An Explicit Three-Term Polak–Ribière–Polyak Conjugate Gradient Method for Bicriteria Optimization

Y. Elboulqe[1] and M. El Maghri[1]

[1]*Department of Mathematics and Computer, Faculty of Sciences Aïn Chock, Hassan II University, Casablanca, BP. 5366, Morocco.*

March 22, 2023

Abstract. We propose in this paper a Polak–Ribière–Polyak conjugate gradient type method for solving bicriteria optimization problems by avoiding scalarization techniques. Two particular advantages in this contribution are to be noted. First, the suggested descent direction common to both criteria may be directly computed by a given formula without solving any intermediate subproblem. Second, the descent property proves not only sufficient but also independent of the line search. The global convergence of the resulting algorithm towards (Pareto) critical points is guaranteed under standard hypotheses, while simply using appropriate Armijo-like stepsizes. Finally, numerical experiments including comparisons with another method of the same type are reported.

Keywords. Multicriteria optimization, Pareto optimality, Pareto critical point, Descent directions, Line search, PRP conjugate gradient method.

1 Introduction

We are concerned with the following unconstrained bicriteria (biobjective) optimization problem:

$$(P) \quad \min_{x\in\mathbb{R}^n} \Big(F_1(x), F_2(x)\Big),$$

in which $F_1,\ F_2 : x \in \mathbb{R}^n \longmapsto \mathbb{R}$ are two continuously differentiable functions. This type of problems arises increasingly in various scientific fields as in economics, medicine, design, transportation and so on. However, in those problems, there is always a conflict between the objectives leading to the unlikely existence of a solution that minimizes all the criteria at the same time. Thus, *Pareto optimality* was introduced as being an *efficient* concept. Recall that a *Pareto optimum* also called an *efficient solution* is a decision variable for which if one objective is improved, then at least one of the others is degraded.

Over the past decades, well-known optimization methods have been directly extended to multicriteria problems. First introduced in [21], the idea was later developed in [10] and then in many other papers (see the references given therein). In this new generation, we no longer need any parametric transformations that could be sensitive to the original multicriteria problem. Briefly, with this new approach, each extended method is developed in much the same way as the corresponding scalar method, thus benefiting from the same tools while developing their own techniques. Our aim in this paper is to similarly develop a Polak–Ribiere–Polyak (PRP) conjugate gradient method to solve (P).

The fundamental Fletcher-Reeves (FR) conjugate gradient (CG) method [9] and all its variants belong to the class of first-order optimization algorithms and are well known for their performance for solving large-scale unconstrained optimization problems:

$$\min_{x\in\mathbb{R}^n} f(x)$$

such that $f:\mathbb{R}^n\to\mathbb{R}$ is continuously differentiable. Globally, a descent algorithm generates a sequence (x^k) following the standard iterative process:

$$x^{k+1}=x^k+t_kd^k,\quad k=0,1,2,\ldots,\tag{1}$$

where $x^0\in\mathbb{R}^n$ is an arbitrary starting guess and $t_k>0$ is a descent steplength computed by some line search procedure along a descent direction $d^k\in\mathbb{R}^n$. In a CG method, the global descent scheme is given by

$$d^k=-g^k+\beta_kd^{k-1},\quad k=1,2,\ldots,\quad d^0=-g^0,\tag{2}$$

where $g^k=\nabla f(x^k)$ is the gradient of f at x^k and $\beta_k\in\mathbb{R}$ is a parameter chosen to satisfy adequate properties. There have been proposed several choices for β_k leading to distinct methods, each one having its computational effectiveness and its convergence properties (see, e.g., [3, 16, 24, 25]).

In this paper, we are particularly interested in the modified Polak–Ribière–Polyak CG method (MPRP) [30] in which the parameter β_k is specified by

$$\beta_k^{PRP}=\frac{\langle g^k,y^{k-1}\rangle}{\|g^{k-1}\|^2},$$

where $\langle\cdot,\cdot\rangle$ stands for the Euclidean inner product, $\|\cdot\|$ its associated norm and $y^{k-1}=g^k-g^{k-1}$. Although the original PRP algorithm performs better than other CG algorithms, generally, it does not provide a descent direction once an Armijo-type line search is selected. This is why the MPRP scheme was introduced since its generates directions satisfying the following sufficient descent property regardless of the line search used:

$$\langle g^k,d^k\rangle\le -c\,\|g^k\|^2,\tag{3}$$

for some $c>0$. The key idea in [30] was essentially based on the fact that the direction schemes of some quasi-Newtonian methods can be viewed as a special three-term CG:

$$d^k=-g^k+\beta_k^{PRP}d^{k-1}-\theta_ky^{k-1},\ k\ge 1,$$

where

$$\theta_k=\frac{\langle g^k,d^{k-1}\rangle}{\|g^{k-1}\|^2}.$$

This approach has been the subject of intensive studies showing its effectiveness (see, e.g., [1, 4, 22, 29, 31]).

Apart from a few multicriteria extensions of some CG methods [13, 14, 23], so far, no study in this direction has been carried out for the three-term one. In fact, we propose in this paper a bicriteria MPRP method (in short, B-MPRP) for solving (P) without scalarization. Our descent scheme is explicitly given and satisfies a sufficient descent for both criteria independently of the line search. The convergence of the B-MPRP algorithm is proved under the standard assumptions while using appropriate Armijo-type stepsizes. Also, experimental results including comparisons with the PRP method in [23] are reported.

