Universal nonmonotone line search method for nonconvex multiobjective optimization problems with convex constraints

Maria Eduarda Pinheiro * Geovani Nunes Grapiglia †

November 14, 2024

Abstract

In this work we propose a general nonmonotone line-search method for nonconvex multiobjective optimization problems with convex constraints. At the kth iteration, the degree of nonmonotonicity is controlled by a vector ν_k with nonnegative components. Different choices for ν_k lead to different nonmonotone step-size rules. Assuming that the sequence $\{\nu_k\}_{k\geq 0}$ is summable, and that the ith objective function has Hölder continuous gradient with smoothness parameter $\theta_i \in (0,1]$, we show that the proposed method takes no more than $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right)$ iterations to find a ϵ-approximate Pareto critical point for a problem with m objectives and $\theta_{\min} = \min_{i=1,\ldots,m}\{\theta_i\}$. In particular, this complexity bound applies to the methods proposed by Drummond and Iusem (Comput. Optim. Appl. 28: 5–29, 2004), by Fazzio and Schuverdt (Optim. Lett. 13: 1365–1379, 2019), and by Mita, Fukuda and Yamashita (J. Glob. Optim. 75: 63–90, 2019). The generality of our approach also allows the development of new methods for multiobjective optimization. As an example, we propose a new nonmonotone step-size rule inspired by the Metropolis criterion. Preliminary numerical results illustrate the benefit of nonmonotone line searches and suggest that our new rule is particularly suitable for multiobjective problems in which at least one of the objectives has many non-global local minimizers.

Keywords: Multiobjective optimization; Projected gradient methods; Nonmonotone line searches; Worst-case complexity

1 Introduction

Motivation

In this work, we consider multiobjective optimization problems with convex constraints. This type of problem appears in several important applications, such as seismic design of buildings [26], planetary exploration [7, 35], maintenance of civil infrastructures [17], planning of cancer treatment [38, 8] and drug design [34, 25]. In most cases, there is not a single point which minimizes all the objective functions at once. This fact motivates the notion of Pareto efficiency. Roughly speaking, a point is

*Trier University, Department of Mathematics, Universitätsring 15, Trier, 54296, Trier, Germany. E-mail: pinheiro@uni-trier.de

†Université catholique de Louvain, ICTEAM/INMA, Avenue Georges Lemaître, 4-6/ L4.05.01, B-1348, Louvain-la-Neuve, Belgium. E-mail: geovani.grapiglia@uclouvain.be

said to be a Pareto efficient solution when from this point it is impossible to obtain an improvement in any of the objective functions without worsening the value of some other objective.

There are many approaches to solve multiobjective optimization problems [28, 27]. In recent years, several methods have been obtained by extending well-known optimization algorithms for single-objective optimization. Notable examples include the steepest descent method proposed in [15], projected gradient methods [11, 18, 19, 1], the Newton's method proposed in [14], proximal-point methods [3, 2], trust-region methods [36, 5], and also nonmonotone line search methods [13, 29].

In the present work, we extend the general nonmonotone method proposed in [33] for single-objective optimization to multiobjective optimization. The method is said to be nonmonotone because it permits an increase in the objectives between consecutive iterations. At the kth iteration, the extent of the increase allowed is governed by a vector ν_k with nonnegative components. Various selections for ν_k result in different nonmonotone step-size rules. Specifically, considering suitable choices for ν_k, our method encompasses instances of the methods proposed in [11, 13, 29]. Assuming that the sequence $\{\nu_k\}_{k\geq 0}$ is summable, and that the ith objective function has Hölder continuous gradient with constant $\bar{H}_i$ and smoothness parameter $\theta_i \in (0, 1]$, we show that the proposed method takes no more than $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right)$ iterations to find a ϵ-approximate Pareto critical point of a problem with m objectives and $\theta_{\min} = \min_{i=1,\ldots,m}\{\theta_i\}$. For the case $m = 1$, this bound agrees in order with the bound proved in [37] for a gradient method that requires the knowledge of H_1 and θ_1. Since our method is fully adaptive with respect to the Hölder constants, we say that it is a *universal method* in the sense of Nesterov [31]. For the case in which $\theta_1 = \ldots = \theta_m = \theta$, i.e., all the objective functions have Hölder continuous gradients *with the same smoothness level* θ, our complexity bound agrees in order with the bound established by [4] for the first-order version of their pth-order method. In addition, when all the objectives have Lipschitz continuous gradients (case $\theta_1 = \ldots = \theta_m = 1$), our bound becomes of $\mathcal{O}\left(\epsilon^{-2}\right)$ which agrees in order with the bounds established in [22] for a multiobjective trust-region method, and in [16] for the multiobjective steepest descent method. Under the weaker assumption that $\{\nu_k\}_{k\geq 0}$ converges to zero, we also prove a liminf-type global convergence result for our method. The generality of our results regarding possible choices of the sequence $\{\nu_k\}_{k\geq 0}$ enables the development of new nonmonotone methods with convergence and worst-case complexity guarantees for multiobjective optimization. As an example, we propose a new Metropolis-based nonmonotone step-size rule inspired by a method recently proposed in [21] for single-objective optimization.

Contents

This paper is organized as follows. In Section 2, we define the problem and review some results about multiobjective optimization. In Section 3 we present our general nonmonotone method and establish its complexity and convergence properties. In Section 4, we analyse some particular instances of the method. In Section 5 we propose a new nonmonotone method that fits into our general scheme. Finally, in Section 6, we report some illustrative numerical results.

Notations

In what follows, $\|\,.\,\|$ denotes the Euclidean norm, $I = \{1, 2, \ldots, m\}$, $\mathbb{R}^m_+ = \{z \in \mathbb{R}^m \,|\, z_i \geq 0,\ i \in I\}$, $\mathbb{R}^m_{++} = \{z \in \mathbb{R}^m \,|\, z_i > 0,\ i \in I\}$ and, given $x \in \Omega$, we consider the set $\Omega - x = \{y \in \mathbb{R}^n \,|\, y + x \in \Omega\}$. In addition, $J_F(x)$ will denote the Jacobian of $F : \mathbb{R}^n \to \mathbb{R}^m$ at x. The relations $\succ$ and $\succ_\omega$ are given

respectively by

$$y \succ x \iff y - x \in \mathbb{R}^m_+ \text{ and } y \succ_\omega x \iff y - x \in \mathbb{R}^m_{++}.$$

Moreover, given a symmetric matrix $B \in \mathbb{R}^{n\times n}$, $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ will denote the smallest and the largest eigenvalues of B. Finally, given a finite set A, $|A|$ denotes the cardinality of A.

2 Problem Definition and Auxiliary Results

In this paper we consider methods for solving the following multiobjective optimization problem

$$\min \quad F(x) = (f_1(x), \ldots, f_m(x))^T, \tag{1}$$

$$\text{s.t.} \quad x \in \Omega, \tag{2}$$

where $\emptyset \neq \Omega \subset \mathbb{R}^n$ is a closed and convex set and $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function, possibly nonconvex. Let us start by recalling the definitions of efficient and (local) weakly efficient solutions of (1)-(2).

Definition 2.1 ([24], page 619)**.** *Given a point $x^* \in \Omega$,*

(a) x^* is said to be a Pareto efficient solution of (1)-(2) when there is no $y \in \Omega$ such that $F(x^*) \succ F(y)$ and $F(y) \neq F(x^*)$;

(b) x^* is said to be a weakly Pareto efficient solution of (1)-(2) when there is no $y \in \Omega$ such that $F(x^*) \succ_w F(y)$; and

(c) x^* is said to be a local (or local weakly) Pareto efficient solution of (1)-(2) when there exists a neighborhood $N(x^*)$ of x^* for which there is no $y \in N(x^*) \cap \Omega$ such that $F(x^*) \succ F(y)$ and $F(y) \neq F(x^*)$ (or, respectively, $F(x^*) \succ_w F(y)$).

The lemma below gives a necessary condition for a point $x^* \in \Omega$ to be a local weakly efficient solution of (1)-(2).

