J. Nonlinear Var. Anal. 8 (2024), No. 4, pp. 625-639
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.8.2024.4.09

CONVERGENCE OF A NEW NONMONOTONE MEMORY GRADIENT METHOD FOR UNCONSTRAINED MULTIOBJECTIVE OPTIMIZATION VIA ROBUST APPROACH

YUSHAN BAI[1], JIAWEI CHEN[1,*], LIPING TANG[2], TAO ZHANG[3,*]

[1]*School of Mathematics and statistics, Southwest University, Chongqing 400715, China*
[2]*National Center for Applied Mathematics in Chongqing, Chongqing 401331, China*
[3]*School of Information and Mathematics, Yangtze University, Jingzhou 434023, China*

Abstract. Robust approach is a special scalarization method to deal with multiobjective optimization problems in the worst-case. In this paper, we propose a new non-monotone gradient type algorithm for solving unconstrained multiobjective optimization problems by the conjugate technique and the robust approach. The proposed method has a memory gradient property since the search direction is constructed by using the current descent direction and the past multi-step iterative descent directions. For this, the search direction is called a memory gradient search direction. The step-size is computed by the non-monotone linear search. A lower bound of the stepsize is presented under some mild conditions. Then the iterative sequence generated by the proposed method is proved to be convergent to a Pareto critical point of the multiobjective optimization problem under some mild conditions. Numerical experiments are reported to show the effectiveness of the proposed method.

Keywords. Multiobjective optimization; Memory gradient direction; Nonmonotone line search; Pareto critical point; Robust approach.

1. INTRODUCTION

Multiobjective optimization is an important topic in the optimization community, and has been widely applied to solve many practical problems arising from economic management, engineering design, machine learning, and so on; see, e.g., [1–4]. Due to the difficulty of several objectives to be minimized simultaneously, various optimal solution notions, such as Pareto efficient solutions, weak efficient solutions, and properly efficient solutions, were introduced according to the different decision situations or decision preferences. The theory of multiobjective optimization has been extensively established; see, e.g., [5–7]. However, the numerical algorithms for multiobjective optimization are still deserved to be developed due to the actual needs of applications. The development of strategies for solving multiobjective optimization

*Corresponding author.
E-mail address: YushanBai3@outlook.com (Y. Bai), j.w.chen713@163.com (J. Chen), tanglipings@163.com (L. Tang), tzhang@yangtzeu.edu.cn (T. Zhang).
Received 10 March 2023; Accepted 5 January 2024; Published 1 May 2024.

problems has attracted wide attention, and many iterative methods for solving scalar optimization problems have been extended to multiobjective optimization, such as the projected gradient method [8, 9], the proximal point method [10–13], the steepest descent method [14, 15], the Newton method [16, 17], the trust-region method [18, 19], and so on.

Among these strategies for solving multiobjective optimization problems, there are mainly two different kinds of approaches. One is linearized scalarization methods such as the weighted normalization method, which aims to convert multiobjective optimization problems into the parameterized single objective ones; see [11, 20, 21]. As described in [16, 22], the drawback of this method is that even when the original vector-valued problem has solutions, the selection of weighted parameters is not easy to determined in advances; and the other one is nonlinear scalarization methods such as oriented distance function and Tammer type scalarization function by which do not require any weighted parameters information; see [8, 14, 17]. By these strategies, the first-order and second-order methods, such as the gradient descent method, the proximal gradient method, the Newton method, and the conjugate-gradient method of optimization problems were extended from one objective to multiobjective. The basic iterative framework is as follows:

$$x^{k+1} = x^k + \rho_k d^k, \quad k = 0, 1, 2, 3, \ldots \tag{1.1}$$

where d^k is the search direction such as steepest descent direction and conjugate gradient direction [14, 23–25], and ρ^k is the stepsize calculated by using 0.618 method, Armijo line search and Wolfe line search; see [8, 9, 14, 26–28]. In [29], the nonlinear conjugate gradient method for vector optimization was proposed in which the search direction d^k was defined as

$$d^k = \begin{cases} v\left(x^k\right), & \text{if } k = 0, \\ v\left(x^k\right) + \alpha^k d^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{1.2}$$

where α^k is a conjugate parameter. Considering the multi-step iterative search directions defined by (3.2), a new conjugate gradient method, termed as the multiobjective memory gradient method (MMG), was proposed to find Pareto critical points of a unconstrained smooth multiobjective optimization problem with Armijio-type monotone line search in [27]. Numerical experiments illustrated that MMG was superior than the multiobjective nonlinear conjugate gradient methods. In [27], the stepsize was obtained by Armijio-type monotone line search and adaptive step size.

As we know, the monotone linear search made the decreasing of objective function values at each iteration, which may slow the convergence rate in the minimization process, especially in the presence of narrow curved valley; see [30]. Considering the limitation of the monotone type linear search, some nonmonotone line search methods have been proposed, which allowed that the value of objective function can be increased in some iterations to improve the rate of convergence in the process of line search; see [26, 30–32]. As pointed out in [31], the Zhang-Hager type nonmonotone line search technique requires the average value of successive functions to be reduced and has been proved to be more efficient than monotone or traditional nonmonotone strategies. A max-type nonmonotone line search method was proposed for solving multiobjective constrained problems in [33]. The nonmonotone line search techniques [30, 31] were extended from the single objective to multiobjective case; see e.g., [9, 26, 28, 32, 34] and the references therein.

Inspired by the above works, we propose a new memory gradient method for a multiobjective optimization problem by the non-monotone line search rule [26, 31, 32, 34] and the robust approach. The search direction of the proposed method is constructed by using the current descent direction and the past multi-step iterative descent directions. Under mild assumptions, we show that the sequence generated via this method converges to a Pareto critical point of the problem. Numerical experiments are reported to verify the effectiveness of the proposed method.

The paper is organized as follows. In Section 2, we recall some definitions and basic results. In Section 3, we propose the multiobjective nonmonotone memory gradient method (MNMG) and its some properties. The convergence of the iterative sequence generated by the MNMG is analyzed in Section 4. In Section 5, some numerical tests are reported to verify the efficiency of the MNMG. Finally, we give the conclusions in 6.