2 The B-MPRP method

Recall that a point $x^* \in \mathbb{R}^n$ is said to be *weak Pareto optimum* or *weakly efficient* for (P), if

$$\nexists x \in \mathbb{R}^n,\ F_i(x) < F_i(x^*), \quad i = 1, 2.$$

The *local weak efficiency* concept is similarly defined replacing above $\mathbb{R}^n$ by a neighbourhood $\mathcal{N}(x^*)$ of x^*.

It is well known that a necessary condition for a point $x^* \in \mathbb{R}^n$ to be locally weakly efficient for (P) is that

$$\nexists d \in \mathbb{R}^2,\ \langle g_i(x^*), d\rangle < 0, \quad i = 1, 2, \tag{4}$$

where $g_i(x^*) = \nabla F_i\,(x^*)$. By the alternative theorem, the last condition is equivalent to

$$\exists \lambda \in [0, 1]\,,\ \lambda\,[g_1\,(x^*) - g_2\,(x^*)] + g_2\,(x^*) = 0. \tag{5}$$

A point $x \in \mathbb{R}^n$ satisfying one of the above necessary efficiency conditions is called *Pareto critical* or *Pareto stationary.* Hence, if x is not a critical point, then there exists $d \in \mathbb{R}^2$ such that

$$\langle g_i(x), d\rangle < 0, \quad i = 1, 2, \tag{6}$$

which implies that the vector d is a *biobjective descent direction* at x, i.e.,

$$\exists t_d > 0,\ \forall t \in \left]0, t_d\right],\ F_i(x + td) < F_i(x), \quad i = 1, 2. \tag{7}$$

2.1 The B-MPRP descent scheme

The B-MPRP iterative process that we propose to solve (P) follows the scheme below:

$$x^{k+1} = x^k + t_k d^k, \quad k = 0, 1, \ldots, \tag{8}$$

where x^k is the current iterate, t_k is a stepsize (which will be defined in the next section) and d^k is the descent direction that we explicitly describe by

$$d^k = \begin{cases} -\,g^k, & \text{if } k = 0, \\ -\,g^k + \beta_k d^{k-1} - \theta_k y^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{9}$$

where

$$g^k = \lambda_k\big(g_1^k - g_2^k\big) + g_2^k, \quad g_i^k = g_i\big(x^k\big), \quad i = 1, 2, \tag{10}$$

$$\lambda_k = \begin{cases} 1, & \text{if} \quad \big(0 \leq a_k \leq -b_k\big) \\ & \text{or} \quad \big(a_k < 0 \ \text{ and } \ a_k \leq -2b_k\big), \\ 0, & \text{if} \quad \big(a_k \geq 0 \ \text{ and } \ b_k > 0\big) \\ & \text{or} \quad \big(a_k < 0 \ \text{ and } \ a_k > -2b_k\big), \\ -\dfrac{b_k}{a_k}, & \text{if} \quad 0 \leq -b_k < a_k, \end{cases} \tag{11}$$

$$a_k = \begin{cases} \left\|g_1^0 - g_2^0\right\|^2, & \text{if } k = 0, \\ \left\|g_1^k - g_2^k\right\|^2 + \dfrac{\left\|g_1^k - g_2^k\right\|^2 \left\langle g_2^k, d^{k-1}\right\rangle}{\left\|g^{k-1}\right\|^2} - \\ \dfrac{\left\langle g_1^k - g_2^k, g_2^k\right\rangle \left\langle g_1^k - g_2^k, d^{k-1}\right\rangle}{\left\|g^{k-1}\right\|^2}, & \text{if } k \geq 1, \end{cases} \tag{12}$$

$$b_k = \begin{cases} \left\langle g_1^0 - g_2^0, g_2^0 \right\rangle, & \text{if } k = 0, \\ \left\langle g_1^k - g_2^k, g_2^k \right\rangle + \dfrac{\left\langle g_1^k, g_2^k - g^{k-1} \right\rangle \left\langle g_2^k, d^{k-1} \right\rangle}{\|g^{k-1}\|^2} - \\ \quad \dfrac{\left\langle g_2^k, g_2^k - g^{k-1} \right\rangle \left\langle g_1^k, d^{k-1} \right\rangle}{\|g^{k-1}\|^2}, & \text{if } k \geq 1, \end{cases} \tag{13}$$

$$\beta_k = \frac{\left\langle g^k, y^{k-1} \right\rangle}{\|g^{k-1}\|^2}, \tag{14}$$

$$\theta_k = \frac{\left\langle g^k, d^{k-1} \right\rangle}{\|g^{k-1}\|^2}, \tag{15}$$

$$y^{k-1} = g^k - g^{k-1}. \tag{16}$$

Now, we shall prove the sufficient descent property of the direction d^k defined in (9). Let us denote by

$$g^k(\lambda) = \lambda\left(g_1^k - g_2^k\right) + g_2^k, \quad \lambda \in \mathbb{R}. \tag{17}$$

Proposition 2.1 *Let x^k be the k^{th} iterate of a sequence generated by the procedure (8)–(16).*

(i) If $g^k \neq 0$, then

$$\left\langle g^k(\lambda), d^k \right\rangle \leq \left\langle g^k, d^k \right\rangle = -\|g^k\|^2, \quad \forall \lambda \in [0,1], \tag{18}$$

or equivalently,

$$\left\langle g_i^k, d^k \right\rangle \leq \left\langle g^k, d^k \right\rangle = -\|g^k\|^2, \quad i = 1, 2. \tag{19}$$

(ii) If $g^k = 0$, then x^k is Pareto critical.