Lemma 2.2. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function and $\Omega \subset \mathbb{R}^n$ be a closed and convex set. If $x^* \in \Omega$ is a local weakly efficient solution of (1)-(2), then*

$$-\mathbb{R}^m_{++} \cap \{J_F(x^*)(x - x^*) \,|\, x \in \Omega\} = \emptyset. \tag{3}$$

Proof. See Theorem 5.1 (item (ii)-(a)) in [24]. □

Lemma 2.2 motivates the following definition.

Definition 2.3. *A point $x^* \in \Omega$ is said to be a Pareto critical point for (1)-(2) if it satisfies condition (3).*

Remark 2.4. *Note that when $m = 1$ and $\Omega = \mathbb{R}^n$, (1)-(2) reduces to the unconstrained scalar optimization problem, and the Pareto criticality condition (3) is equivalent to the classical stationarity condition $\nabla f(x) = 0$.*

Let $x \in \Omega$ be a point which is not Pareto critical. Then, there exists a direction $d \in \Omega - x$ such that

$$J_F(x)d \in -\mathbb{R}^m_{++},$$

that is,

$$J_F(x)d \prec_w 0.$$

In this case, d is called a descent direction for F at x. According to the next lemma, a descent direction of F at x can be obtained by solving the following problem:

$$\min_{d \in \Omega - x} h_x(d) \equiv \max_{i \in I} \left\{ \nabla f_i(x)^T d \right\} + \frac{\|d\|^2}{2}. \tag{4}$$

Lemma 2.5. *The following statements hold:*

(a) The subproblem (4) has only one solution.

(b) If x is a Pareto critical point of F and $s(x) = \arg\min_{d \in \Omega - x} h_x(d)$, then $s(x) = 0$ and consequently $h_x(s(x)) = 0$.

(c) If $x \in \Omega$ is not a Pareto critical point of F, then $s(x) \neq 0$ and $h_x(s(x)) < 0$. In particular, $s(x)$ is a descent direction for F at x.

(d) The mapping $x \longmapsto s(x)$ is continuous.

Proof. See Proposition 3 in [11] and Lemma 1 in [15]. □

By statements (b) and (d) in Lemma 2.5, $\|s(x)\|$ is a suitable Pareto criticality measure for x. This remark motivates the following definition.

Definition 2.6. *Given ϵ, we say that x is an ϵ-approximate Pareto critical point for (1)-(2) when $\|s(x)\| \leq \epsilon$.*

It is worth mentioning that this is not the only way to define ϵ-approximate Pareto criticality. For example, in the case $\Omega = \mathbb{R}^n$, [6] consider x as an ϵ-approximate Pareto critical point when

$$\min_{d \in B[0;1]} \max_{i \in I} \left\{ \nabla f_i(x)^T d \right\} \geq -\epsilon$$

or, equivalently,

$$-\min_{d \in B[0;1]} \max_{i \in I} \left\{ \nabla f_i(x)^T d \right\} \leq \epsilon, \tag{5}$$

where $B[0,1] = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$. Notice that

$$\begin{aligned} \min_{d \in B[0;1]} \max_{i \in I} \left\{ \nabla f_i(x)^T d \right\} &\leq \max_{i \in I} \left\{ \nabla f_i(x)^T \left(\frac{s(x)}{\|s(x)\|} \right) \right\} \\ &= \frac{1}{\|s(x)\|} \max_{i \in I} \left\{ \nabla f_i(x)^T s(x) \right\} \\ &\leq \frac{1}{\|s(x)\|} \left(\max_{i \in I} \left\{ \nabla f_i(x)^T s(x) \right\} + \frac{1}{2} \|s(x)\|^2 \right). \end{aligned} \tag{6}$$

Denote

$$\xi(x) = \max_{i\in I}\left\{\nabla f_i(x)^T s(x)\right\} + \frac{1}{2}\|s(x)\|^2. \tag{7}$$

By Proposition 2 in [29], we have

$$\|s(x)\|^2 = 2(-\xi(x)). \tag{8}$$

Combining (6), (7) and (8), it follows that

$$\min_{d\in B[0;1]}\max_{i\in I}\left\{\nabla f_i(x)^T d\right\} \le -\frac{\|s(x)\|}{2}.$$

Consequently,

$$\|s(x)\| \le 2\left(-\min_{d\in B[0;1]}\max_{i\in I}\left\{\nabla f_i(x)^T d\right\}\right).$$

Therefore, if x is an ϵ-approximate Pareto critical point in the sense of (5), then $\|s(x)\| \le 2\epsilon$.

3 Universal Nonmonotone Line Search Methods

Let us consider the following general algorithm to solve (1)-(2):

Algorithm 1: Universal Nonmonotone Line Search Method

Step 0. Choose $x_0 \in \Omega$, $m_0 \in \{0, 1, \ldots, m\}$, $c_1, c_2 > 0$, and $\beta, \rho \in (0, 1)$. Set $k := 0$.

Step 1. Compute $d(x_k) \in \Omega - x_k$ such that

$$\max_{i\in I}\left\{\nabla f_i(x_k)^T d(x_k)\right\} \le -c_1\|s(x_k)\|^2 \tag{9}$$

and

$$\|d(x_k)\| \le c_2\|s(x_k)\|. \tag{10}$$

Step 2.1. Set $\ell := 0$.

Step 2.2. Choose $\nu_{k,\ell} \in \mathbb{R}^m_+$. If

$$\left|\left\{i \in I \,:\, f_i(x_k + \beta^\ell d(x_k)) \le f_i(x_k) + \rho\beta^\ell \nabla f_i(x_k)^T d(x_k)\right\}\right| \ge m_k, \tag{11}$$

and

$$f_i(x_k + \beta^\ell d(x_k)) \le f_i(x_k) + \rho\beta^\ell \nabla f_i(x_k)^T d(x_k) + [\nu_{k,\ell}]_i \quad \forall i \in I, \tag{12}$$

set $\ell_k = \ell$, and go to Step 3. Otherwise, set $\ell := \ell + 1$ and go back to Step 2.2.

Step 3. Define $\alpha_k = \beta^{\ell_k}$, $x_{k+1} = x_k + \alpha_k d(x_k)$, and $\nu_k = \nu_{k,\ell_k}$. Choose $m_{k+1} \in \{0, 1, \ldots, m\}$, set $k := k + 1$ and go back to Step 1.

Remark 3.1. *A natural choice for the search direction is $d(x_k) = s(x_k)$, which satisfies conditions (9)-(10) with $c_1 = c_2 = 1$. In the case $\Omega = \mathbb{R}^n$, given symmetric positive definite matrices $B_i(x_k) \in$*

$\mathbb{R}^{n\times n}$ *($i = 1, \ldots, m$), we can also use the search direction*

$$d(x_k) = \arg\min_{d\in\mathbb{R}^n} \max_{i\in I} \left\{ \nabla f_i(x_k)^T d + \frac{1}{2} d^T B_i(x_k) d \right\}. \tag{13}$$

Indeed, if $\lambda_{\max}(B_i(x_k)) \leq 1/2c_1$, and $\lambda_{\min}(B_i(x_k)) \geq 1/c_2$, then the vector $d(x_k)$ given in (13) also satisfies (9)-(10)[1]*. In particular, when all objectives are strongly convex with Lipschitz continuous gradients, the Newtonian direction will satisfy (9)-(10) for constants c_1 and c_2 that depend on the extreme eigenvalues of the Hessian matrices of the objectives.*

Remark 3.2. *When $[\nu_{k,\ell}]_i > 0$, condition (12) permits the acceptance of a stepsize β^ℓ even if $f_i(x_k + \beta^\ell d(x_k)) > f_i(x_k)$. In contrast, when $m_k \geq 1$, condition (11) requires a monotonic decrease for at least m_k of the objectives.*

The analysis of Algorithm 1 will be done under the following assumptions.

A1 For each $i = 1, \ldots, m$, the objective function f_i belongs to the class of functions $C^{1,\theta_i}_{H_i}(\Omega)$, $\theta_i \in (0, 1]$, which have Hölder continuous gradients:

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq H_i \|y - x\|^{\theta_i}, \text{ for all } x, y \in \Omega.$$

A2 For each $i \in I$, there exists $f_i^* \in \mathbb{R}$ such that $f_i(x) \geq f_i^*$ for all $x \in \Omega$.

