ω is differentiable at $z_+ = \underset{x}{\operatorname{argmin}}\{\theta(x) + B_\omega(x, z)\}$, then*

$$\theta(x) + B_\omega(x, z) \geq \theta(z_+) + B_\omega(z_+, z) + B_\omega(x, z_+) \quad \forall x \in \operatorname{dom}(\omega).$$

And the following lemma is useful to proof the existence of optimal solution.

Lemma 2.4. *[18] Assume the following:*

- $\omega : \mathbf{R}^n \to (-\infty, \infty]$ *is a proper closed and convex function differentiable over* $\mathrm{dom}(\partial\omega)$.
- $\varphi : \mathbf{R}^n \to (-\infty, \infty]$ *is a proper closed and convex function satisfying* $\mathrm{dom}(\varphi) \subseteq \mathrm{dom}(\omega)$.
- $\omega + \delta_{\mathrm{dom}(\varphi)}$ *is* σ*-strongly convex* ($\sigma > 0$).

Then the minimizer of the problem

$$\min_{x \in \mathbf{R}^n} \{\varphi(x) + \omega(x)\}$$

is uniquely attained at a point in $\mathrm{dom}(\varphi) \cap \mathrm{dom}(\partial\omega)$.

Definition 2.2 (Relative smoothness)**.** *[17] A function f is called L-smooth relative to ω on* $\mathrm{dom}(\omega)$ *if for any $x, y \in$* $\mathrm{dom}(\omega)$*, there is a scalar L for which*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + L B_\omega(y, x).$$

The definition of relative smooth gives an upper approximation of f that is similar to the so-called descend lemma. In fact, it is a special case of relative smooth with $\omega = \frac{1}{2}\|x\|^2$ and $B_\omega(x, y) = \frac{1}{2}\|x - y\|^2$.

In optimization problems, many methods assume that $\nabla f(x)$ satisfies a Lipschitz condition with a constant L, as defined as

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

to ensure associated computational guarantees. But this is a strict condition for functions. In many applications the differentiable function does not have such a property. Even if the condition is satisfied, L may be too large. For example, let $f(x) = -\ln(x) + x^2$ on $Q = \mathbf{R}_{++}$, and consider the level set $\{x : f(x) \leq 10\}$. It still has $L \approx \exp^{20}$ on this level set. Compared to Lipschitz condition, the notion of "relative smoothness" using a function $\omega(x)$ to be a "reference function" is less restrictive. The function $\omega(x)$ does not require the specification of any particular norm and it need not be either strictly or strongly convex.

Definition 2.3 (Relative strongly convexity)**.** *[17] A function f is called μ-strongly convex relative to $h(\cdot)$ on* $\mathrm{dom}(\omega)$ *if for any $x, y \in$* $\mathrm{dom}(\omega)$*, there is a scalar $\mu \geq 0$ for which*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \mu B_\omega(y, x).$$

Now we introduce the Pareto optimal related concept for multi-objective optimization problem. First, we introduce the directional derivative of $f : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ at x in the direction d,

$$f'(x; d) := \lim_{\alpha \searrow 0} \frac{f(x + \alpha d) - h(x)}{\alpha}.$$

Here we follow the notation used in [10]. Recall that $x^* \in \mathbf{R}^n$ is a Pareto optimal point for F, if there is no $x \in \mathbf{R}^n$ such that $F(x) \leq F(x^*)$ and $F(x) \neq F(x^*)$. The set of all Pareto optimal values is called Pareto frontier. Likewise, $x^* \in \mathbf{R}^n$ is a weakly Pareto optimal point for F, if there is no $x \in \mathbf{R}^n$ 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 recall that $\bar{x} \in \mathbf{R}^n$ is Pareto stationary (or critical), if and only if,

$$\max_{i=1,\ldots,m} F_i'(\bar{x}; d) \geq 0 \text{ for all } d \in \mathbf{R}^n.$$

Lemma 2.5. *The following assertions hold:*

1. If $x \in \mathbf{R}^n$ is a weakly Pareto optimal point of F, then x is Pareto stationary.

2. Let every component F_i of F be convex. If $x \in \mathbf{R}^n$ is a Pareto stationary point of F, then x is weakly Pareto optimal.

3. Let every component F_i of F be strictly convex. If $x \in \mathbf{R}^n$ is a Pareto stationary point of F, then x is Pareto optimal.

3 The multi-objective proximal gradient method with Bregman distance

In this section we explain in details the proposed proximal gradient descent method with Bregman distance. For now on, we suppose that the following assumption holds.

Assumption 3.1. *We assume that*

- *The function $\omega : \mathbf{R}^n \to (-\infty, \infty]$ satisfies Assumption 2.1 with $C = \mathrm{dom}(g)$ and $\sigma = 1$.*
- *For all i, $f_i(x)$ is L^i-smooth relative to $\omega(x)$. Moreover, $L = \max_{i=1,\ldots,m} L^i$.*

For each iteration, define $\psi_{x^k} : \mathbf{R}^n \to \mathbf{R}$ as

$$\psi_{x^k}(x) = \max_{i=1\ldots m} \nabla f_i\left(x^k\right)^\top \left(x - x^k\right) + g_i(x) - g_i\left(x^k\right). \tag{3}$$

At each iteration k, we consider the mininization of the following function as a subproblem. Define $\phi_{x^k} : \mathbf{R}^n \to \mathbf{R}$ as

$$\phi_{x^k}(x) = \psi_{x^k}(x) + L_k B_\omega\left(x, x^k\right), \tag{4}$$

where $L_k \geq L$. Moreover, we define the following:

$$x^{k+1} = p_{L_k}(x^k) := \operatorname*{argmin}_{x \in \mathbf{R}^n} \phi_{x^k}(x), \tag{5}$$

$$\theta(x^k) := \min_{x \in \mathbf{R}^n} \phi_{x^k}(x) = \phi_{x^k}\left(p_{L_k}(x^k)\right). \tag{6}$$

