Computational Optimization and Applications
https://doi.org/10.1007/s10589-026-00822-z

A nonmonotone front descent method for bound-constrained multi-objective optimization

Pierluigi Mansueto[1]

Received: 4 March 2026 / Accepted: 30 July 2026

Abstract
This paper introduces a nonmonotone extension of the Front Descent method, a state-of-the-art descent-based algorithmic framework designed to approximate the Pareto front of smooth multiobjective optimization problems. The proposed approach incorporates novel nonmonotone line search strategies that allow temporary increases in some objective functions, potentially accelerating convergence and improving overall efficiency. The theoretical analysis demonstrates that the sequence of sets generated by the algorithm retains the convergence properties of the original framework. Establishing these properties in the nonmonotone setting is nontrivial: the loss of monotonicity across the set sequence introduces substantial analytical challenges, necessitating a careful and rigorous adaptation of the original convergence arguments. Finally, numerical results in the bound-constrained setting are presented to validate the goodness of the proposed approach.

Keywords Multi-objective optimization · Pareto front · Non-monotone line search

Mathematics Subject Classification 90C29 · 90C30

1 Introduction

In this paper, we consider multi-objective optimization problems of the form

$$\min_{x \in \Omega} F(x) = (f_1(x), \dots, f_m(x))^\top, \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is assumed to be continuously differentiable and the feasible set is defined by the box constraints

$$\Omega = [l, u] \subseteq \mathbb{R}^n, \tag{2}$$

✉ Pierluigi Mansueto
pierluigi.mansueto@unifi.it

1 Department of Information Engineering, University of Florence, Via di S. Marta, 3, Florence 50139, Italy

Springer

with $l, u \in \mathbb{R}^n$ representing lower and upper bounds, respectively. Given the vector-valued nature of the objective function, a single point that minimizes all components simultaneously rarely exists. Consequently, the solution of (1) is understood in terms of Pareto optimality, where the goal is to identify points for which no objective can be improved without worsening at least one other. For a comprehensive formal treatment of these concepts, the reader is referred to [1].

Multi-objective optimization has also flourished in recent years due to its potential impact on practical applications and its integration with other mathematical optimization frameworks. Fields where decision-makers must balance conflicting criteria are quite common, including engineering [2, 3], statistics [4], and space exploration [5, 6]. Furthermore, the development of algorithms aimed at approximating the Pareto frontier of multi-objective optimization problems has become an important research area in other branches of mathematical optimization, such as portfolio optimization [7, 8] and black-box optimization [9, 10], where analyzing trade-offs in complex decision-making scenarios is critical.

Among the diverse approaches for multi-objective optimization, descent methods [11–15] have attracted significant and increasing attention. These techniques extend the classical paradigm of scalar gradient-based optimization to the multi-objective setting by identifying directions that ensure a simultaneous decrease for all objectives. While these methods were originally conceived as *single-point* algorithms, i.e., aiming to converge toward an individual Pareto optimal solution, they have more recently been generalized to *front-based* frameworks. Such evolution is designed to compute entire sets of points that collectively approximate the Pareto frontier (see, e.g., [16–19]). By maintaining an approximation of the Pareto frontier throughout the iterative process, these methods have emerged as highly competitive and mathematically rigorous alternatives to traditional scalarization techniques [20, 21] and heuristic-driven evolutionary algorithms [22].

Parallel to these developments, nonmonotone strategies are widely used in scalar optimization to improve efficiency [23, 24]. Their main idea is to relax the requirement of a decrease in the objective function at every iteration, requiring instead a sufficient decrease compared to a benchmark, such as the maximum objective function value over the last $M > 0$ iterations. In scalar problems, allowing temporary increases in the objective function often helps in avoiding slow convergence and escaping local minima [25]. Despite these advantages, nonmonotone methods have only recently been applied to multi-objective optimization [25–27]. These contributions, which are seminal for the introduction of nonmonotone strategies in multi-objective optimization settings, however consist of single-point methodologies that handle sequences of points in order to find a single solution, hopefully lying on the Pareto frontier; therefore, in order to obtain an approximation of the Pareto frontier, such methodologies must be executed in a multi-start fashion, by initializing their execution from multiple different starting points; as one might expect, the effectiveness of obtaining a good approximation of the Pareto frontier strongly depends on such initial points (see, for instance, [16, Section 6]). To our knowledge, the integration of nonmonotone strategies into descent methods designed to handle sequences of sets of points and, as a consequence, to directly return an approximation, as accurate as possible, of the entire Pareto front remains unexplored. Adapting these techniques to *front-based* algorithms is a nontrivial task, as it requires defining suitable benchmarks for an entire set of points rather than a single candidate solution.

In this work, we fill this gap by introducing a nonmonotone extension of the Front Descent method [18] for bound-constrained smooth multi-objective optimization. Front Descent is an advanced front-based descent approach for which recent convergence results concerning the generated sequences of sets have been established [19]. We first introduce a variant of this method, called Front Projected Descent, which is designed to handle box constraints, and we also report the corresponding modifications to the theoretical analysis in [19] required to accommodate this setting. We then propose a set of algorithmic modifications that enable controlled, nonmonotone increases in objective function values within the front-based framework. Crucially, we prove that these modifications preserve the desirable convergence properties established in [19]. Although the resulting convergence results are analogous to those of the monotone case, their derivation requires a substantial and non-trivial analytical effort to account for the lack of monotonicity across the sequence of fronts.

The remainder of the paper is organized as follows. In Sect. 2, we describe the Front Projected Descent method along with its main theoretical properties, and highlight how these properties are adapted with respect to those in [19] to account for box constraints. Section 3 presents our novel nonmonotone variant for bound-constrained multi-objective optimization, including a detailed discussion of the mechanisms of the new proposal and a formal theoretical analysis. Section 4 discusses the results of numerical experiments conducted to validate the framework effectiveness and efficiency w.r.t. the original one. Finally, Sect. 5 offers concluding remarks and suggests directions for future research.

2 The front projected descent method

We introduce the *Front Projected Descent* (FPD) method, a variant of the *Front Descent* (FD) framework [19], originally designed for *unconstrained* multi-objective problems, which is extended here to handle the box constraints of problem (1). Importantly, the modifications introduced in FPD preserve the convergence properties of the original framework.

Given the standard partial ordering in $\mathbb{R}^m$, i.e.,

$$\begin{aligned} u \leq v &\iff u_j \leq v_j, \ \forall j = 1, \dots, m, \\ u < v &\iff u_j < v_j, \ \forall j = 1, \dots, m, \\ u \lneqq v &\iff u \leq v \wedge u \neq v, \end{aligned}$$

the goal of first-order methods, like FPD, is to identify solutions $\bar{x} \in \Omega$, with Ω defined as in Eq. (2), satisfying the following properties listed in order of decreasing strength:

- *Pareto optimal:* $\nexists\, y \in \Omega$ s.t. $F(y) \lneqq F(\bar{x})$;
- *Weak Pareto optimal:* $\nexists\, y \in \Omega$ s.t. $F(y) < F(\bar{x})$;
- *Pareto stationary:*

$$\min_{d \in \mathcal{D}(\bar{x})} \max_{j=1,\dots,m} \nabla f_j(\bar{x})^\top d = 0,$$

where $\mathcal{D}(\bar{x}) = \{v \in \mathbb{R}^n \mid \exists \bar{t} > 0 : \bar{x} + tv \in \Omega \; \forall t \in [0, \bar{t}\,]\}$ is the set of feasible directions at $\bar{x}$.

Pareto optimality is the strongest property, implying the others. Pareto stationarity, instead, is only a necessary first-order condition for weak Pareto optimality; first-order methods are typically proven to converge to Pareto stationary points (see, e.g., [11, 16]); however, this condition becomes sufficient for optimality only under appropriate convexity assumptions. Since minimizing all objectives simultaneously is generally impossible, multiple Pareto optimal solutions (constituting the *Pareto set*) usually arise, forming in the objectives space the *Pareto front*. Decision makers can evaluate these trade-offs *a posteriori* to select the most suitable solutions.

To approximate the entire Pareto front, FPD iteratively updates a set X^k of mutually nondominated solutions, i.e., no $y \in X^k$ exists such that $F(y) \lneqq F(x)$ for $x \in X^k$. The next set X^{k+1} is obtained by search steps from $\hat{x} \in X^k$ along:

- the *projected common descent direction* [12]:

$$v(\hat{x}) = \operatorname*{arg\,min}_{\substack{d \in \mathbb{R}^n \\ \hat{x}+d \in \Omega}} \max_{j=1,\dots,m} \nabla f_j(\hat{x})^\top d + \frac{1}{2}\|d\|^2; \tag{3}$$

- the *projected partial descent directions* [12, 16]: for subset $I \subset \{1, \dots, m\}$,

$$v^I(\hat{x}) = \operatorname*{arg\,min}_{\substack{d \in \mathbb{R}^n \\ \hat{x}+d \in \Omega}} \max_{j \in I} \nabla f_j(\hat{x})^\top d + \frac{1}{2}\|d\|^2. \tag{4}$$

Both problems admit unique solutions due to strong convexity and continuity of the objectives, and convexity of the feasible sets. Their optimal values, $\theta(\hat{x})$ and $\theta^I(\hat{x})$, are negative whenever a descent direction exists. Naturally, v^I and θ^I coincide with v and θ, respectively, when $I = \{1, \dots, m\}$. Moreover, as established in [12], θ is continuous and v satisfies

$$\|v(x)\| \leq 2\|J_F(x)\|, \tag{5}$$

where $J_F(x) \in \mathbb{R}^{m \times n}$ is the Jacobian of $F(x)$. A point $\bar{x}$ is Pareto stationary if and only if $\theta(\bar{x}) = 0$ and $v(\bar{x}) = 0$. Note then that it always holds that $\theta(x) \leq 0$. Finally, defining $D(x, d) = \max_{j=1,\dots,m} \nabla f_j(x)^\top d$, we have $J_F(x)v(x) \leq \mathbf{1}D(x, v(x)) \leq \mathbf{1}\theta(x)$, with $\mathbf{1}$ being the vector of all ones in $\mathbb{R}^m$.

2.1 Algorithmic scheme

The FPD scheme is presented in Algorithm 1. For clarity, three key phases (steps 3-7-21), crucial for our nonmonotone variant, are left in a general form. In FPD, these are handled as follows: in step 3, we set $\hat{X}^k = X^k$; in step 7, we perform the Armijo-type line search for multi-objective optimization

$$\alpha_p^k = \max_{h \in \mathbb{N}} \left\{ \alpha_0 \delta^h \,\middle|\, \begin{array}{l} x_p + \alpha_0 \delta^h v(x_p) \in \Omega \,\wedge \\ F(x_p + \alpha_0 \delta^h v(x_p)) \leq F(x_p) + \mathbf{1}\gamma \alpha_0 \delta^h D(x_p, v(x_p)) \end{array} \right\};$$

finally, in step 21, we set $X^{k+1} = \hat{X}^k$. Note that, in the classical Armijo-type (non-monotone) line search for multi-objective optimization (see, e.g., [11, 25]), the sufficient decrease term depends on $J_F(x_p)v(x_p)$ rather than on $D(x_p, v(x_p))$ as in [19]. Since, as already mentioned at the end of Sect. 2, it holds that $J_F(x_p)v(x_p) \leq \mathbf{1}D(x_p, v(x_p))$, the corresponding sufficient decrease condition is therefore somewhat less restrictive.

Algorithm 1 *Front Projected Gradient*

Require: $F : \mathbb{R}^n \to \mathbb{R}^m$, $X^0 \subset \Omega = [l, u]$ set of mutually nondominated points w.r.t. F, $\alpha_0 \in (0, 1]$, $\delta \in (0, 1)$, $\gamma \in (0, 1)$, $\{\sigma_k\} \subseteq \mathbb{R}_+$.