A3 For each $i \in I$, $\lim_{T\to+\infty} \frac{1}{T} \sum_{k=0}^{T-1} [\nu_k]_i = 0$.

The next lemma establishes that Algorithm 1 is well-defined.

Lemma 3.3. *Suppose that A1 holds. If $s(x_k) \neq 0$, then there exists $\ell \in \mathbb{N}$ such that*

$$f_i(x_k + \beta^\ell d(x_k)) \leq f_i(x_k) + \rho\beta^\ell \nabla f_i(x_k)^T d(x_k) \tag{14}$$

for all $i \in I$.

Proof. Since $s(x_k) \neq 0$, it follows from (9) that

$$\max_{i\in I} \left\{ \nabla f_i(x_k)^T d(x_k) \right\} < -c_1 \|s(x_k)\|^2 < 0.$$

Thus, for any $i \in I$, we have

$$\nabla f_i(x_k)^T d(x_k) < 0. \tag{15}$$

In view of A1, f_i is differentiable, and so

$$\lim_{\alpha\to 0^+} \frac{f_i(x_k + \alpha d(x_k)) - f_i(x_k)}{\alpha} = \nabla f_i(x_k)^T d(x_k) < \rho \nabla f_i(x_k)^T d(x_k), \tag{16}$$

where the last inequality is due to (15) and $\rho \in (0, 1)$. As a consequence of (16), there exists $\delta_i > 0$ such that

$$f_i(x_k + \alpha d(x_k)) \leq f_i(x_k) + \rho\alpha \nabla f_i(x_k)^T d(x_k) \tag{17}$$

for all $\alpha \in (0, \delta_i]$. Thus, defining $\delta = \min_{i\in I} \{\delta_i\}$, it follows that (17) holds for all $i \in I$ as long as $\alpha \in (0, \delta]$. Since $\beta \in (0, 1)$, there exists $\ell \in \mathbb{N}$ such that $\beta^{\ell_k} \leq \delta$. Therefore, for such ℓ, (14) holds for all $i \in I$. □

[1] The proof for this fact follows exactly as in the proof of Proposition 2 in [29], replacing there x with x_k, $\nabla^2 f_i(x)$ with $B_i(x_k)$, ζ with $1/2c_1$, ξ with $1/c_2$; then using the equality $|\xi(x_k)| = \|s(x_k)\|/2$, where $\xi(x_k)$ is defined in (7).

The following lemma gives a lower bound for the sequence $\{\alpha_k\}$.

Lemma 3.4. *Suppose that A1 hold and let $\{x_k\}_{k=0}^T$ be generated by Algorithm 1 with*

$$\|s(x_k)\| > \epsilon, \quad \text{for } k = 0, \dots, T-1, \tag{18}$$

for some $\epsilon \in (0,1)$. Then

$$\alpha_k \geq \kappa_1 \epsilon^{\frac{1-\theta_{\min}}{\theta_{\min}}} \tag{19}$$

for $k = 0, \dots, T-1$, where

$$\kappa_1 = \min\left\{1, \min_{i\in I}\left\{\beta\left[\frac{(1+\theta_i)c_1(1-\rho)}{c_2^{1+\theta_i}H_i}\right]^{\frac{1}{\theta_i}}\right\}\right\} \tag{20}$$

and

$$\theta_{\min} = \min_{i\in I}\{\theta_i\}. \tag{21}$$

Proof. Consider $k \in \{0, \dots, T-1\}$ and consider the index

$$\hat{\ell}_k \equiv \min\{\ell \in \mathbb{N} : (14) \text{ holds for all } i \in I\},$$

which is well defined due to (18) and Lemma 3.3. We will show first that

$$\beta^{\hat{\ell}_k} \geq \kappa_1 \epsilon^{\frac{1-\theta_{\min}}{\theta_{\min}}}. \tag{22}$$

If $\hat{\ell}_k = 0$, then

$$\beta^{\hat{\ell}_k} = 1 > \epsilon^{\frac{1-\theta_{\min}}{\theta_{\min}}} \geq \kappa_1 \epsilon^{\frac{1-\theta_{\min}}{\theta_{\min}}},$$

that is, (22) holds. Now, suppose that $\hat{\ell}_k > 0$. Then, by the definition of $\hat{\ell}_k$, there exists $i \in I$ such that

$$f_i\left(x_k + \beta^{\hat{\ell}_k - 1} d(x_k)\right) - f_i(x_k) > \rho\beta^{\hat{\ell}_k-1}\nabla f_i(x_k)^T d(x_k) + [\nu_{k,\ell}]_i. \tag{23}$$

On the other hand, by A1 and Lemma 1 in [37] we have

$$f_i\left(x_k + \beta^{\hat{\ell}_k-1}d(x_k)\right) \leq f_i(x_k) + \beta^{\hat{\ell}_k-1}\nabla f_i(x_k)^T d(x_k) + \frac{H_i\left(\beta^{\hat{\ell}_k-1}\right)^{1+\theta_i}}{1+\theta_i}\|d(x_k)\|^{1+\theta_i}. \tag{24}$$

Combining (23) and (24), it follows that

$$\rho\beta^{\hat{\ell}_k-1}\nabla f_i(x_k)^T d(x_k) < \beta^{\hat{\ell}_k-1}\nabla f_i(x_k)^T d(x_k) + \frac{H_i\left(\beta^{\hat{\ell}_k-1}\right)^{1+\theta_i}}{1+\theta_i}\|d(x_k)\|^{1+\theta_i},$$

which implies that

$$\left(\beta^{\hat{\ell}_k-1}\right)^{\theta_i} > \frac{(1+\theta_i)(1-\rho)}{H_i}\left(-\frac{\nabla f_i(x_k)^T d(x_k)}{\|d(x_k)\|^{1+\theta_i}}\right). \tag{25}$$

By (9) and (10) we have

$$-\frac{\nabla f_i(x_k)^T d(x_k)}{\|d(x_k)\|^{1+\theta_i}} \geq \frac{c_1\|s(x_k)\|^2}{c_2^{1+\theta_i}\|s(x_k)\|^{1+\theta_i}} = \frac{c_1}{c_2^{1+\theta_i}}\|s(x_k)\|^{1-\theta_i}. \tag{26}$$

Combining (25) and (26), we obtain

$$\beta^{\hat{\ell}_k} = \beta\left(\beta^{\hat{\ell}_k - 1}\right) > \beta\left[\frac{(1+\theta_i)c_1(1-\rho)}{c_2^{1+\theta_i}H_i}\right]^{\frac{1}{\theta_i}} \|s(x_k)\|^{\frac{1-\theta_i}{\theta_i}}.$$

Then, by (18) we have

$$\beta^{\hat{\ell}_k} > \beta\left[\frac{(1+\theta_i)c_1(1-\rho)}{c_2^{1+\theta_i}H_i}\right]^{\frac{1}{\theta_i}} \epsilon^{\frac{1-\theta_i}{\theta_i}} \geq \min_{i\in I}\left\{\beta\left[\frac{(1+\theta_i)c_1(1-\rho)}{c_2^{1+\theta_i}H_i}\right]^{\frac{1}{\theta_i}}\right\} \epsilon^{\frac{1-\theta_{\min}}{\theta_{\min}}} \geq \kappa_1 \epsilon^{\frac{1-\theta_{\min}}{\theta_{\min}}},$$

that is, (22) also holds when $\hat{\ell}_k > 0$. Finally, since $\ell_k \leq \hat{\ell}_k$, it follows from (22) that

$$\alpha_k = \beta^{\ell_k} \geq \beta^{\hat{\ell}_k} \geq \kappa_1 \epsilon^{\frac{1-\theta_{\min}}{\theta_{\min}}}.$$

□

Remark 3.5. *By Lemma 3.4 and the fact that $\alpha_k = \beta^{\ell_k}$, it follows that*

$$\ell_k \leq \frac{\log\left(\kappa_1^{-1}\epsilon^{-\left(\frac{1-\theta_{\min}}{\theta_{\min}}\right)}\right)}{|\log(\beta)|}.$$

This means that, each iteration of Algorithm 1 with $\|s(x_k)\| > \epsilon$ requires the computation of one Jacobian matrix of $F(\cdot)$ and at most $\mathcal{O}\left(\log\left(\epsilon^{-\left(\frac{1-\theta_{\min}}{\theta_{\min}}\right)}\right)\right)$ evaluations of $F(\cdot)$.