From the definition of $\phi_{x^k}(x)$, we know $\phi_{x^k}(x^k) = 0$, so $\theta(x^k) \leq 0$. Then,

$$\psi_{x^k}(x^{k+1}) + L_k B_\omega\left(x^{k+1}, x^k\right) \leq 0.$$

Proposition 3.1. *The function $p_{L_k}(x^k)$ is well-defined for all k.*

Proof. From the definition of $B_\omega(x, x^k)$, we have

$$\begin{aligned}
x^{k+1} &= \underset{x\in\mathbf{R}^n}{\operatorname{argmin}}\left\{\max_{i=1,\ldots,m}\left\{\nabla f_i\left(x^k\right)^\top\left(x-x^k\right)+g_i(x)-g_i\left(x^k\right)+L_kB_\omega\left(x,x^k\right)\right\}\right\}\\
&= \underset{x\in\mathbf{R}^n}{\operatorname{argmin}}\left\{\max_{i=1,\ldots,m}\left\{\nabla f_i\left(x^k\right)^\top\left(x-x^k\right)+g_i(x)-g_i\left(x^k\right)+L_k\left(\omega(x)-\omega\left(x^k\right)\right.\right.\right.\\
&\qquad\left.\left.\left.-\nabla\omega\left(x^k\right)^\top\left(x-x^k\right)\right)\right\}\right\}\\
&= \underset{x\in\mathbf{R}^n}{\operatorname{argmin}}\left\{\max_{i=1,\ldots,m}\left\{\left\langle\nabla f_i\left(x^k\right)-L_k\nabla\omega\left(x^k\right),x-x^k\right\rangle+g_i\left(x\right)-g_i\left(x^k\right)\right.\right.\\
&\qquad\left.\left.+L_k\omega(x)-L_k\omega\left(x^k\right)\right\}\right\}\\
&= \underset{x\in\mathbf{R}^n}{\operatorname{argmin}}\left\{\max_{i=1,\ldots,m}\left\{\left\langle\frac{1}{L_k}\nabla f_i\left(x^k\right)-\nabla\omega\left(x^k\right),x\right\rangle+\frac{1}{L_k}g_i(x)+\omega(x)\right\}\right\}.
\end{aligned}$$

Let $\varphi(x) = \max_{i=1,\ldots,m}\left\{\left\langle\frac{1}{L_k}\nabla f_i\left(x^k\right)-\nabla\omega\left(x^k\right),x\right\rangle+\frac{1}{L_k}g_i(x)\right\}$. The function φ is closed if f, g and ω are closed; it is proper by the fact that $dom(g)\cap dom(\omega)\neq\emptyset$. Since g is convex, φ is convex. To conclude, φ is proper closed and convex function, and hence, by Lemma 2.4 the subproblem has a unique optimal solution in $dom(g)\cap dom(\omega)$. It means that $p_{L_k}(x^k)$ is well defined. □

Lemma 3.1. *Let $\left\{x^k\right\}$ be generated iteratively with (5) and recall the definition of ψ_{x^k} in (3). Then, we have*

$$\psi_{x^k}\left(x^{k+1}\right)\leq -L_k\left\|x^{k+1}-x^k\right\|^2\quad \textit{for all } k. \tag{7}$$

Proof. According to the Lemma 2.3, we have

$$\theta(x)+B_\omega(x,z)\geq\theta\left(z_+\right)+B_\omega\left(z_+,z\right)+B_\omega\left(x,z_+\right)\quad\forall x\in\operatorname{dom}(\omega).$$

By letting $z_+ = x^{k+1}, z = x^k, x = x^k$, and $\theta = L_k^{-1}\psi_{x^k}$, we have

$$\psi_{x^k}(x^k)+L_kB_\omega(x^k,x^k)\geq\psi_{x^k}\left(x^{k+1}\right)+L_kB_\omega\left(x^{k+1},x^k\right)+L_kB_\omega\left(x^k,x^{k+1}\right).$$

From the Lemma 2.1(b) and the fact that $\psi_{x^k}(x^k) = 0$, we have

$$\begin{aligned}
\psi_{x^k}\left(x^{k+1}\right)&\leq -L_kB_\omega\left(x^{k+1},x^k\right)-L_kB_\omega\left(x^k,x^{k+1}\right)\\
&\leq -L_k\left\|x^{k+1}-x^k\right\|^2,
\end{aligned}$$

where the last inequality follows from Lemma 2.1(a). Thus, we obtain,

$$\psi_{x^k}\left(x^{k+1}\right)\leq -L_k\left\|x^{k+1}-x^k\right\|^2.$$

□

From the relative smoothness of f_i, if $L_k \geq L$, we have

$$F_i\left(x^{k+1}\right) - F_i\left(x^k\right) \leq \nabla f_i\left(x^k\right)^\top (x^{k+1} - x^k) + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + L_k B_\omega\left(x^{k+1}, x^k\right). \tag{8}$$

Since x^{k+1} is the optimal solution of (5), the maximum in i of the right right-hand side of (8) is less than or equal to zero. Thus,

$$F_i\left(x^{k+1}\right) \leq F_i\left(x^k\right) \quad \text{for all } k. \tag{9}$$

In the following subsections, we consider two stepsize rules for our method.

3.1 Constant stepsize

Now consider the constant stepsize, which we set as $L_k = \bar{L}$ for all k,with $\bar{L} > L$. Then the proximal gradient method with Bregman distance is given below.

