Chapter 5

A Nonmonotone Conditional Gradient Method for MOPs

5.1 Introduction

The conditional gradient method is used for minimizing differentiable functions on compact convex sets. Its evolution dates back to 1956. A decade later, the technique was further developed to address the minimization of differentiable convex functions over compact convex feasible sets. This extension specifically considers the case where the gradients of these functions exhibit Lipschitz properties [94]. Scientists involved in this field have been intrigued by the method since then. An important factor underlying this interest is its simplicity and straightforward implementation: the method uses a linear minimization oracle to perform linear minimization on a compact convex set at every iteration. Particularly, the low storage cost and ability to take advantage of both separability and sparsity make conditional gradient methods very attractive for application to large-scale problems. It is noteworthy that the recognition of this method has increased in recent years because of its successful use in machine learning applications [78, 89, 91]. That has paved the way for the development of multiple alternative versions of the method and the discovery of various properties of it over time, leading to extensive literature on the subject. This method has been discussed in several papers, for example, [11, 16, 55, 59, 72, 88, 92, 101].

Recently, Assunção et al. [8] proposed the conditional gradient method for constrained MOPs under the compactness and convexity assumption of the constraint set, as well as continuous differentiability of the objective functions. This work demonstrates its asymptotic convergence using the Armijo step size condition. Also, for vector optimization problems, Chen et al. [19] has introduced a conditional gradient methodology that incorporates three-step size strategies, namely adaptive, Armijo, and nonmonotone.

In the past, several methods have been equipped with an Armijo step size rule. In particular, Fliege and Svaiter [49] and Fliege et al. [50] proposed the steepest descent method and Newton's method, respectively, for unconstrained optimization problems. Both techniques allow us to choose the appropriate starting points for finding Pareto optimal solutions. Furthermore, the techniques described in [49, 50] employ an iterative approach, where search directions are determined by solving strongly convex subproblems during each iteration. Additionally, step sizes are selected to satisfy the condition of Armijo. Each iteration results in a decrease in all objective function values. However, as the quantity of objective functions grows, the Armijo-type condition becomes increasingly stringent, potentially causing smaller stepsizes [111].

5.2 Motivation

The choice of the conditional gradient method over sequential programming or the projected gradient method in my research is driven by several key advantages. The conditional gradient method, is particularly well-suited for problems with large-scale or complex feasible regions, as it avoids the need for explicit projections onto the feasible set, which can be computationally expensive or challenging to implement [78]. This method efficiently handles constraints by iteratively moving toward a solution along feasible directions, making it more scalable and practical for high-dimensional optimization problems [54]. Additionally, the conditional gradient method provides flexibility in handling convex and nonconvex problems, which is crucial in my research context, where the problem structure demands an approach that can adapt to varying levels of complexity while maintaining computational efficiency. These advantages make the conditional gradient method a more appropriate choice for achieving accurate and computationally feasible solutions in my specific research domain.

It is well known that in the case of singleobjective optimization, enforcing the function's monotonicity will make the technique take longer time to converge, particularly when the iteration resides at the bottom of a narrow, curved valley [111]. To improve convergence speed, some increases in the function values can be permitted in some iterations during the line search. The strategy has been extensively studied for singleobjective optimization under the name of *nonmonotone* line search. In order to acquire global convergence, the current objective function value must be smaller than a certain benchmark. For instance, Grippo et al. [66] introduced a nonmonotone line search where the maximum of recent objective function values is taken as a benchmark. An average-type nonmonotone procedure was instead considered by Zhang and Hager [156], which was more efficient in practice. Even though nonmonotone techniques are slower for complex problems when there are numerous optimal points, they are beneficial since they can improve the convergence speed of many problems, and they may also be easier to implement. However, nonmonotone line searches are still a rather recent and insufficient research topic in multiobjective descent methods. In the multiobjective setting, nonmonotone line search algorithms were previously considered, for example, in [19, 43, 111, 130, 145].

5.3 Contributions

The following points summarize the contributions of the present chapter:

- The present study draws inspiration from prior research done in [8]. However, rather than utilizing the Armijo line search technique described in [8], we employed more efficient nonmonotone line search technique found in [156].
- We established the asymptotic convergence of the multiobjective conditional gradient method equipped with an average-type nonmonotone line search condition.
- Based on the proposed method, we have shown that any limit point generated by a sequence of iterates is a Pareto critical point.
- A complexity bound for the proposed technique has been established.
- In addition to the theoretical results, numerical experiments have been conducted for convex and non-convex MOPs to highlight the effectiveness of this approach in this novel scenario.
- We have measured the efficacy of the proposed technique by comparing it with the conditional gradient method with Armijo step size [19], the conditional gradient method with nonmonotone line search [19], the Hager-Zhang conjugate gradient method [63] and the steepest descent method [49] for MOPs.
- To check the ability of the proposed algorithm to generate approximations to the complete Pareto front, we compare it with some existing popular solvers based on inverted generational distance (IGD) and hypervolume (HV) indicators for ZDT and DLTZ benchmark suit.

The following points outline the distinctions between the current study and the work conducted by [19].

- The present chapter offers a valuable contribution by providing an iteration complexity bound for the proposed technique, along with demonstrating global convergence. Notably, these aspects were not addressed in the work conducted by [19].

- Our study provides detailed insights into the efficacy of the proposed technique. To assess the capability of the proposed technique in generating approximations to the complete Pareto front, we conduct comparisons with several widely-used solvers based on IGD and HV indicators using the ZDT and DLTZ benchmark suites. Moreover, we further evaluate the performance profile of the proposed method in comparison to the Hager-Zhang conjugate gradient method, the steepest descent method, the monotone conditional gradient method, and a nonomonotone conditional gradient method for MOPs. By conducting these analyses and comparisons, we aim to provide a comprehensive understanding of the proposed technique's performance and its effectiveness in solving MOPs.

In this chapter, our goal is to find all Pareto critical points of the following MOP:

$$\min_{x\in\Omega} F(x) = \big(f_1(x), f_2(x), \ldots, f_r(x)\big)^\top, \tag{5.1}$$

where Ω is a convex compact subset of $\mathbb{R}^n$, $x \in \Omega \subset \mathbb{R}^n$, and $F : \mathbb{R}^n \to \mathbb{R}^r$. We assume that $f_i : \mathbb{R}^n \to \mathbb{R}$ for each $i \in \mathcal{I}$ is bounded from below and continuously differentiable.

Definition 5.1 *The row-wise Lipschitz continuity of the Jacobian JH of the function $F(x)$ is established when there are constants $L_i > 0$ for all $i \in \mathcal{I}$, and it satisfies*

$$\|\nabla f_i(x) - \nabla f_i(y)\|_2 \leq L_i \|x - y\|_2, \qquad \textit{for all } x, y \in \Omega, \qquad \forall i \in \mathcal{I}.$$

To facilitate future references, we use a notation $L = \max\{L_i : i \in \mathcal{I}\}$.

Now, we state a lemma that provides a result related to vector functions following Definition 5.1.

Lemma 5.1 *(See [8]) Assume that JF of the function $F(x)$ follows Definition 5.1.*

Then, for all $x, s \in \Omega$ and $\alpha \in (0, 1]$, the following will always hold

$$F(x + \alpha(s - x)) \preceq F(x) + \big(\mathcal{M}\alpha + \tfrac{L}{2}\|s - x\|_2^2\alpha^2\big)e,$$

where $\mathcal{M} = \max_{i\in\mathcal{I}} \langle \nabla f_i(x), s - x\rangle$, and $e = (1, 1, \dots, 1)^\top \in \mathbb{R}^r$.

The first-order optimality condition for the problem (5.1) can be stated as follows [8]

$$-(\mathbb{R}^r_{++}) \cap JF(\hat{x})(\Omega - \hat{x}) = \emptyset, \tag{5.2}$$

where $\hat{x} \in \Omega$, and $\Omega - \hat{x} = \{v - \hat{x} \ : \ v \in \Omega\}$.

A point $\hat{x} \in \mathbb{R}^n$ is referred to as a *Pareto critical* or *stationary* point of (5.1) if it follows the condition (5.2). Thus, if x is not Pareto critical point, there exists $w \in \Omega$ such that $JF(x)(w - x) \in -(\mathbb{R}^r_{++})$. Each vector $w - x$ meeting this criterion qualifies as a *descent direction* for F at x. In other words, there exists $\epsilon > 0$ for which $F(x + \alpha(w - x)) \prec F(x)$ for any $\alpha \in (0, \epsilon]$.

The Pareto critical or stationary condition (5.2) can be equivalently [8] expressed by

$$\max_{i\in\mathcal{I}} \langle \nabla f_i(\hat{x}), v - \hat{x}\rangle \geq 0, \qquad \forall v \in \Omega.$$

It is important to note that while condition (5.2) is necessary, it alone does not guarantee optimality [8].

The following lemma demonstrates the equivalence between stationarity and weak Pareto optimality in convex cases. This relationship is widely acknowledged in the literature (refer to [38] for further details), and therefore, we will not delve into the proof in this context.

Lemma 5.2 *(See [38]). If F is convex and $\hat{x}$ is a Pareto critical point, then $\hat{x}$ is also a weak Pareto optimal point of Problem* (5.1)*.*

5.4 The Conditional Gradient Method

In the conditional gradient method, the search direction $d(x)$ is defined by $d(x) = s(x) - x$, where $x \in \Omega$ and $s(x)$ represents the optimal solution to the following scalar-valued problem (see [8]):

$$\min_{v \in \Omega} \left(\max_{i \in \mathcal{I}} \langle \nabla f_i(x), v - x \rangle \right). \tag{5.3}$$

Thus,

$$s(x) \in \operatorname*{argmin}_{v \in \Omega} \max_{i \in \mathcal{I}} \langle \nabla f_i(x), v - x \rangle. \tag{5.4}$$

We denote the optimal value of (5.3) by

$$\mathcal{M}(x) = \max_{i \in \mathcal{I}} \langle \nabla f_i(x), s(x) - x \rangle, \tag{5.5}$$

where $s(x) \in \Omega$, is as given in (5.4).

One can notice that the objective function in the equation (5.3) exhibits convexity, while Ω represents a compact convex set. As a result, the problem (5.3) possesses a solution that optimizes the objective function, thereby ensuring the well-definedness of $s(x)$. Although it is a fact that (5.3) is a constrained convex nondifferentiable problem, one can determine its solution by solving the following smooth problem:

$$\left.\begin{array}{ll} \min & \eta \\ \text{subject to} & \langle \nabla f_i(x), v - x \rangle \leq \eta, \quad i \in \mathcal{I} \\ & v \in \Omega. \end{array}\right\}$$

The next result shows the characterization of $\mathcal{M}(x)$ with respect to the stationary point of (5.1).

