



Proximal gradient methods for multiobjective optimization and their applications

Hiroki Tanabe¹ · Ellen H. Fukuda¹ · Nobuo Yamashita¹

Received: 20 July 2018 / Published online: 31 October 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

We propose new descent methods for unconstrained multiobjective optimization problems, where each objective function can be written as the sum of a continuously differentiable function and a proper convex but not necessarily differentiable one. The methods extend the well-known proximal gradient algorithms for scalar-valued nonlinear optimization, which are shown to be efficient for particular problems. Here, we consider two types of algorithms: with and without line searches. Under mild assumptions, we prove that each accumulation point of the sequence generated by these algorithms, if exists, is Pareto stationary. Moreover, we present their applications in constrained multiobjective optimization and robust multiobjective optimization, which is a problem that considers uncertainties. In particular, for the robust case, we show that the subproblems of the proximal gradient algorithms can be seen as quadratic programming, second-order cone programming, or semidefinite programming problems. Considering these cases, we also carry out some numerical experiments, showing the validity of the proposed methods.

Keywords Multiobjective optimization · Proximal gradient method · Robust optimization · Pareto stationarity

✉ Ellen H. Fukuda
ellen@i.kyoto-u.ac.jp

Hiroki Tanabe
tanabehiroki@amp.i.kyoto-u.ac.jp

Nobuo Yamashita
nobuo@i.kyoto-u.ac.jp

¹ Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

1 Introduction

Multiobjective optimization consists in minimizing (or maximizing) more than one objective function at the same time and under possible constraints. In most of the cases, it is not possible to find a single point that minimizes all objective functions at once, so the concept of *Pareto* optimality becomes essential. A point is called Pareto optimal or efficient, 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. Multiobjective optimization problems have applications in many fields. We refer to [15] for a list of applications.

One of the most well-known methods for multiobjective optimization problems is the *scalarization approach* [16,20,23]. It consists in solving one or several parametrized single-objective optimization problems to find a solution of the original multiobjective problem. However, in the non-convex case, we may not necessarily get all Pareto optimal solutions using this approach.

In recent years, descent methods for multiobjective optimization problems have attracted a lot of attention in the optimization community. For example, a *steepest descent method* for differentiable unconstrained multiobjective optimization problems was proposed in [13]. Afterwards, a *proximal point method* [8], that can be applied to nondifferentiable problems, was considered. However, this method is just a conceptual scheme and does not necessarily generate subproblems that are easy to solve. For nondifferentiable problems, a *subgradient method* was also developed in [10]. However, for some particular problems, these methods may not be efficient.

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

$$\begin{aligned} \min \quad & F(x) \\ \text{s.t.} \quad & x \in \mathbf{R}^n, \end{aligned} \tag{1.1}$$

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

$$F_i(x) := f_i(x) + g_i(x), \quad i = 1, \dots, m, \tag{1.2}$$

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

In order to solve (1.1), we propose a proximal gradient method, that combines proximal point and gradient methods [8,13]. The proximal point method solves subproblems iteratively, and their objective functions are defined as the sum of the original function F with regularization terms. However, these subproblems are nonlinear in general, so the computational cost for solving them is possibly high. On the other hand, a proximal gradient method applies a gradient method for the differentiable part f_i and a proximal point method for the convex part g_i . As it can be seen in the single objective function cases [3,22], this method is shown to be efficient when g_i has some special structure.

We also observe that the problem and the proposed methods have many applications. For example, when g_i is an indicator function of a convex set S , (1.1) is equivalent to the optimization problems with constraints $x \in S$. Also, as it can be seen in Sect. 5.2, we can deal with robust optimization problems. These problems include uncertain parameters and basically consists in optimizing under the worst scenario. Although the literature about robust optimization is vast, the studies about robust multiobjective optimization is relatively new [11, 14, 19].

The outline of this paper is as follows. We present some notations and notions of Pareto optimality and Pareto stationarity in Sect. 2. In Sect. 3, we propose the proximal gradient methods for unconstrained multiobjective optimization. We show the global convergence of the proposed algorithms in Sect. 4. In Sect. 5, we apply the proposed method to constrained problems and to robust optimization. Finally, we report some numerical experiments by solving robust multiobjective optimization problems in Sect. 6.

2 Preliminaries

Let us first present some notations that will be used in this paper. Let \mathbf{R} denote the set of real numbers and let \mathbf{N} be the set of positive integers. The symbol $\| \cdot \|$ stands for the Euclidean norm in \mathbf{R}^n . 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 $i = 1, \dots, m$. Moreover, we call

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

the directional derivative of $h: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$ at x in the direction d . It is easy to see that $h'(x; d) = \nabla h(x)^\top d$ when h is differentiable at x , where $\nabla h(x)$ denotes the gradient of h at x . The following well-known result shows a non-decreasing property when h is convex.

Lemma 2.1 [7, Section 4.3] *Let $h: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$ be a convex function. Then, the function $\tilde{h}: (0, +\infty) \rightarrow \mathbf{R}$ defined by*

$$\tilde{h}(\alpha) := \frac{h(x + \alpha d) - h(x)}{\alpha}$$

is non-decreasing. In particular, it follows that

$$h(x + d) - h(x) \geq \frac{h(x + \alpha d) - h(x)}{\alpha} \text{ for all } \alpha \in (0, 1).$$

Now, we introduce the concept of optimality for the multiobjective optimization problem (1.1). 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 say that $\bar{x} \in \mathbf{R}^n$ is *Pareto stationary* (or *critical*), if and only if,

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

Observe that this definition generalizes the one given in [13], because here we have to deal with possibly nondifferentiable F_i (or, in particular, g_i). Moreover, instead of considering subdifferentials as in [10], we use the directional derivative notion. Still, with this definition, we can show in the next lemma that weakly Pareto optimal points are always Pareto stationary, but the converse is not always true. However, if every component F_i is convex, then Pareto stationarity implies weak Pareto optimality. Furthermore, if every component F_i is strictly convex, then Pareto stationary points are also Pareto optimal.

- Lemma 2.2**
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.*

Proof We show the contrapositions of statements 1, 2 and 3.

1. Suppose that x is not Pareto stationary. Then, there exists $d \in \mathbf{R}^n$ such that

$$F'_i(x; d) < 0, \quad i = 1, \dots, m.$$

By the definition of directional derivative, for a sufficiently small scalar $\alpha > 0$ we obtain

$$F_i(x + \alpha d) - F_i(x) < 0, \quad i = 1, \dots, m,$$

which means that x is not weakly Pareto optimal.

2. Suppose that x is not weakly Pareto optimal. Then, there exists a point $y \in \mathbf{R}^n$ such that $y \neq x$ and

$$F_i(y) < F_i(x), \quad i = 1, \dots, m. \tag{2.1}$$

Since F_i is convex, it follows from Lemma 2.1 that

$$\frac{F_i(x + \alpha(y - x)) - F_i(x)}{\alpha} \leq F_i(y) - F_i(x)$$

for an arbitrary i and $\alpha \in (0, 1)$. From the convexity of F_i and Lemma 2.1, the left-hand side is non-decreasing with respect to α . Therefore, we have

$$F'_i(x; y - x) \leq F_i(y) - F_i(x) < 0,$$

where the second inequality follows from (2.1). Since this inequality holds for all $i = 1, \dots, m$, we have

$$\max_{i=1, \dots, m} F'_i(x; d) < 0 \quad \text{for } d = y - x.$$

Consequently, we conclude that x is not Pareto stationary.

