Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 2, 2026, pp 563–589
https://doi.org/10.22067/ijnao.2025.94685.1691
https://ijnao.um.ac.ir/

Research Article

A nonmonotone line search method for solving constrained multiobjective optimization problems

M. Rashidi and E. Khorram*

Abstract

In this paper, we propose a globally convergent sequential quadratic programming method for solving constrained multiobjective optimization problems with inequality constraints. At each iteration, a feasible descent direction is computed by solving an auxiliary quadratic programming subproblem constructed from linear approximations of the objective and constraint functions. Constraint violations are addressed using a nonsmooth exact penalty function. Moreover, the algorithm incorporates a nonmonotone max-type line-search strategy, which improves practical performance by allowing temporary increases in the penalty function value and thereby promoting more effective exploration of the search space. Under mild regularity assumptions, we prove that the sequence generated by the method converges to a weakly or strongly critical point. Numerical experiments on standard benchmark problems demonstrate the effectiveness and robustness of the proposed approach, both in terms of computational performance criteria and the quality of the approximated Pareto front.

AMS subject classifications (2020): 90C30, 49M05, 97N40 90B50.

*Corresponding author

Received 30 July 2025; revised 27 December 2025; accepted 30 December 2025

Mehri Rashidi
Faculty of Mathematics and Computer Sciences, Amirkabir University of Technology, Tehran, Iran. e-mail: mrashidi@aut.ac.ir

Esmaile Khorram
Faculty of Mathematics and Computer Sciences, Amirkabir University of Technology, Tehran, Iran. e-mail: eskhor@aut.ac.ir

How to cite this article
Rashidi, M. and Khorram, E., A nonmonotone line search method for solving constrained multiobjective optimization problems. *Iran. J. Numer. Anal. Optim.*, 2026; 16(2): 563-589. https://doi.org/10.22067/ijnao.2025.94685.1691

Keywords: SQP method, Nonmonotone line search, Penalty method, Mangasarian–Fromovitz constraint qualification, Critical point.

1 Introduction

Multiobjective optimization is essential for addressing complex problems in areas such as game theory, economics, and engineering, where several conflicting objectives must be considered simultaneously [28]. The goal is to identify solutions that provide the best possible compromises among multiple criteria over a given decision space. Because optimizing all objectives simultaneously is generally impossible due to inherent trade-offs, the classical notion of optimality is replaced by efficiency. A point is said to be *efficient* (Pareto optimal) if no objective can be improved without deteriorating at least one other objective, and the collection of such points forms the *nondominated* (Pareto) set. Accurate approximation of this set, as well as its image in the objective space known as the Pareto front, is crucial for understanding the trade-off structure and has motivated extensive research on algorithms for approximating nondominated fronts in multiobjective optimization problems (MOPs) [16, 35].

Scalarization is a classical approach for solving MOPs, where a vector objective is reduced to a single-objective formulation, typically through weighted sums, ε-constraint schemes, or related constructions [19, 20, 32]. Although conceptually appealing, scalarization often requires the choice of user-defined parameters that reflect preferences or trade-off weights, which may be difficult to specify reliably and can significantly affect the obtained solutions. Moreover, depending on the scalarization model and the problem structure, the resulting subproblems may exhibit undesirable behavior such as ill-conditioning or even unboundedness [21].

These limitations have motivated the development of preference-free, direction-based iterative methods that work directly in the multiobjective setting. Starting from a random or feasible point, such methods typically compute a common descent direction by solving an auxiliary subproblem, and then employ a line-search procedure to select a suitable step size, aiming to improve all objectives in the Pareto sense [17, 18, 22]. Within this framework, a variety of algorithmic ideas have been explored, including Newton and quasi-Newton type strategies [21, 30, 40], sequential quadratic programming (SQP)-type methods for constrained MOPs [2, 23, 24, 37], and projection-based schemes on convex feasible sets [6, 43]. In parallel, nonmonotone globalization mechanisms have been increasingly used to enhance robustness in single objective settings, building on nonmonotone line-search principles [11, 27, 41] and extending them to nonconvex and unconstrained multiobjective algorithms [25, 34].

Recent developments in nonmonotone optimization, particularly within trust-region and line-search frameworks, have led to substantial improvements for both constrained and unconstrained problems. In contrast to classical monotone schemes that enforce a decrease in the merit function at every iteration, nonmonotone strategies permit occasional increases, which can improve robustness and promote more effective exploration of the optimization landscape, especially in nonconvex settings [27, 41]. In the context of nonmonotone methods for unconstrained MOPs, several recent contributions have advanced both theory and algorithmic design. Mahdavi-Amiri and Salehi Sadaghiani [30] proposed a nonmonotone quasi-Newton scheme with superlinear convergence, while Ghalavand, Khorram, and Morovati [25] developed adaptive nonmonotone trust-region algorithms to enhance robustness. More recently, Upadhayay et al. [39] introduced a nonmonotone conditional gradient framework for MOPs, and Upadhayay, Ghosh, and Kumar [40] investigated nonmonotone Wolfe-type quasi-Newton methods, providing effective globalization strategies in the unconstrained setting. Despite these advances, constrained MOPs remain particularly challenging, and comparatively few studies have focused on algorithms with strong theoretical guarantees and reliable practical performance in this setting. Notable recent contributions include conditional gradient type methods [8], accelerated proximal gradient schemes [38], away-step Frank–Wolfe strategies for constrained MOPs [26], SQP-type approaches combined with gradient sampling for nonsmooth constrained MOPs [37], and proximal gradient methods designed for convex MOPs [42].

For nonmonotone methods in constrained MOPs, recent contributions include Zhao and Yao [43], who proposed a nonmonotone projected gradient framework; Carrizo, Fazzio, and Schuverdt [6], who developed a nonmonotone projected gradient method on convex sets; Pinheiro and Grapiglia [34], who introduced a universal nonmonotone line-search scheme for nonconvex MOPs with convex constraints; and Chen et al. [7], who presented a nonmonotone proximal gradient algorithm for nonsmooth MOPs, with extensions to robust multiobjective optimization. While nonmonotone line search strategies for unconstrained optimization are well established, their extension to constrained MOPs remains comparatively limited. The main difficulty is the presence of constraints, which typically requires merit or penalty functions, and designing a nonmonotone acceptance criterion that preserves both feasibility progress and convergence guarantees is technically delicate. In particular, nonmonotone line search applied directly to an exact penalty function is still rare. To the best of our knowledge, the most recent work explicitly pursuing this direction is Rashidi, Khorram, and Soleimani-Damaneh [36]. This approach integrates an exact penalty formulation with an SQP-type framework, a nonmonotone line-search mechanism, and rapid infeasibility detection. Moreover, an average-type nonmonotone criterion is employed, and up to two inequality-constrained subproblems may be solved at each iteration. Penalty updates are controlled through trust-region ideas, yielding a hybrid strategy that combines features of both trust-region and line-search globalization.

Ansary and Panda [2] proposed an SQP method for inequality-constrained MOPs. Their scheme computes a descent direction from an auxiliary quadratic subproblem based on linear approximations of the objective and constraint functions, and it handles constraint violations through a nonsmooth exact penalty function.

Motivated by the SQP framework of Ansary and Panda [2] and by the use of nonmonotone globalization mechanisms for penalty-based constrained MOPs in [36], we develop a new globally convergent SQP method for inequality-constrained MOPs. The proposed algorithm preserves the main SQP structure of [2] but introduces a different penalty-function formulation together with a distinct strategy for selecting and updating the penalty parameter. These modifications require a technically justification, both to establish global convergence and to justify the observed numerical behavior. In addition, we incorporate a nonmonotone max-type line search rule, which allows controlled temporary increases of the penalty function and can improve robustness by preventing overly conservative steps. Under mild regularity assumptions, we prove that the sequence generated by the method converges to a weakly or strongly critical point. Numerical results on standard benchmarks confirm the effectiveness and robustness of the proposed approach.

The remainder of the paper is organized as follows. Section 2 introduces the basic definitions and preliminaries. Section 3 presents the proposed method. The convergence analysis is developed in Section 4. Section 5 reports the numerical implementation details and computational results. Finally, Section 6 concludes the paper.