Algorithm 3.1. *Multi-objective proximal gradient method with Bregman distance and constant stepsize*

Step 1 Choose $L_k = \bar{L}$ *with* $\bar{L} > L, \varepsilon > 0, x^0 \in \mathrm{dom}(g) \cap \mathrm{dom}(\partial\omega)$ *and set* $k := 0$.
Step 2 Compute $p_{L_k}(x^k)$ *by solving subproblem (5).*
Step 3 If $\left\|p_{L_k}(x^k) - x^k\right\| < \varepsilon$*, then stop.*
Step 4 Set $x^{k+1} := p_{L_k}(x^k), k := k+1$*, and go to Step 2 .*

3.2 Backtracking procedure

Now we consider backtracking procedure. In the beginning, let $L_{-1} = s$ with $s > 0$. At iteration $k \geq 0$, let $L_k = L_{k-1}$. Then, while exist i such that

$$f_i\left(p_{L_k}(x^k)\right) > f_i\left(x^k\right) + \left\langle \nabla f_i\left(x^k\right), p_{L_k}(x^k) - x^k\right\rangle + \frac{L_k}{2}\left\|p_{L_k}(x^k) - x^k\right\|^2,$$

we set $L_k := \eta L_k$ where $\eta > 1$. In other words, the stepsize is chosen as $L_k = L_{k-1}\eta^{j_k}$, where j_k is the smallest nonnegative integer for which the condition

$$\begin{aligned} j_k := \underset{j \in \mathbf{N}}{\operatorname{argmin}} \Big\{ f\left(p_{L_{k-1}\eta^{j_k}}(x^k)\right) &\leq f\left(x^k\right) + Jf(x^k)^\top (p_{L_{k-1}\eta^j}(x^k) - x^k) \\ &+ \frac{L_{k-1}\eta^j}{2}\left\|p_{L_{k-1}\eta^j}(x^k) - x^k\right\|^2 \Big\}, \end{aligned} \tag{10}$$

is satisfied.

Letting $d^k = x^{k+1} - x^k$, from the so-called descent lemma [19, Proposition A.24], for all $i \in 1, ..., m$, if $L_k \geq L$, we have

$$F_i\left(x^{k+1}\right) - F_i\left(x^k\right) \leq \nabla f_i\left(x^k\right)^\top d^k + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + \frac{L_k}{2}\left\|d^k\right\|^2. \tag{11}$$

The stepsize rule above ensure that (11) is still satisfied at each iteration. In addition, the step L_k that the backtracking procedure produces satisfy the following bounds for all $k \geq 0$:

$$s \leq L_k \leq \max\{\eta L, s\}.$$

The inequality $s \le L_k$ is obvious. To prove the inequality $L_k \le \max\{\eta L, s\}$, we note that either $L_k = s$ or $L_k > s$. In the latter case there exists an index $0 \le k' \le k$ for which the inequality (11) is not satisfied with $k = k'$ and $\frac{L_k}{\eta}$ replacing L_k. By the descent lemma, this implies in particular that $\frac{L_k}{\eta} < L$, and we have thus shown that $L_k \le \max\{\eta L, s\}$. Namely, $L_k \le \alpha L$, where $\alpha = \max\left\{\eta, \frac{s}{L}\right\}$.We also note that the bounds on L_k can be rewritten as

$$\beta L \le L_k \le \alpha L,$$

where

$$\alpha = \begin{cases} \frac{\bar{L}}{L}, & \text{constant,} \\ \max\left\{\eta, \frac{s}{L}\right\}, & \text{backtracking,} \end{cases} \qquad \beta = \begin{cases} \frac{\bar{L}}{L}, & \text{constant} \\ \frac{s}{L}, & \text{backtracking.} \end{cases}$$

So the algorithm with backtracking stepsize is given below.

Algorithm 3.2. *Multi-objective proximal gradient method with Bregman distance and backtracking procedure*

Step 1 Choose $s > 0$, $\eta > 1$, $x^0 \in \mathrm{dom}(g) \cap \mathrm{dom}(\partial\omega)$ *and set* $L_{-1} = s$, $k := 0$.
Step 2 Compute L_k *by solving* (10).
Step 3 Compute $p_{L_k}(x^k)$ *by solving subproblem (5).*
Step 4 If $\left\|p_{L_k}(x^k) - x^k\right\| < \varepsilon$, *then stop.*
Step 5 Set $x^{k+1} := p_{L_k}(x^k)$, $k := k+1$, *and go to Step 2 .*

4 Convergence analysis

In this section, we prove that the sequences generated by Algorithm 3.1 and 3.2 converge to Pareto stationary points and discuss their rate of convergence.

4.1 Convergence to Pareto stationary points

Now we analysis the convergence of the proximal gradient method with bregman distance.

Lemma 4.1. *let* x^k *be generated by Algorithm* 3.1 *or Algorithm* 3.2 *and suppose that* $F_i(x^k)$ *is bouded from below for all* $i = 1, \dots, m$. *Then we have*

$$\lim_{k\to\infty} \left\|x^{k+1} - x^k\right\| = 0. \tag{12}$$

Proof. At the kth iteration,

$$\begin{aligned}
& f_i\left(x^{k+1}\right) + g_i\left(x^{k+1}\right) \\
=& f_i\left(x^k\right) + g_i\left(x^k\right) + f_i\left(x^{k+1}\right) - f_i\left(x^k\right) + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) \\
\leqslant& f_i\left(x^k\right) + g_i\left(x^k\right) + \nabla f_i\left(x^k\right)\left(x^{k+1} - x^k\right) + LB_\omega\left(x^{k+1}, x^k\right) + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) \\
\leqslant& f_i\left(x^k\right) + g_i\left(x^k\right) + \psi_{x^k}\left(x^{k+1}\right) + LB_\omega\left(x^{k+1}, x^k\right) + L_k B_\omega\left(x^{k+1}, x^k\right) - L_k B_\omega\left(x^{k+1}, x^k\right) \\
\leqslant& f_i\left(x^k\right) + g_i\left(x^k\right) + \psi_{x^k}(x) + L_k\left(B_\omega\left(x, x^k\right) - B_\omega\left(x, x^{k+1}\right)\right) + (L - L_k)\, B_\omega\left(x^{k+1}, x^k\right)
\end{aligned}$$