Given $i \in I$, it follows from Assumption A3 that for any $\delta > 0$ there exists $C_i(\delta) > 0$ such that

$$\frac{1}{T}\sum_{k=0}^{T-1}[\nu_k]_i \leq \delta, \quad \forall T \geq C_i(\delta). \tag{27}$$

The theorem below establishes an upper bound for the number of iterations that Algorithm 1 need to find an ϵ-approximate Pareto critical point. The proof is a direct adaptation of the proof of Theorem 2 in [21].

Theorem 3.6. *Suppose that A1-A3 hold and let $\{x_k\}_{k=0}^T$ be generated by Algorithm 1 with*

$$\|s(x_k)\| > \epsilon, \quad \text{for } k = 0, \ldots, T-1, \tag{28}$$

for some $\epsilon \in (0,1)$. Then

$$T \leq \min_{i\in I}\max\left\{C_i\left(\frac{\kappa_2}{2}\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}\right), \frac{2(f_i(x_0)-f_i^*)}{\kappa_2}\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right\} \tag{29}$$

where $C_i(\cdot)$ is defined in (27) and

$$\kappa_2 = c_1\rho\kappa_1, \tag{30}$$

with κ_1 given in (20).

Proof. Let $i \in I$. By (12), (26), (28), Lemma 3.4, and (30), we have

$$\begin{aligned}[\nu_k]_i + f_i(x_k) - f_i(x_{k+1}) &\geq \rho\alpha_k \left(-\nabla f_i(x_k)^T d(x_k)\right) \geq c_1\rho\alpha_k \|s(x_k)\|^2 \\ &\geq c_1\rho\kappa_1\epsilon^{\left(\frac{1-\theta_{\min}}{\theta_{\min}}\right)}\epsilon^2 \\ &= \kappa_2\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)},\end{aligned}$$

for $k = 0, \ldots, T-1$. Now, summing up these inqualities and using A2, we get

$$\sum_{k=0}^{T-1} [\nu_k]_i + f_i(x_0) - f_i^* \geq T\kappa_2\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)},$$

which gives

$$\frac{1}{\kappa_2 T}\sum_{k=0}^{T-1} [\nu_k]_i + \frac{f_i(x_0) - f_i^*}{\kappa_2 T} \geq \epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}. \tag{31}$$

Suppose that

$$T \geq C_i\left(\frac{\kappa_2}{2}\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}\right).$$

In view of the definition of $C_i(\,\cdot\,)$ in (27), this means that

$$\frac{1}{T}\sum_{k=0}^{T-1} [\nu_k]_i \leq \frac{\kappa_2}{2}\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}. \tag{32}$$

In this case, combining (31) and (32), it follows that

$$\frac{f_i(x_0) - f_i^*}{\kappa_2 T} \geq \frac{1}{2}\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}$$

and so

$$T \leq \frac{2(f_i(x_0) - f_i^*)}{\kappa_2}\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}$$

Therefore, in any case we have

$$T \leq \max\left\{C_i\left(\frac{\kappa_2}{2}\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}\right), \frac{2(f_i(x_0) - f_i^*)}{\kappa_2}\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right\}. \tag{33}$$

Since $i \in I$ was arbitrarily chose, it follows that (33) holds for all $i \in I$. Consequently, (29) is true. □

As a consequence of Theorem 3.6 we have the following global convergence result for Algorithm 1.

Corollary 3.7. *Suppose that A1-A3 hold and let $\{x_k\}_{k\geq 0}$ be a sequence generated by Algorithm 1. Then, either exists $\bar{k}$ such that $s(x_{\bar{k}}) = 0$ or*

$$\liminf_{k\to+\infty} \|s(x_k)\| = 0. \tag{34}$$

Proof. Let $\epsilon \in (0,1)$. From Theorem 3.6, if

$$T > \min_{i\in I} \max\left\{ C_i \left(\frac{\kappa_2}{2} \epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)} \right), \frac{2(f_i(x_0) - f_i^*)}{\kappa_2} \epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)} \right\}$$

then

$$\min_{k=0,\ldots,T-1} \|s(x_k)\| \leq \epsilon.$$

Since $\epsilon \in (0,1)$ was chosen arbitrarily, this shows that

$$\lim_{k\to+\infty} \left(\min_{k=0,\ldots,T-1} \|s(x_k)\| \right) = 0.$$

Thus, either there exists $\bar{k}$ such that $s(x_{\bar{k}}) = 0$ or (34) holds. □

4 Particular Cases

In Algorithm 1, different choices for $\nu_{k,\ell} \in \mathbb{R}^m_+$ in Step 2.2 produce different methods. For example, if

$$\nu_{k,\ell} = \nu_k \equiv 0 \text{ for all } k \text{ and } \ell, \tag{35}$$

then Algorithm 1 reduces to an instance of the monotone projected gradient method proposed in [11]. Clearly, this choice gives a sequence $\{\nu_k\}_{k\geq 0}$ that satisfies A3 and for which (27) gives

$$C_i(\delta) = 1 \quad \text{for each } i \in I.$$

Therefore, it follows from Theorem 3.6 that the monotone version of Algorithm 1 takes at most $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right)$ iterations to generate a ϵ-approximate Pareto critical point of (1)-(2).

Let us consider now the choice

$$\nu_{k,\ell} = \nu_k \equiv \begin{cases} 0, & \text{if } k = 0, \\ (1-\delta_k)\left(F(x_{k-1}) + \nu_{k-1} - F(x_k)\right), & \text{if } k \geq 1, \end{cases} \quad \text{for all } \ell \tag{36}$$

with

$$\delta_k \in [\delta_{\min}, 1] \text{ for all } k, \text{ and } \delta_{\min} \in (0,1). \tag{37}$$

For the choice specified by (36) and (37), the next lemma establishes that $\{[\nu_k]_i\}_{k\geq 0}$ is summable for all $i \in I$. The proof is an adaptation of the proof of Lemma Theorem 4 in [20].

Lemma 4.1. *Suppose that A1-A3 hold and let $\{x_k\}_{k\geq 0}$ be generated by Algorithm 1. If $\{\nu_k\}_{k\geq 0}$ is defined by (36) and (37), then $\nu_k \in \mathbb{R}^m_+$ for all k and*

$$\sum_{k=0}^{+\infty} [\nu_k]_i < +\infty \quad \textit{for all } i \in I. \tag{38}$$

Proof. By (36), $\nu_0 = 0 \in \mathbb{R}^m_+$. In view of (12), for all $i \in I$ we have

$$f_i(x_k) + [\nu_k]_i - f_i(x_{k+1}) \geq \rho\alpha_k \left(-\nabla f_i(x_k)^T d(x_k)\right).$$

Consequently, it follows from (37) that

$$(1-\delta_{k+1})\left(f_i(x_k)+[\nu_k]_i-f_i(x_{k+1})\right)\geq(1-\delta_{k+1})\rho\alpha_k\left(-\nabla f_i(x_k)^Td(x_k)\right).$$

Then, by (36) and (26) we get

$$\begin{aligned}[\nu_{k+1}]_i &= (1-\delta_{k+1})\left(f_i(x_k)+[\nu_k]_i-f_i(x_{k+1})\right)\\ &\geq (1-\delta_{k+1})\rho\alpha_k\left(-\nabla f_i(x_k)^Td(x_k)\right)\\ &\geq c_1(1-\delta_{k+1})\rho\alpha_k\|s(x_k)\|^2.\end{aligned}\tag{39}$$