3. Suppose that x is not Pareto optimal. Then, there exists a point $y \in \mathbf{R}^n$ such that $y \neq x$ and $F_i(y) \leq F_i(x)$ for all $i = 1, \dots, m$. Since F_i is strictly convex, for each i we have

$$F_i(x + \alpha(y - x)) < F_i(x) + \alpha(F_i(y) - F_i(x)) \quad \text{for all } \alpha \in (0, 1).$$

Therefore, it follows that

$$\frac{F_i(x + \alpha(y - x)) - F_i(x)}{\alpha} < F_i(y) - F_i(x) \quad \text{for all } \alpha \in (0, 1).$$

In the same way to statement 2, we obtain $F'_i(x; y - x) < 0$ for all $i = 1, \dots, m$, which concludes the proof. □

3 Proximal gradient methods for multiobjective optimization

In this section, we propose two types of proximal gradient methods for unconstrained multiobjective optimization. Both generate some sequence $\{x^k\}$ 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 step size. For the method without line searches, we set $t_k = 1$ in each iteration. Now, define the function $\psi_x: \mathbf{R}^n \rightarrow \mathbf{R}$ by

$$\psi_x(d) := \max_{i=1, \dots, m} \left\{ \nabla f_i(x)^\top d + g_i(x + d) - g_i(x) \right\}, \tag{3.1}$$

where $\nabla f_i(x)$ denotes the gradient of f_i at x . The term inside the maximum represents the approximation of $F_i(x + d) - F_i(x)$. It is clear that ψ_x is convex and $\psi_x(0) = 0$. The following lemma shows an important property of ψ_x .

Lemma 3.1 *For all $d \in \mathbf{R}^n$, the following equality holds:*

$$\psi'_x(0; d) = \max_{i=1, \dots, m} F'_i(x; d).$$

Proof Since $\psi_x(0) = 0$, by the definition of directional derivative, we get

$$\psi'_x(0; d) = \lim_{\alpha \searrow 0} \frac{\psi_x(\alpha d)}{\alpha}.$$

Moreover, the definition of ψ_x in (3.1) shows that

$$\begin{aligned} \lim_{\alpha \searrow 0} \frac{\psi_x(\alpha d)}{\alpha} &= \lim_{\alpha \searrow 0} \max_{i=1, \dots, m} \frac{\nabla f_i(x)^\top(\alpha d) + g_i(x + \alpha d) - g_i(x)}{\alpha} \\ &= \max_{i=1, \dots, m} \lim_{\alpha \searrow 0} \frac{\nabla f_i(x)^\top(\alpha d) + g_i(x + \alpha d) - g_i(x)}{\alpha} \\ &= \max_{i=1, \dots, m} \left\{ \nabla f_i(x)^\top d + g'_i(x; d) \right\} \\ &= \max_{i=1, \dots, m} F'_i(x; d), \end{aligned}$$

where the second equality follows from the continuity of the max function and the third one comes from the definition of directional derivative. Thus, the claim holds. \square

Now, let ℓ be a positive constant. Here, we define $\phi_{\ell,x} : \mathbf{R}^n \rightarrow \mathbf{R}$ as

$$\phi_{\ell,x}(d) := \psi_x(d) + \frac{\ell}{2} \|d\|^2,$$

where the function ψ_x is defined in (3.1). Clearly, $\phi_{\ell,x}$ is strongly convex and $\phi_{\ell,x}(0) = 0$. Using this function, we define the search direction at an iteration k , which we call *proximal gradient direction*, as $d^k = d_\ell(x^k)$, where

$$d_\ell(x) := \operatorname{argmin}_{d \in \mathbf{R}^n} \phi_{\ell,x}(d). \tag{3.2}$$

Remark 3.1 1. When $g_i = 0$, $d_\ell(x)$ given in (3.2) corresponds to the search direction of the multiobjective steepest descent method [13]. On the other hand, when $f_i = 0$, $d_\ell(x)$ given in (3.2) corresponds to the search direction of the multiobjective proximal point method [8].

- 2. Since $\phi_{\ell,x}$ is strongly convex, (3.2) has a unique solution, and so $d_\ell(x)$ is well-defined.
- 3. Since $\phi_{\ell,x}(0) = 0$, we have $\phi_{\ell,x}(d_\ell(x)) \leq 0$.

Let $\beta_\ell(x)$ be the optimal value in (3.2), i.e.,

$$\beta_\ell(x) := \min_{d \in \mathbf{R}^n} \phi_{\ell,x}(d) = \phi_{\ell,x}(d_\ell(x)). \tag{3.3}$$

The following lemma characterizes the stationarity in terms of $d_\ell(\cdot)$ and $\beta_\ell(\cdot)$.

Lemma 3.2 *Let $d_\ell(x)$ and $\beta_\ell(x)$ be defined in (3.2) and (3.3), respectively. Then, the following statements hold.*

- 1. *If x is Pareto stationary, then $d_\ell(x) = 0$ and $\beta_\ell(x) = 0$. Conversely, if $d_\ell(x) = 0$ and $\beta_\ell(x) = 0$, then x is Pareto stationary.*
- 2. *If x is not Pareto stationary, then $d_\ell(x) \neq 0$ and $\beta_\ell(x) < 0$. Conversely, if $d_\ell(x) \neq 0$ and $\beta_\ell(x) < 0$, then x is not Pareto stationary.*
- 3. *The mappings $d_\ell(\cdot)$ and $\beta_\ell(\cdot)$ are continuous.*

Proof 1. Let x be Pareto stationary. Suppose, for the purpose of contradiction, that $d_\ell(x) \neq 0$ or $\beta_\ell(x) < 0$. From statements 2 and 3 in Remark 3.1 it follows that $d_\ell(x) \neq 0$ if and only if $\beta_\ell(x) < 0$. This means that $d_\ell(x) \neq 0$ and $\beta_\ell(x) < 0$. Therefore, we see that

$$\beta_\ell(x) = \psi_x(d_\ell(x)) + \frac{\ell}{2} \|d_\ell(x)\|^2 < 0. \tag{3.4}$$

Since ψ_x is convex and $\psi_x(0) = 0$, we get

$$\begin{aligned} \psi_x(\alpha d_\ell(x)) &= \psi_x(\alpha d_\ell(x) + (1 - \alpha) \cdot 0) \\ &\leq \alpha \psi_x(d_\ell(x)) + (1 - \alpha) \psi_x(0) \\ &= \alpha \psi_x(d_\ell(x)) \\ &< -\frac{\alpha \ell}{2} \|d_\ell(x)\|^2 \quad \text{for all } \alpha \in (0, 1), \end{aligned}$$

where the last inequality follows from (3.4). Thus, for all $\alpha \in (0, 1)$ we have

$$\frac{\psi_x(\alpha d_\ell(x))}{\alpha} < -\frac{\ell}{2} \|d_\ell(x)\|^2.$$

Since $d_\ell(x) \neq 0$ and $\ell > 0$, letting $\alpha \searrow 0$ we obtain

$$\psi'_x(0; d_\ell(x)) \leq -\frac{\ell}{2} \|d_\ell(x)\|^2 < 0.$$

It then follows from Lemma 3.1 that

$$\max_{i=1, \dots, m} F'_i(x; d_\ell(x)) < 0,$$

which contradicts the Pareto stationarity of x .