Proof (i) We first need to prove that λ_k given by (11) is, for each $k \in \mathbb{N}$, a (global) minimum of the following scalar optimization problem:

$$(P_k) \qquad \min_{\lambda \in [0,1]} f_k(\lambda),$$

where the sequence of functions $(f_k)_{k \in \mathbb{N}}$ is defined by

$$f_k(\lambda) = \frac{1}{2} a_k \lambda^2 + b_k \lambda + c_k, \tag{20}$$

a_k, b_k are given respectively by (12) and (13), and c_k is a real number which can be chosen arbitrarily. So, we will perform a case disjunction reasoning. If $a_k > 0$ then the problem is strictly convex and if $a_k < 0$ then it is strictly concave. In the first case, $\lambda_k \in \operatorname{argmin}(P_k)$ iff λ_k satisfies the KKT conditions.

$$\begin{cases} a_k \lambda + b_k - \gamma_1 + \gamma_2 = 0, \\ \gamma_1 \lambda = 0, \\ \gamma_2 (\lambda - 1) = 0, \\ \gamma_i \geq 0, \quad i = 1, 2, \\ \lambda \in [0, 1], \end{cases}$$

which obviously are satisfied, at each one of the following three cases, by a unique solution:

$$\begin{cases} \lambda_k = 1, & \gamma_1 = 0, \quad \gamma_2 = -a_k - b_k, \ \text{if } a_k \leq -b_k, \\ \lambda_k = 0, & \gamma_1 = b_k, \ \gamma_2 = 0, \qquad\qquad \text{if } b_k > 0, \\ \lambda_k = -\dfrac{b_k}{a_k}, & \gamma_1 = 0, \quad \gamma_2 = 0, \qquad\qquad \text{if } 0 \leq -b_k < a_k. \end{cases}$$

While in the second case, it is obvious that the global minimum of (P_k) is $\lambda_k = 1$ if $a_k \leq -2b_k$; otherwise $\lambda_k = 0$. When $a_k = 0$ and $b_k \neq 0$, the function f_k is affine, so that, its minimum value on $[0,1]$ is reached at 0 or 1 according to $b_k > 0$ or $b_k < 0$ respectively. Finally, if $a_k = b_k = 0$, any value in $[0,1]$ is a minimal solution, for instance, $\lambda_k = 1$. Thus, in all cases, we obtain (11).

Let us prove now the inequality in (18). We just see that $\lambda_k \in \operatorname{argmin} f_k([0,1])$. It follows by the minimum variational principle that, for each $k \in \mathbb{N}$,

$$f_k'(\lambda_k)(\lambda - \lambda_k) \geq 0, \quad \forall \lambda \in [0,1]. \tag{21}$$

Simple calculations show that the derivative of f_k at λ_k is given by

$$f_k'(\lambda_k) = -\left\langle g_1^k - g_2^k, d^k \right\rangle. \tag{22}$$

Indeed, for $k = 0$, we have

$$\begin{aligned} f_0'(\lambda_0) &= a_0\lambda_0 + b_0 \\ &= \left\| g_1^0 - g_2^0 \right\|^2 \lambda_0 + \left\langle g_1^0 - g_2^0, g_2^0 \right\rangle \\ &= \left\langle g_1^0 - g_2^0, \lambda_0 \left(g_1^0 - g_2^0 \right) \right\rangle + \left\langle g_1^0 - g_2^0, g_2^0 \right\rangle \\ &= \left\langle g_1^0 - g_2^0, \lambda_0 \left(g_1^0 - g_2^0 \right) + g_2^0 \right\rangle \\ &= -\left\langle g_1^0 - g_2^0, -g^0 \right\rangle \\ &= -\left\langle g_1^0 - g_2^0, d^0 \right\rangle. \end{aligned}$$

If $k \geq 1$, using the definitions of d^k, β_k, θ_k, g^k and y^{k-1}, one has

$$\begin{aligned} &-\left\langle g_1^k - g_2^k, d^k \right\rangle \\ &= \left\langle g_1^k - g_2^k, g^k - \beta_k d^{k-1} + \theta_k y^{k-1} \right\rangle \\ &= \left\langle g_1^k - g_2^k, g^k \right\rangle - \beta_k \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle + \theta_k \left\langle g_1^k - g_2^k, y^{k-1} \right\rangle \\ &= \left\langle g_1^k - g_2^k, g^k \right\rangle - \frac{\left\langle g^k, y^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle + \frac{\left\langle g^k, d^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left\langle g_1^k - g_2^k, y^{k-1} \right\rangle \\ &= \lambda_k \left\| g_1^k - g_2^k \right\|^2 + \left\langle g_1^k - g_2^k, g_2^k \right\rangle - \frac{\lambda_k \left\langle g_1^k - g_2^k, y^{k-1} \right\rangle + \left\langle g_2^k, y^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle \\ &\quad + \frac{\lambda_k \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle + \left\langle g_2^k, d^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left\langle g_1^k - g_2^k, y^{k-1} \right\rangle \\ &= \lambda_k \left\| g_1^k - g_2^k \right\|^2 + \left\langle g_1^k - g_2^k, g_2^k \right\rangle - \frac{\left\langle g_2^k, y^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle + \frac{\left\langle g_2^k, d^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left\langle g_1^k - g_2^k, y^{k-1} \right\rangle \\ &= \lambda_k \left\| g_1^k - g_2^k \right\|^2 + \left\langle g_1^k - g_2^k, g_2^k \right\rangle - \frac{\lambda_k \left\langle g_2^k, g_1^k - g_2^k \right\rangle + \left\langle g_2^k, g_2^k - g^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle \\ &\quad + \frac{\left\langle g_2^k, d^{k-1} \right\rangle}{\|g^{k-1}\|^2} \left(\lambda_k \left\| g_1^k - g_2^k \right\|^2 + \left\langle g_1^k - g_2^k, g_2^k - g^{k-1} \right\rangle \right) \end{aligned}$$