Therefore, $\nu_{k+1}\in\mathbb{R}^m_+$. Now, let $N\geq 1$. Combining (37) and (39) we get

$$\begin{aligned}\sum_{k=0}^{N}[\nu_k]_i &= \sum_{k=0}^{N-1}[\nu_{k+1}]_i=\sum_{k=0}^{N-1}(1-\delta_{k+1})\left(f_i(x_k)+[\nu_k]_i-f_i(x_{k+1})\right)\\ &\leq \sum_{k=0}^{N-1}(1-\delta_{\min})\left(f_i(x_k)-f_i(x_{k+1})+[\nu_k]_i\right)\\ &= (1-\delta_{\min})\left(f_i(x_0)-f_i(x_N)\right)+\sum_{k=0}^{N-1}[\nu_k]_i-\delta_{\min}\sum_{k=0}^{N-1}[\nu_k]_i\\ &\leq (1-\delta_{\min})\left(f_i(x_0)-f_i^*\right)+\sum_{k=0}^{N}[\nu_k]_i-\delta_{\min}\sum_{k=0}^{N-1}[\nu_k]_i,\end{aligned}$$

where the last inequality is due to A2 and thet fact that $[\nu_N]_i\geq 0$. Thus

$$\delta_{\min}\sum_{k=0}^{N-1}[\nu_k]_i\leq(1-\delta_{\min})(f_i(x_0)-f_i^*)$$

and so

$$\sum_{k=0}^{N-1}[\nu_k]_i\leq\left(\frac{1-\delta_{\min}}{\delta_{\min}}\right)(f_i(x_0)-f_i^*).$$

Since $i\in I$ and $N\geq 1$ were taken arbitrarily, this means that (38) is true. □

In view of Lemma 4.1, the sequence $\{\nu_k\}$ defined by (36) and (37) satisfies A3. Moreover, for this sequence, (27) holds with

$$C_i(\delta)=\left(\frac{1-\delta_{\min}}{\delta_{\min}}\right)(f_i(x_0)-f_i^*)\delta^{-1}.$$

Therefore, it follows from Theorem 3.6 that the corresponding instance of Algorithm 1 takes at most $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right)$ iterations to generate a ϵ-approximate Pareto critical point of (1)-(2).

Notice that the sequence $\{\nu_k\}_{k\geq 0}$ given by (36) and (37) is implicitly defined by the choice of the sequence $\{\delta_k\}\subset[\delta_{\min},1]$. One possibility is to use

$$\delta_k=\frac{1}{Q_k},\quad\text{for all } k\geq 1,\tag{40}$$

where

$$Q_0 = 1 \quad \text{and} \quad Q_{k+1} = \eta_k Q_k + 1 \tag{41}$$

for a given sequence $\eta_k \in [\eta_{\min}, \eta_{\max}]$ with $0 \leq \eta_{\min} \leq \eta_{\max} < 1$. In this case we have

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

and so

$$\delta_{k+1} \geq 1 - \eta_{\max} \equiv \delta_{\min}.$$

By (36), we also have

$$F(x_{k+1}) + \nu_{k+1} = (1-\delta_{k+1})(F(x_k) + \nu_k) + \delta_{k+1} F(x_{k+1}). \tag{42}$$

Thus, denoting

$$C_k = F(x_k) + \nu_k,$$

it follows from (42), (40) and (41) that

$$C_{k+1} = \frac{\eta_k Q_k C_k + F(x_{k+1})}{Q_{k+1}}.$$

Moreover, using this notation, condition (12) can be rewritten as

$$f_i(x_k + \beta^\ell d(x_k)) \leq [C_k]_i + \rho \beta^\ell \nabla f_i(x_k)^T d(x_k) \quad \text{for all } i \in I.$$

This means that Algorithm 1 with $\{\nu_k\}$ given by (36), (40) and (41) reduces to:

- an instance of the nonmonotone projected gradient method proposed in [13], when $m_k = 0$ for all $k \geq 0$; and
- the variant of Algorithm 6 proposed in [29] for the case $\Omega = \mathbb{R}^n$, with nonmonotone term inspired by [39].

In particular, it follows from our analysis that these methods also possess an upper complexity bound of $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right)$ iterations.

Note that multiobjective variants of the nonmonotone line search of [23] can also be seen as particular instances of Algorithm 1 with

$$[\nu_{k,\ell}]_i = [\nu_k]_i = \max_{0 \leq j \leq M(k)} f_i(x_{k-j}) - f_i(x_k), \quad \text{for} \quad i = 1, \ldots, m,$$

where $M(k) = \min\{k, M\}$. However, it is not clear whether the corresponding sequence $\{\nu_k\}_{k\geq 0}$ is summable. Thus, the worst-case complexity of this variant remains unknown to us.

5 A New Nonmonotone Method for Multiobjective Optimization

As we saw in the last section, with different choices of $\nu_{k,\ell}$ at Step 2.2 of Algorithm 1, we obtain different methods. From our analysis, to have a globally convergent method it is enough to select $\nu_{k,\ell}$ such that the corresponding sequence $\{\nu_k\}$ satisfies assumption A3. The next lemma establishes that A3 is very mild, in the sense that it holds for any sequence $\{\nu_k\}$ with $\lim_{k\to+\infty}\nu_k = 0$.

Lemma 5.1. *Let $\{\nu_k\}_{k\geq 0} \subset \mathbb{R}^m_+$ with $\nu_k \to 0$. Then $\{\nu_k\}_{k\geq 0}$ satisfies assumption A3.*

Proof. Let $i \in I$. By assumption, $\lim_{k\to+\infty}[\nu_k]_i = 0$. Thus, given $\delta > 0$, there exists $\xi_i(\delta/2) \in \mathbb{N}\setminus\{0\}$ such that

$$[\nu_k]_i \leq \frac{\delta}{2} \quad \text{for all } k \geq \xi_i(\delta/2). \tag{43}$$

Moreover, there exists $M_i > 0$ such that

$$[\nu_k]_i \leq M_i \quad \text{for all } k \geq 0. \tag{44}$$

Consider

$$C_i(\delta) = \max\left\{\frac{2\xi_i(\delta/2)M_i}{\delta}, 1+\xi_i(\delta/2)\right\}. \tag{45}$$

If $T \geq C_i(\delta)$ we have

$$\begin{aligned}
\frac{1}{T}\sum_{k=0}^{T-1}[\nu_k]_i &= \frac{1}{T}\left(\sum_{k=0}^{\xi_i(\delta/2)-1}[\nu_k]_i\right) + \frac{1}{T}\left(\sum_{k=\xi_i(\delta/2)}^{T-1}[\nu_k]_i\right)\\
&\leq \frac{1}{T}\left(\sum_{k=0}^{\xi_i(\delta/2)-1}M_i\right) + \frac{1}{T}\left(\sum_{k=\xi_i(\delta/2)}^{T-1}\frac{\delta}{2}\right)\\
&\leq \frac{1}{T}\xi_i(\delta/2)M_i + \frac{1}{T}\left(\sum_{k=0}^{T-1}\frac{\delta}{2}\right)\\
&= \frac{\delta}{2}+\frac{\delta}{2}\\
&= \delta.
\end{aligned}$$

This shows that

$$\lim_{T\to+\infty}\frac{1}{T}\sum_{k=0}^{T-1}[\nu_k]_i = 0.$$

Since $i \in I$ was taken arbitrarily, we conclude that $\{\nu_k\}_{k\geq 0}$ satisfies A3. □

It follows from Lemma 5.1 and Corollary 3.7 that any instance of Algorithm 1 with $\nu_k \to 0$ is globally convergent for the problem class specified by assumptions A1 and A2. This gives a great deal of freedom for the development of new nonmonotone methods for multiobjective optimization problems. As an example, let us consider the following new instance of Algorithm 1.

Algorithm 2: Metropolis-Based Nonmonotone Line Search Method

Step 0. Choose $x_0 \in \Omega$, $\beta, \rho \in (0,1)$, $c_1, c_2 > 0$, $\sigma \in \mathbb{R}^m_+$, $\gamma > 0$, and $\{\tau_k\}_{k\geq 0} \subset \mathbb{R}_{++}$ with $\tau_k \to 0$. Set $k := 0$.