2 Preliminaries

We consider the following constrained MOP:

$$\begin{aligned} &\min\ f(x) \\ \text{s.t.}\quad &g(x) \leq 0, \end{aligned} \tag{1}$$

where $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))^T$ and $g(x) = (g_1(x), g_2(x), \ldots, g_r(x))^T$. Here, each $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, m$, and each $g_j : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \ldots, r$. The functions $f(\cdot)$ and $g(\cdot)$ are continuously differentiable over the feasible set $X \subseteq \mathbb{R}^n$, which is defined as:

$$X = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0, \text{ for } j = 1, 2, \ldots, r\}.$$

Thus, the optimization problem can be expressed equivalently as

$$\min f(x)$$
$$\text{subject to} \quad x \in X.$$

In MOP, the objective space is a partial order set, while in single-objective optimization, it is a total order set. Let $\mathbb{R}^m_{\geq}$ and $\mathbb{R}^m_{>}$ be defined as

$$\mathbb{R}^m_{\geq} = \{\mathbf{x} \in \mathbb{R}^m : x_k \geq 0, \text{ for all } k\}, \quad \mathbb{R}^m_{>} = \{\mathbf{x} \in \mathbb{R}^m : x_k > 0, \text{ for all } k\}.$$

These sets describe relationships between vector-valued functions $f(x)$ and $f(y)$ based on component-wise comparisons:

$$f(x) \leq f(y) \Leftrightarrow f(y) - f(x) \in \mathbb{R}^m_{\geq}, \quad f(x) < f(y) \Leftrightarrow f(y) - f(x) \in \mathbb{R}^m_{>}.$$

A point $x \in X$ is called a weakly efficient point of problem (1) if there is no $y \in X$ such that $f(y) < f(x)$, in which case $f(x)$ is referred to as a weakly nondominated point [19]. A point $x \in X$ is called an efficient point of problem (1) if there is no $y \in X$ such that $f(y) \leq f(x)$ and $f(y) \neq f(x)$, and $f(x)$ is then referred to as a nondominated point [19]. Furthermore, $\bar{x}$ is a *local weakly efficient* solution of (1) if there exists $\varepsilon > 0$ such that there is no $x \in X \cap B(\bar{x}, \varepsilon)$ satisfying $f(x) < f(\bar{x})$. In this case, $f(\bar{x})$ is called a *local weakly nondominated* point.

Penalty methods convert the constrained MOP (1) into an unconstrained optimization problem by incorporating constraint violations into a merit (penalty) function. In this paper, we adopt a penalization framework and minimize a shifted exact penalty function [5, 36]. The proposed penalty function employs a violation measure of the constraint that is structurally similar to that used by Ansary and Panda [2]. Moreover,

$$\psi^i_\rho(x) := \rho\left(f_i(x) - \underline{f}_i\right) + v(x), \quad i = 1, \dots, m, \tag{2}$$

in which $\rho \geq 0$ is the penalty parameter, and

$$v(x) \ = \ \max\big\{0,\ g_j(x) \mid j = 1, \dots, r\big\},$$

is the violation measure of the constraint. Here, $v(x) = 0$ characterizes feasibility. Moreover, for each i, let $\underline{f}_i := \inf\{f_i(x) : x \in X\}$, which is finite under the boundedness assumption stated in subsection 4.1. Equivalently, $\underline{f}_i$ may be viewed as the infimum of f_i over the smallest convex set containing the iterates $\{x_k\}$ generated by the algorithm. The bound $\underline{f}_i$ is introduced only for theoretical purposes, namely to ensure that the line search procedure is well defined. Its existence follows from assumptions (A1)–(A2). In practice, we do not require $\underline{f}_i$ (or any approximation of it), and the proposed method is not sensitive to its value [36].

Define the *active set*

$$\mathcal{A}(x) \;=\; \{\, j : g_j(x) = v(x)\},$$

collect all constraints attaining the worst violation. Then the true directional derivative of v in any $d \in \mathbb{R}^n$ is

$$v'(x;d) \;=\; \max_{j\in\mathcal{A}(x)} \nabla g_j(x)^T d,$$

which is generally discontinuous as $\mathcal{A}(x)$ changes.

That is, although each g_j is continuously differentiable, the directional derivative $v'(x;d)$ is, in general, *discontinuous with respect to the base point x*. This discontinuity is caused by changes in the active set $\mathcal{A}(x)$ when the maximizer in

$$v(x) = \max\{\, 0,\ g_j(x) \mid j = 1,\dots,m\,\}$$

switches from one constraint to another. For fixed x, however, $v'(x;d)$ is continuous (indeed, piecewise linear) as a function of the direction d. For more clarification, we show this behavior in the following example:

Example 1. Let

$$g_1(x) = x, \qquad g_2(x) = -x, \qquad v(x) = \max\{0, g_1(x), g_2(x)\} = |x|.$$

Then

$$\mathcal{A}(x) = \begin{cases} \{1\}, & x > 0, \\ \{2\}, & x < 0, \\ \{1,2\}, & x = 0, \end{cases}$$

which changes at $x = 0$. Take the direction $d = 1$. For $x > 0$ we have $v'(x;1) = \nabla g_1(x)^T 1 = 1$; for $x < 0$, $v'(x;1) = \nabla g_2(x)^T 1 = -1$; and at $x = 0$,

$$v'(0;1) = \max\{\nabla g_1(0)^T 1, \nabla g_2(0)^T 1\} = \max\{1, -1\} = 1.$$

Thus,

$$v'(x;1) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0, \end{cases}$$

which is clearly discontinuous at $x = 0$. The discontinuity is caused by a switch of the maximizer from g_2 to g_1, that is, by a change in the active set $\mathcal{A}(x)$. Thus, we employ the continuous approximation proposed in [2, 4]:

$$v^*(x; d) = \max\Big\{ g_j(x) + \nabla g_j(x)^T d,\ 0 \ \Big|\ j \in \mathcal{A}(x) \Big\} - v(x).$$

Therefore, we approximate the directional derivative of ψ_ρ^i by the following continuous term:

$$\chi_\rho^i(x; d) = \rho \nabla f_i(x)^T d + v^*(x; d), \quad i = 1, \ldots, m. \tag{3}$$

The following theorem provides necessary optimality conditions for a (weakly) efficient point of problem (1).

Theorem 1 ([32]). Suppose that f_i, $i = 1, \ldots, m$, and g_j, $j = 1, \ldots, r$, are continuously differentiable at $x \in X$. If x is a weakly efficient point of problem (1), then there exist multipliers $\lambda \in \mathbb{R}^m_{\geq 0}$ and $\theta \in \mathbb{R}^r_{\geq 0}$ such that $(\lambda, \theta) \neq (0, 0)$ and

$$\begin{cases} \displaystyle\sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{j=1}^r \theta_j \nabla g_j(x) = 0, \\ \theta_j\, g_j(x) = 0, \qquad j = 1, \ldots, r. \end{cases} \tag{4}$$

The Fritz–John conditions in Theorem 1 are necessary for weak efficiency. In contrast, the Karush–Kuhn–Tucker conditions are necessary under an appropriate constraint qualification [32]. Therefore, in our convergence analysis, we assume that the Mangasarian–Fromovitz constraint qualification (MFCQ) holds at the limit points under consideration.

Definition 1 (Mangasarian–Fromovitz constraint qualification (MFCQ)). [31] The MFCQ is said to hold at a point $x \in \mathbb{R}^n$ if there exists a vector $z \in \mathbb{R}^n$ such that

$$\nabla g_j(x)^T z < 0, \quad \text{for all } j \in \mathcal{A}(x).$$

Definition 2. [2] A feasible point $x \in X$ of the problem (1) is said to be

1. a *strongly critical point* if there exist multipliers $\lambda \in \mathbb{R}^m_+ \setminus \{0\}$ and $\theta \in \mathbb{R}^r_+$ satisfying the conditions (4).

2. a *weakly critical point* if there exists an infeasible sequence $\{x_k\}$ with $x_k \to x$ such that

$$\lim_{k\to\infty} \frac{\displaystyle\max_{d\in D(x_k)} \max_{j=1,\ldots,r} \{ g_j(x_k) + \nabla g_j(x_k)^T d,\ 0\}}{v(x_k)} = 1,$$

where $D(x_k) = \{\, d \in \mathbb{R}^n :\ \nabla f_i(x_k)^T d \le 0,\ i = 1, \dots, m\}$.

Remark 1. The above definition of a *weakly critical point* is proposed because the penalty–SQP scheme may generate *infeasible* iterates, so we need a stationarity concept that remains meaningful outside the feasible set. Geometrically, the condition says that there is no direction d that improves (or at least does not worsen) all objectives while simultaneously decreasing violation measure of the constraint $v(x)$ according to the linearized constraints. Hence, at a weakly critical point the algorithm cannot make simultaneous progress toward feasibility and objective decrease, and such points naturally describe possible limit behavior of infeasible iterates.

More precisely, at an iterate x_k the set $D(x_k) = \{d \in \mathbb{R}^n : \nabla f_i(x_k)^T d \le 0,\ i = 1, \dots, m\}$ is a cone of *nonascending* directions for all objectives. For a given direction d, the quantity $\max_j \max\{g_j(x_k) + \nabla g_j(x_k)^T d, 0\}$ is the worst predicted violation under the linearized constraints. Therefore, the numerator $\max_{d\in D(x_k)} \max_j \max\{g_j(x_k) + \nabla g_j(x_k)^T d, 0\}$ represents the best achievable (i.e., smallest possible) worst-case linearized infeasibility among directions that do not worsen the objectives, while the denominator $v(x_k)$ is the current infeasibility. The limit being equal to 1 means that, asymptotically, even the best direction in $D(x_k)$ cannot yield a meaningful relative decrease in infeasibility: The predicted violation remains essentially the same as the current one.