1: $k = 0$
2: **while** a stopping criterion is not satisfied **do**
3:   Create the set $\hat{X}^k$ based on X^k
4:   **for all** $x_p \in X^k$ **do**
5:     **if** $\nexists y \in \hat{X}^k$ s.t. $F(y) \lneqq F(x_p)$ **then**
6:       **if** $\theta(x_p) < -\sigma_k$ **then**
7:         Run an Armijo-type line search to get α_p^k
8:       **else**
9:         $\alpha_p^k = 0$
10:       **end if**
11:       $z_p^k = x_p + \alpha_p^k v(x_p)$
12:       $\hat{X}^k = \hat{X}^k \setminus \{y \in \hat{X}^k \mid F(z_p^k) \lneqq F(y)\} \cup \{z_p^k\}$
13:       **for all** $I \subset \{1, \ldots, m\}$ s.t. $\theta^I(z_p^k) < 0$ **do**
14:         **if** $z_p^k \in \hat{X}^k$ **then**
15:

$$\alpha_p^I = \max_{h \in \mathbb{N}} \left\{ \alpha_0 \delta^h \,\middle|\, \begin{array}{l} z_p^k + \alpha_0 \delta^h v^I(z_p^k) \in \Omega \wedge \\ \forall y \in \hat{X}^k,\ \exists j : f_j(z_p^k + \alpha_0 \delta^h v^I(z_p^k)) < f_j(y) \end{array} \right\}$$

16:           $\hat{X}^k = \hat{X}^k \setminus \{y \in \hat{X}^k \mid F(z_p^k + \alpha_p^I v^I(z_p^k)) \lneqq F(y)\} \cup \{z_p^k + \alpha_p^I v^I(z_p^k)\}$
17:         **end if**
18:       **end for**
19:     **end if**
20:   **end for**
21:   Create the set X^{k+1} based on $\hat{X}^k$
22:   $k = k + 1$
23: **end while**
24: **return** X^k

At each iteration k, `FPD` updates a list of solutions X^k by processing each point $x_p \in X^k$ sequentially. During this process, new solutions may be generated (see below) and added to the set $\hat{X}^k$. Each point x_p is then processed only if it is not dominated by any solution contained in $\hat{X}^k$ at the time of its processing (see the condition in step 5). The procedure applied to x_p consists of a two-phase optimization scheme:

- Phase 1 (steps 6-12) acts as a refinement step, consisting of a single-point projected descent from x_p. This yields a new point z_p^k that provides a sufficient decrease in all objective functions; if x_p is already approximately Pareto-stationary, no refinement is performed and $z_p^k = x_p$. As in [19], the approximation degree of

Pareto stationarity is controlled by a sequence $\{\sigma_k\} \subseteq \mathbb{R}_+$ (where $\mathbb{R}_+$ denotes the set of nonnegative real numbers), which allows us to formally cover both the asymptotic regime $\sigma_k \to 0$, under which arbitrary accuracy can still be achieved in the limit, and the more practical setting in which a constant $\sigma > 0$ is used for all k. The latter reflects computational practice, where nonzero thresholds are required to ensure robustness in finite-precision arithmetic.

- Phase 2 (steps 13-16) enriches X^k with new nondominated points to explore the objectives space further. Starting at z_p^k, projected partial descent steps are taken along directions with $\theta^I(z_p^k) < 0$, as long as z_p^k is not dominated (i.e., as long as the condition at step 14 is satisfied; if it is not, it means that a solution that dominates z_p^k is added during the for-loop execution at step 16 and, as a consequence, such for-loop started at step 13 can be terminated). The line search considers the entire updated set $\hat{X}^k$ and must produce a point not dominated by any current solution.

Whenever a new point is generated, either in Phase 1 or Phase 2, it is added to the set $\hat{X}^k$ at the end of the corresponding phase (step 12 or step 16). Each time a new solution is added to $\hat{X}^k$, a filtering procedure is also applied at steps 12–16, with the aim of removing from $\hat{X}^k$ all previously stored solutions y that are dominated by the newly added point.

2.2 Theoretical properties

In this section, we present the theoretical analysis of `FPD`, addressing first its well-definedness and then its main convergence properties. This analysis closely follows that of the `FD` framework in [19]; here, we mainly highlight the modifications required to handle the box constraints of problem (1), as well as some definitions and technical lemma needed for the subsequent theoretical study of our non-monotone algorithm.

We begin with three lemmas concerning the finite termination property of the line searches employed in `FPD`, which in turn guarantees the well-definedness of the algorithm.

Lemma 1 *The line search at step 7 of* `FPD` *is well defined if x_p is feasible.*

Proof First, we note that, by definition of the projected common descent direction (3), the feasibility of x_p, the convexity of the feasible set Ω and the definition of α_0 in Algorithm 1, we have that, for all $\alpha \in [0, \alpha_0]$, $x_p + \alpha v(x_p) \in \Omega$.

Then, recalling that $\theta(x_p) < -\sigma_k \leq 0$ by step 6, we have $D(x_p, v(x_p)) < 0$ and thus, by [11, Lemma 4], we get the result. □

Lemma 2 *Step 15 of Algorithm 1 is well defined if z_p^k is feasible and nondominated with respect to points in $\hat{X}^k$.*

Proof Reminding that, by definition of the projected partial descent direction (4), the feasibility of z_p^k, the convexity of the feasible set Ω and the definition of α_0 in Algorithm 1, we have that, for all $\alpha \in [0, \alpha_0]$, $z_p^k + \alpha v^I(z_p^k) \in \Omega$, we obtain the result by following the proof of [18, Proposition 3.2]. □

Lemma 3 *Let X^k be a set of mutually nondominated feasible points. Then*

1. *throughout the entire iteration k, $\hat{X}^k$ is a set of mutually nondominated feasible points;*
2. *the line searches at steps 7-15 of* `FPD` *are always well defined;*
3. *X^{k+1} is a set of mutually nondominated feasible solutions.*

Proof At iteration k, the set $\hat{X}^k$ is initialized with the nondominated, feasible points X^k; then, it is only updated at steps 12 and 16. At step 12, either $z_p^k = x_p$, and the set is not modified, or, by the definition of α_p^k, z_p^k is feasible and dominates x_p, which in turn was nondominated. Thus, the added point z_p^k is nondominated, while all the newly dominated points are removed. At step 16, the added point $z_p^k + \alpha_p^I v^I(z_p^k)$ is feasible and nondominated by the definition of α_p^I; all the newly dominated points are removed. Thus, $\hat{X}^k$ always contains mutually nondominated feasible solutions. Steps 7-15 are therefore always well defined by Lemmas 1-2, respectively. Moreover, since $X^{k+1} = \hat{X}^k$ at the end of the iteration, X^{k+1} inherits the feasibility and nondominance properties from $\hat{X}^k$. □

The convergence properties of `FPD`, concerning the sequence of sets of generated points, are based on the following two definitions and technical lemma. It is worth pointing out that the box constraints of problem (1) do not affect the validity of the lemma nor its proof.

Definition 1 *([19, Definition 5.11])* Let $\zeta \in \mathbb{R}^m$ be a reference point and let $Y \subseteq \mathbb{R}^m$ be a (possibly infinite) set of points. We define the *dominated region* as $\Lambda(Y) = \{y \in \mathbb{R}^m \mid \exists\, \bar{y} \in Y : \ \bar{y} \le y \le \zeta\}$. In addition, the *hyper-volume* [28] associated to Y is defined as the volume (or the Lebesgue measure) of the set $\Lambda(Y)$ and is denoted by $V(Y)$.

It should be noted that a "better" approximation of the Pareto frontier is generally associated with a higher hyper-volume value. For the sake of simplicity, given $X \subset \Omega$, we will use the notation $\Lambda_F(X) = \Lambda(F(X))$ and $V_F(X) = V(F(X))$, with $F(X)$ being the image of X through F. If X only contains mutually non-dominated points, we have $F(X) = (F(X))_{nd} = \{y \in F(X) \mid \nexists \bar{y} \in F(X) \text{ s.t. } \bar{y} \lneqq y\}$ and $F(X)$ is called *stable set*.

Lemma 4 ([19, Lemma 5.12]) *Let $\zeta \in \mathbb{R}^m$ be a reference point and let X be a set of points such that $Y = F(X)$ is a stable set and, for all $y \in Y$, $y \le \zeta$. Let $\bar{x} \in X$ and $\mu \in \mathbb{R}^n$ s.t. $F(\mu) < F(\bar{x})$. Then, the set $Z = F(X \cup \{\mu\} \setminus \{\bar{x}\})$ is such that*

$$V(Z) - V(Y) \ge \prod_{j=1}^{m} (f_j(\bar{x}) - f_j(\mu)) > 0.$$

Definition 2 *([19, Definition 5.13])* Let $\mathcal{X}$ be the set of all sets $X \subseteq \mathbb{R}^n$ of mutually nondominated points w.r.t. F, i.e., $X \in \mathcal{X}$ if $F(X)$ is a stable set. We define the map $\Theta : \mathcal{X} \to \mathbb{R}$ as $\Theta(X) = \inf_{x \in X} \theta(x)$.

The last definition introduces a Pareto-stationarity measure for sets of mutually nondominated points. For finite sets X, the infimum is a minimum, so $\Theta(X)$ returns the common projected descent value θ of the "least" Pareto-stationary point in X. By construction, $\Theta(X)$ is always nonpositive and equals zero if and only if all points in X are Pareto-stationary.

Before stating the convergence properties, we introduce the following assumption, which will also be essential for the theoretical analysis of our nonmonotone variant.

Assumption 1 At each iteration k of Algorithm 1, the first point to be processed in the for loop of steps 4-20 belongs to the set $\arg\min_{x \in X^k} \theta(x)$.

Proposition 5 *Let X^0 be a set of mutually nondominated feasible points and $x_0 \in X^0$ be a point such that the set $\mathcal{L}(x_0) = \bigcup_{j=1}^m \{x \in \Omega \mid f_j(x) \leq f_j(x_0)\}$ is compact. Let $\{X^k\}$ be the sequence of sets of nondominated points produced by* `FPD` *under Assumption 1. Then,*

1. *if $\sigma_k = \sigma > 0$ for all k, there exists $\bar{k}$ such that $\Theta(X^k) \geq -\sigma$ for all $k \geq \bar{k}$;*
2. *if $\sigma_k \to 0$, $\lim_{k\to\infty} \Theta(X^k) = 0$.*

Proof For the sake of brevity, the proof is moved to Appendix B. □

Corollary 6 *Let $\{X^k\}$ be the sequence of sets of nondominated points produced by* `FPD` *under the assumptions of Proposition 5 and with $\sigma_k \to 0$. Let $\{x^k\}$ be any sequence such that $x^k \in X^k$ for all k, then it admits accumulation points and every accumulation point is Pareto stationary for problem (1).*

Proof Following reasoning similar to that used at the beginning of the proof of Proposition 5, we can show that $\{x^k\} \subseteq \mathcal{L}(x_0)$. The remainder of the proof is then identical to that of [19, Corollary 5.18], with Proposition 5 replacing Theorem 5.15. □

Remark 1 As explained in [19], Assumption 1 prevents that `FPD` repeatedly discard points before they undergo phase 1, which would otherwise slow progress toward stationarity despite continual, but insufficient improvements. Moreover, the assumption can arguably be considered non-restrictive, as the gradients and values of θ computed at the beginning of iteration k to search for the $\arg\min_{x \in X^k} \theta(x)$ can be stored in memory and reused later when processing the points in X^k, thus keeping the extra computational cost minimal.

3 A nonmonotone front projected descent algorithm

The nonmonotone variant of `FPD`, which we call `FPD_NMT`, differs from the original algorithm in steps 3–7–21, which are detailed in Procedures 1–2–3, respectively. Appendix A also provides the complete `FPD_NMT` scheme, corresponding to Algorithm 1 with these three procedures integrated. Unlike `FPD`, where the new set X^{k+1} is generated solely based on the previous iterate set X^k, `FPD_NMT` performs the search steps with respect to a new *reference set*.

Definition 3 Let k be an iteration of Algorithm 1. We define the *reference set* C^k w.r.t. X^k as a set of solutions such that:

1. C^k contains only mutually nondominated points;
2. $\forall y \in C^k, \exists x \in X^k$ such that $F(x) \leq F(y)$;
3. $\forall x \in X^k, \exists y \in C^k$ such that $F(x) \leq F(y)$.

The reference set is thus essentially a "worse" set than X^k in terms of the objective functions: for each point in C^k, there exists a point in X^k whose objective function values are equal to or better than those of the point in C^k.

Procedure 1 Step 3 of Algorithm 1 for `FPD_NMT`

1: $X^{l(k)} \in \underset{X \in \{X^k, X^{k-1}, \dots, X^{k-\min\{k,M\}}\}}{\arg\min} V_F(X)$
2: $\bar{C}^k = \{x \in X^{l(k)} \cup C^{k-1} \mid \nexists y \in X^{l(k)} \cup C^{k-1} : F(y) \lneqq F(x)\}$
3: $C^k = \bar{C}^k \cup \{x \in X^k \mid \forall y \in \bar{C}^k \; \exists j \in \{1, \dots, m\} : f_j(y) < f_j(x)\}$
4: $\hat{X}^k = C^k$

Starting from $C^{-1} = \emptyset$, at each iteration k of the `FPD_NMT` algorithm, C^k is created via Procedure 1. In step 1, we select the set with the lowest hyper-volume among X^k and the previous M sets ($M \in \mathbb{N}$ is a parameter of `FPD_NMT`); if $k < M$, only the k previous sets are considered. In order to calculate the hyper-volume metric, we use the reference point r_p such that, for all $j \in \{1, \dots, m\}$, $(r_p)_j = \max_{x \in X_{\text{all}}} f_j(x)$, with X_{all} being the union of all solutions generated up to iteration k. In step 2, the chosen set $X^{l(k)}$ is merged with the previous reference set C^{k-1}, removing the dominated solutions, to form $\bar{C}^k$. Finally, in step 3, C^k is obtained by adding to $\bar{C}^k$ the points from X^k that are strictly worse in at least one objective w.r.t. each point in $\bar{C}^k$. Step 4 initializes $\hat{X}^k$ equal to C^k.