2. PRELIMINARIES

Throughout this paper, if not otherwise stated, let $\mathbb{R}$ be the real number field and $\mathbb{R}^n$ be the n-dimensional Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$, i.e., $\|x\| = \sqrt{\langle x,x\rangle}$ for all $x\in\mathbb{R}^n$. For a positive integer $m\in\mathbb{Z}_+$, $\mathbb{R}^m_+$, and $\mathbb{R}^m_{++}$ denote the nonnegative orthant and positive orthant of $\mathbb{R}^m$, set $\langle m\rangle := \{1,2,\cdots,m\}$, $e := (1,1,\cdots,1)^\top \in\mathbb{R}^m$, where $\top$ denotes the transpose, and denote the zero vector of $\mathbb{R}^n$ by 0_n. We also denote $\chi^+(\chi\in\mathbb{R})$ by

$$\chi^+ = \begin{cases} 0, & \text{if } \chi = 0, \\ \frac{1}{\chi}, & \text{if } \chi\in\mathbb{R}\setminus\{0\}. \end{cases} \tag{2.1}$$

It is easy to see that $\chi\chi^+ \leqslant 1$ and $\chi\chi^+ = 1$ whenever $\chi\neq 0$. The norm of a real matrix $A = \left(A_{i,j}\right)_{m\times n}\in\mathbb{R}^{m\times n}$ is calculated as

$$\|A\| = \max_{x\neq 0}\frac{\|Ax\|_\infty}{\|x\|} = \max_{i\in\langle m\rangle}\|A_{i,\cdot}\| = \max_{i\in\langle m\rangle}\sqrt{\sum_{j=1}^n A_{i,j}^2}\,. \tag{2.2}$$

We give the partial orders that $y\preccurlyeq z\ (z\succcurlyeq y)\ \Leftrightarrow\ z-y\in\mathbb{R}^m_+$ and $y\prec z\ (z\succ y)\ \Leftrightarrow\ z-y\in\mathbb{R}^m_{++}$ for $y,z\in\mathbb{R}^m$.

We in this paper study the following unconstrained multiobjective optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) := (f_1(x), f_2(x),\cdots,f_m(x))^\top, \tag{2.3}$$

where $f:\mathbb{R}^n\to\mathbb{R}^m$ is a continuously differentiable function with the lower boundedness, i.e., f_i are continuously differentiable and bounded from below for all $i\in\langle m\rangle$.

A point $x^*\in\mathbb{R}^n$ is called a *Pareto optimal point* (or *Pareto point*) of (2.3) if there exists no other point $x\in\mathbb{R}^n$ such that $f(x)\preccurlyeq f(x^*)$ and $f(x)\neq f(x^*)$. Similarly, a point $x^*\in\mathbb{R}^n$ is called a *weak Pareto optimal point* (or *weak Pareto*) of (2.3) if there exists no other point $x\in\mathbb{R}^n$ such that $f(x)\prec f(x^*)$.

It is clear that a Pareto optimal point is also a weak Pareto optimal point but not vice versa. The set of the objective values of all Pareto optimal solutions is also called Pareto frontier. Since f is continuously differentiable, the Jacobian of f at $x\in\mathbb{R}^n$ is denoted by

$$\mathfrak{J}f(x) := (\nabla f_1(x),\ldots,\nabla f_m(x))^\top, \tag{2.4}$$

and the image of the Jacobian of F at x is denoted by

$$Im(\mathfrak{J}f(x)) := \{\mathfrak{J}f(x)d\,:\,d\in\mathbb{R}^n\},$$

where $\mathfrak{J}f(x)d := (\langle\nabla f_1(x),d\rangle, \langle\nabla f_2(x),d\rangle, \cdots, \langle\nabla f_m(x),d\rangle)^\top$.

It is well-known [6, 15, 35] that if $x \in \mathbb{R}^n$ is a weak Pareto optimal point to problem (2.3), then the first order optimality necessary condition for problem (2.3) can be characterized by

$$\left(-\mathbb{R}^m_{++}\right) \cap \mathrm{Im}(\mathfrak{J}f(x)) = \emptyset, \tag{2.5}$$

which implies that, for any $d \in \mathbb{R}^n$, there exists at least one $i^* \in \langle m\rangle$ such that

$$(\mathfrak{J}f(x)d)_{i^*} = \langle\nabla f_{i^*}(x),d\rangle \geq 0.$$

So, for each $d \in \mathbb{R}^n$, $\max\limits_{i\in\langle m\rangle}\langle\nabla f_{i^*}(x),d\rangle \geq 0$. We also call the function $\max\limits_{i\in\langle m\rangle}\langle\nabla f_{i^*}(x),d\rangle$ as worst-case function with respect to the index $i \in \langle m\rangle$. Moreover, $x \in \mathbb{R}^n$ is a local weak Pareto optimal point of problem (2.3) if and only if $d = 0_n$ is the solution to the robust optimization problem:

$$\min_{d\in\mathbb{R}^n} \max_{i\in\langle m\rangle} \langle\nabla f_i(x),d\rangle. \tag{2.6}$$

In particular, if f is a convex function, i.e., f_i are convex functions for $i \in \langle m\rangle$, then $x \in \mathbb{R}^n$ is a weak Pareto optimal point to problem (2.3) if and only if $d = 0_n$ is the solution to robust optimization problem (2.6). The robust approach (2.6) is very important method to design numerical algorithms for solving multiobjective optimization problem (2.3), which can also be regarded as a scalarization method.

A point $x \in \mathbb{R}^n$ satisfying (2.5) is said to be a *Pareto critical point* of problem (2.3). Observe that if $x \in \mathbb{R}^n$ is not a Pareto critical point, then there exists a direction $d \in \mathbb{R}^n$ such that $\mathfrak{J}f(x)d \prec 0$, which implies that there exists $\varepsilon > 0$ such that $f(x+\alpha d) \prec f(x)$ for all $\alpha \in (0,\varepsilon]$, i.e., d is a descent direction of f at x.

We now define a function $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as follows

$$\phi(x,d) = \max_{i\in\langle m\rangle} \langle\nabla f_i(x),d\rangle. \tag{2.7}$$

Clearly, ϕ is convex with respect to the second argument. Then problem (2.6) can be equivalently reformulated as the following convex optimization:

$$\min_{d\in\mathbb{R}^n} \phi(x,d). \tag{2.8}$$

As we know, proximal point method is a popular approach to solve the convex optimization problem (2.8) by the proximal function ϕ with respect to the variable d, i.e.,

$$\mathrm{prox}_\phi(0_n) = \arg\min_{d\in\mathbb{R}^n} \phi(x,d) + \frac{1}{2}\|d-0_n\|^2.$$

It is easy to see that $\mathrm{prox}_\phi(0_n)$ is well-defined, single-valued, and dependent on x since $\phi(x,d) + \frac{1}{2}\|d\|^2$ is a strongly convex function, so we set $v(x) := \mathrm{prox}_\phi(0_n)$, i.e.,

$$v(x) = \arg\min_{d\in\mathbb{R}^n} \phi(x,d) + \frac{1}{2}\|d\|^2, \tag{2.9}$$

and set the optimal value

$$\theta(x) := \phi(x,v(x)) + \frac{1}{2}\|v(x)\|^2 = \min_{d\in\mathbb{R}^n} \phi(x,d) + \frac{1}{2}\|d\|^2. \tag{2.10}$$