$$
\begin{aligned}
&= \lambda_k \left(\left\| g_1^k - g_2^k \right\|^2 - \frac{\left\langle g_1^k - g_2^k, g_2^k \right\rangle \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle}{\left\| g^{k-1} \right\|^2} + \frac{\left\langle g_2^k, d^{k-1} \right\rangle \left\| g_1^k - g_2^k \right\|^2}{\left\| g^{k-1} \right\|^2} \right) \\
&+ \left(\left\langle g_1^k - g_2^k, g_2^k \right\rangle - \frac{\left\langle g_2^k, g_2^k - g^{k-1} \right\rangle \left\langle g_1^k - g_2^k, d^{k-1} \right\rangle}{\left\| g^{k-1} \right\|^2} + \frac{\left\langle g_2^k, d^{k-1} \right\rangle \left\langle g_1^k - g_2^k, g_2^k - g^{k-1} \right\rangle}{\left\| g^{k-1} \right\|^2} \right) \\
&= a_k \lambda_k + b_k \\
&= f_k' \left(\lambda_k \right).
\end{aligned}
$$

Hence, by (21) and (22), for all $\lambda \in [0, 1]$, we have

$$\left\langle \lambda \left(g_1^k - g_2^k \right), d^k \right\rangle \leq \left\langle \lambda_k \left(g_1^k - g_2^k \right), d^k \right\rangle.$$

Adding $\left\langle g_2^k, d^k \right\rangle$ in both sides of the last inequality, we get

$$\left\langle g^k(\lambda), d^k \right\rangle \leq \left\langle g^k, d^k \right\rangle, \tag{23}$$

which well proves the inequality in (18).
To complete the proof of (18), it remains to show the equality part. The case $k = 0$ is trivial, while in the case $k \geq 1$, the equality follows immediately from the definitions of d^k, θ_k and β_k:

$$
\begin{aligned}
&\left\langle g^k, d^k \right\rangle \\
&= - \left\| g^k \right\|^2 + \beta_k \left\langle g^k, d^{k-1} \right\rangle - \theta_k \left\langle g^k, y^{k-1} \right\rangle \\
&= - \left\| g^k \right\|^2 + \frac{\left\langle g^k, y^{k-1} \right\rangle}{\left\| g^{k-1} \right\|^2} \left\langle g^k, d^{k-1} \right\rangle - \frac{\left\langle g^k, d^{k-1} \right\rangle}{\left\| g^{k-1} \right\|^2} \left\langle g^k, y^{k-1} \right\rangle \\
&= - \left\| g^k \right\|^2.
\end{aligned}
$$

This completes the proof of (18).
The equivalence between (18) and (19) is obtained by taking $\lambda = 1$ and 0 for sufficiency, and for necessity by taking any convex combination of g_1 and g_2.
(ii) This assertion follows straightforwardly by the very definition (5). □

2.2 Armijo-type steplength

To compute biobjective steplengths, we propose an appropriate condition of the Armijo type. The latter will also be crucial to prove the (global) convergence of the B-MPRP algorithm.

Proposition 2.2 *Let x^k and d^k as defined in (8) and (9) such that $\left\| g^k \right\| \neq 0$. Then, for all $\rho > 0$ (fixed), there exist $t_a > 0$, such that for all $t \in]0, t_a]$, we have that*

$$F_i(x^k + td^k) < F_i(x^k) - \rho t^2 \left\| d^k \right\|^2, \quad i = 1, 2. \tag{24}$$

Proof By differentiability of F_i, for $i = 1, 2$, we have

$$F_i\left(x^k + td^k\right) = F_i\left(x^k\right) + \left\langle g_i^k, td^k \right\rangle + R_i(td^k), \tag{25}$$

with

$$\lim_{t \to 0} \frac{\left| R_i(td^k) \right|}{t \left\| d^k \right\|} = 0.$$

Observe that if $\left\|g^k\right\| \neq 0$, then by (18), we have $\langle g_i^k, d^k \rangle < 0$, and in particular, $d^k \neq 0$. It is therefore clear that, for $i = 1, 2$,

$$-\frac{\langle g_i^k, d^k \rangle}{\|d^k\|} > 0 \quad \text{and} \quad \frac{R_i(td^k) + t^2 \rho \left\|d^k\right\|^2}{t\,\|d^k\|} \underset{t\to 0}{\longrightarrow} 0.$$

Hence the existence of $t_a > 0$ such that for all $t \in \,]0, t_a]$,

$$\left| \frac{R_i(td^k) + t^2 \rho \left\|d^k\right\|^2}{t\,\|d^k\|} \right| < -\frac{\langle g_i^k, d^k \rangle}{\|d^k\|}, \quad i = 1, 2.$$

