Some New Proximal Quasi-Newton Methods for Multiobjective Optimization Problems*

Jian-Wen PENG† Jie REN
(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

(E-mail:jwpeng6@aliyun.com)

Abstract

In this paper, we propose some new proximal quasi-Newton methods with line search or without line search for a special class of nonsmooth multiobjective optimization problems, where each objective function $F_i$ is the sum of a twice continuously differentiable strongly convex function $g_i$ and a proper convex but not necessarily differentiable function $h_i$. In these new proximal quasi-Newton methods, we approximate the Hessian matrices of $g_i$ by using the well known BFGS, self-scaling BFGS, and the Huang BFGS method. We show that each accumulation point of the sequence generated by these new algorithms is a Pareto stationary point of the multiobjective optimization problem. In addition, we give their applications in robust multiobjective optimization, and we show that the subproblems of proximal quasi-Newton algorithms can be regarded as quadratic programming problems. Numerical experiments are carried out to verify the effectiveness of the proposed method.

Key words Multiobjective optimization; Proximal quasi-Newton method; Pareto stationarity; Robust optimization

1 Introduction

Scalarization approach is one of the most effective methods to solve the multi-objective optimization problem, which transforms the multiobjective optimization problem into a single objective mathematical programming problem (see [1-3]). In recent years, the descent methods for multiobjective optimization problems has attracted wide attention in the optimization field [4]. Fliege and Svaiter [5] proposed the steepest descent method for computing a point satisfying first-order nessesary condition for unconstrained multiobjective optimization problems. Bello Cruz, Lucambio Prez and Melo [6] proposed the projection gradient method for quasiconvex multiobjective optimization problems and showed that the sequence generated by the algorithm converges to a stationary point. Iusem and Svaiter [7] proposed a proximal point method for optimization problems of non-differentiable objective functions. Da Cruz Neto, Da Silva, Ferreira, et al.

*Supported by the National Natural Science Foundation of China (11991024).
†Corresponding author.
[8] proposed a subgradient method for quasiconvex multiobjective optimization problems and established the convergence to Pareto optimal points of the sequences produced by the method. Fliege, Grana Drummond and Svaiter [9] introduced the Newton’s method for unconstrained multiobjective optimization problems and showed that the method is locally superlinear convergent to optimal points. Povalej [10] introduced a quasi-Newton method for unconstrained multiobjective optimization problems and showed that the convergence of this method is superlinear.

Tanabe, Fukuda and Yamashita [11] introduced a proximal gradient method with and without line searches for the a special class of nonsmooth multiobjective optimization problem where each objective function is the sum of a continuously differentiable convex function and a proper convex but not necessarily differentiable function. And they showed that each accumulation point of these sequence generated by these algorithms, if exists, is Pareto stationary. They also pointed out in [11] that an interesting topic for future research is to propose a proximal Newton-type algorithm for the above multiobjective optimization problems.

In this paper, the following unconstrained nonsmooth multiobjective optimization problems are studied:

\[
\text{(NMOP)} \quad \min \quad F(x) \\
\text{s.t.} \quad x \in \mathbb{R}^n
\]

(1)

where \( FR^n \rightarrow (\mathbb{R} \cup \{\infty\})^m \) is a vector-valued function \( F = (F_1, ..., F_m)^T \). We assume that each \( F_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\} \) is defined by

\[
F_i(x) = g_i(x) + h_i(x)i = 1, ..., m
\]

(2)

where \( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \) is a twice continuously differentiable strongly convex function \( h_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\} \) is proper convex and continuous but not necessarily differentiable. It is worthy noting that if \( h_i(x) \equiv 0 \) for all \( x \in \mathbb{R}^n \) and \( i = 1, 2, ..., m \), then (1) reduces the multiobjective optimization problems studied in [4–10].

In order to solve the above (NMOP), we propose some new proximal quasi-Newton methods with line search or without line search based on the proximal point method and the quasi-Newton method. Here, the quasi-Newton method is used for the twice continuously differentiable strongly convex function \( g_i \), and the proximal point method is used for the proper convex but not necessarily differentiable function \( h_i \).

The main contents of this paper are as follows: In Section 2, we give some notations and some concepts about Pareto optimality and Pareto stationarity. In Section 3, we propose some new proximal quasi-Newton methods with line search and without line search for the (NMOP). In these new proximal quasi-Newton methods, we approximate the Hessian matrices of \( g_i \) by using the well known BFGS, self-scaling BFGS, and the Huang BFGS method. We prove the global convergence of the proposed algorithms in Section 4. In Section 5, we apply the proposed algorithms to robust multiobjective optimization problems. Finally, in Section 6, we verify the effectiveness of the proposed algorithms through numerical experiments to solve robust multiobjective optimization problems.

2 Preliminaries
For the convenience and brevity of the following discussion, some notations are given in this section, and relevant definitions and lemmas are reviewed.

Let \( R \) denote the set of real numbers and \( N \) denote the set of positive integers. The Euclidean norm in \( R^n \) will be denoted by \( \| \cdot \| \). We define the relationship \( \leq \) (\( < \)) in \( R^n \) as \( u \leq v(u < v) \) if and only if \( u_i \leq v_i(u_i < v_i) \) for all \( i = 1, \ldots, m \).

We call the twice continuously function \( f : R^n \to R \) is strongly convex if for all \( x, y \in R^n \)
\[
(\nabla f(x) - \nabla f(y))^T (x - y) \geq a\|x - y\|^2
\]
for some \( a > 0 \), where \( \nabla f(x) \) denote the gradient of \( f \) at \( x \) (see [12]).

It’s easy to see that (3) is equivalent to
\[
\nabla^2 f(x) \geq aI,
\]
for all \( x \in R^n \), where \( \nabla^2 f(x) \) denote the Hessian matrix of \( f \) at \( x \).

So strong convexity means strict and usual convexity. Hence if \( f_i \) are strongly convex, Hessian matrix \( \nabla^2 f_i(x) \) are positive definite for all \( x \in R^n \) and for all \( i = 1, \ldots, m \).

Let \( f : R^n \to R \cup \{+\infty\} \), and let \( x \) be a point where \( f \) is finite. Then the directional derivative of \( f \) at \( x \) in the direction \( d \in R^n \) is defined to be the limit
\[
f'(x; d) = \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha}
\]
if it exists (see [13]). It’s easy to see that \( f'(x; d) = \nabla f(x)^T d \) when \( f \) is differentiable at \( x \).

**Definition 1** [4, 5] Recall that \( x^* \in R^n \) is a Pareto optimal point for (NMOP), if there is no \( x \in R^n \) 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 \( x^* \in R^n \) is a weakly Pareto optimal point for (NMOP), if there is no \( x \in R^n \) such that \( F(x) < F(x^*) \).

It’s well known that Pareto optimal points are always weakly Pareto optimal, and the converse is not always true.