Proposition 5.1 *(See Proposition 5 of [8]). Let the mapping $\mathcal{M} : \Omega \to \mathbb{R}$ be as given in* (5.5)*. Then, the following is true:*

(i) for all $x \in \Omega$, $\mathcal{M}(x) \leq 0$,

(ii) $\mathcal{M}(x)$ is continuous for all $x \in \Omega$, and

(iii) $\bar{x} \in \Omega$ is a Pareto critical point of (5.1) *iff $\mathcal{M}(\bar{x}) = 0$.*

Proposition 5.1 yields a straightforward implication: if $x' \in \Omega$ and is not a Pareto critical point of problem (5.1), then $\mathcal{M}(x') < 0$ and $s(x') \neq x'$. Consequently, in this scenario, $d(x') = s(x') - x'$ is nonzero and serves as a descent direction for H at x. This implies that for all $i \in \mathcal{I}$, $\langle \nabla f_i(x), s(x) - x \rangle \leq \mathcal{M}(x) < 0$.

5.5 Nonmonotone Line Search

Within this section, we provide a concise explanation of two nonmonotone line search procedures for the MOP (5.1). Numerous research studies have demonstrated that the utilization of nonmonotone schemes can significantly enhance the performance of descent methods, particularly in scalar cases (see [43] and the references therein).

In the conventional monotone line search technique, we choose α_k in a manner that ensures $F(x^{k+1}) \prec F(x^k)$. It means that each iteration results in a lower objective function value. In nonmonotone line search, a certain amount of growth is allowed in the objective function values, i.e., we choose $\alpha_k > 0$, which follows

$$F(x^k + \alpha_k(s(x^k) - x^k)) \preceq C^k + \sigma \alpha_k \mathcal{M}(x^k) e \tag{5.6}$$

with $\sigma \in (0, 1)$ and $C^k \succeq F(x^k)$.

Grippo et al. [66] introduced a nonmonotone line search framework for Newton's method, which is based on taking the maximum of recent objective function values of the previous iterations. Qu et al. [130] proposed two nonmonotone gradient methods formulated on the max-type nonmonotone line search. The first algorithm of [130] can be considered as an extension of the steepest gradient previously investigated by Fliege and Svaiter [49]. The second algorithm of [130] can be seen as an extension of one projected gradient method introduced by Cruz et al. [24]. Mathematically, there is no difference between the two nonmonotone strategies of [130] and [66]. In both strategies, we evaluate the maximum of recent objective function values. In the current investigation, we employ the approach proposed by Zhang and Hager [156] due to its recognized efficiency, particularly in the context of scalar cases. Rather than using the maximum of recent objective function values, it takes the average of the objective function values.

The following Algorithm 7 formalizes an *average-type nonmonotone* conditional gradient method for the MOP (5.1).

Algorithm 7 Conditional gradient method to generate the complete set of Pareto critical points of (5.1)

Aim: To generate a discrete approximation to the complete set of Pareto critical points of problem (5.1)

Provide $F = \big(f_1, f_2, \ldots, f_r\big)$, where each f_i is bounded below and continuously differentiable
Provide $\Omega \subset \mathbb{R}^n$, where Ω is a nonempty compact convex subset of $\mathbb{R}^n$
Choose arbitrarily $\sigma \in (0,1)$, $\rho \in [0,1)$, and $\Gamma_1, \Gamma_2 \in (0,1)$ with $\Gamma_1 < \Gamma_2$
Provide $\mathcal{N}$, the number of initial points to be randomly chosen
Provide the tolerance level $\epsilon > 0$ for the optimum solution of the problem (5.1)
Set Pareto set $\mathcal{S} \leftarrow \emptyset$
for $n = 1 : \mathcal{N}$ **do**
  Choose a random point $x^0 \in \Omega$
  Set $C^0 \leftarrow F(x^0)$, $m_0 \leftarrow 1$, $k \leftarrow 0$
  Compute an optimal solution $s(x^k)$ and an optimal value $\mathcal{M}(x^k)$ by

$$s(x^k) \in \underset{v \in \Omega}{\operatorname{argmin}} \max_{i \in \mathcal{I}} \langle \nabla f_i(x^k), v - x^k \rangle \quad \text{and}$$
$$\mathcal{M}(x^k) = \max_{i \in \mathcal{I}} \langle \nabla f_i(x^k), s(x^k) - x^k \rangle$$

  Compute the search direction $d(x^k) \leftarrow s(x^k) - x^k$
  while $|\mathcal{M}(x^k)| > \epsilon$ **do**
    Set $\alpha_{k_0} \leftarrow 1$ and initialize $\jmath \leftarrow 0$
    while $F\big(x^k + \alpha_{k_\jmath}(s(x^k) - x^k)\big) \succ C^k + \sigma \alpha_{k_\jmath} \mathcal{M}(x^k) e$ **do**
      $\alpha_{k_{\jmath+1}} \in [\Gamma_1 \alpha_{k_\jmath}, \Gamma_2 \alpha_{k_\jmath}]$
      Set $\jmath \leftarrow \jmath + 1$
    end while
    Output $\alpha_{k_\jmath}$
    Set $\alpha_k \leftarrow \alpha_{k_\jmath}$
    $x^{k+1} \leftarrow x^k + \alpha_k d(x^k)$
    Update m_{k+1} and C^{k+1} as follows:

$$m_{k+1} \leftarrow \rho m_k + 1 \quad \text{and} \quad C^{k+1} \leftarrow \frac{\rho m_k}{m_{k+1}} C^k + \frac{1}{m_{k+1}} F(x^{k+1}) \tag{5.7}$$

    Update $k \leftarrow k + 1$
    Compute $s(x)$ and $\mathcal{M}(x)$ at $x = x^k$ using (5.4) and (5.5), respectively
    Set $d(x^k) \leftarrow s(x^k) - x^k$
  end while
  return $\bar{x} = x^k$ as a Pareto critical point
  Update set $\mathcal{S} \leftarrow \mathcal{S} \bigcup \{F(\bar{x})\}$
end for
return $\mathcal{S}$ as a discrete approximation representing the complete Pareto set of the problem (5.1)

In average-type nonmonotone line searches, C^{k+1} is a convex combination of C^k and $F(x^{k+1})$. Since $C^0 = F(x^0)$, C^k acts as a convex combination of all the function values $F(x^0), F(x^1), F(x^2), \ldots, F(x^k)$. The parameter ρ governs the level of nonmonotonicity in C^k. When $\rho = 0$, C^k simplifies to $F(x^k)$, resulting in the nonmonotone line search reverting to the monotone variant.

It is important to note that Algorithm 7 concludes either when reaching a Pareto critical point or generating an infinite sequence x^k of non Pareto critical points. We proceed further with the assumption that the algorithm iterates infinitely. We now recall an important property of the vector C_k, defined in (5.7). The following result shows that for each i, $f_i(x^k)$ and $S_i^k = \frac{1}{k+1}\sum_{j=0}^{k} f_i(x^j)$ represent lower and upper bounds of C_i^k, respectively.

Lemma 5.3 *Let $x^k \in \Omega$ be a non Pareto critical point for $k = 0, 1, 2, \ldots$. In each iteration of Algorithm 7, we obtain $F(x^k) \preceq C^k \preceq S^k$, where $S^k = \frac{1}{k+1}\sum_{j=0}^{k} F(x^j)$.*

Proof: Let us first define a function $\mathcal{B}^k : \mathbb{R} \to \mathbb{R}^r$ by

$$\mathcal{B}^k(t) = \frac{1}{t+1}\big(tC^{k-1} + F(x^k)\big).$$

The derivative of the above function is expressed as follows:

$$\frac{d\mathcal{B}^k(t)}{dt} = \frac{1}{(t+1)^2}\big(C^{k-1} - F(x^k)\big).$$

As x^k is a non Pareto critical point, we have $\mathcal{M}(x^{k-1}) < 0$. It holds due to the nonmonotone Armijo condition (5.6) that $F(x^k) \preceq C^{k-1}$. This implies $\frac{d\mathcal{B}^k(t)}{dt} \succeq 0$ for all $t \neq -1$. Thus, $\mathcal{B}^k$ is nondecreasing for all $t \geq 0$. Using the facts that $\rho \in [0,1]$ and $m_0 = 1$, we obtain $m_k \geq 1$ for all $k = 0, 1, 2, \ldots$. As a result, $\rho m_{k-1} \geq 0$ holds, leading to the following outcome

$$F(x^k) = \mathcal{B}^k(0) \preceq \mathcal{B}^k(\rho m_{k-1}) = C^k.$$

Now, we show that $C^k \preceq S^k$ using induction. For $k = 0$, the inequality trivially satisfies as $C^0 = S^0 = F(x^0)$. Let $C^j \preceq S^j$, for all j with $0 \leq j \leq k-1$. As $\rho \in [0,1]$ and $m_0 = 1$, from (5.7) we can write

$$m_k = 1 + \sum_{\mathcal{J}=1}^{k} \rho^{\mathcal{J}} \leq 1 + k, \quad \forall\, k > 0.$$

Thus, $0 \leq m_k - 1 \leq k$ will follow. As $\mathcal{B}^k$ is nondecreasing for all $t \geq 0$ and $m_k = \rho m_{k-1} + 1$, we get

$$C^k = \mathcal{B}^k(\rho m_{k-1}) = \mathcal{B}^k(m_k - 1) \preceq \mathcal{B}^k(k).$$

By means of the induction process, we additionally obtain

$$\mathcal{B}^k(k) = \frac{1}{(k+1)}\big(kC^{k-1} + F(x^k)\big) \preceq \frac{1}{(k+1)}\big(kS^{k-1} + F(x^k)\big) = S^k,$$

and the conclusion follows. □

The subsequent proposition demonstrates the well-defined nature of Algorithm 7. It asserts that a step size can always be found, satisfying the average-type nonmonotone Armijo-like condition, which allows for the generation of iterates x^k.

Proposition 5.2 *Let $\sigma \in (0,1)$, $x^k \in \Omega$ be a non Pareto critical point generated by Algorithm 7 and $s(x^k)$ and $\mathcal{M}(x^k)$ be computed using (5.4) and (5.5), respectively. Then, there exist $\bar{\alpha} \in (0,1]$ for which*

$$F(x^k + \alpha(s(x^k) - x^k)) \preceq C_k + \sigma\alpha\mathcal{M}(x^k)e, \qquad \forall \alpha \in (0, \bar{\alpha}].$$

Proof: As x^k is a non Pareto critical point, we have $\mathcal{M}(x^k) < 0$. Moreover, $\mathcal{M}(x^k) < \sigma\mathcal{M}(x^k)$ because $\sigma \in (0,1)$. Now, since H is differentiable, we obtain

$$f_i(x^k + \alpha(s(x^k) - x^k)) = f_i(x^k) + \langle \nabla f_i(x^k), \alpha(s(x^k) - x^k)\rangle + o(\alpha), \quad \text{for all } i \in \mathcal{I}. \tag{5.8}$$

Since $\langle \nabla f_i(x^k), s(x^k) - x^k \rangle \leq \mathcal{M}(x^k) < \sigma \mathcal{M}(x^k) < 0$, $\forall i \in \mathcal{I}$, thus there exists $\bar{\alpha} \in (0,1)$ such that for all $i \in \mathcal{I}$, (5.8) can be rewritten as

$$f_i(x^k + \alpha(s(x^k) - x^k)) \leq f_i(x^k) + \sigma\alpha\mathcal{M}(x^k) \quad \text{for all } \alpha \in (0, \bar{\alpha}].$$

Since $h_i(x^k) \leq C_i^k$ for all $i \in \mathcal{I}$ from Lemma 5.3, the above inequality results

$$f_i(x^k + \alpha(s(x^k) - x^k)) \leq C_i^k + \sigma\alpha\mathcal{M}(x^k) \quad \text{for all } \alpha \in (0, \bar{\alpha}].$$

Thus, the result follows. □

5.6 Global Convergence

In this section, we demonstrate global convergence of Algorithm 7. Initially, we establish an auxiliary result for global convergence. In the subsequent lemma, we demonstrate a characteristic of the sequence $\{C^k\}$, which is produced by the average-type nonmonotone line search procedure.

