



Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems

N. S. Fazzio¹ · M. L. Schuverdt¹

Received: 30 January 2018 / Accepted: 3 November 2018 / Published online: 16 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

In this work we consider an extension of the projected gradient method (PGM) to constrained multiobjective problems. The projected gradient scheme for multiobjective optimization proposed by Graña Drummond and Iusem and analyzed by Fukuda and Graña Drummond is extended to include a nonmonotone line search based on the average of the successive previous functions values instead of the traditional Armijo-like rules. Under standard assumptions, stationarity of the accumulation points is established. Moreover, under standard convexity assumptions, we prove full convergence to weakly Pareto optimal solutions of any sequence produced by the proposed algorithm.

Keywords Multiobjective optimization · Projected gradient methods · Nonmonotone line search · Global convergence

1 Introduction

We will consider the constrained multiobjective optimization problem (MOP) of the form:

$$\text{Minimize } F(x) \text{ subject to } x \in C \quad (1)$$

where $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$, $F(x) = (F_1(x), \dots, F_r(x))$ is a continuously differentiable vectorial function in \mathbb{R}^n and $C \subseteq \mathbb{R}^n$ is a closed and convex set.

In a multicriteria setting there are many optimality definitions. Throughout this paper, we are interested in the Pareto and weak Pareto optimality concepts. A feasible point of problem (1) is called Pareto optimum or efficient solution [19] if there is no $x \in C$ such that $F(x) \leq F(x^*)$ and $F(x) \neq F(x^*)$. A point $x^* \in C$ is said to be a weak Pareto optimum point or a weakly efficient solution if there is no $x \in C$

✉ M. L. Schuverdt
schuverd@mate.unlp.edu.ar

N. S. Fazzio
nadiafazzio@gmail.com

¹ CONICET, Department of Mathematics, FCE, University of La Plata, La Plata, Argentina

such that $F(x) < F(x^*)$, where the inequalities $<$ and \leq must be understood in the component-wise sense.

The procedure we propose here is the projected gradient method (PGM) for vector optimization defined by Graña Drummond and Iusem [10] and analyzed by Fukuda and Graña Drummond in [7–9]. In those works, PGM share two essential features: (a) they are all descent methods: the objective value of the vectorial function F decreases at each iteration in the partial order induced by the underlying cones; and (b) full convergence of the sequences generated by the proposed methods is established under reasonable assumptions.

In the present work, we combine the classical PGM for vector optimization with two ingredients from the scalar case: (1) instead of a fixed parameter we propose to associate a variable steplength to compute the search direction; and (2) the classical Armijo line search is replaced by a nonmonotone strategy developed by Zhang and Hager [26].

Projected gradient method for different choices of the stepsize has been extensible studied, in the scalar case, see for example [1,3,4,11,16,23] and references therein. From the practical point of view, the spectral choice of the steplength, introduced by Barzilai and Borwein [1], and later analyzed by Raydan [21] and used in [3], requires little computational work that makes it especially for large-scale problems. From the theoretical point of view, it has been proved that the scheme greatly speeds up the convergence of methods based on gradient ideas. As it is mentioned in [4] it is natural to combine the spectral gradient idea with nonmonotone line search.

Nonmonotone line search techniques have been extensible studied in the scalar case with great success, see for example [6,11–13,18,22,24–26]. In the present work, we propose to use the nonmonotone technique defined in [26] based on the average of successive functional values. We prove that, without convexity assumptions, accumulation points of sequences generated by the proposed algorithm are stationary for problem (1). A feasible point $\bar{x} \in C$ is stationary for F if, and only if,

$$JF(\bar{x})(C - \{\bar{x}\}) \cap [-\mathbb{R}_{++}^n] = \emptyset$$

where $JF(\bar{x})(C - \{\bar{x}\}) := \{JF(\bar{x})(x - \bar{x}) : x \in C\}$. This is a necessary condition for Pareto optimality, [14]. Clearly, if $\bar{x} \in C$ is stationary for F then for all $v \in C - \{\bar{x}\}$, $JF(\bar{x})v \not\prec 0$.

In [5] it is proved that the steepest descent method for smooth (scalar) convex minimization, with stepsize obtained using the Armijo rule implemented by a backtracking procedure, is globally convergent to a solution (essentially, under the sole assumption of existence of optima). Using a similar technique, we extend this result to the multi-objective setting, that is to say, we show that any sequence produced by the proposed algorithm converges to a weakly Pareto optimal point of problem (1), no matter how poor the initial guess might be. Full convergence of PGM for MOP was established in [7–10] under the monotone Armijo line search.

Up to our knowledge, this is the first time that a nonmonotone line search based of [26] is used to analyze the global convergence of a PGM for MOP. In [17] the authors introduce nonmonotone line searches for MOP in the steepest descent methods. In [20] the authors propose two nonmonotone gradient algorithms for vector optimization with

a convex objective function based on the nonmonotone line search given by Grippo et al. [11] in the scalar context.

This paper is organized as follows. In Sect. 2 we present the nonmonotone projected gradient method for MOP. In Sect. 3 it is proved that accumulation points of sequences generated by the proposed algorithm are stationary for (1). In Sect. 4 the convex case is analyzed. Conclusions and lines for future research are given in Sect. 5.

Notation We denote: $\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \geq 0\}$, $\mathbb{R}_{++} = \{t \in \mathbb{R} \mid t > 0\}$, $\mathbb{N} = \{0, 1, 2, \dots\}$, $\|\cdot\|$ an arbitrary vector norm. If x and y are two vectors in \mathbb{R}^n , we write $x \leq y$ if $x_i \leq y_i, i = 1, \dots, n$, and $x < y$ if $x_i < y_i, i = 1, \dots, n$ where v_i is the i -th component of the vector v . If $F : \mathbb{R}^n \rightarrow \mathbb{R}^m, F = (f_1, \dots, f_m), JF(x)$ stands for the Jacobian matrix of F at $x: JF(x)$ is matrix in $\mathbb{R}^{m \times n}$ with entries $(JF(x))_{ij} = \frac{\partial f_i(x)}{\partial x_j}$. If $K = \{k_0, k_1, k_2, \dots\} \subset \mathbb{N} (k_{j+1} > k_j)$ for all j , we denote $\lim_{k \in K} x^k = \lim_{j \rightarrow \infty} x^{k_j}$. The canonical inner product is written as $\langle \cdot, \cdot \rangle$, and the transpose sign is denoted by T .