**Definition 2** [11] We say that \( \bar{x} \in R^n \) is Pareto stationary (or critical) of (NMOP), if and only if
\[
\max_{i=1,\ldots,m} F'_i(\bar{x}; d) \geq 0 \text{ for all } d \in R^n.
\]

It is worthy to noting that Definition 2 generalizes the corresponding ones in [5] and the following important results hold true.

**Lemma 1** [11] (1) If \( x \in R^n \) is a weakly Pareto optimal point of (NMOP)then \( x \) is Pareto stationary.

(2) Let every component \( F_i \) of \( F \) be convex. If \( x \in R^n \) is a Pareto stationary point of (NMOP), then \( x \) is weakly Pareto optimal.

(3) Let every component \( F_i \) of \( F \) be strictly convex. If \( x \in R^n \) is a Pareto stationary point of (NMOP), then \( x \) is Pareto optimal.

The most popular quasi-Newton’s method for nonlinear optimization is BFGS method which was introduced by Broyden, Fletcher, Goldfarb and Shanno [14–17]. It is a line search method with a descent direction
\[
d^k = -(B^k)^{-1} \nabla f(x^k),
\]
where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is the twice continuously differentiable objective function, \( B^k \in \mathbb{R}^{n \times n} \) is an approximation matrix to \( \nabla^2 f(x_k) \), which is updated at every iteration as follows:

\[
B^{k+1} = B^k - \frac{B^k s^k(s^k)^T B^k}{(s^k)^T B^k s^k} + \frac{y^k(y^k)^T}{(s^k)^T y^k},
\]

(4)

where \( s^k = x^{k+1} - x^k \) and \( y^k = \nabla f(x^{k+1}) - \nabla f(x^k) \). As the authors shown in [12], \( B^{k+1} \) remains positive definite whenever \( B^k \) is positive definite. The new iterate is

\[
x^{k+1} = x^k + \lambda_k d_k,
\]

where the step length \( \lambda_k > 0 \).

Next, we will give the updating formulas related to the self-scaling BFGS (in short, SS-BFGS) method [18] of the self-scaling Broyden class.

The updating formula of \( B^k \) for SS-BFGS is as follows:

\[
B^{k+1} = \frac{(s^k)^T y^k}{(s^k)^T B^k s^k} \left( B^k - \frac{B^k s^k(s^k)^T B^k}{(s^k)^T B^k s^k} \right) + \frac{y^k(y^k)^T}{(s^k)^T y^k}.
\]

(5)

Two conditions must be satisfied for the BFGS method and SS-BFGS method to be successful. The first one is a secant equation \( B^{k+1} s^k = y^k \), which we obtain by multiplying (4) by \( s^k \). The second necessary condition is the curvature condition \((s^k)^T y^k > 0\).

Unlike the Broyden and the self-scaling Broyden classes that employ only the gradients and ignore the function value information, the Huang class uses both gradient information and function evaluations for approximating the Hessian matrix. All approximate matrices obtained from the Huang class should satisfy the following equation, namely, the Huang quasi-Newton equation [18, 19] instead of the secant equation:

\[
B^{k+1} s^k = \tilde{y}^k, \quad \tilde{y}^k = y^k + \frac{\theta^k}{(s^k)^T y^k} y^k,\]

where \( \theta^k := 6[f(x^k) - f(x^{k+1})] + 3[\nabla f(x^k) + \nabla f(x^{k+1})]^T s^k \). In this case, the Huang BFGS (in short, H-BFGS) updating formula is as follows:

\[
B^{k+1} = B^k - \frac{B^k s^k(s^k)^T B^k}{(s^k)^T B^k s^k} + \frac{\tilde{y}^k(\tilde{y}^k)^T}{(s^k)^T \tilde{y}^k}.
\]

(6)

Two conditions must be satisfied for the H-BFGS method to be successful. The first one is a Huang quasi-Newton equation \( B^{k+1} s^k = \tilde{y}^k \), which we obtain by multiplying (6) by \( s^k \). The second necessary condition is the curvature condition \((s^k)^T \tilde{y}^k > 0\).

Based on the above statements, if \( f \) is a strongly convex function, then the matrix \( B^{k+1} \) obtained from each of the mentioned updating formulas for approximating the Hessian matrix always preserves positive definiteness. Thus, in this paper, we assume that all \( g_i \) in (NMOP) are strongly convex.

3 Proximal quasi-Newton methods
Throughout the rest of this paper, we always assume that the following assumption holds true.

**Assumption P:** For $i = 1, 2, ..., m$, the function $g_i$ is the twice continuously differentiable strongly convex, and the function $h_i$ is proper convex and continuous but not necessarily differentiable.

In this section, we propose some new proximal quasi-Newton methods for (NMOP) with line search and without line search.

Now we define the function $\theta_x : \mathbb{R}^n \to \mathbb{R}$ by

$$
\theta_x(d) := \max_{i=1,\ldots,m} \{ \nabla g_i(x)^T d + \frac{1}{2} d^T B_i(x) d + h_i(x + d) - h_i(x) \}
$$

where $\nabla g_i(x)$ denotes the gradient of $g_i$ at $x$, $B_i(x)$ is some approximation of $\nabla^2 g_i(x)$, $i = 1, 2, ..., m$. By the convexity of $g_i$ and $h_i$, we get $\theta_x$ is convex and $\theta_x(0) = 0$.

The following lemma shows an important property of $\theta_x$.

**Lemma 2** For all $d \in \mathbb{R}^n$, the following equality holds:

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

**Proof.** Since $\theta_x(0) = 0$, By the definition of directional derivative, we get

$$
\theta'_x(0; d) = \lim_{\alpha \to 0^+} \frac{\theta_x(\alpha d)}{\alpha}
$$

Moreover, the definition of $\theta_x$ in (7) shows that

$$
\lim_{\alpha \to 0^+} \frac{\theta_x(\alpha d)}{\alpha} = \max_{i=1,\ldots,m} \frac{\nabla g_i(x)^T (\alpha d) + \frac{1}{2} (\alpha d)^T B_i(x) (\alpha d) + h_i(x + \alpha d) - h_i(x)}{\alpha}
$$

$$
= \max_{i=1,\ldots,m} \frac{\nabla g_i(x)^T (\alpha d) + \frac{1}{2} (\alpha d)^T B_i(x) (\alpha d) + h_i(x + \alpha d) - h_i(x)}{\alpha}
$$

$$
= \max_{i=1,\ldots,m} \{ \nabla g_i(x)^T d + h'_i(x; d) \}
$$

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

where the second equality follows from the continuity of the max function and the third one comes from the definition of directional derivative.

Let $\omega$ be a positive constant. We define $\varphi_{\omega,x} : \mathbb{R}^n \to \mathbb{R}$ as

$$
\varphi_{\omega,x}(d) := \theta_x(d) + \frac{\omega}{2} \| d \|^2
$$

where the function $\theta_x$ is defined in (7). Clearly, $\varphi_{\omega,x}$ is strongly convex and $\varphi_{\omega,x}(0) = 0$. Using this function, we define the search direction at an iteration $k$, which we call proximal quasi-Newton direction, as $d^k = d_\omega(x^k)$, where

$$
d_\omega(x) := \arg\min_{d \in \mathbb{R}^n} \varphi_{\omega,x}(d).
$$

**Remark 1** (1) Since $\varphi_{\omega,x}$ is strongly convex, (8) has a unique solution $d_\omega(x)$.
Since $\varphi_{\omega,x}(0) = 0$, we have $\varphi_{\omega,x}(d_{\omega}(x)) \leq 0$.