3 Description of the algorithm

This section presents our nonmonotone multiobjective method in the SQP framework (NMOSQP). Our method is developed within the SQP framework of [2]; however, we introduce two essential modifications : (i) a new formulation of the (shifted) exact penalty function together with a different strategy for selecting and updating the penalty parameter, and (ii) a nonmonotone max-type line search rule applied to the penalty function. These changes require very additional technical justification, both in the global convergence analysis and in the numerical investigation.

More specifically, as in [2], NMOSQP computes a common improving direction by solving an auxiliary quadratic subproblem based on linear models of the objective and constraint functions, and then determines the step length by a line search procedure. The first main distinction from [2] lies in the treatment of the penalty parameter, while [2] employs a particular penalty-function structure and update rule, we adopt a different penalty-parameter update mechanism tailored to our nonmonotone line search. The second main distinction concerns globalization. Unlike [2], which uses a monotone line search, NMOSQP employs a nonmonotone max-type line search on the penalty function.

In [36], a descent direction is computed by minimizing a quadratic model of an exact penalty function. In each iteration, the method may require solving up to two inequality-constrained subproblems, while updates of the penalty parameter are governed by steering rules. In [37], an SQP-type method for constrained MOPs with nonconvex, nonsmooth, and locally Lipschitz objectives and constraints is proposed. The approach uses gradient sampling to handle nonsmoothness and employs an exact penalty merit function; its subproblem can be viewed as a nonsmooth counterpart of the one in [36].

We keep all linearized inequality constraints in subproblem (5), rather than restricting to the currently active set as in [33]. In nonconvex constrained MOPs the active set may change abruptly from one iterate to the next, and constraints that are inactive at x_k can become active after a step. Including all constraints improves robustness of the computed direction and helps prevent steps that move toward nearly active constraints and trigger excessive backtracking in the line search. Hence

$$\begin{aligned} \min_{d,z} \quad & z + \frac{1}{2}\|d\|^2, \\ \text{subject to:} \quad & (\nabla f_i(x))^T d \leq z, \quad \text{for all } i = 1, \ldots, m, \\ & g_j(x) + (\nabla g_j(x))^T d \leq z, \quad \text{for all } j = 1, \ldots, r, \\ & (d, z) \in \mathbb{R}^n \times \mathbb{R}. \end{aligned} \tag{5}$$

In this subproblem, z is as an upper bound for all linearized terms, and both z and d are variables. The improving direction is defined as the d-component of a minimizer of subproblem (5). The smooth quadratic subproblem (5) plays an crucial role in the analysis and implementation of the NMOSQP method. The satisfaction of MFCQ for the problem (1) ensures that the subproblem (5) also meets the MFCQ condition [2].

Inspired by the approach proposed in [2], the penalty parameter ρ_k is dynamically updated at each iteration to ensure that the direction d_k remains a descent direction for all penalized functions $\psi^i_{\rho_k}(x_k)$, $i = 1, \ldots, m$. If d_k is already a descent direction for all $\psi^i_{\rho_k}$, more precisely, if $\chi^i_{\rho_k}(x_k; d_k) < 0, i = 1, \ldots, m$, then the penalty parameter remains unchanged, that is, $\rho_{k+1} = \rho_k$. Otherwise, it is updated as

$$\rho_{k+1} = \min\left\{\theta_\rho \rho_k, \min_{i=1,\ldots,m} \frac{v^*(x_k; d_k) + \frac{1}{2} d_k^T d_k}{-\nabla f_i(x_k)^T d_k}\right\}, \tag{6}$$

where $v^*(x_k; d_k)$ is the smooth approximation of the directional derivative of the violation function $v(x)$, and the ratio ensures that d_k becomes a descent direction for each penalized objective $\psi^i_{\rho_k}$.

3.1 Nonmonotone line search

After obtaining the descent direction $d(x)$, the NMOSQP algorithm employs a backtracking line search to compute a suitable step length $\theta > 0$ along this direction. The monotone line search is guided by the Armijo rule, which ensures that the step length satisfies the sufficient decrease condition:

$$\psi^i_{\rho_{k+1}}\big(x_k + \theta_k d_k\big) \leq \psi^i_{\rho_{k+1}}(x_k) + \theta_k \gamma \chi^i_{\rho_{k+1}}(x_k; d_k), \tag{7}$$

for $i = 1, \ldots, m$, where $x_k \in X$ is not a critical point of problem (1), and $\gamma \in (0, 1)$ is a constant. This condition guarantees sufficient improvement in the objective functions, thereby promoting progress in the optimization process.

Nonmonotone line search strategies improve optimization performance by allowing temporary increases in the objective function. This adaptability proves especially effective in escaping narrow or curved valleys and in navigating complex, nonconvex regions of the search space [1, 11]. Such methods typically offer faster convergence and enhanced robustness, making them particularly suitable for solving large-scale and challenging optimization problems. In this paper, we propose a novel nonmonotone line search method for solving constrained MOPs. Our approach extends the classical nonmonotone rule of Grippo, Lampariello, and Lucidi [27], which generalizes the traditional Armijo rule by permitting occasional increases in the function value. This flexibility helps the algorithm escape narrow valleys or avoid excessively small steps in challenging nonlinear problems. Therefore,

$$\psi^i_{\rho_{k+1}}\big(x_k + \theta_k d_k\big) \leq \psi^i_{\rho_k,\mathfrak{r}}(x_k) + \theta_k \gamma \chi^i_{\rho_{k+1}}(x_k; d_k), \tag{8}$$

for $i = 1, \ldots, m$, where

$$\psi^i_{\rho_k,\mathfrak{r}}(x_k) := \max_{0 \leq j \leq m(k)} \left\{ \psi^i_{\rho_{k-j}}(x_{k-j}) \right\}, \qquad \text{for all } k \in \mathbb{N} \cup \{0\}, \tag{9}$$

$m(0) = 0$, $0 \leq m(k) \leq \min\{m(k-1)+1, N\}$, and $N \geq 0$ is an integer. Let $\mathfrak{r}(k)$ be integer such that $k - m(k) \leq \mathfrak{r}(k) \leq k$ and $\psi^i_{\rho_k,\mathfrak{r}}(x_k) = \psi^i_{\rho_{\mathfrak{r}(k)}}(x_{\mathfrak{r}(k)})$. In Lemma 3, we prove that for any $\eta \in (0, 1)$, inequality (8) is satisfied for sufficiently small $\theta > 0$. Now, we can propose NMOSQP algorithm.

3.2 NMOSQP: A Nonmonotone SQP-based algorithm for MOP

1. Choose $x_0 \in \mathbb{R}^n$, some scalars $r \in (0,1)$, $\gamma \in (0,1)$, the initial penalty parameter $\rho > 0$, and an error tolerance ϵ. Set $k := 0$.

2. Solve the quadratic program (5) to find the descent direction (z_k, d_k) with Lagrange multipliers λ_k, θ_k.

3. If $\|d_k\| < \epsilon$, then stop; otherwise, go to Step 4.

4. If $v(x_k) = 0$ or $\chi_{i,\rho_k}(x_k; d_k) < 0,$ for all i, then let $\rho_{k+1} = \rho_k$. Otherwise, update ρ_{k+1} using the rule in (6).

5. Compute the step length θ_k as the first number in the sequence $\{1, r, r^2, \ldots\}$, such that for all $i = 1, \ldots, m$, satisfying (8).

6. Update $x_{k+1} = x_k + \theta_k d_k$. Set $k := k + 1$ and go to Step 2.

The stopping criteria and line search process in Steps 3 and 5 in the NMOSQP algorithm rely on Lemmas 1 and 3.

4 Convergence analysis

4.1 Assumptions

The following assumptions are steadily held during our convergence analysis.

- **(A1)**: The functions $f_i,\ i = 1, 2, \ldots, p$, and $g_j,\ j = 1, 2, \ldots, m$, are twice continuously differentiable over the feasible set $X \subseteq \mathbb{R}^n$ and $\nabla f_j(x)$ and $\nabla g_i(x)$ are Lipschitz continuous for $i = 1, \ldots, m$ and $j = 1, \ldots, r$ with Lipschitz constant L.

- **(A2)**: The sequences $\{x_k\}$ and $\{(z_k, d_k)\}$, which correspond to the sequence of solutions for the subproblem (5), are bounded.

Lemma 1 establishes a connection between the algorithm's stopping criteria and the critical points of problem (1). It shows that $d(x) \neq 0$ implies x is not a critical point and it provides a descent direction of the penalty.