2 The nonmonotone projected gradient method for MOP

The projected gradient method we consider in the present work uses the projected gradient direction $v_\beta(x)$, proposed in [10] and analyzed in [7], that is defined, for a given point $x \in C$, by

$$v_\beta(x) = \operatorname{argmin}_{v \in C - \{x\}} \beta \varphi_x(v) + \frac{\|v\|^2}{2} \tag{2}$$

where $\beta > 0$ is a parameter and

$$\varphi_x(v) = \max_{i=1, \dots, r} \{\nabla F_i(x)^T v\}. \tag{3}$$

We also consider the optimal value function $\theta_\beta : C \rightarrow \mathbb{R}$ given by

$$\theta_\beta(x) = \beta \varphi_x(v_\beta(x)) + \frac{\|v_\beta(x)\|^2}{2}. \tag{4}$$

See for example [9, 10]. It is well established that a point $x \in C$ is stationary for F if, and only if, $\theta_\beta(x) = 0$, [10].

2.1 The nonmonotone line search

It is well-known in the literature that nonmonotone schemes can improve the efficiency of descent methods in the scalar case [4, 6, 15, 18, 22, 24, 26].

The nonmonotone line search framework developed by Grippo et al. [11] is based on the maximum functional value on some previous iterates: given $\sigma \in (0, 1), M \geq 1$ the step α_k must satisfy

$$f(x^k + \alpha_k d^k) \leq M_k + \sigma \alpha_k \nabla f(x^k)^T d^k \tag{5}$$

where $M_k = \max_{0 \leq j \leq m(k)} f(x^{k-j})$ where $m(k) = \min\{m(k-1) + 1, M\}$.

As we have already mentioned in the Introduction, in [20] the authors propose two nonmonotone gradient algorithms based on the nonmonotone line search (5) but we have chosen Zhang and Hager's strategy as it is known to be more efficient, at least in the scalar case. In [26] the authors propose the following scheme: given $\sigma \in (0, 1)$, $C_0 = f(x^0)$, $Q_0 = 1$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $\eta_k \in [\eta_{\min}, \eta_{\max}]$ the step α_k must satisfy

$$f(x^k + \alpha_k d^k) \leq C_k + \sigma \alpha_k \nabla f(x^k)^T d^k$$

where $C_{k+1} = \frac{(\eta_k Q_k C_k + f(x^{k+1}))}{Q_{k+1}}$ and $Q_{k+1} = \eta_k Q_k + 1$.

As it is mentioned in [26], if $\eta_k = 0$ for each k , then the line search is the usual monotone Armijo line search. If $\eta_k = 1$ for each k , then $C_k = A_k$ where $A_k = \frac{1}{k+1} \sum_{i=0}^k f(x^i)$ is the average of the successive functional values. Thus, C_k can be seen as a combination of the functional values $f(x^0), \dots, f(x^k)$ using a parameter η_k that controls the degree of non-monotonicity.

In order to use a nonmonotone line search in the descent method for MOP we propose the following scheme: given $\sigma \in (0, 1)$, $C_0 = F(x^0)$, $Q_0 = 1$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $\eta_k \in [\eta_{\min}, \eta_{\max}]$ and v^k such that $JF(x^k)v^k < 0$ the step α_k must satisfy

$$F(x^k + \alpha_k v^k) \leq C^k + \sigma \alpha_k JF(x^k)v^k \quad (6)$$

where

$$C^{k+1} = \frac{(\eta_k Q_k C^k + F(x^{k+1}))}{Q_{k+1}} \text{ and } Q_{k+1} = \eta_k Q_k + 1. \quad (7)$$

Observe that we can write

$$C^{k+1} = \frac{(\eta_k Q_k + 1) C^k + F(x^{k+1}) - C^k}{Q_{k+1}} = C^k + \frac{F(x^{k+1}) - C^k}{Q_{k+1}}$$

and then

$$C^k - C^{k+1} = \frac{C^k - F(x^{k+1})}{Q_{k+1}}. \quad (8)$$

2.2 The nonmonotone multiobjective projected gradient algorithm

We can now define an extension of the classical PGM using the nonmonotone line search technique (6)–(7) for the constrained MOP (1).

Algorithm 2.1. Choose $\sigma \in (0, 1)$, $Q_0 = 1$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $\eta_0 \in [\eta_{\min}, \eta_{\max}]$, $0 < \beta_{\min} < \beta_{\max} < \infty$ and $\beta_0 \in [\beta_{\min}, \beta_{\max}]$. Let $x^0 \in C$ be an arbitrary initial point. Set $C^0 = F(x^0)$ and $k = 0$.

1. Compute the search direction.

$$v_{\beta_k}(x^k) := \operatorname{argmin}_{v \in C - \{x^k\}} \beta_k \varphi_{x^k}(v) + \frac{\|v\|^2}{2}$$

where

$$\varphi_{x^k}(v) = \max_{i=1, \dots, r} \{\nabla F_i(x^k)^T v\}.$$

2. Stopping criterion. Compute $\theta_{\beta_k}(x^k) = \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2$. If $\theta_{\beta_k}(x^k) = 0$, then stop.

3. Compute the steplength. Choose α_k as the largest $\alpha \in \left\{ \frac{1}{2^j} : j = 0, 1, 2, \dots \right\}$ such that

$$F(x^k + \alpha v_{\beta_k}(x^k)) \leq C^k + \sigma \alpha JF(x^k) v_{\beta_k}(x^k). \tag{9}$$

4. Set $x^{k+1} = x^k + \alpha_k v_{\beta_k}(x^k)$. Define β_{k+1} such that $\beta_{k+1} \in [\beta_{\min}, \beta_{\max}]$.

Choose $\eta_k \in [\eta_{\min}, \eta_{\max}]$ and set Q_{k+1} and C^{k+1} as in (7). Do $k = k + 1$, and go to Step 1.

