A DERIVATIVE-FREE APPROACH TO CONSTRAINED MULTIOBJECTIVE NONSMOOTH OPTIMIZATION[*]

G. LIUZZI[†], S. LUCIDI[‡], AND F. RINALDI[§]

Abstract. In this work, we consider multiobjective optimization problems with both bound constraints on the variables and general nonlinear constraints, where objective and constraint function values can only be obtained by querying a black box. We define a linesearch-based solution method, and we show that it converges to a set of Pareto stationary points. To this aim, we carry out a theoretical analysis of the problem by only assuming Lipschitz continuity of the functions; more specifically, we give new optimality conditions that take explicitly into account the bound constraints, and prove that the original problem is equivalent to a bound constrained problem obtained by penalizing the nonlinear constraints with an exact merit function. Finally, we present the results of some numerical experiments on bound constrained and nonlinearly constrained problems, showing that our approach is promising when compared to a state-of-the-art method from the literature.

Key words. derivative-free multiobjective optimization, Lipschitz optimization, inequality constraints, exact penalty functions

AMS subject classifications. 90C30, 90C56, 65K05, 49J52

DOI. 10.1137/15M1037810

1. Introduction. Many real-world problems can be modeled as the minimization (or maximization) of multiple objective functions, usually conflicting with one another, over a set of constraints. We have a number of relevant examples coming from many fields of science, including engineering, economics, and logistics, where optimal decisions need to be taken in the presence of trade-offs between two or more conflicting objectives (see, e.g., [5] and references therein). When dealing with functions whose derivatives are expensive, unreliable, or impossible to calculate, derivative-free methods (see [4] for a complete overview of derivative-free methods) need to be used in order to get a good solution for the given problem. Methods for multiobjective optimization (MOO) can be classified into two different classes with respect to the moment when preferences relating the objectives are established. This way of classifying MOO approaches also holds in the derivative-free context (see, e.g., [7, 6]), where we have the following:

- *Methods with a priori articulation of preferences*, where objective functions are combined into a single one with a proper aggregation criterion before the optimization starts (see, e.g., [1], [20]). In this case, since the original problem is transformed into a single-objective problem, we will get a single nondominated point.
- *Methods with a posteriori articulation of preferences*, which try to reconstruct the whole Pareto front for the MOO problem under analysis. In this class we have both globally convergent derivative-free methods, like, e.g., direct-

[†]Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti," Via dei Taurini 19, 00185 Rome, Italy (giampaolo.liuzzi@iasi.cnr.it).

[‡]"Sapienza" Università di Roma, Dipartimento di Ingegneria Informatica Automatica e Gestionale "A. Ruberti," Via Ariosto 25, 00185 Rome, Italy (lucidi@dis.uniroma1.it).

[§]Università di Padova, Dipartimento di Matematica, Via Trieste 63, 35121 Padua, Italy (rinaldi@math.unipd.it).

search-type methods [7], and heuristics, like, e.g., genetic algorithms [8] and simulated annealing [24].

In this paper, we are interested in developing new globally convergent derivative-free methods with a posteriori articulation of preferences for nonlinearly constrained multiobjective minimization problems of the following form:

(1.1)
$$\begin{aligned} \min \quad & F(x) = (f_1(x), \ldots, f_q(x))^\top \\ \text{s.t.} \quad & g(x) \leq 0, \\ & l \leq x \leq u, \end{aligned}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, q$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $l, u \in \mathbb{R}^n$ with $l < u$. We denote by X the set defined by simple bounds on the variables, that is,

$$X = \{x \in \mathbb{R}^n : l \leq x \leq u\},$$

and by $\mathcal{F}$ the feasible set of problem (1.1), namely,

$$\mathcal{F} = \{x \in \mathbb{R}^n : g(x) \leq 0\} \cap X.$$

We note that, by definition, X is a compact set.

Into this context, the direct multisearch (DMS) method proposed in [7] represents an appealing method since it extends, from single to multiobjective optimization, classic directional derivative-free methods called direct search [4]. Two different steps characterize the DMS method, namely, the *search step* and the *poll step*. Once a current iterate (a poll center) has been selected from a list of (feasible) nondominated points, the poll step evaluates the objective functions at some neighboring points defined by a positive spanning set and a step size parameter. Then it uses an acceptance criterion based on the concept of Pareto dominance for selecting the new iterates. Finally, the list of nondominated points is updated by using those points generated at the current iteration. The search step is included to further spread the search of points belonging to the Pareto front. The handling of the nonlinear constraints is done by means of an extreme barrier function approach. In practice, the following function

$$\bar{F}(x) = \begin{cases} F(x) & \text{if } x \text{ is feasible,} \\ (+\infty, \ldots, +\infty)^\top & \text{otherwise,} \end{cases}$$

is minimized over $\mathbb{R}^n$. Hence, when a given point is infeasible, the extreme barrier does not evaluate the objective function $F(x)$ and sets the values of $\bar{F}(x)$ to $+\infty$. Thus, a feasible starting point is always needed in an extreme barrier framework. This can be problematic when dealing with real-world problems, since a feasible starting point is not always available, and obtaining such a point can be a very time-consuming task. Furthermore, even though in principle the function $\bar{F}(x)$ can be used in the presence of constraints, in many situations the constraints can be managed in a more efficient way, i.e., when the amount of violation can be quantified.

Inspired by the ideas in [10] and [7], we describe a new exact-penalty-based line-search approach (with sufficient decrease) for nonlinearly constrained MOO problems. In order to study the convergence properties of the proposed method, we carry out a preliminary theoretical analysis of the problem itself. We describe new optimality conditions that take explicitly into account the bound constraints and that are obtained by only assuming Lipschitz continuity of the problem functions. We also prove that the original problem is equivalent to a bound constrained problem obtained by

penalizing the nonlinear constraints with an exact merit function. In particular, we introduce a merit function that penalizes the general nonlinear inequality constraints in each term of $F(x)$ and we resort to the minimization of a penalty function subject to the simple bound constraints. We would like to point out that some exact penalty methods for MOO have already been proposed in the literature (see, e.g., [11, 14] and references therein). Anyway, to the best of our knowledge, this is the first time that a penalty approach with explicit handling of the bound constraints for Lipschitz continuous multiobjective problems is studied. The proposed approach enables us to handle also infeasible starting points. Furthermore, thanks to this exact penalty, we can adapt the derivative-free approach in [10] to the MOO case. This new approach gives us three relevant advantages:

- by means of the sufficient decrease we can avoid the use of integer lattices (see, e.g., [7, 10]);
- the extrapolation phase allows us to better exploit a descent direction and hence to prove, under some density assumptions on the search directions, convergence to a set of Pareto stationary points (i.e., we prove that any accumulation point of the sequences generated by our method is a Pareto stationary point);
- thanks to the exact penalty approach, the starting point can be infeasible.

The fact that any accumulation point of the generated sequences is a Pareto stationary point, is an interesting theoretical result, since it is slightly stronger than the results reported in [7]. We would also like to note that, to the best of our knowledge, the idea of only penalizing the nonlinear constraints is new in the context of multiobjective derivative-free optimization.

In the last part of the paper, we test the linesearch approach on both bound constrained and nonlinearly constrained problems. The aim of the tests on bound constrained problems is understanding to what extent the theoretical properties of our method are helpful in practice. On the other hand, the goal of the tests on nonlinearly constrained problems is to show the effectiveness of the exact penalty approach when embedded in a DFO method for bound constrained multiobjective problems. For this reason we report the results obtained by both our algorithm and the globally convergent version of DMS.

The paper has the following structure. In section 2, we introduce the proposed algorithm along with the penalization idea to manage the general nonlinear constraints. In section 3, we carry out the theoretical analysis of the proposed method. In particular, the analysis is structured as follows:

- in subsection 3.1, we define new optimality conditions for problem (1.1) which explicitly take into account the bound constraints on the variables, and we define "stationary" points for the problem;
- in subsection 3.2, we prove the equivalence between the nonlinearly constrained problem (1.1) and the bound constrained problem obtained by penalizing the nonlinear constraints;
- in subsection 3.3, we prove that (under some suitable assumptions) the proposed algorithm produces sequences of points that converge to stationary points of the original problem.

Numerical results and comparisons are reported and discussed in section 4. Finally, in section 5, we draw some conclusions.

1.1. Notations and preliminary material. Given a vector $v \in \mathbb{R}^n$, a subscript will be used to denote either one of its components (v_i) or the fact that it is an

element of an infinite sequence of vectors (v_k). In case of possible misunderstanding or ambiguities, the ith component of a vector will be denoted by $(v)_i$.

Given two vectors $u, v \in \mathbb{R}^n$, we use the following convention for vector equalities and inequalities:

$$u = v \Leftrightarrow u_i = v_i \ \forall \ i = 1, \dots, n,$$
$$u < v \Leftrightarrow u_i < v_i \ \forall \ i = 1, \dots, n,$$
$$u \leq v \Leftrightarrow u_i \leq v_i \ \forall \ i = 1, \dots, n \text{ and } u \neq v.$$

Note that $u \geq v$ if and only if $-u \leq -v$.

We denote by v^j the generic jth element of a finite set of vectors. Furthermore, vectors $e_1, \dots, e_n$ represent the coordinate unit vectors. Given two vectors $a, b \in \mathbb{R}^n$, we denote by $y = \max\{a, b\}$ the vector such that $y_i = \max\{a_i, b_i\}$, $i = 1, \dots, n$. Furthermore, given a vector v we denote $v^+ = \max\{0, v\}$. The projection of a point x onto the set X will be denoted by $[x]_{[l,u]}$. Finally, we denote the unit sphere in the origin by $S(0, 1) = \{d \in \mathbb{R}^n : \|d\| = 1\}$, and $Co(A)$ indicates the convex hull of the set A. Given a point x and a scalar $\rho > 0$, $\mathcal{B}(x, \rho) = \{y \in \mathbb{R}^n : \|x - y\| \leq \rho\}$. We denote

$$\Gamma = \left\{ \mu \in \mathbb{R}^q : \mu \geq 0, \sum_{i=1}^{q} \mu_i = 1 \right\}.$$

Finally, by $\mathbf{1}$ we denote the vector of all ones of dimension q.

When dealing with several objective functions at a time, the concept of Pareto dominance is usually considered in the comparison of two points.

DEFINITION 1.1 (Pareto dominance). *Given two points, $x, y \in \mathcal{F}$, we say that x dominates y if $F(x) \leq F(y)$.*

Anyway, when coming to optimality, it may not be possible to find a point which is optimal for all the objectives simultaneously. This is the reason why the concept of Pareto dominance is also used to characterize global and local optimality into a multiobjective framework. More specifically, by means of the following two definitions, we are able to identify a set of nondominated points (the so called Pareto front or frontier) which represents the (global or local) optimal solutions of a given multiobjective problem.

DEFINITION 1.2 (global Pareto optimality). *A point $x^\star \in \mathcal{F}$ is a global Pareto optimizer of problem (1.1) if there does not exist a point $y \in \mathcal{F}$ such that $F(y) \leq F(x^\star)$.*

DEFINITION 1.3 (local Pareto optimality). *A point $x^\star \in \mathcal{F}$ is a local Pareto optimizer of problem (1.1) if there does not exist a point $y \in \mathcal{F} \cap \mathcal{B}(x^\star, \rho)$ such that $F(y) \leq F(x^\star)$ for some $\rho > 0$.*

2. The derivative-free algorithm. In this section, we introduce our derivative-free algorithm for the solution of problem (1.1), namely, Algorithm DFMO. It extends to the multiobjective case the approach proposed in [10], and, similarly to DMS in [7], generates a set of points at each iteration. We stress that the algorithm produces a sequence of sets of points $\{L_k\}$ (rather than a sequence of points as is common in the single objective case). More specifically, for each k, L_k is a finite set that can be described as

$$L_k = \{(x_i, \alpha_i), \ x_i \in X, \ \alpha_i > 0, \ i = 1, \dots, r_k\},$$

where $r_k = |L_k|$ and α_i is the trial step size associated with point x_i. Other relevant features of the algorithm are

- a linesearch approach that takes into account the presence of multiple objectives;
- an exact penalty approach for dealing with the nonlinear constraints.

More specifically, given the constrained problem (1.1) and a parameter $\epsilon > 0$, we introduce the following penalty functions,

$$Z_j(x;\epsilon) = f_j(x) + \frac{1}{\epsilon}\sum_{i=1}^{m}\max\{0, g_i(x)\} \qquad \forall\, j = 1,\ldots,q,$$

and define the penalized bound constrained multiobjective problem

$$(2.1) \qquad \begin{aligned} \min \quad & Z(x;\epsilon) = (Z_1(x;\epsilon),\ldots,Z_q(x;\epsilon))^\top \\ s.t. \quad & x \in X. \end{aligned}$$

We report the detailed scheme of Algorithm DFMO.

Algorithm DFMO

Data. $\gamma > 0$, $\theta \in (0,1)$, $\epsilon > 0$, $L_0 = \{(x_i,\alpha_i),\ x_i \in X,\ \alpha_i > 0,\ i = 1,\ldots,|L_0|\}$, a sequence $\{d_k\} \subset \mathbb{R}^n$ such that $\|d_k\| = 1$, for all k.