Procedure 2 Step 7 of Algorithm 1 for `FPD_NMT`

1: **if** $\exists c \in \hat{X}^k : F(x_p) \leq F(c)$ **then**
2:
$$\alpha_p^k = \max_{h \in \mathbb{N}} \left\{ \alpha_0 \delta^h \;\middle|\; \begin{array}{l} x_p + \alpha_0 \delta^h v(x_p) \in \Omega \; \wedge \\ \left(\begin{array}{l} \exists c_p^k \in \hat{X}^k : F(x_p) \leq F(c_p^k) \; \wedge \\ F(x_p + \alpha_0 \delta^h v(x_p)) \leq F(c_p^k) + \mathbf{1} \gamma \alpha_0 \delta^h D(x_p, v(x_p)) \end{array} \right) \end{array} \right\}$$
3: **else**
4:
$$\alpha_p^k = \max_{h \in \mathbb{N}} \left\{ \alpha_0 \delta^h \;\middle|\; \begin{array}{l} x_p + \alpha_0 \delta^h v(x_p) \in \Omega \; \wedge \\ F(x_p + \alpha_0 \delta^h v(x_p)) \leq F(x_p) + \mathbf{1} \gamma \alpha_0 \delta^h D(x_p, v(x_p)) \end{array} \right\}$$
5: **end if**

We now describe Procedure 2, concerning the Armijo-type line search procedure at step 7 of Algorithm 1. There are two cases: (i) there exists a solution $c \in \hat{X}^k$ that is equal to or worse than the current point x_p; note that, since $\hat{X}^k$ is initialized equal to C^k, if C^k is a reference set this condition always holds for the first point $x_p \in X^k$ being processed; in this case, we run step 2 of the procedure, where the goal is to find a new feasible point that achieves sufficient decrease in all the objective functions with respect to at least one point c_p^k satisfying the condition at step 1 (in fact, the point c_p^k may not be unique); (ii) no such solution $c \in \hat{X}^k$ exists, because it was removed during filtering; in this case, the monotone Armijo-type line search is applied (step 4). In both cases, the sufficient decrease term depends on $D(x_p, v(x_p))$, similarly to the monotone setting (see Sect. 2.1).

Procedure 3 Step 21 of Algorithm 1 for `FPD_NMT`

1: $\bar{X}^{l(k)} \in \underset{X \in \{\hat{X}^k\} \cup \{X^k, X^{k-1}, \dots, X^{k-\min\{k, M-1\}}\}}{\arg\min} V_F(X)$

2: $X^{k+1} = \{x \in \hat{X}^k \cup \bar{X}^{l(k)} \mid \nexists y \in \hat{X}^k \cup \bar{X}^{l(k)} : F(y) \lneqq F(x)\}$

After processing all points in X^k, the new set X^{k+1} is generated via Procedure 3. Let $\bar{X}^{l(k)}$ be the set with the lowest hyper-volume value among $\hat{X}^k$ and the previous $M - 1$ sets; X^{k+1} then consists of all mutually non-dominated points in $\hat{X}^k \cup \bar{X}^{l(k)}$. This procedure ensures that no point in X^{k+1} is dominated by any solution contained in the "worst" set in terms of hyper-volume.

3.1 Analysis of the `FPD_NMT` mechanisms

Through the following proposition and corollary, we show that constructing C^k according to Procedures 1-3 ensures that C^k forms a reference set for X^k at each iteration k of the `FPD_NMT` algorithm. Moreover, we prove that the sets X^k and C^k at each iteration consist entirely of feasible solutions.

Proposition 7 *Let k be an iteration of `FPD_NMT` and X^k be a set of mutually non-dominated points where, $\forall y \in C^{k-1} \cup X^{l(k)}$, $\exists x \in X^k$ such that $F(x) \leq F(y)$. Thus, C^k is a reference set w.r.t. X^k. Moreover, X^{k+1} is a set of mutually non-dominated points where, $\forall y \in C^k \cup \bar{X}^{l(k)}$, $\exists x \in X^{k+1}$ such that $F(x) \leq F(y)$.*

Proof First, we prove that C^k satisfies the three properties defining a reference set (Definition 3) one at a time.

1. The set C^k is created through steps 2-3 of Procedure 1. In step 2, we get that $\bar{C}^k \subseteq X^{l(k)} \cup C^{k-1}$, with all the dominated solutions in $X^{l(k)} \cup C^{k-1}$ being filtered out. We then assume, by contradiction, that at step 3 a point $\bar{x} \in X^k$ is added to C^k and there exists $\bar{y} \in \bar{C}^k$ such that $F(\bar{y}) \lneqq F(\bar{x})$. Being $\bar{y} \in \bar{C}^k \subseteq X^{l(k)} \cup C^{k-1}$, we know by hypothesis that there exists $\tilde{x} \in X^k$ such that $F(\tilde{x}) \leq F(\bar{y}) \lneqq F(\bar{x})$. We thus get a contradiction as X^k is assumed to be a set of mutually non-dominated points.
2. By step 3 of Procedure 1, we distinguish two types of points: $y \in C^k \cap \bar{C}^k$ and $y \in C^k \cap X^k$. For the latter, we trivially get that $F(x) \leq F(y)$, with $x = y \in X^k$. Thus, let us consider a point $y \in C^k \cap \bar{C}^k$. Since, by step 2, $\bar{C}^k \subseteq X^{l(k)} \cup C^{k-1}$, by hypothesis there exists $x \in X^k$ such that $F(x) \leq F(y)$.

3. By contradiction, we assume that there exists $x \in X^k$ such that, for all $y \in C^k$, there exists $j \in \{1, \dots, m\}$ satisfying $f_j(y) < f_j(x)$. Thus, it must hold that $x \notin C^k$, i.e., it cannot have been added at step 3 of Procedure 1. Therefore, it must exists $y \in C^k$ such that $F(x) \leq F(y)$. This leads to a contradiction, thereby concluding the proof.

We thus prove that C^k is a reference set w.r.t. X^k.

By step 2 of Procedure 3, we get that $X^{k+1} \subseteq \hat{X}^k \cup \bar{X}^{l(k)}$, with all the dominated points in $\hat{X}^k \cup \bar{X}^{l(k)}$ being filtered out; thus, X^{k+1} is a set of mutually nondominated points. Now, we distinguish two cases: $y \in (\hat{X}^k \cup \bar{X}^{l(k)}) \cap X^{k+1}$ and $y \notin (\hat{X}^k \cup \bar{X}^{l(k)}) \cap X^{k+1}$. In the first case, we trivially get that $F(x) \leq F(y)$, with $x = y \in X^{k+1}$. In the second one, we know that there exists $x \in (\hat{X}^k \cup \bar{X}^{l(k)}) \cap X^{k+1}$ such that $F(x) \lneqq F(y)$. We can then conclude that

$$\forall y \in \hat{X}^k \cup \bar{X}^{l(k)},\ \exists x \in X^{k+1} : F(x) \leq F(y). \tag{6}$$

Let us now consider the set C^k and $\hat{X}^k$. We observe that $\hat{X}^k$ is initialized as C^k at step 4 of Procedure 1 and subsequently updated at steps 12-16 of Algorithm 1, where a filtering operation removes the solutions dominated by the newly added points. Thus, it follows that, throughout the entire iteration k, $\forall y \in C^k$ there exists $x \in \hat{X}^k$ such that $F(x) \leq F(y)$. This result, together with (6), concludes the proof. □

Before stating the corollary, we first need to state an assumption.

Assumption 2 For all iterations $k > 0$ of `FPD_NMT`, the sets $\bar{X}^{l(k-1)}$ and $X^{l(k)}$ are equal.

Corollary 8 *Let Assumption 2 hold and X^0 be a set of mutually nondominated feasible points. Then, for every iteration k of `FPD_NMT`, Proposition 7 holds, and all points in X^{k+1} and C^{k+1} are feasible.*

Proof When $k = 0$, we observe the following: by hypothesis, X^0 is a set of mutually nondominated points; by definition of $X^{l(k)}$ at step 1 of Procedure 1, $X^{l(0)} = X^0$; therefore, given that $C^{-1} = \emptyset$, $C^{-1} \cup X^{l(0)} = X^0$. Thus, all the hypotheses of Proposition 7 are satisfied for $k = 0$.

Now, since $\hat{X}^0 = C^0 = X^0$, the set $\hat{X}^0$ is initialized with feasible solutions and is subsequently updated only at steps 12-16 of Algorithm 1 by adding new points, with feasibility guaranteed by the definitions of the step sizes α_p^k and α_p^l in Procedure 2 and step 15 of Algorithm 1. It thus follows by Procedure 3 that the set X^1 consists entirely of feasible points. The same holds for C^1, which is constructed at the next iteration through Procedure 1 using the sets C^0, X^0, and X^1.

The case of a generic iteration $k > 0$ follows by induction from Assumption 2, Proposition 7 and similar feasibility arguments as in the base case. □

Remark 2 Assumption 2 accounts for the possible non-uniqueness of the minimizers in Procedures 3 (iteration $k-1$) and 1 (iteration k). This assumption is in fact crucial from a theoretical standpoint, since it is essential for guaranteeing that the conclusion of Proposition 7 at iteration $k-1$, namely,

$$\forall y \in C^{k-1} \cup \bar{X}^{l(k-1)}, \exists x \in X^k \text{ such that } F(x) \leq F(y),$$

implies the corresponding hypothesis of the same proposition at iteration k, i.e.,

$$\forall y \in C^{k-1} \cup X^{l(k)}, \exists x \in X^k \text{ such that } F(x) \leq F(y).$$

This implication constitutes the basis of the inductive proof of Corollary 8.

In addition to being essential for the theoretical analysis, this assumption can be also easily satisfied in practice. Indeed, we observe that the collections of sets involved in the two minimization operations between iterations $k-1$ and k differ only in their first element: specifically, the minimization performed at Step 1 of Procedure 3 is over

$$\{\hat{X}^{k-1}\} \cup \{X^{k-1}, X^{k-2}, \dots, X^{k-\min\{k,M\}}\},$$

whereas the minimization at Step 1 of Procedure 1 considers

$$\{X^k\} \cup \{X^{k-1}, X^{k-2}, \dots, X^{k-\min\{k,M\}}\}.$$

Accordingly, two cases may occur.

- If
$$\bar{X}^{l(k-1)} \in \{X^{k-1}, X^{k-2}, \dots, X^{k-\min\{k,M\}}\},$$
then
$$V_F(X^k) \geq \max\{V_F(\hat{X}^{k-1}), V_F(\bar{X}^{l(k-1)})\} \geq V_F(\bar{X}^{l(k-1)}),$$
where the first inequality follows directly from Step 2 of Procedure 3 and the second from the definition of the maximum operator. Hence, X^k is the only set whose hyper-volume may increase w.r.t. the one in $\hat{X}^{k-1}$, while all the remaining sets remain unchanged. Therefore, $\bar{X}^{l(k-1)}$ remains a valid minimizer at iteration k, and it is always possible to choose $X^{l(k)} = \bar{X}^{l(k-1)}$.
- If $\bar{X}^{l(k-1)} = \hat{X}^{k-1}$, then Step 2 of Procedure 3 yields $V_F(X^k) = V_F(\hat{X}^{k-1}) = V_F(\bar{X}^{l(k-1)})$, that is, X^k coincides with $\hat{X}^{k-1}$ and is associated with the minimum hyper-volume value. Consequently, it can be selected at iteration k as $X^{l(k)} = X^k$.

This observation also provides a practical computational advantage. Indeed, the minimizer computed at Step 1 of Procedure 3 can be reused in Step 1 of Procedure 1 at the next iteration, thereby reducing the computational effort associated with these two minimization operations.