for all x. Here, the first inequality follows from the relative smoothness of f_i. The second inequality follows from the definition of $\psi_{x^k}\left(x^{k+1}\right)$. And the third inequality follows from the Lemma 2.3. Letting $x = x^k$, we obtain

$$\begin{aligned}
& f_i(x^{k+1}) + g_i(x^{k+1}) \\
\leqslant & f_i\left(x^k\right) + g_i\left(x^k\right) - L_k B_\omega\left(x^k, x^{k+1}\right) + (L - L_k)\, B_\omega\left(x^{k+1}, x^k\right) \\
\leqslant & f_i\left(x^k\right) + g_i\left(x^k\right) + (L - L_k)\, B_\omega\left(x^{k+1}, x^k\right).
\end{aligned}$$

Since $\left\{F_i\left(x^k\right)\right\}$ is bounded from below from (9), there exists $\tilde{F}_i \leq F_i\left(x^k\right) = f_i\left(x^k\right) + g_i\left(x^k\right)$ for all i, k. Adding up the above inequality from $k = 0$ to $k = \hat{k}$, we obtain

$$\begin{aligned}
& f_i\left(x^{\hat{k}+1}\right) + g_i\left(x^{\hat{k}+1}\right) \\
\leq & f_i\left(x^0\right) + g_i\left(x^0\right) + \sum_{k=0}^{\hat{k}} (L - L_k)\, B_\omega\left(x^{k+1}, x^k\right).
\end{aligned}$$

If $\{x^k\}$ be generated by Algorithm 3.1, then for all k, $L_k = \bar{L}$. Then,

$$\begin{aligned}
& f_i\left(x^{\hat{k}+1}\right) + g_i\left(x^{\hat{k}+1}\right) \\
\leq & f_i\left(x^0\right) + g_i\left(x^0\right) + (L - \bar{L}) \sum_{k=0}^{\hat{k}} B_\omega\left(x^{k+1}, x^k\right).
\end{aligned}$$

Because $L < \bar{L}$, we have

$$\sum_{k=0}^{\hat{k}} B_\omega\left(x^{k+1}, x^k\right) \leq \left(\bar{L} - L\right)^{-1} \left(f_i\left(x^0\right) + g_i\left(x^0\right) - f_i\left(x^{\hat{k}+1}\right) - g_i\left(x^{\hat{k}+1}\right)\right),$$

so

$$\sum_{k=0}^{\hat{k}} B_\omega\left(x^{k+1}, x^k\right) < \infty,$$

which is equal to

$$\lim_{k\to\infty} B_\omega\left(x^{k+1}, x^k\right) = 0,$$

namely,

$$\lim_{k\to\infty} \left\|x^{k+1} - x^k\right\| = 0.$$

Let $\{x^k\}$ be generated by Algorithm 3.2, considering the definition of L and L_k. Since L is finite, the backtracking only needs a finite number of steps to make $L_k > L$. So, without losing the generality, we can assume $L_k > L$. Then the rest of the proof is similar to the case $L_k = \bar{L}$. □

Theorem 4.1. *Every accumulation point of the sequence $\left\{x^k\right\}$ generated by Algorithm 3.1 or 3.2, if it exists, is a Pareto stationary point.*

Proof. From the inequation (11) , we can obtain

$$\begin{aligned}
F_i\left(x^{k+1}\right)-F_i\left(x^k\right) &\le \nabla f_i\left(x^k\right)^\top (x^{k+1}-x^k)+g_i\left(x^{k+1}\right)-g_i\left(x^k\right)+\frac{L_k}{2}\left\|x^{k+1}-x^k\right\|^2\\
&\le \max_{1,\dots,m} \nabla f_i\left(x^k\right)^\top (x^{k+1}-x^k)+g_i\left(x^{k+1}\right)-g_i\left(x^k\right)+\frac{L_k}{2}\left\|x^{k+1}-x^k\right\|^2\\
&= -w_{L_k}\left(x^k\right)\\
&\le 0.
\end{aligned}$$

Namely,

$$F_i\left(x^{k+1}\right)-F_i\left(x^k\right)\le -w_{L_k}\left(x^k\right)\le 0. \tag{13}$$

From Lemma 4.1, we have

$$\lim_{k\to\infty}\left\|x^{k+1}-x^k\right\|=0.$$

Let $x^*\in\mathrm{dom}(g)\cap\mathrm{dom}(\partial\omega)$ be an accumulation point of the sequence $\left\{x^k\right\}$ generated by Algorithm 3.1 or 3.2 and assume that $\lim_{k\to\infty}x^k=x^*$.

If $\left\{x^k\right\}$ is generated by Algorithm 3.1, then $L_k=\bar{L}$. From (13), we get $w_{\bar{L}}(x^*)=0$, and according to [20, Theorem 3.1(ii)] , x^* is the Pareto stationary point of problem (1).

If $\left\{x^k\right\}$ is generated by Algorithm 3.2, then $L_k\le\alpha L$, where $\alpha=\max\left\{\eta,\frac{s}{L}\right\}$. From [21, Theorem 3.2], we have

$$-w_{L_k}\left(x^k\right)\le -w_{\alpha L}\left(x^k\right)\le 0.$$

Combined with (13), we obtain

$$F_i\left(x^{k+1}\right)-F_i\left(x^k\right)\le -w_{\alpha L}\left(x^k\right)\le 0.$$

Then, in the same way, we conclude that x^* is a Pareto stationary point of problem (1). □

4.2 Convergence rate analysis

We introduce two merit functions for the multi-objective optimization problem, and use them to estimate the convergence rate.