Multiplying this inequality by $t\left\|d^k\right\|$, we obtain that for all $t \in]0, t_a]$,

$$R_i\big(td^k\big) < -t\langle g_i^k, d^k \rangle - t^2 \rho \left\|d^k\right\|^2, \quad i = 1, 2.$$

So replacing this in (25), for $i = 1, 2$, we finally get that for all $t \in \,]0, t_a]$,

$$\begin{aligned} F_i\big(x^k + td^k\big) &< F_i\big(x^k\big) + t\langle g_i^k, d^k \rangle - t^2 \rho \left\|d^k\right\|^2 - t\langle g_i^k, d^k \rangle \\ &= F_i\big(x^k\big) - t^2 \rho \left\|d^k\right\|^2, \end{aligned}$$

which shows (24). □

2.3 Algorithm and convergence analysis

Based on the above results, we describe bellow the complete B-MPRP algorithm.

B-MPRP algorithm pseudocode

Step 0: Initialisation.

- Choose $x^0 \in \mathbb{R}^n$, fix the Armijo's constant $\rho > 0$ and put $k = 0$.

Step 1: Descent direction.

- Compute λ_k, g^k and d^k according to (11), (10) and (9) respectively.

Step 2: Stopping criteria.

- If $\left\|g^k\right\| = 0$, then stop: x^k is Pareto critical for (P).

Step 3: Armijo line search.

- Determine a steplength $t_k \in \,]0, 1]$ such that

$$t_k = \max_{p\in\mathbb{N}} \left\{ \frac{1}{2^p} \ : \ F_i\big(x^k + \frac{1}{2^p} d^k\big) < F_i\big(x^k\big) - \frac{\rho}{4^p} \left\|d^k\right\|^2, \text{ for } i = 1, 2 \right\}.$$

Step 4: Updated point.

- Set $x^k := x^k + t_k d^k$, $k = k + 1$ and go to Step 1.

When the stopping criterion $\left\|g^k\right\| \neq 0$ never holds, the following standard assumptions will be required for the convergence of the B-MPRP algorithm:

(A1) The level set

$$\mathcal{L} = \left\{ x \in \mathbb{R}^n \ : \ F_i(x) \leq F_i(x^0), \quad i = 1, 2 \right\}$$

is bounded.

(A2) In some open set $\mathcal{N} \supset \mathcal{L}$, the gradients of the two criteria are L-Lipschitz continuous, namely, there exists a constant $L > 0$ such that, for $i = 1, 2$,

$$\|g_i(x) - g_i(y)\| \leq L \|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{26}$$

Assumption (A1) will ensure the boundedness of the sequence $\left(x^k\right)_{k\in\mathbb{N}}$ generated by the B-MPRP algorithm. Indeed, it is clear that the sequence $\left(F_i\left(x^k\right)\right)_{k\in\mathbb{N}}$, $i = 1, 2$, is decreasing, so $\left(x^k\right)_{k\in\mathbb{N}} \subseteq \mathcal{L}$. While assumption (A2) will guarantee the existence of a constant $\gamma > 0$ such that, for $i = 1, 2$,

$$\|g_i(x)\| \leq \gamma, \quad \forall x \in \mathcal{L}, \tag{27}$$

or equivalently, for all $\lambda \in [0, 1]$,

$$\left\|\lambda\left[g_1(x) - g_2(x)\right] + g_2(x)\right\| \leq \gamma, \quad \forall x \in \mathcal{L}. \tag{28}$$

We first prove the following important lemma.

Lemma 2.1 *Under the assumptions (A1) and (A2), the following assertions hold.*

(i)

$$\lim_{k\to+\infty} t_k \left\|d^k\right\| = 0. \tag{29}$$

(ii) If there exists $\varepsilon > 0$ such that

$$\left\|g^k\right\| \geq \varepsilon, \quad \forall k, \tag{30}$$

then there exists $M > 0$ such that

$$\left\|d^k\right\| \leq M, \quad \forall k. \tag{31}$$

Proof (i) From the Armijo-like condition (24), we have that, for all $k \in \mathbb{N}$,

$$F_i\left(x^k + t_k d^k\right) < F_i\left(x^k\right) - \rho t_k^2 \left\|d^k\right\|^2, \quad i = 1, 2,$$

which, by induction (8), can be rewritten as

$$F_i\left(x^{k+1}\right) < F_i\left(x^0\right) - \sum_{j=0}^{k} \rho t_j^2 \left\|d^j\right\|^2, \quad i = 1, 2.$$

Adding these two inequalities, for $i = 1, 2$, we obtain

$$\frac{1}{2}\sum_{i=1}^{2}\left[F_i\left(x^0\right) - F_i\left(x^{k+1}\right)\right] > \sum_{j=0}^{k} \rho t_j^2 \left\|d^j\right\|^2.$$