In Fig. 1, we show the image sets of X^k and C^k under the mapping F, obtained at iterations $k = 5, 15, 60, 100$ for the CEC09_2 problem [29] ($n = 5$) using the `FPD_NMT` algorithm with $M = 10$. In the early iterations, the two image sets are distant, with $F(C^k)$ composed of points equal to or dominated by at least one point in $F(X^k)$. As the algorithm progresses, the distance between the two sets decreases. Using solutions in C^k for the Armijo-type line search at step 2 of Procedure 2 likely

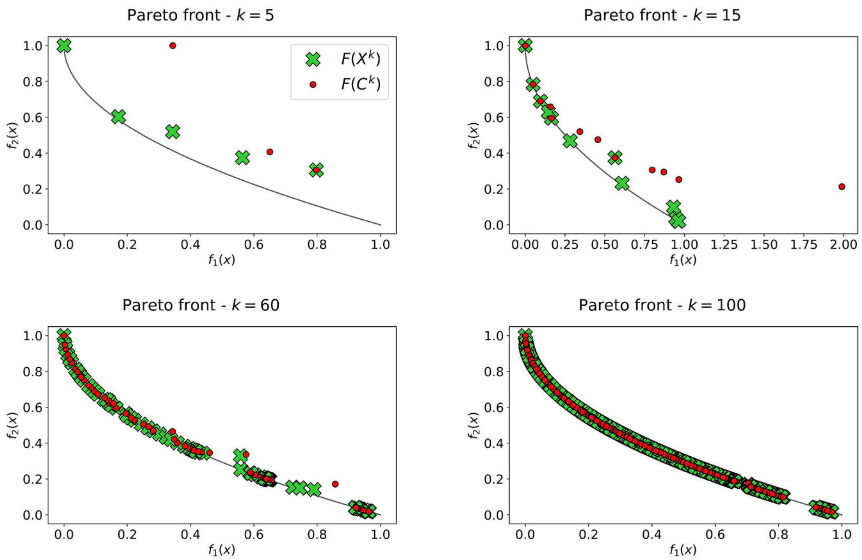

Fig. 1 Image sets of X^k and C^k through the mapping of F for the CEC09_2 problem [29] ($n = 5$) obtained at iteration $k = 5, 15, 60, 100$ by the `FPD_NMT` algorithm with $M = 10$

allows for larger steps compared to the classical procedure at step 4, with some resulting points potentially worsening in terms of (subsets of) objective functions w.r.t. the current points in X^k. This approach is analogous to nonmonotone methods in scalar optimization: avoiding a monotone decrease of the objective functions may improve the overall convergence speed of the algorithm. Note that the line search in step 15 of Algorithm 1 is also performed with respect to the solutions in $\hat{X}^k$, which is initialized as C^k at each iteration k of `FPD_NMT`. Consequently, the selected step sizes are likely to be larger in this phase as well, facilitating a broader exploration of the feasible set.

3.2 Convergence analysis

We begin the analysis of the `FPD_NMT` algorithm by establishing the finite termination properties of the employed line searches, which ensure the overall well-definedness of the method. We then present several technical lemmas required for the convergence analysis, which concludes with the main convergence guarantees.

Lemma 9 *The line searches at steps 2-4 of Procedure 2 are well defined if x_p is feasible.*

Proof First, we note that, by definition of the projected common descent direction (3), the feasibility of x_p, the convexity of the feasible set Ω and the definition of α_0 in Algorithm 1, we have that, for all $\alpha \in [0, \alpha_0]$, $x_p + \alpha v(x_p) \in \Omega$.

Now, since $J_F(x_p)v(x_p) \leq \mathbf{1}D(x_p, v(x_p))$ and the if condition at step 6 of Algorithm 1 ensures that $\theta(x_p) < -\sigma_k \leq 0$, by [11, Lemma 4] there exists $\bar{\alpha} > 0$ such that

$$F(x_p + \alpha v(x_p)) < F(x_p) + \mathbf{1}\gamma\alpha D(x_p, v(x_p)), \quad \forall \alpha \in (0, \bar{\alpha}].$$

Thus, line search at step 4 of Procedure 2 is eventually satisfied. As for the one at step 2 of Procedure 2, we have, for all $\alpha \in (0, \bar{\alpha}]$,

$$F(x_p + \alpha v(x_p)) < F(x_p) + \mathbf{1}\gamma\alpha D(x_p, v(x_p)) \leq F(c_p^k) + \mathbf{1}\gamma\alpha D(x_p, v(x_p)),$$

where c_p^k is well defined by the if condition at step 1 of Procedure 2. Hence, the proof is complete. □

Lemma 10 *Let Assumption 2 hold, X^0 be a set of mutually nondominated feasible points and k be an iteration of* `FPD_NMT` *(Algorithm 1). The set $\hat{X}^k$ contains nondominated feasible points at any time; thus, steps 2-4 of Procedure 2 and step 15 of Algorithm 1 are always well defined.*

Proof At iteration k, the set $\hat{X}^k$ is initialized with the points in C^k, which, by Corollary 8, Proposition 7 and the definition of reference set (Definition 3), are all mutually nondominated and feasible. The set $\hat{X}^k$ is subsequently updated exclusively at steps 12 and 16 of Algorithm 1. In the former step, three cases can occur:

1. $z_p^k = x_p$, which is feasible by Corollary 8 and was guaranteed to be nondominated by the if-condition at step 5 of Algorithm 1;
2. α_p^k is defined at step 2 of Procedure 2; in this case, z_p^k is feasible and dominates $c_p^k \in \hat{X}^k$, which was nondominated as it was not filtered out by earlier executions of steps 12 and 16 of Algorithm 1;
3. α_p^k is defined at step 4 of Procedure 2; in this case, z_p^k is feasible and dominates x_p, which was again nondominated.

Thus, regardless of the case, the added point z_p^k is nondominated, and all the newly dominated points are removed. At step 16 of Algorithm 1, the new point $z_p^k + \alpha_p^I v^I(z_p^k)$ is feasible and nondominated by definition of α_p^I; all the newly dominated points are once again removed. Thus, $\hat{X}^k$ always contains mutually nondominated feasible solutions and, as a consequence, steps 2-4 of Procedure 2 and step 15 of Algorithm 1 are always well defined by Lemmas 9-2, respectively. □

Lemma 11 *Let Assumption 2 hold, X^0 be a set of mutually nondominated points and k be an iteration of* `FPD_NMT`*. Then:*

1. $V_F(C^k) \geq \max(V_F(X^{l(k)}), V_F(C^{k-1}))$;
2. $\forall y \in C^k$, $\forall \tilde{k} \geq k$, $\exists x \in X^{\tilde{k}}$ *such that* $F(x) \leq F(y)$.

Proof First, we prove that, for all $y \in C^{k-1}$, there exists $x \in C^k$ such that $F(x) \leq F(y)$, which is a crucial property to prove the two theses. Note that the case $k = 0$ is trivial by definition of C^{-1} (see Sect. 3).

Let us consider then a point $y \in C^{k-1}$, with $k > 0$. By step 2 of Procedure 1, we have two cases: $y \in C^{k-1} \cap \bar{C}^k$ or $y \notin C^{k-1} \cap \bar{C}^k$. In the first case, since, by step 3 of Procedure 1, $\bar{C}^k \subseteq C^k$, we get that $F(x) \leq F(y)$, with $x = y \in C^k$. In the second one, by step 2 of Procedure 1 we know that there exists $x \in X^{l(k)} \cap \bar{C}^k$ such that $F(x) \lneqq F(y)$. Recalling again that $\bar{C}^k \subseteq C^k$, it holds that $x \in C^k$. Thus, we have proved that, for all $y \in C^{k-1}$, there exists $x \in C^k$ such that $F(x) \leq F(y)$.

Now, let us prove the two properties one at time.

1. By previous result, Proposition 7 and the definition of reference set (Definition 3), we know that
$$(F(C^k \cup C^{k-1}))_{nd} = (F(C^k))_{nd} = F(C^k).$$
Moreover, we observe that $C^{k-1} \subseteq C^{k-1} \cup C^k$ implies that $\Lambda_F(C^{k-1}) \subseteq \Lambda_F(C^{k-1} \cup C^k)$. By the properties of the dominated region (Definition 1), we then have
$$\Lambda_F(C^k) = \Lambda((F(C^k \cup C^{k-1}))_{\mathrm{nd}}) = \Lambda_F(C^k \cup C^{k-1}) \supseteq \Lambda_F(C^{k-1}).$$
Thus, we can also write that $V_F(C^k) \geq V_F(C^{k-1})$. Note that, following reasoning similar to that at the beginning of the proof, we can prove that, for all $y \in X^{l(k)}$, there exists $x \in C^k$ such that $F(x) \leq F(y)$. Thus, using a similar argument as the one presented here, we can obtain that $V_F(C^k) \geq V_F(X^{l(k)})$. By combining this result with $V_F(C^k) \geq V_F(C^{k-1})$, we conclude the proof.
2. By Corollary 8 and definition of the reference set, for all $y \in C^{\tilde{k}}$, there exists $x \in X^{\tilde{k}}$ such that $F(x) \leq F(y)$. If $\tilde{k} = k$, the proof is thus complete; otherwise, by applying an argument analogous to that used at the end of the proof of Proposition 7, we can combine this result with the property established at the beginning of the proof and proceed inductively to obtain the claim.

□

Lemma 12 *After step 12 of* `FPD_NMT` *(Algorithm 1), z_p^k belongs to $\hat{X}^k$. Moreover, there exists $y \in X^{k+1}$ such that $F(y) \leq F(z_p^k)$.*

Proof The result follows as in [18, Lemma 3.1] with $\tilde{k} = k + 1$, recalling that the set X^{k+1} is the result of repeated application of steps 12 and 16 of Algorithm 1 and of the execution of step 2 of Procedure 3, starting from $\hat{X}^k$ at some point when $z_p^k \in \hat{X}^k$. □

We are prepared to state the convergence properties of the algorithm.

Theorem 13 *Let X^0 be a set of mutually nondominated feasible points and $x_0 \in X^0$ be a point such that the set $\mathcal{L}(x_0) = \bigcup_{j=1}^m \{x \in \Omega \mid f_j(x) \leq f_j(x_0)\}$ is compact. Let $\{X^k\}$ be the sequence of sets of nondominated points produced by* `FPD_NMT` *under Assumptions 1-2. Then, there exists at least a subsequence $K \subseteq \{0, 1, \dots\}$ such that*

1. *if $\sigma_k = \sigma > 0$ for all k, there exists $\bar{k} \in K$ s.t. $\Theta(X^k) \geq -\sigma$ for all $k \in K$ s.t. $k \geq \bar{k}$;*
2. *if $\sigma_k \to 0$, $\lim_{k \in K, k \to \infty} \Theta(X^k) = 0$.*

Proof We begin the proof showing that for all $k = 0, 1, \dots$, and for all $x \in X^k$, we have $x \in \mathcal{L}(x_0)$. Let k be a generic iteration and $x \in X^k$. Since Assumption 2 holds and X^0 is a set of nondominated feasible points, by Corollary 8 the set X^k also consists of mutually nondominated feasible solutions. By steps 2-3 of Procedure 1, we have that $x_0 \in C^0$; then, by point 2. of Lemma 11, there exists a point $y_k \in X^k$ such that $F(y_k) \leq F(x_0)$; since y_k does not dominate x, there exists an index h such that $f_h(x) \leq f_h(y_k) \leq f_h(x_0)$, i.e., $x \in \mathcal{L}(x_0)$. Hence, we have that
$$x \in \mathcal{L}(x_0), \quad \forall x \in X^k, \ \forall k. \tag{7}$$

Now, we show that a reference point $\zeta \in \mathbb{R}^m$ exists such that, for all $k = 0, 1, \ldots$, for all $x \in X^k$, $F(x) \le \zeta$. Let $\bar{\zeta} \in \mathbb{R}^m$ such that, for all $j \in \{1, \ldots, m\}$, $\bar{\zeta}_j = \max_{\bar{x} \in \mathcal{L}(x_0)} f_j(\bar{x})$; we observe that this vector is well defined, since F is continuous and $\mathcal{L}(x_0)$ is compact. Let k be any iteration and $x \in X^k$. By Eq. (7), we have $x \in \mathcal{L}(x_0)$. Therefore, for all $j \in \{1, \ldots, m\}$, we have $f_j(x) \le \max_{\bar{x} \in \mathcal{L}(x_0)} f_j(\bar{x}) = \bar{\zeta}_j$. Thus, we can conclude that $F(x) \le \bar{\zeta}$, for all $x \in X^k$ and for all k. Moreover, by definition of the set C^k, we also have that

$$F(x) \le \bar{\zeta}, \quad \text{for all } x \in C^k, \text{ for all } k. \tag{8}$$

Let us consider the subsequence $\{X^{\hat{l}(k)}\}$, with $\hat{l}(k) = l(k) - 1$, and, by contradiction, let us assume that the thesis is false in the two cases $\sigma_k = \sigma > 0$ and $\sigma_k \to 0$, respectively: there exists an infinite subsequence $K_1 \subseteq \{0, 1, \ldots\}$ such that

1. $\forall k \in K_1$ sufficiently large, we have $\Theta(X^{\hat{l}(k)}) < -\sigma$;
2. there exists $\varepsilon > 0$ such that, $\forall k \in K_1$ sufficiently large, we have $\Theta(X^{\hat{l}(k)}) < -\varepsilon$.