Lemma 5.4 *Let $\sigma \in (0,1)$ and $\{x^k\}$ be a sequence obtained by Algorithm 7. Then, for each $i \in \mathcal{I}$, $\{C_i^k\}$ is a non increasing sequence and convergent.*

Proof: By (5.6) and (5.7), for each $i \in \mathcal{I}$ and for all $k = 0, 1, 2, \ldots$, we obtain

$$\begin{aligned} C_i^{k+1} &= \frac{\rho m_k}{m_{k+1}} C_i^k + \frac{1}{m_{k+1}} f_i(x^{k+1}) \leq \frac{\rho m_k}{m_{k+1}} C_i^k + \frac{1}{m_{k+1}} [C_i^k + \sigma\alpha_k \mathcal{M}(x^k)] \\ &= C_i^k + \frac{\sigma\alpha_k}{m_{k+1}} \mathcal{M}(x^k) \leq C_i^k, \end{aligned} \tag{5.9}$$

where the aforementioned inequality is justified by the fact that $\mathcal{M}(x^k) < 0$. Therefore, the sequence $\{C_i^k\}$ is nonincreasing. Since f_i is bounded from below, and Lemma 5.3 gives $f_i(x^k) \leq C_i^k$ for all i and k, we can conclude that $\{C_i^k\}$ converges for each $i \in \mathcal{I}$. □

We proceed to demonstrate the global convergence of the sequence $\{x^k\}$ produced by Algorithm 7. Specifically, we demonstrate that every limit point of $\{x^k\}$ is a Pareto critical point. Proving the global convergence of an optimization algorithm is crucial because it ensures that the algorithm is guaranteed to find the optimal solution, regardless of the starting point. This is mathematically significant as it underscores the reliability and effectiveness of the algorithm in solving a wide range of practical optimization problems.

It should be noted that the presence of limit points is not assured. Nevertheless, we emphasize that the boundedness of the set level $\{x \in \Omega : F(x) \preceq F(x^0)\}$ serves as a sufficient condition to ensure the presence of limit points.

Theorem 5.1 *Consider the sequence $\{x^k\}$ obtained by Algorithm 7. If $\hat{x}$ is a limit point of $\{x^k\}$, it follows that $\hat{x}$ is a Pareto critical point.*

Proof: Let $K = \{k_0, k_1, k_2, \dots\} \underset{\infty}{\subset} \mathbb{N}$ be a sequence of indices such that $\lim_{j\to\infty} x^{k_j} = \hat{x}$. Given that $k_{j+1} \geq k_j + 1$, taking into account the insights provided by Lemmas 5.3 and 5.4 and the inequalities in (5.9), we obtain, for all $j \in \mathbb{N}$,

$$f_i(x^{k_{j+1}}) \leq C_i^{k_{j+1}} \leq C_i^{k_j+1} \leq C_i^{k_j} + \frac{\sigma \alpha_{k_j}}{m_{k_j+1}} \mathcal{M}(x^{k_j}) \leq C_i^{k_j}, \quad \forall\, i \in \mathcal{I}. \tag{5.10}$$

By the continuity assumption on the objective function, we have

$$\lim_{j\to\infty} F(x^{k_j}) = F(\hat{x}).$$

This observation implies that $\{F(x^{k_j})\}$ is bounded, which in turn leads us to conclude, based on (5.10), that $\{C_i^{k_j}\}$ is a bounded monotone sequence for each $i \in \mathcal{I}$. Consequently, each $\{C_i^{k_j}\}$ possesses a limit.

From (5.10), we have

$$C_i^{k_{j+1}} \leq C_i^{k_j} + \frac{\sigma \alpha_{k_j}}{m_{k_j+1}} \mathcal{M}(x^{k_j}), \quad \forall\, i \in \mathcal{I}. \tag{5.11}$$

By (5.11), we can write

$$0 \leq \lim_{j\to\infty} \frac{-\sigma \alpha_{k_j}}{m_{k_j+1}} \mathcal{M}(x^{k_j}) \leq \lim_{j\to\infty} C_i^{k_j} - C_i^{k_{j+1}} = 0. \tag{5.12}$$

Hence,

$$\lim_{j\to\infty} \left(\frac{\alpha_{k_j}}{m_{k_j+1}}\right) \mathcal{M}(x^{k_j}) = 0.$$

We claim that

$$\lim_{j\to\infty} \alpha_{k_j} \mathcal{M}(x^{k_j}) = 0. \tag{5.13}$$

If $\rho = 0$, then $m_{k_j+1} = 1$ for all $j \in \mathbb{N}$, and (5.13) holds as a trivial case. However, assuming that $\rho \in (0,1)$, based on (5.7), we can observe that

$$m_{k_j+1} = 1 + \sum_{\ell=0}^{k_j} \rho^{\ell+1} \leq \sum_{\ell=0}^{\infty} \rho^{\ell} = \frac{1}{1-\rho}. \tag{5.14}$$

Consequently, $\{\frac{1}{m_{k_j+1}}\}$ is bounded from below.

By (5.12) and (5.14), we get

$$0 \leq \lim_{j\to\infty} -(1-\rho)\sigma \alpha_{k_j} \mathcal{M}(x^{k_j}) \leq \lim_{j\to\infty} \frac{-\sigma \alpha_{k_j}}{m_{k_j+1}} \mathcal{M}(x^{k_j}) \leq 0.$$

Thus, (5.13) follows.

Now, by (5.13), there exists a $K' \underset{\infty}{\subset} K$ such that at least one of the two cases holds:

1. $\lim_{k\in K'} \mathcal{M}(x^k) = 0$ or

2. $\lim_{k\in K'} \alpha_k = 0.$

For the above cases, we show that the limit point $\hat{x}$ of the sequence $\{x^k\}$ is a Pareto critical point.

Case 1: Since $\mathcal{M}$ is continuous, $\lim_{k\in K'} \mathcal{M}(x^k) = \mathcal{M}(\hat{x}) = 0$. Thus, from Proposition 5.1 (iii), $\hat{x}$ is Pareto critical.

Case 2: Without loss of generality, let us assume that $\alpha_k < 1$ for all $k \in K'$. Therefore, for all $k \in K'$, by nonmonotone Armijo step size, there exists $\bar{\alpha}_k \in (0, \alpha_k/\Gamma_1]$ such that

$$F(x^k + \bar{\alpha}_k(s(x^k) - x^k)) \not\leq C^k + \sigma\bar{\alpha}_k\mathcal{M}(x^k)e,$$

which implies that there exists at least one $i \in \mathcal{I}$ for which

$$f_i(x^k + \bar{\alpha}_k(s(x^k) - x^k)) > C_i^k + \sigma\bar{\alpha}_k\mathcal{M}(x^k) \geq f_i(x^k) + \sigma\bar{\alpha}_k\mathcal{M}(x^k),$$

where the second inequality is a consequence of the findings in Lemma 5.3.

By employing the mean value theorem, we can ascertain the existence of $\gamma_k \in (0,1)$ for every $k \in K'$ satisfying

$$\langle \nabla f_i(x^k + \gamma_k\bar{\alpha}_k(s(x^k) - x^k)), \bar{\alpha}_k(s(x^k) - x^k)\rangle = f_i(x^k + \bar{\alpha}_k(s(x^k) - x^k)) - h_i(x^k). \tag{5.15}$$

Thus, by (5.5) and (5.15), for all $k \in K'$, we have

$$\langle \nabla f_i(x^k + \gamma_k\bar{\alpha}_k(s(x^k) - x^k)), \bar{\alpha}_k(s(x^k) - x^k)\rangle > \sigma\bar{\alpha}_k\langle \nabla f_i(x^k), s(x^k) - x^k\rangle. \tag{5.16}$$

Recall that Ω is a compact set and $\{s^k\} \subset \Omega$. Thus there exists $\bar{s} \in \Omega$ such that

$$\lim_{k\in K'} s(x^k) = \bar{s}.$$

By dividing both sides of (5.16) by $\bar{\alpha}_k > 0$ and taking limits for $k \in K'$, we get

$$\langle \nabla f_i(\hat{x}), \bar{s} - \hat{x} \rangle \geq \sigma \langle \nabla f_i(\hat{x}), \bar{s} - \hat{x} \rangle.$$

As $\sigma \in (0,1)$, we get

$$\langle \nabla f_i(\hat{x}), \bar{s} - \hat{x} \rangle \geq 0. \tag{5.17}$$

On the other hand, since $\mathcal{M}(x^k) < 0$ for all k, we have $\lim_{k \in K'} \mathcal{M}(x^k) = \max_{i \in \mathcal{I}} \langle \nabla f_i(\hat{x}), \bar{s} - \hat{x} \rangle \leq 0$. Therefore, using (5.17), we conclude that $\lim_{k \in K'} \mathcal{M}(x^k) = 0$. Based on the continuity of $\mathcal{M}$, we can deduce that $\mathcal{M}(\hat{x}) = 0$. Therefore, according to Proposition 5.1(iii), we can infer that $\hat{x}$ is a Pareto critical point, which concludes the proof. □

The following result shows the iteration-complexity bounds for Algorithm 7. Calculating iteration-complexity bounds is crucial for assessing the efficiency of the proposed algorithm, ensuring its practicality, and guiding the choice of the most suitable method for solving specific problems. These bounds, as detailed in the following theorems, provide a quantitative measure to estimate by how many iterations we can certainly reach a Pareto critical point with a given tolerance of error. For the sake of simplicity, we define the following constants:

$$\beta = \min \left\{ \frac{1}{\theta \ \mathrm{diam}(\Omega)}, \frac{2\Gamma_1(1-\sigma)}{L \ \mathrm{diam}(\Omega)^2} \right\}, \quad 0 < \theta = \sup \left\{ \|\nabla h_i(x)\|_2 : i \in \mathcal{I}, x \in \Omega \right\}, \tag{5.18}$$

where $L = \max\{L_i : i \in \mathcal{I}\}$ and $\mathrm{diam}(\Omega) = \max\{\|x - y\|_2 : x, y \in \Omega\}$. Since Ω is a compact set, its diameter is finite.