Observe that by (7) and (9)

$$C^{k+1} = \frac{\eta_k Q_k C^k + F(x^{k+1})}{Q_{k+1}} \leq C^k + \frac{\sigma \alpha_k JF(x^k) v_{\beta_k}(x^k)}{Q_{k+1}} \leq C^k$$

which implies that $\{C^k\}$ is a nonincreasing sequence in \mathbb{R}^r .

The differences between Algorithm 2.1 and a classical extension of PGM for the constrained MOP (see for example Algorithm 4.2 in [9]) rely on: (a) the presence of a variable steplength β_k instead a fixed parameter to compute the search direction in Step 1; and (b) the nonmonotone line search in Step 3.

From the classical characterization of stationarity in terms of $v_\beta(\cdot)$ and $\theta_\beta(\cdot)$ [10], we have that x^k is a stationary point for F if Algorithm 2.1 stops with the stopping criterion in Step 2.

3 Convergence analysis

In this we suppose that Algorithm 2.1 does not have a finite termination and therefore it generates infinite sequences $\{x^k\}$, $\{v_{\beta_k}(x^k)\}$ and $\{\alpha_k\}$.

The following Lemma follows from Lemma 1.1 in [26].

Lemma 1 *If $v_{\beta_k}(x^k) \in C - \{x^k\}$ and $JF(x^k)v_{\beta_k}(x^k) < 0$ for each k , then $F(x^k) \leq C^k \leq A^k$ for each k , where $A^k = \frac{1}{k+1} \sum_{i=0}^k F(x^i)$.*

Lemma 2 *Let $\{x^k\} \subset \mathbb{R}^n$ be a sequence generated by Algorithm 2.1. Then:*

- (a) $\{x^k\} \subset C$.
 (b) Algorithm 2.1 in well defined.

Proof (a) follows from Proposition 5 in [10].

(b) follows from Proposition 1 in [10] and the fact that $F(x^k) \leq C^k$ for each k . \square

Lemma 3 The scalar sequence $\{\|v_{\beta_k}(x)\|\}$ is bounded.

Proof By the definition of $v_{\beta_k}(x)$, we have that

$$\beta_k \varphi_x(v_{\beta_k}(x)) + \frac{\|v_{\beta_k}(x)\|^2}{2} = \theta_{\beta_k}(x) \leq 0$$

where the last inequality holds since $0 \in C - \{x\}$.

Then, by the Cauchy Schwartz inequality we have that, for each $i = 1, \dots, r$

$$-\beta_k \|\nabla F_i(x)\| \|v_{\beta_k}(x)\| + \frac{\|v_{\beta_k}(x)\|^2}{2} \leq 0.$$

Therefore, by the continuous differentiability of F , Lemma 2(a) and the boundedness of $\{\beta_k\}$ we obtain the desired result. \square

Proposition 1 Let $\{x^k\} \subset \mathbb{R}^n$ be a sequence generated by Algorithm 2.1 and $K \subset \mathbb{N}$, $\bar{\beta} > 0$, $\bar{x} \in C$ such that $\lim_{k \in K} x^k = \bar{x}$, $\lim_{k \in K} \beta_k = \bar{\beta}$. Then, there exists $\bar{K} \subset K$ such that

$$\liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k) = \theta_{\bar{\beta}}(\bar{x}). \quad (10)$$

Proof First, let us show that there exists $\bar{K} \subset K$ such that

$$\lim_{k \in \bar{K}} \theta_{\beta_k}(\bar{x}) = \theta_{\bar{\beta}}(\bar{x}). \quad (11)$$

Since $v_{\beta_k}(\bar{x}) \in C - \{\bar{x}\}$, from the definition of $\theta_{\bar{\beta}}(\bar{x})$ we have that

$$\theta_{\bar{\beta}}(\bar{x}) \leq \bar{\beta} \varphi_{\bar{x}}(v_{\beta_k}(\bar{x})) + \frac{1}{2} \|v_{\beta_k}(\bar{x})\|^2 = (\bar{\beta} - \beta_k) \varphi_{\bar{x}}(v_{\beta_k}(\bar{x})) + \theta_{\beta_k}(\bar{x}).$$

Then, by using Lemma 3 (the sequence $\{\|v_{\beta_k}(\bar{x})\|\}$ is bounded) and the continuity of the maximum function $u \mapsto \max\{u_1, \dots, u_r\}$, by taking \liminf in an appropriate subsequence $\bar{K} \subset K$ when $k \in \bar{K}$ goes to infinity we obtain that

$$\theta_{\bar{\beta}}(\bar{x}) \leq \liminf_{k \in \bar{K}} \theta_{\beta_k}(\bar{x}). \quad (12)$$

Now, from the definition of $\theta_{\beta_k}(\bar{x})$, since $v_{\beta_k}(\bar{x}) \in C - \{\bar{x}\}$ we have that $\theta_{\beta_k}(\bar{x}) \leq \beta_k \varphi_{\bar{x}}(v_{\beta_k}(\bar{x})) + \frac{1}{2} \|v_{\beta_k}(\bar{x})\|^2$. Then, by taking \limsup when $k \in \bar{K}$ goes to infinity

we obtain that $\limsup_{k \in \bar{K}} \theta_{\beta_k}(\bar{x}) \leq \theta_{\bar{\beta}}(\bar{x})$ which, together with the inequality (12) proves (11).