We can analyze both cases simultaneously, by assuming there exists $\bar{\sigma} > 0$ such that, for all $k \in K_1$ sufficiently large, $\Theta(X^{\hat{l}(k)}) < -\bar{\sigma}$. Since in case (i) we have $\bar{\sigma} = \sigma_k$, and in case (ii) $\sigma_k \to 0$, it follows that $\Theta(X^{\hat{l}(k)}) < -\sigma_{\hat{l}(k)}$ for all $k \in K_1$ sufficiently large.

Now, let $\{x^{\hat{l}(k)}\}$ be the sequence of points produced by `FPD_NMT` such that, for all k, $x^{\hat{l}(k)} \in X^{\hat{l}(k)}$ is the first point to be processed in the for loop of steps 4-20 of Algorithm 1 at iteration $\hat{l}(k)$. By Assumption 1, $x^{\hat{l}(k)} \in \arg\min_{x \in X^{\hat{l}(k)}} \theta(x)$ for all k, i.e., $\{x^{\hat{l}(k)}\}$ is a sequence of "least Pareto-stationary" points in $\{X^{\hat{l}(k)}\}$, with $\theta(x^{\hat{l}(k)}) = \Theta(X^{\hat{l}(k)})$ for all k. By Eq. (7), we trivially get that $\{x^{\hat{l}(k)}\}_{K_1} \subseteq \{x^{\hat{l}(k)}\} \subseteq \mathcal{L}(x_0)$ and, thus, $\{x^{\hat{l}(k)}\}_{K_1}$ admits accumulation points.

Let $\bar{x}$ be an accumulation point of $\{x^{\hat{l}(k)}\}_{K_1}$, i.e., there exists $K_2 \subseteq K_1$ such that $x^{\hat{l}(k)} \to \bar{x}$ for $k \in K_2$, $k \to \infty$. We define $z^{\hat{l}(k)} = x^{\hat{l}(k)} + \alpha_{\hat{l}(k)} v(x^{\hat{l}(k)})$ as the point obtained at step 11 of Algorithm 1 while processing point $x^{\hat{l}(k)}$. Since F is continuously differentiable and $\{x^{\hat{l}(k)}\}_{K_2}$ is convergent, by Eq. (5) we have that $\{v(x^{\hat{l}(k)})\}_{K_2}$ is bounded. Moreover, $\alpha_{\hat{l}(k)} \in [0, \alpha_0]$ which is a compact set. Therefore, there exists $K_3 \subseteq K_2$ such that $v(x^{\hat{l}(k)}) \to \bar{v}$ and $\alpha_{\hat{l}(k)} \to \bar{\alpha} \in [0, \alpha_0]$ for $k \in K_3$, $k \to \infty$.

By definition of K_1 and the sequences $\{\sigma_{\hat{l}(k)}\} \subseteq \{\sigma_k\}$ and $\{x^{\hat{l}(k)}\}$, we have that, for all $k \in K_1$ sufficiently large,

$$\theta(x^{\hat{l}(k)}) = \Theta(X^{\hat{l}(k)}) < -\bar{\sigma} \le -\sigma_{\hat{l}(k)} \le 0.$$

Also taking into account Assumption 1 and the instructions of the algorithm, it follows that $\alpha_{\hat{l}(k)}$ must necessarily be obtained at step 2 of Procedure 2: there exists $c^{\hat{l}(k)} \in C^{\hat{l}(k)}$ such that

$$F(z^{\hat{l}(k)}) \le F(c^{\hat{l}(k)}) + \mathbf{1}\gamma\alpha_{\hat{l}(k)} D(x^{\hat{l}(k)}, v(x^{\hat{l}(k)})).$$

Since $D(x^{\hat{l}(k)}, v(x^{\hat{l}(k)})) \leq \theta(x^{\hat{l}(k)})$, we reformulate the last result as

$$F(z^{\hat{l}(k)}) \leq F(c^{\hat{l}(k)}) + \mathbf{1}\gamma\alpha_{\hat{l}(k)}\theta(x^{\hat{l}(k)}).$$

By the continuity of θ, we also have the following result: $\theta(\bar{x}) = \lim_{k\in K_3, k\to\infty} \theta(x^{\hat{l}(k)}) \leq -\bar{\sigma} < 0$, i.e., $\bar{x}$ is not Pareto stationary. Incorporating the two last inequalities, we obtain, for $k \in K_3$ sufficiently large,

$$F(z^{\hat{l}(k)}) \leq F(c^{\hat{l}(k)}) - \mathbf{1}\gamma\alpha_{\hat{l}(k)}\bar{\sigma}. \tag{9}$$

By point 1. of Lemma 11, the sequence $\{V_F(C^k)\}$ is monotone nondecreasing and, in fact, it admits limit $\bar{V}$. Similarly to $\bar{\zeta}$, we can define a vector $\bar{\pi} \in \mathbb{R}^m$ such that, for all $j \in \{1, \dots, m\}$, $\bar{\pi}_j = \min_{\bar{x}\in\mathcal{L}(x_0)} f_j(\bar{x})$. It then follows that

$$\Lambda_F(C^k) \subseteq \{y \in \mathbb{R}^m \mid \bar{\pi} \leq y \leq \bar{\zeta}\},$$

where the latter set is compact and therefore has a finite measure M. Consequently, $V_F(C^k) \leq M < \infty$, and the sequence $\{V_F(C^k)\}$ must converge to a finite limit $\bar{V}$.

Now, by Lemma 12 and definition of $\hat{l}(k)$, for all k a point $y^{\hat{l}(k)} \in X^{l(k)}$ exists such that $F(y^{\hat{l}(k)}) \leq F(z^{\hat{l}(k)})$. We then have

$$(X^{l(k)} \cup C^{\hat{l}(k)}) \supseteq \{y^{\hat{l}(k)}\} \cup C^{\hat{l}(k)} \supseteq C^{\hat{l}(k)},$$

which, by point 2. of Lemma 11 and the properties of the dominated region, implies that

$$\Lambda_F(X^{l(k)}) = \Lambda_F(X^{l(k)} \cup C^{\hat{l}(k)}) \supseteq \Lambda_F(\{y^{\hat{l}(k)}\} \cup C^{\hat{l}(k)}) \supseteq \Lambda_F(C^{\hat{l}(k)}),$$

and thus

$$V_F(X^{l(k)}) \geq V_F(\{y^{\hat{l}(k)}\} \cup C^{\hat{l}(k)}) \geq V_F(C^{\hat{l}(k)}). \tag{10}$$

We shall also note that $F(y^{\hat{l}(k)}) \leq F(z^{\hat{l}(k)}) < F(c^{\hat{l}(k)})$, i.e., $c^{\hat{l}(k)}$ is dominated by $y^{\hat{l}(k)}$, and then

$$V_F(\{y^{\hat{l}(k)}\} \cup C^{\hat{l}(k)}) = V_F(\{y^{\hat{l}(k)}\} \cup C^{\hat{l}(k)} \setminus \{c^{\hat{l}(k)}\}). \tag{11}$$

Moreover, by Corollary 8 and definition of reference set, for all k the set C^k contains mutually nondominated points, i.e., $F(C^k)$ is a stable set. Also recalling Eq. (8), it follows that all the assumptions of Lemma 4 are satisfied. In particular, we obtain

$$V_F(\{y^{\hat{l}(k)}\} \cup C^{\hat{l}(k)} \setminus \{c^{\hat{l}(k)}\}) - V_F(C^{\hat{l}(k)}) \geq \prod_{j=1}^{m}(f_j(c^{\hat{l}(k)}) - f_j(y^{\hat{l}(k)})) \geq \prod_{j=1}^{m}(f_j(c^{\hat{l}(k)}) - f_j(z^{\hat{l}(k)})).$$

Recalling point 1. of Lemma 11, and putting together the last result, (9)-(10)-(11), we finally get that

$$
\begin{aligned}
V_F(C^k) - V_F(C^{\hat{l}(k)}) &\geq V_F(X^{l(k)}) - V_F(C^{\hat{l}(k)})\\
&\geq V_F(\{y^{\hat{l}(k)}\} \cup C^{\hat{l}(k)} \setminus \{c^{\hat{l}(k)}\}) - V_F(C^{\hat{l}(k)}) \geq (\gamma\alpha_{\hat{l}(k)}\bar{\sigma})^m.
\end{aligned}
$$

Also recalling that $V_F(C^k) \to \bar{V}$, we take the limits for $k \in K_3$, $k \to \infty$ to obtain $\lim_{k\in K_3, k\to\infty}(\gamma\alpha_{\hat{l}(k)}\bar{\sigma})^m \leq 0$. Since $\gamma > 0$ and $\bar{\sigma} > 0$, we necessarily have that $\alpha_{\hat{l}(k)} \to 0$ for $k \in K_3$, $k \to \infty$.

Since $\alpha_{\hat{l}(k)}$ is defined at step 2 of Procedure 2 and $\alpha_{\hat{l}(k)} \to 0$ for $k \in K_3$, $k \to \infty$, given any $q \in \mathbb{N}$ we must have $\alpha_{\hat{l}(k)} < \alpha_0\delta^q$ for all $k \in K_3$ sufficiently large. Given that the definition of projected common descent direction (3), Lemma 10, the convexity of Ω and the definition of α_0 in Algorithm 1 imply that, for all $\alpha \in [0, \alpha_0]$, $x^{\hat{l}(k)} + \alpha v(x^{\hat{l}(k)}) \in \Omega$, we then deduce that the step $\alpha_0\delta^q$ does not satisfy the Armijo condition: there exists $j_{\hat{l}(k)}$ such that

$$
\begin{aligned}
f_{j_{\hat{l}(k)}}(x^{\hat{l}(k)} + \alpha_0\delta^q v(x^{\hat{l}(k)})) &> f_{j_{\hat{l}(k)}}(c^{\hat{l}(k)}) + \gamma\alpha_0\delta^q D(x^{\hat{l}(k)}, v(x^{\hat{l}(k)}))\\
&\geq f_{j_{\hat{l}(k)}}(x^{\hat{l}(k)}) + \gamma\alpha_0\delta^q D(x^{\hat{l}(k)}, v(x^{\hat{l}(k)})),
\end{aligned}
$$

where the second inequality follows from the definition of $c^{\hat{l}(k)}$ at step 2 of Procedure 2. Taking the limits along a further subsequence such that $j_{\hat{l}(k)} = \bar{j}$, we obtain that

$$
f_{\bar{j}}(\bar{x} + \alpha_0\delta^q\bar{v}) \geq f_{\bar{j}}(\bar{x}) + \gamma\alpha_0\delta^q D(\bar{x}, \bar{v}).
$$

Since q is arbitrary, by [11, Lemma 4] it must be that $D(\bar{x}, \bar{v}) \geq 0$. However, we know that $D(x^{\hat{l}(k)}, v(x^{\hat{l}(k)})) \leq \theta(x^{\hat{l}(k)})$ which, in the limit, turns into $D(\bar{x}, \bar{v}) \leq \theta(\bar{x}) \leq -\bar{\sigma} < 0$. We finally get a contradiction; hence, the proof is complete. □

Corollary 14 *Let $\{X^k\}$ be the sequence of sets of nondominated points produced by* `FPD_NMT` *under the assumptions of Theorem 13 and with $\sigma_k \to 0$. Moreover, let $\{x^k\}$ be any sequence such that $x^k \in X^k$ for all k, and $K \subseteq \{0, 1, \ldots\}$ be the subsequence defined in Theorem 13. Then $\{x^k\}_K$ admits accumulation points and every accumulation point is Pareto stationary.*

Proof By Eq. (7), we have that $\{x^k\} \subseteq \mathcal{L}(x_0)$, and so does $\{x^k\}_K$ which thus admits accumulation points. By the definition of Θ, we have that, for all k, $0 \geq \theta(x^k) \geq \Theta(X^k)$. Thus, taking the limit along any subsequence $K_1 \subseteq K$ such that $x^k \to \bar{x}$ for $k \in K_1$, $k \to \infty$, and recalling Theorem 13, we have $\theta(\bar{x}) = 0$, which concludes the proof. □

Remark 3 It is worth noting that our results are established for a single subsequence. Deriving a stronger result, similar to Proposition 5, would require a forcing function on the distance between consecutive iterates $X^{l(k)}$ and $X^{l(k)-1}$, capturing improvement relative to the "least Pareto-stationary" points. However, such an assumption is rarely satisfied in practice. Future work could further investigate this direction.

