A Reduced Jacobian Scheme with Full Convergence for Multicriteria Optimization

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

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

May 24, 2022

In this paper, we propose a variant of the reduced Jacobian method (RJM) introduced by El Maghri and Elboulqe in [JOTA, 179 (2018) 917–943] for multicriteria optimization under linear constraints. Motivation is that, contrarily to RJM which has only global convergence to Pareto KKT-stationary points in the classical sense of accumulation points, this new variant possesses the full convergence property in the sense that the entire sequence converges whenever the objectives are quasiconvex. Simulations are reported showing the performance of this variant compared to RJM and the evolutionary NSGA-II method.

Keywords: multicriteria optimization; Pareto optimality; nonlinear programming; multiobjective descent direction; reduced gradient method; convergence analysis

AMS Subject Classification: 90C29; 90C30; 90C52

1 Introduction

Many issues encountered today in various scientific areas often have several contradictory objectives for which the design of optimal strategies is a challenge for decision-making. These problems, known as multicriteria or multiobjective optimization problems (MOP), arise in the field of vector optimization. Their particular difficulty is the fact that the order used for preference is generally not total. In this case, there will be no single optimal value, but a set of efficient values called the Pareto front. These solutions first introduced by Pareto as being non-dominated points, are incomparable and are always at the forefront of the feasible image set. Note also that the Pareto front may be infinite or unbounded. Considering further that each Pareto solution is worth another and that none can represent all the others, it becomes important to determine all these solutions or to find good approximations of the diverse regions that well represent the Pareto front.

In this perspective, a consistent way to accomplish this task is to treat the initial MOP as it is, without introducing artificial scalarization parameters, nor other random variables that may be sensitive to the original problem. This idea was first initiated by Mukai [26] in 1980, who suggested a set of protocols for computing a multiobjective descent direction (i.e., a descent direction common to all the criteria). There has been proposed several

*Corresponding author. Email: elmaghri@yahoo.com

algorithms based on different types of finding-direction subproblems inspired from classical methods of scalar optimization. Their global convergence to Pareto critical accumulation points under line search conditions of the Armijo type has been also proved. However, this work remained unknown until the publication of the paper by Fliege and Svaiter [18] in 2000, who distinguished the unconstrained case from the constrained case by presenting the Cauchy method and the Zoutendijk method separately. Then, a series of works followed, e.g., Newton's method [17], projected gradient method [11], Zoutendijk's improved method [25], Wolfe's reduced gradient method [14] and others (see references given therein). Global convergence of these methods has been obtained in the sense of subsequence, and some of them were proved to be fully convergent in the sense of the whole sequence under suitable assumptions.

The reduced gradient method for multicriteria optimization was first introduced and called reduced Jacobian method (in brief, RJM) by El Maghri and Elboulqe [14] (see also [13]). In summary, RJM extends the reduced gradient method (RGM) of Wolfe [36] for linearly constrained single-objective programs without scalarizing the MOP. As in RGM, the constrained MOP is reduced to an unconstrained one depending only on nonbasic variables. This allowed us to easily compute multiobjective reduced descent directions by introducing a simple finding-direction subproblem. As this kind of scheme has only global convergence, we further investigate in order to obtain better convergence properties.

In fact, in this paper we propose a fully convergent reduced Jacobian method, in brief FC-RJM, which is a variant RJM that brings precisely a change to the objective functional of the finding-direction subproblem. We first prove that the new scheme is still globally convergent in the sense of accumulation points, and show in case of quasiconvex criteria, that this convergence becomes full in the sense of the entire sequence, provided that the number of basis changes is finite. This last condition reveals to be not only sufficient but also necessary whenever the Wolfe's criterion is applied, that is, the basis selection strategy favors the largest variables to enter into the basis at each iteration. Our new variant recovers an earlier improved reduced gradient method by Wang [35] for single-objective programs (see also [34]). As can be seen in the digital experiences section, FC-RJM gives good performance on a sample of constrained MOPs, compared to RJM [14], W-RGM [14] (classical weighting method of scalarization by RGM) and NSGA-II [10] (Nondominated sorting genetic algorithm).

2 RJM outline

Let $F : \mathbb{R}^n \to \mathbb{R}^r$ be given by $F(x) = (f_1(x), \ldots, f_r(x))$ a C^1 mapping on a standard polyhedron $S = \{x \in \mathbb{R}^n \ : \ Ax = b, \ x \geq 0\}$, where $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ a matrix of full rank $m < n$.

A point $x^* \in S$ is said to be *locally weakly efficient* or *weak Pareto optimum* for the following multicriteria optimization problems:

$$(MOP) \qquad \underset{x \in S}{\text{Min}} \ F(x)$$

iff, for a certain neighbourhood $\mathcal{N}(x^*)$ of x^*, there is no other $x \in S \cap \mathcal{N}(x^*)$ such that $F(x) < F(x^*)$, or equivalently, $f_j(x) < f_j(x^*)$ for $j = 1 \cdots r$. From the first-order necessary efficiency (Pareto optimality) condition, such a point $x^* \in S$ must satisfy the following Pareto KKT-stationarity system:

$$(KKT) \qquad \begin{cases} JF(x^*)^T \lambda^* + A^T v^* - u^* = 0, \\ u^* \cdot x^* = 0, \end{cases}$$

for some $(\lambda^*, u^*, v^*) \in \mathbb{R}^r_+ \setminus \{0\} \times \mathbb{R}^n_+ \times \mathbb{R}^m$, where $JF(x^*) = \big(\nabla f_1(x^*), \ldots, \nabla f_r(x^*)\big)$ is the Jacobian matrix of F at x^*, $JF(x^*)^T$ means the transpose of the matrix $JF(x^*)$ and the symbol "$\cdot$" stands for the usual inner product.

Any feasible point solution to the previous system is called *Pareto KKT-stationary* or *Pareto critical* for (MOP), and (λ^*, u^*, v^*) are the well-known *Lagrange multipliers*.

Our study in this paper is to find all the Pareto critical points associated to (MOP). This will include the weakly efficient solutions, in particular, the efficient solutions and the properly efficient solutions (see, e.g., [14] for more details).

To this end, we will use Wolfe's reduction strategy. Following this strategy, the matrix A is partitioned to $A = [A_B\ A_N]$ so that the submatrix A_B is invertible (since it is assumed that A is of full rank). The subset $B \subset \{1, \ldots, n\}$ is called a *basis* of A, and obviously $N = \{1, \ldots, n\} \backslash B$. Then, basic variables $x_B = (x_i)_{i \in B}$ and nonbasic variables $x_N = (x_i)_{i \in N}$ are selected in such a way that $x = (x_B, x_N) \geq 0$ and the system $Ax = b$ amounts to $x_B = A_B^{-1}b - A_B^{-1}A_N x_N$. On te other hand, the vector cost F reduces to $H(x_N) = F(A_B^{-1}b - A_B^{-1}A_N x_N, x_N)$. The so-called *reduced Jacobian matrix* of F is in fact the Jacobian matrix $JH(x_N)$ which obviously is given by

$$U_N(x) := JF_N(x) - JF_B(x)A_B^{-1}A_N,$$

where $JF_N(x) = \frac{\partial F}{\partial x_N}(x)$ $\big($resp. $JF_B = \frac{\partial F}{\partial x_B}(x)\big)$ is the Jacobian matrix of F with respect to the nonbasic vector x_N (resp. the basic vector x_B).

Given a nonbasic vector d_N and a basic vector $d_B = -A_B^{-1}A_N d_N$, the vector $d = (d_B, d_N)$ satisfies $Ad = 0$, and by the partition $JF(x) = [JF_B(x)\ JF_N(x)]$, the matrix $JF(x)$ also satisfies :

$$JF(x)d = U_N(x)d_N. \tag{1}$$

A nonbasic vector $d_N \in \mathbb{R}^{n-m}$ is called a *multiobjective reduced descent direction* of F at $x \in S$, if it satisfies:

$$U_N(x)d_N < 0 \quad \text{and} \quad d_i \geq 0,\ \forall i \in I(x) \cap N, \tag{2}$$

where $I(x) = \{i \in \{1, \ldots, n\}\ :\ x_i = 0\}$ is the index set of active variables of x. Note that d_N is not other than a feasible multiobjective descent direction of the reduced cost H at x_N (see [14] for more details about feasible multiobjective descent directions).

Let us recall that a feasible point x is said to be *nondegenerate*, if we can select a nondegenerate basis for x, i.e., there exists a basis B such that $x_B > 0$. The feasible polyhedron S is in turn called nondegenerate, if any $x \in S$ is nondegenerate.

The following result is a consequence of [14, Theorem 2.1] that establishes a geometric characterization of the Pareto KKT-stationarity in terms of reduced directions.

Theorem 2.1 *A nondegenerate point $x \in S$ is Pareto KKT-stationary for (MOP), iff there is no multiobjective reduced descent direction of F at x.*

The descent RJM scheme, as stated in its simple form in [14], consists to take at $x \in S$: $\forall i \in N$,

$$d_i = \begin{cases} -x_i\big(U_N(x)^T\lambda(x)\big)_i & \text{if } \big(U_N(x)^T\lambda(x)\big)_i > 0, \\ -\big(U_N(x)^T\lambda(x)\big)_i & \text{else,} \end{cases} \tag{3}$$

such that

$$\lambda(x) \in \operatorname*{argmin}_{\lambda \in \Lambda} \frac{1}{2} \sum_{i \in N} \Big\lfloor \big(U_N(x)^T\lambda\big)_i \Big\rfloor_-^2 + x_i \Big\lfloor \big(U_N(x)^T\lambda\big)_i \Big\rfloor_+^2,$$

where $\Lambda = \{(\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^r_+ \ : \ \sum_{j=1}^r \lambda_j = 1\}$, $\lfloor a \rfloor_+ = \max\{0, a\}$ is the positive part of the scalar a and $\lfloor a \rfloor_- = \max\{0, -a\}$ is its negative part.