Secondly, since $v_{\bar{\beta}}(\bar{x}) \in C - \{\bar{x}\}$ we have that $v_{\bar{\beta}}(\bar{x}) + \bar{x} - x^k \in C - \{x^k\}$ and from the definition of $\theta_{\beta_k}(x^k)$

$$\theta_{\beta_k}(x^k) \leq \beta_k \varphi_{x^k}(v_{\bar{\beta}}(\bar{x}) + \bar{x} - x^k) + \frac{1}{2} \|v_{\bar{\beta}}(\bar{x}) + \bar{x} - x^k\|^2.$$

Hence, since $\varphi_x(u + v) \leq \varphi_x(u) + \varphi_x(v)$ for all $u, v \in \mathbb{R}^n$ we obtain

$$\begin{aligned} \theta_{\beta_k}(x^k) &\leq \beta_k \varphi_{x^k}(v_{\bar{\beta}}(\bar{x})) + \beta_k \varphi_{x^k}(\bar{x} - x^k) \\ &\quad + \frac{1}{2} \|v_{\bar{\beta}}(\bar{x})\|^2 + \frac{1}{2} \|\bar{x} - x^k\|^2 + \langle v_{\bar{\beta}}(\bar{x}), \bar{x} - x^k \rangle. \end{aligned}$$

By the continuity of the maximum function $u \mapsto \max \{u_1, \dots, u_r\}$ and the continuously differentiability of F , taking \limsup for $k \in \bar{K}$ on the above inequality we have

$$\limsup_{k \in \bar{K}} \theta_{\beta_k}(x^k) \leq \bar{\beta} \varphi_{\bar{x}}(v_{\bar{\beta}}(\bar{x})) + \frac{1}{2} \|v_{\bar{\beta}}(\bar{x})\|^2 = \theta_{\bar{\beta}}(\bar{x}). \tag{13}$$

Once again, since $v_{\beta_k}(x^k) \in C - \{x^k\}$ we have that $v_{\beta_k}(x^k) + x^k - \bar{x} \in C - \{\bar{x}\}$ and

$$\begin{aligned} \theta_{\beta_k}(\bar{x}) &\leq \beta_k \varphi_{\bar{x}}(v_{\beta_k}(x^k) + x^k - \bar{x}) + \frac{1}{2} \|v_{\beta_k}(x^k) + x^k - \bar{x}\|^2 \\ &\leq \beta_k \varphi_{\bar{x}}(v_{\beta_k}(x^k)) + \beta_k \varphi_{\bar{x}}(x^k - \bar{x}) \\ &\quad + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 + \frac{1}{2} \|x^k - \bar{x}\|^2 + \langle v_{\beta_k}(x^k), x^k - \bar{x} \rangle. \end{aligned}$$

So, by Lemma 3 and the continuity of the maximum function, taking \liminf for $k \in \bar{K}$ on both sides of the above inequality, we get, from (11),

$$\begin{aligned} \theta_{\bar{\beta}}(\bar{x}) &\leq \liminf_{k \in \bar{K}} \left[\beta_k \varphi_{\bar{x}}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 + \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) - \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) \right] \\ &= \liminf_{k \in \bar{K}} \left[\theta_{\beta_k}(x^k) + \beta_k \left(\varphi_{\bar{x}}(v_{\beta_k}(x^k)) - \varphi_{x^k}(v_{\beta_k}(x^k)) \right) \right]. \end{aligned}$$

Now, using that the maximum function $u \mapsto \max \{u_1, \dots, u_r\}$ is Lipschitz continuous with constant 1:

$$|\varphi_x(v) - \varphi_y(w)| \leq \|JF(x)v - JF(y)w\|$$

for all $v, w \in \mathbb{R}^n$, we obtain that

$$\begin{aligned} \theta_{\bar{\beta}}(\bar{x}) &\leq \liminf_{k \in \bar{K}} [\theta_{\beta_k}(x^k) + \beta_k \|JF(\bar{x}) - JF(x^k)\| \|v_{\beta_k}(x^k)\|] \\ &\leq \liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k) \end{aligned} \tag{14}$$

where we use Lemma 3 and the continuous differentiability of F in the last inequality. Thus, by combining (13) with (14) we obtain that

$$\limsup_{k \in \bar{K}} \theta_{\beta_k}(x^k) \leq \theta_{\bar{\beta}}(\bar{x}) \leq \liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k)$$

which implies (10). □

The following theorem is the main result of this section: it establishes stationarity of accumulation points of the sequences generated by Algorithm 2.1.

Theorem 1 *Assume that F is bounded from below. Let $\{x^k\} \subset \mathbb{R}^n$ be a sequence generated by Algorithm 2.1. Then, every accumulation point of $\{x^k\}$ is a feasible stationary point of (1).*

Proof Let \bar{x} be an accumulation point of the sequence $\{x^k\}$. Then, there exists $K \subset \mathbb{N}$ such that $\lim_{k \in K} x^k = \bar{x}$. The feasibility of \bar{x} follows combining the fact that C is closed with Lemma 2. The sequence $\{\beta_k\}_{k \in K}$ is bounded, then there exists $K_0 \subset K$ and $\bar{\beta} > 0$ such that $\limsup_{k \in K_0} \beta_k = \bar{\beta}$.

Considering that $\alpha_k \in (0, 1]$ for all $k \in K_0$, we have the following two possibilities: (a) $\limsup_{k \in K_0} \alpha_k > 0$, or (b) $\limsup_{k \in K_0} \alpha_k = 0$.

First, assume that (a) holds. Then, there exist $K_1 \subset K_0$ and $\bar{\alpha} > 0$ such that $\limsup_{k \in K_1} \alpha_k = \bar{\alpha}$. Since $\bar{\alpha} > 0$ there exists $K_2, K_2 \subset K_1$ such that $\alpha_k \geq \epsilon > 0$ for all $k \in K_2$. Therefore, for all $k \in K_2$, by (9) and the fact that $JF(x^k)v_{\beta_k}(x^k) < 0$, we have that

$$F(x^{k+1}) \leq C^k - \sigma\alpha_k(-JF(x^k)v_{\beta_k}(x^k)) \leq C^k - \sigma\epsilon(-JF(x^k)v_{\beta_k}(x^k)).$$

Then

$$\sigma\epsilon(-JF(x^k)v_{\beta_k}(x^k)) \leq C^k - F(x^{k+1}). \tag{15}$$

Now, using (8) and (15) we obtain

$$C^k - C^{k+1} \geq \frac{\sigma\epsilon(-JF(x^k)v_{\beta_k}(x^k))}{Q_{k+1}}. \tag{16}$$

Then, since $\sum_{k=0}^m C^k - C^{k+1} = C^0 - C^{m+1}$ and F is bounded from below we have, by Lemma 1, that $R \leq F(x^k) \leq C^k$, for all k , and we can conclude that C^k is bounded from below. It follows from (16) that

$$\sum_{k=0}^{\infty} \frac{\sigma\epsilon(-JF(x^k)v_{\beta_k}(x^k))}{Q_{k+1}} < \infty. \tag{17}$$

If we assume that $\liminf_{k \in K_2} (-JF(x^k)v_{\beta_k}(x^k)) \neq 0$, then $\liminf_{k \in K_2} (-JF(x^k)v_{\beta_k}(x^k)) \geq c > 0$. Therefore, using the second equality in (7)

$$\sum_{k=0}^{\infty} \frac{(-JF(x^k)v_{\beta_k}(x^k))}{Q_{k+1}} \geq \sum_{k=0}^{\infty} \frac{c}{k+2},$$

in contradiction with (17).