For $k = 0, 1, \ldots$
    Set $\tilde{L}_k = L_k$.
    For $i = 1, 2, \ldots, |L_k|$
        Let (x_i, α_i) be the i-th couple of L_k.
        If $\left(Z([x_i + \alpha_i d_k]_{[l,u]};\epsilon) \not\gtrdot Z(x_j;\epsilon) - \gamma\alpha_i^2\mathbf{1} \right)\ \forall\, x_j \in \tilde{L}_k$ **then**
            Set $\tilde{L}_k = \texttt{Projected Expansion}(x_i, \alpha_i, d_k, \tilde{L}_k, \gamma)$.
        Else
            (Failure Step) **If** $(x_i, \alpha_i) \in \tilde{L}_k$ Set $\tilde{L}_k = \left(\tilde{L}_k \setminus \{(x_i,\alpha_i)\} \right) \cup \{(x_i, \theta\alpha_i)\}$
        Endif
    End For
    Set $L_{k+1} = \tilde{L}_k$
End For

At iteration k, for each point in the set L_k, the algorithm starts a linesearch along direction d_k. If d_k guarantees sufficient decrease (i.e., there exists at least one objective function that reduces enough at the new point generated along d_k for each point in $\tilde{L}_k$, then a "sufficiently" large movement along this direction is performed (by means of a `Projected Expansion` Procedure) and a new set of points is generated.

`Add&Filter` $(L, (x, \alpha_x))$

Set $\tilde{L} = \{(x, \alpha_x)\}$
For each $(y, \alpha_y) \in L$
    If $Z(x;\epsilon) \not\lesssim Z(y;\epsilon)$ **then** $\tilde{L} = \tilde{L} \cup \{(y, \alpha_y)\}$
End For
Return $\tilde{L}$.

The `Add&Filter` procedure, given a list of nondominated pairs L and a pair (x, α_x) which is nondominated by any pair in L, produces the list $\tilde{L}$ of nondominated pairs among those in $L \cup \{(x, \alpha_x)\}$.

Finally, we describe the `Projected Expansion` procedure.

Projected Expansion $(y, \hat{\alpha}, p, \tilde{L}, \gamma)$.

Data. $\delta \in (0, 1)$.

Step 1. Set $\alpha = \hat{\alpha}$.

Step 2. Let $\beta = \alpha/\delta$.

Step 3. If $\left(Z([y + \beta p]_{[l,u]}; \epsilon) \not< Z([y + \alpha p]_{[l,u]}; \epsilon) - \gamma \left(\beta^2 - \alpha^2 \right) \mathbf{1} \right)$ **then**

      Set $\tilde{L} = \mathtt{Add\&Filter}(\tilde{L}, ([y + \alpha p]_{[l,u]}, \alpha))$.

    Endif

Step 4. If $\left(Z([y + \beta p]_{[l,u]}; \epsilon) \not> Z(x_j; \epsilon) - \gamma \beta^2 \mathbf{1} \right) \ \ \forall \, x_j \in \tilde{L}$ **then**

      Set $\alpha = \beta$ and go to Step 2.

    Endif

Step 5. Return $\tilde{L}$.

The `Projected Expansion` Procedure performs a "sufficiently" large movement along direction d_k and updates $\tilde{L}$ by adding and filtering a set of points.

The reader who is particularly interested in computational issues can completely skip section 3 and directly go to section 4 where numerical results of Algorithm DFMO and a comparison with a state-of-the-art code are reported and commented upon.

3. Convergence analysis of DFMO. This section is devoted to the convergence analysis of Algorithm DFMO. In particular, we prove under some suitable conditions that DFMO produces (in the limit) stationary points of the constrained problem (1.1). To this aim, we need to

- define necessary optimality conditions for problem (1.1) that explicitly take into account the bound constraints (see subsection 3.1);
- show that (at least for sufficiently small values of ϵ) solving problem (2.1) is "equivalent" to solving problem (1.1) (see subsection 3.2);
- (finally) prove that Algorithm DFMO converges to stationary points of problem (1.1) (see subsection 3.3).

Throughout the paper we require the following assumption.

ASSUMPTION 1. *The functions f_j, $j = 1, \ldots, q$, and g_i, $i = 1, \ldots, m$, are Lipschitz in X with constants $L_{f_j} > 0$, $j = 1, \ldots, q$, and $L_{g_i} > 0$, $i = 1, \ldots, m$.*

Observe that, since f_j, $j = 1, \ldots, q$, and g_i, $i = 1, \ldots, m$, are Lipschitz continuous, the penalty functions $Z_j(x; \epsilon)$, $j = 1, \ldots, q$, are Lipschitz continuous too, with Lipschitz constants

$$L_j \leq L_{f_j} + \frac{1}{\varepsilon} \sum_{i=1}^{m} L_{g_i}.$$

We denote by $\partial f(x) = \{s \in \mathbb{R}^n : f^{Cl}(x; d) \geq d^T s, \forall \, d \in \mathbb{R}^n\}$ the generalized gradient of f at x, where

$$(3.1) \qquad f^{Cl}(x; d) = \limsup_{y \to x, t \downarrow 0} \frac{f(y + td) - f(y)}{t}.$$

Taking into account the results reported in [3], we can write

$$(3.2) \qquad f^{Cl}(\bar{x}; d) = \max_{\xi \in \partial f(\bar{x})} \xi^\top d.$$

Now we introduce some useful definitions that will be used in the description of our derivative-free algorithm. First, we define the cone of feasible directions related to a point in the set X.

DEFINITION 3.1 (cone of feasible directions). *Given a point $x \in X$, let*

$$D(x) = \{d \in \mathbb{R}^n : d_i \geq 0 \text{ if } x_i = l_i, \ d_i \leq 0 \text{ if } x_i = u_i, d_i \in \mathbb{R} \text{ if } l_i < x_i < u_i, \ i = 1, \ldots, n\}$$

be the cone of feasible directions at x with respect to the simple bound constraints.

Then, given a point $x \in X$, the Clarke–Jahn generalized directional derivative of a function f along the direction $d \in D(x)$ is given by (see [16, section 3.5])

$$(3.3) \qquad f^\circ(x; d) = \limsup_{\substack{y \to x, y \in X \\ t \downarrow 0, y + td \in X}} \frac{f(y + td) - f(y)}{t}.$$

Recall that, according to definitions (3.1) and (3.3), we have, for any $d \in \mathbb{R}^n$,

$$(3.4) \qquad f^{Cl}(x; d) \geq f^\circ(x; d).$$

We finally introduce the following definition of a dense subsequence of directions.

DEFINITION 3.2 (dense sequence). *Let K be an infinite subset of integers (possibly $K = \{0, 1, \ldots\}$). The subsequence of normalized directions $\{d_k\}_{k \in K}$ is said to be dense in the unit sphere $S(0, 1)$ if for any $\bar{d} \in S(0, 1)$ and for any $\epsilon > 0$ there exists an index $k \in K$ such that $\|d_k - \bar{d}\| \leq \epsilon$.*

3.1. Necessary optimality conditions. In this subsection, we extend classic optimality conditions for MOO problems (see, e.g., Chapter 7 in [17]). In particular, we describe some results concerning Lipschitz continuous multiobjective problems with explicit handling of the bound constraints (apart from general nonlinear constraints). These conditions allow us to define stationary points of problem (1.1) which are our desired solutions.

The following proposition extends the result in [13] to the case where additional convex inequality constraints are present. In particular, we want to propose optimality conditions for problem (1.1) which explicitly take into account the feasible directions of the set defined by the bound constraints.

PROPOSITION 3.3 (Pareto–Clarke Fritz–John necessary optimality conditions). *Let $x^\star \in \mathcal{F}$ be a local Pareto point of the problem (1.1). Then, there exist multipliers $\sigma_1^\star, \ldots, \sigma_q^\star, \lambda_1^\star, \ldots, \lambda_m^\star \in \mathbb{R}$, not all zero and such that*

$$(3.5) \qquad \begin{aligned} &\sigma_i^\star \geq 0, &&i = 1, \ldots, q, \\ &\lambda_j^\star \geq 0, \quad \lambda_j^\star g_j(x^\star) = 0, &&j = 1, \ldots, m, \end{aligned}$$

and a vector

$$(3.6) \qquad \bar{\xi} \in \sum_{i=1}^{q} \sigma_i^\star \partial f_i(x^\star) + \sum_{j=1}^{m} \lambda_j^\star \partial g_j(x^\star),$$

such that

$$(3.7) \qquad \bar{\xi}^\top d \geq 0 \quad \text{for every } d \in D(x^\star).$$

Proof. Let us introduce the following functional

$$\Phi(x) = \max\{f_i(x) - f_i(x^\star), g_j(x), (l_h - x_h), (x_h - u_h) :$$
$$i = 1, \ldots, q, j = 1, \ldots, m, h = 1, \ldots, n\}$$

and show that a $\rho > 0$ exists such that, for all $x \in \mathcal{B}(x^\star, \rho)$, $\Phi(x) \geq 0$. Indeed, by contradiction, let us suppose that, for every $\rho > 0$, $\hat{x} \in \mathcal{B}(x^\star, \rho)$ exists such that $\Phi(\hat{x}) < 0$. This implies that $g(\hat{x}) < 0$ and $l < \hat{x} < u$ yielding $\hat{x} \in \mathcal{F}$. Further, for all $i = 1, \ldots, q$, $f_i(\hat{x}) < f_i(x^\star)$. These latter conditions contradict the local Pareto optimality of $x^\star$.

Since $x^\star \in \mathcal{F}$ and $\Phi(x^\star) = 0$, we know that $x^\star$ is a local minimum of $\Phi(x)$ onto $\mathcal{B}(x^\star, \rho)$. Hence, by definition of Clarke stationarity,

$$0 \in \partial\Phi(x^\star),$$

and from [3, Proposition 2.3.12], we have

$$(3.8) \qquad 0 \in \sum_{i=1}^{q} \hat{\sigma}_i \partial f_i(x^\star) + \sum_{j \in I_0(x^\star)} \hat{\lambda}_j \partial g_j(x^\star) - \sum_{h \in I_l(x^\star)} \hat{\mu}_h e_h + \sum_{w \in I_u(x^\star)} \hat{\mu}_w e_w$$

with $I_0(x^\star) = \{j : g_j(x^\star) = 0\}$, $I_l(x^\star) = \{h : x_h^\star = l_h\}$, $I_u(x^\star) = \{w : x_w^\star = u_w\}$, $\hat{\sigma}_i \geq 0$, $i = 1, \ldots, q$, $\hat{\lambda}_j \geq 0$, $j \in I_0(x^\star)$, $\hat{\mu}_h \geq 0$, $h \in I_l(x^\star)$, $\hat{\mu}_w \geq 0$, $w \in I_u(x^\star)$, and

$$(3.9) \qquad \sum_{i=1}^{q} \hat{\sigma}_i + \sum_{j \in I_0(x^\star)} \hat{\lambda}_j + \sum_{h \in I_l(x^\star)} \hat{\mu}_h + \sum_{w \in I_u(x^\star)} \hat{\mu}_w = 1.$$

Observe that by the linear independence of the set $\{e_h, e_w, h \in I_l(x^\star), w \in I_u(x^\star)\}$, we deduce that

$$\sum_{i=1}^{q} \hat{\sigma}_i + \sum_{j \in I_0(x^\star)} \hat{\lambda}_j \neq 0,$$

since, otherwise, we would have from (3.8) that

$$0 = - \sum_{h \in I_l(x^\star)} \hat{\mu}_h e_h + \sum_{w \in I_u(x^\star)} \hat{\mu}_w e_w,$$

which can be satisfied if and only if $\hat{\mu}_h = 0$, $h \in I_l(x^\star)$ and $\hat{\mu}_w = 0$, $w \in I_u(x^\star)$, contradicting (3.9).

Hence, we can normalize multipliers in (3.8) as in

$$(3.10) \qquad 0 \in \sum_{i=1}^{q} \sigma_i \partial f_i(x^\star) + \sum_{j \in I_0(x^\star)} \lambda_j \partial g_j(x^\star) - \sum_{h \in I_l(x^\star)} \mu_h e_h + \sum_{w \in I_u(x^\star)} \mu_w e_w,$$

where now (setting $\sum_{i=1}^{q} \hat{\sigma}_i + \sum_{j \in I_0(x^\star)} \hat{\lambda}_j = \Lambda \neq 0$)

$$\sigma_i = \frac{\hat{\sigma}_i}{\Lambda}, \lambda_j = \frac{\hat{\lambda}_j}{\Lambda} \ \forall \ j \in I_0(x^\star), \ \mu_h = \frac{\hat{\mu}_h}{\Lambda} \ \forall \ h \in I_l(x^\star), \ \mu_w = \frac{\hat{\mu}_w}{\Lambda} \ \forall \ w \in I_u(x^\star),$$

and

$$(3.11) \qquad \sum_{i=1}^{q} \sigma_i + \sum_{j \in I_0(x^\star)} \lambda_j = 1.$$

As a consequence, by (3.11), there exists a vector

$$\xi \in \sum_{i=1}^{q} \sigma_i \partial f_i(x^\star) + \sum_{j \in I_0(x^\star)} \lambda_j \partial g_j(x^\star)$$

which satisfies, from (3.10),

$$\xi \;=\; \sum_{h \in I_l(x^\star)} \mu_h e_h - \sum_{w \in I_u(x^\star)} \mu_w e_w.$$

Recalling the definition of $D(x)$, we have

$$\xi^\top d \geq 0$$

for all $d \in D(x^\star)$.   □

By introducing a constraint qualification, we can give KKT optimality conditions for the multiobjective problem (1.1).