Clearly, $\theta(x) \leq 0$ for all $x \in \mathbb{R}^n$. By the classical convex analysis, we conclude from [28, p. 78] that there exist $\lambda_i(x) \geq 0, i \in \langle m \rangle$ and $\sum_{i=1}^m \lambda_i(x) = 1$ such that $v(x) = -\sum_{i=1}^m \lambda_i(x) \nabla f_i(x)$ and

$$\phi(x, v(x)) = -\|v(x)\|^2, \ \ \theta(x) = -\frac{1}{2}\|v(x)\|^2. \tag{2.11}$$

Proposition 2.1. [36] *For any $x, y, d_1, d_2 \in \mathbb{R}^n$ and $t \geq 0$, the following statements hold:*

(i) $\phi(x, td_1) = t\phi(x, d_1)$;
(ii) $\phi(x, d_1 + d_2) \leq \phi(x, d_1) + \phi(x, d_2)$;
(iii) $|\phi(x, d_1) - \phi(y, d_2)| \leq \|\mathfrak{J}f(x)d_1 - \mathfrak{J}f(y)d_2\|$;
(iv) *$d \in \mathbb{R}^n$ is a descent direction at $x \in \mathbb{R}^n$ if and only if $\phi(x, d) < 0$;*
(v) *$x \in \mathbb{R}^n$ is a Pareto critical point if and only if $\phi(x, d) \geq 0$ for any $d \in \mathbb{R}^n$.*

Proposition 2.2. [14] *Let the functions $v(\cdot)$ and $\theta(\cdot)$ be defined by (2.9) and (2.10). Then the following assertions hold:*

(i) *$v(\cdot)$ and $\theta(\cdot)$ are continuous;*
(ii) *The following statements are equivalent:*
  (a) *$x \in \mathbb{R}^n$ is not a Pareto critical point;*
  (b) *$v(x) \neq 0$;*
  (c) *$\theta(x) < 0$.*

Remark 2.1. From Proposition 2.1 (iv), we deduce that if $v(x) \neq 0$, then $v(x)$ is a descent direction of problem (2.3) at $x \in \mathbb{R}^n$ which results from (2.9), (2.10), and (2.11). Further, Proposition 2.2 (ii) shows a stronger conclusion that $v(x)$ is a descent direction of problem (2.3) at $x \in \mathbb{R}^n$ if and only if $v(x) \neq 0$. If $x \in \mathbb{R}^n$ is not a Pareto critical point, we have

$$\max_{i \in \langle m \rangle} \langle \nabla f_i(x), v(x) \rangle \leq -\frac{\|v(x)\|^2}{2} < 0.$$

Proposition 2.3. [16, Theorem 3.1] *If f is convex and $\bar{x}$ is a Pareto critical point, then $\bar{x}$ is a weakly Pareto optimal point of problem (2.3).*

3. Multiobjective Nonmonotone Memory Gradient Method

In this section, we propose a new multiobjective nonmonotone memory gradient algorithm (MNMG) for problem (2.3) by using the conjugate technique and the robust approach, and then present some basic results of MNMG.

The MNMG generates some sequence $\{x^k\}$ by the following basic iterative procedure:

$$x^{k+1} = x^k + \rho_k d^k, \tag{3.1}$$

where d^k is a combination of the current steepest descent direction and past multi-step descent direction defined as follows:

$$d^k = \begin{cases} \gamma_k v\left(x^k\right), & \text{if } k = 0, \\ \gamma_k v\left(x^k\right) + \sum_{j=1}^{N_k} \alpha_{kj} d^{k-j}, & \text{if } k \geq 1. \end{cases} \tag{3.2}$$

The search direction inherits the memory property of conjugate gradient method.

We now present the multiobjective nonmonotone memory gradient algorithm for problem (2.3) with the nonmonotone Armijo-type line search.

Algorithm 3.1. [Multiobjective Nonmonotone Memory Gradient Method (MNMG)]

Step 1. *Choose parameters* $\sigma \in (0,1)$, $\gamma_0 > 0$, $\mu > 0$, $N \in \mathbb{Z}_+$, *and* $\delta, \varepsilon \in (0,1)$. *Let* $x^0 \in \mathbb{R}^n$ *be an arbitrary initial point. Set* $C^0 = f(x^0)$, $q_0 = 1$, *and* $k = 0$.

Step 2. *Compute the direction* $v(x^k) = \arg\min_{d\in\mathbb{R}^n} \phi(x^k, d) + \frac{1}{2}\|d\|^2$.

Step 3. *If* $v(x^k) = 0$, *STOP. Otherwise, go to Step 4.*

Step 4. *Compute the search direction* d^k *by* (3.2), *where* $0 < \gamma_k < +\infty$, $N_k = \min\{N, k\}$, *and*

$$\alpha_{kj} = -\frac{1}{N_k}\phi(x_k, v(x^k))\omega_{kj}^+,\ \omega_{kj} > \max\left\{\frac{\phi(x^k, d^{k-j})}{\gamma_k}, 0\right\}. \tag{3.3}$$

Step 5. *Compute the step-size* $\rho_k = \mu\varepsilon^{h_k}$, *where* h_k *is the smallest nonnegative integer satisfying:*

$$f(x^k + \rho_k d^k) \preceq C^k + \sigma\rho_k\phi(x^k, d^k)e. \tag{3.4}$$

Set $x^{k+1} = x^k + \rho_k d^k$.

Step 6. *Update* q_k *and* C^k *as follows:*

$$q_{k+1} = \delta q_k + 1, \tag{3.5}$$

$$C^{k+1} = \frac{\delta q_k C^k + f(x^{k+1})}{q_{k+1}}. \tag{3.6}$$

Set $k = k+1$, *return to Step 2.*

Remark 3.1. (i) It is worth noting that the nonmonotone line search techniques have been extended to solve multiobjective optimization problems by using some directions, such as steepest descent direction and Newton descent direction. The MNMG combines nonmonotone Armijo-type line search with memory gradient direction, which is different from the memory gradient method [27] with monotone line search and nonmonotone line search without the memory gradient direction for unconstrained multiobjective optimization problems [28].

(ii) As mentioned in [31], if $\delta = 0$, then $C^k = f(x^k)$ and the line search (3.4) is the usual Armijo line search; if $\delta = 1$, then $C^k = \frac{1}{k+1}\sum_{i=0}^k f_i(x_i)$ is the average of all the previous function values.

For the convergence analysis to MNMG, we give the more stringent condition

$$\phi(x^k, d^k) \leq \tau\phi(x^k, v(x^k)), \tag{3.7}$$

for some $\tau > 0$ and any $k \in \mathbb{N}$. In multiobjective optimization, we say that a direction $d^k \in \mathbb{R}^n$ meets the sufficient descent condition at x^k whenever (3.7) holds.