Remark 4 Unlike classical non-monotone strategies, the notion of "non-monotonicity" adopted here is then not associated with a single quantity (e.g., the worst value(s) of the objective function(s) attained over the previous M iterations, as in the single-objective case [23] and the single-point multi-objective setting [25], or the Pareto front approximation with the lowest hyper-volume value among those generated during the previous M iterations). As shown by the definition of the reference set (Definition 3), Procedure 1 and the example reported in Fig. 1, "non-monotonicity" is instead based on a reference set constructed by combining multiple sources of information: (i) the Pareto front approximation with the lowest hyper-volume value over the previous M iterations; (ii) the previous reference set; and (iii) points from the current Pareto front approximation that, if not included in the reference set, would neither dominate nor coincide with any point already contained in it (see Procedure 1). Omitting any of these sources of information would irremediably compromise the stated convergence properties. For instance, relying only on the Pareto front approximation with the lowest hyper-volume value may result in points in the current front approximation that do not dominate nor coincide with any point in the reference set, thereby violating the validity of Theorem 13, which relies on the fact that at least for the first processed point of X^k such a point in the reference set exists.

The use of the reference set within the non-monotone line search at step 2 of Procedure 2, as well as within the line search associated with the exploration phase at step 15 of Algorithm 1, thus leads to the identification of points that satisfy a decrease condition based on solutions contained in the reference set. As a consequence, such points may exhibit an increase in a subset of the objective functions, or even in all of them, with respect to solutions in the current Pareto front approximation.

4 Numerical results

We present computational experiments assessing the effectiveness and consistency of our proposal. The Python3 code was executed on Ubuntu 22.04 with an Intel Xeon E5-2430 v2 (6 cores, 2.50 GHz) and 32 GB RAM, using Gurobi 12 [30] to solve problems (3)–(4).

4.1 Experimental settings

We compared the nonmonotone `FPD_NMT` to the original `FPD`, with shared parameters $\alpha_0 = 1, \delta = 0.5, \gamma = 10^{-4}$, and a crowding distance [22] heuristic to limit the generation of closely spaced points.

The benchmark is composed by box-constrained problems: CEC09 suite [29], JOS_1 ($\Omega = [0, 100]^n$) [31], MAN [32], ZDT_1 and ZDT_3 [33], mostly with two objectives ($m = 2$), except CEC09_8, CEC09_9, CEC09_10 with $m = 3$. Problem dimensions were $n \in \{5, 6, 8, 10, 12, 15, 17, 20, 25, 30, 35, 40, 45, 50, 100, 200, 300, 500, 1000, 5000\}$. For CEC09 problems, n starting points were

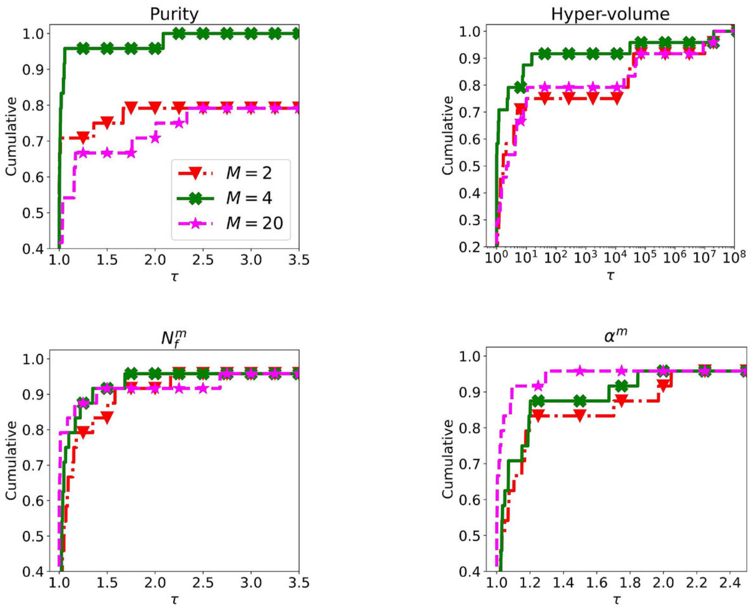

Fig. 2 Performance profiles w.r.t. purity, hyper-volume, N_f^m and α^m obtained by `FPD_NMT` with $M \in \{2, 4, 20\}$ on a subset of problems

uniformly sampled along the hyper-box diagonal; for others, single challenging initial points were considered: $(50, \ldots, 50)^\top$ for JOS_1, $(-10, \ldots, -10)^\top$ for MAN, and $(0.5, \ldots, 0.5)^\top$ for ZDT.

Performance and robustness were assessed via *performance profiles* [34], using standard metrics – *purity* [35] and *hyper-volume* [28] – and two ad-hoc metrics: N_f^m (function evaluations per processed point) and α^m (average Armijo-type line-search stepsize). The reference Pareto set for purity combined the two algorithms' solutions, discarding dominated points; the hyper-volume reference point was computed as in Sect. 3, with X_{all} being the union of both solution sets. Purity, hyper-volume and α^m have increasing values for better solutions: thus, for consistency with performance profiles, purity and α^m values were inverted, while, similarly to [19], hyper-volume ones were transformed to $V_{\mathcal{R}} - V_{\text{solver}} + 10^{-7}$, where $V_{\mathcal{R}}$ is the hypervolume of the reference Pareto set on the considered problem instance.

4.2 Selection of the parameter *M*

The parameter M is undoubtedly the most critical component of `FPD_NMT`; an inappropriate value may indeed significantly compromise the overall performance. Consequently, we conducted preliminary tests on a subset of the problem instances to determine the optimal value for M.

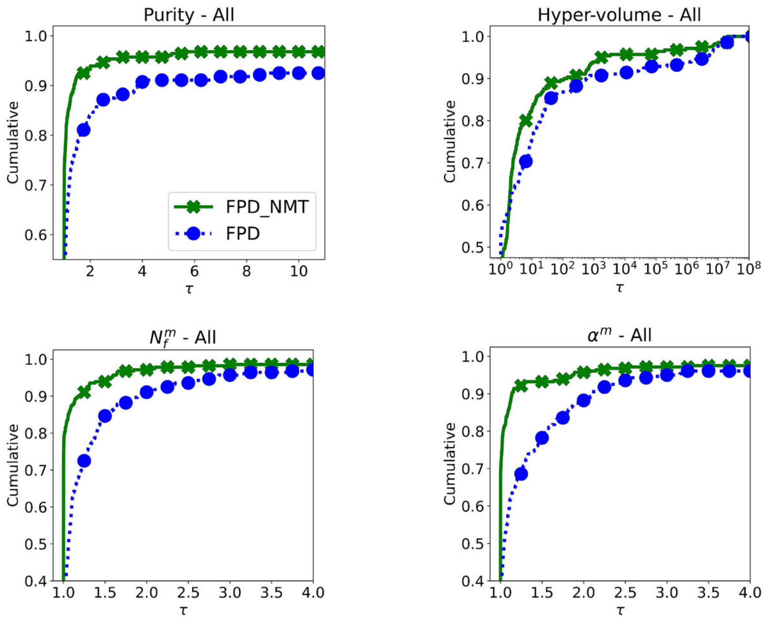

Fig. 3 Performance profiles for FPD and FPD_NMT w.r.t. purity, hyper-volume, N_f^m and α^m across the full problem benchmark

In Fig. 2, we report the performance profiles w.r.t. purity, hyper-volume, N_f^m and α^m obtained by FPD_NMT with $M \in \{2, 4, 20\}$ on CEC09_1, CEC09_3 [29], JOS_1 [31], MAN [32], ZDT_1 and ZDT_3 [33] with $n \in \{5, 50, 500, 5000\}$.

FPD_NMT with $M = 4$ clearly outperformed the other variants in terms of purity and hyper-volume. For N_f^m, the three methods showed similar performance, while the performance profiles for α^m indicate a clear advantage for the variant with $M = 20$. This aligns with the expectations: increasing the memory parameter M makes the sufficient decrease condition in the nonmonotone line search at step 2 of Procedure 2 easier to satisfy with fewer iterations, leading to larger step sizes. A similar trend was observed in the line search at step 15 of Algorithm 1. However, this behavior does not translate into superior performance for purity and hyper-volume.

Notably, FPD_NMT with $M = 4$ remained the second most robust approach with respect to α^m. Its overall performance motivated our choice of this variant for the overall comparisons presented below.

4.3 Overall comparisons

In Fig. 3 performance profiles for FPD and FPD_NMT are shown w.r.t. purity, hyper-volume, N_f^m, and α^m across the full problem benchmark, while in Figs. 4 and 5 performance profiles are calculated based on the low-dimensional problems ($n \leq 30$) and high-dimensional ones ($n > 30$), respectively.

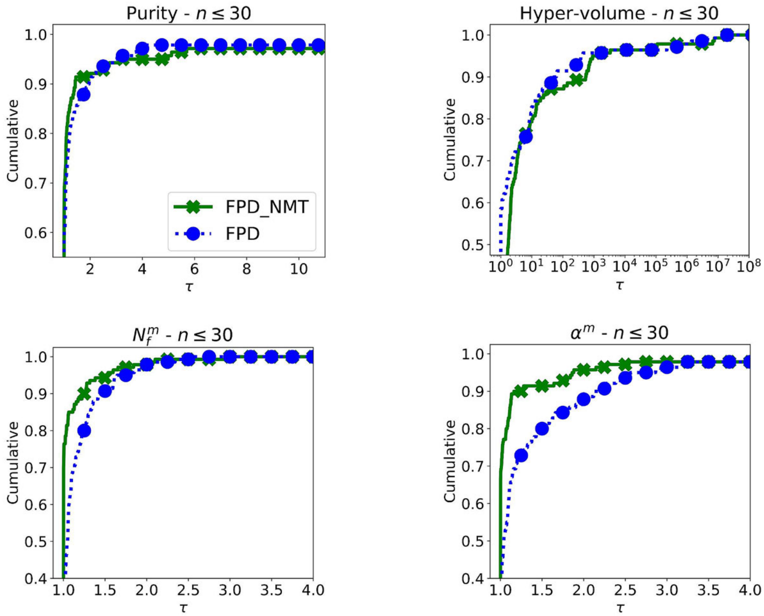

Fig. 4 Performance profiles for FPD and FPD_NMT w.r.t. purity, hyper-volume, N_f^m and α^m on low-dimensional problems ($n \leq 30$)

Considering the full problem benchmark (Fig. 3), FPD_NMT proved to be more effective than FPD in terms of purity; a similar trend has been observed for the hyper-volume metric, albeit with smaller performance differences. For N_f^m and α^m, FPD_NMT clearly outperformed the original approach; improvements on one metric reasonably reflect those on the other, as the nonmonotone algorithm requires fewer line search iterations and thus fewer function evaluations. These savings, however, do not compromise our method effectiveness, as reflected in the performance profiles for purity and hyper-volume.

While the two algorithms perform similarly overall on low-dimensional problems (Fig. 4), their differences become markedly pronounced in high-dimensional settings (Fig. 5). In this scenario, where the increase in the number of variables potentially leads to higher irregularity in the objective functions landscape, FPD_NMT clearly stands out as the superior method, appearing to better escape poor local minima / stationary points compared to the monotone paradigm.