PROPOSITION 3.4 (Pareto–Clarke KKT necessary optimality conditions).
Let $x^\star \in \mathcal{F}$ be a local Pareto minimum of the problem (1.1) and assume that a direction $d \in D(x^\star)$ exists such that for all $j \in \{1, \ldots, m : g_j(x^\star) = 0\}$

(3.12) $(\xi^{g_j})^\top d < 0 \quad \forall\, \xi^{g_j} \in \partial g_j(x^\star).$

Then, there exist multipliers $\sigma_1^\star, \ldots, \sigma_q^\star, \lambda_1^\star, \ldots, \lambda_m^\star \in \mathbb{R}$, such that

$$\sigma_i^\star \geq 0, \quad \sigma^\star \neq 0, \qquad i = 1, \ldots, q,$$
$$\lambda_j^\star \geq 0, \quad \lambda_j^\star g_j(x^\star) = 0, \qquad j = 1, \ldots, m,$$

and a vector

$$\bar{\xi} \in \sum_{i=1}^{q} \sigma_i^\star \partial f_i(x^\star) + \sum_{j=1}^{m} \lambda_j^\star \partial g_j(x^\star),$$

such that

$$\bar{\xi}^\top d \geq 0 \quad \text{for every } d \in D(x^\star).$$

Proof. By Proposition 3.3, there exist multipliers $\sigma_i^\star \geq 0$, $i = 1, \ldots, q$, $\lambda_j^\star \geq 0$, $j = 1, \ldots, m$ with $\lambda_j^\star = 0$ when $g_j(x^\star) < 0$, and a vector $\xi \in \sum_{i=1}^{q} \sigma_i^\star \partial f_i(x^\star) + \sum_{j=1}^{m} \lambda_j^\star \partial g_j(x^\star)$ such that (3.6) and (3.7) hold. In order to prove the proposition, we proceed by contradiction and assume that $\sigma_i^\star = 0$, $i = 1, \ldots, q$. In this case, we would have

$$\xi \in \sum_{j \in I_0(x^\star)} \lambda_j^\star \partial g_j(x^\star),$$

where $I_0(x^\star) = \{j : g_j(x^\star) = 0\}$. Note that the multipliers $\lambda_j^\star$, $j = 1, \ldots, m$, cannot be all zero, since in this case all the multipliers would be zero thus contradicting Proposition 3.3. Then, we can define new multipliers

$$\bar{\lambda}_j = \lambda_j^\star/\beta, \qquad j \in I_0(x^\star),$$

where $\beta = \sum_{j \in I_0(x^\star)} \lambda_j^\star > 0$ and a vector $\bar{\xi} = \xi/\beta$ such that

$$\bar{\xi} \in \sum_{j \in I_0(x^\star)} \bar{\lambda}_j \partial g_j(x^\star).$$

Hence, we have that $\bar{\lambda}_j \geq 0$ and $\sum_{j \in I_0(x^\star)} \bar{\lambda}_j = 1$.

Then, by Proposition 3.3, the following system

$$\begin{aligned} -\bar{\xi}^\top d &> 0, \\ -e_h^\top d &\leq 0 \qquad \forall \ h \in I_l(x^\star), \\ e_w^\top d &\leq 0 \qquad \forall \ w \in I_u(x^\star), \end{aligned}$$

with $I_l(x^\star) = \{h : x_h^\star = l_h\}$, and $I_u(x^\star) = \{w : x_w^\star = u_w\}$, does not have a solution. It is easy to notice that the latter two sets of constraints imply $d \in D(x^\star)$. As a consequence, by Farkas' theorem (see, e.g., [22, Chapter 2]), we have that scalars $\alpha_h \geq 0$, $h \in I_l(x^\star)$ and $\alpha_w \geq 0$, $w \in I_u(x^\star)$, exist such that

(3.13)
$$0 = \bar{\xi} - \sum_{h \in I_l(x^\star)} \alpha_h e_h + \sum_{w \in I_u(x^\star)} \alpha_w e_w.$$

Now, by using condition (3.12), we know that a direction $\bar{d} \in \mathbb{R}^n$ exists such that

$$\begin{aligned} (\xi^{g_j})^\top \bar{d} &< 0 \quad \forall \ \xi^{g_j} \in \partial g_j(x^\star) \quad \forall \ j \in I_0(x^\star), \\ -e_h^\top \bar{d} &\leq 0 \quad \forall \ h \in I_l(x^\star), \\ e_w^\top \bar{d} &\leq 0 \quad \forall \ w \in I_u(x^\star). \end{aligned}$$

Now, by the alternative theorem in [23, Theorem 2.3.4] and [15], there cannot exist multipliers $\hat{\lambda}_j \geq 0$, $j \in I_0(x^\star)$, $\hat{\mu}_h \geq 0$, $h \in I_l(x^\star)$, $\hat{\mu}_w \geq 0$, $w \in I_u(x^\star)$ with

$$\sum_{j \in I_0(x^\star)} \hat{\lambda}_j = 1,$$

such that

$$0 \in \sum_{j \in I_0(x^\star)} \hat{\lambda}_j \partial g_j(x^\star) - \sum_{h \in I_l(x^\star)} \hat{\mu}_h e_h + \sum_{w \in I_u(x^\star)} \hat{\mu}_w e_w.$$

This is in contradiction with (3.13). $\qquad \square$

Now, we can formally define Pareto–Clarke stationary points of problem (1.1).

DEFINITION 3.5 (Pareto–Clarke stationary point). *Given problem* (1.1), *a feasible point* $x^\star \in \mathcal{F}$ *is a Pareto–Clarke stationary point of* (1.1) *if there exist multipliers* $\sigma_1^\star, \ldots, \sigma_q^\star, \lambda_1^\star, \ldots, \lambda_m^\star \in \mathbb{R}$, *such that*

$$\begin{aligned} \sigma_i^\star &\geq 0, \ \sigma^\star \neq 0, \qquad i = 1, \ldots, q, \\ \lambda_j^\star &\geq 0, \quad \lambda_j^\star g_j(x^\star) = 0, \qquad j = 1, \ldots, m, \end{aligned}$$

and a vector

$$\bar{\xi} \in \sum_{i=1}^q \sigma_i^\star \partial f_i(x^\star) + \sum_{j=1}^m \lambda_j^\star \partial g_j(x^\star),$$

such that $\bar{\xi}^\top d \geq 0$ *for every* $d \in D(x^\star)$.

3.1.1. The bound constrained case. In this subsection we restrict our attention to the following bound constrained problem

$$\text{(3.14)} \qquad \begin{aligned} \min \quad & F(x) = (f_1(x), \dots, f_q(x))^\top \\ \text{s.t.} \quad & x \in X. \end{aligned}$$

This is done in order to get some results that will be used in the theoretical analysis of the exact penalty approach described in the next section. The definition of the Pareto–Clarke stationary point for problem (3.14) can be straightforwardly obtained from Definition 3.5.

DEFINITION 3.6. *Given problem* (3.14), *a feasible point* $x^\star \in X$ *is a Pareto–Clarke stationary point of* (3.14) *if there exist nonnegative multipliers* $\sigma_1^\star, \dots, \sigma_q^\star \in \mathbb{R}$ *not all zero and a vector*

$$\bar{\xi} \in \sum_{i=1}^q \sigma_i^\star \partial f_i(x^\star),$$

such that $\bar{\xi}^\top d \geq 0$, *for every* $d \in D(x^\star)$.

We now report a result that characterizes the Pareto–Clarke stationary points in the bound constrained case.

PROPOSITION 3.7. *A point* $\bar{x} \in X$ *is a Pareto–Clarke stationary point of* (3.14) *if and only if for all* $d \in D(\bar{x})$, *an index* $j_d \in \{1, \dots, q\}$ *exists such that*

$$\text{(3.15)} \qquad f_{j_d}^{Cl}(\bar{x}; d) \geq 0.$$

Proof. First, we assume that $\bar{x} \in X$ is a Pareto–Clarke stationary point of (3.14), and we prove that condition (3.15) holds. From the definition of a Pareto–Clarke stationary point, there exist nonnegative multipliers $\bar{\mu}_1, \dots, \bar{\mu}_q \in \mathbb{R}$, not all zero, and a vector $\bar{\xi} \in \sum_{i=1}^q \bar{\mu}_i \partial f_i(\bar{x})$ such that, for any $d \in D(\bar{x})$,

$$\bar{\xi}^T d \geq 0.$$

Then, we can define new multipliers

$$\tilde{\mu}_i = \bar{\mu}_i / \beta, \qquad i = 1, \dots, q,$$

where $\beta = \sum_{i=1}^q \bar{\mu}_i > 0$ (note that $\tilde{\mu}_i \geq 0$ and $\sum_{i=1}^q \tilde{\mu}_i = 1$) and a vector $\tilde{\xi} = \bar{\xi}/\beta$ such that

$$\tilde{\xi} \in \sum_{i=1}^q \tilde{\mu}_i \partial f_i(\bar{x})$$

and

$$\text{(3.16)} \qquad \tilde{\xi}^\top d \geq 0 \qquad \forall\, d \in D(\bar{x}).$$

From (3.16), we have that the following system

$$\begin{aligned} -\tilde{\xi}^\top d &> 0, \\ e_i^\top d &\leq 0 \quad \forall\, i \in I_u(\bar{x}), \\ -e_i^\top d &\leq 0 \quad \forall\, i \in I_l(\bar{x}), \end{aligned}$$

does not have a solution. By using the Farkas lemma, we have

$$\tilde{\xi} + \sum_{i \in I_u(\bar{x})} \rho_i e_i - \sum_{i \in I_l(\bar{x})} \sigma_i e_i = 0$$

with $\rho_i \geq 0$, $i \in I_u(\bar{x})$, and $\sigma_i \geq 0, i \in I_l(\bar{x})$, and so that

$$0 \in \sum_{i=1}^{q} \tilde{\mu}_i \partial f_i(\bar{x}) + \sum_{i \in I_u(\bar{x})} \rho_i e_i - \sum_{i \in I_l(\bar{x})} \sigma_i e_i.$$

Now, by considering the alternative theorem in [23, Theorem 2.3.4] and [15], we have that the system

$$\max \left\{ \xi^\top d : \xi \in \partial f_i(\bar{x}) \right\} < 0 \quad \forall\, i = 1, \ldots, q,$$
$$e_i^\top d \leq 0 \quad \forall\, i \in I_u(\bar{x}),$$
$$-e_i^\top d \leq 0 \quad \forall\, i \in I_l(\bar{x}),$$

where $I_u(\bar{x}) = \{i : \bar{x}_i = u_i\}$ and $I_l(\bar{x}) = \{i : \bar{x}_i = l_i\}$ does not have a solution. This can be equivalently expressed by saying that no direction $d \in D(\bar{x})$ exists such that for all $i = 1, \ldots, q$,

$$\max \left\{ \xi^\top d : \xi \in \partial f_i(\bar{x}) \right\} < 0,$$

which means that, for every $d \in D(\bar{x})$, there exist an index $j_d \in \{1, \ldots, q\}$ such that

$$\max \left\{ \xi^\top d : \xi \in \partial f_{j_d}(\bar{x}) \right\} \geq 0.$$

Then by recalling (3.2), we can equivalently say that for all $d \in D(\bar{x})$, an index $j_d \in \{1, \ldots, q\}$ exists such that
$$f_{j_d}^{Cl}(\bar{x}; d) \geq 0.$$

Now, let us suppose that for all $d \in D(\bar{x})$ there exists a j_d such that condition (3.15) holds.

This fact can be equivalently expressed by saying that no direction $d \in D(\bar{x})$ exists such that for all $i = 1, \ldots, q$,

$$\max \left\{ \xi^\top d : \xi \in \partial f_i(\bar{x}) \right\} < 0.$$

Hence, by recalling the definition of $D(x)$, the preceding means that the following system has no solution:

(3.17) $$\max \left\{ \xi^\top d : \xi \in \partial f_i(\bar{x}) \right\} < 0 \quad \forall\, i = 1, \ldots, q,$$

(3.18) $$e_i^\top d \leq 0 \quad \forall\, i \in I_u(\bar{x}),$$

(3.19) $$-e_i^\top d \leq 0 \quad \forall\, i \in I_l(\bar{x}),$$

where $I_u(\bar{x}) = \{i : \bar{x}_i = u_i\}$ and $I_l(\bar{x}) = \{i : \bar{x}_i = l_i\}$. Now, by considering the alternative theorem in [23, Theorem 2.3.4] and [15], we have that multipliers $\mu \in \Gamma$, i.e., $\mu \geq 0$, $\sum_{i=1}^{q} \mu_i = 1$, and $\rho_i \geq 0$, $i \in I_u(\bar{x})$, $\sigma_i \geq 0$, $i \in I_l(\bar{x})$, exist such that

$$0 \in \sum_{i=1}^{q} \mu_i \partial f_i(\bar{x}) + \sum_{i \in I_u(\bar{x})} \rho_i e_i - \sum_{i \in I_l(\bar{x})} \sigma_i e_i.$$

Hence, there exists a vector $\bar{\xi} \in \sum_{i=1}^{q} \mu_i \partial f_i(\bar{x})$ such that

$$\bar{\xi} = -\left(\sum_{i \in I_u(\bar{x})} \rho_i e_i - \sum_{i \in I_l(\bar{x})} \sigma_i e_i \right).$$

Then, for all $d \in \mathbb{R}^n$, we can write

$$\bar{\xi}^\top d = -\left(\sum_{i \in I_u(\bar{x})} \rho_i e_i - \sum_{i \in I_l(\bar{x})} \sigma_i e_i \right)^\top d.$$

So that, by considering (3.18) and (3.19), for all $d \in D(\bar{x})$ we can write

$$\bar{\xi}^\top d \geq 0 \qquad\qquad\qquad \square$$

so that $\bar{x}$ is a Pareto–Clarke stationary point of (3.14).