Step 1. Compute $d(x_k) \in \Omega - x_k$ such that

$$\max_{i\in I}\left\{\nabla f_i(x_k)^T d(x_k)\right\} \leq -c_1\|s(x_k)\|^2$$

and

$$\|d(x_k)\| \leq c_2\|s(x_k)\|.$$

Step 2.1. Set $\ell := 0$.

Step 2.2. Compute $x^+_{k,\ell} = x_k + \beta^\ell d(x_k)$ and define

$$[\nu_{k,\ell}]_i = \sigma_i \exp\left(-\frac{\max\left\{\gamma, f_i(x^+_{k,\ell}) - f_i(x_k)\right\}}{\tau_k}\right), \quad \forall i \in I. \tag{46}$$

If

$$f_i(x^+_{k,\ell}) \leq f_i(x_k) + \rho\beta^\ell \nabla f_i(x_k)^T d(x_k) + [\nu_{k,\ell}]_i \quad \forall i \in I,$$

set $\ell_k = \ell$ and go to Step 3. Otherwise, set $\ell := \ell + 1$ and repeat Step 2.2.

Step 3. Set $\nu_k = \nu_{k,\ell_k}$, $\alpha_k = \beta^{\ell_k}$, $x_{k+1} = x^+_{k,\ell_k}$, $k := k+1$ and go back to Step 1.

Algorithm 2 is a generalization of the Metropolis-based nonmonotone method proposed in [21] for single-objective optimization. As a consequence of Corollary 3.7 we have the following global convergence result for Algorithm 2.

Theorem 5.2. *Suppose that A1-A2 hold, and let $\{x_k\}_{k\geq 0}$ be a sequence generated by Algorithm 2. Then, either there exists $\bar{k}$ such that $d(x_{\bar{k}}) = 0$ or*

$$\liminf_{k\to+\infty} \|s(x_k)\| = 0.$$

Proof. In view of Corollary 3.7 and Lemma 5.1 it is enough to show that $\{\nu_k\}_{k\geq 0}$ generated by Algorithm 2 converges to $0 \in \mathbb{R}^m$. By Steps 2.2 and 3 in Algorithm 2, for all $i \in I$ and $k \geq 0$, we have

$$0 \leq [\nu_k]_i = \sigma_i \exp\left(-\frac{\max\{\gamma, f_i(x_{k+1}) - f_i(x_k)\}}{\tau_k}\right) \leq \sigma_i \exp\left(-\frac{\gamma}{\tau_k}\right).$$

Since $\gamma, \tau_k > 0$ and $\tau_k \to 0$, it follows that

$$\lim_{k\to+\infty} [\nu_k]_i = 0, \quad \forall i \in I,$$

and so $\nu_k \to 0$. □

From the proof Lemma 5.1 we see that if $\nu_k \to 0$ then (27) holds for

$$C_i(\delta) = \max\left\{\frac{2\xi_i(\delta/2)M_i}{\delta}, 1 + \xi_i(\delta/2)\right\},$$

where M_i is a uniform upper bound to $\{[\nu_k]_i\}_{k\geq 0}$, and $\xi_i(\delta/2)$ is any positive integer such that

$$[\nu_k]_i \leq \frac{\delta}{2}, \quad \forall k \geq \xi_i(\delta/2).$$

Thus, when $\nu_k \to 0$, it follows from Theorem 3.6 that Algorithm 1 takes no more than

$$\left\lceil \min_{i\in I} \max \left\{ \frac{4\xi_i\left(\frac{\kappa_2}{4}\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}\right) M_i}{\kappa_2\epsilon^{\left(1+\frac{1}{\theta_{min}}\right)}}, 1+\xi_i\left(\frac{\kappa_2}{4}\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}\right), \frac{2(f_i(x_0)-f_i^*)}{\kappa_2\epsilon^{\left(1+\frac{1}{\theta_{\min}}\right)}} \right\} \right\rceil \tag{47}$$

iterations to find a ϵ-approximate Pareto critical point of (1)-(2). Therefore, to obtain an explicit iteration complexity bound, all that we need to do is to estimate the rate of decay of $\{[\nu_k]_i\}_{k\geq 0}$, which will allow the identification of M_i and $\xi_i(\delta/2)$. From the proof of Theorem 5.2, we know that the sequence $\{[\nu_k]_i\}_{k\geq 0}$ in Algorithm 2 satisfies

$$0 \leq [\nu_k]_i \leq \sigma_i \exp\left(-\frac{\gamma}{\tau_k}\right), \quad \forall k \geq 0.$$

Let us consider the choice $\tau_k = 1/\ln(k+1)$. In this case, we get

$$[\nu_k]_i \leq \sigma_i \exp(-\gamma \ln(k+1)) = \sigma_i \exp\left(\ln\left((k+1)^{-\gamma}\right)\right) = \frac{\sigma_i}{(k+1)^{\gamma}}. \tag{48}$$

This implies that

$$[\nu_k]_i \leq \sigma_i \equiv M_i, \quad \forall k \geq 0, \tag{49}$$

and

$$[\nu_k]_i \leq \frac{\delta}{2}, \quad \forall k \geq \left(\frac{2\sigma_i}{\delta}\right)^{\frac{1}{\gamma}} \equiv \xi_i(\delta/2). \tag{50}$$

Therefore, it follows from (47), (49) and (50) that Algorithm 2 with $\tau_k = 1/\ln(k+1)$ takes no more than $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)\left(1+\frac{1}{\gamma}\right)}\right)$ iterations to find a ϵ-approximate Pareto critical point of (1)-(2).

For the case $\gamma > 1$, an improved complexity bound can be obtained for Algorithm 2. Indeed, if $\gamma > 1$, then it follows from (48) that

$$\sum_{k=0}^{+\infty} [\nu_k]_i \leq \sigma_i \sum_{k=0}^{+\infty} \frac{1}{(k+1)^{\gamma}} \leq \frac{\sigma_i \gamma}{\gamma - 1}.$$

In this case, (27) is satisfied with $C_i(\delta) = \left(\frac{\sigma_i\gamma}{\gamma-1}\right)\delta^{-1}$. Consequently, by Theorem 3.6, Algorithm 2 with $\gamma > 1$ and $\tau_k = 1/\ln(k+1)$ takes at most $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right)$ iterations to find a ϵ-approximate Pareto critical point.

6 Illustrative Numerical Results

In [21], a variant of Algorithm 2 for single-objective optimization (case $m = 1$) showed a remarkable ability to escape non-global local minimizers. Here, we investigate the performance of Algorithm 2 applied to bi-objective optimization problems where one of the objectives has numerous non-global local minimizers. Specifically, we considered 15 bi-objective problems of the form

$$\begin{array}{ll} \min_{x\in\mathbb{R}^n} & F(x) = (f_1(x), f_2(x)) \\ \text{s.t.} & x \in [-a, a]^n, \end{array}$$

with $a = 5.12$,

$$f_1(x) = 10n + \sum_{i=1}^{n} \left[x_1^2 - 10\cos(2\pi x_i)\right], \quad \forall x \in \mathbb{R}^n, \quad \forall x \in \mathbb{R}^n,$$

and $f_2(\,\cdot\,)$ being one of the 15 functions from the MGH collection [30] whose dimension n can be chosen[2]. Function $f_1(\,\cdot\,)$ is known as the *Rastrigin function*. This function has a large number of spurious minimizers in the hypercube $[-5.12, 5.12]^n$ (see, e.g., [32]). The following `Julia` codes were compared:

- **M**: the monotone version of Algorithm 1, obtained with $\nu_{k,\ell} = 0$ for all k and ℓ, and $m_k = 0$ for all k.
- **N1**: Algorithm 1 with $\nu_{k,\ell}$ given by (36), (40) and (41), $\eta_k = 0.85/(k+1)$, and $m_k = 0$ for all k.
- **N2**: Algorithm 2 with $\tau_k = 1/\ln(k+1)$, $\gamma = 8$, and $\sigma_i = |f_i(x_0)|$ for all $i \in I$.
- **Nh:** The hybrid-type nonmonotone line search as proposed in [29], which uses $m_k = \lceil m/2 \rceil$ for all k.