5 Conclusions

In this paper, we introduced a nonmonotone extension of the state-of-the-art Front Descent framework for box-constrained smooth multi-objective optimization problems, providing convergence guarantees consistent with both the original method [19] and classical nonmonotone theory. To our knowledge, this work represents the first

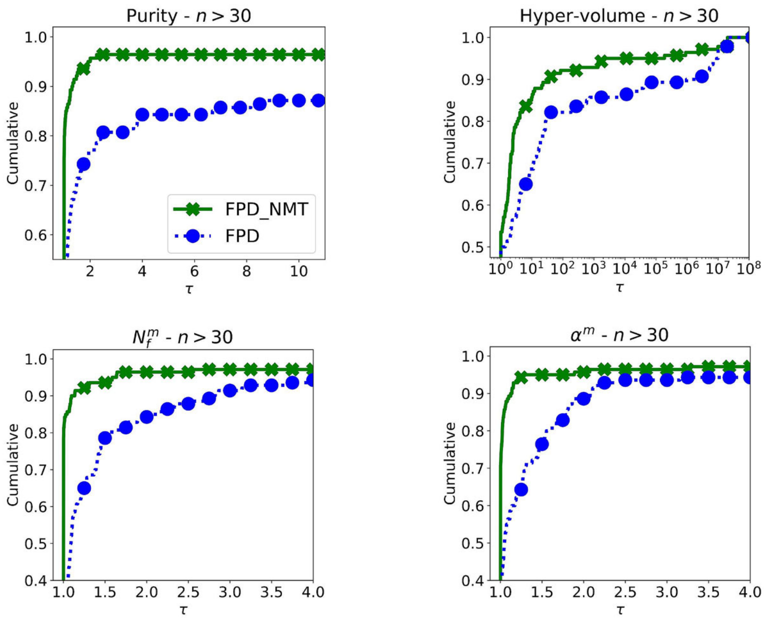

Fig. 5 Performance profiles for `FPD` and `FPD_NMT` w.r.t. purity, hyper-volume, N_f^m and α^m on high-dimensional problems ($n > 30$)

attempt to integrate nonmonotone line searches into a front-based descent approach designed to approximate the Pareto front through sequences of point sets. While the resulting convergence properties align with those of the original framework, the transition to a nonmonotone setting required a dedicated theoretical treatment to account for the absence of monotonic descent. Numerical experiments validate the effectiveness and robustness of the proposed approach in the bound-constrained setting.

Future work may include further theoretical investigations to establish stronger convergence properties, similar to those available for nonmonotone single-point multi-objective methods [25], as well as a complexity analysis analogous to that of [19] for the monotone case, whose extension is far from immediate. Additionally, the framework could be extended to handle more general constraints beyond the box-constrained case.

Acknowledgements The author is grateful to the two anonymous reviewers for their valuable comments, which significantly improved the quality of the manuscript.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. No funding was received for conducting this study.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Code Availability The source code of the `FPD_NMT` algorithm can be found at github.com/pierlumanzu/fpd_nmt.

Declarations

Conflict of interest The author has no conflict of interest to declare that are relevant to the content of this article.

Appendix A Algorithmic scheme of the nonmonotone front projected descent algorithm

In Algorithm A.1, we present the complete scheme of the `FPD_NMT` algorithm, which essentially corresponds to Algorithm 1, with Procedures 1-2-3 respectively replacing steps 3-7-21.

Appendix B Proof of proposition 5

In this appendix, we present the proof of Proposition 5, which establishes the main convergence result of the `FPD` algorithm (Algorithm 1). The proof is adapted from [19, Theorem 5.15] to the box-constrained setting, with most of the argument remaining identical to the original one and only minor modifications arising from the use of the projected common descent direction (3).

Proof We begin the proof showing that for all $k = 0, 1, \dots,$ and for all $\xi \in X^k$, we have that $\xi \in \mathcal{L}(x_0)$. Let k be a generic iteration and consider any $\xi \in X^k$. First, since X^0 is a set of nondominated feasible points and recalling point 3. of Lemma 3, we know that X^k is also a set of mutually nondominated feasible solutions. We now have one of two cases.

- $x_0 \in X^k$; by the nondomination property of X^k, there exists at least one index j such that $f_j(\xi) \leq f_j(x_0)$; thus, $\xi \in \mathcal{L}(x_0)$.
- $x_0 \notin X^k$; by [19, Lemma 5.3], we know that there exists a point $y_k \in X^k$ such that $F(y_k) \leq F(x_0)$; since y_k does not dominate ξ, there exists $h \in \{1, \dots, m\}$ such that $f_h(\xi) \leq f_h(y_k) \leq f_h(x_0)$, i.e., $\xi \in \mathcal{L}(x_0)$ also in this case.

Then, we show that a reference point $\zeta \in \mathbb{R}^m$ exists such that $F(\xi) \leq \zeta$ for all $\xi \in X^k$, for all $k \in \{0, 1, \dots\}$. To this aim, let $\bar{\zeta} \in \mathbb{R}^m$ such that $\bar{\zeta}_j = \max_{x \in \mathcal{L}(x_0)} f_j(x)$, which is well defined being F continuous and $\mathcal{L}(x_0)$ compact. Let k be any iteration and $\xi \in X^k$; we know that $\xi \in \mathcal{L}(x_0)$. Therefore, for all $h \in \{1, \dots, m\}$, we have $f_h(\xi) \leq \max_{x \in \mathcal{L}(x_0)} f_h(x) = \bar{\zeta}_h$, and thus

$$F(\xi) \leq \bar{\zeta} \quad \text{for all } \xi \in X^k, \text{ for all } k. \tag{B.1}$$

Algorithm A.1 Nonmonotone front projected gradient

Require: $F:\mathbb{R}^n \to \mathbb{R}^m$, $X^0 \subset \Omega = [l,u]$, $\alpha_0 \in (0,1]$, $\delta, \gamma \in (0,1)$, $\{\sigma_k\} \subseteq \mathbb{R}_+$, $M \in \mathbb{N}$.

$k = 0$

$C^{-1} = \emptyset$

while a stopping criterion is not satisfied **do**

$\quad X^{l(k)} \in \underset{X \in \{X^k, X^{k-1}, \dots, X^{k-\min\{k,M\}}\}}{\arg\min} V_F(X)$

$\quad \bar{C}^k = \{x \in X^{l(k)} \cup C^{k-1} \mid \nexists y \in X^{l(k)} \cup C^{k-1} : F(y) \lneqq F(x)\}$

$\quad C^k = \bar{C}^k \cup \{x \in X^k \mid \forall y \in \bar{C}^k\ \exists j \in \{1,\dots,m\} : f_j(y) < f_j(x)\}$

$\quad \hat{X}^k = C^k$

$\quad$ **for all** $x_p \in X^k$ **do**

$\quad\quad$ **if** $\nexists y \in \hat{X}^k$ s.t. $F(y) \lneqq F(x_p)$ **then**

$\quad\quad\quad$ **if** $\theta(x_p) < -\sigma_k$ **then**

$\quad\quad\quad\quad$ **if** $\exists c \in \hat{X}^k : F(x_p) \le F(c)$ **then**

$$\alpha_p^k = \max_{h \in \mathbb{N}} \left\{ \alpha_0 \delta^h \;\middle|\; \begin{array}{l} x_p + \alpha_0 \delta^h v(x_p) \in \Omega \wedge \\ \left(\begin{array}{l} \exists c_p^k \in \hat{X}^k : F(x_p) \le F(c_p^k) \wedge \\ F(x_p + \alpha_0 \delta^h v(x_p)) \le F(c_p^k) + \mathbf{1}\gamma\alpha_0\delta^h D(x_p, v(x_p)) \end{array} \right) \end{array} \right\}$$

$\quad\quad\quad\quad$ **else**

$$\alpha_p^k = \max_{h \in \mathbb{N}} \left\{ \alpha_0 \delta^h \;\middle|\; \begin{array}{l} x_p + \alpha_0 \delta^h v(x_p) \in \Omega \wedge \\ F(x_p + \alpha_0 \delta^h v(x_p)) \le F(x_p) + \mathbf{1}\gamma\alpha_0\delta^h D(x_p, v(x_p)) \end{array} \right\}$$

$\quad\quad\quad\quad$ **end if**

$\quad\quad\quad$ **else**

$\quad\quad\quad\quad \alpha_p^k = 0$

$\quad\quad\quad$ **end if**

$\quad\quad\quad z_p^k = x_p + \alpha_p^k v(x_p)$

$\quad\quad\quad \hat{X}^k = \hat{X}^k \setminus \{y \in \hat{X}^k \mid F(z_p^k) \lneqq F(y)\} \cup \{z_p^k\}$

$\quad\quad\quad$ **for all** $I \subset \{1,\dots,m\}$ s.t. $\theta^I(z_p^k) < 0$ **do**

$\quad\quad\quad\quad$ **if** $z_p^k \in \hat{X}^k$ **then**

$$\alpha_p^I = \max_{h \in \mathbb{N}} \left\{ \alpha_0 \delta^h \;\middle|\; \begin{array}{l} z_p^k + \alpha_0 \delta^h v^I(z_p^k) \in \Omega \wedge \\ \forall y \in \hat{X}^k,\ \exists j : f_j(z_p^k + \alpha_0 \delta^h v^I(z_p^k)) < f_j(y) \end{array} \right\}$$

$\quad\quad\quad\quad\quad \hat{X}^k = \hat{X}^k \setminus \{y \in \hat{X}^k \mid F(z_p^k + \alpha_p^I v^I(z_p^k)) \lneqq F(y)\} \cup \{z_p^k + \alpha_p^I v^I(z_p^k)\}$

$\quad\quad\quad\quad$ **end if**

$\quad\quad\quad$ **end for**

$\quad\quad$ **end if**

$\quad$ **end for**

$\quad \bar{X}^{l(k)} \in \underset{X \in \{\hat{X}^k\} \cup \{X^k, X^{k-1}, \dots, X^{k-\min\{k,M-1\}}\}}{\arg\min} V_F(X)$

$\quad X^{k+1} = \{x \in \hat{X}^k \cup \bar{X}^{l(k)} \mid \nexists y \in \hat{X}^k \cup \bar{X}^{l(k)} : F(y) \lneqq F(x)\}$

$\quad k = k + 1$

end while

return X^k

Now let us assume by contradiction that the thesis is false, in the two cases $\sigma_k = \sigma > 0$ for all k and $\sigma_k \to 0$, respectively:

1. there exists an infinite subsequence $K \subseteq \{0, 1, \dots\}$ such that, for all $k \in K$ sufficiently large, we have $\Theta(X^k) < -\sigma$;
2. there exists an infinite subsequence $K \subseteq \{0, 1, \dots\}$ and a value $\epsilon > 0$ such that, for all $k \in K$ sufficiently large, we have $\Theta(X^k) < -\epsilon$.

We can thus analyze the two cases at once, assuming that there exists an infinite subsequence $K \subseteq \{0, 1, \dots\}$ and a value $\bar{\sigma} > 0$ such that, for all $k \in K$ and sufficiently large, we have $\Theta(X^k) < -\bar{\sigma}$. Since either $\bar{\sigma} = \sigma_k$ from case (i) or $\sigma_k \to 0$ from case (ii), we also have that $\Theta(X^k) < -\sigma_k$ for all $k \in K$ sufficiently large.

Now, let $\{x^k\}$ be the sequence of points such that, for all k, x^k is the first point to be processed in the for loop of steps 4-20. By Assumption 1, $x^k \in \arg\min_{x \in X^k} \theta(x)$, i.e., $\{x^k\}$ is a sequence of "least Pareto-stationary points" in $\{X^k\}$, with $\theta(x^k) = \Theta(X^k)$. We know that $\{x^k\} \subseteq \mathcal{L}(x^0)$, and so does $\{x^k\}_K$ which thus admits accumulation points.