Lemma 5.5 *Assume that $F = (f_1, f_2, \ldots, f_r)^\top$ fulfills the conditions specified in Definition 5.1. Then, $\beta \ |\mathcal{M}(x^k)| \leq \alpha_k$, for all $k = 0, 1, 2, \ldots$.*

Proof: Since $0 < \alpha_k \leq 1$ for all $k = 0, 1, 2, \ldots$, we get the following two scenarios:

(i) $\alpha_k = 1$ or

(ii) $0 < \alpha_k < 1$.

For the above cases, we show the desired inequality.

Case (i) By (5.5), we get

$$0 < |\mathcal{M}(x^k)| \leq |\langle \nabla f_i(x^k), s(x^k) - x^k \rangle| \leq \|\nabla f_i(x^k)\|_2 \, \|x^k - s(x^k)\|_2,$$

for all $i \in \mathcal{I}$. Thus, using (5.18), for all k we have $\left|\mathcal{M}(x^k)\right| \leq \theta \operatorname{diam}(\Omega)$, or, equivalently, $|\mathcal{M}(x^k)|/(\theta \operatorname{diam}(\Omega)) \leq 1$. Hence, from the definition of β in (5.18), we have $\beta|\mathcal{M}(x^k)| \leq 1 = \alpha_k$.

Case (ii) Utilizing the Armijo step size strategy, there exists $0 < \bar{\alpha}_k \leq \min\{1, \alpha_k/\Gamma_1\}$ along with $\hat{i} \in \mathcal{I}$ such that

$$f_{\hat{i}}(x^k + \bar{\alpha}_k(s(x^k) - x^k)) > C^k_{\hat{i}} + \sigma\bar{\alpha}_k\mathcal{M}(x^k).$$

Alternatively, by employing Lemma 5.1 with the parameters $\alpha = \bar{\alpha}_k$, $x = x^k$, $s = s(x^k)$, and $\mathcal{M} = \mathcal{M}(x^k)$, we obtain

$$f_i(x^k + \bar{\alpha}_k(s(x^k) - x^k)) \leq C^k_i + \mathcal{M}(x^k)\bar{\alpha}_k + \frac{L}{2}\|s(x^k) - x^k\|_2^2 \, \bar{\alpha}_k^2,$$

for all $i \in \mathcal{I}$. Consequently, by combining the preceding inequalities, we obtain

$$\sigma\bar{\alpha}_k\mathcal{M}(x^k) < \mathcal{M}(x^k)\bar{\alpha}_k + \frac{L}{2}\|s(x^k) - x^k\|_2^2 \, \bar{\alpha}_k{}^2,$$

which implies

$$|\mathcal{M}(x^k)|(1-\sigma) < \frac{L}{2}\|s(x^k) - x^k\|_2^2 \, \bar{\alpha}_k \leq \frac{L\operatorname{diam}(\Omega)^2}{2\Gamma_1}\alpha_k. \tag{5.19}$$

Thus, using (5.18) and (5.19), we get

$$\beta\,|\mathcal{M}(x^k)| \leq \beta\,\frac{L\mathrm{diam}(\Omega)^2}{2\Gamma_1(1-\sigma)}\alpha_k \leq \alpha_k \quad \text{for all} \quad k = 0, 1, \ldots .$$

This completes the proof. □

Theorem 5.2 *Assume that* $F = (f_1, f_2, \ldots, f_r)^\top$ *satisfies Definition 5.1. Then, the following holds:*

(i) $\lim\limits_{k\to+\infty} \mathcal{M}(x^k) = 0$*;*

(ii) for every $\mathcal{N} \in \mathbb{N}$*, we have*

$$\min\left\{|\mathcal{M}(x^k)| : k = 0, 1, \ldots, \mathcal{N}-1\right\} \leq \sqrt{\frac{f_i(x^0) - f_i^{\inf}}{(1-\rho)\sigma\beta\mathcal{N}}} \quad \textit{for all } i \in \mathcal{I},$$

where $f_i^{\inf} = \inf\{f_i(x^k) : k = 0, 1, 2, \ldots\}$.

Proof: By (5.7) and the inequality (5.6), we have

$$C^{k+1} = \frac{\rho m_k}{m_{k+1}}C^k + \frac{1}{m_{k+1}}F(x^{k+1}) \preceq \frac{\rho m_k}{m_{k+1}}C^k + \frac{1}{m_{k+1}}(C^k + \sigma\alpha_k\mathcal{M}(x^k)e) = C^k + \frac{\sigma\alpha_k}{m_{k+1}}\mathcal{M}(x^k)e.$$

Taking into account the fact that $\mathcal{M}(x^k) < 0$, the above inequality implies that

$$\frac{\sigma\alpha_k}{m_{k+1}}|M(x^k)|e \preceq C^k - C^{k+1}. \tag{5.20}$$

Thus, by Lemma 5.5 and (5.20), we get

$$0 \prec \frac{\sigma\beta}{m_{k+1}}(\mathcal{M}(x^k))^2 e \preceq C^k - C^{k+1}. \tag{5.21}$$

Therefore, by Lemma 5.4, we have

$$\lim_{k\to\infty}\frac{\sigma\beta}{m_{k+1}}(\mathcal{M}(x^k))^2 = 0. \tag{5.22}$$

If $\rho = 0$, then $m_{k+1} = 1$ for all $k \in \mathbb{N}$, and item (i) trivially holds. Assuming that $\rho \in (0,1)$, by (5.7), we have

$$m_{k+1} = 1 + \sum_{\ell=0}^{k} \rho^{\ell+1} \leq \sum_{\ell=0}^{\infty} \rho^{\ell} = \frac{1}{1-\rho}.$$

Therefore, $\{\frac{1}{m_{k+1}}\}$ is bounded from below, and by (5.22) we get

$$\lim_{k \to +\infty} (\mathcal{M}(x^k))^2 = 0,$$

i.e., item (i) holds.

To demonstrate item (ii), we sum both sides of the second inequality in (5.21), for $k = 0, 1, 2, \ldots, \mathcal{N} - 1$, which gives

$$\sum_{k=0}^{\mathcal{N}-1} \frac{\sigma\beta}{m_{k+1}} (\mathcal{M}(x^k))^2 \leq C_i^0 - C_i^{\mathcal{N}} \quad \text{for all } i \in \mathcal{I}.$$

Since $\{\frac{1}{m_{k+1}}\}$ is bounded from below, we get

$$\sum_{k=0}^{\mathcal{N}-1} (\mathcal{M}(x^k))^2 \leq \frac{1}{(1-\rho)\sigma\beta} (C_i^0 - C_i^{\mathcal{N}}).$$

Since $C_i^0 = h_i(x^0)$, and by Lemma 5.3, we can write

$$\sum_{k=0}^{\mathcal{N}-1} (\mathcal{M}(x^k))^2 \leq \frac{1}{(1-\rho)\sigma\beta} (f_i(x^0) - f_i(x^{\mathcal{N}})). \tag{5.23}$$

As we have assumed that $h_i(x)$ is bounded below for each $i \in \mathcal{I}$, we get

$$\min\{(\mathcal{M}(x^k))^2 : k = 0, 1, \ldots, \mathcal{N} - 1\} \leq \frac{1}{(1-\rho)\sigma\beta\mathcal{N}} (f_i(x^0) - f_i^{\inf}).$$

And so

$$\min\left\{|\mathcal{M}(x^k)| : k = 0, 1, \ldots, \mathcal{N} - 1\right\} \leq \sqrt{\frac{f_i(x^0) - f_i^{\inf}}{(1-\rho)\sigma\beta\mathcal{N}}} \quad \text{for all } i \in \mathcal{I},$$

which completes the proof. □

Theorem 5.3 *Assume that $F = (f_1, f_2, \ldots, f_r)^\top$ satisfies Definition 5.1 and $\epsilon > 0$. If for every $\mathcal{N} \in \mathbb{N}$ we define the set $B(\epsilon) = \{k : |\mathcal{M}(x^k)| > \epsilon, k = 0, 1, 2 \ldots \mathcal{N} - 1\}$. Then,*

$$|B(\epsilon)| \leq \frac{(f_i(x^0) - f_i^{\inf})}{(1-\rho)\sigma\beta}\frac{1}{\epsilon^2},$$

where $|B(\epsilon)|$ denotes the number of elements of $B(\epsilon)$.

Proof: By (5.23), we can write

$$\epsilon^2|B(\epsilon)| \leq \sum_{k=0}^{\mathcal{N}-1}(\mathcal{M}(x^k))^2 \leq \frac{1}{(1-\rho)\sigma\beta}(f_i(x^0) - f_i(x^{\mathcal{N}})).$$

As we have assumed that $f_i(x)$ is bounded below for each $i \in \mathcal{I}$, we get

$$|B(\epsilon)| \leq \frac{(f_i(x^0) - h_i^{\inf})}{(1-\rho)\sigma\beta}\frac{1}{\epsilon^2}.$$

□

Theorem 5.4 *Suppose that $F = (f_1, f_2, \ldots, f_r)^\top$ satisfies Definition 5.1 and $\epsilon > 0$. Let us consider an iteration k, where $F(x^k)$ is known. If $|\mathcal{M}(x^k)| > \epsilon$, then the nonmonotone Armijo-like line search requires at most $1 + \left\lceil \frac{\ln(\beta\epsilon)}{\ln(\Gamma_2)} \right\rceil$ evaluations of F to evaluate the step size α_k.*

Proof: In Algorithm 7, $\mathfrak{T}_k$ represents the count of inner iterations, while $v(k)$ denotes the number of evaluations performed on F to compute α_k. Then, by Algorithm 7, we have

$$v(k) = \mathfrak{T}_k + 1 \quad \text{and} \quad \alpha_k \leq \Gamma_2^{\mathfrak{T}_k}\alpha_{k_0} = \Gamma_2^{\mathfrak{T}_k} \quad \text{since } \alpha_{k_0} = 1.$$

Hence using Lemma 5.5 and the above inequality, it follows that $\beta|\mathcal{M}(x^k)| \leq \Gamma_2^{\mathfrak{T}_k}$. Since $|\mathcal{M}(x^k)| \geq \epsilon$, we have $\beta\epsilon \leq \Gamma_2^{\mathfrak{T}_k}$.

Therefore, due to the fact $0 < \Gamma_2 < 1$, we get $\mathfrak{T}_k \leq \frac{\ln(\beta\epsilon)}{\ln(\Gamma_2)}$, which completes the proof. □

In the following theorem, we establish the convergence of the proposed method, considering the convexity assumption. Combining Lemma 5.2 and Theorem 5.1, we obtain the following result.

Theorem 5.5 *Let F be convex on Ω and $\{x_k\}$ be a sequence produced by Algorithm 7. If $\hat{x}$ is a limit point of $\{x_k\}$, it follows that $\hat{x}$ is a weakly Pareto optimal point of the problem* (5.1).