In all implementations, we considered $d(x_k) = s(x_k)$ (for which $c_1 = c_2 = 1$), and the parameters $\rho = 10^{-4}$ and $\beta = 0.5$. Focusing on the case $n = 4$, for each problem we tested 81 choices for the starting point $x_0 \in \mathbb{R}^4$, namely,

$$x_0 = [-2a + ia, -2a + ja, -2a + ka, -2a + \ell a]^T, \quad i, j, k, \ell \in \{1, 2, 3\}.$$

This resulted in a total of 1215 pairs (problem, starting point). We applied the three solvers in all these pairs with stopping criterion

$$\|s(x_k)\| \le \epsilon \equiv 10^{-4}, \tag{51}$$

allowing a maximum of 1000 iterations for each solver. All the experiments were performed with `Julia` 1.7.2 on a PC with Intel(R) Core(TM) i7-10510U with microprocessor 1.8 GHz and 32 GB RAM. We use `Gurobi` and `JuMP` [12] to compute $s(x_k)$.

Codes are compared using performance profile [10]. Given a set of solvers $\mathcal{S}$, and a set of test problems $\mathcal{P}$, denote by $t_{p,s} > 0$ the performance of the solver $s \in \mathcal{S}$ applied to problem $p \in \mathcal{P}$. Then the performance profile of solver s is the graph of the function $\gamma_s : [1, \infty) \to [0, 1]$ given by

$$\gamma_s(\tau) = \frac{1}{|P|}\left|\left\{p \in \mathcal{P} : \frac{t_{p,s}}{\min\{t_{p,s} : s \in \mathcal{S}\}} \le \tau\right\}\right|.$$

Note that, $\gamma_s(1)$ is the percentage of problems for which solver s wins over the rest of the solvers.

[2]Namely, `Extended Rosenbrock`, `Extended Powell Singular`, `Penalty I`, `Penalty II`, `Variably Dimensioned`, `Trigonometric`, `Discrete Boundary Value`, `Discrete Integral Equation`, `Broyden Tridiagonal`, `Broyden Banded`, `Brown Almost Linear`, `Linear`, `Linear-1`, `Linear-0`, `Chebyquad`.

6.1 Number of Iterations to achieve ϵ-Pareto Criticality

Figure 1 shows the performance profiles of solvers M, N1, N2 and Nh, considering $\mathcal{P} = \{(\text{problem, starting point})\}$ and $t_{p,s}$ as the number of iterations that solver s applied to the pair $p \in \mathcal{P}$ requires to generate x_k such that (51) holds. As we can see, N2 was the only nonmonotone method more efficient than the monotone method. In addition, N1 and N2 were more robust than M and Nh.

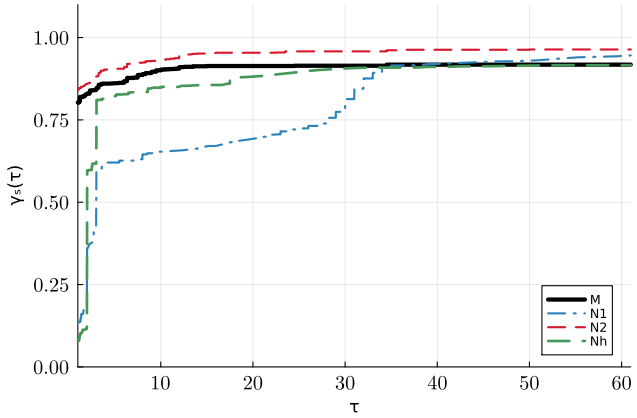

Figure 1: Performance profile with respect to the number of iterations

6.2 Purity Metric

In the context of multiobjective optimization, other performance measures are also relevant. For example, it is particularly interesting to quantify the ability of the solvers to find Pareto efficient solutions, also known as *Pareto front*. For that, we need first to identify an approximation for this set. By applying solver s to problem p starting from N different initial points, we obtain a set of N approximate Pareto critical points in Ω. Removing from this set all the dominated points[3], we get the set $PF_{p,s}$. Let PF_p be the set obtained by removing all the dominated points from $\cup_{s\in S} PF_{p,s}$. Then, PF_p can be seen as an approximation to the Pareto front of problem p. The Purity metric of solver s applied to problem p is defined as

$$t_{p,s} = \begin{cases} 1/\bar{t}_{p,s}, & \text{if } \bar{t}_{p,s} \neq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

where

$$\bar{t}_{p,s} = \frac{|PF_{p,s} \cap PF_p|}{|PF_p|}.$$

When using the Purity metric, as suggested in [9], we compare the algorithms in pairs. As we can see in Figure 2, our new method N2 outperformed both M, N1 and Nh with respect to the Purity metric.

[3]We say that a point x_1 is dominated by a point x_2 when $f_i(x_2) \leq f_i(x_1)$, $\forall i \in I$ and $f_j(x_2) < f_j(x_1)$ for some $j \in I$.

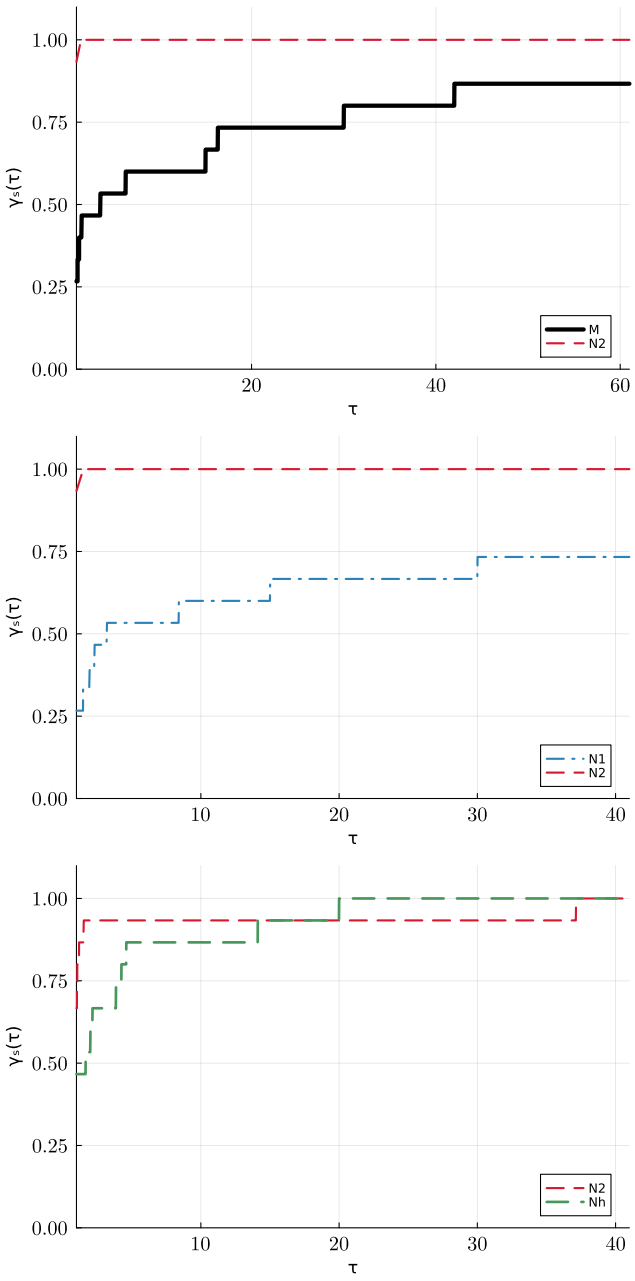

Figure 2: Performance profile with respect to the purity metric

7 Conclusion

In this work, we proposed a universal nonmonotone line search method for nonconvex multiobjective problems with convex constraints. The method is *universal* because it does not require any knowledge about the constants that define the smoothness level of the objectives. Specifically, the step-sizes are selected using a relaxed Armijo condition, allowing the increase of the objectives between consecutive