Let $\beta_{\ell}(x)$ be the optimal value in (8), i.e.,

$$
\beta_{\omega}(x) := \min_{d \in \mathbb{R}^n} \varphi_{\omega,x}(d) = \varphi_{\omega,x}(d_{\omega}(x)).
$$

(9)

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

**Lemma 3** Let $d_{\omega}(x)$ and $\beta_{\omega}(x)$ be defined in (8) and (9), respectively. Then, the following statements hold.

(1) If $x$ is Pareto stationary, then $d_{\omega}(x) = 0$ and $\beta_{\omega}(x) = 0$. Conversely, if $d_{\omega}(x) = 0$ and $\beta_{\omega}(x) = 0$, then $x$ is Pareto stationary.

(2) If $x$ is not Pareto stationary, then $d_{\omega}(x) \neq 0$ and $\beta_{\omega}(x) < 0$. Conversely, if $d_{\omega}(x) \neq 0$ and $\beta_{\omega}(x) < 0$, then $x$ is not Pareto stationary.

(3) The mappings $d_{\omega}(\cdot)$ and $\beta_{\omega}(\cdot)$ are continuous.

**Proof.** (1) (Necessity) Let $x$ be Pareto stationary. Suppose, for the purpose of contradiction, that $d_{\omega}(x) \neq 0$ or $\beta_{\omega}(x) < 0$. From statements (1) and (2) in Remark 1 it follows that $d_{\omega}(x) \neq 0$ if and only if $\beta_{\omega}(x) < 0$. This means that $d_{\omega}(x) \neq 0$ and $\beta_{\omega}(x) < 0$.

Therefore, we see that

$$
\beta_{\omega}(x) = \theta_x(d_{\omega}(x)) + \frac{\omega}{2}\|d_{\omega}(x)\|^2 < 0.
$$

(10)

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

$$
\theta_x(\eta d_{\omega}(x)) = \theta_x(\eta d_{\omega}(x) + (1 - \eta) \cdot 0)
\leq \eta \theta_x(d_{\omega}(x)) + (1 - \eta) \theta_x(0)
= \eta \theta_x(d_{\omega}(x))
< - \frac{\eta \omega}{2}\|d_{\omega}(x)\|^2, \forall \eta \in (0, 1)
$$

where the last inequality follows from (10).

Thus, for all $\eta \in (0, 1)$ we have

$$
\frac{\theta_x(\eta d_{\omega}(x))}{\eta} < - \frac{\omega}{2}\|d_{\omega}(x)\|^2.
$$

Since $d_{\omega}(x) \neq 0$ and $\ell > 0$ letting $\eta \to 0^+$ we obtain

$$
\theta_x'(0; d_{\omega}(x)) \leq - \frac{\omega}{2}\|d_{\omega}(x)\|^2 < 0.
$$

It then follows from Lemma 3 that

$$
\max_{i=1,\ldots,m} F_i'(x; d_{\omega}(x)) < 0
$$

which contradicts the Pareto stationarity of $x$.

(Sufficiency) Let us now prove the converse. Then, suppose that $d_{\omega}(x) = 0$ and $\beta_{\omega}(x) = 0$. From the definition of $\beta_{\omega}(x)$ given in (9), we have

$$
\varphi_{\omega,x}(d) = \theta_x(d) + \frac{\omega}{2}\|d\|^2 \geq \beta_{\omega}(x) = 0 \text{ for all } d.
$$
Let $\eta \in (0, 1)$. We get
\[
\frac{\theta_x(\eta d) + \frac{\omega}{2}\|\eta d\|}{\eta} \geq 0 \text{ for all } d.
\]

Letting $\eta \to 0^+$ and using Lemma 3, we obtain
\[
\max_{i=1,\ldots,m} F'_i(x; d) \geq 0,
\]
which is our claim.

(2) This statement is equivalent to statement (1).

(3) It is easy to see that the function
\[
\max_{i=1,\ldots,m} \{\nabla g_i(x)^T d + \frac{1}{2} d^T B_i(x)d + h_i(x + d) - h_i(x)\} + \frac{\omega}{2}\|d\|^2
\]
is continuous with respect to $x$ and $d$. Therefore, the optimal value function $\beta_\omega(\cdot)$ is also continuous from [20, Maximum Theorem]. Moreover, since the optimal set mapping $d_\omega(\cdot)$ is unique, $d_\omega(\cdot)$ is continuous from [21, Corollary 8.1].

3.1 A proximal quasi-Newton method with line searches

Now, we present the proposed proximal quasi-Newton method with line searches for the nonsmooth multiobjective optimization problem (1). To compute the step length $\lambda_k > 0$, we use an Armijo rule. Let $\tau \in (0, 1)$ be a prespecified constant. The condition to accept $\lambda_k$ is given by
\[
F_i(x^k + \lambda_k d^k) \leq F_i(x^k) + \lambda_k \tau \theta_x^i(d^k), i = 1, \ldots, m.
\]

We begin with $\lambda_k = 1$ and while (11) is not satisfied, we update
\[
\lambda_k := \zeta \lambda_k
\]
where $\zeta \in (0, 1)$. The following lemma demonstrates the finiteness of this procedure.

**Lemma 4** Let $d^k$ be defined in (8) with $x = x^k$ and $\tau \in (0, 1)$ If $x^k$ is not Pareto stationary, then there exists some $\bar{\lambda}_k > 0$ such that
\[
F_i(x^k + \lambda d^k) \leq F_i(x^k) + \lambda \tau \theta_x^i(d^k), i = 1, \ldots, m
\]
for any $\lambda \in (0, \bar{\lambda}_k]$.