Lemma 1. [2] Let (z_k, d_k) be the solution of the subproblem (5). Then the following statements hold:

(i) The value of z_k satisfies
$$z_k \leq v(x_k) - \frac{1}{2} d_k^T d_k. \tag{10}$$

(ii) If $d_k = 0$ and MFCQ holds at x_k, then x_k is a strongly critical point of the problem (1).

(iii) If $d_k \neq 0$, then for sufficiently small ρ_k, then the vector d_k is a descent direction of the penalty function $\psi^i_{\rho_k}(x_k)$ at x_k.

Proof. Since the subproblem (5) is identical to that considered in [2], the proof follows similarly from [2, Lemma 1] and is therefore omitted here. □

Remark 2. The proof strategy in lemma 3 follows a standard line of argument used in related single-objective analyses (see, e.g., [41]); nevertheless, we present it here in a form tailored to our exact penalty function and the constrained MOP setting.

Lemma 2. During iteration k of the NMOSQP algorithm, we have

$$\psi^i_{\rho_{k+1}}(x_{k+1}) \leq \psi^i_{\rho_{k+1},\mathfrak{r}}(x_{k+1}) \leq \psi^i_{\rho_k,\mathfrak{r}}(x_k), \tag{11}$$

for all $k = 0, 1, 2, \ldots$. Moreover, the sequence $\psi^i_{\rho_k,\mathfrak{r}}(x_k)$ is convergent.

Proof. Let x_k be not a critical point of (1). The first inequality is obvious from (9). From (9) and $m(k+1) \leq m(k) + 1$ we have

$$\begin{aligned}
\psi^i_{\rho_{k+1},\mathfrak{r}}(x_{k+1}) &= \max_{0\leq j\leq m(k+1)} \left\{ \psi^i_{\rho_{k+1-j}}(x_{k+1-j}) \right\} \qquad (12)\\
&\leq \max_{0\leq j\leq m(k)+1} \left\{ \psi^i_{\rho_{k+1-j}}(x_{k+1-j}) \right\}\\
&= \max \left\{ \max_{0\leq j\leq m(k)} \left\{ \psi^i_{\rho_{k-j}}(x_{k-j}) \right\}, \psi^i_{\rho_{k+1}}(x_{k+1}) \right\}\\
&= \max \left\{ \psi^i_{\rho_k,\mathfrak{r}}(x_k), \psi^i_{\rho_{k+1}}(x_{k+1}) \right\}\\
&= \psi^i_{\rho_k,\mathfrak{r}}(x_k). \qquad (13)
\end{aligned}$$

From Step 4 of the NMOSQP algorithm, we have $\chi^i_{\rho_{k+1}}(x_k; d_k) < 0$ for all $i = 1, \ldots, m$. Therefore, the last equality is concluded from (8). This means the sequence $\psi^i_{\rho_k,\mathfrak{r}}(x_k)$ is nonincreasing for all $i = 1, \ldots, m$. Consequently, we obtain

$$0 \;\le\; \psi^i_{\rho_{k+n}}(x_{k+n}) \;\le\; \psi^i_{\rho_{k+n},\mathfrak{r}}(x_{k+n}) \;\le\; \cdots \;\le\; \psi^i_{\rho_k,\mathfrak{r}}(x_k). \tag{14}$$

Thus, the sequence $\{\psi^i_{\rho_k,\mathfrak{r}}(x_k)\}$ is monotonically non-increasing and bounded below by zero, which implies that it must converge. $\square$

Lemma 3. Suppose that (z_k, d_k) is an optimal solution for subproblem (5) at iteration k of the NMOSQP algorithm. If x_k is not a weakly or strongly critical point of problem (1), then there exists θ_k that satisfies the nonmonotone line search (8).

Proof. Since x_k is not a weakly or strongly critical point of problem (1), from Lemma 1 we have $d(x_k) \neq 0$. Under Assumption $(A1)$, it follows from [4, Lemma 2.1(3)] that there exists a constant $L > 0$ such that for $i = 1, \ldots, m$,

$$\psi^i_{\rho_{k+1}}(x_k + \theta d_k) \le \psi^i_{\rho_{k+1}}(x_k) + \theta \chi^i_{\rho_{k+1}}(x_k; d_k) + \frac{1}{2}(\rho_{k+1} + 1)Lt^2\|d_k\|^2, \tag{15}$$

for $\theta \in [0,1]$,and let $\gamma \in (0,1)$ be a constant as specified in Step 1 of NMOSQP algorithm. Since $\rho_{k+1} \le \rho_k$, and $\left(f_i(x_k) - \underline{f}\right) \ge 0$, from (2), we have

$$\rho_{k+1}\left(f_i(x_k) - \underline{f}\right) + v(x_k) \le \rho_k \left(f_i(x_k) - \underline{f}\right) + v(x_k), \quad i = 1, \ldots, m, \tag{16}$$

which means

$$\psi^i_{\rho_{k+1}}(x_k) \le \psi^i_{\rho_k}(x_k). \tag{17}$$

Then, by applying inequality (15), we have

$$\begin{aligned} &\psi^i_{\rho_{k+1}}(x_k + \theta d_k) - \psi^i_{\rho_k}(x_k) - \gamma\,\theta\,\chi^i_{\rho_{k+1}}(x_k; d_k) \\ &\le (1-\gamma)\,\theta\,\chi^i_{\rho_{k+1}}(x_k; d_k) + \tfrac{1}{2}\left(\rho_{k+1} + 1\right) L\,\theta^2\,\|d_k\|^2, \end{aligned} \tag{18}$$

for $i = 1, \ldots, m$. Step 4 of the NMOSQP algorithm implies

$$\chi^i_{\rho_{k+1}}(x_k; d_k) < 0,\; i = 1, \ldots, m.$$

Since $d_k \neq 0$, from (18), it follows that the right-hand side becomes negative for sufficiently small values of $\theta > 0$, ensuring that the monotone Armijo's rule (7) is satisfied.

Given that $\psi^i_{\rho_k}(x_k) \le \psi^i_{\rho_k,\mathfrak{r}}(x_k)$, condition (7) implies the validity of (8). Consequently, there exists a scalar $\theta_k > 0$ for which inequality (8) holds. $\square$

Lemma 4. Let (z_k, d_k) be the solution to the subproblem (5). If $x_k \to x^*$ as $k \to \infty$, then the sequence $\{(z_k, d_k)\}$ converges to (z^*, d^*), where (z^*, d^*) solves subproblem (5) at x^*. In particular, if $d_k \to 0$ and the MFCQ holds at every x_k, then x^* is a strong critical point of the problem (1).

Proof. Since the subproblem (5) is identical to that considered in [2], the proof follows directly from [2, Lemma 3] and is therefore omitted. □

Lemma 5. Suppose that $\rho_k = \rho > 0$ for all sufficiently large k. Assume that $\{x_k\}_{k\in K}$ is a convergent subsequence. Then $d_k \to 0$ as $k \to \infty$ and $k \in K$.

Proof. Without loss of generality, assume that $\rho_k = \rho > 0$ for all $k \in K$. If possible suppose that there exist an infinite subset $K' \subset K$ and a positive constant η such that

$$\|d_k\| \geq \eta, \quad k \in K'. \tag{19}$$

First, we will show that there exists $\delta > 0$ such that $\theta_k \geq \delta$ holds for every k, where θ_k is the step length obtained in Step 5 of NMOSQP algorithm. From this step either $\theta_k = 1$ or $\theta_k = r^{k_1}$ holds for some $k_1 \in \mathbb{N}$. If $\theta_k = r^{k_1}$ holds then there exists $i \in \{1,\dots,m\}$ satisfying,

$$\psi^i_\rho(x_k + r^{k_1-1}d_k) - \psi^i_{\rho,\mathfrak{r}}(x_k) > r^{k_1-1}\gamma\chi^i_\rho(x_k; d_k).$$

Then from (18),

$$r^{k_1-1}(1-\gamma)\chi^i_\rho(x_k; d_k) + \frac{1}{2}(1+\rho)Lr^{2(k_1-1)}\|d_k\|^2 > 0.$$

Thus,

$$\frac{1}{2}(1+\rho)Lr^{2(k_1-1)}\|d_k\|^2 \geq -r^{k_1-1}(1-\gamma)\chi^i_\rho(x_k; d_k). \tag{20}$$

If $\exists j \in \{1,\dots,r\}$ such that $g_j(x_k) + \nabla g_j(x_k)^T d_k > 0$, then from Lemma 1 and subproblem (5), we have

$$\max_{j=1,\dots,r}\{g_j(x_k) + \nabla g_j(x_k)^T d_k, 0\} - v(x_k) \leq v(x_k) - \frac{1}{2}d_k^T d_k - v(x_k) = -\frac{1}{2}d_k^T d_k. \tag{21}$$

If $g_j(x_k) + \nabla g_j(x_k)^T d_k \leq 0,$ for all $j = 1,\dots,r$, then

$$\max_{j=1,\dots,r}\{g_j(x_k) + \nabla g_j(x_k)^T d_k, 0\} - v(x_k) = -v(x_k). \tag{22}$$

Thus, by the definition of v^* and $\mathcal{A}(x) \subseteq \{1,\dots,r\}$, we have

$$v^*(x_k; d_k) \leq \max_{j=1,\dots,r}\{g_j(x_k) + \nabla g_j(x_k)^T d_k, 0\} - v(x_k) \leq \max\{-\frac{1}{2}d_k^T d_k, -v(x_k)\}. \tag{23}$$

The last inequality follows from (5), (21), and (22).