Let us now prove the converse. Then, suppose that $d_\ell(x) = 0$ and $\beta_\ell(x) = 0$. From the definition of $\beta_\ell(x)$ given in (3.3), we have

$$\phi_{\ell,x}(d) = \psi_x(d) + \frac{\ell}{2} \|d\|^2 \geq \beta_\ell(x) = 0 \quad \text{for all } d.$$

Let $\alpha \in (0, 1)$. We get

$$\frac{\psi_x(\alpha d) + \frac{\ell}{2} \|\alpha d\|^2}{\alpha} \geq 0 \quad \text{for all } d.$$

Letting $\alpha \searrow 0$ and using Lemma 3.1, we obtain

$$\max_{i=1, \dots, m} F'_i(x; d) \geq 0,$$

which is our claim.

- 2. This statement is equivalent to statement 1.
- 3. It is easy to see that the function

$$\max_{i=1,\dots,m} \left\{ \nabla f_i(x)^\top d + g_i(x + d) - g_i(x) \right\} + \frac{\ell}{2} \|d\|^2 \tag{3.5}$$

is continuous with respect to x and d . Therefore, the optimal value function $\beta_\ell(\cdot)$ is also continuous from [5, Maximum Theorem]. Moreover, since the optimal set mapping $d_\ell(\cdot)$ is unique, $d_\ell(\cdot)$ is continuous from [18, Corollary 8.1]. \square

3.1 A proximal gradient method with line searches

In this section, we present the proposed method with line searches. To compute the step length $t_k > 0$, we use an Armijo rule. Let $\rho \in (0, 1)$ be a prespecified constant. The condition to accept t_k is given by

$$F_i(x^k + t_k d^k) \leq F_i(x^k) + t_k \rho \psi_{x^k}(d^k), \quad i = 1, \dots, m. \tag{3.6}$$

We begin with $t_k = 1$ and while (3.6) is not satisfied, we update

$$t_k := \xi t_k,$$

where $\xi \in (0, 1)$. The following lemma demonstrates the finiteness of this procedure.

Lemma 3.3 *Let d^k be defined in (3.2) with $x = x^k$ and assume that x^k is not Pareto stationary. If $\rho \in (0, 1)$, then there exists some $\bar{t}_k > 0$ such that*

$$F_i(x^k + t d^k) \leq F_i(x^k) + t \rho \psi_{x^k}(d^k), \quad i = 1, \dots, m$$

for any $t \in (0, \bar{t}_k]$.

Proof Let $t \in (0, 1]$. Since g_i is convex for all $i = 1, \dots, m$, we have

$$\begin{aligned} g_i(x^k + t d^k) - g_i(x^k) &= g_i((1 - t)x^k + t(x^k + d^k)) - g_i(x^k) \\ &\leq (1 - t)g_i(x^k) + t g_i(x^k + d^k) - g_i(x^k) \\ &= t(g_i(x^k + d^k) - g_i(x^k)). \end{aligned}$$

Therefore, from the differentiability of f we obtain

$$\begin{aligned} &f_i(x^k + t d^k) + g_i(x^k + t d^k) \\ &\leq f_i(x^k) + g_i(x^k) + t \nabla f_i(x^k)^\top d^k + t \left(g_i(x^k + d^k) - g_i(x^k) \right) + o(t) \\ &\leq f_i(x^k) + g_i(x^k) + t \psi_{x^k}(d^k) + o(t), \end{aligned}$$

where the last inequality comes from the definition (3.1) of ψ_{x^k} . Since x^k is not Pareto stationary, we have $\psi_{x^k}(d^k) < 0$ from Lemma 3.2. Thus, if $\rho \in (0, 1)$, then there exists some $\bar{t}_k > 0$ such that

$$f_i(x^k + td^k) + g_i(x^k + td^k) \leq f_i(x^k) + g_i(x^k) + t\rho\psi_{x^k}(d^k), \quad i = 1, \dots, m$$

for any $t \in (0, \bar{t}_k]$. □

Finally, based on the previous discussions, we state below the proposed method, considering line searches.

Algorithm 3.1

Step 1 Choose $\ell > 0$, $\rho \in (0, 1)$, $\xi \in (0, 1)$, $x^0 \in \mathbf{R}^n$ and set $k := 0$.

Step 2 Compute d^k by solving subproblem (3.2) with $x = x^k$.

Step 3 If $d^k = 0$, then stop.

Step 4 Compute the step length $t_k \in (0, 1]$ as the maximum of

$$T_k := \{t = \xi^j \mid j \in \mathbf{N}, F_i(x^k + t_k d^k) \leq F_i(x^k) + t_k \rho \psi_{x^k}(d^k), i = 1, \dots, m\}$$

Step 5 Set $x^{k+1} := x^k + t_k d^k$, $k := k + 1$, and go to Step 2.

Observe that from Lemma 3.2, the algorithm stops at Step 3 with a Pareto stationary point or produces an infinite sequence of nonstationary points $\{x^k\}$. If Step 4 is reached in some iteration k , it means that in Step 3, $d^k \neq 0$, or equivalently, $\beta_\ell(x^k) < 0$. Thus, we have $\psi_{x^k}(d^k) < 0$. From the Armijo condition, we conclude that all objective functions decrease, i.e.,

$$F_i(x^k + t_k d^k) \leq F_i(x^k), \quad i = 1, \dots, m.$$

3.2 A proximal gradient method without line searches

In this section, we assume that ∇f_i is Lipschitz continuous with constant L . When we set $\ell > L/2$ in (3.2), we can fix the step length $t_k = 1$ for every iteration. We then state below the proposed method.

Algorithm 3.2

Step 1 Choose $\ell > L/2$, $x^0 \in \mathbf{R}^n$ and set $k := 0$

Step 2 Compute d^k by solving subproblem (3.2) with $x = x^k$.

Step 3 If $d^k = 0$, then stop.

Step 4 Set $x^{k+1} := x^k + d^k$, $k := k + 1$, and go to Step 2.

Similarly to Algorithm 3.1, it stops with a Pareto stationary point or generates an infinite sequence of nonstationary points. Moreover, as we can see in Lemma 4.3, the objective function values also decrease in each iteration.

4 Convergence analysis

In this section, we prove that the sequences generated by Algorithm 3.1 and Algorithm 3.2 converge to Pareto stationary points, respectively. From now on, we assume that an infinite sequence is generated. To begin with, we recall the so-called three points property, which is a key to show convergence of proximal point type algorithms.

Theorem 4.1 (Three points property) [9, Lemma 3.2] *Let $\theta: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$ be proper convex and define*

$$x^* = \operatorname{argmin}_{x \in \mathbf{R}^n} \left\{ \theta(x) + \frac{1}{2} \|x - y\|^2 \right\}.$$

Then, for all $z \in \mathbf{R}^n$, we have

$$\theta(x^*) - \theta(z) \leq -\frac{1}{2} \|z - x^*\|^2 - \frac{1}{2} \|y - x^*\|^2 + \frac{1}{2} \|z - y\|^2.$$

In order to show the convergence of the search direction, we first prove the following lemma.