**Proof.** Let $\lambda \in (0, 1]$. Since $h_i$ is convex for all $i = 1, \ldots, m$, we have
\[
h_i(x^k + \lambda d^k) - h_i(x^k) = h_i((1 - \lambda)x^k + \lambda(x^k + d^k)) - h_i(x^k)
\leq (1 - \lambda)h_i(x^k) + \lambda h_i(x^k + d^k) - h_i(x^k)
= \lambda(h_i(x^k + d^k) - h_i(x^k)).
\]
Therefore, from the second-order Taylor expansion of \( g_i \) we obtain

\[
g_i(x^k + \lambda d^k) + h_i(x^k + \lambda d^k)
\leq g_i(x^k) + \lambda \nabla g_i(x^k)^T d^k + \frac{1}{2}(\lambda d^k)^T B_i(x^k)(\lambda d^k) + h_i(x^k) + \lambda(h_i(x^k + \lambda d^k) - h_i(x^k)) + o(\lambda^2)
\]

\[
= g_i(x^k) + h_i(x^k) + \lambda[\nabla g_i(x^k)^T d^k + \frac{1}{2}(d^k)^T B_i(x^k)(d^k) + h_i(x^k + \lambda d^k) - h_i(x^k)] + o(\lambda^2)
\]

\[
\leq g_i(x^k) + h_i(x^k) + \lambda(\nabla g_i(x^k)^T d^k + \frac{1}{2}(d^k)^T B_i(x^k)(d^k) + h_i(x^k + \lambda d^k) - h_i(x^k)] + o(\lambda^2)
\]

\[
\leq g_i(x^k) + h_i(x^k) + \lambda \theta_{x^k}(d^k) + o(\lambda^2)
\]

where \( B_i(x^k) \) is some approximation of \( \nabla^2 g_i(x^k), i = 1, ..., m \), the second inequality follows from the positive definiteness of \( B_i(x^k) \) and \( \lambda \in (0, 1) \), and the third one comes from the definition (7) of \( \theta_{x^k} \). Since \( x^k \) is not Pareto stationary, we have \( \theta_{x^k}(d^k) < 0 \) from Lemma 4. Thus, \( \tau \in (0, 1) \), then there exists some \( \tilde{\lambda}_k > 0 \) such that

\[
g_i(x^k + \lambda d^k) + h_i(x^k + \lambda d^k) \leq g_i(x^k) + h_i(x^k) + \lambda \tau \theta_{x^k}(d^k), i = 1, ..., m, \forall \lambda \in (0, \tilde{\lambda}_k].
\]

To simplify the notation we will use \( B_i^k \) to denote \( B_i(x^k) \) for all \( i = 1, ..., m \) and \( k = 0, 1, 2, ..., \)

Based on the previous discussions, now we state the proposed new proximal quasi-Newton method with line searches for the multiobjective optimization problem (1) as follows:

**Algorithm 1**

Step 1 Choose \( \omega > 0 \), \( \tau \in (0, 1) \), \( \zeta \in (0, 1) \) \( x^0 \in \mathbb{R}^n \), symmetric positive definite matrix \( B_i^0 \in \mathbb{R}^{n \times n}, i = 1, ..., m \) and set \( k := 0 \).

Step 2 Compute \( d^k \) by solving subproblem (8) with \( x = x^k \).

Step 3 If \( d^k = 0 \), then stop. Otherwise, proceed to the next step.

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

\[
\Lambda_k := \{\lambda = \zeta^j | j \in N, F_i(x^k + \lambda d^k) \leq F_i(x^k) + \lambda \tau \theta_{x^k}(d^k), i = 1, ..., m\}
\]

Step 5 Set \( x^{k+1} = x^k + \lambda_k d^k \), update \( \{B_i^k\} \) by either one of following three formula

\[
B_i^{k+1} = B_i^k - \frac{B_i^k s^k (s^k)^T B_i^k}{(s^k)^T B_i^k s^k} + \frac{y_i^k (y_i^k)^T}{(s^k)^T y_i^k} \tag{12}
\]

\[
B_i^{k+1} = \frac{(s^k)^T y_i^k}{(s^k)^T B_i^k s^k} \left( B_i^k - \frac{B_i^k s^k (s^k)^T B_i^k}{(s^k)^T B_i^k s^k} \right) + \frac{y_i^k (y_i^k)^T}{(s^k)^T y_i^k} \tag{13}
\]

or

\[
B_i^{k+1} = B_i^k - \frac{B_i^k s^k (s^k)^T B_i^k}{(s^k)^T B_i^k s^k} + \frac{y_i^k (y_i^k)^T}{(s^k)^T y_i^k} \tag{14}
\]
where \( s^k = x^{k+1} - x^k = \lambda_k d^k \), \( y^k_i = \nabla g_i(x^{k+1}) - \nabla g_i(x^k) \), \( \hat{y}^k_i = y^k_i + \frac{\theta^k_i}{(s^k)^T y^k_i} s^k \) and \( \theta^k_i := 6[g_i(x^k) - g_i(x^{k+1})] + 3[\nabla g_i(x^k) + \nabla g_i(x^{k+1})]^T s^k \). Set \( k := k + 1 \), and go to Step 2.

It is worthy noting that we call Algorithm 1 to be BFGS method, self-scaling BFGS (in short, SS-BFGS) method and Huang BFGS (in short, H-BFGS) method with line searches for (NMOP), respectively, when one updates \( \{B^k_i\} \) using (12), (13) and (14), respectively.

Observe that from Lemma 3, 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, \( \beta_x(x^k) < 0 \). Thus, we have \( \theta_{x}(d^k) < 0 \). It follows from the Armijo condition that objective values sequence \( \{F(x^k)\} \) is \( R_m^+ \)-decrease, i.e.,

\[
F(x^{k+1}) < F(x^k) \text{ for all } k.
\]

### 3.2 A proximal quasi-Newton method without line searches

In this section, we assume that \( \nabla g_i \) is Lipschitz continuous with constant \( L \) for all \( i = 1, 2, ..., m \). Let the step length \( \lambda_k \equiv 1 \) for \( k = 0, 1, 2, ... \). Now we introduce some new proximal quasi-Newton method without line searches for the multiobjective optimization problem (1).

**Algorithm 2**

Step 1 Choose \( \omega > L/2 \), \( x^0 \in \mathbb{R}^n \), symmetric positive definite matrix \( B^0_i \in \mathbb{R}^{n \times n} \), \( i = 1, ..., m \) and set \( k := 0 \).

Step 2 Compute \( d^k \) by solving subproblem (8) with \( x = x^k \).

Step 3 If \( d^k = 0 \), then stop. Otherwise, proceed to the next step.

Step 4 Set \( x^{k+1} = x^k + d^k \), update \( B^{k+1}_i \) by either one of the formula (12), (13) or (14). Set \( k := k + 1 \), and go to Step 2.

Similarly to Algorithm 1, we call Algorithm 2 to be BFGS method, SS-BFGS method and H-BFGS method without line searches for (NMOP), respectively, when one updates \( \{B^k_i\} \) using (12), (13) and (14), respectively. And it is easy to see that the algorithm 2 stops at Step 3 with a Pareto stationary point or generates an infinite sequence of nonstationary points \( \{x^k\} \). Moreover, as we can see from the proof of Lemma 7, the objective function values also decrease in each iteration, i.e.,

\[
F_i(x^k + d^k) < F_i(x^k), \text{ } i = 1, ..., m.
\]

### 4 Convergence analysis

In this section, we prove that the sequences generated by Algorithm 1 and Algorithm 2 converge to Pareto stationary points, respectively.