Finally, with reference to problem (3.14), we can introduce the following slightly stronger definition of stationarity.

DEFINITION 3.8 (Pareto–Clarke–Jahn stationary point).   *Given the problem (3.14), $\bar{x}$ is a Pareto–Clarke–Jahn stationary point of (3.14) if, for all $d \in D(\bar{x})$, an index $j_d \in \{1, \ldots, q\}$ exists such that*

$$f_{j_d}^\circ(\bar{x}; d) \geq 0.$$

3.2. Exact penalization of the constraints. In this subsection we analyze the equivalence between the original constrained problem (1.1) and the (penalized) bound constrained problem (2.1).

More specifically, we prove that there exists a correspondence between Pareto–Clarke stationary points of the penalized problem and Pareto–Clarke stationary points of the original constrained problem we want to solve. Furthermore, we show that the two problems share the same global Pareto optimal solutions.

These results are at the basis of Algorithm DFMO which makes use of the penalty approach to manage the hard nonlinear constraints, and that explicitly handles the simple bound constraints defining the set X.

In order to prove the main results, we also need an extended version of the Mangasarian–Fromowitz constraint qualification (EMFCQ) condition for nonsmooth problems.

ASSUMPTION 2 (EMFCQ). *Given problem (1.1), for any $x \in X \setminus \overset{\circ}{\mathcal{F}}$ a direction $d \in D(x)$ exists such that*

$$(\xi^{g_i})^\top d < 0$$

for all $\xi^{g_i} \in \partial g_i(x)$, $i \in \{1, \ldots, m : g_i(x) \geq 0\}$.

We remark that the above assumption is an extension to the nonsmooth case of the EMFCQ for differentiable problems (see, e.g., [12, 21]). It concerns the behavior of the constraint functions outside the feasible set and is somewhat connected to the feasibility of the original problem. In particular, it is a sufficient condition to guarantee that the feasible set is not empty, which becomes necessary for convex feasible sets (see [21]). We further note that this assumption is particularly important to study convergence properties of algorithms that approach a solution by (possibly) producing infeasible points.

PROPOSITION 3.9. *Let Assumption* 2 *hold. Given problem* (1.1) *and considering problem* (2.1), *a threshold value* $\epsilon^\star > 0$ *exists such that, for every* $\epsilon \in (0, \epsilon^\star]$, *problem* (2.1) *has no Pareto–Clarke stationary points in* $X \setminus \mathcal{F}$.

Proof. We proceed by contradiction and assume that, for any integer k, an $\epsilon_k \leq 1/k$, and a stationary point for problem (2.1) $x_k \in X \setminus \mathcal{F}$ exist. Then, let us consider a limit point of this sequence $\bar{x}$ belonging to the closure of $X \setminus \mathcal{F}$ (hence $\bar{x} \notin \overset{\circ}{\mathcal{F}}$) and let us relabel the corresponding subsequence again $\{x_k\}$.

Since $\bar{x} \notin \overset{\circ}{\mathcal{F}}$, Assumption 2 guarantees that a direction $\bar{d} \in D(\bar{x})$ exists such that

$$(\xi^{g_i})^\top \bar{d} < 0 \qquad \forall\, \xi^{g_i} \in \partial g_i(\bar{x}),\ i \in I(\bar{x}),$$

where $I(x) = \{i \in \{1, \ldots, m\} : g_i(x) \geq 0\}$. The above property can be equivalently expressed by saying that a positive scalar $\eta > 0$ exists, such that

$$(3.20) \qquad \max_{\substack{\xi^{g_i} \in \partial g_i(\bar{x}) \\ i \in I(\bar{x})}} (\xi^{g_i})^\top \bar{d} = -\eta.$$

Recalling that, for k sufficiently large, $D(\bar{x}) \subseteq D(x_k)$ (see, e.g., [19]), so that $\bar{d} \in D(x_k)$, we get, by considering Proposition 3.7 and that x_k is a Pareto–Clarke stationary point of problem (2.1), that an index $j^k \in \{1, \ldots, q\}$ (depending on $\bar{d}$ and x_k) must exist such that

$$(3.21) \qquad Z_{j^k}^{Cl}(x_k; \epsilon, \bar{d}) \geq 0.$$

Since, by definition, $Z_{j^k}^{Cl}(x; \epsilon, \bar{d}) = \max_{\xi \in \partial Z_{j^k}(x; \epsilon)} \xi^\top \bar{d}$, and we know that

$$\partial Z_{j^k}(x; \epsilon) \subseteq \partial f_{j^k}(x) + \frac{1}{\epsilon} \sum_{i=1}^{m} \partial(\max\{0, g_i(x)\})$$

and, by recalling the definition of $I(x)$, we have

$$\partial f_{j^k}(x) + \frac{1}{\epsilon} \sum_{i=1}^{m} \partial(\max\{0, g_i(x)\}) \subseteq \partial f_{j^k}(x) + \frac{1}{\epsilon} \sum_{i \in I(x)} \partial g_i(x),$$

thus, inequality (3.21) can be written as

$$(3.22) \qquad \left(\xi_k^{(f_{j^k})} + \frac{1}{\epsilon_k} \sum_{i \in I(x_k)} \xi_k^{g_i} \right)^\top \bar{d} \geq 0$$

with $\xi_k^{(f_{j^k})} \in \partial f_{j^k}(x_k)$, $\xi_k^{g_i} \in \partial g_i(x_k)$.

Now, recalling that q and m are finite numbers, there exist $\bar{j} \in \{1, \ldots, q\}$ and $\bar{I} \subseteq \{1, \ldots, m\}$, and we can consider the subsequence of $\{x_k\}$, where $j^k = \bar{j}$ and $I(x_k) = \bar{I}$.

Then, since the generalized gradient of a locally Lipschitz continuous function is locally bounded, we have that all the considered sequences $\{\xi_k^{f_{\bar{j}}}\}$, $\{\xi_k^{g_i}\}$, $i \in \bar{I}$, where

$\xi_k^{f_{\bar{j}}} \in \partial f_{\bar{j}}(x_k)$, $\xi_k^{g_i} \in \partial g_i(x_k)$, $x_k \in X$, are bounded.[1] Hence, we get that

(3.23a) $$\xi_k^{f_{\bar{j}}} \to \bar{\xi}^{f_{\bar{j}}},$$
(3.23b) $$\xi_k^{g_i} \to \bar{\xi}^{g_i} \qquad \forall\, i \in \bar{I}.$$

Further, since $\partial f_{\bar{j}}$ and ∂g_i, $i \in \bar{I}$, are upper semicontinuous at $\bar{x}$ (see Proposition 2.1.5 in [3]), it results that $\bar{\xi}^{f_{\bar{j}}} \in \partial f_{\bar{j}}(\bar{x})$, $\bar{\xi}^{g_i} \in \partial g_i(\bar{x})$, $i \in \bar{I}$.

Now, since by continuity of the problem functions we have for k sufficiently large

$$\{i : g_i(\bar{x}) < 0\} \subseteq \{i : g_i(x_k) < 0\},$$

we have, for k sufficiently large,

$$\{i : g_i(x_k) \geq 0\} = I(x_k) \subseteq I(\bar{x}) = \{i : g_i(\bar{x}) \geq 0\},$$

so that

(3.24) $$\bar{I} \subseteq I(\bar{x}).$$

Now, by considering that, for all $i \in \bar{I}$,

$$(\xi_k^{g_i})^\top \bar{d} \leq \max_{\substack{\vartheta_k^{g_i} \in \partial g_i(x_k) \\ i \in \bar{I}}} (\vartheta_k^{g_i})^\top \bar{d},$$

by (3.20), (3.23), and (3.24), we get, for k sufficiently large,

(3.25) $$(\xi_k^{g_i})^\top \bar{d} \leq -\frac{\eta}{2} \qquad \forall\, i \in \bar{I}.$$

Now, by multiplying (3.22) by ϵ_k we have

$$\left(\epsilon_k \xi_k^{f_{\bar{j}}} + \sum_{i \in \bar{I}} \xi_k^{g_i} \right)^\top \bar{d} \geq 0,$$

which, by (3.25), yields

$$0 \leq \left(\epsilon_k \xi_k^{f_{\bar{j}}} + \sum_{i \in \bar{I}} \xi_k^{g_i} \right)^\top \bar{d} \leq \left(\epsilon_k \xi_k^{f_{\bar{j}}} \right)^\top \bar{d} - \frac{\eta}{2}.$$

Finally, the above expression, considering (3.23a), gives rise to a contradiction when $\epsilon_k \to 0$. $\qquad \square$

In order to give stationarity results for problem (2.1), we have the following proposition.

PROPOSITION 3.10. *Let Assumption* 2 *hold. Then, for any $\epsilon > 0$, every Pareto–Clarke stationary point $\bar{x}$ of problem* (2.1), *such that $\bar{x} \in \mathcal{F}$, is also a Pareto–Clarke stationary point of problem* (1.1).

---

[1] This result follows by considering that a finite covering of X by bounded sets exists and that any $\xi_k^{f_{\bar{j}}}, \xi_k^{g_i}$, $i \in \bar{I}$, are bounded on the latter sets.

Proof. Since $\bar{x}$ is, by assumption, a Pareto–Clarke stationary point of problem (2.1), then by Definition 3.6 we know that a vector of nonnegative multipliers $\mu \in \Gamma$, i.e., $\mu \geq 0$, $\sum_{i=1}^{q} \mu_i$, not all zero, and a vector $\xi^{\star} \in \sum_{i=1}^{q} \mu_i \partial Z_i(\bar{x}, \epsilon)$ exist such that, for all $d \in D(\bar{x})$,

$$(\xi^{\star})^{\top} d \geq 0.$$

Now, we recall that

$$\partial Z_i(x, \epsilon) \subseteq \partial f_i(x) + \frac{1}{\epsilon} \sum_{j \in I(x)} \partial g_j(x),$$

where $I(x) = \{i : g_i(x) \geq 0\}$. Hence, we have that $\xi^{*} \in \sum_{i=1}^{q} \mu_i \Big(\partial f_i(\bar{x}) + \frac{1}{\epsilon} \sum_{j \in I(\bar{x})} \partial g_j(\bar{x}) \Big)$, $\mu \in \Gamma$. Then, denoting $\lambda_j = \left(\sum_{i=1}^{q} \mu_i \right) / \epsilon$,

$$\max \left\{ \xi^{\top} d : \ \xi \in \sum_{i=1}^{q} \mu_i \partial f_i(\bar{x}) + \sum_{j \in I(\bar{x})} \lambda_j \partial g_j(\bar{x}) \right\} \geq 0$$

for all $d \in D(\bar{x})$ with $\lambda_j \geq 0$, $j \in I(\bar{x})$. The above condition shows that a $\bar{\xi} \in \sum_{i=1}^{q} \mu_i \partial f_i(\bar{x}) + \sum_{j \in I(\bar{x})} \lambda_j \partial g_j(\bar{x})$ exists such that $\bar{\xi}^{\top} d \geq 0$ for all $d \in D(\bar{x})$. Hence, by recalling that $\bar{x} \in \mathcal{F}$ and $I(\bar{x}) = I_0(\bar{x})$ when $\bar{x} \in \mathcal{F}$, and setting $\lambda_j = 0$ when $j \notin I_0(\bar{x})$, it is proved that $\bar{x}$ is a Pareto–Clarke stationary point of problem (1.1). $\square$

Now, we introduce an intermediate result which basically states that, for ϵ sufficiently small, every Pareto–Clarke–Jahn stationary point of problem (2.1) is a Pareto–Clarke stationary point of problem (1.1).

PROPOSITION 3.11. *Let Assumption 2 hold. Given problem (1.1) and considering problem (2.1), a threshold value $\epsilon^{\star} > 0$ exists such that, for every $\epsilon \in (0, \epsilon^{\star}]$, every Pareto–Clarke–Jahn-stationary point of problem (2.1) is a Pareto–Clarke stationary point of problem (1.1).*

Proof. Let $\bar{x} \in X$ be Pareto–Clarke–Jahn stationary for problem (2.1). By (3.4), we also have that $\bar{x}$ is Pareto–Clarke stationary for problem (2.1). Now, the proof follows by considering Propositions 3.9 and 3.10. $\square$

Finally, we show the correspondence between the global Pareto minimizers of problem (1.1) and those of problem (2.1).