The next property shows that the search direction d^k given by (3.2) is a descent direction.

Proposition 3.1. *Let the direction* d^k *be given by* (3.2). *Then it is a descent direction for all* $k \in \mathbb{N}$.

Proof. The proof is the same as that of [27, Lemma 3.1], so it is omitted. □

Proposition 3.2. *Let the direction* d^k *be given by* (3.2). *Assume that there exists a constant* $\gamma^* > 0$ *such that* $\gamma_k \geq \gamma^*$ *and* ω_{kj} *has the following property:*

$$\omega_{kj} > \frac{\phi(x^k, d^{k-j}) + \|\mathfrak{J}f(x^k)\|\|d^{k-j}\|}{\gamma_k}.$$

Then d^k *satisfies the sufficient descent condition* (3.7) *at* x^k *with* $\tau = \frac{\gamma^*}{2} > 0$ *for all* $k \in \mathbb{N}$.

Proof. The proof is the same as that of [27, Lemma 3.2], so it is omitted. □

Proposition 3.3. *For the iterative sequence $\{x^k\}$ generated by the MNMG, $f(x^k) \preceq C^k$ holds for all $k \in \mathbb{N}$.*

Proof. The proof is the same as that of [31, Lemma 1.1], so it is omitted. □

The following result which shows that the line search in the MNMG is well-defined.

Proposition 3.4. *For the iterative sequence $\{x^k\}$ generated by the MNMG, if x^k is not Pareto critical point, then there exists a stepsize ρ_k that satisfies the nonmonotone Armijo-type line search condition* (3.4).

Proof. It follows from Proposition 3.1 that $\phi(x^k, d^k) < 0$, so $\nabla f_i(x^k)^\top d^k < 0$ for all $i \in \langle m \rangle$. Since f is continuously differentiable and $\sigma \in (0,1)$, then, for any $i \in \langle m \rangle$, there exists some $\bar{\rho}_k \in (0,1)$ such that

$$\begin{aligned} f_i(x^k + \rho d^k) &= f_i(x^k) + \rho \nabla f_i(x^k)^\top d^k + \mathrm{o}(\rho) \\ &\le f_i(x^k) + \sigma\rho\phi(x^k, d^k) \\ &\le C_i^k + \sigma\rho\phi(x^k, d^k),\ \forall\, \rho \in (0, \bar{\rho}_k], \end{aligned}$$

where the last inequality follos from Proposition 3.3. Moreover, one has

$$f(x^k + \rho d^k) \preceq C^k + \sigma\rho\phi(x^k, d^k)e, \forall \rho \in (0, \bar{\rho}_k].$$

It therefore implies that there exists a stepsize ρ_k that satisfies the nonmonotone Armijo-type line search condition (3.4). □

We next show the well-definedness of the MNMG.

Proposition 3.5. *Assume that all conditions of Proposition 3.2 are satisfied. Then the MNMG is well-defined.*

Proof. By the strong convexity of $\phi(x^k, \cdot) + \frac{1}{2}\|\cdot\|^2$, the direction $v(x^k)$ can be uniquely existed. It follows from Propositions 3.1 and 3.2 that $d^k \in \mathbb{R}^n$ is a descent direction. Besides, Proposition 3.4 yields that there exists a stepsize ρ_k that satisfies the nonmonotone Armijo-type line search condition (3.4). Consequently, the MNMG is well-defined. □

We now present the usual assumption which will be used to estimate the stepsize ρ_k.

Assumption 3.1. For all $k \in \mathbb{N}$, if $\rho_k < \mu\varepsilon$, $\nabla f_i\,(i \in \langle m \rangle)$ satisfy the following Lipschitz condition with Lipschitz constant ℓ:

$$\|\nabla f_i(x) - \nabla f_i(x^k)\| \le \ell \|x - x^k\|, \tag{3.8}$$

for all x on the line segment connecting x^k and $x^k + \frac{1}{\varepsilon}\rho_k d^k$.

Remark 3.2. It should be pointed out that Assumption 3.1 was applied to show the boundedness from below of the stepsize of a non-monotone gradient method in [26]. In particular, if $f_i\,(i \in \langle m \rangle)$ are continuously differentiable with gradient Lipschitz continuity, Assumption 3.1 is natural.

4. Convergence Analysis

In this section, we analyze the convergence of MNMG under some suitable assumptions.

From the iterative procedure of MNMG, we know that MNMG stops when $v(x^k) = 0$, which implies that x^k is a Pareto critical point by Proposition 2.2 (ii). If MNMG stops at x^k with finite steps k, then x^k is a Pareto critical point. We consider the case that the iterative number of the sequence $\{x^k\}$ is infinite in the rest of this paper.

Proposition 4.1. *Let d^k be defined by (3.2) and the sequence $\{x^k\}$ be generated by the MNMG. Assume that the conditions of Proposition 3.2 are satisfied. Then*

$$\phi(x^k, d^k) \leq -\frac{\tau}{2}\|v(x^k)\|^2 = \tau\theta(x^k),$$

where $\tau = \frac{\gamma^*}{2} > 0$.

Proof. It follows from Proposition 3.2 that the direction d^k satisfies the sufficient descent condition (3.7) at x^k. Due to $\theta(x^k) \leq 0$, we have

$$\phi(x^k, d^k) \leq \tau\phi(x^k, v(x^k)) \leq -\frac{\tau}{2}\|v(x^k)\|^2,$$

which together with (2.11) shows that $\phi(x^k, d^k) \leq -\frac{\tau}{2}\|v(x^k)\|^2 = \tau\theta(x^k)$. □

We next present a nonincreasing property of the sequences $\{C_i^k\}$.