By continuity, the functions F_i, $i = 1, 2$, are bounded over the bounded set $\mathcal{L}$, and since $\left(x^k\right)_{k\in\mathbb{N}} \subset \mathcal{L}$, then there exists $M > 0$ such that, for all $k \in \mathbb{N}$,

$$\frac{1}{2}\sum_{i=1}^{2}\left[F_i\left(x^0\right) - F_i\left(x^{k+1}\right)\right] < M,$$

which implies that

$$\sum_{j\geq 0} \rho t_j^2 \left\|d^j\right\|^2 < +\infty.$$

In particular,

$$\lim_{k\to+\infty} t_k \left\|d^k\right\| = 0.$$

(ii) Using successively (9), (26), (28) with $\lambda = \lambda_k$, and, (30), one easily obtain that

$$\begin{aligned}\left\|d^k\right\| &\leq \left\|g^k\right\| + \frac{2\left\|g^k\right\|\left\|y^{k-1}\right\|\left\|d^{k-1}\right\|}{\|g^{k-1}\|^2} \\ &\leq \gamma + \frac{2\gamma L\, t_{k-1}\left\|d^{k-1}\right\|}{\varepsilon^2}\left\|d^{k-1}\right\|.\end{aligned}$$

Since, by (29), $t_k \left\|d^k\right\| \to 0$ as $k \to +\infty$, then for all $\eta \in]0,1[$, there exists $k_0 \in \mathbb{N}$ such that

$$\frac{2\gamma L}{\varepsilon^2}\, t_{k-1}\left\|d^{k-1}\right\| \leq \eta, \quad \forall k \geq k_0.$$

It follows that, for any $k > k_0$,

$$\begin{aligned}\left\|d^k\right\| &\leq \gamma + \eta\left\|d^{k-1}\right\| \\ &\leq \gamma \sum_{l=0}^{k-k_0-1} \eta^l + \eta^{k-k_0}\left\|d^{k_0}\right\| \\ &\leq \frac{\gamma}{1-\eta} + \left\|d^{k_0}\right\|.\end{aligned}$$

By setting

$$M = \max\left\{\left\|d^1\right\|, \left\|d^2\right\|, \ldots, \left\|d^{k_0}\right\|, \frac{\gamma}{1-\eta} + \left\|d^{k_0}\right\|\right\},$$

we get the desired result. □

The global convergence of the B-MPRP algorithm can now be stated as follows.

Theorem 2.1 *Under the assumptions (A1) and (A2), its holds that*

$$\liminf_{k\to+\infty} \left\|g^k\right\| = 0.$$

Proof Suppose, by a way of contradiction, that there exists $\varepsilon > 0$ such that

$$\left\|g^k\right\| \geq \varepsilon, \quad \forall k \in \mathbb{N}.$$

Then, according to (31), $\left(d^k\right)_{k\in\mathbb{N}}$ would be bounded. Two cases are to be distinguished.
First case: $\liminf_{k\to+\infty} t_k > 0$. Using (29), we would have that $\liminf_{k\to+\infty} \left\|d^k\right\| = 0$. But according to (19), we always have

$$\left\|g^k\right\| \leq \left\|d^k\right\|, \quad \forall k \in \mathbb{N}.$$

Passing to the "lim inf", we get the contradiction $\varepsilon \leq 0$.
Second case: $\liminf_{k\to+\infty} t_k = 0$. This case would imply that $t_k < 1$ for infinitely many k, and then, $2t_k$ could not satisfy the Armijo-like condition (24), i.e.,

$$F_{i_k}\left(x^k + 2t_k d^k\right) - F_{i_k}\left(x^k\right) > -4\rho t_k^2 \left\|d^k\right\|^2, \tag{32}$$

with $i_k = 1$ or 2. On the other hand, for all $\delta \in [0,1]$, one has

$$\left\| x^k - \left(x^k + 2\delta t_k d^k \right) \right\| \leq 2t_k \left\| d^k \right\|.$$

Since $x^k \in \mathcal{L} \subset \mathcal{N}$ and $\lim_{k\to\infty} t_k \left\| d^k \right\| = 0$ (because of (29)), it follows that, for all k sufficiently large,

$$x^k + 2\delta t_k d^k \in \mathcal{N}, \quad \forall \delta \in [0,1].$$

Using the mean value theorem together with (26) and (19), we deduce the existence of $\delta_k \in \,]0,1[$ such that

$$\begin{aligned}
& F_{i_k}\left(x^k + 2t_k d^k\right) - F_{i_k}\left(x^k\right) \\
& = 2t_k \left\langle g_{i_k}\left(x^k + 2\delta_k t_k d^k\right), d^k \right\rangle \\
& = 2t_k \left\langle g_{i_k}(x^k), d^k \right\rangle + 2t_k \left\langle g_{i_k}\left(x^k + 2\delta_k t_k d^k\right) - g_{i_k}(x^k), d^k \right\rangle \\
& \leq 2t_k \left\langle g_{i_k}(x^k), d^k \right\rangle + 2t_k \left\| g_{i_k}(x^k + 2\delta_k t_k d^k) - g_{i_k}(x^k) \right\| \left\| d^k \right\| \\
& \leq 2t_k \left\langle g_{i_k}^k, d^k \right\rangle + 4L\delta_k t_k^2 \left\| d^k \right\|^2 \\
& < 2t_k \left\langle g_{i_k}^k, d^k \right\rangle + 4L t_k^2 \left\| d^k \right\|^2 \\
& = -2t_k \left\| g^k \right\|^2 + 4L t_k^2 \left\| d^k \right\|^2.
\end{aligned}$$

Substituting the last inequality into (32), it would follow that, for infinitely many k,

$$\left\| g^k \right\|^2 \leq 2(L+\rho) t_k \left\| d^k \right\|^2.$$

Passing to the "lim inf" while using the fact that $\left(d^k\right)_{k\in\mathbb{N}}$ would be bounded, we also get a contradiction ($\varepsilon = 0$). This completes the proof of the theorem. □

3 Numerical experiments and comparison

In this section, some numerical experiments on the proposed method are reported. A comparison is also made between B-MPRP and the PRP method proposed in [23] for solving multicritera problems. The B-MPRP algorithm was coded in SCILAB 6.1.0., while we used the PRP code written in Fortran 90, which is made by the authors themselves of [23], freely available on: https://lfprudente.ime.ufg.br/. All codes were executed on the same machine equipped with 1.90 GHz Intel(R) Core(TM) i5 CPU and 16 Go memory.