LEMMA 3.12. *Given problem (1.1), if a feasible point $\hat{x}$ is not a global Pareto minimizer, then a global Pareto minimizer x^{*} exists that dominates $\hat{x}$.*

Proof. Since $\hat{x}$ is not a global Pareto minimizer, a feasible point y exists such that

(3.26a) $f_i(y) \leq f_i(\hat{x}) \qquad \forall \, i \in \{1, \ldots, q\},$

(3.26b) $f_{\ell}(y) < f_{\ell}(\hat{x}) \qquad$ for some $\ell \in \{1, \ldots, q\}.$

Then, let us define the following (single objective) optimization problem

(3.27)
$$\begin{aligned} \min \quad & \sum_{i=1}^{q} f_i(x) \\ \text{s.t.} \quad & f_i(x) \leq f_i(\hat{x}), \qquad i = 1, \ldots, q. \\ & x \in \mathcal{F}. \end{aligned}$$

We recall that $\hat{x}$ is a feasible solution of the above problem. Furthermore, since the feasible set is compact and the objective function is continuous, the above problem admits a global solution x^* such that

$$\sum_{i=1}^{q} f_i(x^*) < \sum_{i=1}^{q} f_i(\hat{x}),$$

where the above inequality follows from (3.26). We claim that x^* is a global Pareto minimizer of problem (1.1). Indeed, if this was not the case, a point $\bar{x} \in \mathcal{F}$ would exist such that

$$f_i(\bar{x}) \le f_i(x^*) \qquad \forall\, i \in \{1, \ldots, q\},$$
$$f_j(\bar{x}) < f_j(x^*) \qquad \text{for some } j \in \{1, \ldots, q\}.$$

But this would imply that

$$\sum_{i=1}^{q} f_i(\bar{x}) < \sum_{i=1}^{q} f_i(x^*),$$

thus contradicting the fact that x^* is the global minimizer of (3.27). Hence, the proof is concluded. $\square$

PROPOSITION 3.13. *Let Assumption 2 hold. Then, given problem* (1.1) *and considering problem* (2.1)*, a threshold value $\epsilon^\star > 0$ exists such that, for every $\epsilon \in (0, \epsilon^\star]$, any global Pareto minimizer of problem* (2.1) *is a global Pareto minimizer of problem* (1.1)*, and conversely.*

Proof. Let us first suppose that x^* is a global Pareto minimizer of problem (2.1). We proceed by contradiction and assume that, for any integer k and $\epsilon_k \le 1/k$, a point x_k, which is a global Pareto minimizer of problem (2.1) but not of problem (1.1), exists. From the definition of global Pareto optimality, it follows that there cannot exist any $y \in X$ such that

$$Z(y, \epsilon_k) \le Z(x_k, \epsilon_k).$$

In particular, the above inequality must not hold for any $y \in \mathcal{F}$ which, considering that $Z(y, \epsilon_k) = F(y)$ for $y \in \mathcal{F}$, yields

$$F(y) \le Z(x_k, \epsilon_k).$$

Further, from Proposition 3.9 and considering that x_k is also a stationary point of problem (2.1), for k sufficiently large, we have that $x_k \in \mathcal{F}$. Hence, there cannot exist $y \in \mathcal{F}$ such that

$$F(y) \le F(x_k),$$

that is, x_k is a global Pareto minimizer of problem (1.1).

Let us now suppose that $\hat{x} \in \mathcal{F}$ is a global Pareto minimizer of problem (1.1). Let us assume that it is not a global Pareto minimizer of problem (2.1). Then, by Lemma 3.12 applied to problem (2.1), a global Pareto minimizer x^* of problem (2.1) exists such that x^* dominates $\hat{x}$. If $x^* \in \mathcal{F}$, then $Z(x^*, \epsilon) = F(x^*)$ and this would contradict the fact that $\hat{x}$ is a global Pareto minimizer of (1.1). Otherwise, if $x^* \in X \setminus \mathcal{F}$, for ϵ sufficiently small, this would be in contrast with Proposition 3.9, thus concluding the proof. $\square$

3.3. Stationarity result for DFMO. With reference to Algorithm DFMO, we give a definition that will be used throughout the subsection.

DEFINITION 3.14. *Let $\{L_k\}$ with $L_k = \{(x_j, \alpha_j), \ j = 1, \ldots, |L_k|\}$, be the sequence of sets of nondominated points produced by DFMO. We define a linked sequence as a sequence $\{(x_{j_k}, \alpha_{j_k})\}$ such that for any $k = 1, 2, \ldots,$ the pair $(x_{j_k}, \alpha_{j_k}) \in L_k$ is generated at iteration $k - 1$ of DFMO by the pair $(x_{j_{k-1}}, \alpha_{j_{k-1}}) \in L_{k-1}$.*

We highlight that one of the following two cases can happen.

1. Success step: the point $\left[x_{j_{k-1}} + \frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}}} d_{k-1}\right]_{[l,u]}$ is such that, for an integer $s_{k-1} > 0$, it results in

(3.28)
$$Z\left(\left[x_{j_{k-1}} + \frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}+1}} d_{k-1}\right]_{[l,u]}; \epsilon\right) \not< Z\left(\left[x_{j_{k-1}} + \frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}}} d_{k-1}\right]_{[l,u]}; \epsilon\right)$$
$$-\gamma\left(\left(\frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}+1}}\right)^2 - \left(\frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}}}\right)^2\right)\mathbf{1}.$$

In this case we have

(3.29a)
$$x_{j_k} = \left[x_{j_{k-1}} + \frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}}} d_{k-1}\right]_{[l,u]},$$

(3.29b)
$$\alpha_{j_k} = \frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}}}.$$

2. Failure step: we have

(3.30)
$$Z([x_{j_{k-1}} + \alpha_{j_{k-1}} d_{k-1}]_{[l,u]}; \epsilon) > Z(x_{\ell_j}; \epsilon) - \gamma(\alpha_{j_{k-1}})^2 \mathbf{1}$$

for at least a point $x_{\ell_j} \in \tilde{L}_{k-1}$, that is, $Z([x_{j_{k-1}} + \alpha_{j_{k-1}} d_{k-1}]_{[l,u]}; \epsilon)$ is dominated by $Z(x_{\ell_j}; \epsilon) - \gamma(\alpha_{j_{k-1}})^2 \mathbf{1}$. This implies that $[x_{j_{k-1}} + \alpha_{j_{k-1}} d_{k-1}]_{[l,u]}$ is a "bad" point so that the step $\alpha_{j_{k-1}}$ associated with $x_{j_{k-1}}$ is shrunk by the constant factor θ, i.e.,

(3.31a)
$$x_{j_k} = x_{j_{k-1}},$$
(3.31b)
$$\alpha_{j_k} = \theta\alpha_{j_{k-1}}.$$

Now we state a simple theoretical result.

LEMMA 3.15. *Let $\alpha, \beta \in \mathbb{R}$, with $\alpha, \beta \neq 0$, be such that $sign(\alpha) = sign(\beta)$ and let $x \in X$ and $p \in \mathbb{R}^n$. Then, we have*

$$[x + (\alpha + \beta)p]_{[l,u]} = \left[[x + \alpha p]_{[l,u]} + \beta p\right]_{[l,u]}.$$

Proof. We first recall that

$$[x_i + \alpha p_i]_{[l_i,u_i]} \begin{cases} l_i & \text{if} \quad x_i + \alpha p_i < l_i, \\ x_i + \alpha p_i & \text{if} \quad l_i \leq x_i + \alpha p_i \leq u_i, \\ u_i & \text{if} \quad x_i + \alpha p_i > u_i \end{cases}$$

for all $i = 1, \ldots, n$. In the case $l_i \leq x_i + \alpha p_i \leq u_i$ it is easy to see that

$$[x_i + (\alpha + \beta)p_i]_{[l_i,u_i]} = \left[[x_i + \alpha p_i]_{[l_i,u_i]} + \beta p_i\right]_{[l_i,u_i]}.$$

Now, we consider the case $x_i + \alpha p_i < l_i$. Since $x \in X$, we have $\alpha p_i < 0$ and, consequently, $\beta p_i \leq 0$. Thus, we can write

$$x_i + \alpha p_i + \beta p_i < l_i + \beta p_i \leq l_i,$$

which means

$$[x_i + (\alpha + \beta)p_i]_{[l_i, u_i]} = \left[[x_i + \alpha p_i]_{[l_i, u_i]} + \beta p_i\right]_{[l_i, u_i]}.$$

Similar reasoning can be done for the case $x_i + \alpha p_i > u_i$. □

By using Lemma 3.15 and (3.29a) and (3.29b), it results in

$$\left[x_{j_{k-1}} + \frac{\alpha_{j_{k-1}}}{\delta^{s_{k-1}+1}} d_{k-1}\right]_{[l,u]} = \left[x_{j_k} + \frac{(1-\delta)\alpha_{j_k}}{\delta} d_{k-1}\right]_{[l,u]},$$

so that (3.28) can be rewritten as

$$(3.32) \qquad Z\left(\left[x_{j_k} + \frac{(1-\delta)\alpha_{j_k}}{\delta} d_{k-1}\right]_{[l,u]}; \epsilon\right) \not< Z(x_{j_k}; \epsilon) - \gamma\left(\frac{1-\delta^2}{\delta^2}\right)\alpha_{j_k}^2 \mathbf{1}.$$

We show in the following proposition that the *projected expansion* cannot cycle.

PROPOSITION 3.16. *The* projected expansion *is such that*
(i) *the test at Step* 4 *is satisfied a finite number of times, i.e., the procedure cannot infinitely cycle;*
(ii) *the test at Step* 3 *is satisfied at least once, i.e.,* `Add&Filter` *is called at least once*

Proof. Point (i). We proceed by contradiction and assume that the test at Step 4 is always satisfied, i.e., a monotonically increasing sequence of positive numbers $\{\beta^j\}$ exists such that

$$Z([y + \beta^j p]_{[l,u]}; \epsilon) \not> Z(x_i; \epsilon) - \gamma(\beta^j)^2 \mathbf{1} \qquad \forall\, x_i \in \tilde{L},$$

and, in particular,

$$Z([y + \beta^j p]_{[l,u]}; \epsilon) \not> Z(y; \epsilon) - \gamma(\beta^j)^2 \mathbf{1}.$$

This means that, for any given j, an index $i \in \{1, \ldots, q\}$ exists such that

$$Z_i([y + \beta^j p]_{[l,u]}; \epsilon) \leq Z_i(y; \epsilon) - \gamma(\beta^j)^2.$$

Then, since q is finite, we can extract a subsequence of $\{\beta^j\}$ such that $i = \bar{\imath}$, i.e., we have

$$Z_{\bar{\imath}}([y + \beta^j p]_{[l,u]}; \epsilon) \leq Z_{\bar{\imath}}(y; \epsilon) - \gamma(\beta^j)^2.$$

The above relation contradicts the compactness of set X.

Point (ii). Let us consider a generic iteration of the *projected expansion*. Then, either $\alpha = \hat{\alpha}$ or $\alpha = \hat{\alpha}/\delta^r$ (with $r \in \mathbb{N}$, $r \geq 1$) and

$$\left(Z([y + \alpha p]_{[l,u]}; \epsilon) \not> Z(x_j; \epsilon) - \gamma\alpha^2 \mathbf{1}\right) \forall\, x_j \in \tilde{L}.$$

The above relations imply that, for all $x_j \in \tilde{L}$, an index $\ell_j \in \{1, \ldots, q\}$ exists such that

$$(3.33) \qquad Z_{\ell_j}([y + \alpha p]_{[l,u]}; \epsilon) \leq Z_{\ell_j}(x_j; \epsilon) - \gamma\alpha^2.$$

Furthermore, let us assume that the test at Step 3 is not satisfied, which means that

$$Z([y + \beta p]_{[l,u]}; \epsilon) < Z([y + \alpha p]_{[l,u]}; \epsilon) - \gamma \left(\beta^2 - \alpha^2\right) \mathbf{1}$$

and, in particular, in view of (3.33), for all $x_j \in \tilde{L}$, an index $\ell_j \in \{1, \ldots, q\}$ exists such that

$$Z_{\ell_j}([y + \beta p]_{[l,u]}; \epsilon) < Z_{\ell_j}([y + \alpha p]_{[l,u]}; \epsilon) - \gamma \left(\beta^2 - \alpha^2\right) \le Z_{\ell_j}(x_j; \epsilon) - \gamma \beta^2.$$

The above relation implies that the test at Step 4 is satisfied so that the procedure will perform a further iteration.

Now, let us assume by contradiction that the test at Step 3 is never satisfied. Then, the above reasoning would imply that the procedure infinitely cycles, which contradicts Point i). $\qquad\square$

Now, we prove a result concerning convergence to zero of the step sizes produced by algorithm DFMO.

PROPOSITION 3.17. *Let $\{L_k\}$ with $L_k = \{(x_j, \alpha_j), \ j = 1, \ldots, |L_k|\}$ be the sequence of sets of nondominated pairs produced by Algorithm DFMO. Then every linked sequence $\{(x_{j_k}, \alpha_{j_k})\}$ is such that*

$$\lim_{k \to \infty} \alpha_{j_k} = 0.$$

Proof. Let us first define the following set:

$$\chi_\epsilon = \{z \in \mathbb{R}^q : z = Z(x; \epsilon) \ \forall \ x \in X\}.$$

Note that the compactness of X and the continuity of the penalty functions $Z_j(x; \epsilon)$, $j = 1, \ldots, q$, imply that χ_ϵ is compact as well.