Lemma 4.1 *Let $\{d^k\}$ be generated by Algorithms 3.1 or 3.2 and recall the definition of ψ_x in (3.1). Then, we have*

$$\psi_{x^k}(d^k) \leq -\ell \|d^k\|^2 \text{ for all } k.$$

Proof Defining $\theta := \psi_{x^k} / \ell$ we can rewrite (3.2) with $x = x^k$ as

$$d^k = \operatorname{argmin}_{d \in \mathbf{R}^n} \left\{ \theta(d) + \frac{1}{2} \|d - 0\|^2 \right\}.$$

Thus, substituting $x^* = d^k$ and $y = z = 0$ into Theorem 4.1, we get $\theta(d^k) - \theta(0) \leq -\|d^k\|^2$. Therefore, recalling that $\psi_{x^k}(0) = 0$, we have

$$\psi_{x^k}(d^k) \leq -\ell \|d^k\|^2 \text{ for all } k. \quad \square$$

4.1 Convergence of Algorithm 3.1

Let us recall the algorithm with line searches. We first show the convergence of the length of steps $\|x^{k+1} - x^k\|$.

Lemma 4.2 *Let $\{d^k\}$ be generated by Algorithm 3.1 and suppose that $\{F_i(x^k)\}$ is bounded from below for all $i = 1, \dots, m$. Then, it follows that*

$$\sum_{k=0}^{\infty} t_k \|d^k\|^2 < \infty.$$

Proof From Lemma 4.1 and from (3.6) it follows that

$$F_i(x^k + t_k d^k) \leq F_i(x^k) - t_k \rho \ell \|d^k\|^2, \quad i = 1, \dots, m.$$

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

$$F_i(x^{\tilde{k}+1}) \leq F_i(x^0) - \rho \ell \sum_{k=0}^{\tilde{k}} t_k \|d^k\|^2.$$

Thus, we have

$$\sum_{k=0}^{\tilde{k}} t_k \|d^k\|^2 \leq \frac{1}{\rho \ell} (F_i(x^0) - \tilde{F}_i).$$

Taking $\tilde{k} \rightarrow \infty$, we have $\sum_{k=0}^{\infty} t_k \|d^k\|^2 < \infty$. □

The next theorem is our main result. Basically, if the sequence produced by Algorithm 3.1 has accumulation points, then they are all Pareto stationary.

Theorem 4.2 *Every accumulation point of the sequence $\{x^k\}$ generated by Algorithm 3.1, if it exists, is a Pareto stationary point. In particular, if the level set of each F_i is bounded, then $\{x^k\}$ has accumulation points and they are all Pareto stationary.*

Proof The second statement follows immediately from the first. Let \bar{x} be an accumulation point of $\{x^k\}$ and let $\{x^{k_j}\}$ be a subsequence converging to \bar{x} . From statement 3 of Lemma 3.2, we have $d^{k_j} = d_\ell(x^{k_j}) \rightarrow d_\ell(\bar{x})$. Here, it is sufficient to show that $d_\ell(\bar{x}) = 0$ because of statements 1 and 3 of Lemma 3.2. Suppose for contradiction that $d_\ell(\bar{x}) \neq 0$. Then, it follows from Lemma 4.2 that $t_{k_j} \rightarrow 0$ since the existence of an accumulation point of $\{x^k\}$ implies boundedness of $\{F_i(x^k)\}$ for all i . Therefore, by the definition of t_{k_j} in Step 4 of Algorithm 3.1, for sufficiently large j there exists some $i_{k_j} \in \{1, \dots, m\}$ such that

$$F_{i_{k_j}}(x^{k_j} + \xi^{-1} t_{k_j} d^{k_j}) \geq F_{i_{k_j}}(x^{k_j}) + \xi^{-1} t_{k_j} \rho \psi_{x^{k_j}}(d^{k_j}).$$

Since i only takes finite number of values $\{1, \dots, m\}$, we can assume that $i_{k_j} = \bar{i}$ without loss of generality. We thus obtain

$$\frac{F_{\bar{i}}(x^{k_j} + \xi^{-1} t_{k_j} d^{k_j}) - F_{\bar{i}}(x^{k_j})}{\xi^{-1} t_{k_j}} \geq \rho \psi_{x^{k_j}}(d^{k_j}). \tag{4.1}$$

Recall that $0 < \xi^{-1}t_{k_j} < 1$. It follows from the definition (3.1) of $\psi_{x^{k_j}}$ that

$$\begin{aligned} &\psi_{x^{k_j}}(d^{k_j}) \\ &\geq \{\nabla f_i(x^{k_j})^\top d^{k_j} + g_i^-(x^{k_j} + d^{k_j}) - g_i^-(x^{k_j})\} \\ &\geq \frac{\xi^{-1}t_{k_j} \nabla f_i(x^{k_j})^\top d^{k_j} + g_i^-(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - g_i^-(x^{k_j})}{\xi^{-1}t_{k_j}} \\ &= \frac{f_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) + g_i^-(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - f_i(x^{k_j}) - g_i^-(x^{k_j}) + o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}} \\ &= \frac{F_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - F_i(x^{k_j})}{\xi^{-1}t_{k_j}} + \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}}, \end{aligned}$$

where the second inequality comes from the convexity of g_i and Lemma 2.1, and the first equality follows from the differentiability of f . Therefore, we get

$$\psi_{x^{k_j}}(d^{k_j}) \geq \frac{F_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - F_i(x^{k_j})}{\xi^{-1}t_{k_j}} + \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}}. \tag{4.2}$$

From (4.1) and (4.2), we have

$$\begin{aligned} &\frac{F_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - F_i(x^{k_j})}{\xi^{-1}t_{k_j}} \\ &\geq \rho \frac{F_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - F_i(x^{k_j})}{\xi^{-1}t_{k_j}} + \rho \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}}. \end{aligned}$$

We thus get

$$\frac{F_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - F_i(x^{k_j})}{\xi^{-1}t_{k_j}} \geq \left(\frac{\rho}{1-\rho}\right) \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}}. \tag{4.3}$$

On the other hand, Lemma 4.1 yields

$$\psi_{x^{k_j}}(d^{k_j}) \leq -\ell\|d^{k_j}\|^2.$$

Since $d^{k_j} \rightarrow d_\ell(\bar{x}) \neq 0$, there exists some $\delta > 0$ such that

$$\begin{aligned} -\delta &\geq \psi_{x^{k_j}}(d^{k_j}) \\ &\geq \frac{F_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - F_i(x^{k_j})}{\xi^{-1}t_{k_j}} + \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}}, \end{aligned}$$

where the last inequality comes from (4.2). Therefore, we obtain

$$\frac{F_i(x^{k_j} + \xi^{-1}t_{k_j}d^{k_j}) - F_i(x^{k_j})}{\xi^{-1}t_{k_j}} \leq -\delta - \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}}. \tag{4.4}$$

From (4.3) and (4.4), it follows that

$$\left(\frac{\rho}{1-\rho}\right) \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}} \leq -\delta - \frac{o(\xi^{-1}t_{k_j}\|d^{k_j}\|)}{\xi^{-1}t_{k_j}}.$$

Taking $j \rightarrow \infty$, we have $0 \leq -\delta$, which contradicts the fact that $\delta > 0$. Therefore, we conclude that $d_\ell(\bar{x}) = 0$. □

4.2 Convergence of Algorithm 3.2

Let us now show the convergence of the search direction for the algorithm without line searches. Recall that we have to assume Lipschitz continuity of ∇f_i , with a constant $L > 0$.