For $r = 1$, we retrieve the well-known continuous scheme of Wolfe for single-objective programs (see, e.g., [2] or [22]) which may be given under the simpler compact form :

$$d_i = \left\lfloor u_i(x) \right\rfloor_- - x_i \left\lfloor u_i(x) \right\rfloor_+, \tag{4}$$

where $u_i(x)$ stands for the ith component of the reduced gradient of the objective at x.

There has been shown that the vector d_N given by (3) possesses multiobjective descent properties: if $d_N \neq 0$, then it is a multiobjective reduced descent direction of F at x; otherwise, x is Pareto KKT-stationary for (MOP).

Algorithm RJM:

1. Choose a nondegenerate point $x = (x_B, x_N) \in S$ and fix an Armijo's constant $\beta \in]0, 1[$.

2. Compute a multiobjective reduced descent direction d_N for F at x according to (3).

3. If $\|d_N\| = 0$, then STOP: the current point x is Pareto KKT-stationary for (MOP); else, form the direction $d = (d_B, d_N)$ with $d_B = -A_B^{-1} A_N d_N$, and choose a steplength $t \in \left]0, \bar{t}\right]$ moving along d, that satisfies:

$$t = \max_{p \in \mathbb{N}} \left\{ \frac{1}{2^p} \ : \ F(x + \frac{1}{2^p} d) < F(x) + \beta \frac{1}{2^p} U_N(x) d_N \right\},$$

   where $\bar{t} = 1$ if $d \geq 0$, and, $\bar{t} = \min \left\{ -\frac{x_i}{d_i} \ : \ d_i < 0, \ i = 1, \ldots, n \right\}$ otherwise.

4. Update with $x := x + td$ and change the basis B in case of degeneracy.

Remark 2.1 In this procedure, the bound $\bar{t}$ ensures the positivity of the new iterate when $d \not\geq 0$, and, the nondegeneracy assumption ensures that $\bar{t} > 0$, so that, $]0, \bar{t}]$ is nonempty. If it happens that $t = \bar{t} = -\frac{x_i}{d_i}$ for some $i \in B$, then the new iterate will be degenerate and a basis change procedure is necessary in this case; otherwise, nondegeneracy is preserved (see [14, Lemma 4.1]). On the other hand, the existence of the steplength $t \in]0, \bar{t}]$ in Step 3 is guaranteed according to the following extended Armijo property for vector mappings (see, e.g., [26, Proposition 6] or [18, Lemma 4]):

$$\exists t_a > 0, \quad \forall t \in]0, t_a], \quad F(x + td) < F(x) + \beta t JF(x) d, \tag{5}$$

whenever d satisfies $JF(x)d < 0$. Let us recall, by (1), that $JF(x)d = U_N(x)d_N$.

The RJM scheme as defined above is however only globally convergent, i.e., every accumulation point of the sequence produced by the algorithm is Pareto KKT-stationary. As said in the introduction, we aim to extend this convergence to the entire sequence to obtain the so-called full convergence property. In the next section, we will see that modifying the reduced descent scheme will lead to a variant of RJM with the desired property.

3 The FC-RJM variant

Consider, for $x \in S$, the following scalar optimization subproblem:

$$(\mathcal{P}_x) \qquad \underset{\lambda\in\Lambda}{\text{Min}}\ f(\lambda,x) := \frac{1}{2}\left(\sum_{i\in N}\Big\lfloor x_i - \big(U_N(x)^T\lambda\big)_i\Big\rfloor_+^2\right) + x_N\cdot\big(U_N(x)^T\lambda\big)\,.$$

It is obvious that $(\mathcal{P}_x)$ is convex continuously differentiable and always has an optimal solution $\lambda(x)$. Moreover, it is easy to verify that the gradient of the functional $f(.,x)$ is given by

$$\nabla_\lambda f(\lambda,x) = -U_N(x)\delta_N(\lambda,x),$$

where

$$\delta_N(\lambda,x) := \left(\Big\lfloor x_i - \big(U_N(x)^T\lambda\big)_i\Big\rfloor_+ - x_i\right)_{i\in N}.$$

The following lemma characterizing the solutions of $(\mathcal{P}_x)$ will be useful not only for proving the descent properties but also for the convergence analysis.

Lemma 3.1 *$\lambda(x)\in\Lambda$ is optimal for $(\mathcal{P}_x)$, iff*

$$\Big(U_N(x)\delta_N(\lambda(x),x)\Big)\cdot\lambda \le \Big(U_N(x)\delta_N(\lambda(x),x)\Big)\cdot\lambda(x) \qquad (\forall\lambda\in\Lambda),$$

or equivalently,

$$U_N(x)\delta_N(\lambda(x),x) \le \Big[\Big(U_N(x)\delta_N(\lambda(x),x)\Big)\cdot\lambda(x)\Big]e, \qquad e=(1,\dots,1)\in\mathbb{R}^r.$$

Proof By the minimum principle applied to the differentiable convex program $(\mathcal{P}_x)$,

$$\lambda(x)\in\text{argmin}(\mathcal{P}_x) \quad\Leftrightarrow\quad \nabla_\lambda f\big(\lambda(x),x\big)\cdot\big(\lambda-\lambda(x)\big)\ge 0 \qquad (\forall\lambda\in\Lambda).$$

So the first inequality of the lemma follows straightforwardly using the expression of $\nabla_\lambda f\big(\lambda(x),x\big)$. The second inequality follows necessarily by taking successively $\lambda = e^j$ $(j=1\cdot\cdot\, r)$, where e^j is the jth vector of the canonical basis of $\mathbb{R}^r$, while sufficiency follows easily using the fact that $\lambda\ge 0$ and $\lambda\cdot e = 1$. □

The descent FC-RJM scheme consists to take, at $x\in S$, $d_N := \delta_N(\lambda(x),x)$, where $\lambda(x)\in\text{argmin}(\mathcal{P}_x)$, so that: $\forall i\in N$,

$$d_i = \begin{cases} -x_i & \text{if } \big(U_N(x)^T\lambda(x)\big)_i > x_i,\\ -\big(U_N(x)^T\lambda(x)\big)_i & \text{else.}\end{cases} \tag{6}$$

The following descent properties hold.

Proposition 3.1 *Let the nonbasic vector d_N be given by (6). Then,*

(i) *$U_N(x)d_N \le -\|d_N\|^2 e$, where $e=(1,\dots,1)\in\mathbb{R}^r$.*

(ii) *$d_N\neq 0 \Longrightarrow d_N$ is a multiobjective reduced descent direction of F at x.*
If furthermore $x_B>0$ (i.e., x is nondegenerate) with $B=\{1,\dots,n\}\setminus N$, then x is not Pareto KKT-stationary for (MOP).

(iii) *$d_N = 0 \Longrightarrow x$ is Pareto KKT-stationary for (MOP).*

Proof (i) According to the second inequality of Lemma 3.1 and the very definition of d_N, it suffices to prove that

$$\Big(U_N(x)\delta_N(\lambda(x),x)\Big)\cdot\lambda(x)\leq-\left\|d_N\right\|^2.$$

Indeed, if we denote by $\tilde{I}=\left\{i\in N\ :\ \left(U_N(x)^T\lambda(x)\right)_i>x_i\right\}$, then by virtue of (6),

$$\begin{aligned}\Big(U_N(x)\delta_N(\lambda(x),x)\Big)\cdot\lambda(x) &= d_N\cdot\left(U_N(x)^T\lambda(x)\right)\\ &= -\sum_{i\in\tilde{I}}x_i\left(U_N(x)^T\lambda(x)\right)_i-\sum_{i\in N\setminus\tilde{I}}\left(U_N(x)^T\lambda(x)\right)_i^2\\ &\leq -\sum_{i\in\tilde{I}}x_i^2-\sum_{i\in N\setminus\tilde{I}}\left(U_N(x)^T\lambda(x)\right)_i^2\\ &= -\left\|d_N\right\|^2.\end{aligned}$$

(ii) The first part of this assertion follows straightforwardly from the assertion (i) and the fact that, for $i\in I(x)\cap N$, $d_i=\delta_i(\lambda(x),x)=\Big\lfloor-\left(U_N(x)^T\lambda(x)\right)_i\Big\rfloor_+\geq 0$. The second part follows from Theorem 2.1.

(iii) Suppose that $d_N=0$ and put $u_N^*:=U_N(x)^T\lambda(x)$. Then, by (6), it follows that

$$u_N^*\geq 0\quad\text{and}\quad u_N^*\cdot x_N=0.$$

Clearly, for $\lambda^*:=\lambda(x)$, $u^*:=(0,u_N^*)$ and $v^*:=-(A_B^{-1})^TJF_B(x)^T\lambda^*$, it holds that $\lambda^*\in\mathbb{R}^r_+\setminus\{0\}$, $u^*\in\mathbb{R}^n_+$, $u^*=JF(x)^T\lambda^*+A^Tv^*$ and $u^*\cdot x=0$, which express the Pareto KKT-stationarity for (MOP). □

In the scalar case ($r=1$), the new FC-RJM scheme (6) coincides with the reduced gradient scheme by Wang [35]: $\forall i\in N$, $d_i=-x_i$ if $u_i(x)>x_i$, $d_i=-u_i(x)$ otherwise, where $u_i(x)$ is the ith component of the reduced gradient of the objective. Note that his scheme may also be rewritten under the compact form:

$$d_i=\left\lfloor x_i-u_i(x)\right\rfloor_+-x_i. \tag{7}$$

Algorithm FC-RJM:

1. Choose a nondegenerate point $x=(x_B,x_N)\in S$ and fix an Armijo's constant $\beta\in]0,1[$.

2. Compute a multiobjective reduced descent direction d_N for F at x according to (6).

3. If $\|d_N\|=0$, then STOP: the current point x is Pareto KKT-stationary for (MOP); else, form the direction $d=(d_B,d_N)$ with $d_B=-A_B^{-1}A_Nd_N$, and choose a steplength $t\in\left]0,\bar{t}\right]$ moving along d, that satisfies:

$$t=\max_{p\in\mathbb{N}}\left\{\frac{1}{2^p}\ :\ F(x+\frac{1}{2^p}d)<F(x)+\beta\frac{1}{2^p}U_N(x)d_N\right\},$$

where $\bar{t}=1$ if $d_B\geq 0$, and, $\bar{t}=\min\left\{-\frac{x_i}{d_i}\ :\ d_i<0,\ i\in B\right\}$ otherwise.