Now, we split the iteration sequence into two sets, namely, K_1 and K_2 such that
  1. for every $k \in K_1$, a success step is performed, i.e., (3.29) holds;
  2. for every $k \in K_2$, a failure step is performed, i.e., (3.31) holds.
We carry out the proof by showing that:

$$\lim_{k \to \infty, k \in K_1} \alpha_{j_k} = 0, \qquad \text{provided that } K_1 \text{ is infinite},$$

$$\lim_{k \to \infty, k \in K_2} \alpha_{j_k} = 0, \qquad \text{provided that } K_2 \text{ is infinite}.$$

First, note that K_1 and K_2 cannot both be finite. Then, let us first assume that K_1 is infinite and assume by contradiction that a subsequence $\bar{K} \subseteq K_1$ such that, for all $k \in \bar{K}$ sufficiently large, $\alpha_{j_k} \ge \bar{\alpha} > 0$. If this is the case, the algorithm generates infinitely many iterates x_{j_k} such that (3.32) is satisfied. This implies that the corresponding points $Z(x_{j_k}; \epsilon) \in \chi_\epsilon$ have a distance not smaller than $\gamma(\bar{\alpha})^2$ from each other. In this case, considering the compactness of χ_ϵ, we have

(3.34) $$\lim_{k \to \infty, k \in K_1} \alpha_{j_k} = 0.$$

Now, let us suppose that K_2 is an infinite subset and let m_k be the biggest integer such that $m_k < k$ with $m_k \in K_1$. Then $\alpha_{j_k} = \theta^{k - m_k} \alpha_{j_{m_k}}$ (we can assume $m_k = 0$ if the index m_k does not exist, that is, K_1 is empty). Now we can distinguish two cases:
  • $m_k \to \infty$ (namely, K_1 is an infinite subset). Then, by (3.34) and $\alpha_{j_k} = \theta^{k - m_k} \alpha_{j_{m_k}}$, we obtain

(3.35) $$\lim_{k \to \infty, k \in K_2} \alpha_{j_k} = 0.$$

- $(k - m_k) \to \infty$ (namely, K_1 is finite, i.e., m_k is finite). Then $\theta^{k-m_k} \to 0$, so that again

(3.36)
$$\lim_{k\to\infty, k\in K_2} \alpha_{j_k} = 0. \qquad \square$$

Hence, the proof is concluded by considering (3.34), (3.35), and (3.36).

Then, we report the following technical lemma that is used in the convergence proof of Algorithm DFMO.

LEMMA 3.18 (See [10, Lemma 2.6]).   *Let $\{x_k\}$ be a sequence of points, $\{d_k\}$ be the sequence of search directions used by DFMO, and let $\{\eta_k\}$ be a sequence such that $\eta_k > 0$ for all k. Further, let K be a subset of indices such that*

$$\lim_{k\to\infty, k\in K} x_k = \bar{x},$$
$$\lim_{k\to\infty, k\in K} d_k = \bar{d},$$
$$\lim_{k\to\infty, k\in K} \eta_k = 0$$

with $\bar{x} \in X$ and $\bar{d} \in D(\bar{x})$, $\bar{d} \neq 0$. Then,
   (i) *for all $k \in K$ sufficiently large,*

$$[x_k + \eta_k d_k]_{[l,u]} \neq x_k;$$

   (ii) *the following limit holds,*
$$\lim_{k\to\infty, k\in K} v_k = \bar{d},$$

   where

(3.37)
$$v_k = \frac{[x_k + \eta_k d_k]_{[l,u]} - x_k}{\eta_k}.$$

Finally, we can prove the main convergence result related to Algorithm DFMO. We would like to highlight the fact that, according to the following proposition, any accumulation point of any linked sequence generated by DFMO is a Pareto–Clarke KKT stationary point for problem (1.1).

PROPOSITION 3.19. *Let $\{L_k\}$ with $L_k = \{(x_j, \alpha_j), \ j = 1, \ldots, |L_k|\}$ be the sequence of sets of nondominated pairs produced by Algorithm DFMO. Let Assumption 2 hold. Let $\{(x_{j_k}, \alpha_{j_k})\}$ be a linked sequence and $\bar{x}$ be any limit point of $\{x_{j_k}\}$, i.e.,*

$$\lim_{k\to\infty, k\in K} x_{j_k} = \bar{x}$$

for a subset K of indices. Then, an $\epsilon^\star > 0$ exists such that for all $\epsilon \in (0, \epsilon^\star]$, if the subsequence $\{d_k\}_{k\in K}$ is dense in the unit sphere, $\bar{x}$ is a Pareto–Clarke KKT stationary point for problem (1.1).

   Proof. First we recall that, by Definition 3.8, $\bar{x}$ is a Pareto–Clarke–Jahn stationary point for problem 2.1 if, for all $\bar{d} \in D(\bar{x})$, an index $j_{\bar{d}}$ exists such that the following condition holds:

$$Z_{j_{\bar{d}}}^\circ(\bar{x}; \epsilon, \bar{d}) = \limsup_{\substack{y \to \bar{x}, y \in X, \\ t \downarrow 0, y + t\bar{d} \in X}} \frac{Z_{j_{\bar{d}}}(y + t\bar{d}; \epsilon) - Z_{j_{\bar{d}}}(y; \epsilon)}{t} \geq 0.$$

Taking into account Proposition 3.17, we have

$$\lim_{k \to \infty} \alpha_{j_k} = 0.$$

Then we prove that $\bar{x}$ is a Pareto–Clarke–Jahn stationary point according to Definition 3.8. To this aim, we proceed by contradiction and assume that a direction $\bar{d} \in D(\bar{x}) \cap S(0,1)$ exists such that, for all indices $j \in \{1, \dots, q\}$,

$$(3.38) \qquad Z_j^{\circ}(\bar{x}; \epsilon, \bar{d}) = \limsup_{\substack{x_{j_k} \to \bar{x}, x_{j_k} \in X, \\ t \downarrow 0, x_{j_k} + t\bar{d} \in X}} \frac{Z_j(x_{j_k} + t\bar{d}; \epsilon) - Z_j(x_{j_k}; \epsilon)}{t} < 0.$$

Moreover, by Proposition 3.17, and recalling that, by assumption, $\{d_k\}_{k \in K}$ is dense in the unit sphere and $\lim_{k \to \infty, k \in K} x_{j_k} = \bar{x}$, a subset $\check{K} \subseteq K$ exists such that

$$(3.39a) \qquad \lim_{k \to \infty, k \in \check{K}} x_{j_{k-1}} = \bar{x},$$

$$(3.39b) \qquad \lim_{k \to \infty, k \in \check{K}} d_{k-1} = \bar{d},$$

$$(3.39c) \qquad \lim_{k \to \infty, k \in \check{K}} \alpha_{j_k} = 0,$$

where $d_{k-1} \in D(x_{j_{k-1}})$. Considering a failure step, and taking into account (3.30) and (3.31), we can write

$$(3.40) \qquad Z_j\left(\left[x_{j_k} + \frac{\alpha_{j_k}}{\theta} d_{k-1}\right]_{[l,u]}; \epsilon\right) > Z_j(x_{\ell_j}; \epsilon) - \gamma\left(\frac{\alpha_{j_k}}{\theta}\right)^2$$

for $x_{\ell_j} \in \tilde{L}_{k-1}$ and for all $j \in \{1, \dots, q\}$.

Furthermore, considering a success step, and taking into account (3.32), we have that there exists an index $j \in \{1, \dots, q\}$ such that

$$(3.41) \qquad Z_j\left(\left[x_{j_k} + \frac{(1-\delta)\alpha_{j_k}}{\delta} d_{k-1}\right]_{[l,u]}; \epsilon\right) \geq Z_j\left(x_{j_k}; \epsilon\right) - \gamma\left(\frac{1-\delta^2}{\delta^2}\right)\alpha_{j_k}^2,$$

which can be rewritten as

$$Z_j\left(\left[x_{j_k} + \frac{(1-\delta)\alpha_{j_k}}{\delta} d_{k-1}\right]_{[l,u]}; \epsilon\right) \geq Z_j\left(x_{j_k}; \epsilon\right) - \gamma\frac{1-\delta^2}{(1-\delta)^2}\frac{(1-\delta)^2}{\delta^2}\alpha_{j_k}^2.$$

Then, we can define

$$\eta_{j_k} = \begin{cases} \dfrac{\alpha_{j_k}}{\theta} & \text{if (3.40) holds,} \\[2mm] \dfrac{(1-\delta)\alpha_{j_k}}{\delta} & \text{if (3.41) holds,} \end{cases}$$

and, considering expression (3.37) of Lemma 3.18, we also define

$$v_{j_k} = \frac{[x_{j_k} + \eta_{j_k} d_{k-1}]_{[l,u]} - x_{j_k}}{\eta_{j_k}}.$$

Furthermore, by point (i) of Lemma 3.18, we have, for all k sufficiently large, that

$$v_{j_k} \neq 0,$$

and, by point (ii) of Lemma 3.18,

$$\lim_{k\to\infty,k\in\breve{K}} v_{j_k} = \bar{d}.$$

Hence, for all $k \in \breve{K}$, the instructions of the Algorithm DFMO imply that, for at least an index $j \in \{1,\dots,q\}$,

$$Z_j(x_{j_k} + \eta_{j_k}v_{j_k};\epsilon) \ge Z_j(x_{j_k};\epsilon) - \gamma'\eta_{j_k}^2,$$

where $\gamma' = \gamma$ if j_k is a failure step and $\gamma' = \gamma(1-\delta^2)/(1-\delta)^2$ if j_k is a success step. Now, by considering that q is finite, we can extract a further subset of indices $\bar{K} \subset \breve{K}$ and an index $\bar{\jmath} \in \{1,\dots,q\}$, such that

$$Z_{\bar{\jmath}}(x_{j_k} + \eta_{j_k}v_{j_k};\epsilon) \ge Z_{\bar{\jmath}}(x_{j_k};\epsilon) - \gamma'\eta_{j_k}^2,$$

that is,

(3.42) $$\frac{Z_{\bar{\jmath}}(x_{j_k} + \eta_{j_k}v_{j_k};\epsilon) - Z_{\bar{\jmath}}(x_{j_k};\epsilon)}{\eta_{j_k}} \ge -\gamma'\eta_{j_k}.$$

Then,

$$\limsup_{\substack{y\to\bar{x},y\in X,\\ t\downarrow 0, y+t\bar{d}\in X}} \frac{Z_{\bar{\jmath}}(y + t\bar{d};\epsilon) - Z_{\bar{\jmath}}(y;\epsilon)}{t} \ge \limsup_{k\to\infty,\ k\in\bar{K}} \frac{Z_{\bar{\jmath}}(x_{j_k} + \eta_{j_k}\bar{d};\epsilon) - Z_{\bar{\jmath}}(x_{j_k};\epsilon)}{\eta_{j_k}}$$

$$= \limsup_{k\to\infty,\ k\in\bar{K}} \frac{Z_{\bar{\jmath}}(x_{j_k} + \eta_{j_k}\bar{d};\epsilon) + Z_{\bar{\jmath}}(x_{j_k} + \eta_{j_k}v_{j_k};\epsilon) - Z_{\bar{\jmath}}(x_{j_k} + \eta_{j_k}v_{j_k};\epsilon) - Z_{\bar{\jmath}}(x_{j_k};\epsilon)}{\eta_{j_k}}$$

$$\ge \limsup_{k\to\infty,\ k\in\bar{K}} \frac{Z_{\bar{\jmath}}(x_{j_k} + \eta_{j_k}v_{j_k};\epsilon) - Z_{\bar{\jmath}}(x_{j_k};\epsilon)}{\eta_{j_k}} - L_{\bar{\jmath}}\|\bar{d} - v_{j_k}\|,$$

where $L_{\bar{\jmath}}$ is the Lipschitz constant of $Z_{\bar{\jmath}}$. By (3.39a), (3.39b), and (3.42) and the above relation, we get

$$\limsup_{\substack{y\to\bar{x},y\in X,\\ t\downarrow 0, y+t\bar{d}\in X}} \frac{Z_{\bar{\jmath}}(y + t\bar{d};\epsilon) - Z_{\bar{\jmath}}(y;\epsilon)}{t} \ge 0,$$

which contradicts (3.38) and proves that $\bar{x}$ is a Pareto–Clarke–Jahn stationary point of problem (2.1). Then the proof follows by recalling Proposition 3.11. □

4. Numerical results. This section is devoted to the numerical experimentation of the DFMO algorithm both on bound constrained and nonlinearly constrained problems. The DFMO code is freely available at the URL http://www.dis.uniroma1.it/~lucidi/DFL. The aim of the tests on bound constrained problems is to understand if the theoretical properties of our linesearch approach (i.e., the fact that any accumulation point of any linked sequence is a Pareto stationary point) have some practical interest. For this reason the comparison is carried out with the globally convergent version of DMS [7], namely, the one using asymptotically dense sets of directions. In the second part, the effectiveness of the exact penalization technique for general constraints is analyzed. Again we study the behavior of DFMO in comparison with

DMS. To this aim, since the original version of the DMS code cannot handle nonlinear constraints, we embedded the penalty approach in the DMS framework. This also enabled us to show the versatility of the penalty approach, in that it can be easily used within algorithms for bound constraints multiobjective problems.

In Algorithm DFMO, each couple $(x_i, \alpha_i) \in \tilde{L}_k$ is explored by using the set of directions $\{D_k, -D_k\}$, where D_k is an orthonormal basis in $\mathbb{R}^n$ constructed starting from d_k, in place of the direction d_k alone. The directions belonging to D_k are used as in Algorithm CS-DFN from [10]. In particular, they are explored one after the other in search of a sufficent improvement. When such improvement cannot be obtained along d_k, $-d_k$ is explored before passing to the next direction. Obviously, using a positive basis rather than the single random direction d_k should enable us to get a better sampling of the functions and hence to guarantee better performance.

We also implemented a simplified version of DFMO, called DFMO$_\oplus$, which consists in always setting $D_k = \{I\}$, thus (possibly) considering the set of directions $\{I, -I\}$.

4.1. Bound constrained problems. We used the collection of problems defined in [7], i.e., a set of 100 multiobjective problems with number of variables $n \in [1, 30]$ and number of objectives $q = 2, 3$, and 4 (note that $q = 4$ only for one test problem, namely, FES3). The problems coded in FORTRAN90 are available at the URL http://www.dis.uniroma1.it/~lucidi/DFL (the original problems in AMPL format can be found at the URL http://www.mat.uc.pt/dms).

All the results described in this subsection have been obtained by allowing a maximum number of 20,000 function evaluations.