**Theorem 1** (Three points property) [22, Lemma 3.2] Let \( \sigma : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \) be proper convex and define

\[
x^* = \arg \min \left\{ \sigma(x) + \frac{1}{2} \|x - y\|^2 \right\}.
\]

Then, for all \( z \in \mathbb{R}^n \), we have

\[
\sigma(x^*) - \sigma(z) \leq -\frac{1}{2} \|z - x^*\|^2 - \frac{1}{2} \|y - x^*\|^2 + \frac{1}{2} \|z - y\|^2.
\]
Lemma 5 Let \{d^k\} be generated by Algorithms 1 or 2 and recall the definition of \(\theta_x\) in (7). Then, we have

\[ \theta_x(d^k) \leq -\omega \|d^k\|^2 \text{ for all } k. \]

**Proof.** Defining \(\sigma := \theta_x(d^k)/\omega\), we can rewrite (8) with \(x = x^k\) as

\[ d^k = \arg \min_{d \in \mathbb{R}^n} \left\{ \sigma(d) + \frac{1}{2} \|d - 0\|^2 \right\}. \]

Thus, substituting \(x^* = d^k\) and \(y = z = 0\) into Theorem 1, we get

\[ \sigma(d^k) - \sigma(0) \leq -\|d^k\|^2. \]

Therefore, recalling that \(\theta_x(0) = 0\), we have

\[ \theta_x(d^k) \leq -\omega \|d^k\|^2 \text{ for all } k. \]

\[ \square \]

4.1 Convergence of Algorithm 1

Lemma 6 Let \{d^k\} be generated by Algorithm 1 and suppose that \(\{F_i(x^k)\}\) is bounded from below for all \(i = 1, ..., m\). Then, it follows that

\[ \lim_{k \to \infty} \lambda_k \|d^k\|^2 = 0. \]

**Proof.** It follows from Lemma 6 and step 4 of Algorithm 1 that

\[ F_i(x^k + \lambda_k d^k) \leq F_i(x^k) - \lambda_k \tau \omega \|d^k\|^2, \quad i = 1, ..., m. \]

Adding up the above inequality from \(k = 0\) to \(k = \hat{k}\), where \(\hat{k}\) is a positive integer, we obtain

\[ F_i(x^{\hat{k}+1}) \leq F_i(x^0) - \tau \omega \sum_{k=0}^{\hat{k}} \lambda_k \|d^k\|^2. \] (15)

Since \(\{F_i(x^k)\}\) is bounded from below for all \(i = 1, ..., m\), there exists \(\hat{F}_i \in \mathbb{R}\) such that \(\hat{F}_i \leq F_i(x^k)\) for all \(i\) and \(k\).

It follows from (15) that

\[ \sum_{k=0}^{\hat{k}} \lambda_k \|d^k\|^2 \leq \frac{1}{\tau \omega} (F_i(x^0) - F_i(x^{\hat{k}+1})) \]

\[ \leq \frac{1}{\tau \omega} (F_i(x^0) - \hat{F}_i). \]

Taking \(\hat{k} \to \infty\) we have \(\sum_{k=0}^{\infty} \lambda_k \|d^k\|^2 < \infty\)
and hence \(\lim_{k \to \infty} \lambda_k \|d^k\| = 0. \)

\[ \square \]
Theorem 2 (i) Suppose that \{F_i(x^k)\} is bounded from below for all \(i = 1, \ldots, m\). Then every accumulation point of the sequence \{x^k\} generated by Algorithm 1, if it exists, is a Pareto stationary point.

(ii) If the level set of \(F\) in the sense that \(\{x \in \mathbb{R}^n \mid F(x) \leq F(x_0)\}\) is bounded, then \{x^k\} has accumulation points and they are all Pareto stationary.

**Proof.** We now prove the first statement.

Let \(\bar{x}\) be an accumulation point of \{x^k\} and let \{x^k\} be a subsequence converging to \(\bar{x}\). From statement (3) of Lemma 4, we have \(d^k = d_\infty(x^k) \to d_\infty(\bar{x})\). Here, it is sufficient to show that \(d_\infty(\bar{x}) = 0\) because of statements (1) and (3) of Lemma 4. Suppose for contradiction that \(d_\infty(\bar{x}) \neq 0\). Then, it follows from Lemma 7 that \(\lambda_k \to 0\). Therefore, by the definition of \(\lambda_k\) in Step 4 of Algorithm 1, for sufficiently large \(j\) there exists some \(i_k \in \{1, \ldots, m\}\) such that

\[
F_{i_k}(x^k) + \zeta^{-1}\lambda_{k_j}d^k) > F_{i_k}(x^k) + \zeta^{-1}\lambda_{k_j}\theta_{x^k_j}(d^k).
\]

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

\[
\frac{F_i(x^k) + \zeta^{-1}\lambda_{k_j}d^k) - F_i(x^k)}{\zeta^{-1}\lambda_{k_j}} > \tau \theta_{x^k_j}(d^k).
\]

(16)

Recall that \(0 < \zeta^{-1}\lambda_{k_j} < 1\). It follows from the definition (7) of \(\theta_{x^k_j}\) that

\[
\theta_{x^k_j}(d^k) = \nabla g_i(x^k)^T d^k + \frac{1}{2}(d^k)^T B_i(x^k)(d^k) + h_i(x^k) + h_i(x^k + d^k) - h_i(x^k)
\]

\[
\geq \zeta^{-1}\lambda_{k_j} \nabla g_i(x^k)^T d^k + \frac{1}{2}\zeta^{-1}\lambda_{k_j} (d^k)^T B_i(x^k)(d^k) + h_i(x^k) + \zeta^{-1}\lambda_{k_j} d^k - h_i(x^k)
\]

\[
= g_i(x^k) + \zeta^{-1}\lambda_{k_j} d^k + h_i(x^k) - h_i(x^k) - h_i(x^k) + o((\zeta^{-1}\lambda_{k_j} d^k)^2)
\]

\[
= F_i(x^k) - F_i(x^k) + \frac{o((\zeta^{-1}\lambda_{k_j} d^k)^2)}{\zeta^{-1}\lambda_{k_j}}
\]

(17)

where \(B_i(x^k)\) is some approximation of \(\nabla^2 g_i(x^k)\), the second inequality comes from the convexity of \(h_i\) and Lemma 1, and the first equality follows from the second-order Taylor expansion of \(g_i\). Therefore, we get

\[
\theta_{x^k_j}(d^k) \geq \frac{F_i(x^k) + \zeta^{-1}\lambda_{k_j} d^k) - F_i(x^k)}{\zeta^{-1}\lambda_{k_j}} + \frac{o((\zeta^{-1}\lambda_{k_j} d^k)^2)}{\zeta^{-1}\lambda_{k_j}}.
\]

(17)

From (16) and (17), we get

\[
\frac{F_i(x^k) + \zeta^{-1}\lambda_{k_j} d^k) - F_i(x^k)}{\zeta^{-1}\lambda_{k_j}} > \tau \frac{F_i(x^k) + \zeta^{-1}\lambda_{k_j} d^k) - F_i(x^k)}{\zeta^{-1}\lambda_{k_j}} + \frac{o((\zeta^{-1}\lambda_{k_j} d^k)^2)}{\zeta^{-1}\lambda_{k_j}}.
\]