iterations. The degree of nonmonotonicity allowed for the objective $f_i(\cdot)$ is controlled by a nonnegative parameter $[\nu_k]_i$. Assuming that $\nabla f_i(\cdot)$ is θ_i-Hölder continuous, and that $\{[\nu_k]_i\}_{k\geq 0}$ is summable for every $i \in \{1,\ldots,m\}$, we proved that our method takes no more than $\mathcal{O}\left(\epsilon^{-\left(1+\frac{1}{\theta_{\min}}\right)}\right)$ iterations to find a ϵ-approximate critical point of a problem with m objectives and $\theta_{\min} = \min_{i=1,\ldots,m}\{\theta_i\}$. Under the weaker assumption that $\lim_{k\to+\infty}[\nu_k]_i = 0$ for every $i \in \{1,\ldots,m\}$, we also proved a liminf-type global convergence result. In particular, we showed that our complexity bound applies to some existing monotone and nonmonotone line search methods. In addition, exploring the generality of our assumptions about $\{\nu_k\}_{k\geq 0} \subset \mathbb{R}^m_+$, we proposed a new Metropolis-based nonmonotone method that fits into our general scheme. Our preliminary numerical results indicate that this new method performs favorably in comparison to the monotone projected gradient method by Drummond and Iusem [11], the nonmonotone projected gradient method by Fazzio and Schuverdt [13], and a projected variant of the nonmonotone method by Mita, Fukuda, and Yamashita [29], on problems where at least one objective has numerous non-global local minimizers.

Funding

G.N. Grapiglia was partially supported by the National Council for Scientific and Technological Development (CNPq) - Brazil (grant 312777/2020-5). M.E. Pinheiro was partially supported by the Coordination for the Improvement of Higher Education Personnel (CAPES) - Brazil.

References

[1] J. Bello Cruz, L. Lucambio Pérez, and J. Melo. Convergence of the projected gradient method for quasiconvex multiobjective optimization. *Nonlinear Analysis*, 74:5268–5273, 2011.

[2] G. Bento, J. Cruz Neto, and A. Soubeyran. A proximal point-type method for multicriteria optimization. *Set-Valued and Variational Analysis*, 22:557–573, 09 2014.

[3] H. Bonnel, A. Iusem, and B. Svaiter. Proximal methods in vector optimization. *SIAM Journal on Optimization*, 15:953–970, 01 2005.

[4] L. Calderón, M. Diniz-Ehrhardt, and J. Martínez. On high-order model regularization for multiobjective optimization. *Optimization Methods and Software*, 37:175–191, 2022.

[5] G. Carrizo, P. Lotito, and M. Maciel. Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. *Mathematical Programming*, 159:339–369, 10 2016.

[6] G. Cocchi, , and M. Lapucci. An augmented lagrangian algorithm for multi-objective optimization. *Computational Optimization and Applications*, 77:29–56, 2020.

[7] V. Coverstone-Carroll, J. Hartmann, and W. Mason. Optimal multi-objective low-thrust spacecraft trajectories. *Computer Methods in Applied Mechanics and Engineering*, 186:387–402, 06 2000.

[8] D. Craft, T. Halabi, H. Shih, and T. Bortfeld. An approach for practical multiobjective IMRT treatment planning. *International Journal of Radiation, Oncology, Biology, Physics*, 69:1600–1607, 01 2007.

[9] A. Custódio, J. Madeira, A. Vaz, and L. Vicente. Direct multisearch for multiobjective optimization. *SIAM Journal on Optimization*, 21:1109–1140, 07 2011.

[10] E. Dolan and J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91:201–213, 03 2002.

[11] L. Drummond and A. Iusem. A projected gradient method for vector optimization problems. *Computational Optimization and Applications*, 28:5–29, 04 2004.

[12] I. Dunning, J. Huchette, and M. Lubin. Jump: A modeling language for mathematical optimization. *SIAM Review*, 59:295–320, 2017.

[13] N. Fazzio and M. Schuverdt. Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems. *Optimization Letters*, 13:1365–1379, 09 2019.

[14] J. Fliege, L. Drummond, and B. Svaiter. Newton's method for multiobjective optimization. *SIAM Journal on Optimization*, 20:602–626, 2009.

[15] J. Fliege and B. Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51:479–494, 2000.

[16] J. Fliege, A. Vaz, and L. Vicente. Complexity of gradient descent for multiobjective optimization. *Optimization Methods and Software*, 34:949–959, 08 2019.

[17] D. Frangopol and M. Liu. Maintenance and management of civil infraestructure based on condition, safety, optimization, and life-cycle cost. *Structure and Infraestructure Engineering*, 3:29–41, 2007.

[18] E. Fukuda and L. Drummond. On the convergence of the projected gradient method for vector optimization. *Optimization*, 60:1009–1021, 12 2011.

[19] E. Fukuda and L. Drummond. Inexact projected gradient method for vector optimization. *Computational Optimization and Applications*, 54:473–493, 04 2013.

[20] G. Grapiglia and E. Sachs. On the worst-case evaluation complexity of non-monotone line search algorithms. *Computational Optimization and Applications*, 68:555–577, 07 2017.

[21] G. Grapiglia and E. Sachs. A generalized worst-case complexity analysis for non-monotone line searches. *Numerical Algorithms*, 87:779–796, 06 2021.

[22] G. Grapiglia, J. Yuan, and Y. Yuan. On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization. *Mathematical Programming*, 152:491–520, 07 2015.

[23] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for newton's method. *SIAM Journal on Numerical Analysis*, 23:707–716, 1986.

[24] A. Guerraggio and D. Luc. Optimality conditions for $C^{1,1}$ vector optimization problems. *Journal of Optimization Theory and Applications*, 109:615–629, 06 2001.

[25] J. Handl, D. Kell, and J. Knowles. Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4:279–292, 05 2007.

[26] M. Liu, S. Burns, and Y. Wen. Multiobjective optimization for performance-based seismic design of stell moment frame structures. *Earthquake Engineering and Structural Dynamics*, 34:289–306, 2005.

[27] R. Marler and J. Arora. Survey of multiobjective optimization methods for engineering. *Structural and Multidiciplinary Optmization*, 26:369–395, 2004.

[28] K. Miettinen. *Nonlinear multiobjective optimization.* Kluwer Academic Publishers, Boston, USA, 1998.

[29] K. Mita, E. Fukuda, and N. Yamashita. Nonmonotone line searches for unconstrained multiobjective optimization problems. *Journal of Global Optimization*, 75:63–90, 2019.

[30] J. Moré, B. Garbow, and K. Hillstrom. Testing unconstrained optimization software. *ACM Transactions on Mathematical Software*, 7:17–41, 1981.

[31] Y. Nesterov. Universal gradient methods for convex optimization problems. *Mathematical Programming*, 152:381–404, 2015.

[32] L. Rastrigin. Extremal control systems. *Theoretical Foundations of Engineering Cybernetics Series*, 1974.

[33] E. Sachs and S. Sachs. Nonmonotone line searches for optimization algorithms. *Control and Cybernetics*, 40:1059–1075, 01 2011.

[34] G. Schneider and U. Fechner. Computer-based *de novo* design of drug-like molecules. *Nature Reviews Drug Discovery*, 4:649–663, 2005.

[35] M. Tavana. A subjective assessment of alternative mission architectures for the human exploration of Mars at NASA using multicriteria decision making. *Computers & Operations Research*, 31:1147–1164, 06 2004.

[36] K. Villacorta, P. Oliveira, and A. Soubeyran. A trust-region method for unconstrained multiobjective problem with applications in satiscicing processes. *Journal of Optimization Theory and Applications*, 160:865–889, 2014.

[37] M. Yashtini. On the global convergence rate of the gradient descent method for functions with Hölder continuous gradients. *Optimization Letters*, 10:1361–1370, 08 2016.

[38] Y. Yu. Multiobjective decision theory for computational optimization in radiation therapy. *Medical Physics*, 24:1445–1454, 1997.

[39] H. Zhang and W. W. Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.*, 14:1043–1056, 2004.