The comparison that we report is between version 0.2 of DMS [7] and DFMO. We remark that DMS has been run by using its default[2] settings except for: `tol_stop` $= 10^{-9}$ and `dir_dense` $= 1$. Hence, the variant of DMS we are using is the one referenced as DMS(n,line) in [7].

The results are reported in Figure 1 in terms of the purity [2], spread metrics Γ and Δ (both metrics as defined in [7]), and the hypervolume metric [25] by using performance profiles [9].

By taking a look at the figures, we notice that DFMO gives good performances in terms of purity, spread metric Γ, and hypervolume metric (both in terms of efficiency and robustness), while DMS is more efficient when considering spread Δ. This could be explained by the fact that the theoretical properties of our algorithm somehow help to generate a larger number of nondominated points (with respect to DMS). As a consequence this enables us to get a higher percentage of nondominated points (purity and hypervolume) and to reduce the maximum size of the holes in the Pareto front (spread Γ). On the other hand, the selection and the updating strategies used by DMS to evolve the list of nondominated points seem to allow a better distribution of the points in the Pareto front (spread Δ). A possible explanation of this feature is that the computational burden (in terms of function evaluations) connected with the exploration carried out by DMS is better balanced than the one related to exploration used in DFMO (which is anyway needed to guarantee the theoretical properties of the algorithm). More specifically, at each iteration DMS analyzes a single point before updating the list, while DFMO must visit all the points in the list before the updating.

---

[2]It should be noted that by "default" setting we mean the setting of DMS given by the file parameters_dms.m included in the DMS software package.
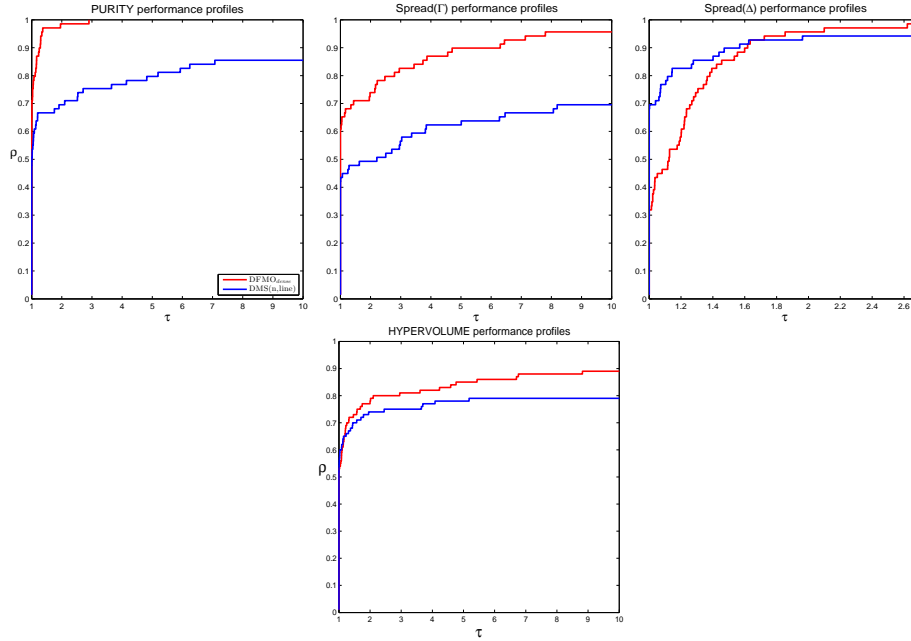
Fig. 1. *Comparison between DFMO and DMS(n,line) on the bound constrained problems by using performance profiles and by means of the four metrics, i.e., purity (left), spread(Γ) (center), spread(Δ) (right), and hypervolume (bottom).*

As a further experiment, we run DMS(n,line) in its default settings, except for `tol_stop` $= 10^{-9}$ (that is, we leave `dir_dense` $= 0$). This variant of DMS, which we denote by DMS(n,line)$_\oplus$, is the one using the coordinate directions as search directions. Then, we adopt the same strategy and use algorithm DFMO$_\oplus$. The results of this comparison are reported in Figure 2. As can be seen, the situation is rather different than the previous one. Indeed, now DMS(n,line)$_\oplus$ is better than DFMO$_\oplus$ except that DFMO$_\oplus$ is slightly more robust in terms of the purity metric. Hence, it is clear that the choice of search directions has a considerable impact on the performance of the two algorithms. Then, to better understand the influence of the search directions on the algorithms we did a further comparison which we report in Figure 3. Namely, we compare the four methods, DMS(n,line), DMS(n,line)$_\oplus$, DFMO$_{dense}$, and DFMO$_\oplus$, against each other. As can be noted, the use of the constant set of coordinate directions as search directions, at least in the case of bound constrained problems, seems to be sufficient to collect enough information on the objective functions.

4.2. Nonlinearly constrained problems. To the best of our knowledge, in the literature there is no standard collection of nonlinearly constrained multiobjective test problems. Hence, we defined a new collection by coupling a subset of the bound constrained problems [7] used in the preceding subsection with six families of constraints proposed in [18]. In more detail, we selected 51 bound constrained problems, i.e., all the problems with $n \geq 3$ variables. Then, we defined a set of 306 constrained multiobjective problems by adding to each problem the following six families of nonlinear constraints, namely,
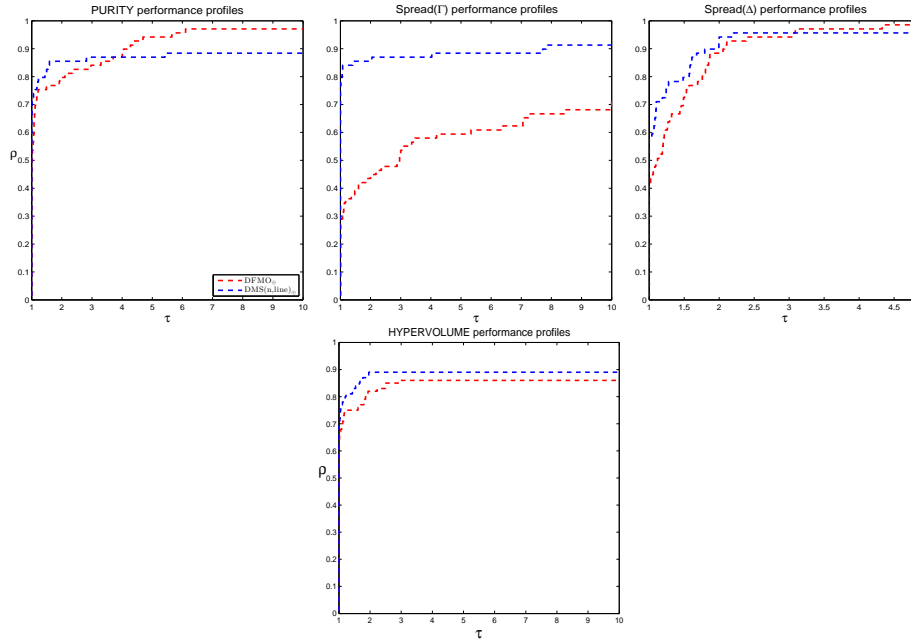
FIG. 2. *Comparison between* $DFMO_\oplus$ *and* $DMS(n,line)_\oplus$ *on the bound constrained problems by using performance profiles and by means of the three metrics, i.e., purity (left), spread(*Γ*) (center), spread(*Δ*) (right), and hypervolume (bottom).*
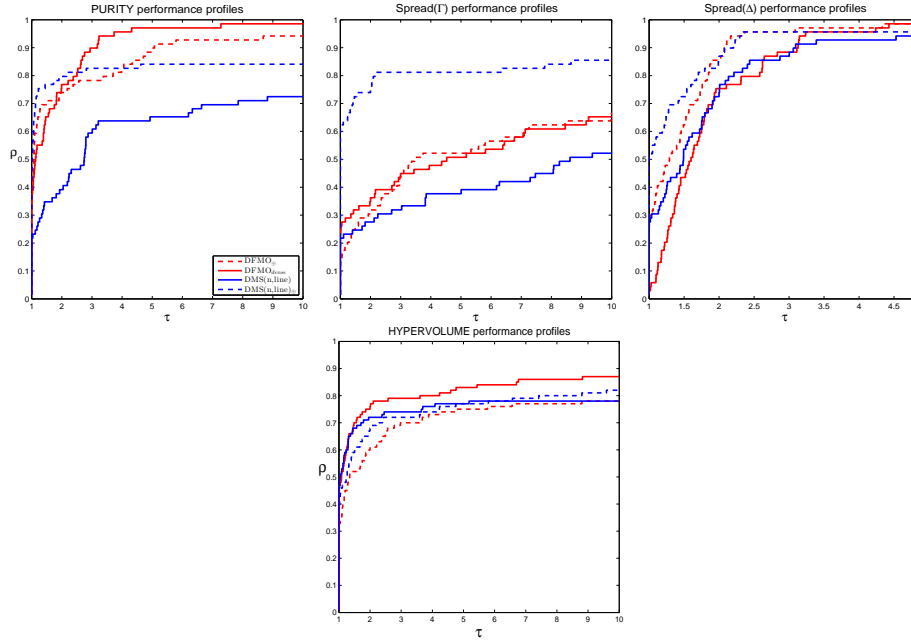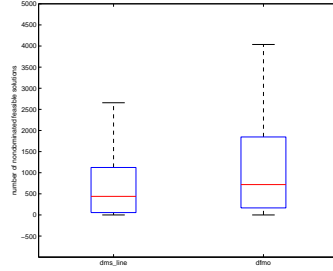

FIG. 3. *Comparison between* $DMS(n,line)$, $DMS(n,line)_\oplus$, $DFMO_{dense}$, *and* $DFMO_\oplus$ *on the bound constrained problems by using performance profiles and by means of the three metrics, i.e., purity (left), spread(*Γ*) (center), spread(*Δ*) (right), and hypervolume (bottom).*

FIG. 4. *Comparison between DFMO and DMS(n,line)—number of nondominated feasible solutions.*

$$g_j(x) = (3-2x_{j+1})x_{j+1}-x_j-2x_{j+2}+1 \le 0 \qquad \forall\, j=1,\ldots,m,\ m=n-2;$$
$$g_j(x) = (3-2x_{j+1})x_{j+1}-x_j-2x_{j+2}+2.5 \le 0 \qquad \forall\, j=1,\ldots,m,\ m=n-2;$$
$$g_j(x) = x_j^2 + x_{j+1}^2 + x_j x_{j+1}-2x_j-2x_{j+1}+1 \le 0 \qquad \forall\, j=1,\ldots,m,\ m=n-1;$$
$$g_j(x) = x_j^2 + x_{j+1}^2 + x_j x_{j+1}-1 \le 0 \qquad \forall\, j=1,\ldots,m,\ m=n-1;$$
$$g_j(x) = (3-0.5x_{j+1})x_{j+1}-x_j-2x_{j+2}+1 \le 0 \qquad \forall\, j=1,\ldots,m,\ m=n-2;$$
$$g_j(x) = \sum_{i=1}^{n+1}((3-0.5x_{j+1})x_{j+1}-x_j-2x_{j+2}+1) \le 0 \quad \forall\, j=1,\ldots,m,\ m=1.$$

Finally, by a preliminary investigation we found out that 92 problems out of the 306 are inherently infeasible, thus we dropped them from the test set. This is due to the combination of the nonlinear constraints with the bound constraints of the original problems. Hence, we have a set of 214 nonlinearly constrained multiobjective problems with $n \in [3,30]$, $m \in [1,29]$, and $q \in [2,4]$.

In order to manage the general constraints, we used a vector of penalty parameters $\epsilon \in \mathbb{R}^m$ and considered the penalty functions

$$Z_j(x;\epsilon) = f_j(x) + \sum_{i=1}^{m} \frac{1}{\epsilon_i}\max\{0, g_i(x)\} \quad \forall\, j=1,\ldots,q,$$

which trivially preserve all the theoretical results proved in section 3.2. The vector of penalty parameters is set as follows:

$$\epsilon_i = \left\{ \begin{array}{ll} 10^{-3} & \text{if } \max\{0, g_i(x_0)\} < 1, \\ 10^{-1} & \text{otherwise,} \end{array} \right. \qquad i=1,\ldots,m.$$

Once again, all the results have been obtained by allowing a maximum number of 20,000 function evaluations.