Proposition 4.2. *Let the sequence $\{x^k\}$ be the sequence generated by the MNMG. Then $\{C^k\}$ is nonincreasing (i.e., $C^{k+1} \preceq C^k$) and has a limit as $k \to \infty$.*

Proof. The proof is similar to that of [28, 32], so it is omitted. □

Theorem 4.1. *Let the sequence $\{x^k\}$ be generated by the MNMG. Then,*

$$\lim_{k\to\infty} \rho_k\theta(x^k) = 0.$$

Proof. From Proposition 4.2, one has

$$\lim_{k\to\infty} \left(C^k - C^{k+1}\right) = 0.$$

It follows from (3.4) that

$$C^{k+1} = C^k + \frac{f(x^{k+1}) - C^k}{q_{k+1}} \preceq C^k + \frac{\sigma\rho_k\phi(x^k, d^k)}{q_{k+1}}e.$$

Taking into account that $\sigma \in (0,1)$, $\frac{\sigma\rho_k}{q_{k+1}} \geq 0$ and $\phi(x^k, d^k) \leq 0$, one has

$$0 \leq \lim_{k\to\infty} \frac{-\rho_k}{q_{k+1}}\phi(x^k, d^k) \leq \lim_{k\to\infty} \frac{C_i^k - C_i^{k+1}}{\sigma} = 0,\ i \in \langle m\rangle,$$

which implies that

$$\lim_{k\to\infty} \frac{\rho_k}{q_{k+1}}\phi(x^k, d^k) = 0. \tag{4.1}$$

In view of $\delta \in (0,1)$, we have $0 < 1-\delta < 1$. Using (3.5) yields that

$$1+\delta \leq q_{k+1} = 1 + \sum_{j=0}^{k} \delta^{j+1} \leq \sum_{j=0}^{\infty} \delta^j = \frac{1}{1-\delta}, \tag{4.2}$$

so $0 < 1-\delta \leq \frac{1}{q_{k+1}} \leq \frac{1}{1+\delta}$ for all $k \in \mathbb{N}$. In turn, we conclude from Proposition 4.1 that

$$\frac{\rho_k}{q_{k+1}}\phi(x^k,d^k) \leq \frac{\rho_k}{q_{k+1}}\tau\theta(x^k) \leq 0. \tag{4.3}$$

Therefore, taking the limit in both sides of the above formula (4.3) as $k \to \infty$, which together with (4.1) shows that $\lim_{k\to\infty} \rho_k\theta(x^k) = 0$. □

We present a lower bound estimation of the stepsize ρ_k generated by (3.4).

Proposition 4.3. *Let the sequence $\{x^k\}$ be generated by the MNMG, d^k be defined by (3.2), and the stepsize ρ_k be defined by (3.4). Assume that Assumption 3.1 and the conditions of Proposition 3.2 hold. Then, for each k,*

$$\rho_k \geq \min\left\{\mu\varepsilon, \frac{\varepsilon(\sigma-1)\phi(x^k,d^k)}{\ell\|d^k\|^2}\right\}. \tag{4.4}$$

Proof. Since $\phi(x^k,d^k) \leq 0$ and $\varepsilon,\sigma \in (0,1)$ for each k, then $\frac{\varepsilon(\sigma-1)\phi(x^k,d^k)}{\ell\|d^k\|^2+\varepsilon^k} \geq 0$. If $\rho_k \geq \mu\varepsilon$, then (4.4) holds. If $\rho_k < \mu\varepsilon$, then $\mu\varepsilon^{h_k} = \rho_k < \mu\varepsilon$, so $h_k \geq 2$. Set $s_k := \frac{1}{\varepsilon}\rho_k = \mu\varepsilon^{h_k-1}$. By (3.4), we have

$$f(x^k + s_kd^k) \not\preceq C^k + \sigma s_k\phi(x^k,d^k)e.$$

Clearly, $d^k \neq 0$ because of $f(x^k) \not\preceq C^k$ and $\phi(x^k,0) = 0$. Thus there exists $j_k \in \langle m\rangle$ such that

$$f_{j_k}(x^k+s_kd^k) > C^k_{j_k} + \sigma s_k\phi(x^k,d^k) \geq f_{j_k}(x^k) + \sigma s_k\phi(x^k,d^k).$$

Since f is continuously differentiable, by the Mean-Value Theorem, there exists $\iota_k \in [0,1]$ such that $f_{j_k}(x^k+s_kd^k) - f_{j_k}(x^k) = \langle\nabla f_{j_k}(x^k+s_k\iota_kd^k), s_kd^k\rangle$, so

$$\langle\nabla f_{j_k}(x^k+s_k\iota_kd^k), d^k\rangle > \sigma\phi(x^k,d^k) \geq \sigma\langle\nabla f_{j_k}(x^k), d^k\rangle,$$

where the second inequality results from (2.7). It follows that

$$\begin{aligned}
0 \leq (\sigma-1)\phi(x^k,d^k) &< \langle\nabla f_{j_k}(x^k+s_k\iota_kd^k), d^k\rangle - \phi(x^k,d^k) \\
&\leq \langle\nabla f_{j_k}(x^k+s_k\iota_kd^k), d^k\rangle - \langle\nabla f_{j_k}(x^k), d^k\rangle \\
&\leq \|\nabla f_{j_k}(x^k+s_k\iota_kd^k) - \nabla f_{j_k}(x^k)\|\|d^k\| \\
&< \ell s_k\|d^k\|^2 = \frac{1}{\varepsilon}\rho_k\ell\|d^k\|^2,
\end{aligned}$$

so $\rho_k \geq \frac{\varepsilon(\sigma-1)\phi(x^k,d^k)}{\ell\|d^k\|^2}$. Thus the stepsize ρ_k has a bounded below satisfying (4.4). □

Theorem 4.2. *Let the sequence $\{x^k\}$ be generated by the MNMG and all conditions of Proposition 3.2 be satisfied. Then the following assertions are true*

(i) *each accumulation point of the sequence $\{x^k\}$ is Pareto critical point of problem (2.3);*
(ii) *assume that the level set $\{x \in \mathbb{R} : f(x) \preceq f(x^0)\}$ is bounded. Then there exists at least one Pareto critical point $\bar{x}$ of problem (2.3) such that $\bar{x}$ is an accumulation point of iterative sequence $\{x^k\}$. Further, if f is convex, then there exists at least one Pareto optimal point $\bar{x}$ of problem (2.3) such that $\bar{x}$ is an accumulation point of $\{x^k\}$.*

Proof. (i) Let $\bar{x}$ be an accumulation point of $\{x^k\}$. Then there exists a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ such that $x^{k_j} \to \bar{x}$ as $j \to \infty$. From Proposition 2.2 we conclude that $v(\cdot)$ and $\theta(\cdot)$ are continuous. It follows that $\lim_{j\to\infty} v(x^{k_j}) = \lim_{j\to\infty} v(\bar{x})$ and $\lim_{j\to\infty} \theta(x^{k_j}) = \theta(\bar{x})$. In view of Theorem 4.1, one has

$$0 = \lim_{j\to\infty} \rho_{k_j} \theta(x^{k_j}) = \lim_{k\to\infty} \rho_{k_j} \lim_{j\to\infty} \theta(x^{k_j}) = \theta(\bar{x}) \lim_{k\to\infty} \rho_{k_j}.$$

Then there exists at least $\theta(\bar{x}) = 0$ or $\lim\limits_{k\to\infty} \rho_{k_j} = 0$. It therefore follows from Proposition 2.2 that $\bar{x}$ is a Pareto critical point of problem (2.3) when $\theta(\bar{x}) = 0$.