Lemma 4.3 *Let $\{d^k\}$ be generated by Algorithm 3.2 and suppose that $\{F_i(x^k)\}$ is bounded from below for all $i = 1, \dots, m$. Then, we have*

$$\lim_{k \rightarrow \infty} \|d^k\| = 0.$$

Proof From the so-called descent Lemma [6, Proposition A.24] and by Lipschitz continuity of ∇f_i , we obtain

$$f_i(x^k + d^k) \leq f_i(x^k) + \nabla f_i(x^k)^\top d^k + \frac{L}{2} \|d^k\|^2. \tag{4.5}$$

At the k th iteration, we have

$$\begin{aligned} & f_i(x^k + d^k) + g_i(x^k + d^k) \\ &= f_i(x^k) + g_i(x^k) + f_i(x^k + d^k) - f_i(x^k) + g_i(x^k + d^k) - g_i(x^k) \\ &\leq f_i(x^k) + g_i(x^k) + \nabla f_i(x^k)^\top d^k + g_i(x^k + d^k) - g_i(x^k) + \frac{L}{2} \|d^k\|^2 \\ &\leq f_i(x^k) + g_i(x^k) + \psi_{x^k}(d^k) + \frac{L}{2} \|d^k\|^2 \\ &\leq f_i(x^k) + g_i(x^k) + \frac{L-2\ell}{2} \|d^k\|^2. \end{aligned}$$

Here, the first inequality follows from (4.5), the second one follows from the definition (3.1) of ψ_{x^k} , and the third one comes from Lemma 4.1. Since $\{F_i(x^k)\}$ is bounded from below, there exists $\tilde{F}_i \leq F_i(x^k) = f_i(x^k) + g_i(x^k)$ for all i, k . Adding up the above inequality from $k = 0$ to $k = \tilde{k}$, we obtain

$$f_i(x^{\tilde{k}+1}) + g_i(x^{\tilde{k}+1}) \leq f_i(x^0) + g_i(x^0) + \frac{L - 2\ell}{2} \sum_{k=0}^{\tilde{k}} \|d^k\|^2.$$

Since $\ell > L/2$, we have

$$\sum_{k=0}^{\tilde{k}} \|d^k\|^2 \leq \frac{2}{2\ell - L} (f_i(x^0) + g_i(x^0) - \tilde{F}_i).$$

Taking $\tilde{k} \rightarrow \infty$, we obtain

$$\sum_{k=0}^{\infty} \|d^k\|^2 < \infty$$

and hence $\lim_{k \rightarrow \infty} \|d^k\| = 0$. □

The next theorem directly follows from Lemmas 4.3 and 3.2.

Theorem 4.3 *Every accumulation point of the sequence $\{x^k\}$ generated by Algorithm 3.2, if it exists, is a Pareto stationary point. In particular, if the level set of each F_i is bounded, then $\{x^k\}$ has accumulation points and they are all Pareto stationary.*

5 Applications

In this section, we consider two applications of multiobjective optimization problem (1.1) with (1.2), and discuss how to solve subproblems (3.2) in these particular applications.

5.1 Application to constrained multiobjective optimization

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

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in S, \end{aligned} \tag{5.1}$$

where $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ is a vector-valued function with $f := (f_1, \dots, f_m)^\top$ and $S \subset \mathbf{R}^n$ is convex. Suppose that each component f_i of f is continuously differentiable. Let $g: \mathbf{R}^n \rightarrow \mathbf{R}^m$ be a vector-valued function with $g := (g_1, \dots, g_m)^\top$, where each g_i is the indicator function of S , i.e.,

$$g_i(x) = \begin{cases} 0 & x \in S, \\ \infty & x \notin S. \end{cases}$$

Then, we can rewrite the search direction given in (3.2) with $x = x^k$ as

$$d^k := \operatorname{argmin}_{d \in S-x^k} \left\{ \max_{i=1, \dots, m} \{ \nabla f_i(x^k)^\top d \} + \frac{\ell}{2} \|d\|^2 \right\},$$

which coincides with the projected gradient direction for multiobjective optimization [17].

5.2 Application to robust multiobjective optimization

Now, let us apply the proposed algorithms to robust multiobjective optimization. Here, we suppose that the problems include uncertain parameters. Moreover, suppose that we can estimate the set of these uncertain parameters. Then, we try to optimize by considering the worst scenario. We observe that studies about robust multiobjective optimization is relatively new [11, 14, 19].

Here, we consider the convex function g_i defined as follows:

$$g_i(x) := \max_{u \in \mathcal{U}_i} \hat{g}_i(x, u). \tag{5.2}$$

We call $\mathcal{U}_i \subseteq \mathbf{R}^n$ an *uncertainty set*. From now on, we assume $\mathcal{U}_i \subset \mathbf{R}^n$ and $\hat{g}_i: \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}$ to be convex with respect to x . It is easy to see that g_i is also convex. However, g_i is not necessarily differentiable even if \hat{g}_i is differentiable. First, let us reformulate the subproblem (3.2) by using an extra variable $\gamma \in \mathbf{R}$ as

$$\begin{aligned} \min_{\gamma, d} \quad & \gamma + \frac{\ell}{2} \|d\|^2 \\ \text{s.t.} \quad & \nabla f_i(x)^\top d + g_i(x+d) - g_i(x) \leq \gamma, \quad i = 1, \dots, m. \end{aligned}$$

Note that g_i is not easy to calculate, and thus, the subproblem is difficult to solve. When \hat{g}_i and \mathcal{U}_i have some special structure, the constraints can be written as explicit formulae by using the duality of (5.2). Now, assume that the dual problem of the maximization problem (5.2) is written as follows:

$$\begin{aligned} \min_{w_i} \quad & \tilde{g}_i(x, w_i) \\ \text{s.t.} \quad & w_i \in \tilde{\mathcal{U}}_i(x), \end{aligned}$$

where $\tilde{g}_i: \mathbf{R}^n \times \mathbf{R}^m \rightarrow \mathbf{R}$ and $\tilde{\mathcal{U}}_i: \mathbf{R}^n \rightarrow 2^{\mathbf{R}^m}$. If strong duality holds, then we see that the subproblem (3.2) is equivalent to

$$\begin{aligned} \min_{\gamma, d, w_i} \quad & \gamma + \frac{\ell}{2} \|d\|^2 \\ \text{s.t.} \quad & \nabla f_i(x)^\top d + \tilde{g}_i(x+d, w_i) - g_i(x) \leq \gamma, \\ & w_i \in \tilde{\mathcal{U}}_i(x+d), \quad i = 1, \dots, m. \end{aligned} \tag{5.3}$$

When \tilde{g}_i and \tilde{U}_i have some explicit form, this problem is tractable. As we mention below, in this case, we can convert the above subproblem to some well-known convex optimization problems. This idea can be also seen in [4]. In the following, we will introduce some robust multiobjective optimization problems where the subproblems can be written as quadratic programming, second-order cone programming or semidefinite programming problems.