5.7 Numerical Experiments

This section shows the effectiveness and ability to generate Pareto frontiers properly for some benchmark multiobjective test problems. We perform our experiments using MATLAB software (version R2018b). Regarding the algorithmic parameters, we choose $\sigma = 10^{-4}$, $\rho = 0.85$, $\Gamma_1 = 0.3$, $\Gamma_2 = 0.5$. For each starting point, the iteration stops at x^k, satisfying $|\mathcal{M}(x^k)| \leq \epsilon$, where $\epsilon = 10^{-6}$.

In Table 5.1, we have listed 32 multiobjective test problems, which include convex and nonconvex problems. We consider the constrained set Ω as a box-type constraint, i.e., $\Omega = \{x \in \mathbb{R}^n : lb \leq x \leq ub\}$. Table 5.1 contains the major features of the test problems. Columns of Table 5.1 are:

(i) Problem: the name of the test problem,

(ii) Source: source references of the test problem,

(iii) n: number of decision variables in the test problem,

(iv) r: count of objectives in the test problem,

(v) Convex: whether the problem is convex or not–Y for convex and N for nonconvex,

(vi) $lb^\top$: lower bound vector of the decision variable $x \in \mathbb{R}^n$, and

(vii) $ub^\top$: upper bound vector of the decision variable $x \in \mathbb{R}^n$.

To evaluate the proposed technique's performance, a comparative analysis is conducted against alternative methods for MOPs. For comparison, we have considered the following methods:

(i) HZ: the Hager Zhang method [63] under the standard Wolfe line search condition,

(ii) SD: the steepest descent method [49] under the standard Wolfe line search condition,

(iii) NMCG: the conditional gradient method under the average-type nonmonotone line search condition, i.e., Algorithm 7,

(iv) CGArm: the conditional gradient method [19] under the Armijo line search, and

(v) CGNon: the conditional gradient method [19] under a nonmonotone line search.

For each method, a total of 100 iterations were performed to solve all test problems. The starting points for these iterations were randomly selected from the box of lower and upper bounds as shown in columns "$lb^\top$" and "$ub^\top$", respectively. Table 5.2 shows the comparison of the proposed technique with HZ and SD methods for MOPs. The columns in Table 5.2 can be interpreted as follows:

(i) Problem: the name of test problem,

(ii) NI: average count of iterations,

(iii) NF: average count of function evaluations,

(iv) NG: average count of gradient evaluations, and

(v) NT: the total count of average function evaluations and gradient evaluations (see (2.39)).

During the performance evaluation, we compare the HZ and SD methods with the NMCG method in the following manner: for each i^{th} test problem, we determine the total count of average function evaluations and gradient evaluations by the j^{th} solver using the equation (2.39) with $m = 5$. The value is denoted as $NT(i, j)$. Subsequently, we calculate the ratio $R(i, j) = \frac{NT(i,j)}{NT(i,NMCG)}$.
The geometric mean of these ratios for j^{th} solver across all the test problems are computed as follows:

$$R(j) = \left(\prod_{i \in P} R(i, j) \right)^{\frac{1}{|P|}}, \tag{5.24}$$

which is also known as *relative efficiency*, where the set of the test problems under consideration is denoted as P, and the cardinality of P is denoted by $|P|$.

Problem	n	r	Convex	$lb^\top$	$ub^\top$	Source
FON	2	2	N	$(-4,-4,\ldots,-4)$	$(4,4,\ldots,4)$	[164]
Hill	2	2	N	$(0,0)$	$(1,1)$	[74]
SK2	4	2	N	$(-10,-10,\ldots,-10)$	$(10,10,\ldots,10)$	[75]
PNR	2	2	Y	$(-1,-1)$	$(1,1)$	[126]
MOP3	2	2	N	$(-\pi,-\pi)$	(π,π)	[75]
KW2	2	2	N	$(-3,-3)$	$(3,3)$	[87]
FAR1	2	2	N	$(-1,-1)$	$(1,1)$	[98]
FF1	2	2	N	$(-1,-1)$	$(1,1)$	[75]
VU1	2	2	N	$(-3,-3)$	$(3,3)$	[75]
VU2	2	2	Y	$(-3,-3)$	$(3,3)$	[75]
SP1	2	2	Y	$(-10,-10)$	$(10,10)$	[75]
JOS1	1000	2	Y	$(-10^4,-10^4,\ldots,-10^4)$	$(10^4,10^4,\ldots,10^4)$	[75]
AP1	2	3	Y	$(-100,-100)$	$(100,100)$	[6]
AP3	2	2	N	$(-100,-100)$	$(100,100)$	[6]
AP4	3	3	Y	$(-100,-100,-100)$	$(100,100,100)$	[6]
SLCDT1	2	2	N	$(-5,-5)$	$(5,5)$	[136]
SLCDT2	10	3	Y	$(-100,-100,\ldots,-100)$	$(100,100,\ldots,100)$	[136]
DD1	5	2	N	$(-20,-20,\ldots,-20)$	$(20,20,\ldots,20)$	[32]
COMET	2	3	N	$(1,-2)$	$(3.5,2)$	[75]
VIENNET	2	3	N	$(-3,-3)$	$(3,3)$	[75]
DLTZ2	10	3	N	$(0,0,\ldots,0)$	$(1,1,\ldots,1)$	[164]
VFM1	2	3	Y	$(-2,-2)$	$(2,2)$	[75]
IKK1	2	3	Y	$(-50,-50)$	$(50,50)$	[75]
BK1	2	2	Y	$(-5,-5)$	(1010)	[75]
LRS1	2	2	Y	$(-50,-50)$	$(50,50)$	[75]
SSFYY1	2	2	Y	$(-100,-100)$	$(100,100)$	[75]
FDS#1	2000	3	Y	$(-2,-2,\ldots,-2)$	$(2,2,\ldots,2)$	[50]
FDS#2	3000	3	Y	$(-2,-2,\ldots,-2)$	$(2,2,\ldots,2)$	[50]
FDS#3	5000	3	Y	$(-2,-2,\ldots,-2)$	$(2,2,\ldots,2)$	[50]
MMR5#1	2000	2	N	$(-5,-5,\ldots,-5)$	$(5,5,\ldots,5)$	[107]
MMR5#2	3000	2	N	$(-5,-5,\ldots,-5)$	$(5,5,\ldots,5)$	[107]
MMR5#3	5000	2	N	$(-5,-5,\ldots,-5)$	$(5,5,\ldots,5)$	[107]

Table 5.1: List of test problems taken for numerical comparison

	HZ				SD				CGArm				CGNon				NMCG			
Problem	NI	NF	NG	NT	NI	NF	NG	NT	NI	NF	NG	NT	NI	NF	NG	NT	NI	NF	NG	NT
FON	7	54	48	294	8.5	61	49	306	117.6	932.9	899.3	5429.4	65.9	268.1	233.9	1437.6	3.2	6.4	4.2	27.4
HILL	9	62	54	332	10	66	56.5	348	11.5	112	101.8	621	11.7	107.3	88.1	547.8	1.9	3.9	2.9	18.8
SK2	29	93	951	568	38	117.5	119	712.5	19	34.3	101.7	542.8	16.3	29.5	81.5	437	12.4	24.9	13.4	92.3
PNR	3	13.5	13.5	81	11	52	44.5	274.5	22.3	329.5	265.2	1655.5	21.9	280.8	178.9	1175.3	2.6	5.2	3.6	23.4
MOP3	9	35.5	35	210.5	8.5	36.5	33	201.5	34.3	1024.4	1088.2	6465.4	23.4	722.4	638.7	3915.9	12.8	25.7	13.8	95
KW2	12	87.5	75	462.5	11	81.5	70.5	434	762.3	27268.3	25366.7	154101.8	761.8	27198	23516.6	144781	15.2	30.5	16.2	112
FAR1	36	208.5	182.5	1121	34.5	199	174	1069	59.2	869.2	761.5	4676.7	50.5	722.1	685.5	4149.6	30.1	175	157.8	964
FF1	7.5	53	48	293	12	76.5	66	406.5	42.3	560.9	501.2	3066.9	45.2	583.4	522.8	3197.4	11.5	23.1	12.5	86.1
VU1	784	2394	2396	14374	780	2380	2382	14290	253.4	3786.5	3012.3	18848	252.6	3779.6	2967.4	18616.6	103.5	207	106.8	741.2
VU2	6.9	16.2	8.9	60.7	7.5	21	11.3	77.5	2.3	3.9	3.2	19.9	2.28	3.9	2.6	16.9	6.5	13	7.5	50.7
SP1	3	6	8	46	10	20	22	130	100.9	3290.8	2996.1	18271.3	78.7	2185.5	1967.3	12022	6.5	13	9.8	62.3
JOS1	1	2	4	22	1	2	4	22	33.3	271.9	278.4	1663.9	39.6	288.3	302.4	1800.3	1	2	4	22
AP1	10.5	104	85	529	11	104	85	529	221.5	1906.3	1205.3	7932.8	151	1101.8	955.1	5877.3	9.8	19.7	10.8	74
AP3	13.5	129.5	118	719.5	43	263	219	1358	493.5	11495.6	9454.2	58766.6	323.3	9078.9	7032.1	44239.4	10.1	23	17.1	108.5
AP4	11	108	101	613	20	151	142	861	201.5	567	345	2292	195	501.7	289.4	1948.7	9.7	93.7	99	588.7
SLCDT1	4	22.5	22	132.5	4	22.5	22	132.5	6.3	37.5	28	177.5	6.1	33.7	20.3	135.2	3.3	3.3	4.3	24.8
SLCDT2	18	190.5	173.5	1058	20.5	181.5	159.5	979	23	221.7	194.2	1192.7	21	199.5	177.4	1086.5	11	131.3	144.7	854.8
DD1	74.5	230	232	1390	74.5	225.5	227.5	1363	22.7	65.2	53.5	332.7	22	61.3	55.3	337.8	13.9	27.9	17.2	114.3
COMET	74.4	74.4	443.4	2291.4	253.1	257	509.7	2805.5	27.3	33.6	30.4	185.6	28.2	37	35.1	212.5	7.9	15.9	11.2	72.3
VIENNET	127.1	145	762.6	3958	151.1	185.7	303.2	1701.7	996.2	1022.3	1733.4	9689.3	885.1	956.6	1003.2	5972.6	16.6	33.2	17.6	121.4
DLTZ2	41.7	161.5	190.3	1113	87.8	272.6	339	1967.6	101.3	541.8	558	3331.8	77.1	378.2	372.4	2240.2	13.9	31.2	21	136.2
VFM1	5.9	16.2	29	161.2	12.9	27.2	41	232.2	12.9	23.7	44	243.7	7.9	19.2	36	199.2	3.2	6.4	4.2	27.6
IKK1	15.7	33	45.9	262.5	26.4	43	57.2	329	581.3	20878.4	14004.3	90899.9	531.4	19071.9	12278.3	80463.4	3.71	40	4.7	63.6
BK1	6.1	12	8.9	56.5	11.6	27.1	25.3	153.6	11.6	27.4	27.2	163.4	8.7	19.3	21.5	126.8	4.7	9.5	5.7	38.2
LRS1	17.7	25	29.8	174	22.2	31.6	37	216.6	10.3	12	11.9	71.5	9.2	11.2	10.3	62.7	5.4	10.8	6.4	42.8
SSFYY1	22.1	36.5	29	181.5	31	42.9	36.1	223.4	195	103.5	78.2	494.5	167	95.3	67.2	431.3	6.8	13.7	7.8	53.2
FDS#1	33	257	234	1427	69	481	442	2691	233	1057.1	967.4	5894.1	176.3	784.3	667.3	4120.8	18.2	148.4	133.5	815.9
FDS#2	35	262	241	1467	72	499	460	2799	321.5	1305.4	1067.3	6641.9	318	1105.3	1004.3	6126.8	27	191	169	1036
FDS#3	36	271	248.5	1513.5	74	514	473	2879	111.5	667.3	665.2	3993.3	89.3	433.2	410.4	2485.2	31	211	188	1151
MMR5#1	25.5	237.5	218	1327.5	7.5	87	83.5	504.5	21.3	166.3	123.3	782.8	19.3	149.4	115.4	726.4	15.7	140.3	133.6	808.3
MMR5#2	38	370	339	2065	30	298.5	289.5	1746	108	670.4	596.2	3651.4	97.5	398.1	339.6	2096.1	31.2	285	294.4	1757
MMR5#3	25	244	236	1424	15	161	162	971	125.6	344.3	201.4	1351.3	115.7	261.6	187	1196.6	11	141.7	165	966.7