We now consider the case $\lim\limits_{k\to\infty} \rho_{k_j} = 0$. If $\bar{x}$ is not Pareto critical point of problem (2.3), then $v(\bar{x}) \neq 0$. For sufficiently large j, there exist $\bar{i} \in \langle m \rangle$ and $\varepsilon \in (0,1]$ such that $0 < \rho_{k_j} < \varepsilon$ and

$$f_{\bar{i}}(x^{k_j} + \varepsilon^{-1}\rho_{k_j} d^{k_j}) > C_{\bar{i}}^{k_j} + \sigma\varepsilon^{-1}\rho_{k_j}\phi(x^{k_j}, d^{k_j}) \geq f_{\bar{i}}(x^{k_j}) + \sigma\varepsilon^{-1}\rho_{k_j}\phi(x^{k_j}, d^{k_j}),$$

where the first inequality results from (3.4) and the second inequality follows from Proposition 3.3. It implies that

$$\begin{aligned}
\frac{f_{\bar{i}}(x^{k_j} + \varepsilon^{-1}\rho_{k_j} d^{k_j}) - f_{\bar{i}}(x^{k_j})}{\varepsilon^{-1}\rho_{k_j}} &\geq \sigma\phi(x^{k_j}, d^{k_j}) \\
&\geq \sigma \left\langle \nabla f_{\bar{i}}(x^{k_j}), d^{k_j} \right\rangle \\
&= \frac{\sigma\varepsilon^{-1}\rho_{k_j} \nabla f_{\bar{i}}\left(x^{k_j}\right)^\top d^{k_j}}{\varepsilon^{-1}\rho_{k_j}} \\
&= \sigma \left(\frac{f_{\bar{i}}(x^{k_j} + \varepsilon^{-1}\rho_{k_j} d^{k_j}) - f_{\bar{i}}(x^{k_j})}{\varepsilon^{-1}\rho_{k_j}} + \frac{o(\varepsilon^{-1}\rho_{k_j}\|d^{k_j}\|)}{\varepsilon^{-1}\rho_{k_j}} \right).
\end{aligned}$$

Moreover, we have

$$\frac{f_{\bar{i}}(x^{k_j} + \varepsilon^{-1}\rho_{k_j} d^{k_j}) - f_{\bar{i}}(x^{k_j})}{\varepsilon^{-1}\rho_{k_j}} \geq \frac{\sigma}{1-\sigma} \frac{o(\varepsilon^{-1}\rho_{k_j}\|d^{k_j}\|)}{\varepsilon^{-1}\rho_{k_j}},$$

and

$$\phi(x^{k_j}, d^{k_j}) \geq \frac{1}{1-\sigma} \frac{o(\varepsilon^{-1}\rho_{k_j}\|d^{k_j}\|)}{\varepsilon^{-1}\rho_{k_j}}. \tag{4.5}$$

Since $v(\cdot)$ is continuous, then $\|v(\cdot)\|^2$ is continuous. Again, from $v(\bar{x}) \neq 0$, it implies that there exists some $\xi > 0$ such that $\|v(x^{k_j})\|^2 \geq \xi$. Then we deduce from (4.5) and Proposition 4.1 that

$$0 > -\frac{\tau\xi}{2} \geq \lim_{j\to\infty} \phi(x^{k_j}, d^{k_j}) \geq \lim_{j\to\infty} \frac{1}{1-\sigma} \frac{o(\varepsilon^{-1}\rho_{k_j}\|d^{k_j}\|)}{\varepsilon^{-1}\rho_{k_j}} = 0,$$

which is a contradiction. Consequently, we obtain $v(\bar{x}) = 0$, so $\bar{x}$ is a Pareto critical point of problem (2.3).

(ii) It follows from Propositions 3.3 and 4.2 that $\{C_i^k\}$ is nonincreasing and $f(x^k) \preceq C^k$ for all $k \in \mathbb{N}$. Thus $f(x^k) \preceq C^k \preceq \cdots \preceq C^0 = f(x^0)$ for all $k \in \mathbb{N}$. Since the level set $\{x \in \mathbb{R} : f(x) \preceq f(x^0)\}$ is bounded, then $\{x^k\}$ is bounded, which implies that $\{x^k\}$ has at least one accumulation point. It thus follows from (i) that the accumulation point of $\{x^k\}$ is a Pareto critical point of problem (2.3), which together with Proposition 2.3 yields that the accumulation point of $\{x^k\}$ is a Pareto optimal point of problem (2.3) when f is convex. □

5. Numerical Experiments

In this section, we present some numerical experiments to illustrate the potential practical advantages of the MNMG for solving multiobjective problems by comparing MNMG, multiobjective memory gradient (MMG) [27] and multiobjective nonmonotone steepest descent (MNSD) [28] method.

The codes are edited in Matlab programming language, and implemented in Matlab R2021b on a computer CPU Intel(R) Core(TM) i5-10210U 1.60GHz and 8GB RAM. We use $\sigma = 10^{-4}$, $\mu = 1$, $\varepsilon = 0.5$, and $N = 5$ for all $k \in \mathbb{N}$, and we set

$$\omega_{kj} = \frac{\phi(x^k, d^{k-j}) + \|\mathfrak{J}f(x^k)\|\|d^{k-j}\| + 1}{\gamma_k}. \tag{5.1}$$

It follows that the value of ω_{kj} for $k \in \mathbb{N}, j \in \langle N_k \rangle$ meet with the condition in Lemma 3.2, and we consider the stop condition $v(x^k) < 10^{-6}$, or maximum number of iterations equals to 1000.

Example 5.1. Let $f : \mathbb{R}^n \to \mathbb{R}^2$ be a vector-valued function with the form of $f(x) = (f_1(x), f_2(x))$, where $f_1(x) := \|Ax - b\|_2^2$, $f_2(x) := \|x\|_2^2$, $A \in \mathbb{R}^{n\times n}$, and $b \in \mathbb{R}^n$.

In the experiments, we let $n = 2$, A, and b be randomly generated by the RAND function in Matlab. We consider two cases of the choice of γ_k firstly:

(i) $\gamma_k = 1$ for all $k \in \mathbb{N}$;
(ii) $\gamma_k = 1$ for $k = 0$ and for $k \geq 1$, we have

$$\gamma_k = \begin{cases} 1, & \text{if } \frac{\|x^k - x^{k-1}\|}{\|v^k - v^{k-1}\|} < \gamma^*, \\ \frac{\|x^k - x^{k-1}\|}{\|v^k - v^{k-1}\|}, & \text{otherwise }, \end{cases} \tag{5.2}$$

where $\gamma^* = 10^{-10}$.