Since $(z_k, d_k) = (0, 0)$ is feasible for subproblem (5), we have $z_k + \frac{1}{2}{d_k}^T d_k \leq 0$. Then, (3) and (5) imply

$$\begin{aligned}\chi^i_\rho(x_k; d_k) &\leq \rho z_k + v^*(x_k; d_k)\\ &\leq \rho\big(-\tfrac{1}{2}\, d_k^T d_k\big) + \max\big\{-\tfrac{1}{2}\, d_k^T d_k,\ -v(x_k)\big\}\\ &\leq -\tfrac{\rho}{2}\, d_k^T d_k.\end{aligned}$$

From (20) we have

$$\frac{1}{2}(1+\rho)Lr^{k_1-1}\|d_k\|^2 \geq \frac{\rho}{2}(1-\gamma)\|d_k\|^2.$$

This implies $r^{k_1} \geq \frac{r(1-\gamma)\rho}{(1+\rho)L}$. Choose $\delta = \min\left\{1, \frac{r(1-\gamma)\rho}{(1+\rho)L}\right\}$. Then $\theta_k \geq \delta$ holds for every k. Now, from (7) and (20), we have

$$\begin{aligned}\psi^i_\rho(x_{k+1}) - \psi^i_\rho(x_0) &= \sum_{l=0}^{k}\big[\psi^i_\rho(x_{l+1}) - \psi^i_\rho(x_l)\big] < -\tfrac{\gamma\rho}{2}\sum_{l=0}^{k}\theta_l\|d_l\|^2\\ &\leq -\tfrac{1}{2}(k+1)\,\delta\,\gamma\,\rho\,\eta^2, \qquad \text{for all } k \in K'.\end{aligned}$$

This implies $\psi^i_\rho(x_k + \theta_k d_k) \to -\infty$ as $k \to \infty$ and $k \in K'$ (since $\theta_0\gamma\rho\eta^2 > 0$). This contradicts the assumption that $\{x_k\}$, $\{(z_k, d_k)\}$ are bounded as ψ^i_ρ is a continuous function. So there exists no any $K' \subset K$ and $\eta > 0$ such that (19) holds. $\square$

Lemma 6. If $\rho_k \to 0$, then $\lim_{k\to\infty} v(x_k) = 0$.

Proof. Suppose that $\lim_{k\to\infty} v(x_k)$ exists if $\rho_k \to 0$. By the definition of $v(x_k)$, we have $\lim_{k\to\infty} v(x_k) \geq 0$. Suppose $\lim_{k\to\infty} v(x_k) > 0$. Then there exists a constant $c_1 > 0$ such that for k large enough $v(x_k) > c_1$. We consider two cases separately.

Case 1: There exists a positive constant $c_2 > 0$ such that, for k large enough, $\|d_k\| > c_2$. Thus, from (23) and under assumptions $(A1)$ and $(A2)$, the following inequality holds for k large enough:

$$v^*(x_k; d_k) \leq \max\{-\frac{1}{2}d_k^T d_k, -v(x_k)\} \leq -c_3,$$

where $c_3 = \min\{\frac{1}{2}c_2^2, c_1\} > 0$. Then, from (3), (5), and Lemma 1, we have

$$\begin{aligned}\chi^i_{\rho_k}(x_k; d_k) &\leq \rho_k\, z_k + v^*(x_k; d_k)\\ &\leq \rho_k(v(x_k) - \tfrac{1}{2}d_k^T d_k) + \max\{-\frac{1}{2}d_k^T d_k, -v(x_k)\}.\end{aligned}$$

Thus, if $\rho_k \to 0$, then $\chi^i_{\rho_k}(x_k; d_k) \leq -c_3$. From Step 4 of the NMOSQP algorithm, this would imply that $\rho_{k+1} = \rho_k$ for all sufficiently large k, which contradicts the assumption that $\rho_k \to 0$.

Case 2: Case 1 does not hold. Thus there exists an infinite index set K such that $\{d_k\}_{k\in K}$ converges to zero. Since K is an infinite index set, by Assumption $(A2)$, $\{x_k\}_{k\in K}$ has at least one accumulation point. Let $K' \subseteq K$ such that $\{x_k\}_{k\in K'}$ converges to $\bar{x}$. By Lemma 4, $\bar{x}$ is a strong stationary point of (1). Hence, $v(\bar{x}) = 0$, which implies that $v(x_k) < c_1$ holds for $k \in K'$ large enough. This contradicts $v(x_k) > c_1$. □

Theorem 2. Let $\{x_k\}$ be a sequence generated by NMOSQP algorithm, and let the MFCQ be satisfied at every x_k. Then any accumulation point of $\{x_k\}$ is a critical point (either strongly or weakly critical point) of problem (1).

Proof. (i) Convergence to a strongly critical point:
Let K be an infinite index set such that $x_k \to x^*$ as $k \to \infty$ and $k \in K$. Let $\{(z_k, d_k)\}$ be the solution of subproblem (5). If $d_k \to 0$ as $k \to \infty$, then by Lemma 4, x^* is a strongly critical point of P.
(ii) Convergence to a weakly critical point:
If there exists a constant $c_0 > 0$ such that $\|d_k\| \geq c_0$ for large $k \in K$, then from Lemma 5, $\rho_k \to 0$ as $k \to \infty$. Since the sequences $\{x_k\}, \{(z_k, d_k)\}$ are bounded, so from Lemma 6, $\lim_{k\to\infty} v(x_k) = 0$. Hence from (10) and subproblem (5),

$$\nabla f_j(x_k)^T d_k \leq z_k < 0, \quad \text{for all } i = 1, \ldots, m.$$

This implies, $d_k \in D(x_k)$ for large k. We follow the proof by contradiction. Suppose that x_k is not a weakly critical point. From Definition 2, there exists a constant $\eta_2 > 0$ such that for sufficiently large k,

$$\max_{d\in D(x_k)} \max_{j=1,\ldots,r} \{g_j(x_k) + \nabla g_j(x_k)^T d; 0\} \leq v(x_k) - \eta_2.$$

Suppose that $\hat{d}_k$ maximizes $\max_{j=1,\ldots,r}\{g_j(x_k) + \nabla g_j(x_k)^T d; 0\}$. From (23) we have

$$v^*(x_k; d_k) \leq \max\bigl\{-\tfrac{1}{2}\, d_k^T d_k,\, -v(x_k)\bigr\} \leq \max_{d\in D(x_k)} \max\bigl\{-\tfrac{1}{2}\, d^T d,\, -v(x_k)\bigr\} \leq -\eta_2.$$

Then, from (3), (5), and Lemma 1, we have

$$\begin{aligned} \chi^i_{\rho_k}(x_k; d_k) &\leq \rho_k\, z_k + v^*(x_k; d_k) \\ &\leq \rho_k(v(x_k) - \tfrac{1}{2} d_k^T d_k) + \max\{-\frac{1}{2} d_k^T d_k, -v(x_k)\}. \end{aligned}$$

This means that if $\rho_k \to 0$, then $\chi^i_{\rho_k}(x_k; d_k) \leq -\eta_2$. From Step 4 of the NMOSQP algorithm, it follows that $\rho_{k+1} = \rho_k$ for all sufficiently large k, which contradicts that $\rho_k \to 0$. As a result, the sequence $\{x_k\}$ converges to a weakly critical point. Hence, any accumulation point of $\{x_k\}$ is either a strongly critical point or a weakly critical point. □

5 Numerical results

In this section, we describe the MATLAB R2021a implementation of NMOSQP and assess its performance on a diverse collection of standard test problems from the MOPs literature. The test set covers a broad range of features, including convex and nonconvex instances, biobjective and three-objective models, and both box-constrained and generally constrained MOPs. Table 1 summarizes the test problems: The first column reports the problem name; n denotes the decision-space dimension (set from 2 to 200); and n_f and n_g represent the numbers of objective functions and constraints, respectively. A dash in the n_g column of Table 1 indicates that the problem has no explicit constraints. The bound vectors x_L and x_U are also reported, and the last column provides the corresponding references.

Table 1: Details of box-constrained and general constrained test problems.