4. Update with $x:=x+td$ and change the basis B in case of degeneracy.

Remark 3.1 It is worth noting that in this procedure and contrarily to the RJM scheme, the positivity bound $\bar{t}$ does not require $i \in N$, which is considerable for computational purpose. Indeed, according to the new scheme (6), one has that for each $i \in N$ such that $d_i < 0$,

$$x_i + td_i = \begin{cases} x_i - tx_i & \text{if } \left(U_N(x)^T\lambda(x)\right)_i > x_i, \\ x_i - t\left(U_N(x)^T\lambda(x)\right)_i & \text{else.} \end{cases}$$

Since in the second case, $\left(U_N(x)^T\lambda(x)\right)_i > 0$ because $d_i < 0$, it is clear that $x_i + td_i \geq 0$ for all $i \in N$ and $t \in]0, 1]$.

4 Convergence analysis

From now one, we will assume that the sequence produced by the FC-RJM algorithm is infinite, since otherwise, according to Proposition 3.1(iii), the algorithm will stop after a finite number of iterations at a Pareto KKT-stationary point.

We shall also consider the well-known basis property hypothesis which guarantees preservation of nondegeneracy at any limit point non Pareto KKT-stationary:

$$(\mathcal{H}) \qquad \left\{ \begin{array}{l} x^k \underset{k\in K\subseteq\mathbb{N}}{\longrightarrow} x^* \text{ non Pareto KKT-stationary,} \\ \\ \exists B \text{ basis, } \forall k \in K,\ x_B^k > 0. \end{array} \right\} \Longrightarrow x_B^* > 0.$$

The hypothesis above was already used by many authors for the convergence analysis of several RGM algorithms. An example given in [24] shows that this hypothesis is essential. However, it has been also established that using some special basis selection procedures like the Wolfe's criterion, namely, the strategy that favors the largest variables to enter into the basis at each iteration, the convergence of RGM then does not require this hypothesis (see, e.g., algorithms proposed in [2, 22]). In particular, Smeers [30] proposed a basis selection procedure that guarantees the result of $(\mathcal{H})$. Also in [37], a pivotal operation for basis changes was performed at each iteration ensuring in somewhat the same result. Of course, these latter procedures are in counterpart expensive from the point of view of computational costs.

4.1 Global convergence

The following theorem gives a global convergence result. Its proof closely follows the convergence lines of [14, Theorem 4.1], but let us take this opportunity to restore the forgotten hypothesis $(\mathcal{H})$ in the latter theorem, which in fact, is required for the proof.

Theorem 4.1 *Assume that S is nondegenerate and $(\mathcal{H})$ is satisfied. Let $(x^k)_{k\in\mathbb{N}}$ be the sequence produced by the FC-RJM algorithm. Then,*

(i) *$x^k \in S$ and $F(x^{k+1}) < F(x^k)$ for all $k \in \mathbb{N}$;*

(ii) *any accumulation point x^* of $(x^k)_k$ is a Pareto KKT-stationary point for (MOP), and, $\lim\limits_{k\to+\infty} F(x^k) = F(x^*)$.*

Proof (i) At the iteration k, one has $x^{k+1} = x^k + t_k d^k$. The fact that $Ax^{k+1} = b$ comes from the fact that $Ad^k = 0$, since by construction, $d^k_{B_k} = -A^{-1}_{B_k} A_{N_k} d^k_{N_k}$ where B_k is the nondegenerate basis for the kth iterate x^k. While the positivity of x^{k+1}, it follows from

$t_k \in]0, \bar{t}_k]$ where $\bar{t}_k$ is defined in Step 3 of the algorithm (see Remarks 2.1 and 3.1). The second part of this assertion follows directly from the Armijo's condition (5): $\forall k \in \mathbb{N}$,

$$F(x^{k+1}) - F(x^k) < \beta t_k U_{N_k}(x^k) d^k_{N_k} < 0, \tag{8}$$

since the algorithm is assumed to be infinite, i.e., $d^k_{N_k}$ are always reduced descent directions. (ii) By hypothesis, there exists a subsequence $x^k \underset{k\in K_1}{\longrightarrow} x^*$ with $K_1 \subseteq \mathbb{N}$ an infinite subset. The closeness of S implies that x^* remains feasible as $(x^k)_k$. Now, by continuity of F and the previous assertion (i), it follows that $F(x^k) \underset{k\in\mathbb{N}}{\longrightarrow} F(x^*)$. On the other hand, since the index set B_k is included in the finite set $\{1, \ldots, n\}$, we can assume without loss of generality that $B_k = B$ (hence $N_k = N$) for all $k \in K_1$. Now, observe that while passing onto the limit in (8), we get

$$\lim_{\substack{k\in K_1 \\ k\to+\infty}} t_k U_N(x^k) d^k_N = 0. \tag{9}$$

Also, as $\big(\lambda(x^k)\big)_k$ is bounded and the sequence $\big(U_N(x^k)\big)_k$ is convergent, then by virtue of (6), $\big(d^k_N\big)_{k\in K_1}$ is bounded. This shows that $\lambda(x^k) \underset{k\in K_2}{\longrightarrow} \lambda^* \in \Lambda$ and $d^k_N \underset{k\in K_2}{\longrightarrow} d^*_N$ for some $K_2 \subseteq K_1$ an infinite subset. In fact, we shall first prove that d^*_N is given by (6) at x^*. Indeed, according to Lemma 3.1, one has: $\forall k \in K_2$,

$$U_N(x^k)\delta_N(\lambda(x^k), x^k) \leq \Big[\Big(U_N(x^k)\delta_N(\lambda(x^k), x^k)\Big) \cdot \lambda(x^k)\Big] e.$$

Passing to the limit, as $k \nearrow +\infty$, this leads to

$$U_N(x^*)\delta_N(\lambda^*, x^*) \leq \Big[\Big(U_N(x^*)\delta_N(\lambda^*, x^*)\Big) \cdot \lambda^*\Big] e.$$

Once again, by virtue of Lemma 3.1, $\lambda^* \in \operatorname{argmin}(\mathcal{P}_{x^*})$. Also, passing to the limit in the formulae $d^k_N = \delta_N(\lambda(x^k), x^k)$, as $k \nearrow +\infty$, we obtain that

$$d^*_N = \delta_N(\lambda^*, x^*).$$

Hence d^*_N is given by (6) at x^*. Now, suppose by contradiction way, that x^* is not Pareto KKT-stationary. Then, according to Proposition 3.1(ii)-(iii), d^*_N would be a multiobjective reduced descent direction at $x^* \in S$, that is,

$$U_N(x^*) d^*_N < 0. \tag{10}$$

By (9), we would have $t_k \underset{k\in K_2}{\longrightarrow} 0$, which would imply that for any fixed $p \in \mathbb{N}$ and any $k \in K_2$ sufficiently large, $t_k < \frac{1}{2^p}$. Moreover, there exists p_0 such that for any $p \geq p_0$, $\frac{1}{2^p} < \inf\limits_{k\in K_2} \bar{t}_k$, where the steplength bound $\bar{t}_k$ is defined in Step 3 of the FC-RJM algorithm. Indeed, if $\inf\limits_{k\in K_2} \bar{t}_k = 0$, as B is a finite set, we would have (without loss of generality) that $\bar{t}_k = -\frac{x^k_{i_0}}{d^k_{i_0}} \underset{k\in K_2}{\longrightarrow} 0$ for some $i_0 \in B$. Since $d^k_B = -A_B^{-1} A_N d^k_N$, the sequence $(d^k_{i_0})_{k\in K_2}$ is bounded as $(d^k_N)_k$, and we would obtain that $x^k_{i_0} \underset{k\in K_2}{\longrightarrow} 0 = x^*_{i_0}$, which would contradict the hypothesis $(\mathcal{H})$. Hence, for any fixed $p \geq p_0$ and any $k \in K_2$ sufficiently large, $t_k < \frac{1}{2^p} < \bar{t}_k$. This means that, at the iteration k, the Armijo's condition would not be satisfied at $t = \frac{1}{2^p}$. That is, for all $k \in K_2$ sufficiently large,

$$F\left(x^k + \frac{1}{2^p} d^k\right) \not< F(x^k) + \frac{\beta}{2^p} U_N(x^k) d^k_N.$$

At the limit, when $k \nearrow +\infty$, we would obtain that for infinitely many p,

$$F\left(x^* + \frac{1}{2^p}d^*\right) \not< F(x^*) + \frac{\beta}{2^p}U_N(x^*)d_N^*,$$

where $d^* := (d_B^*, d_N^*)$ with $d_B^* := A_B^{-1}A_N d_N^*$. But, by virtue of (10), this contradicts the necessary Armijo's condition (5). Hence, x^* is well Pareto KKT-stationary point for (MOP). □

Remark 4.1 An accumulation point of the generated sequence $\left(x^k\right)_k$ exists, for instance, if the feasible level subset $\left\{x \in S :\ F(x) \leq F(x^0)\right\}$ is bounded, since it contains $\left(x^k\right)_k$. Thus, if such a point exists, then by replacing the stopping criterion $\left\|d_N^k\right\| = 0$ in Step 3 of the algorithm by $\left\|d_N^k\right\| \leq \varepsilon$ with any prespecified precision $\varepsilon > 0$, the algorithm will stop after a finite number of iterations according to Theorem 4.1(ii).