Eighty well-known unconstrained bicriteria problems that we think are sufficient to reflect the essential aspects of the proposed method are selected: JOS1, SP1, Far1, MLF2, MOP2, MOP3, SK2 and VU1 from [18]; Lov1, Lov3 and Lov4 from [19]; SLC1 and SLC2 from [27]; AP3 from [2]; Hil1 from [17], MMR1 and MMR5 from [20].

To make correct comparisons, two factors were taken into account. First, each test problem were solved by the two methods starting with the same initial population of 200 individuals. Second, a similar stopping criterion was considered:

$$\left\| g^k \right\|^2 < 10^{-6},$$

where g^k is defined by (10)-(11) for B-MPRP, while for PRP as defined in [23], $g^k = g^k(\lambda_k)$ (the mapping $g^k(\cdot)$ being given by (17)) such that

$$\lambda_k \in \operatorname*{argmin}_{\lambda\in[0,1]} \frac{1}{2}\left\| g^k(\lambda) \right\|^2.$$

During running, we prescribed as Armijo's constant $\rho = 10^{-4}$ for B-MPRP.

In order to graphically analyse the experiment results and easily compare the two methods, we opted for the performance measures: purity (P), hypervolume (HV) and generational distance (GD). Let us recall that P and GD measure the convergence of the approximated sets towards the exact Pareto front, while HV measures the dispersion (density) of the approximated solutions (see, e.g., [8] for more details). Fig. 1 shows the performance of the two methods with respect to CPU time and the metrics P, HV and GD, which we evaluated using the well-known profiles suggested in [5].

Note that $\rho(1)$ gives the largest number of problems among the best solved by a method with respect to a given performance measure. However, a value $\rho(\alpha)$ equal to 1 means that all the problems have been solved by the method at the threshold α. Thus, the best overall performance of a method is the one that achieves large values of $\rho(\alpha)$ for small values of α. So, comparing from Fig. 1, it is clearly seen that the two methods are very close in terms of CPU time, while in terms of HV, P and GD metrics, B-MPRP seems to perform better than PRP. Let us finally recall the practical interest of the new method proposed with both an explicit scheme and the attractive property of sufficient descent regardless of the line search, so that, the use of any type of stepsize is left free.

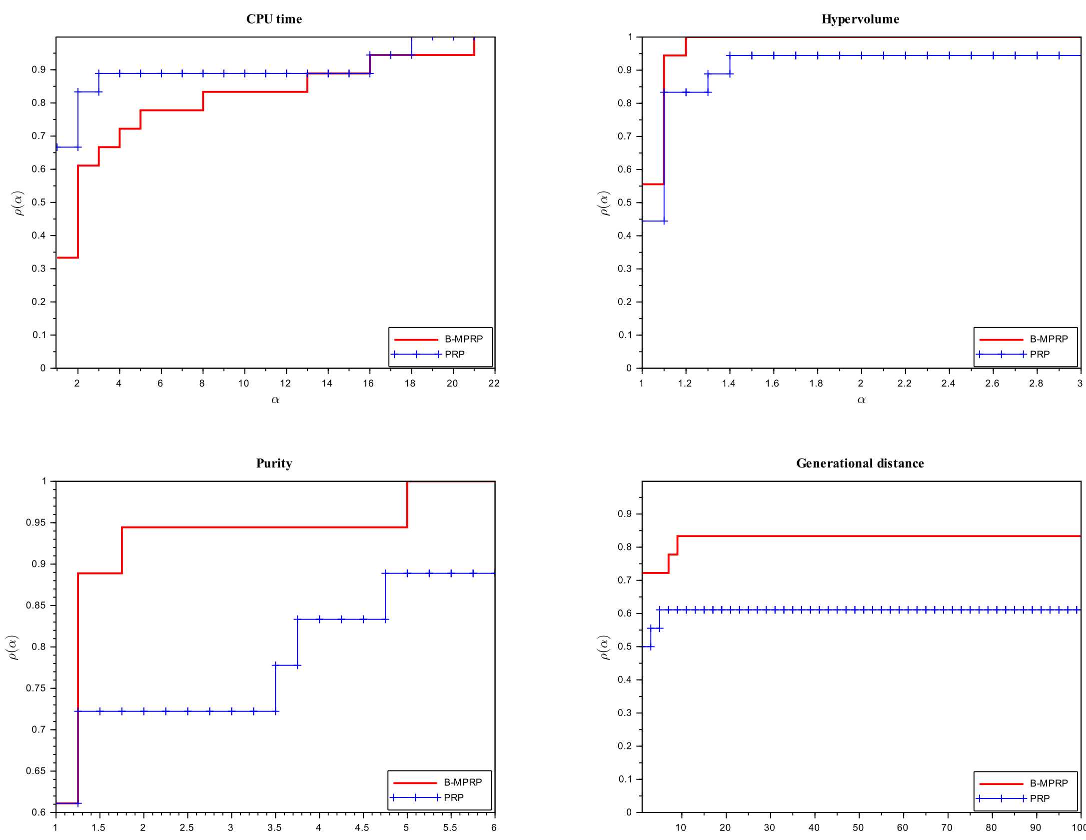

Figure 1: Performance profiles.