Name	n	nf	ng	x_L	x_U	Reference
ABC-comp	2	2	3	$[0,0]$	$[10,10]$	[28]
BNH	2	2	2	$[0,0]$	$[5,3]$	[12]
DTLZ0	3	3	2	$[0,0,0]$	$[2,1,1]$	[14]
Eichfelder2	2	2	1	$[0,0]$	$[100,100]$	[20]
Eichfelder6	3	3	1	$[0,0,0]$	$[100,100,100]$	[20]
Eichfelder7	3	3	1	$[0,0,1.2]$	$[\pi,100,100]$	[20]
FDSa	5	3	–	$-[2,\dots,2]$	$[2,\dots,2]$	[14]
FDSb	10	3	–	$-[2,\dots,2]$	$[2,\dots,2]$	[14]
FDSc	50	3	–	$-[2,\dots,2]$	$[2,\dots,2]$	[14]
FDSd	100	3	–	$-[2,\dots,2]$	$[2,\dots,2]$	[14]
Hanne2	2	2	1	$[0,0]$	$[10,10]$	[9]
Hanne4	2	2	1	$[0,0]$	$[10,10]$	[9]
HLL1	2	2	1	$[0,0]$	$[100,100]$	[28]
JOS1a	50	2	–	$-[2,\dots,2]$	$[2,\dots,2]$	[12]
JOS1c	200	2	–	$-[2,\dots,2]$	$[2,\dots,2]$	[20]
JOS1h	200	2	–	$-100[1,\dots,1]$	$100[1,\dots,1]$	[14]
P11	10	3	3	$-[2,\dots,2]$	$[2,\dots,2]$	[36]
P12	50	3	3	$-[2,\dots,2]$	$[2,\dots,2]$	[36]
P13	200	3	3	$-[2,\dots,2]$	$[2,\dots,2]$	[36]
TNK	2	2	2	$[0,0]$	$[\pi,\pi]$	[13]

To ensure that the sequence generated by NMOSQP remains inside the box constraints $x_L \le x \le x_U$, these constraints are added to subproblem (5), which leads to the following subproblem:

$$\min \ q(v,z) = z + \frac{1}{2}\|v\|^2,$$

$$
s.t. \begin{cases} (\nabla f_i(x))^T v \leq z, & for,\ i = 1, 2, \ldots, m, \\ g_j\,(x) + (\nabla g_j(x))^T v \leq z, & for\ j = 1, \ldots, r, \\ x_L - x \leq v \leq x_U - x, \\ (v, z) \in \mathbb{R}^n \times \mathbb{R}. \end{cases} \tag{24}
$$

In our implementation, we utilize subproblem (24) instead of (5) to ensure practical applicability. The stopping criterion is defined as $\|d_k\| \leq \varepsilon$, where $\varepsilon = 10^{-4}$. The Armijo rule parameter is set to $\gamma = 0.1$. To handle potential nonconvergence in nonconvex problems, we limit the maximum number of iterations to 500. We employ $N = 10$ in the nonmonotone term (9). At each iteration, we solve subproblem (24) using MATLAB's `quadprog` function.

Both MOSQP and NMOSQP start from arbitrary initial points, which we generate using MATLAB's `randn` function. To ensure a comprehensive evaluation, we generate 100 initial points for each method. Employing a multi-start strategy [21], we apply each algorithm from all initial points to approximate the nondominated frontier.

To comprehensively compare NMOSQP and MOSQP, we conduct two experiments, each focusing on a different performance aspect. In the first experiment, we examine the capability of NMOSQP to approximate the Pareto front and compare its results with those obtained by MOSQP as well as with the reference (true) Pareto front. Figure 1 shows the nondominated fronts generated by NMOSQP and MOSQP together with the reference front. For the Hanne2 problem, NMOSQP does not cover the entire front; however, it produces a dense and well-distributed approximation in the middle region, which is noticeably superior to MOSQP. In contrast, MOSQP performs poorly on Hanne2 and even generates dominated points. For the ABC-comp and BNH problems, both methods exhibit similar and satisfactory behavior relative to the reference front. In the TNK problem, NMOSQP generates a larger nondominated portion of the front than MOSQP; nevertheless, the two approximations are complementary, and their union covers a broader part of the true front, yielding a more well-distributed overall approximation.

In the second experiment, we use the Dolan and Moré performance profiles [15] to compare NMOSQP and MOSQP across the specified performance criteria.

The performance profile is a more reliable visualization tool for comparing the results obtained by different solvers, especially when the number of problems and solvers increases. Following is a simplified description of the performance profile of a specific criterion for a solver associated with a problem [15]. Consider $t_{p,s}$ as the performance of the solver $s \in S$ in solving the problem $p \in P$. Accordingly, the performance ratio $r_{p,s}$ is defined as $r_{p,s} := t_{p,s}/\min\{t_{p,s} : s \in S\}$. Based on this definition, the probability of the performance ratio of the solver $s \in S$ to be within the interval $[1, \tau)$ is denoted by $\rho_s(\tau)$ and defined as follows:

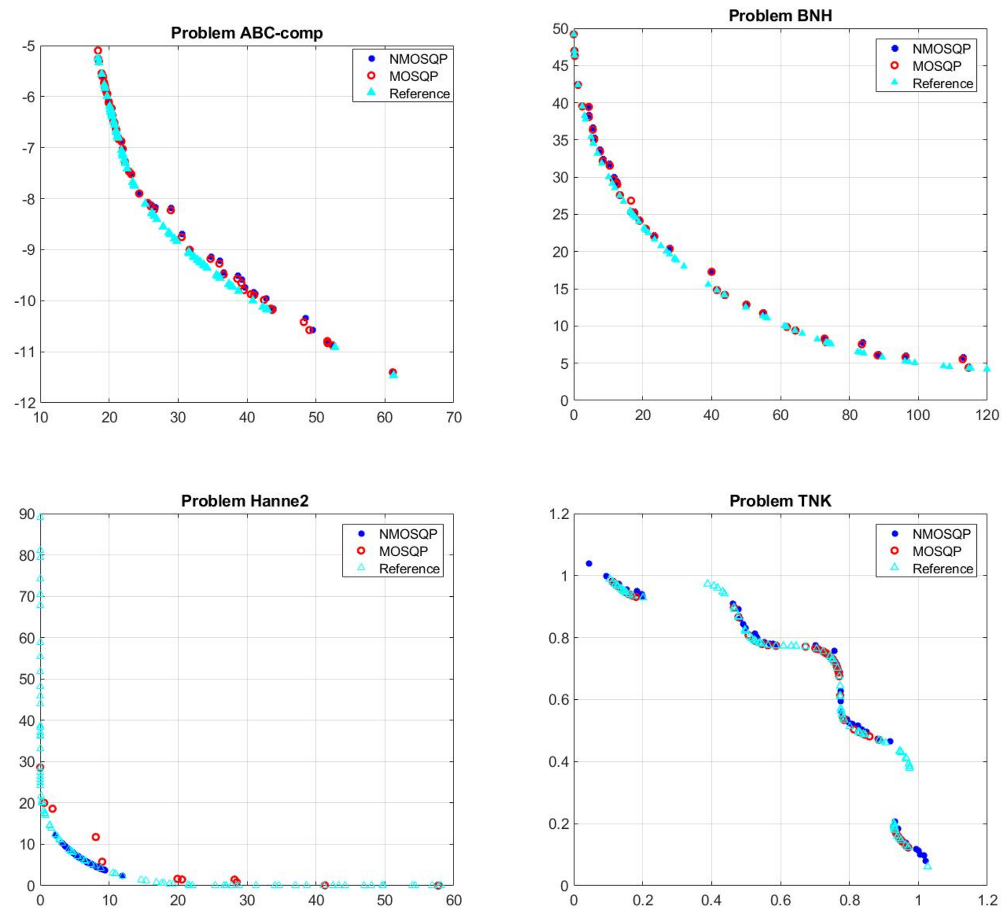

Figure 1: Nondominated fronts generated by NMOSQP and MOSQP on some test problems.

$$\rho_s(\tau) := \frac{1}{|P|} \left| p \in P : r_{p,s} \leq \tau \right|,$$

where ρ_s is the cumulative distribution function for the performance ratio [15]. The performance profiles depict the function $\rho_s(\tau)$ for each solver $s \in S$. According to the above explanation about $\rho_s(\tau)$, the value of $\rho_s(1)$ is the probability that the solver s will overcome the remaining ones. Thus, $\rho_s(1)$ shows the win percentage of solver s compared to the other solvers.

Our evaluation is based on two complementary aspects: (i) computational effort and (ii) quality of the produced Pareto front approximation. The criteria related to computational effort (and hence the speed of convergence) include the number of iterations, the numbers of objective and constraint function evaluations, the numbers of gradient evaluations for both objectives and constraints, and the wall-clock time. For multiobjective methods, however, it is equally important to assess the quality of the approximated Pareto front. Therefore, in addition to the above measures, we report the number of generated nondominated points and three widely used multiobjective indicators: The epsilon indicator [29], the spread [3, 10], and the hyper-volume (HV) [44]. The epsilon indicator provides a measure of how far the computed front is from the reference Pareto front, the spread evaluates how evenly the points are distributed along the front, and HV quantifies the portion of the objective space dominated (covered) by the obtained nondominated set (larger HV indicates a better approximation).