We embed in both DFMO and DMS(n,line) the penalty function. Also in this case DMS(n,line) has been run by using its default settings except for `tol_stop` $= 10^{-9}$ and `dir_dense` $= 1$. First of all, we notice that both DFMO and DMS(n,line) exhibit a considerable ability to produce feasible points on all the considered problems. This is confirmed by the box plots, reported in Figure 4, which are related to the distribution of the number of nondominated feasible solutions found by each algorithm. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers were dropped out. By observing the figure, we notice that both methods are able to find a consistent number of nondominated feasible solutions, which, in our opinion, confirms the efficiency of the proposed penalty approach.
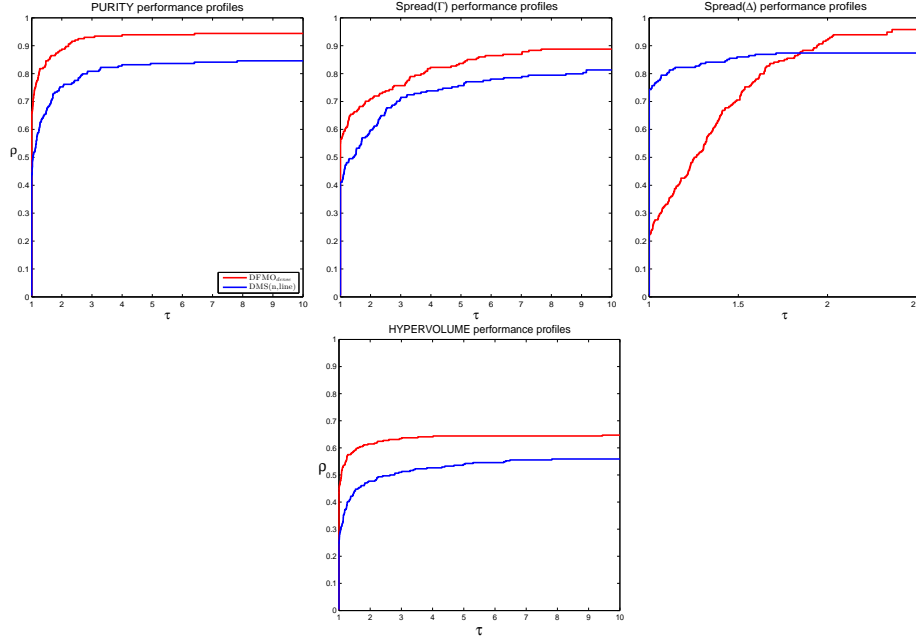
FIG. 5. *Comparison between* $DFMO_{dense}$ *and DMS(n,line) on the constrained problems by using performance profiles and by means of the three metrics, i.e., purity (left), spread(Γ) (center), spread(Δ) (right), and hypervolume (bottom).*
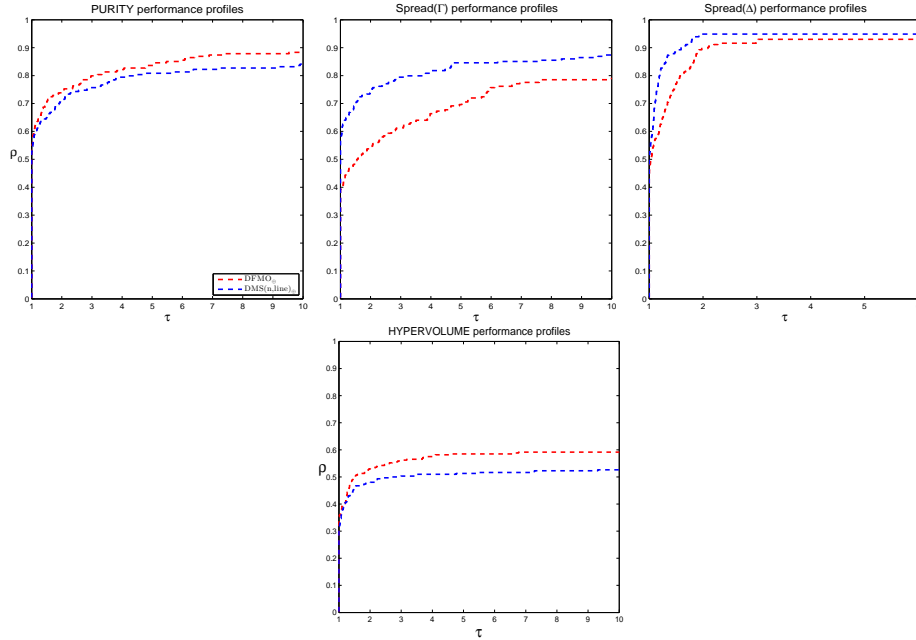

FIG. 6. *Comparison between* $DFMO_{\oplus}$ *and* $DMS(n,line)_{\oplus}$ *on the constrained problems by using performance profiles and by means of the three metrics, i.e., Purity (left), Spread(Γ) (center), Spread(Δ) (right), and hypervolume (bottom).*
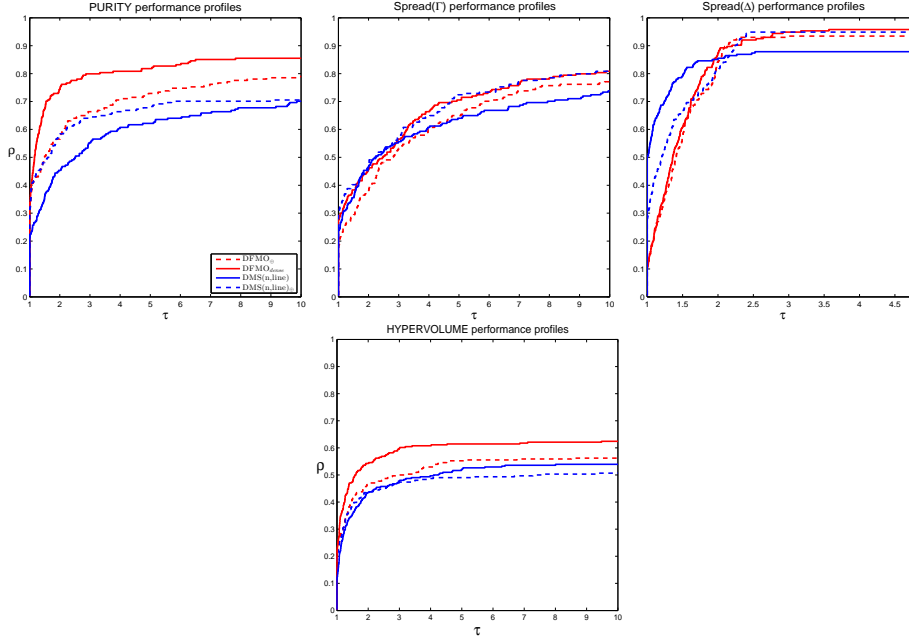
Fig. 7. *Comparison between DMS(n,line), DMS(n,line)$_\oplus$, DFMO$_{dense}$ and DFMO$_\oplus$ on the constrained problems by using performance profiles and by means of the three metrics, i.e., Purity (left), Spread(Γ) (center), Spread(Δ) (right), and hypervolume (bottom).*

In Figure 5, we report the results in terms of the purity, spread metrics Γ and Δ, and hypervolume metric by using performance profiles. The figures show that DFMO$_{dense}$ and DMS(n,line) have a behavior similar to the bound constrained case. Anyway, we notice an improvement of the performances of DMS. This, in our opinion, could be due to the fact that, as noticed in the previous subsection, DMS has the ability to uniformly distribute the points on the Pareto front. As a further experiment, we compared DFMO$_\oplus$ and DMS(n,line)$_\oplus$, i.e., the versions using a constant set of coordinate directions as the set of search directions. The results of this latter comparison are reported in Figure 6. Looking at the figure, we provide some observations. First of all, the good ability of DMS(n,line)$_\oplus$ to uniformly spread the generated points on the computed Pareto front still emerges, as in the bound constrained case. Second, it can be noted that, for constrained problems, DFMO$_\oplus$ is better than DMS(n,line)$_\oplus$ at generating nondominated points.

Finally, to better understand the influence of the search directions on the algorithms, in Figure 7 we report the comparison of the four methods, DMS(n,line), DMS(n,line)$_\oplus$, DFMO$_{dense}$, and DFMO$_\oplus$, against each other. The first thing that we note is that, in the constrained case, the use of a richer set of search directions considerably helps the methods to produce good nondominated points.

5. Conclusions. In this paper, we described a new exact-penalty-based derivative-free approach (based on linesearches with sufficient decrease) for nonlinearly constrained MOO problems. In order to develop the method, we first analyzed the theoretical properties of the class of problems to handle. In particular, we described new optimality conditions that explicitly take into account the bound constraints and that are obtained by only assuming Lipschitz continuity of the problem

functions. We further proved that the original problem is equivalent to a bound constrained problem obtained by penalizing the nonlinear constraints with an exact merit function. Furthermore, thanks to this exact penalty, we were able to adapt the derivative-free approach in [10] to the MOO case. The sufficient decrease gave us the chance to avoid the use of integer lattices. The extrapolation phase also allowed us to prove, under some density assumptions on the search directions, convergence to a set of Pareto stationary points (i.e., we prove that any accumulation point of the sequences generated by our method is a Pareto stationary point). The exact penalty approach gave us the freedom to choose an infeasible starting point (which can be a big advantage in practice).

Finally, we tested our approach on both bound constrained and nonlinearly constrained problems. The goal of the tests on bound constrained problems was understanding how much the theoretical properties of our method are helpful in practice. On the other hand, the tests on nonlinearly constrained problems were carried out to show the effectiveness of the exact penalty approach when embedded on a DFO method for bound constrained multiobjective problems. The reported results both showed that, when dealing with MOO problems, the linesearch approach can be very helpful in practice in identifying a consistent number of good nondominated points, and that the merit function introduced to handle the nonlinear constraints, when embedded on a globally convergent DFO method, guarantees good performance and enables us to find a significant number of feasible points.

REFERENCES

[1] C. AUDET, G. SAVARD, AND W. ZGHAL, *Multiobjective optimization through a series of single-objective formulations*, SIAM J. Optim., 19 (2008), pp. 188–210.

[2] S. BANDYOPADHYAY, S. K. PAL, AND B. ARUNA, *Multiobjective GAs, quantitative indices, and pattern classification*, IEEE Trans. Syst. Man Cybernet. Part B Cybernet. 34 (2004), pp. 2088–2099.

[3] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[4] A. CONN, K. SCHEINBERG, AND L. N. VICENTE, *Introduction to Derivative-Free Optimization*, MPS-SIAM Ser. Optim. 8, SIAM, Philadelphia, 2009.

[5] A. L. CUSTÓDIO, M. EMMERICH, AND J. F. A. MADEIRA, *Recent developments in derivative-free multiobjective optimization*, Comput. Technol. Rev., 5 (2012), pp. 1–30.

[6] A. L. CUSTÓDIO, J. F. A. MADEIRA, A. I. F. VAZ, AND L. N. VICENTE, *Errata to Direct Multisearch for Multiobjective Optimization*, http://www.mat.uc.pt/~lnv/papers/errata-dms.pdf.

[7] A. L. CUSTÓDIO, J. F. A. MADEIRA, A. I. F. VAZ, AND L. N. VICENTE, *Direct multisearch for multiobjective optimization*, SIAM J. Optim., 21 (2011), pp. 1109–1140.

[8] K. DEB, A. PRATAP, S. AGARWAL, AND T. A. M. T. MEYARIVAN, *A fast and elitist multiobjective genetic algorithm:* IEEE Trans. Evol. Comput., 6 (2002), pp. 182–197.

[9] E. D. DOLAN AND J. J. MORÉ, *Benchmarking optimization software with performance profiles*, Math. Program., 91 (2002), pp. 201–213.

[10] G. FASANO, G. LIUZZI, S. LUCIDI, AND F. RINALDI, *A linesearch-based derivative-free approach for nonsmooth constrained optimization*, SIAM J. Optim., 24 (2014), pp. 959–992.

[11] E. H. FUKUDA, L. M. GRAÑA DRUMMOND, AND F. M. P. RAUPP, *An external penalty-type method for multicriteria*, TOP, 24 (2016), pp. 493–513.

[12] T. GLAD AND E. POLAK, *A multiplier method with automatic limitation of penalty growth*, Math. Program., 17 (1979), pp. 140–155.

[13] J. B. HIRIART-URRUTY, *On optimality conditions in nondifferentiable programming*, Math. Program., 14 (1978), pp. 73–86.

[14] X. X. HUANG AND X. Q. YANG, *Nonlinear Lagrangian for multiobjective optimization and applications to duality and exact penalization*, SIAM J. Optim., 13 (2002), pp. 675–692.

[15] Y. ISHIZUKA AND K. SHIMIZU, *Necessary and sufficient conditions for the efficient solutions of nondifferentiable multiobjective problems*, IEEE Trans. Syst. Man Cybernet., SMC-14 (1984), pp. 624–629.

[16] J. JAHN, *Introduction to the Theory of Nonlinear Optimization*, Springer, Berlin, 1996.

[17] J. JAHN, *Vector Optimization*, Springer, Berlin, 2009.

[18] N. KARMITSA, *Test Problems for Large-Scale Nonsmooth Minimization*, Technical report No., B. 4/2007, Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland, 2007.

[19] C.-J. LIN, S. LUCIDI, L. PALAGI, A. RISI, AND M. SCIANDRONE, *Decomposition algorithm model for singly linearly-constrained problems subject to lower and upper bounds*, J. Optim. Theory Appl., 141 (2009), pp. 107–126.

[20] G. LIUZZI, S. LUCIDI, F. PARASILITI, AND M. VILLANI, *Multiobjective optimization techniques for the design of induction motors*, IEEE Trans. Magnet., 39 (2003), pp. 1261–1264.

[21] S. LUCIDI, *New results on a continuously differentiable exact penalty function*, SIAM J. Optim., 2 (1992), pp. 558–574.

[22] O. L. MANGASARIAN, *Nonlinear Programming*, Classics Appl. Math., SIAM, Philadelphia, 1994.

[23] K. SHIMIZU, Y. ISHIZUKA, AND J. F. BARD, *Nondifferentiable and Two-level Mathematical Programming*, Kluwer, Norwell, MA, 1997.

[24] A. SUPPAPITNARM, K. A. SEFFEN, G. T. PARKS, AND P. J. CLARKSON, *A simulated annealing algorithm for multiobjective optimization*, Eng. Optim., 33 (2000), pp. 59–85.

[25] E. ZITZLER AND L. THIELE, *Multiobjective Optimization using Evolutionary Algorithms — A Comparative Case Study*, in Parallel Problem Solving from Nature — PPSN V: 5th International Conference Amsterdam, The Netherlands, A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, eds., Springer, Berlin 1998, pp. 292–301.