Let $\bar{x}$ be such an accumulation point, i.e., there exists $K_1 \subseteq K$ such that $x^k \to \bar{x}$ for $k \in K_1, k \to \infty$. Let us define $z^k = x^k + \alpha_k v(x^k)$ as the point obtained at step 11 of the algorithm while processing point x^k. Since F is continuously differentiable and $\{x^k\}_{K_1}$ is convergent, by Equation (5) we have that $\{v(x^k)\}_{K_1}$ is bounded. Moreover, $\alpha_k \in [0, \alpha_0]$, which is a compact set. Therefore, there exists a further subsequence $K_2 \subseteq K_1$ such that, for $k \in K_2, k \to \infty$,

$$\lim_{\substack{k\to\infty \\ k\in K_2}} \alpha_k = \bar{\alpha} \in [0, \alpha_0], \qquad \lim_{\substack{k\to\infty \\ k\in K_2}} v(x^k) = \bar{v}, \qquad \lim_{\substack{k\to\infty \\ k\in K_2}} z^k = \bar{x} + \bar{\alpha}\bar{v} = \bar{z}.$$

By the definition of $\{x^k\}$, K and $\{\sigma_k\}$, we have for all $k \in K$ sufficiently large that $\theta(x^k) = \Theta(X^k) < -\bar{\sigma} \le -\sigma_k \le 0$. Then, α_k has certainly be obtained, at step 7, by the line search along the direction $v(x^k)$, and we have

$$F(z^k) \le F(x^k) + \mathbf{1}\gamma\alpha_k D(x^k, v(x^k)).$$

Since $D(x^k, v(x^k)) \le \theta(x^k)$, we reformulate the last result as

$$F(z^k) \le F(x^k) + \mathbf{1}\gamma\alpha_k\theta(x^k).$$

By the continuity of θ, we also have the following result: $\theta(\bar{x}) = \lim_{k\in K_2, k\to\infty} \theta(x^k) \le -\bar{\sigma} < 0$, i.e., $\bar{x}$ is not Pareto stationary. Incorporating the two last inequalities, we obtain, for $k \in K_2$ sufficiently large,

$$F(z^k) \le F(x^k) - \mathbf{1}\gamma\alpha_k\bar{\sigma}. \tag{B.2}$$

We now consider the behavior of the hypervolume metric along the sequence of sets $\{X^k\}$. By the instructions of the algorithm, a point is removed from the current set of solutions only when a new point is added that dominates it. In other words, if $\xi \in X^k$ and $\xi \notin X^{k+1}$, there exists $\xi_1 \in X^{k+1}$ such that $F(\xi_1) \lneqq F(\xi)$. Therefore,

$(F(X^{k+1} \cup X^k))_{\mathrm{nd}} = (F(X^{k+1}))_{\mathrm{nd}} = F(X^{k+1})$, where the last equality follows from the nondominance property of $\{X^k\}$ (Lemma 3); moreover, it is straightforward to observe that, since $X^k \subseteq (X^{k+1} \cup X^k)$, we have $\Lambda_F(X^k) \subseteq \Lambda_F(X^{k+1} \cup X^k)$. Thus, by the properties of the dominated region, we have:

$$\Lambda_F(X^{k+1}) = \Lambda((F(X^{k+1} \cup X^k))_{\mathrm{nd}}) = \Lambda_F(X^{k+1} \cup X^k) \supseteq \Lambda_F(X^k).$$

Then, we can also write $V_F(X^{k+1}) \geq V_F(X^k)$, i.e., the sequence $\{V_F(X^k)\}$ is monotone nondecreasing and thus it either diverges to $+\infty$ or admits limit $\bar{V}$. We can note that this is, in fact, the second case, with $\bar{V}$ being finite. Indeed, similarly as $\bar{\zeta}$, let $\pi \in \mathbb{R}^m$ be such that $\pi_j = \min_{x \in \mathcal{L}(x_0)} f_j(x)$. We then observe that $\Lambda_F(X^k) \subseteq \{y \in \mathbb{R}^m \mid \pi \leq y \leq \bar{\zeta}\}$, the latter set being a compact set and thus having a finite measure M. Therefore, $V_F(X^k) \leq M < \infty$ and the sequence $\{V_F(X^k)\}$ must be then convergent to a finite limit $\bar{V}$.

Now, by [19, Lemma 5.3], a point y^k exists, for all k, such that $y^k \in X^{k+1}$ with $F(y^k) \leq F(z^k)$. We have $(X^{k+1} \cup X^k) \supseteq X^k \cup \{y^k\} \supseteq X^k$, so that, by the properties of the dominated region, we can also write:

$$\Lambda_F(X^{k+1}) = \Lambda_F(X^{k+1} \cup X^k) \supseteq \Lambda_F(\{y^k\} \cup X^k) \supseteq \Lambda_F(X^k).$$

and thus

$$V_F(X^{k+1}) \geq V_F(\{y^k\} \cup X^k) \geq V_F(X^k). \tag{B.3}$$

We shall now note that $F(y^k) \leq F(z^k) < F(x^k)$; thus, x^k is dominated by y^k and

$$V_F(\{y^k\} \cup X^k) = V_F(\{y^k\} \cup X^k \setminus \{x^k\}). \tag{B.4}$$

We shall also note that, by Lemma 3, for all k the set X^k contains mutually nondominated points, i.e., $F(X^k)$ is a stable set. Recalling (B.1), we have that all the assumptions of Lemma 4 are satisfied when we evaluate $V_F(\{y^k\} \cup X^k \setminus \{x^k\})$. In particular, we get

$$V_F(\{y^k\} \cup X^k \setminus \{x^k\}) - V_F(X^k) \geq \prod_{j=1}^{m}(f_j(x^k) - f_j(y^k)) \geq \prod_{j=1}^{m}(f_j(x^k) - f_j(z^k)). \tag{B.5}$$

Putting together (B.2), (B.3), (B.4) and (B.5), we finally obtain that

$$V_F(X^{k+1}) - V_F(X^k) \geq V_F(\{y^k\} \cup X^k \backslash \{x^k\}) - V_F(X^k) \geq (\gamma \alpha_k \bar{\sigma})^m .$$

Recalling that $V_F(X^k) \to \bar{V}$, we can take the limits for $k \in K_2$, $k \to \infty$ to obtain

$$\lim_{\substack{k \to \infty \\ k \in K_2}} (\gamma \alpha_k \bar{\sigma})^m \leq 0.$$

Since $\gamma > 0$ and $\bar{\sigma} > 0$, we necessarily have that $\alpha_k \to 0$ for $k \in K_2$, $k \to \infty$.

Since α_k is defined at step 7 and, for $k \in K_2$, $\alpha_k \to 0$, given any $q \in \mathbb{N}$, for all $k \in K_2$ large enough we necessarily have $\alpha_k < \alpha_0 \delta^q$. Since the definition of projected common descent direction (3), Lemma 3, the convexity of Ω and the definition of α_0 in Algorithm 1 imply that, for all $\alpha \in [0, \alpha_0]$, $x^k + \alpha v(x^k) \in \Omega$, we then deduce that the step $\alpha = \alpha_0 \delta^q$ does not satisfy the Armijo condition $F\left(x^k + \alpha v(x^k)\right) \leq F(x^k) + \mathbf{1}\gamma\alpha D(x^k, v(x^k))$, i.e., for some $\tilde{h}(k)$ we have

$$f_{\tilde{h}(k)}\left(x^k + \alpha_0\delta^q v(x^k)\right) > f_{\tilde{h}(k)}(x^k) + \gamma\alpha_0\delta^q D(x^k, v(x^k)).$$

Taking the limits along a suitable subsequence such that $\tilde{h}(k) = \tilde{h}$, we get

$$f_{\tilde{h}}\left(\bar{x} + \alpha_0\delta^q \bar{v}\right) \geq f_{\tilde{h}}(\bar{x}) + \gamma\alpha_0\delta^q D(\bar{x}, \bar{v}).$$

Being q arbitrary and recalling [11, Lemma 4], it must be $D(\bar{x}, \bar{v}) \geq 0$. However, we know that $D(x^k, v(x^k)) \leq \theta(x^k)$ which, in the limit, turns into $D(\bar{x}, \bar{v}) \leq \theta(\bar{x}) \leq -\bar{\sigma} < 0$. We finally get a contradiction; hence, the proof is complete. □

References

1. Eichfelder, G.: Twenty years of continuous multi-objective optimization in the twenty-first century. EURO J. Comput. Optim. **9**, 100014 (2021)
2. Campana, E.F., Diez, M., Liuzzi, G., Lucidi, S., Pellegrini, R., Piccialli, V., Rinaldi, F., Serani, A.: A multi-objective direct algorithm for ship hull optimization. Comput. Optim. Appl. **71**(1), 53–72 (2018)
3. Sun, Y., Ng, D.W.K., Zhu, J., Schober, R.: Multi-objective optimization for robust power efficient and secure full-duplex wireless communication systems. IEEE Trans. Wireless Commun. **15**(8), 5511–5526 (2016)
4. Carrizosa, E., Frenk, J.B.G.: Dominating sets for convex functions with some applications. J. Optim. Theory Appl. **96**(2), 281–295 (1998)
5. Palermo, G., Silvano, C., Valsecchi, S., Zaccaria, V.: A system-level methodology for fast multi-objective design space exploration. In: Proceedings of the 13th ACM Great Lakes Symposium on VLSI, pp. 92–95. ACM (2003)
6. Tavana, M.: A subjective assessment of alternative mission architectures for the human exploration of Mars at NASA using multicriteria decision making. Comput. Oper. Res. **31**(7), 1147–1164 (2004)
7. Chen, B., Lin, Y., Zeng, W., Xu, H., Zhang, D.: The mean-variance cardinality constrained portfolio optimization problem using a local search-based multi-objective evolutionary algorithm. Appl. Intell. **47**(2), 505–525 (2017)
8. Annunziata, A., Lapucci, M., Mansueto, P., Pucci, D.: On the computation of the efficient frontier in advanced sparse portfolio optimization. 4OR (2025)
9. De Ath, G., Everson, R.M., Rahat, A.A.M., Fieldsend, J.E.: Greed is good: exploration and exploitation trade-offs in Bayesian optimisation. ACM Trans. Evol. Learn. Optim. **1**(1) (2021)
10. Carciaghi, F., Magistri, S., Mansueto, P., Schoen, F.: A bi-objective optimization based acquisition strategy for batch Bayesian global optimization. Comput. Optim. Appl. **92**(1), 81–123 (2025)
11. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**(3), 479–494 (2000)
12. Drummond, L.M.G., Iusem, A.N.: A projected gradient method for vector optimization problems. Comput. Optim. Appl. **28**(1), 5–29 (2004)
13. Fliege, J., Drummond, L.M.G.N., Svaiter, B.F.: Newton's method for multi-objective optimization. SIAM J. Optim. **20**(2), 602–626 (2009)
14. Prudente, L.F., Souza, D.R.: Global convergence of a BFGS-type algorithm for nonconvex multi-objective optimization problems. Comput. Optim. Appl. **88**(3), 719–757 (2024)

15. Lapucci, M., Mansueto, P.: A limited memory quasi-newton approach for multi-objective optimization. Comput. Optim. Appl. **85**(1), 33–73 (2023)
16. Cocchi, G., Liuzzi, G., Lucidi, S., Sciandrone, M.: On the convergence of steepest descent methods for multi-objective optimization. Comput. Optim. Appl. **77**(1), 1–27 (2020)
17. Cocchi, G., Lapucci, M., Mansueto, P.: Pareto front approximation through a multi-objective augmented Lagrangian method. EURO J. Comput. Optim. **9**, 100008 (2021)
18. Lapucci, M., Mansueto, P.: Improved front steepest descent for multi-objective optimization. Oper. Res. Lett. **51**(3), 242–247 (2023)
19. Lapucci, M., Mansueto, P., Pucci, D.: Effective front-descent algorithms with convergence guarantees. SIAM J. Optim. **36**(2), 597–625 (2026)
20. Pascoletti, A., Serafini, P.: Scalarizing vector optimization problems. J. Optim. Theory Appl. **42**(4), 499–524 (1984)
21. Eichfelder, G.: An adaptive scalarization method in multi-objective optimization. SIAM J. Optim. **19**(4), 1694–1718 (2009)
22. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)
23. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for newton's method. SIAM J. Numer. Anal. **23**(4), 707–716 (1986)
24. Zhang, H., Hager, W.W.: A nonmonotone line search technique and its application to unconstrained optimization. SIAM J. Optim. **14**(4), 1043–1056 (2004)
25. Mita, K., Fukuda, E.H., Yamashita, N.: Nonmonotone line searches for unconstrained multi-objective optimization problems. J. Global Optim. **75**(1), 63–90 (2019)
26. Qu, S., Ji, Y., Jiang, J., Zhang, Q.: Nonmonotone gradient methods for vector optimization with a portfolio optimization application. Eur. J. Oper. Res. **263**(2), 356–366 (2017)
27. Fazzio, N.S., Schuverdt, M.L.: Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems. Optim. Lett. **13**(6), 1365–1379 (2019)
28. Zitzler, E., Thiele, L.: Multiobjective optimization using evolutionary algorithms—a comparative case study. In: Eiben, A.E., Bäck, T., Schoenauer, M., Schwefel, H.-P. (eds.) Parallel Problem Solving from Nature—PPSN V, pp. 292–301. Springer, Berlin, Heidelberg (1998)
29. Zhang, Q., Zhou, A., Zhao, S., Suganthan, P., Liu, W., Tiwari, S.: Multiobjective optimization test instances for the CEC 2009 special session and competition. Mech. Eng. (2008)
30. Gurobi Optimization, LLC: Gurobi Optimizer Reference Manual (2024)
31. Jin, Y., Olhofer, M., Sendhoff, B.: Dynamic weighted aggregation for evolutionary multi-objective optimization: why does it work and how? In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 1042–1049 (2001)
32. Lapucci, M., Mansueto, P., Schoen, F.: A memetic procedure for global multi-objective optimization. Math. Program. Comput. (2022)
33. Zitzler, E., Deb, K., Thiele, L.: Comparison of multiobjective evolutionary algorithms: empirical results. Evol. Comput. **8**(2), 173–195 (2000)
34. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201–213 (2002)
35. Custódio, A.L., Madeira, J.F.A., Vaz, A.I.F., Vicente, L.N.: Direct multisearch for multiobjective optimization. SIAM J. Optim. **21**(3), 1109–1140 (2011)