The first merit function is the simple function $u_0:\mathbf{R}^n\to\mathbf{R}\cup\{\infty\}$ defined as follows:

$$u_0(x):=\sup_{y\in\mathbf{R}^n}\min_{i\in\{1,\dots,m\}}\left\{F_i(x)-F_i(y)\right\}.$$

The second type is the regularized and partially linearized merit function $w_\ell:\mathbf{R}^n\to\mathbf{R}$, given by

$$w_\ell(x):=\max_{y\in\mathbf{R}^n}\min_{i\in\{1,\dots,m\}}\left\{\nabla f_i(x)^\top(x-y)+g_i(x)-g_i(y)-\frac{\ell}{2}\|x-y\|^2\right\},$$

where $\ell>0$ is a given constant.

Lemma 4.2. *Let $\{x^k\}$ be a sequence generated by Algorithm 3.1 or 3.2. Then, the following inequality holds for all k:*

$$w_{L_k}(x^k) \geq -\phi_{x^k}(x^{k+1}).$$

Proof. Let $\varphi_i(x, y) = \nabla f_i(x)^\top (y - x) + g_i(y) - g_i(x)$, then

$$\begin{aligned} w_\ell(x^k) &= \max_{y \in \mathbf{R^n}} \min_{i \in \{1,\ldots,m\}} \left\{ \nabla f_i(x^k)^\top (x^k - y) + g_i(x^k) - g_i(y) - \frac{\ell}{2}\|x^k - y\|^2 \right\}. \\ &= \max_{y \in \mathbf{R^n}} \min_{i \in \{1,\ldots,m\}} \left\{ -\varphi_i(x^k) - \frac{\ell}{2}\|y - x^k\|^2 \right\} \\ &= \max_{y \in \mathbf{R^n}} \left\{ - \max_{i \in \{1,\ldots,m\}} \varphi_i(x^k) - \frac{\ell}{2}\|y - x^k\|^2 \right\} \\ &= \max_{y \in \mathbf{R^n}} \left\{ -\psi_{x^k}(y) - \frac{\ell}{2}\|y - x^k\|^2 \right\} \\ &= -\min_{y \in \mathbf{R^n}} \left\{ \psi_{x^k}(y) + \frac{\ell}{2}\|y - x^k\|^2 \right\}. \end{aligned}$$

According to the Lemma 2.1(a),

$$\begin{aligned} \phi_{x^k}(x^{k+1}) &= \psi_{x^k}(x^{k+1}) + L_k B_\omega\left(x^{k+1}, x^k\right) \\ &= \max_{i \in \{1,\ldots,m\}} \nabla f_i\left(x^k\right)^\top \left(x^{k+1} - x^k\right) + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + L_k B_\omega\left(x^{k+1}, x^k\right) \\ &\geqslant \max_{i \in \{1,\ldots,m\}} \nabla f_i\left(x^k\right)^\top \left(x^{k+1} - x^k\right) + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + \frac{L_k}{2}\left\|x^{k+1} - x^k\right\|^2 \\ &= \psi_{x^k}(x^{k+1}) + \frac{L_k}{2}\|x^{k+1} - x^k\|^2 \\ &\geqslant -w_{L_k}\left(x^k\right), \end{aligned}$$

that is,

$$w_{L_k}(x^k) \geq -\phi_{x^k}(x^{k+1}).$$

□

4.2.1 The non-convex case

Here we use the function $w_\ell(x)$ to analyze the convergence rate.

Theorem 4.2. *Suppose that there exists some nonempty set $\mathcal{J} \subseteq \{1, \ldots, m\}$ such that if $i \in \mathcal{J}$ then $F_i(x)$ has a lower bound $F_i^{\min}$ for all $x \in \mathbf{R}^n$. Let $F^{\min} := \min_{i \in \mathcal{J}} F_i^{\min}$ and $F_0^{\max} := \max_{i \in \{1,\ldots,m\}} F_i\left(x^0\right)$. Then, the Algorithm 3.1 (or Algorithm 3.2) generates a sequence $\left\{x^k\right\}$ such that*

$$\min_{0 \leq j \leq k-1} w_1\left(x^j\right) \leq \frac{\left(F_0^{\max} - F^{\min}\right) \max\{1, \alpha L\}}{k},$$

where $\alpha = \frac{\bar{L}}{L}$ in the constant stepsize setting and $\alpha = \max\left\{\eta, \frac{s}{L}\right\}$ if the backtracking rule is employed.

Proof. Let $i \in \mathcal{J}$. From (11), we have

$$F_i\left(x^{k+1}\right) - F_i\left(x^k\right) \le \nabla f_i\left(x^k\right)^\top d^k + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + \frac{L_k}{2}\left\|d^k\right\|^2$$
$$\le \max_{i\in\{1,\dots,m\}}\left\{\nabla f_i\left(x^k\right)^\top d^k + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + \frac{L_k}{2}\left\|d^k\right\|^2\right\} = -w_{L_k}\left(x^k\right),$$

Adding up the above inequality from $k = 0$ to $k = \tilde{k} - 1$ yields that

$$F_i\left(x^{\tilde{k}}\right) - F_i\left(x^0\right) \le -\sum_{k=0}^{\tilde{k}-1} w_\ell\left(x^k\right) \le -\tilde{k}\min_{0\le k\le\tilde{k}-1} w_{L_k}\left(x^k\right).$$