Table 5.2: Performance of HZ, SD, CGArm, CGNon, and NMCG methods on the set of test problems given in Table 5.1

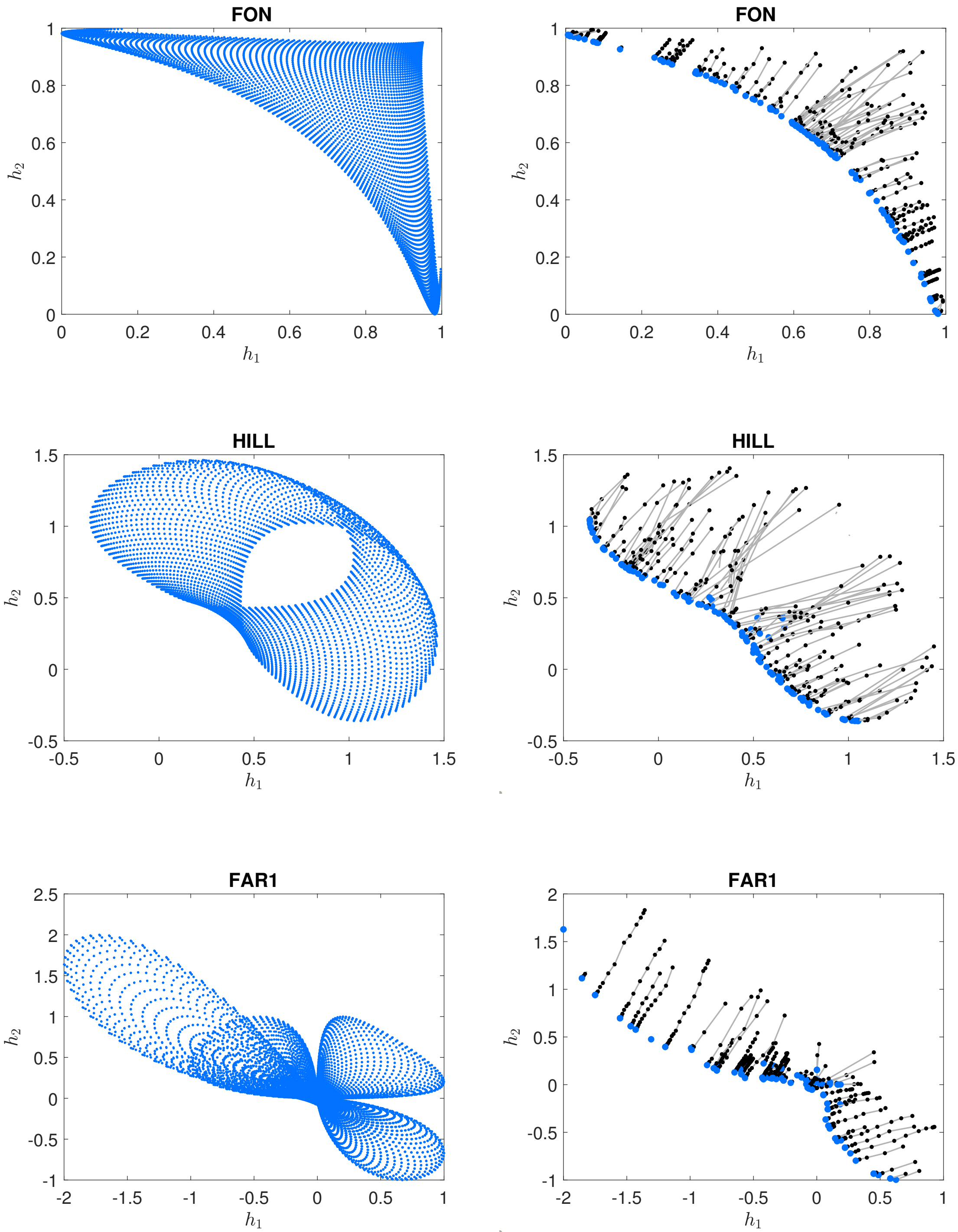

Figure 5.1: (Left column). Objective feasible region of the objective problems FON, HILL FAR1, PNR, MOP3, SLCDT1, FF1, VU2, and BK1. (Right column). For each of the randomly taken 100 initial points, Algorithm 7 moves through black points and finally reaches to a blue point