(a) Linearly constrained quadratic programming problem case

Suppose that $\hat{g}_i(x, u) = u^\top x$ and $\mathcal{U}_i = \{u \in \mathbf{R}^n \mid A_i u \leq b_i\}$, where $A_i \in \mathbf{R}^{d \times n}$ and $b_i \in \mathbf{R}^d$, that is, \hat{g}_i is linear in x , and \mathcal{U}_i is a polyhedron. Suppose also that \mathcal{U}_i is nonempty and bounded. Then, we can rewrite (5.2) as the following linear programming problem:

$$\begin{aligned} \max_u \quad & x^\top u \\ \text{s.t.} \quad & A_i u \leq b_i. \end{aligned} \tag{5.4}$$

Its dual problem is given by

$$\begin{aligned} \min_w \quad & b_i^\top w \\ \text{s.t.} \quad & A_i^\top w = x, \\ & w \geq 0. \end{aligned}$$

Since the strong duality holds, we can convert the subproblem (3.2) [or, equivalently (5.3)] to a linearly constrained quadratic programming problem:

$$\begin{aligned} \min_{\gamma, d, w_i} \quad & \gamma + \frac{\ell}{2} \|d\|^2 \\ \text{s.t.} \quad & \nabla f_i(x)^\top d + b_i^\top w_i - g_i(x) \leq \gamma, \\ & A_i^\top w_i = x + d, \\ & w_i \geq 0, \quad i = 1, \dots, m. \end{aligned} \tag{5.5}$$

(b) Second-order cone programming problem case

Suppose that $\hat{g}_i(x, u) = u^\top x$ and $\mathcal{U}_i = \{a_i + P_i v \in \mathbf{R}^n \mid \|v\| \leq 1, v \in \mathbf{R}^n\}$, where $a_i \in \mathbf{R}^n$ and $P_i \in \mathbf{R}^{n \times n}$, that is, \hat{g}_i is once again linear in x and \mathcal{U}_i is an ellipsoid. Then, for all $i = 1, \dots, m$ we have

$$\begin{aligned} g_i(x) &= \max_{u \in \mathcal{U}_i} \hat{g}_i(x, u) \\ &= \max_{v: \|v\| \leq 1} (a_i + P_i v)^\top x \\ &= a_i^\top x + \max_{v: \|v\| \leq 1} (P_i^\top x)^\top v. \end{aligned}$$

If $P_i^\top x = 0$, then $\max_{v: \|v\| \leq 1} (P_i^\top x)^\top v = 0 = \|P_i^\top x\|$. If $P_i^\top x \neq 0$, then $\frac{P_i^\top x}{\|P_i^\top x\|}$ is a solution of $\max_{v: \|v\| \leq 1} (P_i^\top x)^\top v$, and hence $\max_{v: \|v\| \leq 1} (P_i^\top x)^\top v = \|P_i^\top x\|$. Consequently, we have

$$g_i(x) = a_i^\top x + \|P_i^\top x\|.$$

Therefore, introducing slack variables $\gamma \in \mathbf{R}$ and $\tau \in \mathbf{R}$, the subproblem (3.2) can be written as

$$\begin{aligned} & \min_{\tau, \gamma, d} \tau \\ & \text{s.t. } \nabla f_i(x)^\top d + \|P_i^\top(x + d)\| - \|P_i^\top x\| + a_i^\top d \leq \gamma, \quad i = 1, \dots, m, \\ & \quad \gamma + \frac{\ell}{2} \|d\|^2 \leq \tau. \end{aligned}$$

Note that convex quadratic constraints can be converted to second-order cone constraints. Using the expression given in [1, Section 2.1], we get the following second-order cone programming problem (SOCP):

$$\begin{aligned} & \min_{\tau, \gamma, d} \tau \\ & \text{s.t. } \begin{bmatrix} -(\nabla f_i(x) + a_i)^\top d + \gamma + \|P_i^\top x\| \\ P_i^\top(x + d) \end{bmatrix} \in \mathcal{K}_{n+1}, \\ & \quad \begin{bmatrix} 1 - \gamma + \tau \\ 1 + \gamma - \tau \\ \sqrt{2\ell}d \end{bmatrix} \in \mathcal{K}_{n+2}, \end{aligned} \tag{5.6}$$

where $\mathcal{K}_q := \{(y_0, \bar{y}) \in \mathbf{R} \times \mathbf{R}^{q-1} \mid y_0 \geq \|\bar{y}\|\}$ is the second-order cone in \mathbf{R}^q . The above SOCP can be solved efficiently with an interior point method [1].

(c) Semidefinite programming problem case

Suppose that¹ $\hat{g}_i(x, u) = (x + u)^\top A_i(x + u)$ and $\mathcal{U}_i = \{a_i + P_i v \in \mathbf{R}^n \mid \|v\| \leq 1\}$, where $A_i \in \mathbf{R}^{n \times n}$ and $A_i \succeq O$, $a_i \in \mathbf{R}^n$ and $P_i \in \mathbf{R}^{n \times n}$. Then, there exists a matrix $M_i \in \mathbf{R}^{n \times n}$ such that $A_i = M_i M_i^\top$. Note that \hat{g}_i is convex quadratic and \mathcal{U}_i is an ellipsoid. Here, without loss of generality we can assume that A is a symmetric matrix since $(x + u)^\top A_i(x + u) = (x + u)^\top \tilde{A}_i(x + u)$, where $\tilde{A}_i := (A_i + A_i^\top)/2$. Then, $g_i(x)$ can be given as

¹ We denote $A \succeq (>)O$ when A is positive semidefinite (positive definite). Also, $A \succeq (>)B$ if and only if $A - B \succeq (>)O$.

$$g_i(x) = \max_{v: \|v\| \leq 1} (x + a_i + P_i v)^\top A_i (x + a_i + P_i v). \tag{5.7}$$

Since problem (5.7) is a maximization problem of a convex function, it is not a convex optimization problem. Fortunately, it can be seen as a subproblem of a trust region method, so its optimal value $g_i(x)$ can be obtained efficiently. Considering (5.7), we observe that

$$g_i(x + d) = \max_{v: \|v\| \leq 1} (x + d + a_i + P_i v)^\top A_i (x + d + a_i + P_i v). \tag{5.8}$$