Statements and Declarations The authors have no pertinent declarations concerning conflicts of interest, financial or non-financial interests, competing interests or other statements to disclose.

References

[1] Al-Baali, M., Narushima, Y., Yabe, H.: A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization. Comput. Optim. Appl., 60, 89–110 (2015)

[2] Ansary, M. A., Panda, G.: A modified quasi-Newton method for vector optimization problem. Optim., 64, 2289–2306 (2015)

[3] Dai, Y.H., Yuan, Y.: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim., 10, pp. 177–182 (1999)

[4] Dai, Z.F., Tian, B.S.: Global convergence of some modified PRP nonlinear conjugate gradient methods. Optim. Lett., 5, 615–630 (2011)

[5] Dolan, E. D., Moré, J. J.: Benchmarking optimization software with performance profiles. Math. Program., 91, 201–213 (2002)

[6] El Maghri M., Elboulqe, Y.: Reduced Jacobian method. J. Optim. Theory Appl., 179, 917–943 (2018)

[7] El Maghri M., Elboulqe, Y.: Correction to: reduced Jacobian method. J. Optim. Theory Appl., 187, 304 (2020)

[8] El Maghri M., Elboulqe, Y.: A reduced Jacobian scheme with full convergence for multicriteria optimization (submitted).
https://optimization-online.org/?p=18970

[9] Fletcher, R., Reeves, C. M.: Function minimization by conjugate gradients. The Computer Journal, 7, 149–154 (1964)

[10] Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res., 51, 479–494 (2000)

[11] Fliege, J., Drummond, L.M.G., Svaiter, B.F.: Newton's method for multiobjective optimisation. SIAM J. Optim., 20, 602–626 (2009)

[12] Fukuda, E.H., Drummond, L.M.G.: A survey on multiobjective descent methods. Pesqui. Operacional, 34, 585–620 (2014)

[13] Fukuda, E.H., Drummond, L.M.G., Masuda, A.M.: A conjugate directions-type procedure for quadratic multiobjective optimization. Optim., 71, 419–437 (2022)

[14] Gonçalves, M.L.N., Prudente, L.F.: On the extension of the Hager–Zhang conjugate gradient method for vector optimization. Comput. Optim. Appl., 76, 889–916 (2020)

[15] García, P.U.M., Burguillo, R.J.C., González C.F.J.: Explicit gradient information in multiobjective optimization. Oper. Res. Lett., 36, 722–725 (2008)

[16] Hestenes, M.R., Stiefel, E.: Methods for conjugate gradients for solving linear systems. J. Res. Natl. Bur. Stand., 49, 409–436 (1952)

[17] Hillermeier, C.: Generalized homotopy approach to multiobjective optimization. J. Optim. Theory Appl., 110, 557–583 (2001)

[18] Huband, S., Hingston, P., Barone, L., While, L: A review of multiobjective test problems and a scalable test problem toolkit. IEEE Trans. Evol. Comput, 10, 477–506 (2006)

[19] Lovison, A.: Singular continuation: Generating piecewise linear approximations to Pareto sets via global analysis. SIAM J. Optim., 21, 463–490 (2011)

[20] Miglierina, E., Molho, E., Recchioni, M.: Box-constrained multi-objective optimization: A gradient-like method without "a priori" scalarization. Eur. J. Oper. Res., 188, 662–682 (2008)

[21] Mukai, H.: Algorithms for multicriterion optimization. IEEE Trans. Autom. Control, 25, 177–186 (1980)

[22] Narushima, Y., Yabe, H., Ford, J.A.: A three-term conjugate gradient method with sufficient descent property for unconstrained optimization. SIAM J. Optim., 21, 212–230 (2011)

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

[24] Polak, E., Ribière, G.: Note sur la convergence de méthodes de directions conjuguées. Rev. Fr. Inform. Rech. Opér., 3, 35–43 (1969)

[25] Polyak, B.T.: The conjugate gradient method in extreme problems. USSR Comput. Math. Math. Phys., 9, pp. 94–112 (1969)

[26] Qu, S., Goh, M., Chan, F.T.: Quasi-Newton methods for solving multiobjective optimization, Oper. Res. Lett., 39, 397–399 (2011)

[27] Schütze, O., Lara, A., Coello, C.A.C.: The directed search method for unconstrained multi-objective optimization problems, In: Proceedings of the EVOLVE–a bridge between probability, set oriented numerics, and evolutionary computation, pp. 1–4 (2011)

[28] Svaiter, B.F.: The multiobjective steepest descent direction is not Lipschitz continuous, but is Hölder continuous, Oper. Res. Lett., 46, 430–433 (2018)

[29] Yuan, G., Zhang, M.: A three-terms Polak–Ribière–Polyak conjugate gradient algorithm for large–scale nonlinear equations, J. Comput. Appl. Math., 286, 186–195 (2015)

[30] Zhang, L., Zhou, W., Dong-Hui, L.: A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence, IMA J. Numer. Anal., 26, 629–640 (2006)

[31] Zhang, L., Zhou, W., Li, D.H.: Some descent three-term conjugate gradient methods and their global convergence, Optim. Methods Softw., 22, 697–711 (2007)