From the definitions of $F^{\min}$ and $F_0^{\max}$, we obtain

$$\min_{0\le k\le\tilde{k}-1} w_{L_k}\left(x^k\right) \le \frac{F_0^{\max} - F^{\min}}{\tilde{k}}.$$

Finally, from [21, Theorem 3.2], we get

$$\min_{0\le k\le\tilde{k}-1} w_1\left(x^k\right) \le \frac{\left(F_0^{\max} - F^{\min}\right)\max\{1, L_k\}}{\tilde{k}}.$$

Namely ,

$$\min_{0\le j\le k-1} w_1\left(x^j\right) \le \frac{\left(F_0^{\max} - F^{\min}\right)\max\{1, \alpha L\}}{k},$$

where $\alpha = \frac{\bar{L}}{L}$ in the constant stepsize setting and $\alpha = \max\left\{\eta, \frac{s}{L}\right\}$ if the backtracking rule is employed. □

4.2.2 The convex case

Here we use the function $u_0(x^k)$ to analyze the convergence rate. First we give the following lemma. Note that we state it with f_i and g_i having general convexity parameters, which turn out to be zero in this subsection. It will not make a difference here, but it will be important in the discussion of the next subsection.

Lemma 4.3. *Assume that f_i is μ_i-strongly convex relative to ω and g_i has convexity parameters $\nu_i \in \mathbf{R}$, and write $\mu := \min_{i\in\{1,\dots,m\}} \mu_i$ and $\nu := \min_{i\in\{1,\dots,m\}} \nu_i$. Then, for all $x \in \mathbf{R}^n$ it follows that*

$$\sum_{i=1}^m \beta_i^k\left(F_i\left(x^{k+1}\right) - F_i(x)\right) \leqslant L_k\left(B_\omega\left(x, x^k\right) - B_\omega\left(x, x^{k+1}\right)\right)$$
$$- \mu B_\omega\left(x, x^k\right) - \nu B_\omega\left(x, x^{k+1}\right),$$

where β_i^k satisfies the following conditions:

(i) There exists $\eta_i^k \in \partial g_i\left(x^{k+1}\right)$ such that

$$\sum_{i=1}^{m} \beta_i^k \left(\nabla f_i\left(x^k\right) + \eta_i^k\right) + L_k\left(\nabla\omega\left(x^{k+1}\right) - \nabla\omega\left(x^k\right)\right) = 0,$$

(ii) $\sum_{i=1}^m \beta_i^k = 1, \beta_i^k \geq 0 \left(i \in \mathcal{I}_{x^k}\left(x^{k+1}\right)\right)$ and $\beta_i^k = 0 \left(i \notin \mathcal{I}_{x^k}\left(x^{k+1}\right)\right)$,
where $\mathcal{I}_{x^k}(x^{k+1}) := \left\{i \in \{1,\ldots,m\} \mid \psi_{x^k}(x^{k+1}) = \nabla f_i(x^k)^\top d + g_i(x^{k+1}) - g_i(x^k)\right\}$.

Proof. As we know, for all i, from the relative smoothness of f_i, we have

$$F_i\left(x^{k+1}\right) - F_i\left(x^k\right) \leq \nabla f_i\left(x^k\right)^\top\left(x^{k+1} - x^k\right) + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + L_k B_\omega\left(x^{k+1}, x^k\right). \tag{14}$$

The above inequality and relative strong convexity of f_i with modulus μ_i give

$$\begin{aligned}
&F_i\left(x^{k+1}\right) - F_i(x) = \left(F_i\left(x^k\right) - F_i(x)\right) + \left(F_i\left(x^{k+1}\right) - F_i\left(x^k\right)\right) \\
&\leq \left(\nabla f_i\left(x^k\right)^\top\left(x^k - x\right) - \frac{\mu_i}{2} B_\omega\left(x, x^k\right) + g_i\left(x^k\right) - g_i(x)\right) \\
&\quad + \left(\nabla f_i\left(x^k\right)^\top\left(x^{k+1} - x^k\right) + g_i\left(x^{k+1}\right) - g_i\left(x^k\right) + L_k B_\omega\left(x^{k+1}, x^k\right)\right) \\
&\leq \nabla f_i\left(x^k\right)^\top\left(x^{k+1} - x\right) + g_i\left(x^{k+1}\right) - g_i(x) - \mu B_\omega\left(x, x^k\right) + L_k B_\omega\left(x, x^k\right) \\
&\leq \left(\nabla f_i\left(x^k\right) + \eta_i^k\right)^\top\left(x^{k+1} - x\right) - \mu B_\omega\left(x, x^k\right) - \nu B_\omega\left(x, x^{k+1}\right) + L_k B_\omega\left(x^{k+1}, x^k\right),
\end{aligned}$$

where the second inequality follows from the definition of μ and the last one comes from the convexity of g_i. Multiplying the above inequality by β_i^k and summing for all