From [2, Section 3], the Lagrangian dual of the maximization problem (5.8) is given by

$$\begin{aligned} \min_{\alpha, w} \quad & -w \\ \text{s.t.} \quad & \begin{bmatrix} -P_i^\top A_i P_i & -P_i^\top A_i (x + d + a_i) \\ -(x + d + a_i)^\top A_i^\top P_i & -(x + d + a_i)^\top A_i (x + d + a_i) - w \end{bmatrix} \\ & \succeq \alpha \begin{bmatrix} -I_n & 0 \\ 0 & 1 \end{bmatrix}, \\ & \alpha \geq 0, \end{aligned} \tag{5.9}$$

where I_n stands for the identity matrix of dimension n . Let (α^*, w^*) be an optimal solution of (5.9) and assume that² $\dim(\ker(A_i + \alpha^* I_n)) \neq 1$. Since both (5.8) and (5.9) have strictly feasible solutions and $I_n \succ O$, then the strong duality holds from [2, Theorem 3.5]. Therefore, recalling (5.3), the subproblem (3.2) is equivalent to

$$\begin{aligned} \min_{\gamma, d, w_i, \alpha_i} \quad & \gamma + \frac{\ell}{2} \|d\|^2 \\ \text{s.t.} \quad & \nabla f_i(x)^\top d - w_i - g_i(x) \leq \gamma, \\ & \begin{bmatrix} -P_i^\top A_i P_i + \alpha_i I_n & -P_i^\top A_i (x + d + a_i) \\ -(x + d + a_i)^\top A_i^\top P_i & -(x + d + a_i)^\top A_i (x + d + a_i) - w_i - \alpha_i \end{bmatrix} \\ & \succeq O, \\ & \alpha_i \geq 0, \quad i = 1, \dots, m. \end{aligned}$$

Now, by using slack variables $\tau \in \mathbf{R}$ and $\zeta_i \in \mathbf{R}$ and converting the convex quadratic constraints to second-order cone ones, we get the following semidefinite programming problem:

² Here, \dim denotes dimension of a space and \ker means kernel of a matrix.

$$\begin{aligned}
 & \min_{\tau, \alpha_i, w_i, \gamma, d} \quad \tau \\
 & \text{s.t.} \quad \nabla f_i(x)^\top d - w_i - g_i(x) \leq \gamma, \\
 & \quad \begin{bmatrix} 1 - \gamma + \tau \\ 1 + \gamma - \tau \\ \sqrt{2\ell}d \end{bmatrix} \in \mathcal{K}_{n+2}, \\
 & \quad \begin{bmatrix} -P_i^\top A_i P_i + \alpha_i I_n & -P_i^\top A_i(x + d + a_i) \\ -(x + d + a_i)^\top A_i^\top P_i & \zeta_i \end{bmatrix} \succeq O, \quad (5.10) \\
 & \quad \begin{bmatrix} \frac{1 - \zeta_i - w_i - \alpha_i}{1 + \zeta_i + w_i + \alpha_i} \\ \frac{2}{M_i^\top(x + d + a_i)} \end{bmatrix} \in \mathcal{K}_{n+2}, \\
 & \quad \alpha_i \geq 0, \quad i = 1, \dots, m,
 \end{aligned}$$

where O stands for a zero matrix with appropriate dimension. Note that the second-order cone constraints can be converted further into semidefinite constraints.

6 Numerical experiments

In this section, we present some numerical results using Algorithm 3.2 for the problems in Sect. 5.2. The experiments are carried out on a machine with a 1.8 GHz Intel Core i5 CPU and 8GB memory, and we implement all codes in MATLAB R2017a. We consider the problem (1.1), where $n = 5$, $m = 2$, $f_i(x) = \frac{1}{2}x^\top A_i x + a_i^\top x$, $g_i(x) = \max_{u \in \mathcal{U}_i} \hat{g}_i(x, u)$, $A_i \in \mathbf{R}^{n \times n}$, $a_i \in \mathbf{R}^n$, and $\hat{g}_i: \mathbf{R}^n \rightarrow \mathbf{R}$, $i = 1, \dots, m$. Here, we assume that each A_i is positive semidefinite, so it can be decomposed as $A_i = M_i M_i^\top$, where $M_i \in \mathbf{R}^{n \times n}$. We generate M_i and a_i by choosing every component randomly from the standard normal distribution. To implement Algorithm 3.2, we make the following choices.

- Remark 6.1** – Every component of x^0 is chosen randomly from the standard normal distribution.
- In Experiments 1 and 3, we set the constant $\ell = 5$. In Experiment 2, we set the constant $\ell = 7$.
 - The terminate criteria is replaced by $\|d^k\| < \varepsilon := 10^{-6}$.

Also, we run each one of the following experiments 100 times from different initial points, and with $\delta = 0, 0.05, 0.1$. Naturally, when $\delta = 0$, no uncertainties are considered.

Experiment 1

In the first experiment, we solve the problem of Sect. 5.2(a). We assume that $g_i(x) = \max_{u \in \mathcal{U}_i} u^\top x$, $i = 1, 2$, where $\mathcal{U}_1 = \{u \in \mathbf{R}^5 \mid -\delta \leq u_i \leq \delta, i = 1, \dots, 5\}$ and $\mathcal{U}_2 = \{u \in \mathbf{R}^5 \mid -\delta \leq (Bu)_i \leq \delta, i = 1, \dots, 5\}$. Here, every component of

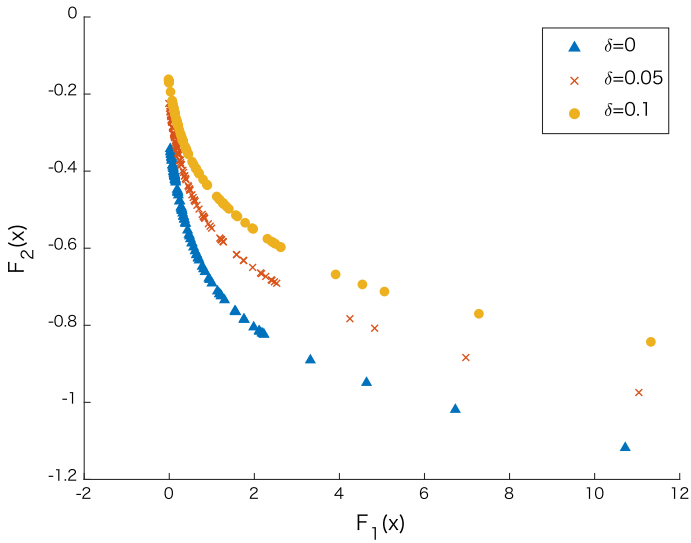


Fig. 1 Result for Experiment 1

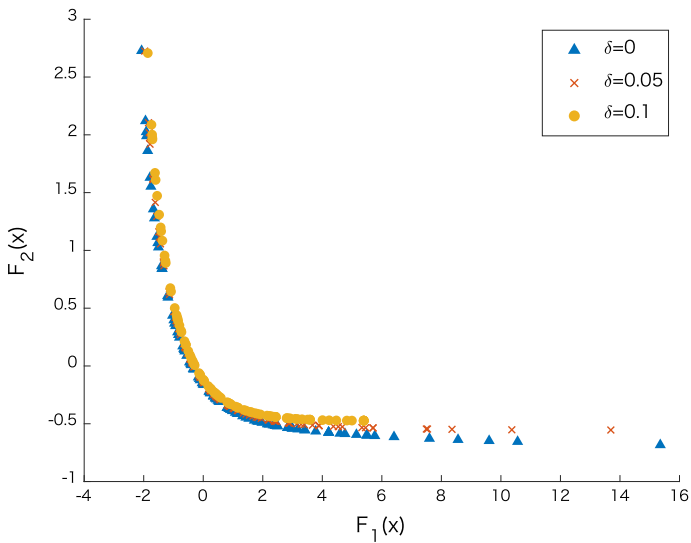


Fig. 2 Result for Experiment 2

$B \in \mathbf{R}^{5 \times 5}$ is chosen randomly from the standard normal distribution and $\delta \geq 0$. We use the MATLAB solver *linprog* to solve (5.4) and *quadprog* to solve (5.5). Figure 1 is the result for this experiment. For each δ , we obtained part of the Pareto frontier, and as δ gets smaller the objective values become smaller.