In this study, the HV indicator is computed using a common reference point selected separately for each test problem. For a given problem, we first form the union of all solutions generated by all algorithms over all runs. Then, for each objective, we determine the best observed value (the smallest one, since all problems are minimization) and the worst observed value (the largest one); the difference between these values defines the observed objective range. The HV reference point is chosen slightly worse than the worst observed value by adding 5% of this range to each objective component. This choice ensures that the reference point is dominated by all observed solutions and, consequently, makes the HV values fair and comparable across algorithms. With this setting, larger HV values indicate better performance. Finally, because Dolan and Moré performance profiles are typically constructed for measures where smaller values are better, we transform HV and the number of nondominated points (both larger-is-better criteria) into smaller-is-better quantities before plotting, so that algorithms achieving larger HV or producing more nondominated points are correctly identified as superior in the performance profiles.

Figure 2 compares the two methods with respect to the computational criteria. It contains four subplots reporting the number of iterations, wall-clock time, the number of objective function evaluations, and the number of constraint function evaluations.

Figure 3 summarizes the quality of the obtained Pareto front approximations by reporting the number of generated nondominated points for both methods, epsilon indicator, spread, and the HV metrics.

We provide a detailed comparison of NMOSQP and MOSQP using the Dolan and Moré performance profiles across the selected computational and quality criteria. For the *number of iterations* and *gradient evaluations*, NMOSQP clearly dominates and is the best performer across all test instances. Regarding *function evaluations*, the two methods are competitive on roughly 80% of the test instances; nevertheless, NMOSQP attains the best performance more frequently and shows an overall advantage. A similar behavior is observed for *wall-clock time*: Although the methods compete in about 80% of the instances, NMOSQP achieves the best performance on approximately 95% of the test problems, indicating a clear improvement in

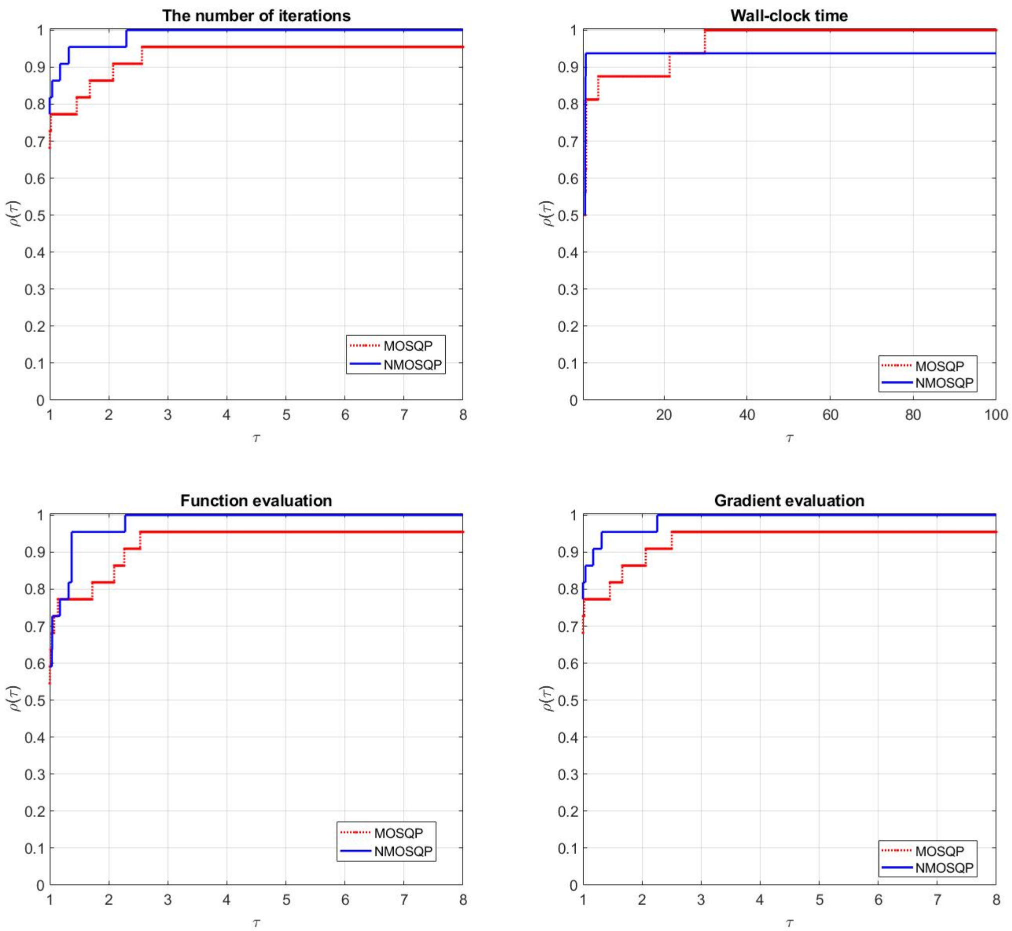

Figure 2: Iterations, Function evaluations, Gradient evaluations, and Wall-clock time performance profiles.

practical efficiency. Turning to the quality-related performance profiles, NMOSQP and MOSQP are largely competitive in terms of *front generation* and the *epsilon indicator*. For the *spread* metric, the methods still compete in about 85% of the instances, but NMOSQP more often ranks first and thus provides a better overall distribution of points along the front. Finally, the *HV* profile shows a significant superiority of NMOSQP over MOSQP, indicating that NMOSQP produces nondominated sets that dominate a larger portion of the objective space and hence yield higher-quality Pareto-front approximations.

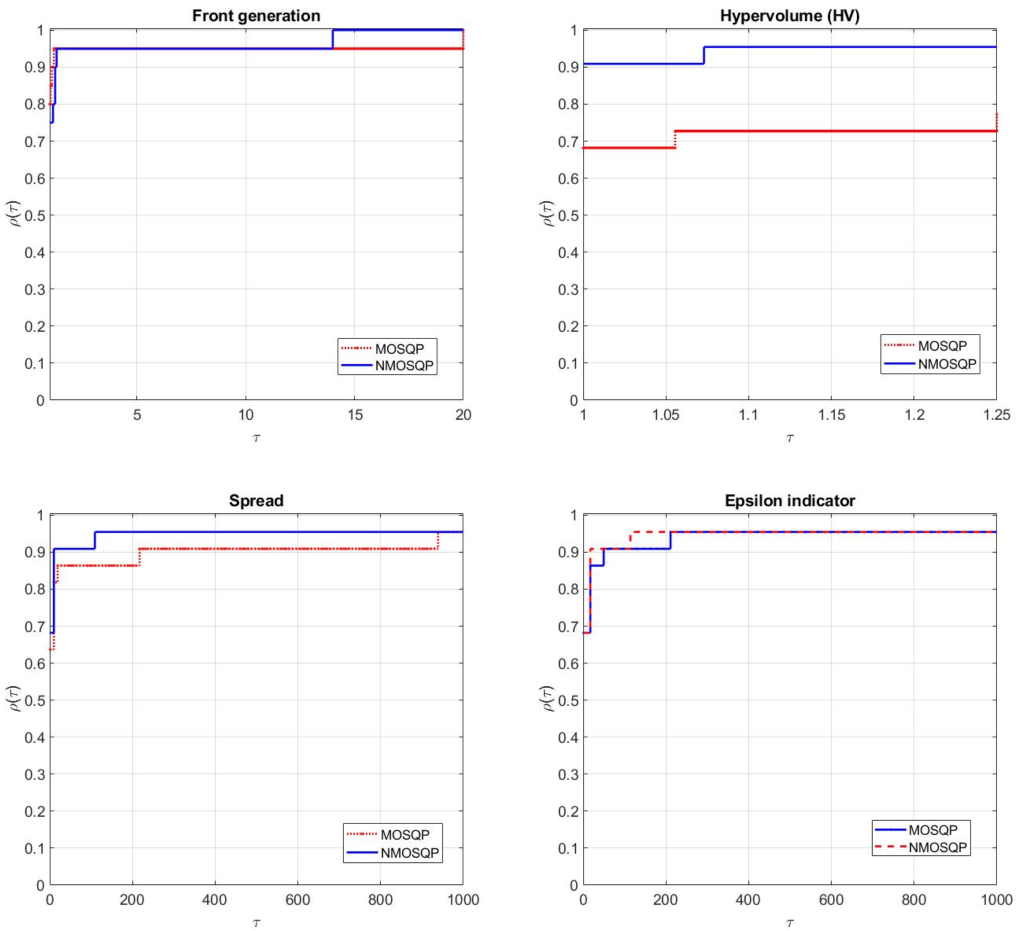

Figure 3: Number of nondominated points, Hypervolume(HV), Spread, and Epsilon indicator performance profiles.