Obviously, the value of γ_k for all $k \in \mathbb{N}$ satisfies the condition in Lemma 3.2, which together with the choice of ω_{kj} for $k \in \mathbb{N}, j \in \langle N_k \rangle$, suggests that d^k for all $k \in \mathbb{N}$ is sufficient descent directions.

Table 1. Average number and average runtime of iterations of the MNMG with different choices of γ_k for Example 5.1.

	AverIter	**AverTime**
Case (i)	32.56	5.36
Case (ii)	12.49	1.24

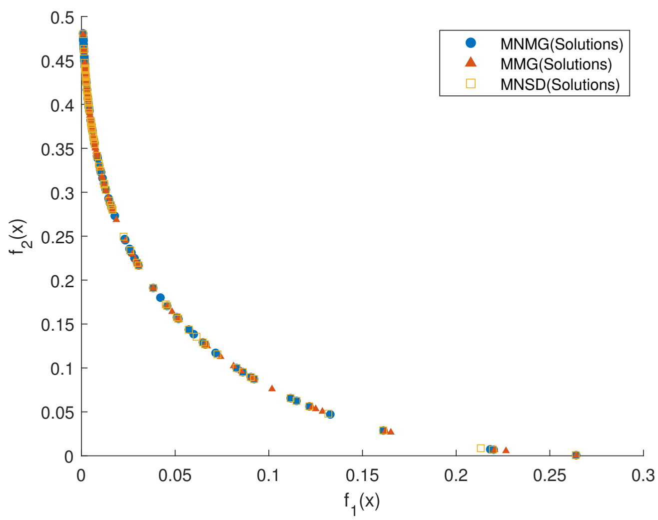
We solve the problem using different 100 initial points from a random distribution. Table 1 shows the average runtime and average number of iterations of the MNMG with two choices of γ_k for Example 5.1, in which Column "AverIter" stands for the average number of iterations, and Column "AverTime" states the average runtime. It follows from Table 1 that Case (ii) is obviously better than Case (i). We take γ_k as in (5.2) in the subsequent discussion.

Furthermore, to show the superior of MNMG, we also compare the numerical performance of MNMG with that of MMG and MNSD. Table 2 shows the average number and average runtime of iterations of the MNMG, MMG and MNSD for Example 5.1. From Table 2, one can see that in terms of the average runtime and the average number of iterations, MMG is pretty better than

TABLE 2. Average number and average runtime of iterations of the MNMG, MMG and MNSD for Example 5.1.

	MNMG	**MMG**	**MNSD**
AverIter	12.49	35.92	31.28
AverTime	1.24	3.55	3.08

FIGURE 1. The Pareto frontiers generated by MNMG, MMG and MNSD for solving Example 5.1.



MMG and MNSD. The Pareto frontiers generated by MNMG, MMG and MNSD for solving Example 5.1 are shown in the Figure 1.

Example 5.2. Let $f(x) = (f_1(x), f_2(x))^\top$ for $x \in \mathbb{R}^2$, where $f_1(x) := (x_1-1)^2 + (x_1-x_2)^2$ and $f_2(x) := (x_2-3)^2 + (x_1-x_2)^2$.

We solve the problem using different 100 initial points from a random distribution to compare the numerical performance of MNMG with that of MMG and MNSD. Table 3 shows the average number and average runtime of iterations of the MNMG, MMG and MNSD for Example 5.2. From Table 3, one can see that MMG is significantly better than MMG and MNSD in terms of the average runtime and the average number of iterations. The Pareto frontiers generated by MNMG, MMG and MNSD for solving Example 5.2 are shown in the Figure 2.

TABLE 3. Average number and average runtime of iterations of the MNMG, MMG and MNSD for Example 5.2.

	MNMG	**MMG**	**MNSD**
AverIter	6.36	12.52	48.49
AverTime	0.69	1.30	4.43

Example 5.3. Let $f : \mathbb{R}^n \to \mathbb{R}^2$, where $f_1(x) := \sum_{i=1}^n \frac{x_i^2}{i}$ and $f_2(x) := \sum_{i=1}^n \frac{(x_i-i)^2}{i}$.

For $n = 5, 10, 20, 50$, we solve the problem using different 100 initial points from a random distribution respectively to compare the numerical performance of MNMG with that of MMG

FIGURE 2. The Pareto frontiers generated by MNMG, MMG and MNSD for solving Example 5.2.

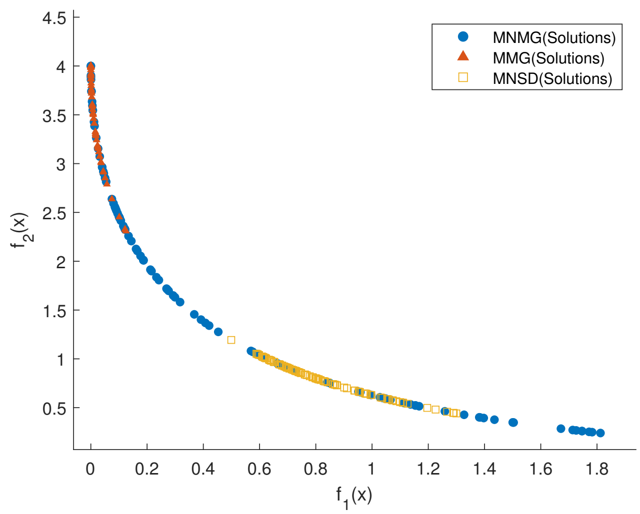


TABLE 4. Average number and average runtime of iterations of the MNMG, MMG and MNSD for Example 5.3.

	MNMG		MMG		MNSD	
	AverIter	AverTime	AverIter	AverTime	AverIter	AverTime
n=5	13.20	1.23	14.01	1.44	43.05	5.74
n=10	22.59	2.33	23.47	2.94	78.84	12.44
n=20	34.77	7.04	36.61	8.63	149.58	46.11
n=50	59.77	11.82	63.56	20.42	356.20	186.96

and MNSD. Table 4 shows the average number and average runtime of iterations of the MNMG, MMG and MNSD with $n = 5, 10, 20, 50$ respectively for Example 5.3. From Table 4, one can see that MMG is slightly better than MMG and obviously better than MNSD in terms of the average runtime and the average number of iterations.

6. CONCLUSION

A new multiobjective nonmonotone memory gradient method was proposed to solve unconstrained smooth multiobjective optimization problem. The proposed method was constructed by the worst-case function and conjugate gradient method in which the search direction has memory property. The asymptotically convergence of the sequences generated by MNMG was established under suitable conditions. Numerical experiments were reported to show the effectiveness of the MNMG. In the future research, it is interesting to study the convergence rate of MNMG under some suitable conditions.