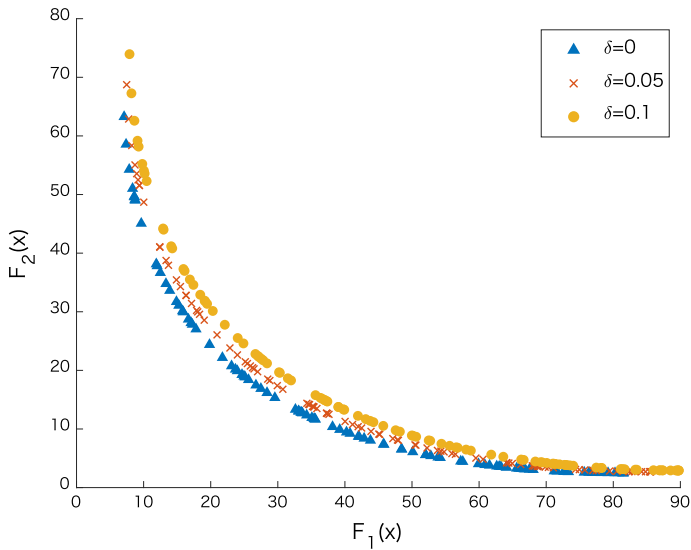


Fig. 3 Result for Experiment 3

Experiment 2

In the second experiment, we solve the problem of Sect. 5.2(b). We assume that $g_i(x) = \max_{u \in \mathcal{U}_i} u^\top x$, where $\mathcal{U}_i = \{u \in \mathbf{R}^5 \mid \|u\| \leq \delta\}$, $i = 1, 2$. We use the MATLAB solver *SeDuMi* [21] to solve (5.6). Figure 2 is the result for this experiment. Once again, we obtained part of the Pareto frontier for the problems with and without uncertainties.

Experiment 3

Now, in the last experiment, we solve the problem of Sect. 5.2(c). We assume that $g_i(x) = \max_{u \in \mathcal{U}_i} (u + x)^\top B B^\top (u + x)$, where $\mathcal{U}_i = \{u \in \mathbf{R}^5 \mid \|u\| \leq \delta\}$, $i = 1, 2$. Here, once again, every component of $B \in \mathbf{R}^{5 \times 5}$ is chosen randomly from the standard normal distribution and $\delta \geq 0$. We use the MATLAB solver *fmincon* to solve (5.7) and *SeDuMi* to solve (5.10). As it can be seen in Fig. 3, we also obtained the Pareto frontier in this case.

7 Conclusion

We proposed proximal gradient methods for unconstrained multiobjective optimization problems with, and without line searches. Under reasonable assumptions, we proved that each accumulation points of the sequences generated by the proposed algorithms are Pareto stationary. Moreover, we presented some applications in constrained optimization and robust multiobjective optimization. In constrained optimization, the

proposed search direction is equivalent to the projected gradient direction, and in some robust optimization problems we can convert the subproblems to well-known optimization problems. Finally, we carried out some numerical experiments for robust multiobjective optimization problems and we observed that the Pareto frontier changes when the uncertainty set is modified.

We have not analyzed the proposed methods in view of the convergence rate. It is known that scalar-valued proximal gradient method is sublinear convergent under reasonable assumptions [3]. In recent years, faster methods such as Newton's method [12] for differentiable multiobjective optimization problem have been also proposed. Therefore, an interesting topic for future research is to investigate the convergence rate of the proposed methods and to propose a proximal Newton-type algorithm for multiobjective optimization.

Acknowledgements This work was supported by the Kyoto University Foundation, and the Grant-in-Aid for Scientific Research (C) (17K00032) from Japan Society for the Promotion of Science. We are also grateful to the anonymous referees for their useful comments.

References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. *Math. Program.* **95**(1), 1–52 (2001)
2. Beck, A., Eldar, Y.: Strong duality in nonconvex quadratic optimization with two quadratic constraints. *SIAM J. Optim.* **17**(3), 844–860 (2006)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
4. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. *Math. Oper. Res.* **23**(4), 769–805 (1998)
5. Berge, C., Patterson, E.M.: *Topological Spaces*. Dover Publications, Edinburgh (1963)
6. Bertsekas, D.P.: *Nonlinear Programming*, 2nd edn. Athena Scientific, Belmont (1999)
7. Bertsekas, D.P.: *Convex Analysis and Optimization*. Athena Scientific, Belmont (2003)
8. Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. *SIAM J. Optim.* **15**(4), 953–970 (2005)
9. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.* **3**(3), 538–543 (1993)
10. Cruz Neto, J.X., Silva, G.J.P., Ferreira, O.P., Lopes, J.O.: A subgradient method for multiobjective optimization. *Comput. Optim. Appl.* **54**(3), 461–472 (2013)
11. Ehrgott, M., Ide, J., Schöbel, A.: Minmax robustness for multi-objective optimization problems. *Eur. J. Oper. Res.* **239**(1), 17–31 (2014)
12. Fliege, J., Graña Drummond, L.M., Svaiter, B.F.: Newton's method for multiobjective optimization. *SIAM J. Optim.* **20**(2), 602–626 (2009)
13. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. *Math. Methods Oper. Res.* **51**(3), 479–494 (2000)
14. Fliege, J., Werner, R.: Robust multiobjective optimization and applications in portfolio optimization. *Eur. J. Oper. Res.* **234**(2), 422–433 (2014)
15. Fukuda, E.H., Graña Drummond, L.M.: A survey on multiobjective descent methods. *Pesquisa Operacional* **34**(3), 585–620 (2014)
16. Geoffrion, A.M.: Proper efficiency and the theory of vector maximization. *J. Math. Anal. Appl.* **22**(3), 618–630 (1968)
17. Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. *Comput. Optim. Appl.* **28**(1), 5–29 (2004)
18. Hogan, W.: Point-to-set maps in mathematical programming. *SIAM Rev.* **15**(3), 591–603 (1973)
19. Morishita, M.: A descent method for robust multiobjective optimization in the presence of implementation errors. Master's thesis, Kyoto University (2016)

20. Saul, G., Thomas, S.: The computational algorithm for the parametric objective function. *Naval Res. Logist. Q.* **2**(12), 39–45 (1955)
21. Sturm, J.F.: Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **11**(1–4), 625–653 (1999)
22. Tseng, P.: Approximation accuracy, gradient methods, and error bound for structured convex optimization. *Math. Program.* **125**(2), 263–295 (2010)
23. Zadeh, L.: Optimality and non-scalar-valued performance criteria. *IEEE Trans. Autom. Control* **8**(1), 59–60 (1963)