4.2 Full convergence

In this subsection we shall prove the full convergence of the FC-RJM algorithm when the vector objective F is assumed to be quasiconvex. Recall that the vector mapping $F : \mathbb{R}^n \to \mathbb{R}^r$ is said to be *quasiconvex* on the convex set S, if for all $y \in \mathbb{R}^r$, the feasible level set $\{x \in S : F(x) \leq y\}$ is convex. Note that F is quasiconvex on S iff, F is componentwise quasiconvex on S (see [4]). As vector convexity also means componentwise, it follows that vector quasiconvexity is more general than vector convexity. Since here, we assume that the vector mapping F is differentiable, we will rather use the following characterization whose proof can be found in [5].

Proposition 4.1 The vector mapping $F : S \subset \mathbb{R}^n \longrightarrow \mathbb{R}^r$ is quasiconvex on S, iff

$$\forall x, x' \in S, \quad F(x') < F(x) \Longrightarrow JF(x)(x' - x) \leq 0.$$

Before stating the main theorem, we first prove the following lemma.

Lemma 4.1 *Assume that F is quasiconvex on S nondegenerate and let $(x^k)_{k\in\mathbb{N}}$ be the sequence produced by the FC-RJM algorithm. Then for any $x' \in S$ and $k \in \mathbb{N}$ such that $F(x') < F(x^k)$, we have*

$$\sum_{j=1}^{r} f_j(x^{k+1}) + \frac{\beta}{2}\left\|x_{N_k}^{k+1} - x'_{N_k}\right\|^2 < \sum_{j=1}^{r} f_j(x^k) + \frac{\beta}{2}\left\|x_{N_k}^{k} - x'_{N_k}\right\|^2,$$

where N_k is the nonbasic index set generated by the algorithm at the iteration k, and, $\beta \in]0,1[$ is the Armijo's constant.

Proof It is not difficult to verify that the nonbasic vector $d_{N_k}^k$ generated by the algorithm at the iteration k, is an optimal solution to the convex problem:

$$\underset{v \geq -x_{N_k}^k}{\text{Min}} \ \psi(v,\ x^k) := \frac{1}{2}\left\|v + U_{N_k}(x^k)^T\lambda(x^k)\right\|^2,$$

and then, satisfies the minimum variational principle: $\forall v \geq -x_{N_k}^k$,

$$0 \leq \nabla_v\psi\big(d_{N_k}^k,\ x^k\big) \cdot \big(v - d_{N_k}^k\big) = \Big(d_{N_k}^k + U_{N_k}(x^k)^T\lambda(x^k)\Big) \cdot \big(v - d_{N_k}^k\big).$$

By taking $v = x'_{N_k} - x^k_{N_k}$ for any $x' \in S$ such that $F(x') < F(x^k)$, it follows that

$$\begin{aligned}
\left\|x^{k+1}_{N_k} - x'_{N_k}\right\|^2 &= \left\|x^k_{N_k} + t_k d^k_{N_k} - x'_{N_k}\right\|^2 \\
&= \left\|x^k_{N_k} - x'_{N_k}\right\|^2 + 2t_k d^k_{N_k} \cdot \left(x^k_{N_k} - x'_{N_k}\right) + t_k^2 \left\|d^k_{N_k}\right\|^2 \\
&\leq \left\|x^k_{N_k} - x'_{N_k}\right\|^2 + 2t_k d^k_{N_k} \cdot \left(x^k_{N_k} - x'_{N_k}\right) + 2t_k \left\|d^k_{N_k}\right\|^2 \\
&= \left\|x^k_{N_k} - x'_{N_k}\right\|^2 - 2t_k d^k_{N_k} \cdot \left(x'_{N_k} - x^k_{N_k} - d^k_{N_k}\right) \\
&\leq \left\|x^k_{N_k} - x'_{N_k}\right\|^2 + 2t_k \Big(U_{N_k}(x^k)^T \lambda(x^k)\Big) \cdot \left(x'_{N_k} - x^k_{N_k} - d^k_{N_k}\right) \\
&= \left\|x^k_{N_k} - x'_{N_k}\right\|^2 - 2t_k d^k_{N_k} \cdot \Big(U_{N_k}(x^k)^T \lambda(x^k)\Big) \\
&\qquad + \quad 2t_k \Big(U_{N_k}(x^k)^T \lambda(x^k)\Big) \cdot \left(x'_{N_k} - x^k_{N_k}\right).
\end{aligned}$$

On the other hand, by Armijo's condition (5) multiplied by the positive vector $\lambda(x^k)$: $\forall k \in \mathbb{N}$,

$$\begin{aligned}
-\frac{2}{\beta}\Big(F(x^{k+1}) - F(x^k)\Big) \cdot \lambda(x^k) &> -2t_k \Big(U_{N_k}(x^k) d^k_{N_k}\Big) \cdot \lambda(x^k) \\
&= -2t_k d^k_{N_k} \cdot \Big(U_{N_k}(x^k)^T \lambda(x^k)\Big).
\end{aligned}$$

It follows that

$$\begin{aligned}
\left\|x^{k+1}_{N_k} - x'_{N_k}\right\|^2 < \left\|x^k_{N_k} - x'_{N_k}\right\|^2 &- \frac{2}{\beta}\lambda(x^k) \cdot \Big(F(x^{k+1}) - F(x^k)\Big) \\
&+ \quad 2t_k \Big(U_{N_k}(x^k)^T \lambda(x^k)\Big) \cdot \left(x'_{N_k} - x^k_{N_k}\right).
\end{aligned}$$

According to Proposition 4.1 applied to x' and x^k since $F(x') < F(x^k)$, it is easy to see that

$$\begin{aligned}
0 &\geq \Big(JF(x^k)(x' - x^k)\Big) \cdot \lambda(x^k) \\
&= U_{N_k}(x^k)\left(x'_{N_k} - x^k_{N_k}\right) \cdot \lambda(x^k) \\
&= \Big(U_{N_k}(x^k)^T \lambda(x^k)\Big) \cdot \left(x'_{N_k} - x^k_{N_k}\right).
\end{aligned}$$

We finally obtain that

$$\left\|x^{k+1}_{N_k} - x'_{N_k}\right\|^2 < \left\|x^k_{N_k} - x'_{N_k}\right\|^2 - \frac{2}{\beta}\lambda(x^k) \cdot \Big(F(x^{k+1}) - F(x^k)\Big).$$

Using now that $\big(F(x^k)\big)_k$ is decreasing (see Theorem 4.1) and that $\big(\lambda(x^k)\big)_k \subset [0,1]^r$, we get the desired inequality:

$$\left\|x^{k+1}_{N_k} - x'_{N_k}\right\|^2 < \left\|x^k_{N_k} - x'_{N_k}\right\|^2 - \frac{2}{\beta}\sum_{j=1}^{r} f_j(x^{k+1}) - f_j(x^k).$$

□

Remark 4.2 Since in this section the algorithm is assumed to be infinite, then no x^k can be weakly efficient, and therefore, by the very definition, there is always $x' \in S$ such that $F(x') < F(x^k)$.

The theorem below gives necessary and sufficient conditions for the full convergence of the FC-RJM algorithm.

Theorem 4.2 *Assume that F is quasiconvex on S nondegenerate and $(\mathcal{H})$ is satisfied. Let $(x^k)_{k\in\mathbb{N}}$ be the sequence produced by the FC-RJM algorithm.*

(i) *If the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded and the number of basis changes is finite, then the whole sequence $(x^k)_{k\in\mathbb{N}}$ converges to a Pareto KKT-stationary point for (MOP).*

(ii) *Conversely, if the basis selection strategy that favors the largest variables to enter into the basis at each iteration (the so-called Wolfe's criterion) is used, then the number of basis changes is finite.*

Proof (i) By hypothesis, there exists a subsequence $x^k \underset{k\in K}{\longrightarrow} x^*$ with $K \subseteq \mathbb{N}$, and, there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$, $B_k = B$ (hence $N_k = N$), where $(B_k)_k$ is the sequence of the basic index sets generated by the algorithm. By Theorem 4.1, the sequence $\left(F(x^k)\right)_{k\in\mathbb{N}}$ is decreasing, $x^* \in S$ is a Pareto KKT-stationary point for (MOP) and $F(x^k) \underset{k\in\mathbb{N}}{\longrightarrow} F(x^*)$. Thus, $F(x^*) < F(x^k)$ for all $k \in \mathbb{N}$. It follows by Lemma 4.1 that for all $k \geq k_0$,

$$\sum_{j=1}^{r} f_j(x^{k+1}) + \frac{\beta}{2}\left\|x_N^{k+1} - x_N^*\right\|^2 < \sum_{j=1}^{r} f_j(x^k) + \frac{\beta}{2}\left\|x_N^k - x_N^*\right\|^2,$$

which means that the sequence

$$\left(\sum_{j=1}^{r} f_j(x^k) + \frac{\beta}{2}\left\|x_N^k - x_N^*\right\|^2\right)_{k\geq k_0}$$

is decreasing, and as it is bounded below, then it is convergent. Therefore,

$$\lim_{k\to+\infty}\sum_{j=1}^{r} f_j(x^k) + \frac{\beta}{2}\left\|x_N^k - x_N^*\right\|^2 = \lim_{\substack{k\in K\\ k\to+\infty}}\sum_{j=1}^{r} f_j(x^k) + \frac{\beta}{2}\left\|x_N^k - x_N^*\right\|^2 = \sum_{j=1}^{r} f_j(x^*).$$

Hence,

$$x_N^k \underset{k\in\mathbb{N}}{\longrightarrow} x_N^* \quad \text{and} \quad x_B^k = A_B^{-1}b - A_B^{-1}A_N x_N^k \underset{k\in\mathbb{N}}{\longrightarrow} A_B^{-1}b - A_B^{-1}A_N x_N^* = x_B^*,$$

which proves the convergence of the entire sequence.
(ii) Suppose, by a contradiction way, that the basis changes are infinite. Then, there would exist an index $i_0 \in \{1,\ldots,n\}$ that goes in and out of the generated bases infinitely many times. This would imply the existence of a subsequence $(x^k)_{k\in K}$ such that $x_{i_0}^k = 0$ for all $k \in K$. But the whole sequence $(x^k)_k$ is assumed to be convergent, say to x^*. Then, we would obtain that $x_{i_0}^k \underset{k\in\mathbb{N}}{\longrightarrow} x_{i_0}^* = 0$. Since by Theorem 4.1, $(x^k)_k$ is feasible, its limit x^* feasible too. Now, by nondegeneracy assumption on S, there exists a basis B such that $x_B^* > 0$. This would imply that

$$\exists k_0 \in \mathbb{N}, \quad \forall k \geq k_0, \qquad x_i^k > x_{i_0}^k \geq 0 \qquad \forall i \in B.$$