Acknowledgments
The authors would like to express their sincere thanks to the associated editor and anonymous referees for their valuable comments and helpful suggestions, which help to improve the paper. This paper was partially supported by the the Natural Science Foundation of China (12071379,

12171060, and 62373066), and the Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0925, and cstc2022ycjh-bgzxm0097).

REFERENCES

[1] R.T. Marler, J.S. Arora, Survey of multi-objective optimization methods for engineering, Struct. Multid. Optim. 26 (2004), 369-395.
[2] J. Fliege, OLAF–A general modeling system to evaluate and optimize the location of an air polluting facility: OLAF–ein Modellierungssystem zur Standortevaluierung und-optimierung einer luftverschmutzenden Anlage, OR-Spektrum, 23 (2001), 117-136.
[3] A. Hasani, S.M.H. Hosseini, A bi-objective flexible flow shop scheduling problem with machine-dependent processing stages: Trade-off between production costs and energy consumption. Appl. Math. Comput. 386 (2020), 125533.
[4] Y. Jin, Multi-objective Machine Learning, Springer Science & Business Media, 2006.
[5] D. T. Luc, Theory of Vector Optimization. Lecture Notes in Economics and Mathematical Systems, Springer, Berlin, 1989.
[6] J. Jahn, Vector Optimization Theory, Applications and Extensions, Springer, Berlin, 2004.
[7] Q. H. Ansari, E. Köbis, J.C. Yao, Vector Variational Inequalities and Vector Optimization: Theory and Applications, Springer, Cham, 2018.
[8] L.M.G. Drummond, A.N. Iusem, A projected gradient method for vector optimization problems, Comput. Optim. Appl. 28 (2004), 5-29.
[9] X. Zhao, M.A. Köbis, Y. Yao, J.C. Yao, A projected subgradient method for nondifferentiable quasiconvex multiobjective optimization problems, J. Optim. Theory Appl. 190(2021), 82-107.
[10] L.C. Ceng, J.C. Yao, Approximate proximal methods in vector optimization, European J. Oper. Res. 183 (2007), 1-19.
[11] E.A.P. Quiroz, H.C.F. Apolinario, K.D. Villacorta, R. Paulo, A linear scalarization proximal point method for quasiconvex multiobjective minimization, J. Optim. Theory Appl. 183 (2019), 1028-1052.
[12] E.A. Papa Quiroz, N. Baygorrea Cusihuallpa, N. Maculan, Inexact proximal point methods for multiobjective quasiconvex minimization on Hadamard manifolds, J. Optim. Theory Appl. 186 (2020), 879-898.
[13] E.A. Papa Quiroz, S. Cruzado, An inexact scalarization proximal point method for multiobjective quasiconvex minimization, Ann. Oper. Res. 316 (2022), 1445-1470.
[14] J. Fliege, B.F. Svaiter, Steepest descent methods for multicriteria optimization, Math. Meth. Oper. Res. 51 (2000), 479-494.
[15] G.C. Bento, J.X.D.C. Neto, P.R. Oliveira, A. Soubeyran, The self regulation problem as an inexact steepest descent method for multicriteria optimization, European J. Oper. Res. 235 (2014), 494-502.
[16] J. Fliege, L.M.G. Drummond, B.F. Svaiter, Newton's method for multiobjective optimization, SIAM J. Optim. 20 (2009), 602-626.
[17] J. Wang, Y. Hu, C.K.W. Yu, C. Li, X. Yang, Extended Newton methods for multiobjective optimization: majorizing function technique and convergence analysis, SIAM J. Optim. 29 (2019), 2388-2421.
[18] G.A. Carrizo, P.A. Lotito, M.C. Maciel, Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem, Math. Program. 159 (2016), 339-369.
[19] J. Thomann, G. Eichfelder, A trust-region algorithm for heterogeneous multiobjective optimization, SIAM J. Optim. 29 (2019), 1017-1047.
[20] H.C.F. Apolinário, E.A. Papa Quiroz, P.R. Oliveira, A scalarization proximal point method for quasiconvex multiobjective minimization, J. Global Optim. 64 (2016), 79-96.
[21] R.S. Burachik, C.Y. Kaya, M.M. Rizvi, A new scalarization technique and new algorithms to generate Pareto fronts, SIAM J. Optim. 27 (2017), 1010-1034.
[22] L.M.G. Drummond, N. Maculan, B.F. Svaiter, On the choice of parameters for the weighting method in vector optimization, Math. Program. 111 (2008), 201-216.
[23] Y. Narushima, H. Yabe, Global convergence of a memory gradient method for unconstrained optimization, Comput. Optim. Appl. 35 (2006), 325-346.

[24] Y. Zheng, Z. Wan, A new variant of the memory gradient method for unconstrained optimization, Optim. Lett. 6 (2012), 1643-1655.

[25] Y. Ou, Y. Liu, A nonmonotone supermemory gradient algorithm for unconstrained optimization, J. Appl. Math. Comput. 46 (2014), 215-235.

[26] X. Zhao, J.C. Yao, Linear convergence of a nonmonotone projected gradient method for multiobjective optimization, J. Global Optim. 82 (2022), 577-594.

[27] W. Chen, X. Yang, Y. Zhao, Memory gradient method for multiobjective optimization, Appl. Math. Comput. 443 (2023), 127791.

[28] K. Mita, E.H. Fukuda, N. Yamashita, Nonmonotone line searches for unconstrained multiobjective optimization problems, J. Global Optim. 75 (2019), 63-90.

[29] L.R. Lucambio Pérez, L.F. Prudente, Nonlinear conjugate gradient methods for vector optimization, SIAM J. Optim. 28 (2018), 2690–2720.

[30] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's method, SIAM J. Numer. Anal. 23 (1986), 707-716.

[31] H. Zhang, W.W. Hager, A nonmonotone line search technique and its application to unconstrained optimization, SIAM J. Optim. 14 (2004), 1043-1056.

[32] J.W. Chen, J. Liu, X. Qin, J.C. Yao, A non-monotone proximal gradient method for nonsmooth multiobjective optimization problems with application to robust optimization, J. Comput. Appl. Math. Submitted.

[33] S. Qu, Y. Ji, J. Jiang, Q. Zhang, Nonmonotone gradient methods for vector optimization with a portfolio optimization application, European J. Oper. Res. 263 (2017), 356-366.

[34] X.P. Zhao, J.C. Yao, Y. Yao, A nonmonotone gradient method for constrained multiobjective optimization problems, J. Nonlinear Var. Anal. 6 (2022), 693-706.

[35] J. Chen, Y. Dai, Multiobjective optimization with least constraint violation: Optimality conditions and exact penalization, J. Global Optim. 87 (2023), 807–830.

[36] E.H. Fukuda, L.M.G. Drummond, A survey on multiobjective descent methods, Pesquisa Oper. 34 (2014), 585-620.