It follows that

\[
\frac{F_i(x^k) + \zeta^{-1}\lambda_{k_j} d^k) - F_i(x^k)}{\zeta^{-1}\lambda_{k_j}} > \frac{(1 - \tau) o((\zeta^{-1}\lambda_{k_j} d^k)^2)}{\zeta^{-1}\lambda_{k_j}}.
\]

(18)
On the other hand, Lemma 6 yields

$$\theta_{x_k j}(d^{k_j}) \leq -\omega \|d^{k_j}\|^2.$$  

Since $d^{k_j} \to d_\omega(x) \neq 0$, it follows from the above inequality and (17) that there exists $\gamma = \omega \|d_\omega\|^2 > 0$ such that

$$-\gamma \geq \theta_{x_k j}(d^{k_j})$$

for sufficiently large $j$. Therefore, for sufficiently large $j$, the following inequality holds.

$$F_i(x^{k_i} + \zeta^{-1}\lambda_k d^{k_j}) - F_i(x^{k_i}) \leq -\gamma - \frac{o((\zeta^{-1}\lambda_k \|d^{k_j}\|)^2)}{\zeta^{-1}\lambda_k}.$$  

(19)

From (18) and (19), we know that for sufficiently large $j$

$$\left(\frac{\tau}{1-\tau}\right) \frac{o((\zeta^{-1}\lambda_k \|d^{k_j}\|)^2)}{\zeta^{-1}\lambda_k} < -\gamma - \frac{o((\zeta^{-1}\lambda_k \|d^{k_j}\|)^2)}{\zeta^{-1}\lambda_k}.$$  

Taking $j \to \infty$, we have $0 < -\gamma$, which contradicts the fact that $\gamma > 0$. Therefore, we conclude that $d_\omega(x) = 0$.

We now prove the second statement. It is easy to see that the set $\{x \in \mathbb{R}^n \mid F(x) \leq F(x_0)\}$ is bounded and that objective values sequence $\{F(x^k)\}$ is $R^m_+$-decrease. Therefore, the sequence $\{x_n\}$ generated by Algorithm 1 is contained in the above set and so it is also bounded and has at least one accumulation point, which is a stationary point of (NMOP) according to the first statement.

4.2 Convergence of Algorithm 2

**Lemma 7** Let $\{d^k\}$ be generated by Algorithm 2 and suppose that $\{F_i(x^k)\}$ is bounded from below for all $i = 1, ..., m$. Then, we have

$$\lim_{k \to \infty} \|d^k\|^2 = 0.$$  

**Proof.** From the so-called descent Lemma [23, Proposition A.24] and by Lipschitz continuity of $\nabla g_i$, we obtain

$$g_i(x^k + d^k) \leq g_i(x^k) + \nabla g_i(x^k)^T d^k + \frac{L}{2} \|d^k\|^2.$$  

Moreover, since the positive definiteness of $B_i(x^k)$ implies $(d^k)^T B_i(x^k)(d^k) > 0$ for all $i = 1, ..., m$. Therefore, we get

$$g_i(x^k + d^k) < g_i(x^k) + \nabla g_i(x^k)^T d^k + \frac{1}{2}(d^k)^T B_i(x^k)(d^k) + \frac{L}{2} \|d^k\|^2.$$  

(20)
At the $k$th iteration, we have for $i = 1, 2, \ldots, m$,
\[
g_i(x^k + d^k) + h_i(x^k + d^k) \\
= g_i(x^k) + h_i(x^k) + g_i(x^k + d^k) - g_i(x^k) + h_i(x^k + d^k) - h_i(x^k) \\
< g_i(x^k) + h_i(x^k) + \nabla g_i(x^k)^T d^k + \frac{1}{2}(d^k)^T B_i(x^k)(d^k) + h_i(x^k + d^k) - h_i(x^k) + \frac{L}{2}\|d^k\|^2 \\
\leq g_i(x^k) + h_i(x^k) + \theta x^i(d^k) + \frac{L}{2}\|d^k\|^2 \\
\leq g_i(x^k) + h_i(x^k) + \frac{L - 2\omega}{2}\|d^k\|^2.
\]

Here, the first inequality follows from (20), the second one follows from the definition (7) of $\theta x$, and the third one comes from Lemma 5. Since $\{F_i(x^k)\}$ is bounded from below, there exists $\hat{F}_i \in R$ such that $\hat{F}_i \leq F_i(x^k) = g_i(x^k) + h_i(x^k)$ for all $i, k$. Adding up the above inequality from $k = 0$ to $k = \hat{k}$, where $\hat{k}$ is a positive integer, we obtain
\[
g_i(x^{k+1}) + h_i(x^{k+1}) < g_i(x^0) + h_i(x^0) + \frac{L - 2\omega}{2}\sum_{k=0}^{\hat{k}}\|d^k\|^2.
\]

Since $\omega > L/2$, we have
\[
\sum_{k=0}^{\hat{k}}\|d^k\|^2 < \frac{2}{2\omega - L}(g_i(x^0) + h_i(x^0) - (g_i(x^{\hat{k}+1}) + h_i(x^{\hat{k}+1})) \\
< \frac{2}{2\omega - L}(g_i(x^0) + h_i(x^0) - \hat{F}_i).
\]

Taking $\hat{k} \to \infty$, we obtain
\[
\sum_{k=0}^{\infty}\|d^k\|^2 < \infty
\]
and hence $\lim_{k \to \infty} \|d^k\|^2 = 0$. $\square$

**Theorem 3** (i) Suppose that $\{F_i(x^k)\}$ is bounded from below for all $i = 1, \ldots, m$. Then every accumulation point of the sequence $\{x^k\}$ generated by Algorithm 2 is a Pareto stationary point.

(ii) If the level set of $F$ in the sense that $\{x \in R^n \mid F(x) \leq F(x_0)\}$ is bounded, then $\{x^k\}$ has accumulation points and they are all Pareto stationary.

**Proof.** The second statement follows immediately from the first. Let $\hat{x}$ be an accumulation point of $\{x^k\}$ and let $\{x^k\}$ be a subsequence converging to $\hat{x}$. From statement (3) of Lemma 3, we have $d^k = d^\omega(x^k) \to d^\omega(\hat{x})$. Here, it is sufficient to show that $d^\omega(\hat{x}) = 0$ because of statements (1) and (3) of Lemma 3. Suppose for contradiction that $d^\omega(\hat{x}) \neq 0$, which contradicts the fact that Lemma 7. Therefore, we conclude that $d^\omega(\hat{x}) = 0$. $\square$

5 Application to robust multiobjective optimization
In this section, we consider an application of the nonsmooth multiobjective optimization problem (1) with (2), and discuss how to solve subproblems (8) in a particular application.

Now, let us apply the proposed proximal quasi-Newton algorithms to the robust multiobjective optimization including uncertain parameters which is exactly the multiobjective optimization problem discussed in [11]. In other words, we will solve the (NMOP) with the convex function \( h_i \) defined as follows:

\[
h_i(x) := \max_{u \in U_i} \hat{h}_i(x, u)
\]  

where \( U_i \subseteq \mathbb{R}^n \) is an uncertainty set, and \( \hat{h}_i : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \) is convex with respect to the first argument.