Hence $\liminf_{k \in K_2} (-JF(x^k)v_{\beta_k}(x^k)) = 0$, which, component-wise, can be written as

$$\liminf_{k \in K_2} (-\nabla F_i(x^k)^T v_{\beta_k}(x^k)) = 0, \text{ for all } i = 1, \dots, r. \tag{18}$$

Therefore, using (18) and (10) there exists $\bar{K} \subset K_2$ such that

$$\begin{aligned} 0 &= \liminf_{k \in \bar{K}} \beta_k \nabla F_i(x^k)^T v_{\beta_k}(x^k) \leq \liminf_{k \in \bar{K}} \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 \\ &= \liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k) = \theta_{\bar{\beta}}(\bar{x}) \end{aligned}$$

and $\theta_{\bar{\beta}}(\bar{x}) = 0$. Thus, since $\bar{\beta}$ is fixed, in view of what was mentioned immediately after (4) we conclude that \bar{x} is a stationary point of (1).

Now assume that (b) holds. Due to Lemma 3, there exist $K_1 \subset K_0$ and $\bar{v} \in \mathbb{R}^n$ such that $\lim_{k \in K_1} v_{\beta_k}(x^k) = \bar{v}$ and $\lim_{k \in K_1} \alpha_k = 0$. Note that we have $\max_{i=1, \dots, r} \{\beta_k \nabla F_i(x^k)^T v_{\beta_k}(x^k)\} \leq \theta_{\beta_k}(x^k) < 0$ so, letting $k \in K_1$ go to infinity we get

$$\max_{i=1, \dots, r} \{\bar{\beta} \nabla F_i(\bar{x})^T \bar{v}\} \leq \theta_{\bar{\beta}}(\bar{x}) \leq 0. \tag{19}$$

Take now a fixed but arbitrary positive integer q . Since $\alpha_k \rightarrow 0$, for $k \in K_1$ large enough, we have $\alpha_k < \frac{1}{2^q}$, which means that the nonmonotone condition in Step 3 of Algorithm 2.1 is not satisfied for $\alpha = \frac{1}{2^q}$ at x^k :

$$F(x^k + \left(\frac{1}{2}\right)^q v_{\beta_k}(x^k)) \not\leq C^k + \sigma \left(\frac{1}{2}\right)^q JF(x^k)v_{\beta_k}(x^k).$$

So for each $k \in K_1$ there exists $i = i(k) \in \{1, \dots, r\}$ such that

$$F_i(x^k + \left(\frac{1}{2}\right)^q v_{\beta_k}(x^k)) \geq C_i^k + \sigma \left(\frac{1}{2}\right)^q \nabla F_i(x^k)^T v_{\beta_k}(x^k)$$

where $C_i^k = \frac{(\eta_{k-1} Q_{k-1} C_i^{k-1} + F_i(x^k))}{Q_k}$. Since $\{i(k)\}_{k \in K_1} \subset \{1, \dots, r\}$, there exist $K_2 \subset K_1$ and an index i_0 such that $i_0 = i(k)$ for all $k \in K_2$,

$$F_{i_0}(x^k + \left(\frac{1}{2}\right)^q v_{\beta_k}(x^k)) \geq C_{i_0}^k + \sigma \left(\frac{1}{2}\right)^q \nabla F_{i_0}(x^k)^T v_{\beta_k}(x^k)$$

and since, by Lemma 1, $F(x^k) \leq C^k$ we have

$$F_{i_0}(x^k + \left(\frac{1}{2}\right)^q v_{\beta_k}(x^k)) \geq F_{i_0}(x^k) + \sigma \left(\frac{1}{2}\right)^q \nabla F_{i_0}(x^k)^T v_{\beta_k}(x^k).$$

Taking the limit when $k \in K_2$ goes to infinity in the above inequality, we obtain $F_{i_0}(\bar{x} + (1/2)^q \bar{v}) \geq F_{i_0}(\bar{x}) + \sigma (1/2)^q \nabla F_{i_0}(\bar{x})^T \bar{v}$. Since this inequality holds for any positive integer q and for i_0 (depending on q), by Proposition 1 in [10] it follows that $JF(\bar{x})\bar{v} \neq 0$, then $\max_{i=1, \dots, r} \{\nabla F_i(\bar{x})^T \bar{v}\} \geq 0$, which, together with (19) implies $\theta_{\bar{\beta}}(\bar{x}) = 0$.

Therefore, we conclude that \bar{x} is a stationary point of (1). □

4 The convex case

The objective of this section is to prove, under convexity assumptions, that in the case in which Algorithm 2.1 does not have a finite termination, the sequence $\{x^k\}$ converges to a weak Pareto optimum point.

We say that the mapping $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ is \mathbb{R}^r_+ -convex if

$$F(\lambda x + (1 - \lambda)z) \leq \lambda F(x) + (1 - \lambda)F(z)$$

for all $x, z \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$. Clearly, $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ is \mathbb{R}^r_+ -convex if and only if its components $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are all convex.