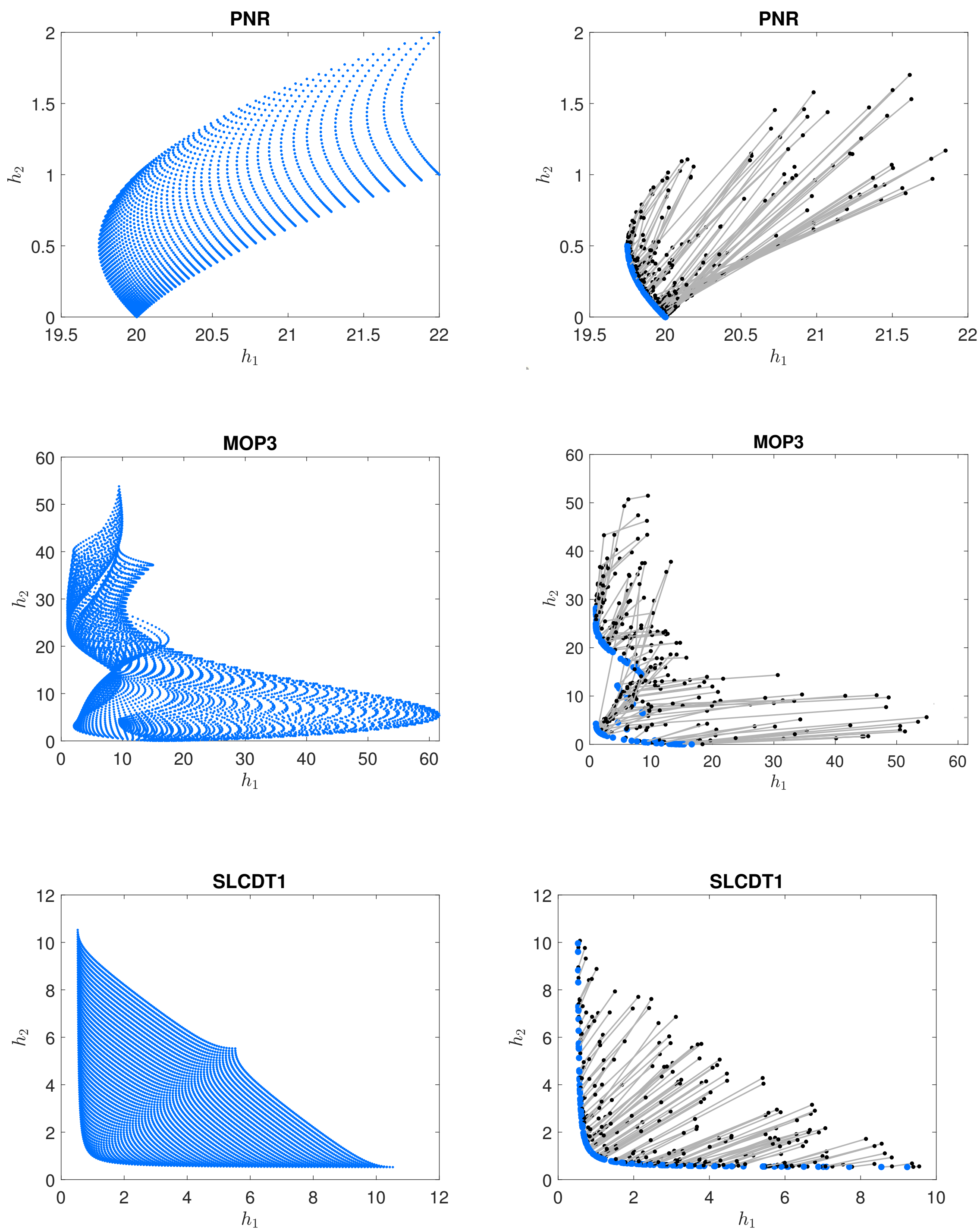

Figure 5.1: continued from Figure 5.1

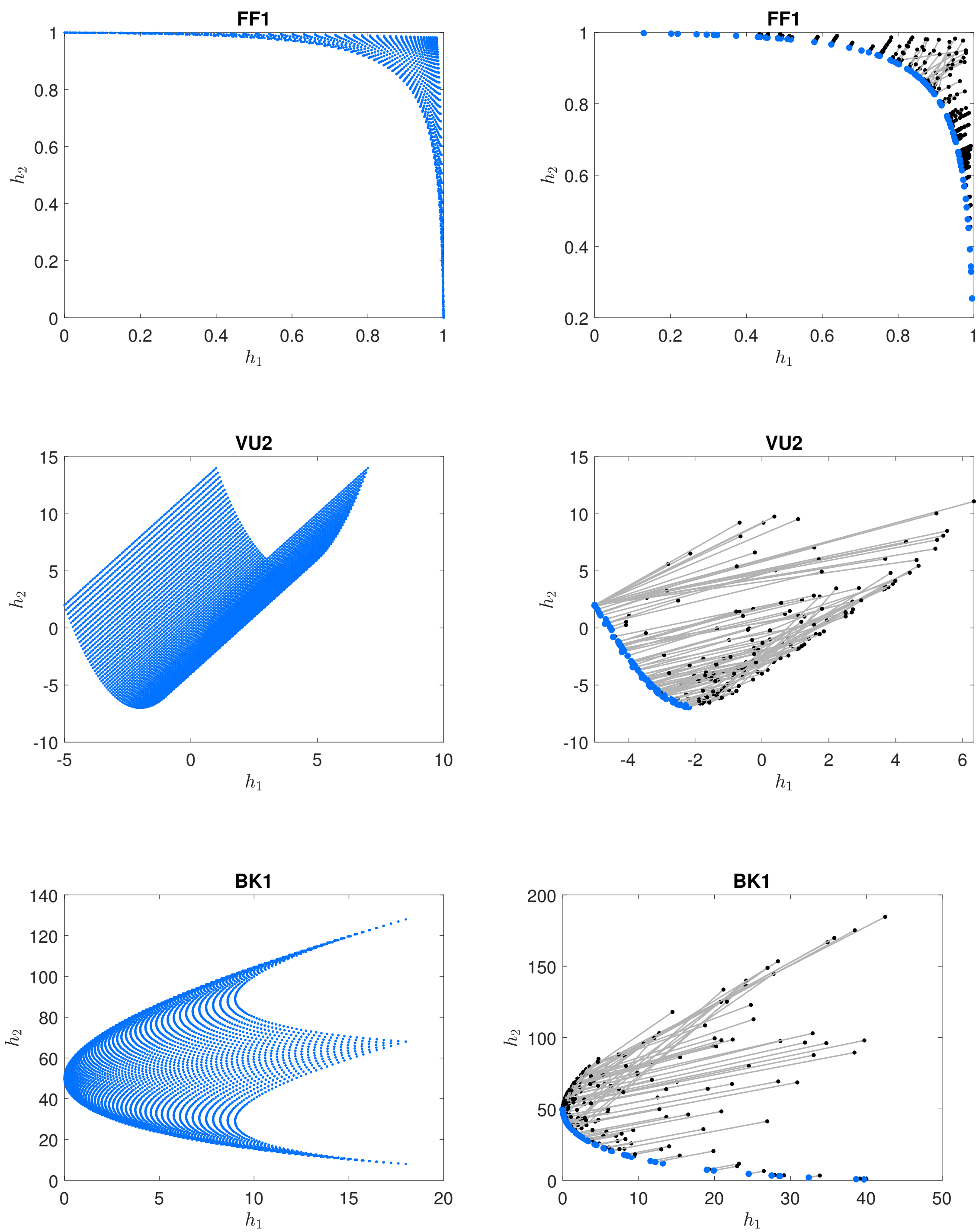

Figure 5.1: continued from Figure 5.1

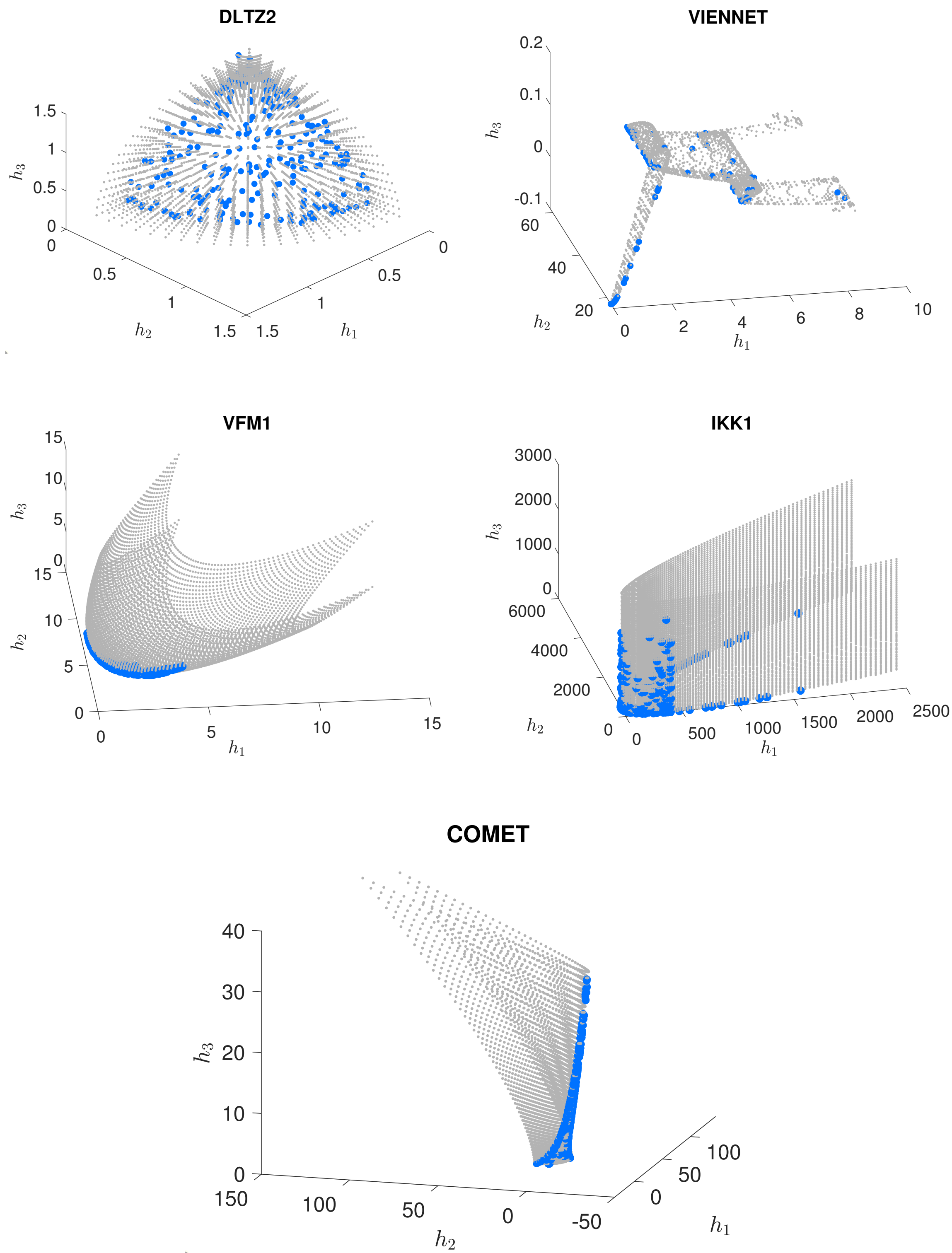

Figure 5.2: Gray points denote the objective feasible region of the tri-objective problems DLTZ2, VIENNET, VFM1, IKK1, and COMET. The blue points represent the final iterates obtained by Algorithm 7

HZ	SD	CGArm	CGNon	NMCG
3.537	4.298	15.473	12.228	1

Table 5.3: Relative efficiency (see (5.24)) of HZ, SD, CGArm, CGNon, and NMCG methods

The comparison between the considered methods is relative and not influenced by a small number of problems that may require a significant amount of function evaluations and gradient evaluations. Referring to the rule stated in (5.24), it is apparent that $R(\text{NMCG}) = 1$. Also, the lesser R, the better the performance. Table 5.3 displays values of $R(\text{HZ})$, $R(\text{SD})$, $R(\text{CGArm})$ and $R(\text{CGNon})$. Table 5.3 indicates that the NMCG method produces the best average performances.

Figures 5.1 and 5.2 display the output of the test results obtained by the proposed method. To check the ability to generate a Pareto front by the proposed method, we have considered fourteen multiobjective test problems of both convex and non-convex types:

(i) Biobjective test problems: FON, HILL FAR1, PNR, MOP3, SLCDT1, FF1, VU2 and BK1.

(ii) Tri-objective test problems: DLTZ2, VIENNET, VFM1, IKK1 and Comet.

For each of the biobjective test problems, there are two graphs. The graphs in the left column depict the image points obtained by discretizing the respective boxes Ω using a finely spaced grid and plotting them. These graphs offer precise depictions of the image spaces of H within Ω and provide a geometric depiction of the Pareto set of F. Figures in the right column of Figure 5.1 were obtained by running the proposed algorithm for each considered problem using 100 times randomly obtained starting points from the corresponding sets Ω. In these graphs, the blue points represent the final iterates while the black points represent the intermediate iterates obtained by Algorithm 7. Figures

5.1 and 5.2 demonstrate that by using an appropriate number of starting points and for the selected set of test problems, the conditional gradient method successfully produces an adequate portrayal of the Pareto front.

Based on the test results reported in Table 5.2, we show the performance profile given by Dolan and Moré [37] to compare the performance of the considered methods. For detailed explanation of the performance profile given by Dolan and Moré [37], see Chapter 2.

Figures 5.3a-5.3d show that the NMCG method outperform the HZ, SD, CGArm, and CGNon methods for all four scenarios, i.e., NI, NF, NG, and NT.

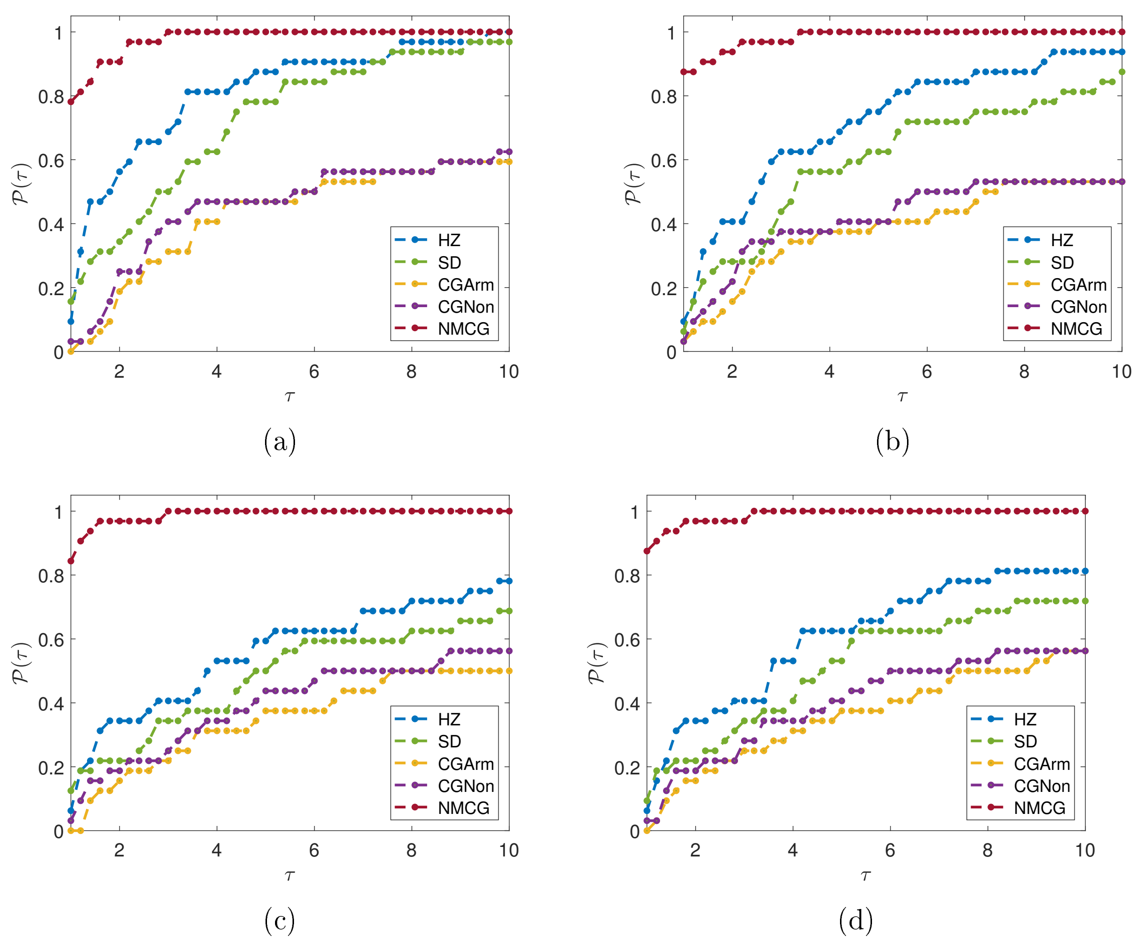

Figure 5.3: Performance profile of HZ, SD, CGArm, CGNon and NMCG measured by (a) NI (b) NF (c) NG, and (d) NT

5.7.1 Performance Metrics

Within this subsection, we conduct a comparison between Algorithm 7 and several heuristic methods applied to the ZDT and DLTZ benchmark suits. The purpose is to evaluate the effectiveness of the proposed algorithm in generating an approximation of the Pareto front. To assess the quality of the solution set produced by Algorithm 7, two performance metrics, namely HV and IGD [9, 141], are employed. The IGD value indicates that lower values are preferable, while for the HV value, higher values are more desirable.