$i \in \{1, \ldots, m\}$, the conditions (i) and (ii) give

$$
\begin{aligned}
&\sum_{i=1}^{m} \beta_i^k \left(F_i\left(x^{k+1}\right) - F_i\left(x\right)\right) \\
&\leq L_k \left(\nabla\omega\left(x^{k+1}\right) - \nabla\omega\left(x^{k}\right)\right)^\top \left(x^{k+1} - x\right) + L_k B_\omega\left(x^{k+1}, x^k\right) - \mu B_\omega\left(x, x^k\right) - \nu B_\omega\left(x, x^{k+1}\right) \\
&= L_k \left(\nabla\omega\left(x^{k}\right)^\top \left(x^{k+1} - x\right) - \nabla\omega\left(x^{k+1}\right)^\top \left(x^{k+1} - x\right) + B_\omega\left(x^{k+1}, x^k\right)\right) - \mu B_\omega\left(x, x^k\right) \\
&\quad - \nu B_\omega\left(x, x^{k+1}\right) \\
&= L_k \left(\nabla\omega\left(x^{k}\right)^\top x^{k+1} - \nabla\omega\left(x^{k}\right)^\top x + \nabla\omega\left(x^{k+1}\right)^\top \left(x - x^{k+1}\right) + \omega\left(x^{k+1}\right) - \omega\left(x^k\right)\right. \\
&\quad \left. - \nabla\omega\left(x^{k}\right)^\top \left(x^{k+1} - x^k\right)\right) - \mu B_\omega\left(x, x^k\right) - \nu B_\omega\left(x, x^{k+1}\right) \\
&= L_k \left(\omega(x) - \omega\left(x^k\right) - \nabla\omega\left(x^{k}\right)^\top \left(x - x^k\right) - \omega(x) + \omega\left(x^{k+1}\right) + \nabla\omega\left(x^{k+1}\right)^\top \left(x - x^{k+1}\right)\right) \\
&\quad - \mu B_\omega\left(x, x^k\right) - \nu B_\omega\left(x, x^{k+1}\right) \\
&= L_k \left(B_\omega\left(x - x^k\right) - B_\omega\left(x - x^{k+1}\right)\right) - \mu B_\omega\left(x, x^k\right) - \nu B_\omega\left(x, x^{k+1}\right),
\end{aligned}
$$

where the third and last equalities follow from the definition of Bregman distance.

□

Assumption 4.1. *Let X^* be the set of weakly Pareto optimal points for the multiobjective problem, and define the level set of F for $\alpha \in \mathbf{R}^m$ by $\Omega_F(\alpha) := \{x \in S \mid F(x) \leq \alpha\}$. Then, for all $x \in \Omega_F\left(F\left(x^0\right)\right)$ there exists $x^* \in X^*$ such that $F\left(x^*\right) \leq F(x)$ and*

$$
R := \sup_{F^* \in F\left(X^* \cap \Omega_F\left(F\left(x^0\right)\right)\right)} \inf_{x \in F^{-1}(\{F^*\})} B_\omega\left(x, x^0\right) < \infty.
$$

Theorem 4.3. *Assume that F_i is convex for all $i \in \{1, \ldots, m\}$. Under Assumption 4.1, Algorithm 3.1 (or Algorithm 3.2) generates a sequence $\left\{x^k\right\}$ such that*

$$
u_0\left(x^k\right) \leq \frac{\alpha L R}{k} \quad \textit{for all } k \geq 1,
$$

where $\alpha = \frac{\bar{L}}{L}$ in the constant stepsize setting and $\alpha = \max\left\{\eta, \frac{s}{L}\right\}$ if the backtracking rule is employed.

Proof. From Lemma 4.3 and the convexity of f_i and g_i, for all $x \in \mathbf{R}^n$ we have

$$
\sum_{i=1}^{m} \beta_i^k \left(F_i\left(x^{k+1}\right) - F_i(x)\right) \leq L_k \left(B_\omega\left(x - x^k\right) - B_\omega\left(x - x^{k+1}\right)\right).
$$

Adding up the above inequality from $k = 0$ to $k = \hat{k}$, we obtain

$$\sum_{k=0}^{\hat{k}} \sum_{i=1}^{m} \beta_i^k \left(F_i\left(x^{k+1}\right) - F_i(x)\right) \leq L_k \left(B_\omega\left(x, x^0\right) - B_\omega\left(x, x^{\hat{k}+1}\right)\right)$$
$$\leq L_k B_\omega\left(x, x^0\right).$$

The rest of the proof follows similarly to the proof of [20, Theorem 5.2]. □

4.2.3 The strongly convex case

Here, we show that $\{x^k\}$ generated by Algorithms 3.1 and 3.2 converges linearly to a Pareto optimal point in the strongly convex case.

Theorem 4.4. *Let f_i and g_i have convexity parameters $\mu_i \in \mathbf{R}$ and $\nu_i \in \mathbf{R}$, respectively, and write $\mu := \min_{i\in\{1,\dots,m\}} \mu_i$ and $\nu := \min_{i\in\{1,\dots,m\}} \nu_i$. Assume that $\nabla\omega$ is Lipschitz continuous with parameter q and $1 \leq p < \frac{\beta L+\nu}{\alpha L-\mu}$. If $L_k > L$, then there exists a Pareto optimal point $x^* \in \mathbf{R}^n$ such that for each iteration k,*

$$\left\|x^{k+1} - x^*\right\| \leqslant \sqrt{\frac{q\,(\alpha L - \mu)}{\beta L + \nu}} \left\|x^k - x^*\right\|,$$

where

$$\alpha = \begin{cases} \frac{\bar{L}}{L}, & \text{constant,} \\ \max\left\{\eta, \frac{s}{L}\right\}, & \text{backtracking,} \end{cases} \qquad \beta = \begin{cases} \frac{\bar{L}}{L}, & \text{constant} \\ \frac{s}{L}, & \text{backtracking.} \end{cases}$$

Thus, we have

$$\left\|x^k - x^*\right\| \leq \left(\sqrt{\frac{q\,(\alpha L - \mu)}{\beta L + \nu}}\right)^k \left\|x^0 - x^*\right\|.$$

Proof. Since each F_i is strongly convex, the level set of every F_i is bounded. Thus, $\left\{x^k\right\}$ has an accumulation point $x^* \in \mathbf{R}^n$. Note that x^* is Pareto stationary, from Lemma 4.3, we have

$$\sum_{i=1}^{m} \beta_i^k \left(F_i\left(x^{k+1}\right) - F_i(x^*)\right) \leqslant L_k \left(B_\omega\left(x^*, x^k\right) - B_\omega\left(x^*, x^{k+1}\right)\right)$$
$$- \mu B_\omega\left(x^*, x^k\right) - \nu B_\omega\left(x^*, x^{k+1}\right),$$