Clearly, by the special basis change strategy, this means that the index i_0 would never enter in any basis after the iteration k_0, which contradicts the starting hypothesis. This finally proves the finiteness of the basis changes that favors the largest variables. □

				FC-RJM			RJM			W-RGM		
Pb	Ref	m	n	Iter	Feval	Beval	Iter	Feval	Beval	Iter	Feval	Beval
EL1	[14]	1	2	0.66	4.04	**0**	**0.595**	**3.72**	**0**	4.115	38.475	0.33
EL2	[14]	2	6	**0.995**	**1.99**	0.555	**0.995**	**1.99**	**0.55**	2.025	3.03	1.53
Constr	[9]	2	6	**1**	**2**	0.52	**1**	**2**	**0.415**	3.44	15.62	1.345
Hanne2	[7]	2	5	**1**	**2**	0.165	**1**	**2**	**0.135**	8.815	92.33	0.33
Hanne3	[7]	2	5	**1**	**2**	**0.195**	**1**	**2**	0.2	2	3	0.535
GBG	[20]	2	7	1.77	3.525	**0.455**	**1.75**	**3.485**	0.495	2.245	3.525	1.655
TLK1	[31]	2	8	**11.14**	**56.57**	**0.55**	11.5	60.42	0.57	13.185	110.3	0.8
OF	[27]	2	5	**0.325**	**0.65**	**0.125**	**0.325**	**0.65**	**0.125**	5.02	32.845	1.12
PF	[29]	4	5	5.385	27.33	**0.28**	**4.31**	**19.095**	0.31	9.46	41.765	0.77
MOP3	[21]	2	4	3.835	26.055	0.15	**2.78**	**14.15**	**0.14**	3.475	18.43	0.38
SK2	[21]	4	8	**2.535**	**3.585**	**0.89**	3.415	4.485	**0.89**	4.425	10.28	1.635
TRS	[8]	12	26	**9.79**	**10.79**	**8.215**	12.43	13.43	8.415	14.94	15.94	11.25
JOS1	[21]	50	100	1.98	8.485	**0**	**1.9**	8.185	**0**	3.075	**7.16**	0.88
		100	200	**1.42**	**5.245**	**0**	1.86	7.06	**0**	2.535	6.22	0.7
		200	400	**1.25**	**4.43**	**0**	1.89	7.465	**0**	2.45	5.7	0.73
ZDT1	[39]	50	100	1.81	3.39	**0**	**1.26**	**2.59**	**0**	24.735	35.695	20.25
		100	200	2.405	3.665	**0**	**1.125**	**2.28**	**0**	45.245	60.445	40
		200	400	4.165	5.245	**0**	**1.085**	**2.175**	**0**	166.82	176.55	164
ZDT2	[39]	50	100	1.135	2.34	**0**	**1**	**2**	**0**	28.435	29.435	27.515
		100	200	1.34	2.34	**0**	**1**	**2**	**0**	60.32	61.32	59.51
		200	400	1.945	2.945	**0**	**1**	**2**	**0**	165.35	166.35	165
ZDT3	[39]	50	100	3.135	9.775	**0**	**2.26**	**8.635**	**0**	593.88	971.676	47.64
		100	200	3.525	9.435	**0**	**2.61**	**8.84**	**0**	721.45	1054.941	53.98
		200	400	9.25	19.5	**0**	**4.25**	**16.3125**	**0**	841.23	1353.22	67.8

Table 1: Standard performance measurements.

Remark 4.3 It has been shown in [14] that if the vector objective F is pseudoconvex (resp. strictly pseudoconvex) on S, then any Pareto KKT-stationary point of the MOP is a weakly efficient (resp. efficient) point. On the other hand, it is well known that in the convex case, any Pareto KKT-stationary point associated with a positive vector Lagrange multiplier, say λ^*, is actually properly efficient (see [12], e.g.). Furthermore, it is worth noting, according to the proofs of Proposition 3.1(iii) and Theorem 4.1(ii), that λ^* is not other than an accumulation point of the sequence $\lambda(x^k)$ generated by the FC-RJM algorithm. Hence, under these considerations and the assumptions of the previous theorem, the limit of the iterates x^k by the variant FC-RJM will belong to the corresponding type of efficient set.

5 Numerical experiments

In this section, experiments carried out on the proposed FC-RJM method are described. This includes comparisons with the original RJM version [14], the weighting method of scalarization denoted in [14] by W-RGM, and the evolutionary NSGA-II method [10]. All codes are implemented in MATLAB R2021a on a machine equipped with 1.90 GHz Intel(R) Core(TM) i5 CPU and 16 Go memory.

Sixteen well-known test MOPs with linear constraints are experimented (see Table 1). In order to graphically investigate the approximated Pareto fronts, we restricted ourselves to bi-criteria and three-criteria problems that we think are sufficient to reflect essential aspects of the compared methods. All these problems were solved with the same selected initial population of 200 individuals (starting points). With the exception of NSGA-II which can give less than 200 approximate non-dominated solutions, the FC-RJM, RJM

and W-RGM iterative descent methods necessarily provide the same number as the initial population. To make a correct comparison, we have set "PopulationSize" equal to 571 for NSGA-II, in order to get the same number 200 of final solutions for each problem solved. All other settings of NSGA-II were used as predefined by the solver in MATLAB. For the reduced methods, each feasible polyhedron must be set under the standard form: $Ax = b,\ x \geq 0$. After a basis B is selected, nonbasic coordinates x_N are randomly chosen between a positive lower bound and an upper bound of the feasible polyhedron. The vector basic variable is of course computed according to the relationship $x_B = A_B^{-1}b - A_B^{-1}A_N x_N$, so that $Ax = b$, and while $x_B \not> 0$, the initialization procedure is repeated.

The W-RGM method consists to apply Wolfe's continuous RGM scheme (4) to the scalarized problem:

$$\underset{x \in S}{\text{Min}}\ \lambda \cdot F(x),$$

where the weighting vector of scalarization $\lambda = (\lambda_1, \ldots, \lambda_r) \in \Lambda$ are of course chosen at random. In the bi-criteria case, for example, each qth individual, $q = 1, \ldots, 200$, can have a different weight chosen at random in an equidistant partition of $[0, 1]$, namely, $\lambda_1 \in \left[\frac{q-1}{200}, \frac{q}{200}\right]$, and obviously, $\lambda_2 = 1 - \lambda_1$.

During the numerical experiments, we prescribed $\beta = 0.25$ as Armijo's constant, and we started the Armijo procedure with the initial guess $t_0 = \overline{t}$. The finding-direction programs for FC-RJM and RJM were in turn solved by RGM schemes (7) and (4) respectively. It should be noted that these two problems are convex differentiable having as many variables as the number of objectives and only one linear equality constraint, so that, a nondegenerate guess in their feasible set Λ is trivial.

During running, the stopping criterion was: $\|d_N\| < 10^{-5}$.

We started by analysing the standard measures of computational effort that are specific to descent methods, namely, the number of iterations, the number of evaluations of vector functions and the number of basis changes, denoted here respectively by: "Iter ", "Feval" and "Beval". Table 1 summarizes the results obtained by FC-RJM, RJM, and W-RGM on the sixteen test problems considered, where "Pb", "Ref", "m" and "n" mean respectively the name of the test problem, its reference, the number of its original variables and the number of its variables with standard form. The values presented are the average number per population of 200 individuals. The best scores are in bold.

According to Tab.1, one can observe that the reduced Jacobian methods FC-RJM and RJM give the solutions within only a few iterations, and with an equally reasonable number of function evaluations. Therefore, this implies a relatively low cost of computation of steplengths, while noting a slight advantage of the RJM over FC-RJM with regard to both number of iterations and function evaluations. One can also see that the two reduced Jacobian approaches perform, on average, generally less than a single basis change, except for the instance TRS. It is also worth noting that both FC-RJM and RJM work quite well on large size issues, and they are not very sensitive to increased dimension n, as we can see it with JOS1 and the ZDT instances. Comparing these results with those of W-RGM, show that W-RGM spent in both number of iterations, function evaluations and basis changes, especially, for the ZDT problems. Note also that W-RGM did not completely recover the Pareto front even in the convex case, as clearly seen in Fig.1 with the Constr problem. Moreover, from all figures corresponding to the nonconvex cases, including the pseudoconvex problem EL2, we see that W-RGM has found only the extreme efficient points.