The proof of the full convergence to a weak Pareto optimum point follows from the analysis made in [7]. From the theory of convex analysis and optimization we know that, from (2), there exists $w(x) \in \mathbb{R}^r$ with

$$w(x) \geq 0, \quad \sum_{i=1}^r w_i(x) = 1 \tag{20}$$

such that $v_{\beta}(x) = P_C(x - \beta JF(x)^T w(x)) - x$, where $P_C(z)$ is the orthogonal projection of z on C , satisfying

$$\langle \beta JF(x)^T w(x) + v_{\beta}(x), v - v_{\beta}(x) \rangle \geq 0 \tag{21}$$

for all $v \in C - \{x\}$. See [2,8,9].

We will use the following technical result in order to prove the full convergence of the sequence $\{x^k\}$ generated by Algorithm 2.1.

Lemma 4 *Suppose that F is \mathbb{R}^r_+ -convex and let $\{x^k\}$ be an infinite sequence generated by Algorithm 2.1. Let $w^k := w(x^k) \in \mathbb{R}^r$ such that (20) holds and $v^k := v_{\beta_k}(x^k) = P_C(x^k - \beta_k JF(x^k)^T w^k) - x^k$. If for $x^* \in C$ and $k \geq 0$, we have $F(x^*) \leq C^k$, then*

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + 2\beta_{\max} \alpha_k |\langle w^k, JF(x^k)v^k \rangle| \\ &\quad + 2\beta_{\max} \alpha_k \langle w^k, C^k - F(x^k) \rangle. \end{aligned} \tag{22}$$

Proof Since $x^{k+1} = x^k + \alpha_k v^k$, we have

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - x^* + \alpha_k v^k\|^2 \\ &= \|x^k - x^*\|^2 + \alpha_k^2 \|v^k\|^2 - 2\alpha_k \langle v^k, x^* - x^k \rangle. \end{aligned} \tag{23}$$

By using (21) with $x = x^k$ and $\beta = \beta_k$ we get

$$\langle \beta_k JF(x^k)^T w^k + v^k, v - v^k \rangle \geq 0 \text{ for all } v \in C - \{x^k\}.$$

Taking $v = x^* - x^k \in C - \{x^k\}$ on the above inequality, we obtain

$$-\langle v^k, x^* - x^k \rangle \leq \beta_k \langle w^k, JF(x^k)(x^* - x^k) \rangle - \beta_k \langle w^k, JF(x^k)v^k \rangle - \|v^k\|^2. \tag{24}$$

By the convexity of F we know that $JF(x^k)(x^* - x^k) \leq F(x^*) - F(x^k)$. Then, this fact, together with $w^k \geq 0$ and $F(x^*) \leq C^k$, imply

$$\langle w^k, JF(x^k)(x^* - x^k) \rangle \leq \langle w^k, F(x^*) - F(x^k) \rangle \leq \langle w^k, C^k - F(x^k) \rangle.$$

Now, since x^k is a nonstationary point, we have that $JF(x^k)v^k < 0$. Then, $\langle w^k, JF(x^k)v^k \rangle < 0$ and from (24) it follows that

$$\begin{aligned} -\langle v^k, x^* - x^k \rangle &\leq \beta_k \langle w^k, C^k - F(x^k) \rangle + \beta_k |\langle w^k, JF(x^k)v^k \rangle| - \|v^k\|^2 \\ &\leq \beta_{\max} \langle w^k, C^k - F(x^k) \rangle + \beta_{\max} |\langle w^k, JF(x^k)v^k \rangle| - \|v^k\|^2. \end{aligned}$$

Thus, from (23)

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + \alpha_k^2 \|v^k\|^2 + 2\alpha_k \beta_k \langle w^k, JF(x^k)(x^* - x^k) \rangle \\ &\quad + 2\alpha_k \beta_k |\langle w^k, JF(x^k)v^k \rangle| - 2\alpha_k \|v^k\|^2 \leq \|x^k - x^*\|^2 \\ &\quad + 2\alpha_k \beta_{\max} \langle w^k, C^k - F(x^k) \rangle + 2\alpha_k \beta_{\max} |\langle w^k, JF(x^k)v^k \rangle|. \end{aligned}$$

And the result is obtained because $\alpha_k^2 - 2\alpha_k \leq 0$ since $0 < \alpha_k \leq 1$. □

The following theorem is established in order to prove the full convergence of the sequence regardless of the initial feasible iterate.

Theorem 2 Assume that $F : \mathbb{R}^n \rightarrow \mathbb{R}^f$ is \mathbb{R}_+^f -convex. Let us suppose that $\{x^k\}$ is a sequence generated by Algorithm 2.1 by considering $\eta_{\max} < 1$ and let us define

$$L := \{x \in C : F(x) \leq C^k \text{ for all } k\}. \tag{25}$$

Assume that $L \neq \emptyset$. Then, the sequence $\{x^k\}$ is quasi-Féjer convergent to L , which means that, for every $x^* \in L$ there exists a sequence $\{\epsilon_k\} \subset \mathbb{R}$, $\epsilon_k \geq 0$ for all k such that $\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + \epsilon_k$ with $\sum_{k=0}^\infty \epsilon_k < \infty$.

Proof By the hypothesis there is $x^* \in L$. Since F is \mathbb{R}^r_+ -convex it follows from Lemma 4 that, for all k

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + 2\beta_{\max}\alpha_k|\langle w^k, JF(x^k)v^k \rangle| \\ &\quad + 2\beta_{\max}\alpha_k\langle w^k, C^k - F(x^k) \rangle. \end{aligned} \tag{26}$$

By considering the canonical basis of \mathbb{R}^r $\{e_1, \dots, e_r\}$, we can write $w^k = \sum_{i=1}^r w_i^k e_i$ for each k . Then, from (26) and the facts that $0 \leq w_i^k \leq 1$ for all i and k and $C^k \geq F(x^k)$ we have that, for all k ,

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 \\ &\quad + 2\beta_{\max}\alpha_k \sum_{i=1}^r \left(|\langle e_i, JF(x^k)v^k \rangle| + \langle e_i, C^k - F(x^k) \rangle \right). \end{aligned}$$

Defining

$$\epsilon_k := 2\beta_{\max}\alpha_k \sum_{i=1}^r \left(|\langle e_i, JF(x^k)v^k \rangle| + \langle e_i, C^k - F(x^k) \rangle \right)$$

we have that $\epsilon_k \geq 0$ and $\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + \epsilon_k$. Thus, let us prove that $\sum_{k=0}^\infty \epsilon_k < \infty$. To do that, let us define $\epsilon_k^1 = 2\beta_{\max}\alpha_k \sum_{i=1}^r |\langle e_i, JF(x^k)v^k \rangle|$ and $\epsilon_k^2 = 2\beta_{\max}\alpha_k \sum_{i=1}^r \langle e_i, C^k - F(x^k) \rangle$. We are aware that ϵ_k^1 depends on β_k but our intention is to demonstrate that ϵ_k^1 can be independently bounded from the steplength β_k and that is the reason why we prefer to write it in that way.