Since the left-hand side is nonnegative because of (9), we obtain

$$0 \leq L_k \left(B_\omega\left(x^*, x^k\right) - B_\omega\left(x^*, x^{k+1}\right)\right) - \mu B_\omega\left(x^*, x^k\right) - \nu B_\omega\left(x^*, x^{k+1}\right).$$

Namely,

$$(L_k + \nu)\, B_\omega\left(x^*, x^{k+1}\right) \leq (L_k - \mu)\, B_\omega\left(x^*, x^k\right). \tag{15}$$

From the so-called descent lemma [19, Proposition A.24] and by Lipschitz continuity of $\nabla\omega$, we obtain for all x, y,

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

Combined with the definition of $B_\omega(x, y)$, we have

$$\frac{1}{2}\|x-y\|^2 \leqslant B_\omega(x,y) \leqslant \frac{q}{2}\|x-y\|^2.$$

The above inequality, Lemma 2.1(a) and (15) give

$$\frac{L_k+\nu}{2}\left\|x^*-x^{k+1}\right\|^2 \leqslant \frac{q(L_k-\mu)}{2}\left\|x^*-x^k\right\|^2, \tag{16}$$

which is equivalent to

$$\left\|x^{k+1}-x^*\right\| \leqslant \sqrt{\frac{q(L_k-\mu)}{L_k+\nu}}\left\|x^k-x^*\right\|.$$

Namely,

$$\left\|x^{k+1}-x^*\right\| \leqslant \sqrt{\frac{m(\alpha L-\mu)}{\beta L+\nu}}\left\|x^k-x^*\right\|,$$

where $\alpha = \frac{\bar{L}}{L}$ in the constant stepsize setting and $\alpha = \max\left\{\eta, \frac{s}{L}\right\}$ if the backtracking rule is employed. □

5 Conclusion

We proposed a proximal gradient method with Bregman distance for multi-objective optimization problems. We also used two step size strategies: the constant stepsize and the backtracking strategy. We prove that the sequence generated by the algorithms can converge to a Pareto stationary point and further analyze its convergence rate through some merit functions. Finally, we proved the convergence rates for non-convex $(O(\sqrt{1/k}))$, convex $(O(1/k))$, and strongly convex $\left(O\left(r^k\right)\right.$ for some $r \in (0,1))$ problems. In the future, we can consider trying other stepsize strategies to improve the algorithm, carry out numerical experiments to show the validity of out method, and compare it with other methods.

Acknowledgment

This work was supported by the Grant-in-Aid for Scientific Research (C) (17K00032 and 19K11840).

References

[1] Arthur M Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications*, 22(3):618–630, 1968.

[2] Jörg Fliege, Christoph Heermann, and Bernd Weyers. A new adaptive algorithm for convex quadratic multicriteria optimization. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2005.

[3] Jörg Fliege. An efficient interior-point method for convex multicriteria optimization problems. *Mathematics of Operations Research*, 31(4):825–845, 2006.

[4] Jörg Fliege and Benar Fux Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.

[5] Joerg Fliege, LM Grana Drummond, and Benar Fux Svaiter. Newton's method for multiobjective optimization. *SIAM Journal on Optimization*, 20(2):602–626, 2009.

[6] Ellen H Fukuda and LM Graña Drummond. Inexact projected gradient method for vector optimization. *Computational Optimization and Applications*, 54(3):473–493, 2013.

[7] LM Graña Drummond and Alfredo N Iusem. A projected gradient method for vector optimization problems. *Computational Optimization and applications*, 28(1):5–29, 2004.

[8] João Xavier da Cruz Neto, Gustavo de Jesus Pires Da Silva, Orizon Pereira Ferreira, and Jurandir O Lopes. A subgradient method for multiobjective optimization. *Computational Optimization and Applications*, 54(3):461–472, 2013.

[9] Zhe Chen, XueXiang Huang, and XiaoQi Yang. Generalized proximal point algorithms for multiobjective optimization problems. *Applicable Analysis*, 90(6):935–949, 2011.

[10] Hiroki Tanabe, Ellen H Fukuda, and Nobuo Yamashita. Proximal gradient methods for multiobjective optimization and their applications. *Computational Optimization and Applications*, 72(2):339–361, 2019.

[11] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[12] Yair Censor and Arnold Lent. An iterative row-action method for interval convex programming. *Journal of Optimization Theory and Applications*, 34(3):321–353, 1981.

[13] Gong Chen and Marc Teboulle. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

[14] Jérôme Bolte, Shoham Sabach, Marc Teboulle, and Yakov Vaisbourd. First order methods beyond convexity and lipschitz gradient continuity with applications to quadratic inverse problems. *SIAM Journal on Optimization*, 28(3):2131–2151, 2018.

[15] Marc Teboulle. A simplified view of first order methods for optimization. *Mathematical Programming*, 170(1):67–96, 2018.

[16] Heinz H Bauschke, Jérôme Bolte, and Marc Teboulle. A descent lemma beyond lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348, 2017.

[17] Haihao Lu, Robert M Freund, and Yurii Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28(1):333–354, 2018.

[18] Amir Beck. *First-Order Methods in Optimization*. SIAM, 2017.

[19] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[20] Hiroki Tanabe, Ellen H Fukuda, and Nobuo Yamashita. Convergence rates analysis of multiobjective proximal gradient methods. *arXiv preprint arXiv:2010.08217*, 2020.

[21] Hiroki Tanabe, Ellen H Fukuda, and Nobuo Yamashita. New merit functions and error bounds for non-convex multiobjective optimization. *arXiv preprint arXiv:2010.09333*, 2020.