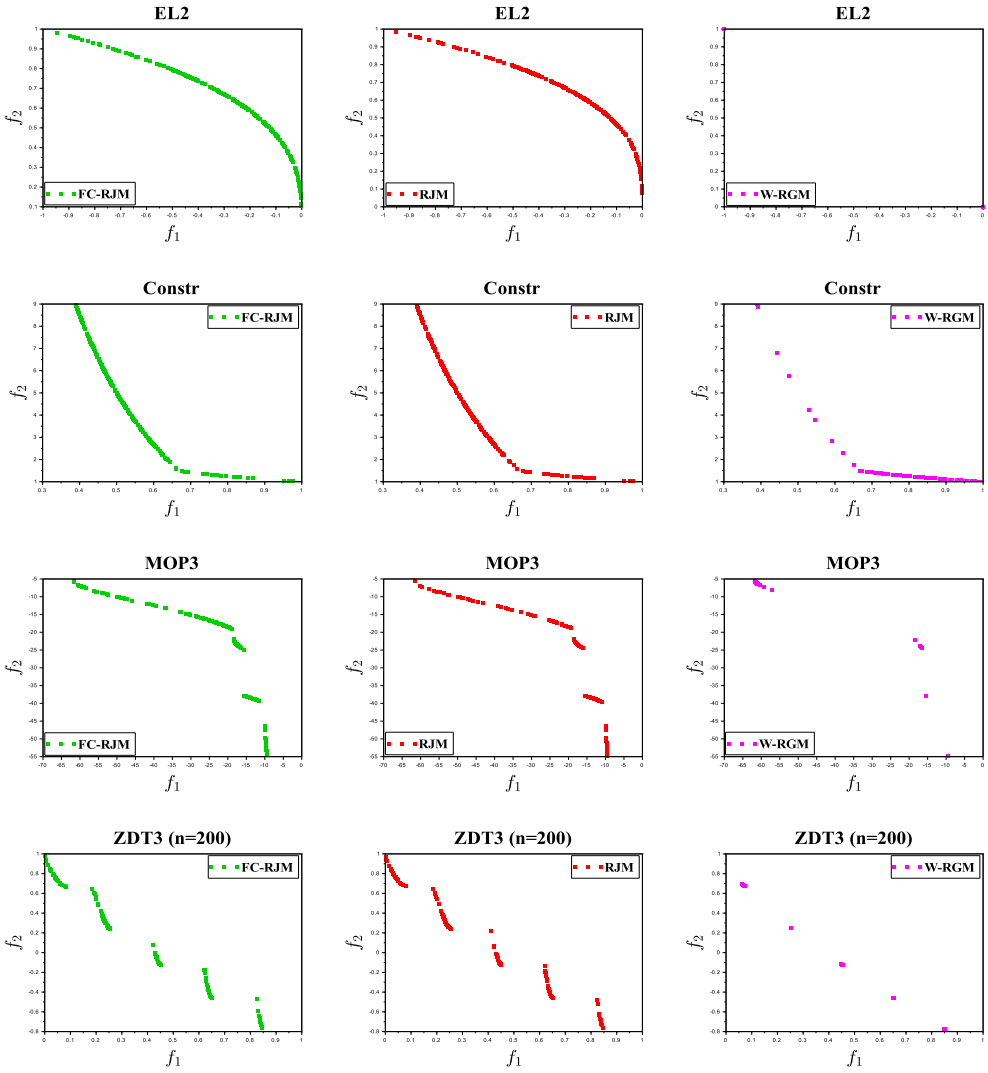

Figure 1: Pareto fronts of some test problems by FC-RJM, RJM and W-RGM.

In order to analyse the graphical representations and to have convenient comparisons between the methods, two essential factors are considered: the convergence of approximate sets towards the true Pareto front, and the diversity of solutions. These two tasks can be measured with many performance metrics suggested in the literature. Here, we have opted for the three performance measures: Purity metric (P) [1], Hypervolume metric (HV) [40], and the Generational distance (GD) [32]. Since the true Pareto front may not be readily available in real-life problems, these measures are rather defined with the help of the so-called reference Pareto front, which is not other than a concatenation of the non-dominated solutions provided by all the competing methods, of course, after eliminating the extra-dominated points. More precisely, if we denote by $\mathcal{S}$ the set of considered solvers and $\mathcal{P}$ the set of tested problems, and if we call $F_{p,s}$ the approximated Pareto front of problem $p \in \mathcal{P}$ obtained by method $s \in \mathcal{S}$, then the reference Pareto front associated to the problem p is given by

$$F_p = \left\{ y^* \in \bigcup_{s\in\mathcal{S}} F_{p,s} \ : \ \nexists y \in \bigcup_{s\in\mathcal{S}} F_{p,s}, \quad y < y^* \right\}.$$

– **Purity** (P). This metric consists in measuring the proportion of $F_{p,s}$ points admitted in F_p:

$$P_{p,s} = \frac{|F_{p,s} \cap F_p|}{|F_{p,s}|}.$$

Clearly $P_{p,s} \in [0,1]$ and the extreme values are significant in the sense that a value $P_{p,s}$ close to 1 indicates better performance, while a value equal to 0 implies that the algorithm is unable to generate any point of F_p.

– **Hypervolume** (HV). The hypervolume tries to estimate the size of the feasible region dominated by $F_{p,s}$. For example, in the case of a bicriteria problem, this corresponds to the area delimited by $F_{p,s}$ and a reference point. Often the latter is taken as the vector whose ith component is the maximum value of the ith objective observed for the given problem. Thus, the higher the hypervolume of solver s for problem p, the better the dispersion of the solutions in $F_{p,s}$.

– **Generational Distance** (GD). Measuring the convergence, this metric represents how far $F_{p,s}$ is from F_p:

$$GD_{p,s} = \frac{1}{|F_{p,s}|}\sqrt{\sum_{i=1}^{|F_{p,s}|} d_i^2}$$

where d_i is the minimum distance between the ith solution in $F_{p,s}$ and the reference Pareto front F_p. Obviously, lower values of $GD_{p,s}$ are requested.

Table 2 summarizes the results of the performance measures realized by FC-RJM, RJM, W-RGM and NSGA-II on the sixteen test problems, in addition to the CPU times (in sec). To simplify the interpretation of these results and to be able to analyse the overall performance of each solver on all the test problems, then look for significant differences between them, the performance profile is used (see, e.g., [38]). Recall that a performance profile is the graphical representation of the (cumulative) distribution function of measures obtained by a solver on a set of problem tests with respect to a performance metric. More precisely, given a performance measure $m_{p,s}$ by solver s for solving p, and the performance ratio

$$r_{p,s} := \frac{m_{p,s}}{\min\limits_{s \in \mathcal{S}} m_{p,s}}$$

with respect to a decreasing metric[1], the distribution function associated to s is given by

$$\rho_s(\alpha) = \frac{\big|\{p \in \mathcal{P} \ : \ r_{p,s} \leq \alpha\}\big|}{|\mathcal{P}|}.$$

Note that at the threshold $\alpha = 1$, $\rho_s(\alpha)$ gives us the largest number of problems among the best solved by s according to the analysed performance. However, a value $\rho_s(\alpha)$ attaining 1 means that all the problems $p \in \mathcal{P}$ have been solved by solver s at the threshold α. Thus, the best overall performance of a solver is that which reaches the value 1 for the smallest value of α.

[1]Here, the metric $m_{p,s}$ may be $1/P$, $1/HV$, GD or CPU, so that, all these metrics have the same asymptotic behaviour in a sense that the smaller the measure, the better the solver.

	FC-RJM				RJM			
Pbs	CPU	P	HV	GD	CPU	P	HV	GD
EL1	**0.0188**	0.985	0.582282	0.0000005	0.0248	0.985	0.5823362	0.0000005
EL2	0.0331	0.995	0.2442221	0.0004188	0.0403	0.995	0.2442227	0.0004188
Constr	0.0451	**1**	3.8048019	**0**	0.0588	**1**	3.805181	**0**
Hanne2	0.0116	**1**	8672.2525	**0**	0.0173	**1**	8671.8278	**0**
Hanne3	0.0121	**1**	4.0376345	**0**	0.0160	**1**	4.0581354	**0**
GBG	0.0110	1	35.101078	**0**	0.0143	1	34.789304	**0**
TLK1	0.0595	0.235	0.2904914	0.0001144	0.0801	0.21	0.2937795	0.0001138
OF	**0.0168**	**1**	31.21749	**0**	0.0213	**1**	31.217491	**0**
PF	0.0441	0.855	0.0021093	0.0000003	0.0483	0.86	0.0021105	0.0000003
MOP3	0.2889	0.995	**560.13286**	0.0000012	0.2638	0.99	558.92114	**0**
SK2	0.0337	**1**	10.0984	**0**	0.0591	0.97	9.7412008	2.112D-08
TRS	0.1511	**1**	**4093.9733**	**0**	0.2233	**1**	4093.3217	**0**
JOS1	0.0276	0.89	1.6601522	**4.305D-08**	0.0331	0.915	1.6600265	0.0000002
	0.0257	0.945	1.6840092	**0.0000011**	0.0346	0.73	1.6848459	0.0000228
	0.0430	**0.975**	1.716233	**0.0000005**	0.0585	0.75	1.7158919	0.0000394
ZDT1	0.0720	**1**	**0.9829638**	**0**	0.0387	**1**	0.9828619	**0**
	0.1343	**1**	**2.7570317**	**0**	0.0434	**1**	2.7557124	**0**
	0.2557	**1**	**5.1332833**	**0**	0.0634	**1**	5.1326993	**0**
ZDT2	0.0297	**1**	0.7416248	**0**	0.0348	**1**	**0.7426255**	**0**
	0.0478	**1**	0.7078539	**0**	0.0451	**1**	**0.7080764**	**0**
	0.1273	**1**	**1.5383919**	**0**	**0.0854**	**1**	1.5380217	**0**
ZDT3	0.2909	**0.665**	2.248064	**0.0105172**	0.1346	0.65	**2.249226**	0.011813
	0.9059	**0.755**	**2.285064**	**0.0107885**	0.3356	0.705	2.251236	0.0189156
	1.9823	0.755	**2.287016**	**0.0099565**	1.3566	**0.85**	2.281667	0.0189157