It is easy to see that \( h_i \) is also convex. However, \( h_i \) is not necessarily differentiable even if \( \hat{h}_i \) is differentiable. First, let us reformulate the subproblem (8) by using an extra variable \( \mu \in \mathbb{R} \) as

\[
\begin{align*}
\min_{\mu, d} & \quad \mu + \frac{\omega}{2} \|d\|^2 \\
\text{s.t.} & \quad \nabla g_i(x)^T d + \frac{1}{2} d^T B_i(x) d + h_i(x + d) - h_i(x) \leq \mu, \ i = 1, \ldots, m.
\end{align*}
\]

It is worthy noting that \( h_i \) is not easy to calculate, and thus, the subproblem is difficult to solve. When \( \hat{h}_i \) and \( U_i \) have some special structure, the constraints of the above problem can be written as explicit formulae by using the duality of (21). Now, assume that the dual problem of the maximization problem (21) is written as follows:

\[
\begin{align*}
\min_{w_i} & \quad \tilde{h}_i(x, w_i) \\
\text{s.t.} & \quad w_i \in \tilde{U}_i(x)
\end{align*}
\]

where \( \tilde{h}_i : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \) and \( \tilde{U}_i : \mathbb{R}^n \rightarrow 2^{\mathbb{R}^m} \). If strong duality holds, then we see that the subproblem (8) is equivalent to

\[
\begin{align*}
\min_{\mu, d, w_i} & \quad \mu + \frac{\omega}{2} \|d\|^2 \\
\text{s.t.} & \quad \nabla g_i(x)^T d + \frac{1}{2} d^T B_i(x) d + \tilde{h}_i(x + d, w_i) - h_i(x) \leq \mu, \\
& \quad w_i \in \tilde{U}_i(x + d), \ i = 1, \ldots, m.
\end{align*}
\]

When \( \tilde{h}_i \) 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 [11]. In the following, we will introduce a robust multiobjective optimization problem where the subproblem can be written as a quadratic programming.

Suppose that \( \hat{h}_i(x, u) = u^T x \) and \( U_i = \{u \in \mathbb{R}^n | A_i u \leq b_i \} \), where \( A_i \in \mathbb{R}^{d \times n} \) and \( b_i \in \mathbb{R}^d \), that is, \( \hat{h}_i \) is linear in \( x \), and \( U_i \) is a polyhedron. Suppose also that \( U_i \) is nonempty and bounded. Then, follow the ideas of Tanabe, Fukuda and Yamashita [11], problem (21) can be rewritten as the following linear programming problem:

\[
\begin{align*}
\max_u & \quad x^T u \\
\text{s.t.} & \quad A_i u \leq b_i
\end{align*}
\]
And its dual problem is given by
\[
\min_{w} \quad b_i^T w \\
\text{s.t.} \quad A_i^T w = x, \\
w \geq 0.
\]

Since the strong duality holds, we can convert the subproblem (8) [or, equivalently (22)] to a linearly constrained quadratic programming problem:
\[
\begin{align*}
\min_{\mu, d, w} \quad & \mu + \frac{\omega}{2} \|d\|^2 \\
\text{s.t.} \quad & \nabla g_i(x)^T d + \frac{\omega}{2} d^T B_i(x) d + b_i^T w_i - h_i(x) \leq \mu \\
& A_i^T w_i = x + d \\
& w_i \geq 0, i = 1, \ldots, m.
\end{align*}
\]

(24)

6 Numerical experiments

In this section, we present some numerical results using Algorithms 1 and 2 for the linearly constrained quadratic programming problem in Section 5. The experiments are carried out on a machine with a 2.2GHz Intel Core i3 CPU and 6GB memory, and we implement all codes in MATLAB R2018b. We consider the problem (1), where \( n = 5, m = 2, g_i(x) = \frac{1}{2} x^T Q_i x + q_i^T x h_i(x) = \max_{u \in U_i} u^T x, \) \( Q_i \in \mathbb{R}^{n \times n}, q_i \in \mathbb{R}^n, \) and \( h_i : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \ldots, m. \) Here, we assume that each \( Q_i \) is positive definite, so it can be decomposed as \( Q_i = M_i M_i^T, \) where \( M_i \in \mathbb{R}^{n \times n}. \) We generate \( M_i \) and \( q_i \) by choosing every component randomly from the standard normal distribution. To implement Algorithms 1 and 2, we make the following choices.

Remark 2
(1) Every component of \( x^0 \) is chosen randomly from the standard normal distribution.
(2) \( B_i^0 \) is the identity matrix, i.e., \( B_i^0 = I, i = 1, \ldots, m. \)
(3) We set the constant \( \omega = 5, \tau = \frac{1}{2}, \) and \( \varsigma = \frac{1}{2}. \)
(4) 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.

In order to solve the linear constrained quadratic programming problem in Section 5, the following numerical experiments are performed. We assume that \( h_i(x) = \max_{u \in U_i} u^T x, i = 1, 2, \) where \( U_1 = \{ u \in R^5 \mid -\delta \leq u_i \leq \delta, i = 1, \ldots, 5 \} \) and \( U_2 = \{ u \in R^5 \mid -\delta \leq (Bu)_i \leq \delta, i = 1, \ldots, 5 \}. \) Here, every component of \( B \in R^{5 \times 5} \) is chosen randomly from the standard normal distribution and \( \delta \geq 0. \)

For simplicity of the notation, from now on, the proximal gradient method for (NMOP), which introduced in [11] is denoted by PGM. BFGS, self-scaling BFGS, and Huang BFGS proximal quasi-Newton methods for multiobjective optimization are denoted by PQNM(BFGS), PQNM(SS-BFGS), and PQNM(H-BFGS), respectively. We will use the toolbox of convex optimization in MATLAB to solve (23) and (24).
6.1 Numerical experiment of algorithm 1

On one hand, the experimental result obtained by using PGM with line search (i.e., Algorithm 3.1 in [11]) with deferent $\delta$ for (NMOP) is shown in Fig. 1. The experimental results obtained by using PQNM(BFGS), PQNM(SS-BFGS), and PQNM(H-BFGS) with line search for (NMOP) with deferent $\delta$ is shown in Fig. 2, Fig. 3 and Fig. 4, respectively. For each $\delta$, we obtained part of the Pareto frontier. And it is easy to see that the Pareto frontier of (NMOP) becomes lower when $\delta$ is smaller. On the other hand, Fig. 5 shows the comparison among the (PGM), PQNM(BFGS), PQNM(SS-BFGS) and PQNM(H-BFGS) with line search for (NMOP) when $\delta$ is 0, 0.05 and 0.1, respectively. When $\delta$ is same, we observed that the Pareto frontier of PQNM(H-BFGS) with line search for (NMOP) is the lowest, the Pareto frontier of PGM with line search for (NMOP) is the highest, and the Pareto frontier of PQNM(SS-BFGS) with line search for (NMOP) is lower than that of the PQNM(BFGS) with line search for (NMOP). And so PQNM(H-BFGS) with line search for (NMOP) is the most effective method and the introduced proximal quasi-Newton methods with line search for (NMOP) are more effective than PGM with line search for (NMOP).