From one side, from the line search condition in Step 3, we obtain $-\alpha_k \langle e_i, JF(x^k)v^k \rangle \leq \frac{1}{\sigma}(C_i^k - F_i(x^{k+1}))$. Since x^k is nonstationary, we also have for all i and k that $\langle e_i, JF(x^k)v^k \rangle = \nabla F_i(x^k)^T v^k < 0$, which means that $-\alpha_k \langle e_i, JF(x^k)v^k \rangle = \alpha_k |\langle e_i, JF(x^k)v^k \rangle|$. Hence, by the definition of ϵ_k^1 , $\epsilon_k^1 \leq \frac{2\beta_{\max}}{\sigma} \sum_{i=1}^r (C_i^k - F_i(x^{k+1}))$ and, by (8) $\epsilon_k^1 \leq \frac{2\beta_{\max}}{\sigma} \sum_{i=1}^r Q_{k+1}(C_i^k - C_i^{k+1})$. Since

$$Q_{k+1} = 1 + \sum_{i=0}^k \prod_{m=0}^i \eta_{k-m} \leq 1 + \sum_{i=0}^k \eta_{\max}^{i+1} \leq \sum_{i=0}^\infty \eta_{\max}^i \leq \frac{1}{1 - \eta_{\max}}$$

we obtain that

$$0 \leq \epsilon_k^1 \leq \frac{2\beta_{\max}}{\sigma(1 - \eta_{\max})} \sum_{i=1}^r (C_i^k - C_i^{k+1}). \tag{27}$$

On the other side, by (8), $C_i^k = C_i^{k-1} + \frac{F_i(x^k) - C_i^{k-1}}{Q_k}$, for all $k \geq 1$ and we have that $C_i^k - F_i(x^k) = (1 - \frac{1}{Q_k})(C_i^{k-1} - F_i(x^k))$. Then, by using (8) again:

$$\begin{aligned} \epsilon_k^2 &= 2\beta_{\max}\alpha_k \frac{(Q_k - 1)}{Q_k} \sum_{i=1}^r (C_i^{k-1} - F_i(x^k)) \\ &= 2\beta_{\max}\alpha_k (Q_k - 1) \sum_{i=1}^r (C_i^{k-1} - C_i^k). \end{aligned}$$

Since $Q_k - 1 \leq \frac{\eta_{\max}}{1 - \eta_{\max}}$ and $\alpha_k \leq 1$ we get

$$\epsilon_k^2 \leq \frac{2\beta_{\max}\eta_{\max}}{(1 - \eta_{\max})} \sum_{i=1}^r (C_i^{k-1} - C_i^k). \tag{28}$$

Adding up from $k = 0$ to $k = N$ at the inequalities (27) and (28) where N is any positive integer and defining $C^{-1} = C^0$, we get

$$\sum_{k=0}^N \epsilon_k \leq \frac{2\beta_{\max}}{\sigma(1 - \eta_{\max})} \sum_{i=1}^r (C_i^0 - C_i^{N+1}) + \frac{2\beta_{\max}\eta_{\max}}{(1 - \eta_{\max})} \sum_{i=1}^r (C_i^0 - C_i^N).$$

Since N is an arbitrary positive integer, $\beta_{\max} > 0$, $\sigma > 0$, $\eta_{\max} < 1$ and $F(x^*) \leq C^k$ for all k , we conclude that $\sum_{k=0}^{\infty} \epsilon_k < \infty$ as we wanted to prove. \square

The previous theorem formulates conditions under which the sequence generated by the nonmonotone Algorithm 2.1 is a quasi-Féjer sequence. Some comments concerning the assumption $L \neq \emptyset$ are in order. When the Armijo line search is considered the assumption used is the following [8,9]:

Assumption Every \mathbb{R}_+^r -decreasing sequence $\{y^k\} \subset F(C) := \{F(x) : x \in C\}$ is \mathbb{R}_+^r -bounded from below by an element of $F(C)$, i.e., for any $\{y^k\}$ contained in $F(C)$ with $y^{k+1} < y^k$ for all k , there exists $\bar{x} \in C$ such that $F(\bar{x}) \leq y^k$ for all k .

It is well-known that in the classical unconstrained (convex) optimization case, this condition, known as \mathbb{R}_+^r -completeness, is equivalent to existence of solutions of the optimization problem and it is standard for ensuring existence of efficient points for vector optimization problems.

Thus, in the monotone case [8,9], since the sequence $\{F(x^k)\}$ is \mathbb{R}_+^r -decreasing, by the above assumption, there exists a point $\hat{x} \in \{x \in C : F(x) \leq F(x^k)\}$ and an analogous of Lemma 4 is proved with the inequality $\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\beta_{\max}\alpha_k |\langle w^k, JF(x^k)v^k \rangle|$, instead of (22).

In the nonmonotone case, we consider the set L defined by (25) by considering C^k instead of $F(x^k)$ since $\{C^k\}$ is a \mathbb{R}_+^r -decreasing sequence. This fact is the responsible for the appearance of the last term in (22) and for ϵ_k^2 in the proof of Theorem 2.

Observe finally that the assumption $L \neq \emptyset$ is verified when F is a linear mapping or being nonlinear $F(C)$ is a convex set or even when the original problem has a solution.

Theorem 3 Assume that $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ is \mathbb{R}_+^r -convex and that the hypotheses of the previous theorem hold. Then, any sequence $\{x^k\}$ generated by Algorithm 2.1 converges to a weak Pareto optimum point.