	W-RGM				NSGA-II			
Pbs	CPU	P	HV	GD	CPU	P	HV	GD
EL1	0.0223	0.985	0.5633661	**0.0000001**	0.0366	**0.99**	**0.5827026**	0.0000003
EL2	**0.0121**	**1**	1.508D-10	**0**	2.0116	0.805	**0.2454173**	0.0000162
Constr	**0.0153**	**1**	3.6466906	**0**	2.7901	0.615	**3.8055397**	0.0000882
Hanne2	**0.0103**	**1**	**8676.4909**	**0**	0.5548	0.645	8675.526	0.0006059
Hanne3	**0.0043**	**1**	3.0455804	**0**	0.5632	0.755	**4.1011699**	0.0000886
GBG	**0.0046**	**1**	0.0792	**0**	0.5396	0.77	**35.551297**	0.0003792
TLK1	**0.0418**	0.155	0.2967715	0.0001436	1.7514	**1**	**0.2986474**	**0**
OF	0.0176	0.95	31.527816	**0**	1.4782	0.81	**31.600599**	0.0000829
PF	**0.0271**	**0.94**	0.0021371	**0.0000002**	0.9598	0.445	**0.0021496**	0.0000089
MOP3	0.0303	**1**	356.21129	**0**	**0.0234**	0.875	558.92114	0.000615
SK2	**0.0198**	**1**	8.9025976	**0**	0.0297	0.785	**10.944865**	0.0001213
TRS	**0.0857**	**1**	3578.0001	**0**	1.6461	0	1799.8796	0.4927958
JOS1	**0.0244**	**0.995**	**2.1388898**	0.0000004	0.0435	0.495	2.0123538	0.000119
	0.0603	**0.98**	**2.1436591**	0.0000182	0.0581	0.33	2.1164373	0.0001171
	0.3903	**0.975**	**2.1113116**	0.0000177	0.0759	0.345	2.0682378	0.0001069
ZDT1	0.4637	**1**	0.8953741	**0**	**0.0281**	0.04	0.9548326	0.0005029
	4.5067	**1**	2.4190441	**0**	**0.0352**	0.045	2.727446	0.0011227
	131.7077	**1**	4.5152886	**0**	**0.0582**	0	5.1018942	0.0156109
ZDT2	0.6056	**1**	0.4169525	**0**	**0.0296**	0	0.5934084	0.0028143
	7.7111	**1**	0.3811394	**0**	**0.0423**	0	0.6151129	0.001585
	152.7394	**1**	1.2119	**0**	0.1079	0	1.4488972	0.0027216
ZDT3	0.7346	0.26	2.10378	0.012013	**0.0222**	0	2.237751	0.0171232
	0.8386	0.213	2.10378	0.015037	**0.0227**	0	2.257581	0.0156127
	2.7684	0.113	2.10375	0.0133765	**0.1245**	0	2.256591	0.0154432

Table 2: Multiobjective performance measurements.

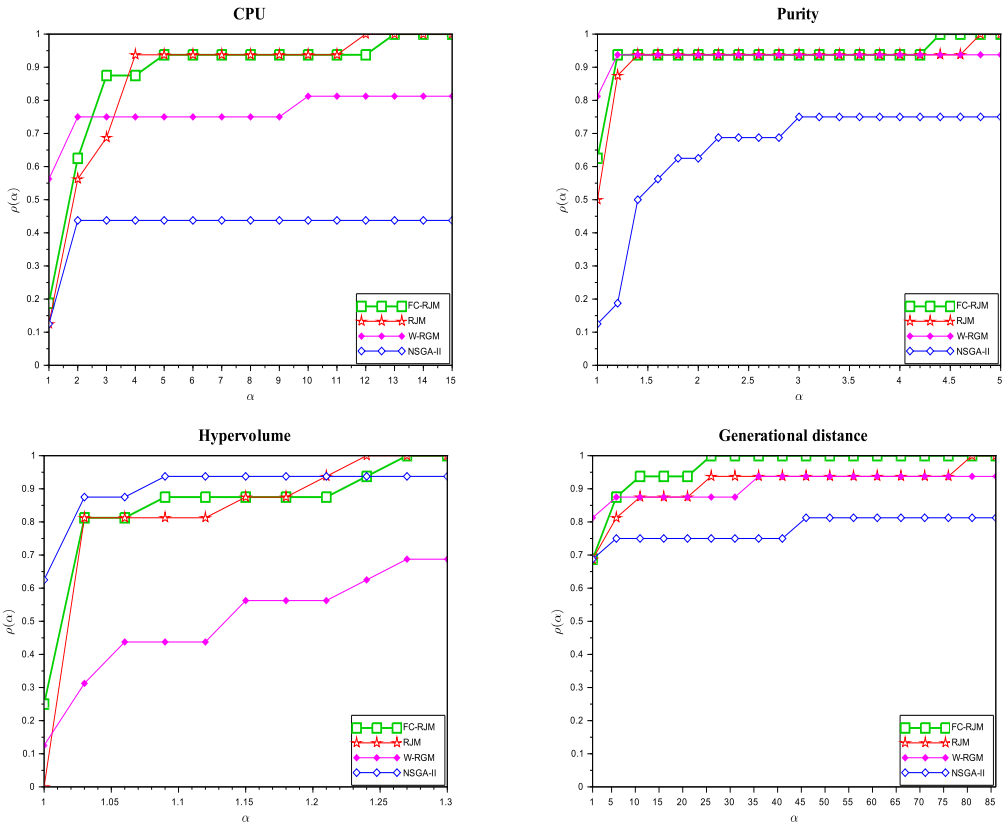

Figure 2: Performance profiles.

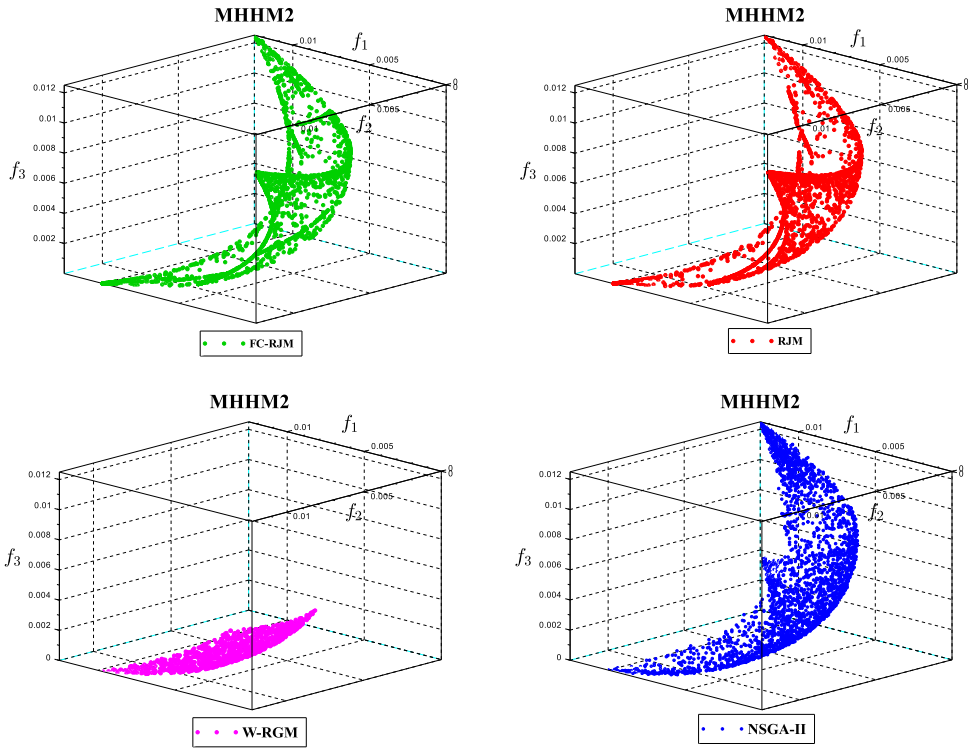

Figure 3: Pareto fronts of the 3-criteria test problem MHHM2 [23] by FC-RJM, RJM, W-RJM and NSGA-II with an initial population of 10000 individuals.

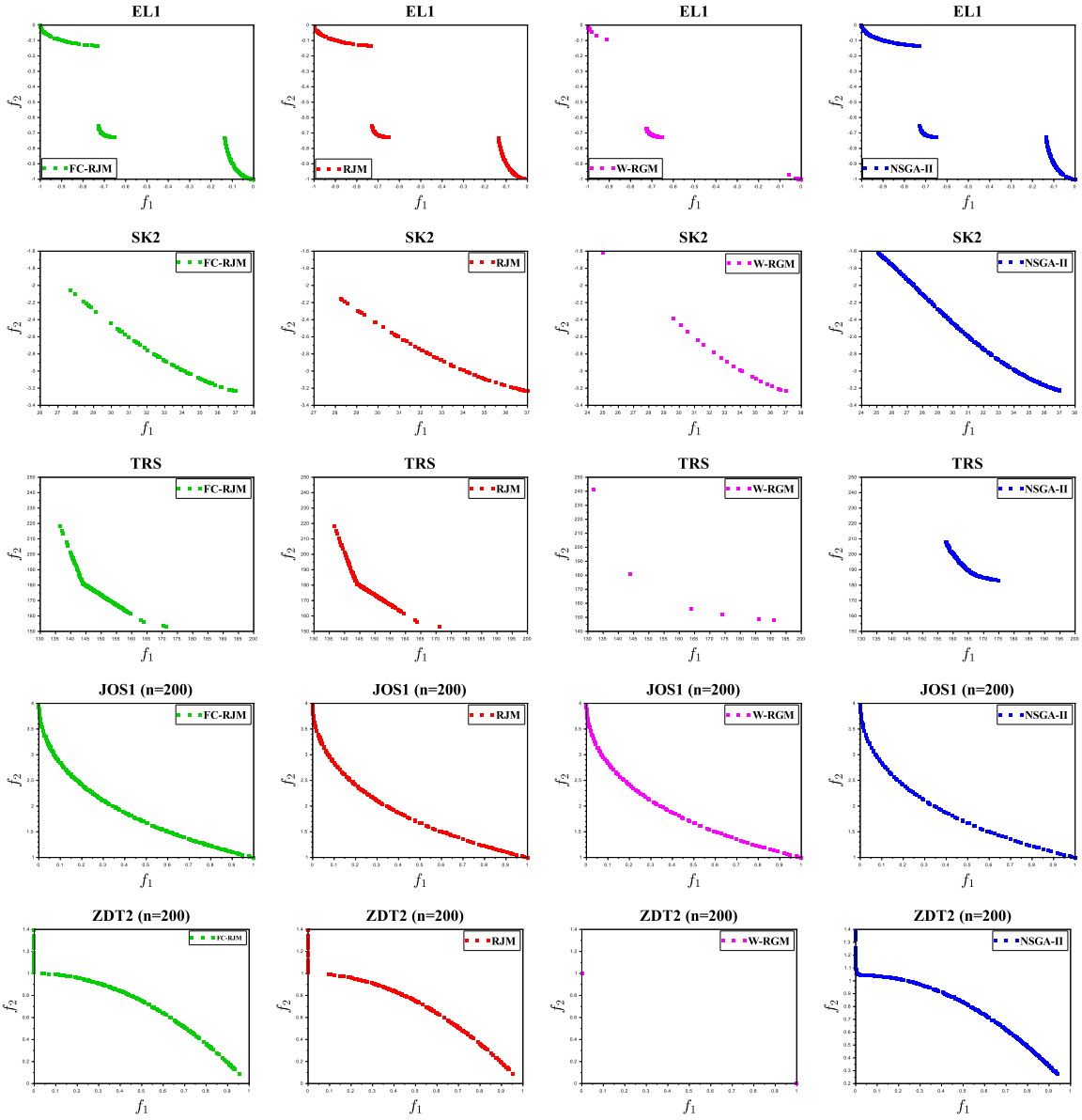

Figure 4: Pareto fronts of some test problems by FC-RJM, RJM, W-RGM and NSGA-II.