The test problems under consideration are presented in Table 5.4. Table 5.4 displays the parameters m and n, which correspond to the count of objectives and decision variables, respectively. The term "$lb^\top$" represents the vector of lower bounds for the decision variable $x \in \mathbb{R}^n$, while "$ub^\top$" represents the vector of upper bounds for the decision variable $x \in \mathbb{R}^n$. The column "nondominated set type" indicates whether the nondominated set of the test problem is convex or non-convex. The column "nondominated set location" provides information on whether the nondominated solution is priori known or not. The column "Source" provides the source of the test problems.

Problem	m	n	$lb^\top$	$ub^\top$	Nondominated set type	Nondominated set location	Source
ZDT1	2	10	(0,0)	(1,1)	Convex	Known	[164]
ZDT2	2	10	(0,0)	(1,1)	Non-convex	Known	[164]
ZDT3	2	10	(0,0)	(1,1)	Nonconvex	Known	[164]
ZDT4	2	10	(0,-5)	(1,5)	Convex	Known	[164]
DTLZ1	3	5	(0,0,0)	(1,1,1)	Convex	Known	[164]
DTLZ2	3	5	(0,0,0)	(1,1,1)	Nonconvex	Known	[164]
DTLZ3	3	5	(0,0,0)	(1,1,1)	Nonconvex	Known	[164]
DTLZ5	3	5	(0,0,0)	(1,1,1)	Nonconvex	Known	[60, 61]

Table 5.4: The data used for evaluating the performance of Algorithm 7 on the test problems

Table 5.5 presents the median values of the IGD for the DLTZ and ZDT test suites (listed in Table 5.4) obtained from Algorithm 7 and other existing efficient solvers. Similarly, Table 5.6 displays the median values of HV for the obtained solution sets

Problem	MOEA/D (TE)	MOEA/D (WS)	NSGA-II	MOEA/D (PBI)	$pa\lambda$-MOEA/D	NMCG
DTLZ1	$6.93E-4$	$4.03E-3$	$7.91E-4$	$4.21E-4$	$4.52E-4$	$2.04E-6$
DTLZ2	$7.42E-4$	$5.35E-3$	$7.58E-4$	$6.15E-4$	$5.78E-4$	$1.39E-4$
DTLZ3	$1.24E-3$	$1.42E-2$	$3.59E-3$	$1.90E-3$	$1.18E-3$	$1.01E-3$
DTLZ5	$5.22E-5$	$1.38E-3$	$1.89E-5$	$1.11E-4$	$2.79E-5$	$2.71E-6$
ZDT1	$6.84E-4$	$5.42E-4$	$7.94E-4$	$1.14E-3$	$5.80E-4$	$1.02E-5$
ZDT2	$5.84E-4$	$1.30E-2$	$8.16E-4$	$7.03E-4$	$6.02E-4$	$2.47E-6$
ZDT3	$2.01E-3$	$4.93E-3$	$1.20E-3$	$2.06E-3$	$1.97E-3$	$1.04E-3$
ZDT4	$6.54E-4$	$7.02E-3$	$8.14E-4$	$7.86E-4$	$5.94E-4$	$1.34E-6$

Table 5.5: Median values of IGD obtained by different algorithms are computed for ZDT and DTLZ benchmark suite

Problem	MOEA/D (TE)	MOEA/D (WS)	NSGA-II	MOEA/D (PBI)	$pa\lambda$-MOEA/D	NMCG
DTLZ1	0.7434	0.2259	0.7262	0.7835	0.7814	0.8987
DTLZ2	0.3777	0.0000	0.3766	0.3812	0.4074	0.5117
DTLZ3	0.3605	0.0000	0.1901	0.2633	0.3817	0.3922
DTLZ5	0.0894	0.0000	0.0930	0.0779	0.0916	0.2023
ZDT1	0.6392	0.6521	0.6381	0.6057	0.6412	0.6721
ZDT2	0.3097	0.0000	0.3060	0.2957	0.3107	0.3129
ZDT3	0.4807	0.4863	0.5066	0.4642	0.4873	0.5217
ZDT4	0.6360	0.3534	0.6359	0.6244	0.6391	0.6398

Table 5.6: Median values of HV obtained by different algorithms are computed for ZDT and DTLZ benchmark suite

through Algorithm 7 and the other solvers. The median values of IGD and HV for MOEA/D(TE), MOEA/D(WS), NSGA-II, MOEA/D(PBI), and $pa\lambda$-MOEA/D on the ZDT and DTLZ problems are obtained from [142]. As reported in [142], the stopping criteria for these heuristic methods are as follows: the maximum number of function evaluations is 25,000 for the ZDT problem and 50,000 for the DTLZ problem.

It is evident from Table 5.5 that NMCG achieves the lowest values across all the test problems. Table 5.6 demonstrates that NMCG attains the highest values across all the test problems. Hence, the NMCG method surpasses the performance of the existing efficient heuristic methods for the DTLZ and ZDT benchmark suites.

NMCG method often surpasses the performance of heuristic methods in terms of

IGD and HV values due to several reasons:

- NMCG method has well-defined convergence criteria, allowing users to control the quality of the solution obtained. Heuristic methods may lack such clear convergence criteria and may require a significant amount of tuning to achieve competitive IGD and HV values.

- NMCG method provides a consistent set of solutions for evaluation, making it easier to compare results across different instances and datasets, which is important when comparing IGD and HV values.

- NMCG method produces consistent results given the same input, while heuristics may provide different solutions in different runs. This consistency can lead to a more reliable evaluation of IGD and HV.

The performance of the proposed method is further evaluated using relative efficiency, as indicated by the results presented in Table 5.5. To compare the efficiency of the NMCG method with other solvers (MOEA/D(TE), MOEA/D(WS), NSGA-II, MOEA/D(PBI), and $pa\lambda$-MOEA/D), we compare the median of the IGD values (MIGD) for each test problem using each solver. Subsequently, we compute the ratio $R(i,j)$ as follows:

$$R(i,j) = \frac{\text{MIGD}(i,j)}{\text{MIGD}(i,\text{NMCG})},$$

where $\text{MIGD}(i,j)$ represents the median IGD value obtained by the j^{th} solver for the i^{th} test problem, and MIGD(i,NMCG) represents the median IGD value obtained by the NMCG method for the i^{th} test problem. The relative efficiency, also known as the geometric mean of these ratios for the j^{th} solver across all test problems, is evaluated using the equation (5.24).

MOEA/D(TE)	MOEA/D(WS)	NSGA-II	MOEA/D(PBI)	$pa\lambda$-MOEA/D	NMCG
29.879	209.959	31.376	35.622	24.443	1

Table 5.7: Relative efficiency of MOEA/D(TE), MOEA/D(WS), NSGA-II, MOEA/D(PBI), $pa\lambda$-MOEA/D and NMCG methods based on the median values of IGD

According to the rule given in (5.24), it is evident that $R(\text{NMCG}) = 1$. Table 5.7 presents the mentioned values of R(MOEA/D(TE)), R(MOEA/D(WS)), R(NSGA-II), R(MOEA/D(PBI)) and R($pa\lambda$-MOEA/D). Table 5.7 indicates that the NMCG method produces the best median of IGD values among all the considered methods.

By considering the median values of HV presented in Table 5.6, we further compute the relative efficiency in relation to the median values of HV. Table 5.8 indicates that the NMCG method produces the best median of HV values among all the considered methods.

MOEA/D(TE)	MOEA/D(WS)	NSGA-II	MOEA/D(PBI)	$pa\lambda$-MOEA/D	NMCG
0.824	0	0.766	0.770	0.848	1

Table 5.8: Relative efficiency of MOEA/D(TE), MOEA/D(WS), NSGA-II, MOEA/D(PBI), $pa\lambda$-MOEA/D and NMCG methods based on the median values of HV

5.8 Conclusion

This chapter has introduced an NMCG algorithm as a tool for MOPs without relying on the assumption of convexity on the objective function. The main innovation of this approach lies in the utilization of an average-type Armijo-like nonmonotone line search rule, deviating from the traditional monotone line search strategies. Theoretical analysis, specifically Theorem 5.1, demonstrates that, under common assumptions, the proposed method generates a sequence of iterates that converge to a Pareto critical point. Complexity bounds for the method are established in Theorems 5.2, 5.3, and

5.4. Theorem 5.5 illustrates that under the convexity assumption on the objective function, a sequence produced by Algorithm 7 converges to a weakly Pareto optimal solution for Problem (5.1).

In order to evaluate the efficacy of the proposed method compared to alternative approaches, a comparative analysis with the HZ, SD, CGArm and CGNon methods is performed. The efficiency of the proposed method is evaluated through empirical analysis, which includes the computation of relative efficiency (see Tables 5.2 and 5.3) and the generation of performance profiles (see Figures 5.3a-5.3d) using the methodology developed by Dolan and Moré [37]. Additionally, the proposed method is benchmarked against popular solvers based on inverted generational distance and hypervolume indicators. The numerical findings point out that the NMCG method surpassed the HZ and SD methods for the multiobjective optimization test problems considered in this study.

In future, we aim to investigate the theoretical properties as well as the experimental performance of a max-type nonmonotone conditional gradient method for addressing constrained vector optimization problems within the framework of a partial order created by a convex, closed, and pointed cone possessing a nonempty interior. As future research, we can also focus on the implementation of average-type and max-type nonmonotone conditional gradient methods for set optimization problems. These methods can provide valuable insights and solutions to complex optimization scenarios involving sets. Implementing these techniques in the context of set optimization problems could enhance the efficiency and effectiveness of optimization algorithms, opening up new avenues for solving challenging real-world problems.

****