6 Conclusion

This paper proposed a nonmonotone penalty method in the SQP framework for solving general inequality-constrained MOPs. In contrast to standard monotone penalty-based schemes, which enforce a sufficient decrease of the penalty function at every iteration, the proposed approach employs a nonmonotone max-type line search that allows controlled temporary increases of the penalty function. This additional flexibility can prevent overly conservative steps and improve practical efficiency on difficult constrained instances. We compare the proposed nonmonotone method (NMOSQP) with the monotone baseline (MOSQP). Dolan and Moré performance profiles indicate that NMOSQP is consistently faster and more efficient across the considered test set. Moreover, under mild regularity assumptions, we establish global convergence results showing that, for any starting point, the sequence generated by NMOSQP converges to a weakly or strongly critical point. Finally, as a direction for future research, it is of interest to investigate alternative nonmonotone line search rules within the proposed penalty-SQP framework to further enhance performance.

Acknowledgements

The authors are grateful to the anonymous referees and the Editor for their constructive comments.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Amini, K. and Rashidi, M. On determining radius in nonmonotone trust-region approaches, J. Math. Modeling, 11(3) (2023), 507–526.

[2] Ansary, M.A.T. and Panda, G. A sequential quadratic programming method for constrained multiobjective optimization problems, J. Appl. Math. Comput. 64(1) (2020), 379–397.

[3] Bandyopadhyay, S., Pal, S.K. and Aruna, B. Multiobjective GAs, quantitative indices, and pattern classification, IEEE Trans. Syst. Man Cybern. B Cybern. 34(5) (2004), 2088–2099.

[4] Bazaraa, M.S. and Goode, J.J. An algorithm for solving linearly constrained minimax problems, Eur. J. Oper. Res. 11(2) (1982), 158–166.

[5] Burke, J.V., Curtis, F.E. and Wang, H. A sequential quadratic optimization algorithm with rapid infeasibility detection, SIAM J. Optim. 24(2) (2014), 839–872.

[6] Carrizo, G.A., Fazzio, N.S. and Schuverdt, M.L. A nonmonotone projected gradient method for multiobjective problems on convex sets, J. Oper. Res. Soc. China, 12(2) (2024), 410–427.

[7] Chen, J. Liu, J. Qin, X. and Yao, J.-C. A nonmonotone proximal gradient algorithm for solving nonsmooth multiobjective optimization problems with an extending application to robust multiobjective optimization, J. Comput. Appl. Math. 460 (2025).

[8] Chen, W. Yang, X. and Zhao, Y. Conditional gradient method for vector optimization, Comput. Optim. Appl. 85(3) (2023), 857–896.

[9] Collette, Y. and Siarry, P. Multiobjective optimization: Principles and case studies, Springer Science & Business Media, 2003.

[10] Custódio, A.L., Madeira, J.F., Vaz, A.I. and Vicente, L. N. Direct multisearch for multiobjective optimization, SIAM J. Optim. 21 (2011), 1109–1140.

[11] Dai, Y.-H. On the Nonmonotone Line Search, J. Optim. Theory Appl. 112 (2002), 315–330.

[12] Deb, K. Multiobjective optimization using evolutionary algorithms, John Wiley & Sons, 2001.

[13] Deb, K., Pratap, A. and Meyarivan, T. Constrained test problems for multiobjective evolutionary optimization, in Evolutionary Multi-Criterion Optimization, Springer, Berlin, Heidelberg, 2001, 284–298.

[14] Deb, K., Thiele, L., Laumanns, M. and Zitzler, E. Scalable test problems for evolutionary multiobjective optimization, Springer, London, 2005.

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

[16] Dolatnezhadsomarin, A. and Khorram, E. Two efficient algorithms for constructing almost even approximations of the Pareto front in multiobjective optimization problems, Eng. Optim. 51(4) (2019), 567–589.

[17] Drummond, L.M.G. A projected gradient method for vector optimization problems, Comput. Optim. Appl. 28 (2004), 5–29.

[18] Drummond, L.M.G. and Svaiter, B.F. A steepest descent method for vector optimization, J. Comput. Appl. Math. 175 (2005), 395–414.

[19] Ehrgott, M. Multicriteria optimization, Springer, Berlin, 2005.

[20] Eichfelder, G. Adaptive scalarization methods in multiobjective optimization, Springer, Berlin, 2008.

[21] Fliege, J., Drummond, L.M.G. and Svaiter, B.F. Newton's method for multiobjective optimization, SIAM J. Optim. 20(2) (2009), 602–626.

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

[23] Fliege, J. and Vaz, A.I.F. A method for constrained multiobjective optimization based on SQP techniques, SIAM J. Optim. 26(4) (2016), 2091–2119.

[24] Gebken, B., Peitz, S. and Dellnitz, M. A descent method for equality and inequality constrained multiobjective optimization problems, in Numerical and Evolutionary Optimization – NEO 2017, Springer International Publishing, 2018, 29–61.

[25] Ghalavand, N., Khorram, E. and Morovati, V. Two adaptive nonmonotone trust-region algorithms for solving multiobjective optimization problems, Optimization, 73(9) (2023), 2953–2985.

[26] Gonçalves, D.S., Gonçalves, M.L.N. and Melo, J.G. An away-step Frank–Wolfe algorithm for constrained multiobjective optimization, Comput. Optim. Appl. 88(3) (2024), 759–781.

[27] Grippo, L., Lampariello, F. and Lucidi, S. A nonmonotone line search technique for Newton's method, SIAM J. Numer. Anal. 23 (1986), 707–716.

[28] Hwang, C.-L. and Masud, A.S.M. Multiple objective decision making—Methods and applications, Lecture Notes in Economics and Mathematical Systems, vol. 164, Springer, Berlin, 1979.

[29] Knowles, J. Thiele, L. and Zitzler, E. A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers, TIK Report 214, Computer Engineering and Networks Laboratory, ETH Zurich, 2006.

[30] Mahdavi-Amiri, N. and Salehi Sadaghiani, F. A superlinearly convergent nonmonotone quasi-Newton method for unconstrained multiobjective optimization, Optim. Methods Softw. 35(6) (2020), 1223–1247.

[31] Mangasarian, O.L. and Fromovitz, S. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints, J. Math. Anal. Appl. 17(1) (1967), 37–47.

[32] Miettinen, K. Nonlinear multiobjective optimization, International Series in Operations Research and Management Science, vol. 12, Springer Science and Business Media, 2012.

[33] Morovati, V. and Pourkarimi, L. Extension of Zoutendijk method for solving constrained multiobjective optimization problems, Eur. J. Oper. Res. 273(1) (2019), 44–57.

[34] Pinheiro, M.E. and Grapiglia, G.N. Universal nonmonotone line search method for nonconvex multiobjective optimization problems with convex constraints, Comput. Appl. Math. 44(1) (2025), 56.

[35] Pirouz, B. and Khorram, E. A computational approach based on the ε-constraint method in multiobjective optimization problems, Adv. Appl. Stat. 49(6) (2016), 453–483.

[36] Rashidi, M., Khorram, E. and Soleimani-Damaneh, M. An exact penalty method with nonmonotone line search and rapid infeasibility detection for constrained multiobjective optimization: Application in supervised machine learning, Comput. Oper. Res. 188 (2026), 107351.

[37] Rashidi, M. and Soleimani-damaneh, M. MultiSQP-GS: a sequential quadratic programming algorithm via gradient sampling for nonsmooth constrained multiobjective optimization, Comput. Optim. Appl. 89 (2024), 729–767.

[38] Tanabe, H., Fukuda, E.H., and Yamashita, N. An accelerated proximal gradient method for multiobjective optimization, Comput. Optim. Appl. 86(2) (2023), 421–455.

[39] Upadhayay, A., Ghosh, D., Jauny, Yao, J.C. and Zhao, X. A nonmonotone conditional gradient method for multiobjective optimization problems, Soft Comput. 28(17-18) (2024), 9609–9630.

[40] Upadhayay, A., Ghosh, D. and Kumar, K. Nonmonotone Wolfe-type quasi-Newton methods for multiobjective optimization problems, Optimization, (2025), 1–33.

[41] Zhang, H.C. and Hager, W.W. A nonmonotone line search technique for unconstrained optimization, SIAM J. Optim. 14(4) (2004), 1043–1056.

[42] Zhao, X., Raushan, R., Ghosh, D., Yao, J.-C. and Qi, M. Proximal gradient method for convex multiobjective optimization problems without Lipschitz continuous gradients, Comput. Optim. Appl. 91(1) (2025), 27–66.

[43] Zhao, X. and Yao, J.-C. Linear convergence of a nonmonotone projected gradient method for multiobjective optimization, J. Global Optim. 82(3) (2022), 577–594.

[44] Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M. and da Fonseca, V.G. Performance assessment of multiobjective optimizers: an analysis and review, IEEE Trans. Evol. Comput. 7(2) (2003), 117–132.