According to Figure 2, FC-RJM and RJM obtain overall good scores in both CPU, P, HV and GD compared to W-RGM and NSGA-II. Indeed, we can see that the best value $\rho(\alpha) = 1$ could be reached at a fairly small threshold α. For example, in terms of convergence, we can clearly see from Fig.2 that the performance profiles P and GD indicate that NSGA-II as well as W-RGM could not globally reach the reference front for the set of the test problems, especially, because of TLK1, PF, TRS, JOS1, ZDT1, ZDT2 and ZDT3 (see Tab.2 for more precisions). In contrast, the observed HV dispersion profile shows that FC-RJM and RJM achieve a slight superiority against NSGA-II, although this latter being generally efficient with regard to this metric. Finally, comparing FC-RJM and RJM, it can be also seen from Fig.2 that the new variant proposed in this paper clearly exceeds its rival RJM in terms of the convergence indicators GD and P, noting however a slight advantage of RJM in CPU and HV.

As we know, the graphical representation of approximated Pareto fronts does not reflect the real landscape of the true fronts. However, those exposed in Fig.4, for instance, confirms our interpretations made on the new FC-RJM about its ability together with RJM

and NSGA-II to better explore Pareto fronts even in non-convex cases, and to handle some serious difficulties, such as discontinuity or nonuniform density. Unlike W-RGM where we must underline the difficulty of the classical scalarizing approach of being able to identify the Pareto fronts, especially, in the non-convex cases.

References

[1] S. Bandyopadhyay, S.K. Pal, and B. Aruna, *Multiobjective GAs, quantative indices, and pattern classification*, IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics 34 (2004), pp. 2088–2099.

[2] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, New York, 2006.

[3] J.Y. Bello-Cruz, L.R. Lucambio Prez, and J.G. Melo, *Convergence of the projected gradient method for quasiconvex multiobjective optimization*, Nonlinear Analysis 74 (2011), pp. 5268–5273.

[4] J. Benoist, J.M. Borwein, and N. Popovici, *A Characterization of quasiconvex vector-valued functions*, Proceedings of the American Mathematical Society 131 (2003), pp. 1109–1113.

[5] G.C. Bento, J.X. Cruz Neto, P.R. Oliveira, and A. Soubeyran, *The self regulation problem as an inexact steepest descent method for multicriteria optimization*, European Journal of Operational Research 235 (2014), pp. 494–502.

[6] R. Burachik, L.M. Graa Drummond, A.N. Iusem, and B.F. Svaiter, *Full convergence of the steepest descent method with inexact line searches*, Optimization 32 (1995), pp. 137–146.

[7] Y. Collette, and P. Siarry, *Optimisation Multiobjectif*, Eyrolles, Paris, 2002.

[8] S.K. Das, A. Goswami, and S.S. Alam, *Multiobjective transportation problem with interval cost, source and destination parameters*, European Journal of Operational Research 117 (1999), pp. 100–112.

[9] K. Deb, *Multi-objective Optimization Using Evolutionary Algorithms*, Wiley, Chichester, 2001.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, *A fast and elitist multiobjective genetic algorithm: NSGA-II*, IEEE Transactions on Evolutionary Computation 6 (2002), pp. 182–197.

[11] L.M.G. Drummond, and A.N. Iusem, *A projected gradient method for vector optimization problems*, Computational Optimization and Applications 28 (2004), pp. 5–29.

[12] M. Ehrgott, *Multicriteria Optimization*, Springer, Berlin, 2005.

[13] M. El Maghri, and Y. Elboulqe, *Correction to : Reduced Jacobian method*, Journal of Optimization Theory and Applications 187 (2020), pp. 304.

[14] M. El Maghri, and Y. Elboulqe, *Reduced Jacobian method*, Journal of Optimization Theory and Applications 179 (2018), pp. 917–943.

[15] M. El Maghri, *A propos de la différentiabilité*, Quadrature 109 (2018), pp. 23–25.

[16] M. El Maghri, *A free reduced gradient scheme*, Asian Journal of Mathematics and Computer Research 9 (2016), pp. 228–239.

[17] J. Fliege, L.M.G. Drummond, and B.F. Svaiter, *Newton's method for multiobjective optimisation*, SIAM Journal on Optimization 20 (2009), pp. 602–626.

[18] J. Fliege, and B.F. Svaiter, *Steepest descent methods for multicriteria optimization*, Mathematical Methods of Operations Research 51 (2000), pp. 479–494.

[19] E.H. Fukuda, and L.M.G. Drummond, *On the convergence of the projected gradient method for vector optimization*, Optimization 60 (2011), pp. 1009–1021.

[20] U.M. García-Palomares, J.C. Burguillo-Rial, and F.J. González-Castaño, *Explicit gradient information in multiobjective optimization*, Operation Research Letters 36 (2008), pp. 722–725.

[21] S. Huband, P. Hingston, L. Barone, and L. While, *A review of multiobjective test problems and a scalable test problem toolkit*, IEEE Transactions on Evolutionary Computation 10 (2006), pp. 477–506.

[22] D.G. Luenberger, and Y. Ye, *Linear and Nonlinear Programming*, 3rd Edition, Springer, New York, 2008.

[23] J. Mao, K. Hirasawa, J. Hu, and J. Murata, *Genetic symbiosis algorithm for multiobjective optimization problem*, in Proc. IEEE Int. Workshop on Robot and Human Interactive Commun. (2000), pp. 137–142.

[24] H. Mokhtar-Kharroubi, *Sur la convergence théorique de la méthode du gradient réduit généralisé*, Numerische Mathematik 31 (1980), pp. 13–35.

[25] V. Morovati, and L. Pourkarimi, *Extension of Zoutendijk method for solving constrained multiobjective optimization problems*, European Journal of Operational Research 273 (2019), pp. 44–57.

[26] H. Mukai, *Algorithms for multicriterion optimization*, IEEE Transactions on Automatic Control 2 (1980), pp. 177–186.

[27] S.L.C. Oliveira, and P.A.V. Ferreira, *Bi-objective optimization with multiple decision makers: A convex approach to attain majority solutions*, Journal of the Operational Research Society 51 (2000), pp. 333–340.

[28] S. Qu, M. Goh, and F.T.S. Chand, *Quasi-Newton methods for solving multiobjective optimization*, Operation Research Letters 39 (2011), pp. 397–399.

[29] A. Radhakrishnan, *Evolutionary algorithms for multiobjective optimization with applications in portfolio optimization*, Master's thesis, North Carolina State University, 2007.

[30] Y. Smeers, *Generalized reduced gradient method as an extension of feasible direction methods*, Journal of Optimization Theory and Applications 22 (1977), pp. 209–226.

[31] T.N. Thang, D.T. Luc, and N.T.B. Kim, *Solving generalized convex multiobjective programming problems by a normal direction method*, Optimization 65 (2016), pp. 2269–2292.

[32] D.A. Van Veldhuizen, and G.B. Lamont, *On measuring multiobjective evolutionary algorithm performance*, in Congress on Evolutionary Computation (CEC2000), A. Zalzala and R. Eberhart, Eds. Piscataway, NJ, IEEE Press 1 (2000), pp. 204–211.

[33] R. Viennet, C. Fonteix, and I. Marc, *Multicriteria optimization using a genetic algorithm for determining a Pareto set*, International Journal of Systems Science 27 (1996), pp. 255–260.

[34] C.Y. Wang, *On convergence property of an improved reduced gradient method*, Kexue Tongbao 28 (1983), pp. 577–582.

[35] C.Y. Wang, *Convergence characterizations of both new pivote method and simplified Levitin–Polyak gradient projection method*, Acta Mathematicae Applicatae Sinica 4 (1981), pp. 37–52, in Chinese.

[36] P. Wolfe, *Methods of nonlinear programming*, in Recent Advances in Mathematical Programming, R.L. Graves, P. Wolfe Eds., McGraw-Hill, 1963, pp. 76–77.

[37] M. Yue, and J. Han, *A new reduced gradient method*, Scientia Sinica 22 (1979), pp. 1099–1113.

[38] E. Zitzler, J. Knowles, and L. Thiele, *Quality assessment of Pareto set approximations*, in Multiobjective Optimization, J. Branke, K. Deb, K. Miettinen, and R. Sowiski, Eds., Springer, Berlin, Lecture Notes in Computer Science 5252 (2008), pp. 373–404.

[39] E. Zitzler, K. Deb K, and L. Thiele, *Comparison of multiobjective evolutionary algorithms: empirical results*, Evolutionary Computation 8 (2000), pp. 173–195.

[40] E. Zitzler, L. Thiele, *Multiobjective evolutionary algorithms : A comparative case study and the strength Pareto approach*, IEEE Transaction. Evolutionary Computation 3 (1999), pp. 257–271.