![Fig.1 Result for PGM](image1)

![Fig.2 Result for PQNM(BFGS)](image2)

![Fig.3 Result for PQNM(SS-BFGS)](image3)

![Fig.4 Result for PQNM(H-BFGS)](image4)

6.2 Numerical experiment of algorithm 2
Firstly, the experimental result of the PGM without line search (i.e., Algorithm 3.2 in [11]) for (NMOP) is shown in Fig. 6. And the experimental results obtained by using PQNM(BFGS), PQNM(SS-BFGS), and PQNM(H-BFGS) without line search for (NMOP) with deferent $\delta$ is shown in Fig. 7, Fig. 8, Fig. 9, respectively. It is obvious that the objective values become smaller as $\delta$ gets smaller. Secondly, Fig.10 shows the comparison among the (PGM), PQNM(BFGS), PQNM(SS-BFGS) and PQNM(H-BFGS) without line search for (NMOP) when $\delta$ is 0, 0.05 and 0.1, respectively. When $\delta$ is same, we also observed that the Pareto frontier of PQNM(H-BFGS) without line search for (NMOP) is the lowest, the Pareto frontier of PGM without line search for (NMOP) is the highest, and the Pareto frontier of PQNM(SS-BFGS) without line search for (NMOP) is lower than that of the PQNM(BFGS) without line search for (NMOP). And so PQNM(H-BFGS) without line search for (NMOP) is the most effective method and the introduced proximal quasi-Newton methods without line search for (NMOP) are more effective than PGM without line search for (NMOP).

Fig.5 Comparison of PGM, PQNM(BFGS), PQNM(SS-BFGS) and PQNM(H-BFGS) when $\delta$ is 0, 0.05 and 0.1, respectively
7 Conclusion

First, for a special type of unconstrained nonsmooth multiobjective optimization problems, where each objective function is the sum of a twice continuously differentiable strongly convex function and a proper convex but not necessarily differentiable function, the proximal quasi-Newton method with and without line search is proposed. Secondly, under appropriate conditions, we prove that each cluster point of the sequence generated by these two types of algorithms is the Pareto stationary point of the multiobjective optimization problem. Thirdly, we give their applications in robust multiobjective optimization, and we show that the subproblems of the proximal quasi-Newton method can be regarded as quadratic programming problems. Finally, numerical experiments are performed to verify the effectiveness of the proposed algorithms. In the future, we will analyze the convergence rate of the proposed algorithms.
Fig. 10 Comparison of PGM, PQNM(BFGS), PQNM(SS-BFGS) and PQNM(H-BFGS) when $\delta$ is 0, 0.05 and 0.1, respectively

References

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

[2] Gass S., Saaty T.: The computational algorithm for the parametric objective function[J]. Naval Research Logistics, 1955, 2(1-2):39-45.

[3] Zadeh L. A.: Optimality and non-scalar-valued performance criteria[J]. IEEE Transactions on Automatic Control, 1963, 8(1):59-60.

[4] Ellen H.F., Luis M.G.D.: A Survey on Multiobjective Descent Method[J]. Pesquisa Operacional, (2014)34(3):585-620.

[5] Fliege J., Svaiter B. F.: Steepest descent methods for multicriteria optimization[J]. Mathematical Methods of Operations Research, 2000, 51(3):479-494.

[6] Bello Cruz J.Y., Lucambio Prez L. R., Melo J.G.: Convergence of the projected gradient method for quasiconvex multiobjective optimization[J]. Nonlinear Analysis, 2011, 74(16):5268-5273.
[7] Bonnel H., Iusem A. N., Svaiter B. F.: Proximal Methods in Vector Optimization[J]. SIAM Journal on Optimization, 2005, 15:953-970.

[8] Da Cruz Neto J. X., Da Silva G. J. P., Ferreira O. P., et al.: A subgradient method for multiobjective optimization[J]. Computational Optimization and Applications, 2013, 54(3):461-472.

[9] Fliege J., Grana Drummond L. M., Svaiter B. F.: Newton’s Method for Multiobjective Optimization[J]. SIAM Journal on Optimization, 2009, 20(2):602-626.

[10] Povalej Ž.: Quasi-Newton’s method for multiobjective optimization[J]. Journal of Computational and Applied Mathematics, 2014, 255:765-777.

[11] Tanabe H., Fukuda E. H., Yamashita N.: Proximal gradient methods for multiobjective optimization and their applications[J]. Computational Optimization and Applications, 2019, 72:339-361.

[12] Nocedal J., Wright S. J.: Numerical Optimization, second ed., Springer Science+Business Media, LLC, New York, 2006.

[13] Rockafellar R. T.: Convex analysis, University Press, Princeton, 1970.

[14] Broyden C G.: A new double-rank minimization algorithm[J]. Notices of the American Mathematical Society, 1969, 16:670.

[15] Fletcher, R. A new approach to variable metric algorithms[J]. Computer J, 1970, 13(3):317-322.

[16] Goldfarb D.: A family of variable-metric methods derived by variational means[J]. Mathematics of Computation, 1970, 24(109):23-26.

[17] D, F, Shanno. Conditioning of quasi-Newton methods for function minimization[J]. Math. Comp. 1970(24):647-656.

[18] Sun W, Yuan Y X. Optimization Theory and Methods[M]. Springer, 2006.

[19] Xu Z C. Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations[J]. Journal of Computational and Applied Mathematics, 2001, 137(2):269-278.

[20] Berge C, Patterson E M. Topological spaces[M]. Dover Publications, Edinburgh, 1963.

[21] Hogan W W. Point-to-Set Maps in Mathematical Programming[J]. Siam Review, 1973, 15(3):591-603.

[22] Chen G, Teboulle M. Convergence Analysis of a Proximal-Like Minimization Algorithm Using Bregman Functions[J]. Siam Journal on Optimization, 1993, 3(3):538-543.

[23] Bertsekas D P. Nonlinear Programming, 2nd edn[M]. Athena Scientific, Belmont, 1999.
[24] Ehrgott M, Ide J, A Schobel. Minmax robustness for multi-objective optimization problems[J]. European Journal of Operational Research, 2014, 239(1):17-31.

[25] Fliege J, Werner R. Robust multiobjective optimization and applications in portfolio optimization[J]. European Journal of Operational Research, 2014, 234(2):422-433.

[26] Morishita M. A descent method for robust multiobjective optimization in the presence of implementation errors. Masters thesis, Kyoto University, 2016.

[27] Ben-Tal A, Nemirovski A. Robust Convex Optimization[J]. Mathematics of Operations Research, 1998, 23(4):769-805.