Proof By Theorem 2, since $\{x^k\}$ is a quasi-Féjer sequence convergent to L , it follows, from Theorem 1 in [5], that $\{x^k\}$ is bounded. So, since C is closed, $\{x^k\} \subset C$ has at least one feasible limit point. Thus, there exist $x^* \in C$ and $K \subset \mathbb{N}$ such that $\lim_{k \in K} x^k = x^*$. By Theorem 1, x^* is stationary for problem (1). Using the \mathbb{R}_+^r -convexity of F , from Lemma 5.2 in [7], it follows that $x^* \in C$ is a weak Pareto optimum solution of (1).

Let us prove now that $x^* \in L$. Since $C^{k+1} \leq C^k$ for all k , then, for each k_0 fixed we have that $C^k \leq C^{k_0}$ for all $k > k_0$. Then, $F(x^k) \leq C^{k_0}$ for all $k > k_0$. Thus, by taking limit for $k \in K$ and using the continuity of F we obtain that $F(x^*) \leq C^{k_0}$. Then, $F(x^*) \leq C^k$ for all $k \in K$. Thus $x^* \in L$. Then, by Theorem 1 in [5], since $\{x^k\}$ has a cluster limit point $x^* \in L$ we conclude that $\{x^k\}$ converges to $x^* \in C$ as we wanted to prove. \square

5 Conclusions

We have presented a new algorithm for multiobjective optimization with convex constraints. At each iteration the search direction was computed by considering a variable steplength instead of a fixed parameter. The novelty feature has been the use of a nonmonotone line search technique instead of the classical Armijo strategy.

Stationarity of the accumulation points of the sequences generated by the proposed algorithm has been established in the general case, and, under standard convexity assumptions, full convergence to weak Pareto points of any sequence has been proved.

Future work will be focused on the study of a numerical implementation of Algorithm 2.1 by incorporating a variable steplength β_k which may be updated using the ideas underlying the spectral parameter from the scalar case [1,3,21]. The use of the nonmonotone PGM as a method to solve the subproblem inside of an Augmented Lagrangian type method to solve multiobjective problems with general constraints is also left as a topic for a future work.

Acknowledgements The authors are indebted to the anonymous referees whose comments helped a lot to improve the quality of the paper.

References

1. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. *IMA J. Numer. Anal.* **8**, 141–148 (1988)
2. Bertsekas, D.P.: *Convex Analysis and Optimization*. Athena Scientific, Belmont, MA (2003)
3. Birgin, E.G., Martínez, J.M., Raydan, M.: Inexact spectral projected gradient methods on convex sets. *IMA J. Numer. Anal.* **23**(4), 539–559 (2003)
4. Birgin, E.G., Martínez, J.M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.* **10**(4), 1196–12119 (1999)
5. Burachik, R., Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F.: Full convergence of the steepest descent method with inexact line searches. *Optimization* **32**(2), 137–146 (1995)
6. Dai, Y.H.: On the nonmonotone line search. *J. Optim. Theory Appl.* **112**(2), 315–330 (2002)
7. Fukuda, E.H., Graña Drummond, L.M.: On the convergence of the projected gradient method for vector optimization. *Optimization* **60**, 1009–1021 (2011)

8. Fukuda, E.H., Graña Drummond, L.M.: Inexact projected gradient method for vector optimization. *Comput. Optim. Appl.* **54**, 473–493 (2013)
9. Fukuda, E.H., Graña Drummond, L.M.: A survey on multiobjective descent methods. *Pesqui. Oper.* **34**, 585–620 (2014)
10. Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. *Comput. Optim. Appl.* **28**, 5–29 (2004)
11. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for newton's method. *SIAM J. Numer. Anal.* **23**, 707–716 (1986)
12. Grippo, L., Lampariello, F., Lucidi, S.: A class on nonmonotone stabilization methods in unconstrained optimization. *Numer. Math.* **59**(8), 779–806 (1991)
13. Hager, W.W., Zhang, H.: Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. *Assoc. Comput. Mach. ACM Trans. Math. Softw.* **32**(1), 113–137 (2006)
14. Luc, D.: *Theory of Vector Optimization*. Springer, Berlin (1989)
15. Lucidi, S., Rochetich, F., Roma, M.: Curvilinear stabilization techniques for truncated newton methods in large-scale unconstrained optimization. *SIAM J. Opt.* **8**, 916–939 (1998)
16. McCormick, G.P., Tapia, R.A.: The gradient projection method under mild differentiability conditions. *SIAM J. Control Optim.* **10**, 93–98 (1972)
17. Mita, K., Fukuda, E.H., Yamashita, N.: On using nonmonotone line search techniques in steepest descent methods for multiobjective optimization. In: *Proceedings of the 61st Annual Conference of the Institute of Systems, Control and Information Engineers (ISCIE), Kyoto* (2017)
18. Panier, E.R., Tits, A.L.: Avoiding the maratos effect by means of a nonmonotone line search. I: general constrained problems. *SIAM J. Numer. Anal.* **28**(4), 1183–1195 (1991)
19. Pareto, V.: *Manual D'economie Politique*. F. Rouge, Lausanne (1896)
20. Qu, S., Ji, Y., Jiang, J., Zhang, Q.: Nonmonotone gradient methods for vector optimization with a portfolio optimization application. *Eur. J. Oper. Res.* **263**(2), 356–366 (2017)
21. Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. *IMA J. Numer. Anal.* **13**, 321–326 (1993)
22. Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.* **7**(1), 26–33 (1997)
23. Schwartz, A., Polak, E.: Family of projected descent methods for optimization problems with simple bounds. *J. Optim. Theory Appl.* **92**, 1–31 (1997)
24. Toint, P.L.: An assessment of nonmonotone linesearch techniques for unconstrained optimization. *SIAM J. Sci. Comput.* **17**(3), 725–739 (1996)
25. Wan, Z., Feng, D.: Investigation on a nonmonotone cautious BFGS algorithm. *Math. Numer. Sin.* **33**(4), 387–396 (2011)
26. Zhang, H., Hager, W.W.: A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.* **14**(4), 1043–1056 (2004)